

Developing Multivariate Survival Trees with a Proportional Hazards Structure

Feng Gao¹, Amita K. Manatunga², Shande Chen³

¹Washington University, ²Emory University and ³University of North Texas

Abstract: In this paper, a tree-structured method is proposed to extend *Classification and Regression Trees* (CART) algorithm to multivariate survival data, assuming a proportional hazard structure in the whole tree. The method works on the marginal survivor distributions and uses a sandwich estimator of variance to account for the association between survival times. The Wald-test statistics is defined as the splitting rule and the survival trees are developed by maximizing between-node separation. The proposed method intends to classify patients into subgroups with distinctively different prognosis. However, unlike the conventional tree-growing algorithms which work on a subset of data at every partition, the proposed method deals with the whole data set and searches the global optimal split at each partition. The method is applied to a prostate cancer data and its performance is also evaluated by several simulation studies.

Key words: CART, multivariate survival data, survival tree, WLW model.

1. Introduction

In recent years, many efforts have been made in the application of tree-structured methods to the censored data. The thrust of tree-structured methods comes from the desire to extract meaningful subgroups. Although proportional hazard model (Cox 1972) is a very useful tool in quantifying the effects of covariates on survival time, it is usually problematic for prognostic classification. For example, we may create prognostic groups based on “risk score” which is calculated from linear combinations of coefficients and covariate values. However, this approach may be problematic: (a) the “risk score” depends on the correct specification of the model structure, which may be hard to check when many covariates are involved; (b) the regression model becomes difficult to interpret for the purpose of classification when there exist many interaction terms; (c) the resulting risk groups are arbitrary, and (d) sometime a risk group becomes empty because no patients are included in some covariate profiles.

In contrast, tree-structured methods recursively split the covariate space and directly lead to homogeneous subgroups. Tree-structured method was first proposed by Morgon and Sonquist (1963) and became popular due to the monograph *Classification and Regression Tree* (CART) by Breiman *et al* (1984). Since Gordon and Olshen (1985) gave the first adaptation of CART algorithm to censored data, a rich amount of literature has been dedicated to survival data. Davis and Anderson (1989) grew trees by assuming survival times to be exponential within a given node. LeBlanc and Crowley (1992) proposed a method to grow trees via an one-step full log-likelihood estimation. Huang, Chen and Soong (1998) extended Davis and Anderson's method to the time-dependent case, assuming the survival times to be piecewise exponential. The above methods grew trees rewarding within-node homogeneity. Segal (1988) presented a method to grow trees by goodness of split. Instead of maximizing within-node homogeneity, Segal's method maximized between-node separation, using log-rank test statistics as the splitting rule. Bacchetti and Segal (1995) extended Segal's method to survival data with left-truncation and time-dependent covariates. LeBlanc and Crowley (1993) developed an efficient pruning algorithm for trees that maximized between-node separation.

All these methods were proposed for univariate survival data which assumes survival times to be independent. However, this assumption may be violated in multivariate (clustered) survival data. In the ophthalmology studies, for example, it may be inappropriate to assume that the survival patterns of the two eyes from the same subject are independent. There are very few published tree methods regarding correlated survival data. Gao, Manatunga and Chen (2004) and Su and Fan (2004) extended CART algorithm to multivariate censored data in a similar manner. Both of them introduced a gamma distributed frailty to account for the dependence among survival times. The former used Wald-test statistics as the splitting function while the later was based on a likelihood ratio test. Though survival trees developed above were computationally convenient, they suffered a potential drawback that the overall model structure was unclear. When a partition was done, for example, the above methods assume hazards within the two daughter nodes to be conditionally proportional given the unobserved frailty, but the hazards structures for any two nodes from different parents are unclear. In this paper, we propose an algorithm that develops multiple survival trees assuming a proportional hazards structure within the whole tree. The proposed method can be viewed as a hybrid of CART algorithm and WLW model (Wei, Lin and Weissfeld 1989). Instead of fitting a model for formal inference, it automatically develops a set of prognostics groups while assuming a proportional hazards structure. This paper is organized as follows. In section 2, we describe the algorithm for classification. In section 3, we apply the method to a prostate

cancer data. We assess the performance of the method by simulations in section 4. Finally, we conclude with a discussion in section 5.

