Research Article

Statistics in Medicine

Received 16 July 2013, Accepted 20 October 2014 Published online 10 November 2014 in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/sim.6360

Optimal survival time-related cut-point with censored data

Xinhua Liu*† and Zhezhen Jin

In biomedical research and practice, continuous biomarkers are often used for diagnosis and prognosis, with a cut-point being established on the measurement to aid binary classification. When survival time is examined for the purposes of disease prognostication and is found to be related to the baseline measure of a biomarker, employing a single cut-point on the biomarker may not be very informative. Using survival time-dependent sensitivity and specificity, we extend a concordance probability-based objective function to select survival time-related cut-points. To estimate the objective function with censored survival data, we adopt a non-parametric procedure for time-dependent receiver operational characteristics curves, which uses nearest neighbor estimation techniques. In a simulation study, the proposed method, when used to select a cut-point to optimally predict survival at a given time within a specified range, yields satisfactory results. We apply the procedure to estimate survival time-dependent cut-point on the prognostic biomarker of serum bilirubin among patients with primary biliary cirrhosis.

Keywords: censored data; concordance probability; cut-point; sensitivity; specificity

1. Introduction

In biomedical research and practice, a biomarker related to the progression of a disease is commonly used to predict its prognosis. When the biomarker measure is continuous and a binary outcome variable for a good or poor prognosis is well defined, then we can select a cut-point on the biomarker to aid binary classification. We define sensitivity and specificity at each possible cut-point as the probability of correctly classifying patients into the poor-prognosis and the good-prognosis categories, respectively; one may use the receiver operational characteristics (ROC) curve [1,2], with sensitivity versus 1 − specificity (or the false positive rate), to show how sensitivity changes with changes in specificity across cut-points. There are several ROC curve-related criteria that are used in optimal cut-point selection. The most commonly used is the Youden index [3], defined, at each cut-point, as the sum of the associated sensitivity and specificity minus one. The cut-point maximizing the index will maximize the total correct classification rate, or equivalently minimize the sum of error rates, the false negative rate (= 1 − sensitivity), and the false positive rate [4]. Less commonly used is the closest-to-(0, 1) criterion [5], which selects the cut-point minimizing the distance between a point on the ROC curve and an ideal point (0, 1) representing perfect specificity and sensitivity. A recently proposed criterion selects the cut-point maximizing the classification concordance probability, which is equivalent to a product of the associated sensitivity and specificity [6]. If an ROC curve-related criterion is to be used for selecting the cut-point on a continuous biomarker and if prognosis is measured quantitatively, then the measure must be divided into two groups to distinguish between good and poor prognosis. A clinically meaningful dichotomization on the prognostic measure should be based on clinical considerations.

When a disease is diagnosed, survival time or time to a serious event (T) is often used for disease prognosis. The survival curve, which describes the distribution of survival time, can be quantified both as a function of time t, survival rate $r = P(T > t)$ for time t, and as the quantile of survival time $T(r)$ for survival rate r. Practically, clinicians can use either for the prognosis of a specific disease. For example, a 5-year survival rate estimates the probability of a patient surviving for 5 or more years, and median

Department of Biostatistics, Columbia University, New York, NY 10032, U.S.A.
**Correspondence to: Xinhua Liu, Department of Biostatistics, Columbia University, New York, NY 10032, U.S.A.*
†E-mail: xl26@columbia.edu

survival time indicates the amount of time that a patient would survive beyond, with probability of 0.5. Dichotomizing survival time T at specified value t means classifying a sample of patients into two groups, with the $D = 1$ group having experienced the event up to time t ($T \leqslant t$), and the $D = 0$ group experiencing it after time t ($T > t$). Defining sensitivity and specificity for each possible cut-point on X, we can draw an ROC curve for t. Using a preferred ROC curve-related criterion or objective function at the time, we can select a cut-point on the biomarker X. Consequently, the selected cut-point depends on the dichotomization of the survival time. In studies of disease prognosis, survival time is likely to be subject to right censoring because not all patients who provide biomarker data at the baseline may experience the event before the end of the study and some of the patients may drop out from the study during follow up. Depending upon the censoring pattern, attrition in the sizes of the dichotomized groups may also have impact on cut-point selection. Heagerty *et al.* [7] investigated methods for estimating time-dependent ROC curves with censored data in evaluating the performance of continuous biomarkers in predicting time-dependent outcomes. Building on their work, we extend an ROC curve-related criterion in order to select an optimal survival time-related cut-point corresponding to a specified survival time or survival rate.

