Statistics and Its Interface Volume 10 (2017) 199–205

New tree structured survival analysis for hip fracture of SOF data

Hua Jin* and Ying Lu

Osteoporosis is a common disease among postmenopausal women and older men. It is critical to accurately predict osteoporotic fracture risk so that high-risk subjects can receive appropriate treatments before fractures occur. We want to establish classification algorithms for identifying subjects at high risk of hip fracture based on prospective cohort data from the Study of Osteoporotic Fracture (SOF). We propose a new algorithm that is similar to the traditional forward regression for survival analysis on the basis of restricted mean lifetime and apply it to the SOF data to form a classification tree. We also construct the second tree based on log-rank test statistic and the third tree using the martingale-type residuals from the Cox proportional hazards model without including any covariates of interest. We compare the three trees to each other. All the results suggest that the classification rule based upon our new method provide a good prognostic staging system for hip fracture both in classification efficiency and stability. Our proposed method may be a competitive alternative to conventional tree-structured survival analysis that uses multiple risk factors to provide powerful and understandable classification procedures.

Keywords and phrases: Log-rank test, Osteoporotic fracture, Restricted mean lifetime, Forward tree-structured survival analysis.

1. INTRODUCTION

Osteoporosis is a common disease among postmenopausal women and older men, which will lead to an increase in fracture risk. It is critical to predict osteoporotic fracture risk so that high-risk subjects can receive appropriate treatments before fractures occur. Instead of the traditional Cox proportional hazards model, Lu et al. [1, 2, 3] used tree-structured survival analysis (TSSA) based on log-rank test statistic [2] (denoted as LRTSSA) to study the effects of age and bone mineral density (BMD) measurements on time to hip fracture for the first time. One limitation of their paper was that no other clinical predictors except age and BMD measurements were used in the data analysis.

Although conventional TSSA has the advantages of effective classification and simple interpretation, especially in a clinical setting, it lacks a formal statistical framework allowing statistical inference. It grows a tree in a way similar to the traditional forward selection, but it is not based on hypothesis testing theory in variable selection. We want to make sure that a node splitting is not by chance. The current procedures provide few options in this respect.

Lu, Jin and Mi [3] introduced an index DOS, which is defined explicitly by Equation (2.2) in Section 2.2, based on restricted mean lifetime to measure the efficiency in prognostic separation by a classification method, and established a test framework to compare the efficiency of two classification methods with survival time as the endpoint. The method was successfully applied to studies of prediction of time to osteoporotic fractures and survival time of ovarian cancer patients. Like tree based regression analysis that uses variance as a criterion for node partition and pruning, the variance of restricted mean lifetime between different groups can be an alternative index to the log-rank test statistic in construction of survival trees.

In this paper, we propose an algorithm that is similar to the traditional forward regression for survival analysis. First, we explain the methodological details for node splitting and stopping under a statistical framework. In the next section, we apply the proposed method to the data from the Study of Osteoporotic Fractures (SOF) [4, 5, 6]. In the third section, we compare our classification rule to that of the conventional LRTSSA. In the fourth section, we present results of a simulation study that compared our method with LRTSSA, which is similar to RTSSA. In the last section, we present our discussion and conclusion.

2. METHOD

2.1 Subjects

Our data includes 7,784 women. They all had forearm, calcaneus, hip and AP spine BMD measurements, with values presented as T-scores according to reference values developed by Lu et al. [7]. In addition to the BMD measurements, other factors that have been previously identified as predictive variables at baseline, such as height loss since age 25, weight, body mass index, walking speed, functional score, and vision depth, were also investigated. Time to hip fracture after BMD measurement was also recorded and was treated as the outcome variable. For women without a hip fracture, the last examination was considered as a censoring time for hip fracture.

*Corresponding author.

2.2 Statistical method

Let T be the random variable of survival time for a random subject from a population and the corresponding survival function be

$$S(t) = P(T \geq t)$$

Suppose we classify the population into g mutually exclusive and exhaustive groups. Let G be the random variable of group indicator and $p_j = P(G = j), j = 1, 2, ..., g$, be the corresponding probability of a patient assigned to the jth group, where $\sum_j p_j = 1$.