2. Methodology

The method is proposed for multivariate (clustered) survival data where several items (or units of system) are followed simultaneously and we only consider the data with a fixed cluster size. The simplest clustered data to consider is the pair of survival times observed in two eyes in an ophthalmologic study. Suppose that data includes N clusters (subjects) and each cluster includes J items (eyes). Let $i = 1, 2, \dots, N$ index clusters and $j = 1, 2, \dots, J$ index the items within each cluster. Let $T_{ij} = \min(X_{ij}, C_{ij})$ be the observed time for the j^{th} item in the i^{th} cluster, where X_{ij} is the true failure time and C_{ij} is the true censoring time. Let δ_{ij} be the corresponding survival indicator with $\delta_{ij} = 1$ if $X_{ij} \leq C_{ij}$ and $\delta_{ij} = 0$ otherwise. In this paper we assume that the covariate vector is common to items within the same cluster and let Z_i denote the covariate vector for i^{th} cluster. Then, the data can be denoted as $\{(t_{ij}, \delta_{ij}, z_i), i = 1, 2, \dots, N; j = 1, 2, \dots, J\}$. We further assume that the survival time and censoring time are independent. Given covariate Z , the survival times between clusters are independent and the survival times within the same cluster may be dependent.

The proposed method intends to extract meaningful subgroups via a tree-like algorithm. Similar to the procedure in Gao *et al* (2004), trees are developed to maximize between-node separation as measured by a Wald-test statistic and the partition is performed on a single covariate at a time. A large tree is developed first and Segal's pruning method (1988) is adopted to determine the proper tree size. Finally a marginal Kaplan-Meier survival curve is prepared separately (i.e., for each eye) and thus to summarize the performance of the classification. However, unlike other tree-growing algorithms which work on a subset of data at a partition, the proposed method deals with the whole data set at each split and assumes a global proportional hazard structure in whole tree.

The proposed method models on the marginal distributions of survival times, specifying the hazard function for j^{th} event to be,

$$h_j(t|z_i) = h_{0j}(t)g(z_i), \quad j = 1, 2, \dots, J \quad (2.1)$$

where $h_{0j}(\cdot)$ is the baseline hazard function for j^{th} event and $g(\cdot)$ is an arbitrary non-negative function of covariate vector.

The classification will be performed in a stepwise manner. Let Γ_k denote the k subgroups obtained at k^{th} step. We assume that these subgroups are consisted of two disjoint subsets A and B , where $A = \{A_1, A_2, \dots, A_{k_1}\}$ contains these subgroups permitting further splitting while $B = \{B_1, B_2, \dots, B_{k_2}\}$ includes those

that cannot be further split, with $k = k_1 + k_2$. In terms of non-splitting subgroup, we mean that either the covariate space within the subgroup becomes homogeneous or the sample size (number of events) of the subgroup is less than a pre-specified value. Suppose that, at $(k+1)^{th}$ step, a subgroup $\gamma \in A$ is further split, and let Γ_{k+1} represent the $k+1$ subgroups at current step. For mathematical convenience, we rearrange the subgroups as $\Gamma_{k+1} = \{C_1, C_2, \dots, C_k, C_{k+1}\}$, specifically letting C_1, C_k and C_{k+1} being the reference group and the two newly obtained groups respectively. The reference group will be arbitrarily chosen (say, B_1) from subgroups that do not permit further splitting. Since the splitting statistics (see below) is based on the relative difference of the two new subgroups (C_k and C_{k+1}) rather than their absolute effects, the comparisons among candidate partitions will be valid as long as the two new subgroups have a common reference level.

In order to rate over all possible splits, the following working model is fitted,

$$h_j(t|z_i) = h_{0j}(t)e^{\beta I(z_i)}. \quad (2.2)$$

That is, the function $g(z_i)$ in model (2.1) is approximated by $g(z_i) = e^{\beta I(z_i)} = e^{\sum_{m=1}^{k+1} \beta_m I(z_i \in C_m)}$, where $I(z_i \in C_m)$ is an indicator function with $I(z_i \in C_m) = 1$ if a cluster (subject) falls into subgroup C_m and 0 otherwise. In approximation (2.2), $\beta = \{\beta_1 \equiv 0, \beta_2, \beta_3, \dots, \beta_{k+1}\}$ and the first subgroup is set to be the reference level. The unknown parameters can be estimated by maximizing the following partial likelihood function (Wei, Lin and Weissfeld 1989),