Recently, Gönen and Sima [8] considered a single cut-point model for the selection of an optimal cut-point with censored data. They examined five criteria, including the Wald, log-rank, and partial likelihood ratio statistics, three chi-square-based metrics, and two related to concordance probabilities. The authors defined the concordance probability-based criteria either non-parametrically or semi-parametrically using a proportional hazards model with a single predictor, which was also assumed for the criteria using the Wald and partial likelihood ratio statistics. Through simulations performed under a variety of scenarios, the authors showed that the partial likelihood ratio test statistic has the best performance. However, the authors did not provide a way to check their fundamental assumption that the cut-point on the biomarker is unique.

It is more realistic to relax the single cut-point assumption and to not restrict the pattern of association between a continuous biomarker X and survival time T. To evaluate the contribution of X in predicting prognosis, one may either use statistical models or some summary measure such as Harell's C-index or the concordance probability [9, 10]. Supposing that a smaller X is associated with a larger T, the concordance probability $C = P(T_1 < T_2 \mid X_1 > X_2)$ or $C^* = P(X_1 > X_2 \mid T_1 < T_2)$ describes the discrimination accuracy of X in predicting survival time. A value above 0.50 may indicate some relationship between X and T, while a value above 0.60 would be considered meaningful. It is known that when T is a binary variable, the concordance probability C^* is equivalent to the area under the ROC curve. For a continuous T, the concordance probabilities are well defined with censored data and can be consistently estimated [11], even when censoring is random [12, 13].

As each cut-point on the baseline biomarker is established through a specific dichotomization of survival time, it is natural to expect that the cut-points corresponding to different time points of follow-up may not be the same. For example, for a given disease, the cut-point on a biomarker related to survival time longer than 3 years can be different from that related to survival time longer than 5 years. Multiple survival time-related cut-points, together with their respective sensitivities and specificities, would yield an informative profile for use in evaluating a patient's prognosis.

In this paper, we propose a survival time-related objective function. The criterion is based on a concordance probability related to time-dependent sensitivity and specificity which does not require parametric models. To select a cut-point optimizing the objective function for a given survival time or survival rate, we adopt a non-parametric procedure and conduct a simulation study to examine its finite sample performance. We also apply the method to select survival time-related cut-points on the biomarker of serum bilirubin to be used for prognosis of primary biliary cirrhosis.

2. Method

2.1. *Objective function for optimal survival time-related cut-point*

We consider an extension to an ROC curve-related criterion used for the selection of a cut-point for binary classification. Let $X^{(D)}$ be a continuous measure of biomarker X in group D with distribution functions $F_D(x)$ for $D = 0, 1$. Suppose that the group ($D = 1$) tends to have larger values for X, then for a cut-point x, the specificity $\mathrm{Spe}(x) = P(X^{(D=0)} \leqslant x) = F_0(x)$, and the sensitivity $\mathrm{Sen}(x) = P(X^{(D=1)} > x) = 1 - F_1(x) = S_1(x)$, where $F_1(x)$ is the false negative error rate $\mathrm{FN}(x)$, and $1 - F_0(x)$ is the false positive error rate $\mathrm{FP}(x)$. Among the functions of cut-point x, the most commonly used criterion is the Youden

index, $J(x) = \text{Sen}(x) + \text{Spe}(x) - 1 = 1 - (\text{FN}(x) + \text{FP}(x))$, which takes value in $[-1, 1]$. As Gönen and Sima [8] have shown, maximizing $J(x)$ is equivalent to maximizing the area under the ROC curve for an indicator variable obtained by dichotomizing the continuous biomarker, that is, the area under the two lines connecting the three points of the ROC curve with the two lines starting from the ends of the chance line and meeting at a point associated with the dichotomization. In contrast, the recently proposed criterion $\text{Cc}(x) = P(X^{(0)} \leqslant x < X^{(1)}) = \text{Sen}(x)\ \text{Spe}(x)$ takes value in [0,1]. This concordance probability for binary classification can be expressed as a rectangular area with width and length being the sensitivity and specificity associated with a cut-point on continuous X, such that its vertex $(1 - \text{Spe}(x), \text{Sen}(x))$ lies on the ROC curve. A cut-point maximizing the concordance probability will thus maximize the rectangle [6].