Furthermore, let $S_j(t) = P(T \geq t \mid G = j)$ and $m_j = \int_0^\infty S_j(t)dt$ be the corresponding survival function and mean survival time of the jth group respectively. Then the mean survival time of the population is easily shown to be

$$m = \sum_j m_j p_j \tag{2.1}$$

Considering that the larger the between-group variance in mean survival times, the larger the difference in mean survival times between prognostic groups, Lu, Jin and Mi [3] defined the degree of separation (DOS) index as the variance of group mean survival times, i.e.,

$$DOS = \sum_j (m_j - m)^2 P_j \tag{2.2}$$

Estimation of mean survival time based on sample data may be unfeasible because of the presence of censoring [8, 9, 10]. Several authors [11, 12, 13, 14, 15] suggested the use of restricted mean life time with a suitably chosen finite time LT, to replace the overall mean survival time in practical applications. Then the restricted mean lifetime for group j, j = 1, 2, . . . , g, is defined as $m_j(LT) = \int_0^{LT} S_j(t)dt$, and the restricted mean lifetime of the population can be expressed as

$$m(LT) = \sum_j m_j(LT) p_j \tag{2.1}'$$

In such a case, the corresponding index to measure the degree of the separation becomes

$$DOS[LT] = \sum_j [m_j(LT) - m(LT)]^2 p_j \tag{2.2}'$$

Note that Equation $(2.1)'$ and $(2.2)'$ are similar to Equations (2.1) and (2.2), respectively, except that the latter uses a restricted time limit. In the rest of this article, we skip LT in our notation for m and DOS.

2.2.1 Splitting rule

At each new separation (or split), all of the included predictive variables are reexamined. In order to make the splitting result simple and robust, the observed values of the continuous predictors are first rounded to the nearest tenth, and then used as cutoff values for partition: Subjects with values less than or equal to a specific value fall into the left daughter node, whereas other subjects go to the right. Changing the cutoff points of each predictive covariate generates many possible separations. We choose the optimal split which results in the largest change of DOS. The details are presented as follows:

From Equation (2.2), the contribution of group j (*i.e. node j*) to the overall separation (*i.e. DOS*) is $(m_j - m)^2 p_j$. If we partition the group j into two daughter nodes, the change of DOS is

$$\begin{aligned} CDS_j &= (m_{j1} - m)^2 p_{j1} + (m_{j2} - m)^2 p_{j2} - (m_j - m)^2 p_j \\ &= \frac{(m_{j1} - m_{j2})^2 p_{j1} p_{j2}}{(p_{j1} + p_{j2})} \end{aligned}$$

where p_{j1} and p_{j2} are the probabilities of a subject in group j being further classified into the left and right daughter nodes, and m_{j1} and m_{j2} are the corresponding restricted mean survival times, which satisfies $p_{j1} + p_{j2} = p_j$ and $m_{j1}p_{j1} + m_{j2}p_{j2} = m_j p_j$. So partitioning a node does not affect any nodes other than its ancestors.

Suppose that there are n_{j1} and n_{j2} subjects in the left and right daughter nodes of the jth group, then the maximum likelihood estimates of p_{j1} and p_{j2} are $\hat{p}_{j1} = n_{j1}/n$ and $\hat{p}_{j2} = n_{j2}/n$ respectively, where n is the total number of subjects in the study. The estimates of the restricted mean lifetime for the two daughter nodes of the jth group are respectively $\hat{m}_{j1} = \int_0^{LT} \hat{S}_{j1}(t)dt$ and $\hat{m}_{j2} = \int_0^{LT} \hat{S}_{j2}(t)dt$, where $\hat{S}_{j1}(t)$ and $\hat{S}_{j2}(t)$ are the Kaplan-Meier estimators of the corresponding survival functions. Then a natural estimation of CDS_j is given by

$$\hat{CDS}_j = \frac{(\hat{m}_{j1} - \hat{m}_{j2})^2 \hat{p}_{j1} \hat{p}_{j2}}{\hat{p}_{j1} + \hat{p}_{j2}}$$

For this node, we select a split that results in the largest change of DOS (i.e. $\hat{CDS}_j$). This maximizes the difference in the restricted mean survival times between the two daughter nodes.

2.2.2 Stopping rule

Most TSSA use cross-validation to prune a large tree or are based on the plot of the log-rank test statistic against the tree size. They are either computationally intensive or subjective. Here we adopt a direct stopping rule: Stop partitioning node j if either the number of subjects in the node is less than a pre-specified minimum node size n_0 (for example, 100 for a large study), or there is no significant change of DOS to justify splitting the node, i.e. $CDS_j \leq lm$, where lm is a pre-specified limit, such as 0.1% or 0.2% of variance of the restricted mean lifetime.