$$L(\beta) = \prod_{i=1}^N \prod_{j=1}^J \left\{ \frac{e^{\beta I(z_i)}}{\sum_{l=1}^N Y_{lj}(t_{ij}) e^{\beta I(z_l)}} \right\}^{\delta_{ij}},$$

where $Y_{lj}(t)$ is an indicator for whether j^{th} item in l^{th} cluster is at risk at time t , with $Y_{lj}(t) = 1$ if $t_{lj} \geq t$ and 0 otherwise. The corresponding score function and the information matrix of β respectively are,

$$U(\beta) = \frac{d \log L(\beta)}{d\beta} = \sum_{i=1}^N \sum_{j=1}^J \delta_{ij} \left\{ I(z_i) - \frac{S_j^{(1)}(\beta, t_{ij})}{S_j^{(0)}(\beta, t_{ij})} \right\},$$

$$A(\beta) = -\frac{d \log^2 L(\beta)}{d\beta^2} = \sum_{i=1}^N \sum_{j=1}^J \delta_{ij} \left\{ \frac{S_j^{(2)}(\beta, t_{ij})}{S_j^{(0)}(\beta, t_{ij})} - \frac{S_j^{(1)}(\beta, t_{ij}) S_j^{(1)}(\beta, t_{ij})'}{S_j^{(0)}(\beta, t_{ij})^2} \right\},$$

where $S_j^{(0)}(\beta, t) = \sum_{l=1}^N Y_{lj}(t) e^{\beta I(z_l)}$, $S_j^{(1)}(\beta, t) = \sum_{l=1}^N Y_{lj}(t) e^{\beta I(z_l)} I(z_l)$, and $S_j^{(2)}(\beta, t) = \sum_{l=1}^N Y_{lj}(t) e^{\beta I(z_l)} I(z_l) I(z_l)'$. At a given step, model (2.2) is essentially a WLW model (Wei, Lin and Weissfeld 1989) with a set of indicators for the

resultant subgroups. The estimator $\hat{\beta}$ will be consistent for β but $A^{-1}(\beta)$ may not provide a valid variance-covariance estimator (Lin and Wei, 1989). Then, a “sandwich” estimator (Lin and Wei, 1989) can be used to account for the within-cluster association,

$$\hat{\Sigma} = A^{-1}(\hat{\beta})B(\hat{\beta})A^{-1}(\hat{\beta}), \text{ where}$$

$$B(\hat{\beta}) = \sum_{i=1}^N \sum_{j=1}^J \sum_{l=1}^J \{W_{ij}(\hat{\beta})\}' \{W_{il}(\hat{\beta})\},$$

$$W_{ij}(\hat{\beta}) = \delta_{ij} \left\{ I(z_i) - \frac{S_j^{(1)}(\hat{\beta}, t_{ij})}{S_j^{(0)}(\hat{\beta}, t_{ij})} \right\} - \sum_{l=1}^N \frac{\delta_{lj} Y_{ij}(t_{lj}) e^{\hat{\beta} I(z_i)}}{S_j^{(0)}(\hat{\beta}, t_{lj})} \left\{ I(z_i) - \frac{S_j^{(1)}(\hat{\beta}, t_{ij})}{S_j^{(0)}(\hat{\beta}, t_{ij})} \right\},$$

and $A(\hat{\beta}) = -\frac{d \log^2 L(\beta)}{d\beta^2}$ is the information matrix of β evaluated at $\hat{\beta}$.

The splitting function to evaluate and compare any candidate partitions is defined as $\Phi = \left\{ \frac{(\hat{\beta}_k - \hat{\beta}_{k+1})^2}{\text{var}(\hat{\beta}_k - \hat{\beta}_{k+1})} \right\}$, the robust Wald-test statistic for testing the hypothesis whether the two new subgroups are different $H_0 : \beta_k = \beta_{k+1}$. We note that, as in all tree-based methods, the models are developed conditionally in that the latter split is always conditional on previous ones. In such a case, the asymptotic distributions of estimated parameters become unclear and a formal statistical inference may be problematic. However, Φ still quantifies the separation between the two new subgroups and can be used to rate over the candidate splits. The best partition is chosen such that Φ is maximized among all candidate splits in subgroups that permit further partition. In the special case where all terminal nodes permitting further splitting (the set B is empty), first we need to randomly pick up one node from set A (A_1 , say) as the reference group to evaluate the splits among all other subgroups. Then, we fix one of the other subgroups as the reference to evaluate all possible splits within node A_1 .