Depending on the distributions of $F_D(x)$, the two criteria based on the Youden index and the concordance probability may or may not produce the same optimum. Between the two, we prefer the latter for the following reasons. For an ROC curve above the chance line, the area under the curve always covers the area represented by $\text{Cc}(x)$; this can be expressed as $\text{Cc}(x) \leqslant P(X^{(0)} < X^{(1)})$ for all possible values of x. In contrast, part of the $J(x)$-related area under the two lines for the dichotomized variable may not always be under the ROC curve, which suggests inconsistency in some cases. Comparing the cut-points x_J and x_C, which maximizes the Youden index and the concordance probability, respectively, the concordance probability $\text{Cc}(x_J)$ is always smaller than $\text{Cc}(x_C)$, except when $x_J = x_C$, in which case $\text{Cc}(x_J) = \text{Cc}(x_C)$. When $X^{(D)} \sim N(\mu_D, \sigma^2)$, the two criteria will have the same optimum, that is, $x_\text{J} = x_\text{C}$, but the variance of the non-parametric cut-point estimator optimizing the Youden index may be much larger than that of the concordance probability optimizer, $\text{Var}\left(\hat{x}_J\right) > \text{Var}\left(\hat{x}_C\right)$ [6].

Dichotomizing survival time T at t for two classes, $D(t) = 0$ if $T > t$ and $D(t) = 1$ if $T \leqslant t$, as used by Heagerty *et al.* [7], the time-dependent sensitivity and specificity for cut-point x on biomarker X are defined as

$$\begin{aligned}\text{Sen}(x \mid t) &= P(X^{(D(t)=1)} > x) = P(X > x \mid T \leqslant t),\\ \text{Spe}(x \mid t) &= P(X^{(D(t)=0)} \leqslant x) = P(X \leqslant x \mid T > t).\end{aligned}$$

Then, the concordance probability for the binary outcome will be

$$\text{Cc}(x \mid t) = P(X^{(D(t)=0)} \leqslant x < X^{(D(t)=1)}) = \text{Spe}(x \mid t)\ \text{Sen}(x \mid t).$$

Because the survival curve describes the distribution of survival time T for a particular disease, $r(t)$, the survival rate or probability of surviving beyond time t is useful for articulating the disease prognosis. Similarly useful is a quantile of the survival time $T(r)$, for example, $T(0.5)$, the median survival time. Therefore, an alternative dichotomization for binary outcome can be based on the survival rate, that is,

$$\begin{aligned}\text{Sen}(x \mid r) &= P(X^{(D(r)=1)} > x) = P(X > x \mid r(T) > r),\\ \text{Spe}(x \mid r) &= P(X^{(D(r)=0)} \leqslant x) = P(X \leqslant x \mid r(T) \leqslant r).\end{aligned}$$

The concordance probability is then

$$\text{Cc}(x \mid r) = P(X^{(D(r)=0)} \leqslant x < X^{(D(r)=1)}) = \text{Sen}(x \mid r)\ \text{Spe}(x \mid r).$$

The optimal cut-point can be obtained by maximizing either $\text{Cc}(x \mid t)$ or $\text{Cc}(x \mid r)$. However, the maximization may not give a stable cut-point estimate, especially when X and T are not highly correlated, the data are subject to censoring, and the sample size is not large. To avoid this problem, we construct the objective function by using locally time-averaged criterion with concordance probabilities in the neighborhood of time t or r,

$$Q(x \mid t) = \int_{s \in H(t)} \text{Cc}(x \mid s) wt(s) ds, \quad \text{or} \quad Q(x \mid r) = \int_{e \in G(r)} \text{Cc}(x \mid e) wr(e) de.$$