Specifically, we construct a test for $H_0 : CDS_j \leq lm$ versus $H_1 : CDS_j > lm$ with the statistic

$$Z_j = \frac{\hat{CDS}_j - lm}{\sqrt{\hat{Var}(\hat{CDS}_j)}}$$

where

$$\hat{Var}(C\hat{D}S_j) = (C\hat{D}S_j)^2\left[\left(\frac{\hat{p}_{j1}^2 - \hat{p}_{j1}\hat{p}_{j2} + \hat{p}_{j2}^2}{\hat{p}_{j1}\hat{p}_{j2}(\hat{p}_{j1}+\hat{p}_{j2})} - 1\right)\frac{1}{n} + 4\frac{\hat{\sigma}_{j1}^2/n_{j1} + \hat{\sigma}_{j2}^2/n_{j2}}{(\hat{m}_{j1}-\hat{m}_{j2})^2}\right]$$

by δ method, and $\hat{\sigma}_{ji}^2$ is a consistent estimator of $\sigma_{ji}^2 = -\int_0^{LT}\frac{(\int_t^{LT} S_{ji}(u)du)^2}{S_{ji}^2(t)C_{ji}^-(t)}dS_{ji}(t) < \infty$ by substituting the Kaplan-Meier estimators for S_{ji} and $C_{ji}^-(t)$ in it, and letting $C_{ji}^-(t)$ be the left continuous version of $C_{ji}(t)$ which is the survival function for the independent and identically distributed censoring random variables in the subnode ji, i = 1,2. More details are available in Lu, Jin and Mi [3], Gill [9] and/or Pepe and Fleming [16, 17].

We will partition node j if and only if the node size $n_j \geq n_0$ and $Z_j \geq z_{1-\alpha}$, where $z_{1-\alpha}$ is the $1-\alpha$ percentile for the standard normal distribution. The same procedure can be applied to other nodes until no further split satisfies the above criterion.

2.2.3 Grouping rule

To facilitate clinical applications, we group those terminal nodes (TNs, nodes in a tree that have no daughter nodes) with similar survival profiles over the whole follow-up period. Here we employ the log-rank test to determine if there are any differences between the survival curves of those terminal nodes at the significant level of 0.05. We combine those terminal nodes with similar survival curves into one group.

The resulting tree is denoted as DOSTSSA tree in this paper.

3. RESULTS

3.1 The classification tree and rule by DOSTSSA

Considering the exploratory nature of TSSA, we randomly divide our data into two parts: the training data (5,875 with 453 fractured) is used to generate survival trees and the validation data (1,909 with 152 fractured) used to compare the recursive partitioning algorithms.

Using the DOSTSSA method with the restricted time $LT = 12$ (years), the significance level $\alpha = 0.1$, the prespecified limit $lm = 0.2\%{}^*Var(T)$ and minimum node size $n_0 = 60$, we construct a classification tree using three predictors–femoral neck BMD, age and walking speed, which is called the DOSTSSA tree and is presented in Figure 1.

Note that we group those terminal nodes with similar survival curves for clinical use: terminal node TN2 is the worst risk group (Group 1); TN1, TN3 and TN5 are combined to form Group 2; TN4 forms Group 3; and TN6 has the best prognosis (Group 4). The final classification of all the subjects into four different risk groups is presented in Table 1, denoted as the DOSTSSA classification rule.

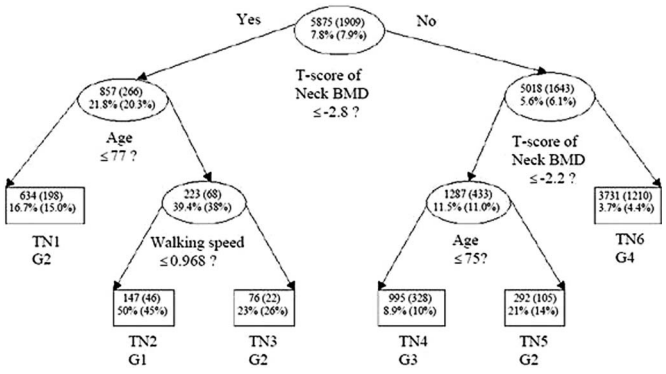

Figure 1. The survival tree by DOSTSSA.The number of subjects in each resulting subgroup, and the corresponding probability (in %) of hip fracture in a 12 year follow up period without parentheses resulted from the generating data used for construction of the trees and within parentheses resulted from the validation data.