In summary, the proposed method incorporates a tree-like algorithm into the conventional WLW model. At each step all competing splits are evaluated, and the one maximizing the splitting function is chosen and executed. As is conventional practice in tree-based methods, we choose a conservative stopping criterion to avoid missing important data structures. The partition process is applied repeatedly until either all the subgroups can not be further split or $\Phi \leq 0.45$, which corresponds to 50% percentile of an X^2 distribution with 1 degree of freedom. As mentioned previously, however, the asymptotic distributions of estimated parameters become unclear due to the conditionality of tree-based methods. The choosing of 0.45 is rather arbitrary and could be adjusted by readers to avoid a tree with too large size. Usually a large tree will be developed using such a conservative stopping criterion, and then we select the tree with a proper size by adopting Segal’s pruning algorithm (1988) detailed below. We

initially grow a very large tree and identify all possible branches of the tree in a bottom-up approach. Within each branch, we find the maximal splitting statistic that reflects the strength of the link of that branch to the tree. We then rank these maximal statistics and prune off the branch with the smallest value (the branch with the weakest link). Thus, we get the first pruned tree plus its weakest linking statistic. The second tree can be obtained by applying this process to the first pruned tree. Repeating the above steps until the pruned tree contains only the root node, we obtain a sequence of nested trees together with their weakest linking statistics. Finally we plot the weakest linking statistics against the tree sizes. Because a plateau in the curve indicates that the strength of the link does not decrease as the tree size increases, we choose the best tree as the one corresponding to the characteristic “kink” in the curve.

Table 1: Clinical characteristics for the patients with prostate cancer

Characteristics	Number of cases (%)
PrePSA (ng/ml)	537
< 4.0	39 (7.3%)
4.1-10.0	359 (66.9%)
10.1-20.0	108 (20.1%)
> 20.0	31 (5.8%)
Gleason Score	529
2-4	90 (17.0%)
5-6	325 (61.4%)
7	92 (17.4%)
8-10	22 (4.2%)
Clinical Stage	537
T1A-T1C	238 (44.3%)
T2A	156 (29.1%)
T2B	106 (19.7%)
T2C	37 (6.9%)

3. Application: Prostate Cancer Data

Disease-free survival rate is an important measurement in the evaluation of treatment effect for prostate cancer patients. Currently there are two types of definitions for disease freedom in prostate cancer treatment. One specifies a non-detectable level of prostate specific antigen (PSA) and the other, by the American Society of Therapeutic Oncology (ASTRO), defines a non-rising PSA

level. From August 1992 through October 1997, 537 patients with prostate cancer were treated by transperineal I-125 implant and followed up for 3 to 8 years, with a median follow-up of 6 years (Critz 2002). The outcome of interest for this study is the time to recurrence. Two outcomes were defined using follow-up PSA levels. One records the time from treatment to the detection of ≥ 0.2 ng/ml PSA level. The other is the time to 3 consecutive increasing in PSA as measured 6 months apart. At the end of study, 81 patients recurred based on the definition of undetectable PSA and 59 patients recurred according to ASTRO definition. Three clinical characteristics were also collected: pre-treatment PSA (PrePSA), clinical stage, and Gleason Score. All of them were ordinal covariates with 4 categories (Table 1). Eight patients were excluded due to missing values in Gleason Score and these 529 patients constituted the cohort for our study. The original goal of the study was to check the agreement between the two definitions for recurrence. In this paper, we are interested in identifying risk groups for recurrence.