Here, H(t) is a set of survival time in the neighborhood of t with length of $LH(t) = \int_{s \in H(t)} wt(s) ds$, and $G(r)$ is a set of survival rate in the neighborhood of r with length of $LG(r) = \int_{e \in G(r)} wr(e) de$, while wt()

and wr() are weight functions. The optimal cut-point X_t for a given t will maximize the objective function over set Ξ containing candidate cut-point values, that is, $X_t = \max_{x\in\Xi} Q(x \mid t)$. Similarly, the optimal cut-point X_r for a given r maximizes the objective function over set Ξ, that is, $X_r = \max_{x\in\Xi} Q(x \mid r)$. The objective function, established from locally time-averaged concordance probability, can help reduce the variation in the cut-point estimates for a specified time point.

When it is necessary to set a constant specificity in the selection process, for example, $\mathrm{Spe}(x \mid t) = 0.8$, the concordance function can be defined by the sensitivity for x satisfying the condition. Particularly, $\mathrm{Cc}(x \mid t) = 0.8\ \mathrm{Sen}(x \mid t)$ for x satisfying $\mathrm{Spe}(x \mid t) = 0.8$, or $\mathrm{Cc}(x \mid r) = 0.8\ \mathrm{Sen}(x \mid r)$ for x satisfying $\mathrm{Spe}(x \mid r) = 0.8$. Consequently, maximizing the objective function $Q(x \mid t)$ or $Q(x \mid r)$ has to be over a subset of Ξ with possible cut-points satisfying the condition.

Once a cut-point x_t for survival time t is obtained, we can define the sensitivity and specificity at the point as $\mathrm{Sen}(x_t) = \int_{t\in H(t)} \frac{\mathrm{Sen}(x_t \mid s)wt(s)}{LH(s)} ds$ and $\mathrm{Spe}(x_t) = \int_{s\in H(t)} \frac{\mathrm{Spe}(x_t \mid s)wt(s)}{LH(s)} ds$, respectively. Alternatively, we can define the sensitivity and specificity at cut-point x_r for survival rate r as $\mathrm{Sen}(x_r) = \int_{e\in G(r)} \frac{\mathrm{Sen}(x_r \mid e)wr(e)}{LG(e)} de$ and $\mathrm{Spe}(x_r) = \int_{e\in G(r)} \frac{\mathrm{Spe}(x_r \mid e)wr(e)}{LG(e)} de$.

Likewise, we define the positive predictive value (PPV) and the negative predictive value (NPV) with cut-point x_t for survival time t as

$$PPV(x_t) = P(T \leqslant t \mid X > x_t) = \int_{s\in H(t)} \frac{\mathrm{Sen}(x_t \mid s)F_T(s)wt(s)}{LH(s)(1 - F_X(x_t))} ds \text{ and}$$

$$NPV(x_t) = P(T > t \mid X \leqslant x_t) = \int_{s\in H(t)} \frac{\mathrm{Spe}(x_t \mid s)(1 - F_T(s))wt(s)}{LH(s)F_X(x_t)} ds,$$

where $F_T()$ and $F_X()$ are the distribution functions of T and X, respectively. Similarly, we can define the predictive values PPV and NPV with cut-point x_r for survival rate r.

2.2. *Estimation of survival time-related cut-point*

A typical dataset will include a baseline biomarker, along with the observed survival time and event indicator of subject i, (X_i, Y_i, d_i), $i = 1, \ldots, n$. Suppose that censoring time U is a random variable. The observed time $Y_i = T_i\ d_i + U_i(1 - d_i)$ with event indicator $d_i = I(T_i \leqslant U_i)$. In theoretical derivation, we assume that the censoring time U is independent of the true survival time T. To minimize the loss of information from the data, we can use Akritas' nearest neighbor estimation method [14], as recommended by Heagerty *et al.* [7], to estimate the bivariate distribution function of (X, T) to obtain consistent estimators of time-dependent sensitivity and specificity while allowing the censoring process to depend on X. Suppose that $\hat{F}(x)$ is the empirical distribution function of X. Using a nearest neighbor kernel function with smoothing parameter λ_n, we can obtain a semi-parametric efficient estimator $\hat{S}_{\lambda n}(x, t)$ for $S(x, t) = P(X > x,\ T > t)$. Let $\hat{S}_{\lambda n}(t) = \hat{S}_{\lambda n}(-\infty, t)$, then the estimators of time-dependent sensitivity and specificity are