For our DOSTSSA rule, the differences in the probabilities of no hip fracture (survival probability) among the four groups increase with the length of follow-up, as shown by Figure 2. At the end of 10 years, the probability of hip fracture is 49.8% for Group 1, compared to 3.7% for Group 4. Their ratio is the relative risk presented in Table 1. Similarly, we present ratios of probabilities of hip fracture for Groups 2 and 3 compared to Group 4 in Table 1, where we also give the restricted mean survival time with LT = 10 for each group.

The results of the DOSTSSA classification tree and rule are reproducible on the validation data set as shown in Table 2 and by comparison of Figures 2 and 3.

From Table 2, we know that the DOS index (with $LT =$ 12 years) of the DOSTSSA classification rule based on the generating data is about 0.34 ($year^2$) while the corresponding DOS from the validation data reduces to 0.25 ($year^2$). If we use the ratio of two DOSs to describe the reproduce performance in validation data sets, its efficiency may be about $\frac{0.25}{0.34} = 73.5\%$, which maybe happen to be somewhat low. In order to see more clearly how well our method can reproduce its performance for validation data, we employ a 6-fold cross-validation (CV) to quantify the reliability of this new method. We randomly divide the 7,784 women into 6 subsets such that each subset has almost the same number of subjects. We use 5 subsets to generate the tree (with $LT = 12$ years) and left one subset for validation. Repeat the procedures 6 times and the results are presented in Table 3. The corresponding efficiency (Ratio of DOS) varies from 76.9% to 89.5%, which suggest that our reproduce performance may be good.

Table 1. Classification of subjects by risk of hip fractures based on the DOSTSSA rule

Risk Group	Terminal Node	percentage of Subjects (N = 5,875)	Definition	RR* and 95% C.I.	RMST** (years)
G1	TN2	2.5%	femoral neck BMD <= -2.8 and age > 77 and walking speed <= 1.0	13.5 (8.6, 21)	7.7
G2	TN1	17.1%	neck BMD <= -2.8 and age <= 77,	5.0 (3.8, 6.4)	9.2
	TN3		neck BMD <= -2.8 and age > 77 and walking speed > 1.0		
	or		or		
	TN5		-2.8 < neck BMD <= -2.2 and age > 75		
G3	TN4	16.9%	-2.8 < neck BMD <= -2.2 and age <= 75	2.4 (1.8, 3.2)	9.7
G4	TN6	63.5%	neck BMD > -2.2	1	9.9

The original terminal nodes TN1, TN3 and TN5 were combined into Group 2 due to their similar survival profiles.
* Relative Risk (RR) for hip fracture within 10 years follow-up when risk group 4 was the reference group.
** Restricted Mean Survival Time (RMST) was computed with LT = 10 years.

Table 2. Statistical Utility for Classification Rules

Rules	DOSTSSA	LRTSSA	RTSSA
Log-rank test statistic*	590	527	592
	(123)	(130)	(109)
DOS (LT = 10 years)*	0.16	0.13	0.15
	(0.13)	(0.13)	(0.13)
DOS (LT = 12 years)*	0.34	0.30	0.33
	(0.25)	(0.25)	(0.24)

* Log-rank test statistic and DOS indices without parentheses resulted from the generating data used for construction of the rules and within parentheses resulted from the validation data.

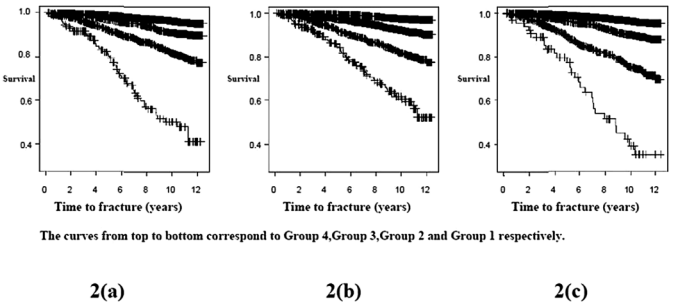

Figure 2. Kaplan-Meier survival curves for hip fractures of (a) the Survival Tree by DOSTSSA, (b) the Tree by LRTSSA and (c) the Tree by RTSSA based on generation data. The differences in survival probability in the four groups increase with the length of follow up. After 10 years, the probability of no hip fracture was (a) 50.2% for G1 versus 96.3% for G4, (b) 60.6% for G1 versus 97.6% for G4 and (c) 39.0% for G1 versus 96.3% for G4.