Table 2: Results of marginal model analysis for prostate cancer data

Characteristics	$\hat{\beta}$	$se(\hat{\beta})$	z	p
PrePSA (ng/ml)				
4.1-10.0 vs < 4.0	0.422	0.520	0.572	0.570
10.1-20.0 vs < 4.0	1.635	0.523	2.204	0.027
> 20.0 vs < 4.0	1.596	0.566	2.023	0.043
Gleason Score				
5-6 vs 2-4	0.142	0.273	0.377	0.710
7 vs 2-4	0.776	0.290	2.002	0.045
8-10 vs 2-4	1.017	0.394	1.895	0.058
Clinical Stage				
T2A vs T1A-T1C	0.224	0.237	0.714	0.470
T2B vs T1A-T1C	0.680	0.220	2.247	0.025
T2C vs T1A-T1C	0.882	0.266	2.508	0.012

For comparison, we first analyze the data by WLW model (Wei, Lin and Weissfeld 1989), treating the first category of each covariate as the reference level. The results show that those patients being in stage T2B and T2C have higher risk of recurrence, and those patients with higher level of pre-treatment PSA or larger values of Gleason Score are more likely to experience recurrence (Table 2). Though the above model is very useful in quantifying prognostic effects, it does not directly lead to risk groups. Next, we apply the proposed tree-structured method to this data and set the minimal size of the terminal nodes to be 20

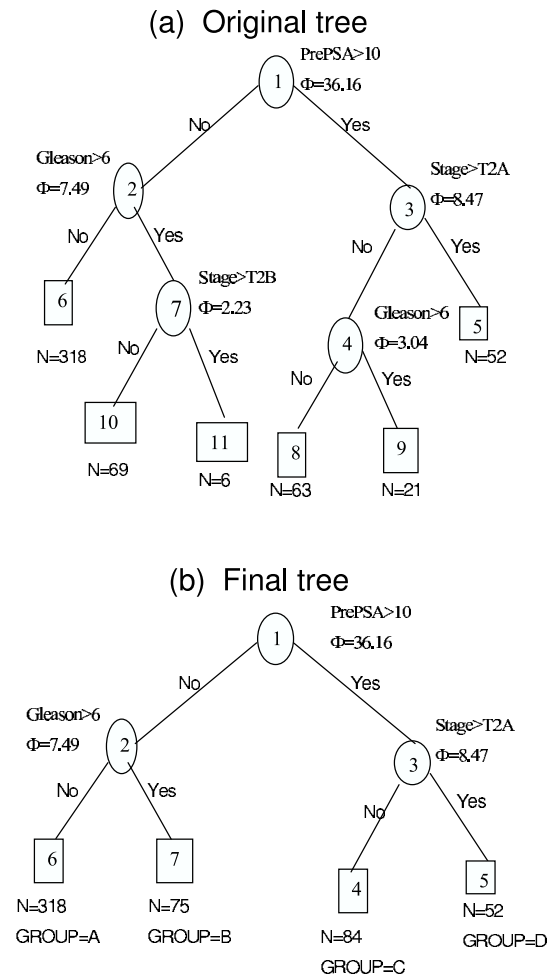


Figure 1: Multivariate survival trees for prostate cancer data, where Φ is the splitting statistics quantifying the separation between two subgroups and N represents the number of patients in the resultant prognostic groups.

failures. The method first splits on whether pre-treatment PSA is larger than 10 or not. Then Gleason Score (> 6 versus ≤ 6) and clinical stage (T1A-T2A versus T2B-T2c) are chosen as the best partitions respectively. After 5 partitions, an original tree with 6 terminal nodes is developed (Figure 1(a)), where circles and squares represent internal nodes and terminal nodes respectively, and Φ denotes the splitting function for each partition. After pruning process, we end up a final tree with 4 terminal nodes which lead to 4 risk groups (Figure 1(b)): PrePSA ≤ 10.0 and Gleason Score ≤ 6 (Group A, node 6), PrePSA ≤ 10.0 and Gleason Score > 6 (Group B, node 7), PrePSA > 10.0 and clinical stage being T1A-T1C or T2B (Group C, node 4), and PrePSA > 10.0 and clinical stage being T2B or T2C (Group D, node 5). Because larger values of covariates associated with

higher risk of recurrence, we anticipate that patients in Group A have the lowest recurrent risk and patients in Group D have the highest risk. The marginal Kaplan-Meier plots of these 4 prognostic groups are presented in Figure 2, and the curves show that these subgroups are well separated in each definition for disease-freedom. The risks of recurrence are in the order of Group A < Group B < Group C < Group D.