$$\hat{\mathrm{Sen}}_{\lambda n}(x \mid t) = \frac{1 - \hat{F}(x) - \hat{S}_{\lambda n}(x, t)}{1 - \hat{S}_{\lambda n}(t)} \quad \text{and} \quad \hat{\mathrm{Spe}}_{\lambda n}(x \mid t) = 1 - \frac{\hat{S}_{\lambda n}(x, t)}{\hat{S}_{\lambda n}(t)},$$

which are monotonically increasing and decreasing with x, respectively. The kernel function and smoothing parameter λ_n can be chosen in the usual way to warrant that ROC curve estimates are consistent and invariant to monotonic transformations of X. Using the available R-function "survivalROC" provided by Heagerty [15] with the option of some kernel functions and $\lambda_n \leqslant O(n^{-1/3})$, we can easily calculate the concordance probability estimator $\hat{\mathrm{Cc}}(x \mid t) = \hat{\mathrm{Sen}}(x \mid t)\hat{\mathrm{Spe}}(x \mid t)$.

To estimate the objective function, we need to define $\mathrm{H}(t)$ or $\mathrm{G}(r)$. Based on the Kaplan–Meier estimate of the survival curve [16], we have the estimate $\hat{r}(t)$ for $r(t)$. Using the estimated standard error (se) of $\hat{r}(t)$ we can have an interval estimate $\left(\hat{r}_l,\ \hat{r}_u\right)$ for $r(t)$, and we can define $\mathrm{H}(t)$ with a time interval (t_l, t_u) where $t_l = \hat{T}\left(\hat{r}_u\right)$ and $t_u = \hat{T}\left(\hat{r}_l\right)$. Alternatively, for the $(1 - r)$th quantile of survival time $T(r)$, we can obtain a point estimate $\hat{T}(r)$ from the estimated survival curve, which is the survival time with estimated survival rate $\hat{r}' = r$. Then, the se $(\hat{r})$-based interval $\left(\hat{r}'_l,\ \hat{r}'_u\right)$ has a corresponding interval for the survival

time, with interval bounds $\hat{T}\left(\hat{r}'_u\right)$ and $\hat{T}\left(\hat{r}'_l\right)$, which can also be estimated using the survival curve. We use the interval to define G(r). Then, with a specified weight function, we can use

$$\hat{Q}(x \mid t) = \sum_{s \in H(t)} \hat{\mathrm{C}}\mathrm{c}(x \mid s) wt(s) \text{ and } \hat{Q}(x \mid r) = \sum_{e \in G(r)} \hat{\mathrm{C}}\mathrm{c}(x \mid e) wr(e)$$

to estimate the objective functions we proposed in Section 2.1. Applying a non-parametric searching procedure, we can find the optimal cut-point maximizing the objective function at a specific survival time t, that is, $\hat{X}_t = \max_{x \in W} \hat{Q}(x \mid t)$. Maximizing the objective function at survival rate r, we have $\hat{X}_r = \max_{x \in W} \hat{Q}(x \mid r)$, where set W contains possible cut-point values. It should be noted that it is likely to underestimate the cut-point when the selection process includes the points with sensitivity estimates of one. In the case when a high survival rate or a short survival time is given for selecting a cut-point, there are few subjects experiencing the event, the sensitivity estimates can be one for some biomarker values. To avoid the underestimation, the selection can be restricted to biomarker values with sensitivity less than the value 1 (e.g., 0.99).