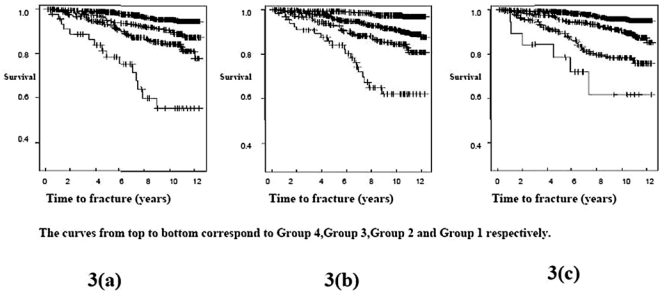

Figure 3. Kaplan-Meier survival curves for hip fractures of (a) the Survival Tree by DOSTSSA, (b) the Tree by LRTSSA and (c) the Tree by RTSSA based on validation data. The differences in survival probability in the four groups increase with the length of follow up. After 10 years, the probability of no hip fracture was (a) 55.5% for G1 versus 95.6% for G4, (b) 62.0% for G1 versus 97.5% for G4 and (c) 61.4% for G1 versus 95.6% for G4.

Table 3. Comparison of the DOS indices with $LT = 12$ years based on 6-fold CV

DOS	DOS0*	DOS1**	DOS1/ DOS0
1	0.320	0.268	83.8%
2	0.305	0.256	83.9%
3	0.321	0.254	79.1%
4	0.311	0.265	85.2%
5	0.325	0.250	76.9%
6	0.314	0.281	89.5%

* The DOS index DOS0 was computed using the generating data
** The DOS index DOS1 was computed using the validation data

To know more about how well the results are reproduced in the validation, we directly make a node-by-node comparison of the resulting DOS tree from the validation data to that one based on the generating data. For each of the six terminal nodes, we list the percentage of total subjects and the corresponding restricted mean survival time with $LT = 12$ in Table 4. It follows from the table that there is little difference between them, and the corresponding six terminal nodes match well in the ordering of the restricted mean survival times with $LT = 12$.

Table 4. Node-by-node Comparison of the two DOS trees from the generating data and the validation data

Terminal Node	RMST* (years)	percentage of Subjects (%)
TN1	10.9	10.8
	(10.9)	(10.4)
TN2	8.6	2.5
	(8.9)	(2.4)
TN3	10.6	1.3
	(10.3)	(1.2)
TN4	11.5	16.9
	(11.3)	(11.7)
TN5	10.6	5.0
	(10.9)	(5.5)
TN6	11.8	63.5
	(17.1)	(63.4)
Total	11.4	100
	(11.4)	(100)

* Restricted Mean Survival Time (RMST) was computed with LT = 12 years.
The numbers without parentheses resulted from the generating data used for construction of the rules and within parentheses resulted from the validation data.

3.2 Comparisons to other two classification trees

We also apply TSSA based on log-rank test statistic to the same data (Figure 4, the LRTSSA tree). In addition, we use the martingale-type residuals from the Cox proportional hazards model without including any covariates of interest [18, 19, 20] to construct another tree (Figure 5, i.e. RTSSA tree). Both of them use only age and femoral neck BMD. However, it seems that TSSA based on log-rank test statistic leads to further splitting for patients with higher neck BMD values (> -2.2) while residual-based TSSA (denoted as RTSSA) results in more partitioning for those with lower neck BMD values (< -2.8).

For a direct comparison, we list log-rank test statistics and DOS indexes with different restricted times for the three classification rules resulting from DOSTSSA, LRTSSA and RTSSA in Table 2. The corresponding indexes of the DOSTSSA rule are among the best of all. In fact, the DOSTSSA and RTSSA rules were better than the LRTSSA rule for the generation data while the DOSTSSA and LRTSSA rules appeared slightly better than the RTSSA rule for the validation data. Furthermore, this is reinforced by comparison of the corresponding Kaplan-Meier survival curves based on both generation data and validation data, presented in Figures 2 and 3, where (a), (b) and (c) represent the DOSTSSA, LRTSSA and RTSSA rules, respectively. On the one hand, our DOSTSSA rule can separate more patients with higher risk from others than the RTSSA rule. On the other hand, it can also result in the highest risk group being further separated from others than the LRTSSA rule. The classification is improved by incorporating additional variable of walking speed in the sense that the corresponding DOS index increases about 13.0% and 7.2% respectively for the generation data and validation data.