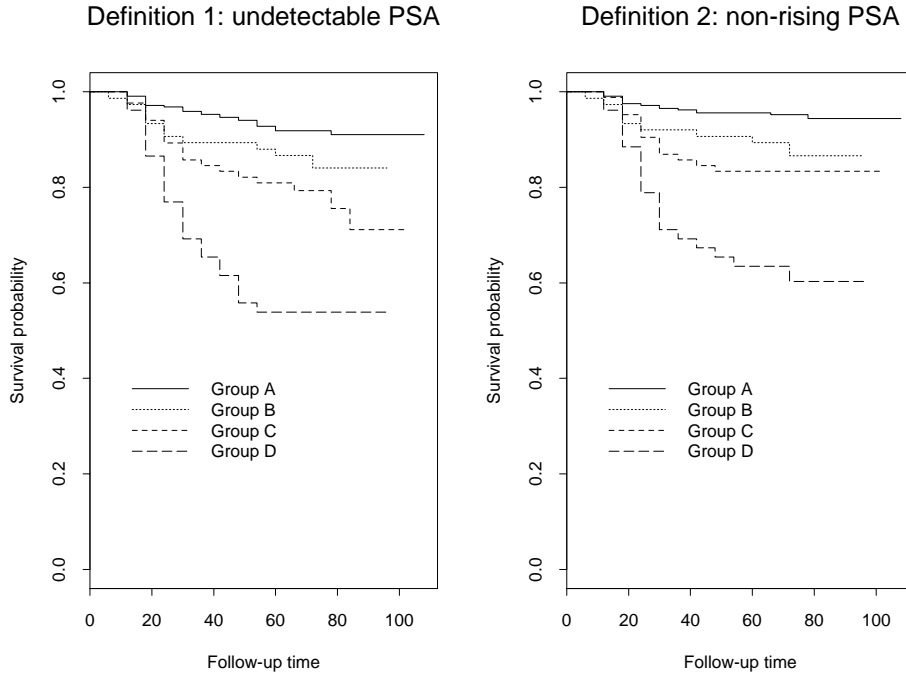


Figure 2: Kaplan-Meier plots for the 4 prognostic groups derived from the survival tree (A separate plot was prepared for each definition of disease freedom)

4. Simulation Studies

Several simulation studies are performed to assess the proposed method. Data sets are generated from the bivariate Clayton-Oakes model (Oakes 1982), with the joint survivor function $S(t_1, t_2) = Pr(T_1 > t_1, T_2 > t_2)$ given by,

$$S(t_1, t_2) = \{ \{G(t_1)\}^{(1-\theta)} + \{H(t_2)\}^{(1-\theta)} - 1 \}^{\frac{1}{1-\theta}},$$

where $G(t_1) = Pr(T_1 > t_1)$ and $H(t_2) = Pr(T_2 > t_2)$ are marginal survivor functions for survival times T_1 and T_2 respectively, and the parameter θ governs the association between T_1 and T_2 . As $\theta \rightarrow 1+$, $S(t_1, t_2) \rightarrow G(t_1)H(t_2)$ and the two

survival times become independent. As $\theta \rightarrow \infty$, $S(t_1, t_2) \rightarrow \min(G(t_1), H(t_2))$ and the bivariate distribution shows maximal association. In the simulation studies, different values of θ are set to introduce the model a low to moderate association, indexed by Kendall's $\tau=0.2, 0.4$, and 0.6 respectively.

In the simulations, for simplicity we consider $G(\cdot) = H(\cdot)$ and assume the hazard function to be