With cut-point estimate $\hat{X}_t$ and $\hat{LH}(t) = \sum_{t_j \in H(t)} wt(t_j)$, we can also estimate the associated sensitivity and the specificity using $\hat{\mathrm{Sen}}\left(\hat{x}_t\right) = \sum_{s \in H(t)} \frac{\hat{\mathrm{Sen}}\left(\hat{x}_t \mid s\right) wt(s)}{\hat{LH}(t)}$ and $\hat{\mathrm{Spe}}\left(\hat{x}_r\right) = \sum_{s \in H(t)} \frac{\hat{\mathrm{Spe}}\left(\hat{x}_t \mid s\right) wt(s)}{\hat{LH}(t)}$, respectively. Furthermore, with the estimated survival rate function $\hat{r}(t)$ and the empirical distribution function $\hat{F}(x)$, we can estimate the PPV by $PPV\left(\hat{x}_t\right) = \sum_{s \in H(t)} \frac{\hat{\mathrm{Sen}}\left(\hat{x}_t \mid s\right)(1-\hat{r}(s)) wt(s)}{\hat{LH}(t)\left(1-\hat{F}\left(\hat{x}_t\right)\right)}$ and estimate the NPV by $NPV\left(\hat{x}_t\right) = \sum_{s \in H(t)} \frac{\hat{\mathrm{Spe}}\left(\hat{x}_t \mid s\right)\hat{r}(s) wt(s)}{\hat{LH}(t)\hat{F}\left(\hat{x}_t\right)}$. With $\hat{X}_r$ and $\hat{LG}(r) = \sum_{t_j \in G(r)} wr(t_j)$, we can estimate the associated sensitivity and the specificity, as well as the predictive values in a similar way.

Heagerty *et al.* pointed out that the bootstrap method [17] can be used to estimate the confidence interval (CI) of a time-dependent ROC curve. Similarly, we propose to use the bootstrap method to evaluate variation in the cut-point estimates. Using a simple random sampling scheme with replacement, we can draw n_b independent bootstrap samples from the study sample. With each bootstrap sample, we estimate the cut-point for a given survival time or survival rate. Based on the empirical distribution of the survival time-related cut-point estimates, we can calculate the summary statistics of mean and standard deviation and construct a 95% CI of the survival time-related cut-point. In particular, we will use the mean as the bootstrap estimator of the cut-point, $\hat{X}_b(t)$ or $\hat{X}_b(r)$, and use the standard deviation for the se of the estimator, $S_b(t)$ or $S_b(r)$. Then, with a reasonably large n_b, we can use $\hat{X}_b(t) \pm 1.96\ S_b(t)$ for the bounds of the 95% CI of X_t or use $\hat{X}_b(r) \pm 1.96\ S_b(r)$ for the 95% CI bounds of X_r.

3. A simulation study

We conducted a simulation study to examine the finite sample performance of the non-parametric cut-point selection procedure, with samples of size $n = 100$ and 150 for censoring proportions of 20% and 40%. For the continuous biomarker X and the true survival time T, we assumed that $\log(X)$ and $\log(T)$ follow bivariate normal distribution with mean $\mu = (0, 1)$, standard deviation of $\sigma = (0.4, 1)$, and correlation coefficient of ρ with the value specified so that the concordance probability $C = P(T_1 < T_2 \mid X_1 > X_2)$ equals to 0.70. For the independent censoring variable U, we assumed that it has a uniform distribution $U(0, \theta)$ with pre-specified values of θ to meet the preset censoring proportion. Let $d = I(T < U)$, taking the value of one if T is observed or zero if T is censored by U. The observed time variable $Y = Td + U(1-d)$. Let $t(r)$ be the time point with survival rate r and $D(r) = I(T \leqslant t(r))$. The true cut-point X_r is calculated by maximizing the concordance probability $P(X^{(D(r)=0)} \leqslant x < X^{(D(r)=1)})$, for $r = 0.3$, 0.4, 0.5, 0.6, 0.7, and 0.8. A 0/1 nearest neighbor kernel function is used in estimating time-dependent sensitivities and specificities with the nearest neighbor method, that is, $K_{\lambda n}(x_i - x_j) = I\left(\left|\hat{F}(x_i) - \hat{F}(x_j)\right| < \lambda_n\right)$ with $2\lambda_n \in (0, 1)$ for the percentage of observations included in each neighborhood. The objective function is evaluated with equal weights and G(r) specified by $\hat{r} \pm 0.67\mathrm{se}\,(\hat{r})$ for the bounds of a 50% CI of $T(r)$. For each setting, we applied the procedure to the 500 generated datasets. We calculated the se of the cut-point estimator and coverage probability by the bootstrap method. Particularly, from the ith dataset ($i = 1, \ldots, 500$), we drew 100 bootstrap samples using simple random sampling with replacement to

Table I. Survival time-related cut-point estimates from 500 simulated datasets ($CA = 0.70$).