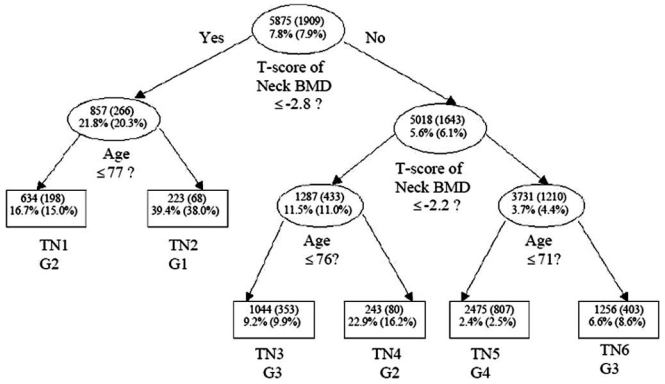

Figure 4. The survival tree by LRTSSA. The number of subjects in each resulting subgroup, and the corresponding probability of hip fracture in a 12 year follow up period without parentheses resulted from the generating data used for construction of the trees and within parentheses resulted from the validation data.

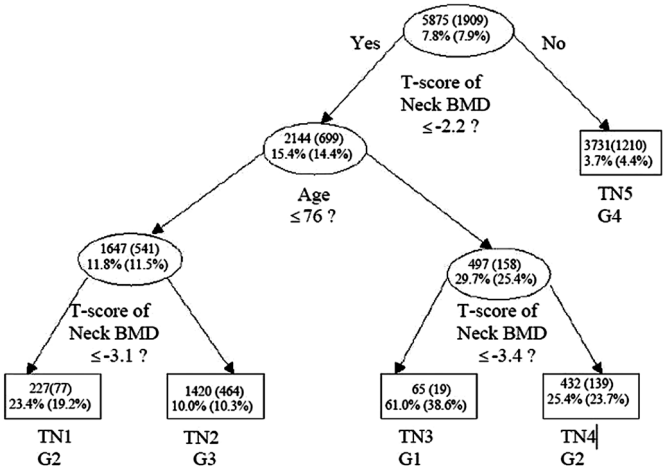

Figure 5. The survival tree RTSSA. The number of subjects in each resulting subgroup, and the corresponding probability of hip fracture in a 12 year follow up period without parentheses resulted from the generating data used for construction of the trees and within parentheses resulted from the validation data.

All methods construct trees based on 43 predictors of interests, but only DOSTSSA method can pick up three of them to build the classification rule while the other two cannot. In particular, DOSTSSA method can further split subjects of the relatively high fracture risk with neck $BMD \leq -2.8$ and $age > 77$ whereas the LRTSSA misses this partition. From this view, our DOSTSSA is able to provide information missed by LRTSSA.

In summary, all these results suggest that the classification rule based upon the DOSTSSA method provides a good

prognostic staging system for hip fracture both in classification efficiency and stability.

4. A SIMULATION STUDY

Comparing the statistical utility of 2 approaches applied to a single data set is insufficient. To test further the validity of our conclusions, we performed a limited simulation study to investigate 1) the dependence of performance of the DOSTSSA method on the censoring patterns and 2) comparison of the effectiveness in selecting splits between DOSTSSA and LRTSSA/RTSSA. Because the differences between LRTSSA and RTSSA are minimal both in theory and in reality [20, 21], we compare DOSTSSA only with LRTSSA in this simulation study.

Because the RP algorithm is based on sequential splits, comparison of a 1-step split between 2 methods is sufficient. In our design of this experiment, we introduced 3 covariates, X_1, X_2 and X_3, that were generated independently from uniform distribution on interval $(0, 1)$. Survival times were generated from exponential distributions with the hazard functions of $-\frac{\log(0.3I_{(X_1>0.5)}+0.4)}{10}$. Thus, survival time depended only on X_1 whereas X_2 and X_3 were nuisance variables. For $X_1 > 0.5$, the mean survival time was 28, and for $X_1 \leq 0.5$, the mean survival time was 11. Furthermore, we investigated 3 types of censoring distributions: 1) a uniform distribution on interval (0,18), 2) a distribution with density function of $\frac{(18-x)I_{(0<X<18)}}{162}$, and 3) a distribution with a density function of $\frac{xI_{(0<X<18)}}{162}$. The 2nd case implied heavy censoring early on, whereas the 3rd used heavy censoring late in the process.