$$h_j(t|z) = h_0(t)g(z), \quad j = 1, 2$$

where the baseline hazard $h_0(t)$ is chosen such that the baseline survivor function has a Weibull distribution $S_0(t) = e^{-(\eta t)^\kappa}$ with a shape parameter $\kappa = 1.5$ and a scale parameter $\eta = 1$. The function $g(z)$ is defined as $g(z) = e^{\beta_1 I(z_1 > 2) + \beta_2 I(z_2 > 1)}$, with $\beta_1 = 1.0$, $\beta_2 = 0.75$ and $I(\cdot)$ to be the indicator function of covariates. Therefore, the above model leads to 4 prognostic groups. To have a better evaluation of the ability to correctly identify these prognostic groups, two extraneous variables (z_3 and z_4), which are not related to the outcome, are also included into the data. Among these 4 covariates, z_1 and z_4 have discrete uniform distributions taking values from 1 to 5, and z_2 and z_3 are dichotomous variables which take values 1 and 2 with the same probability. In addition to the complete data, censoring times are also introduced independently from a uniform distribution $[0, u]$. Approximately 30% censoring is introduced by adjusting the value of u . Each simulation has 1000 repetitions and each repetition consists of 400 pairs of survival times. The simulations are implemented via statistical package R (version 1.5). In simulation studies, the minimal size for a terminal node is set to be 40 failures.

Table 3: Simulations investigating the capability in identifying data structures

Given values		Correct	Over-sized	Partial-recognized	Failed
τ^*	censoring				
0.2	0%	0.980	0.011	0.009	0.000
	30%	0.957	0.024	0.019	0.000
0.4	0%	0.939	0.034	0.023	0.004
	30%	0.917	0.045	0.028	0.010
0.6	0%	0.897	0.048	0.039	0.016
	30%	0.868	0.061	0.048	0.023

*: Kendall's τ , which index the within-cluster association

The final trees are categorized according to their ability to identify correct data structures. A tree will be classified as "correct" if all 4 prognostic groups are identified as the terminal nodes of the pruned tree. A tree will be classified as

“over_sized” if all 4 prognostic groups are included in the tree, but not all of them are terminal nodes of the tree. A tree will be classified as “partial_recognized” if only 2 of the 4 prognostic groups are included in the tree. Otherwise the tree will be classified as “failed”. Table 3 summarizes the simulations, where τ reflects the correlation between survival times and a larger value of τ is associated with a higher correlation. The results show that the proposed method performs quite well in identifying correct data structures. In all the simulations, the prognostic groups in nearly 90% of the simulated data are correctly identified. However, the capability of identification decreases slightly in the presence of censoring. Because the power to detect the difference decreases as the correlation increases (Lee, Wei and Amato 1992), the capability of identification also decreases as the correlation increases.

5. Discussion

Tree-structured methods have provided a superior way for prognostic classification. Under classical CART algorithm, trees are grown by maximizing within-node homogeneity which is usually measured with log-likelihood. In multivariate survival data, writing a log-likelihood function is problematic due to the unknown marginal survivor functions as well as the presence of association among events. In contrast, the idea of growing trees rewarding between-node separation by Segal (1988) can be extended to develop survival trees in multivariate survival data. In the two survival trees by Gao *et al* (2004) and Su and Fan (2004), for example, a gamma distributed frailty was introduced to account for the dependence among survival times and the trees were developed maximizing relative risks as indexed by Wald test and likelihood ratio test respectively. In trees constructed by conventional methods, all the resultant subgroups are completely unrelated. However, this may not be true for the above multivariate survival tree because the two daughter nodes share a common baseline hazard function. That is, two components (relative risk and baseline hazards) are involved in the model, but only one (relative risk β) is used for tree construction and the other (baseline hazards) is left unspecified. To address this problem, we advocate a method to develop classification trees under proportional hazards structure. The proposed method works on marginal distributions of the survival times, and uses the “sandwich” estimator of variance to account for within-cluster association. The splitting function is based on the robust Wald-test statistics of WLW model (Wei, Lin and Weissfeld 1989) which is fast in computation. The proposed method can be used as a complimentary tool to the usual parametric or semi-parametric models. The method is useful in preliminary data analysis to identify the potential data structures, thus facilitating a better understanding of the survival data and a more efficient statistical modeling. We evaluate the method by several simula-

tion studies. The results show that the method performs quite well in prognostic classification. In all the simulations, data structures in nearly 90% data sets can be correctly identified.