		Survival time	Cut-point $X(r)$		
20% censored	Rate r	$\hat{T}(r)$ Bias (SD)	$\hat{X}(r)$ Bias $\left(\sqrt{\text{MSE}}\right)$	$\hat{X}_b(r)$ Bias $\left(\sqrt{\text{MSE}}\right)$	Coverage %
$n = 150$	0.8	−0.0076 (0.1313)	−0.0639 (0.0977)	−0.0530 (0.0802)	90.6
	0.7	0.0042 (0.1640)	−0.0407 (0.0775)	−0.0313 (0.0620)	94.4
	0.6	−0.0051 (0.2040)	−0.0227 (0.0660)	−0.0166 (0.0538)	95.0
	0.5	−0.0085 (0.2707)	−0.0078 (0.0624)	−0.0034 (0.0501)	95.8
	0.4	−0.0064 (0.3606)	0.0066 (0.0634)	0.0094 (0.0510)	95.0
	0.3	−0.0105 (0.5083)	0.0217 (0.0676)	0.0230 (0.0561)	95.2
$n = 100$	0.8	−0.0018 (0.1748)	−0.0580 (0.1060)	−0.0436 (0.0798)	96.0
	0.7	0.0047 (0.2199)	−0.0425 (0.0892)	−0.0284 (0.0661)	96.4
	0.6	−0.0047 (0.2765)	−0.0244 (0.0752)	−0.0140 (0.0583)	97.0
	0.5	0.0088 (0.3613)	−0.0079 (0.0707)	−0.0007 (0.0552)	97.4
	0.4	−0.0063 (0.4446)	0.0081 (0.0713)	0.0123 (0.0569)	96.6
	0.3	−0.0077 (0.6176)	0.0216 (0.0782)	0.0250 (0.0636)	96.2
40% censored					
$n = 150$	0.8	−0.0048 (0.1413)	−0.0645 (0.0987)	−0.0543 (0.0811)	90.8
	0.7	0.0042 (0.1828)	−0.0413 (0.0795)	−0.0336 (0.0643)	94.0
	0.6	0.0046 (0.2333)	−0.0211 (0.0669)	−0.0172 (0.0551)	95.2
	0.5	0.0080 (0.3031)	−0.0051 (0.0653)	−0.0022 (0.0518)	96.6
	0.4	−0.0022 (0.4032)	0.0076 (0.0657)	0.0116 (0.0537)	96.6
	0.3	0.0332 (0.6706)	0.0213 (0.0732)	0.0244 (0.0600)	97.4
$n = 100$	0.8	−0.0010 (0.1763)	−0.0585 (0.1076)	−0.0440 (0.0804)	96.0
	0.7	−0.0079 (0.2249)	−0.0414 (0.0888)	−0.0294 (0.0668)	96.2
	0.6	0.0008 (0.2915)	−0.0240 (0.0794)	−0.0151 (0.0587)	97.4
	0.5	−0.0111 (0.3850)	–0.0086 (0.0717)	−0.0020 (0.0551)	97.8
	0.4	0.0041 (0.4905)	0.0075 (0.0737)	0.0123 (0.0574)	97.4
	0.3*	0.0190 (0.7914)	0.0280 (0.0838)	0.0289 (0.0659)	97.8

SD, standard deviation; MSE, mean squared error.
Bias = Mean − True.
*One dataset and some bootstrap samples were excluded as they yielded least survival rate estimates above 0.3.

obtain 100 cut-point estimates for X. We used the bootstrap mean $\hat{X}_b(r)$ and bootstrap standard deviation $S_b(r)$ to construct a 95% CI with bounds $\hat{X}_b(r) \pm 1.96\ S_b(r)$. We estimated the coverage probability using the proportion of the 95% CIs that covered the true cut-point values.