In our simulation, the top node has 500 observations. The minimum node size of a daughter node was 20 observations, the restricted time limit LT was selected as 10 and the pre-specified limit $lm = 0.2\%^*Var(T)$ for the DOSTSSA method. Splits could go as deep as possible for both TSSA methods, although there was only 1 true split. 30 trials were conducted for each experimental condition. A method was successful in a trial when it generated only 1 split based on variable X_1. The number of successful trials among 30 trials is the summary statistic for our comparison.

When the censoring time is uniformly distributed, the DOSTSSA method succeeded in 14 of 30 trials, whereas LRTSSA succeeded only 8 times. There was no trial in which LRTSSA succeeded DOSTSSA failed. The only failure for DOSTSSA was no splitting. There was neither a trial successful for 1 separation but using the nuisance variables or using X_1 in a wrong direction, nor a trial that had more than 2 splits. However, several types of failure occurred for LRTSSA, including no splitting and 2 or more splits. An exact McNemar test suggested that DOSTSSA has a significantly higher success rate than LRTSSA ($P = 0.03$). The results were the same whether there was heavy censoring early or late: DOSTSSA had an overall success rate of about 44%, almost twice that of LRTSSA. This simulation suggests that the DOSTSSA approach is not as dependent on the censoring mechanism and can split nodes successfully more often than LRTSSA, under our experimental setting.

5. DISCUSSION

Based upon the variance of the restricted mean lifetime between groups, we have established a formal statistical framework for tree construction, under which a nonparametric TSSA different from the conventional methods is proposed. The new method employs DOS for node partition and uses a direct stopping rule with obvious statistical meanings rather than the usual cross-validation method.

The index DOS measures the difference of the restricted mean lifetime between groups. It is similar to the Least Squares criterion in CART for regression tree, or log-rank test statistics in LRTSSA. However, it is more intuitive for node partitioning and tree construction for survival data. Gordon and Olshen [22, 23, 24, 25] have previously developed TSSA based on the distance between the two Kaplan-Meier curves using Lp Wasserstein metrics. Although their approach enjoys a theoretical advantage of convex function of distance function and compares survival profiles over the whole range of interest, the numerical implementation is difficult. Our approach is much easier to calculate.

Our method has a format of forward regression analysis. However, there is a big difference between DOSTSSA and forward regression. In forward regression, we must check type III p-values for all the covariates in the equation after a new one has been entered. Different than this situation, our statistical test for node splitting has no impact on any other nodes except its ancestors and hence, no type 3 p-value reassessment is necessary.

The limiting time LT is critical to node partition and tree construction as well as comparison of two DOS indices. However, we believe the result of the final tree will remain stable if a suitable LT is selected. Because of the nonparametric nature of our approach and limitations introduced by censoring, we could not estimate mean survival times beyond the LOL (largest observed lifetime) [13]. Naturally, LOL could be chosen as Lt. Lu and others [3] discussed some possible choices of a suitable Lt. In our example in this article, the survival probability of the censoring time drops sharply from about 40% at 11 years to 18% at 12 years. We set Lt at 12 years, at which time the survival function of the censoring time is relatively large so that the DOS indices are still robust. In fact, we also construct a survival tree with LT set at 10 years. It resembles the tree presented in Figure 3 in this paper, where LT is selected as 12 years.

We would like to point out that the final tree size depends not on the significance level of the hypothesis test, but on the pre-specified limit lm. The larger the significance level, the more nodes we may split. On the other hand, the bigger the pre-specified limit, the smaller the tree we will construct. Hence, there is a trade off between these two parameters. For example, if we select the significance level to be 0.15, then we have one more node split. However, the SOF data suggests 0.10 be a reasonable choice for the significance level.

In conclusion, we propose a new method for nonparametric TSSA that first directly connects single node partition to total tree construction based upon a formal statistical test framework. It can result in a classification tree competitive with conventional methods such as TSSA based on log-rank test statistic or residual-based TSSA. Our example shows its useful application in medical research.

ACKNOWLEDGEMENTS

The study is supported by grants from the National Institutes of Health R03 AR47104, R01EB004079 and from National Bureau of Statistics of China LX:2006B45.