The proposed algorithm for multivariate survival trees shares some inherent features of the univariate survival tree methods. Therefore, it is straightforward to further extend this algorithm in a variety of ways, i.e., incorporating surrogate splits (Breiman *et al*, 1984) to handle missing values in covariates or using boosting and bagging method (Hothorn *et al* 2004) to stabilize its performance. Comparing to the conventional WLW model (Wei, Lin and Weissfeld 1989) or the trees developed via frailty model (Gao *et al* 2004, Su and Fan 2004), the proposed method possesses some unique advantages. First, instead of presenting an algebraic formula, this method provides a set of subgroups directly. In addition, the prognostic groups are developed within the framework of proportional hazards model. Instead of assuming a locally proportional hazards structure, this algorithm assumes a proportional hazards structure within the whole tree and thus presents a clear model structure. As a consequence, a global optimal split is obtained at each partition because the best split is searched over the whole tree. Though the above advantages are gained in the price of intensive searching, computational time has become a minor issue with the availability of high performed personal computers nowadays.

Acknowledgments

The authors would like to thank Dr. FRANK A. CRITZ from the Radiotherapy Clinics of Georgia for allowing us to use the data set for demonstration. This work was supported by National Institutes of Health Grants R01-ES012458-01 and M01-RR00039 for Amita Manatunga.

References

- Bacchetti, P. and Segal, M. R. (1995). Survival Trees with Time-dependent Covariate: 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 Stone, C. J. (1984). Classification and Regression Tree. Chapman and Hall.
- Cox, D. R. (1972). Regression Models and Life Tables (with discussion). *Journal of Royal Statistical Society, Series B* **34**, 187-220.
- Critz, F. A (2002). A standard definition of disease freedom is needed for prostate cancer: undetectable prostate specific antigen compared with the American Society of Therapeutic Radiology and Oncology consensus definition. *The Journal of Urology* **167**, 1310-1313.

-
- Davis, R. B. and Anderson, J. R. (1989). Exponential Survival Trees. *Statistics in Medicine* **8**, 947-961.
- Gao, F., Manatunga, A. K., and Chen, S., (2004). Identification of prognostic factors for multivariate survival data. *Computational Statistics and Data Analysis* **45**, 813-824.
- Gordon, L. and Olshen, R. A. (1985). Tree-structured survival analysis. *Cancer Treatment Reports* **69**, 1065-1068.
- Hothorn, T., Lausen, B., Benner, A and Radespiel-Troger, M. (2004). Bagging survival trees. *Statistics in Medicine* **23**, 77-91.
- Huang, X., Chen, S., and Soong, S. (1998). Piecewise Exponential Survival Trees with Time-dependent covariates. *Biometrics* **54**, 1420-1433.
- LeBlanc, M. and Crowley, J. (1993). Survival Trees by Goodness of Split. *Journal of the American Statistical Association* **88**, 457-467.
- LeBlanc, M. and Crowley, J. (1992). Relative Risk Trees for Censored Survival Data. *Biometrics* **48**, 411-425.
- Lee, E. W., Wei, L. J., and Amato, D. A. (1992). Cox-type regression analysis for large numbers of small groups of correlated failure time observations. In *Survival Analysis: State of the Art* (Edited by J. P. Klein and P. K. Goel, P. K.), 237-247. Kluwer Academic Publishers.
- Lin, D. Y. and Wei, L. J. (1989). The robust inference for the Cox proportional hazards model. *Journal of the American Statistical Association* **84**, 1074-1078.
- Morgon, J. N. and Sonquist, J. A. (1963). Problems in the analysis of survey data and a proposal. *Journal of the American Statistical Association* **58**, 415-434.
- Oakes, D. (1982). A model for association in bivariate survival data. *Journal of Royal Statistical Society, Series B* **44**, 414-422.
- Segal, M. R. (1988). Regression Tree for Censored Data. *Biometrics* **44**, 35-47.
- Su, X. and Fan, J. (2004). Multivariate survival trees: A maximum likelihood approach based on frailty models. *Biometrics* **60**, 93-99.
- Wei, L. J., Lin, D. Y., and Weissfeld, L. (1989). Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *Journal of the American Statistical Association* **84**, 1065-1073.

Feng Gao
Division of Biostatistics
Washington University School of Medicine,
Campus Box 8067
660 S. Euclid Ave.
St. Louis, MO 63110, USA
feng@wubios.wustl.edu

Amita K. Manatunga
Department of Biostatistics
Rollins School of Public Health,
Emory University
1518 Clifton Road
Atlanta, GA 30322, USA

Shande Chen
Department of Biostatistics
School of Public Health
University of North Texas Health Science Center
Fort Worth, TX 76107, USA