Received 28 May 2015

REFERENCES

[1] Lu Y, Black D, Mathur AK, et al. (2003) Study of hip fracture risk using tree structured survival analysis. J Miner Stoffwechs 10(1):11–16.
[2] Segal M (1988) Regression trees for censored data. Biometrics 44:35–47.
[3] Lu Y, Jin H, and Mi J (2004) On comparison of two classification methods with survival endpoints. In Balakrishnan N and Rao CR (eds) (2004) Handbook of Statistics, Volume 23: Advances in Survival Analysis. North–Holland, Elsevier, pp 43–59. MR2065760
[4] Cummings S, Black D, Nevitt M, et al. (1993) Bone density at various sites for prediction of hip fractures: The study of osteoporotic fractures. Lancet 341:72–75.
[5] Cummings S, Nevitt M, Browner W, et al. (1995) Risk factors for hip fracture in white women. Study of osteoporosis research group [see comments]. N Engl J Med 332:767–773.
[6] Cummings S, Browner W, Bauer D, et al. (1998) Endogenous hormones and the risk of hip and vertebral fractures among older women. Study of Osteoporotic Fractures Research Group. N Engl J Med 339(1):767–768.
[7] Lu Y, Fuerst T, Hui S, et al. (2001) Standardization of bone mineral density at femoral neck, trochanter and Ward's triangle. Osteoporos Int 12(6):438–444.
[8] Susarla V, Van Ryzin J (1980) Large sample theory for an estimator of the mean survival time from censored samples. Ann Stat 8:1002–1016. MR0585699
[9] Gill R (1983) Large sample behaviour of the product–limit estimator on the whole line. Ann Stat 11:49–58. MR0684862
[10] Zheng Z (1995) Two methods of estimating the mean survival time from censored samples. Sankhya, Series A, Indian Journal of Statistics 57:126–136. MR1392637
[11] Irwin JO (1949) The standard error of an estimate of expectational life. Journal of Hygiene 47:188–189.
[12] Kaplan EL, Meier P (1958) Nonparametric estimation from incomplete observations. J Am Stat Assoc 53:457–481. MR0093867
[13] Karrison T (1987) Restricted mean life with adjustment for covariates. J Am Stat Assoc 82:1169–1176. MR0922182
[14] Zucker DM (1998) Restricted mean life with covariates: Modification and extension of a useful survival analysis method. J Am Stat Assoc 93:702–709. MR1631365
[15] Chen P–Y, Tsiatis AA (2001) Causal inference on the difference of the restricted mean lifetime between two groups. Biometrics 57(4):1030–1038. MR1950418
[16] Pepe MS, and Fleming, Thomas R (1989) Weighted Kaplan–Meier statistics: A class of distance tests for censored survival data. Biometrics 45:497–507. MR1010515
[17] Pepe MS, Fleming TR (1991) Weighted Kaplan–Meier statistics: Large sample and optimality considerations. J Roy Stat Soc B 53(2):341–352. MR1108331
[18] Therneau T, Grambsch, P, and Fleming, T (1990) Martingale based residuals for survival models. Biometrika 77:147–160. MR1049416
[19] LeBlanc M, Crowley J (1992) Relative risk trees for censored survival data. Biometrics 48:411–425.
[20] Keles S, Segal MR (2002) Residual-based tree-structured survival analysis. Stat Med 21:313–326.
[21] Zhang H, Singer D (1999) Recursive Partitioning Tree and Application in Health Sciences. New York: Springer–Verlag.
[22] Gordon L, Olshen RA (1978) Asymptotically efficient solutions to the classification problem. Ann Stat 6:515–533. MR0468035
[23] Gordon L, Olshen RA (1980) Consistent nonparametric regression from recursive partitioning schemes. J Multivariate Anal 10:611–627. MR0599694
[24] Gordon L, Olshen RA (1984) Almost surely consistent nonparametric regression from recursive partitioning schemes. J Multivariate Anal 15:147–163. MR0763592
[25] Gordon L, Olshen RA (1985) Tree-structured survival analysis. Cancer Treatment Reports 69:1065–1069.

Hua Jin
School of Mathematical Sciences
South China Normal University
Guangzhou
China
E-mail address: jinh1@163.com

Ying Lu
Department of Health Research and Policy
Stanford University
CA
USA

Palo Alto Cooperative Studies Program Center
Palo Alto VA Health Care System
USA
E-mail address: ylu1@stanford.edu