Table I presents the bias (= Mean – True) and square root of mean squared error with the proposed estimator $\hat{X}(r)$ and the bootstrap estimator $\hat{X}_b(r)$ and the estimated coverage probability in each of the four settings with specified censoring proportion and sample size. As expected, the mean squared errors decrease with increased sample size and decreased censoring proportion, while the pattern is not apparent with the magnitude of bias. The cut-point estimators for median survival time have consistently smaller bias and mean squared error than the cut-point estimators for the other survival time. Compared to $\hat{X}(r)$, the bootstrap estimator $\hat{X}_b(r)$ performs better, as it has consistently smaller mean squared errors. Except for $r = 0.8$, the bias of $\hat{X}_b(r)$ is not large, and the estimated coverage probability by the bootstrap method ($n_b = 100$) has a reasonable range of $94.0 \sim 97.8\%$. In the case with an increased sample size ($n = 150$) and decreased censoring proportion (20%), the estimated coverage probability is improved to be $94.4 \sim 95.8\%$ for $0.3 \leqslant r \leqslant 0.7$.

In each of the four settings, the mean squared error of the estimated cut-point is found to be the largest for $r = 0.8$ and becomes smaller for lower survival rates. This is consistent with the fact that, given a high survival rate of 0.8, only a small number of subjects experiencing the event can contribute to the estimation of the sensitivity for the objective function, resulting in a larger bias and mean squared error

with the cut-point estimates. On the other hand, when the smallest survival rate estimate is above specified survival rate r_0, the cut-point for r_0 is not estimable. In our case with sample size of 100 and censoring proportion of 40%, one dataset and some of the bootstrap samples are excluded in estimating cut-point for $r_0 = 0.3$ when the lowest survival rate estimates are above 0.3. When the least survival rate estimate is below but close to 0.3, as seen in some of the samples with 40% censored data, the interval of G(0.3) is narrower, resulting in a relatively larger mean squared error.

Although the estimable survival time-related cut-point is limited by the range of estimable survival rates, the procedure had a fairly good finite sample performance in cut-point estimation, particularly, in the case where the bootstrap estimator $\hat{X}_b(r)$ is employed. The simulation study also suggests that the use of a larger sample is helpful when the censoring proportion is not small. For 40% censoring, $n \geqslant 150$ is preferred.

4. Application

We applied the methods to a dataset of 418 patients with primary biliary cirrhosis, presented in Fleming and Harrington [18], to estimate survival time-related cut-points on the prognostic biomarker of serum bilirubin. In the sample, the patients ranged in age between 26 and 78 years, with a mean age of 51 years; 89.5% were female. The longest follow up time was 13.13 years. During the follow up period, 38.5% of the patients ($n = 161$) died. Figure 1 displays the estimated survival curve along with the 95% confidence band, while Figure 2 shows the number of patients at risk and cumulative deaths over time. At baseline, blood samples were taken from all patients, and a set of serum variables including bilirubin was measured. The serum bilirubin levels ranged between 0.3 and 28 mg/dl with a median of 1.40 mg/dl and had a skewed distribution (Figure 3). In the prediction of prognosis, higher serum bilirubin was related to shorter survival time; the discrimination accuracy $C = P(T_1 < T_2 \mid X_1 > X_2)$ had an estimate of 0.7482 and $C^* = P(X_1 < X_2 \mid T_1 > T_2) = 0.7507$.

Because the lowest survival rate estimate of 0.3534 was lower than 0.4, we applied the proposed non-parametric procedure to estimate the cut-point on the biomarker for survival rates ranging between 0.4 and 0.9. We used the bootstrap method to estimate the survival time-related cut-points and 95% CI. Specifically, we used a simple random sampling scheme with replacement to draw 500 bootstrap samples from the study sample. With each sample, we estimated cut-points for survival rates of 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, and cut-points for survival times of 3 and 5 years. Based on the empirical distribution of the cut-point estimates, we calculated the bootstrap estimate and 95% CI for each survival time-related cut-point. With each estimated cut-point, we also calculated the associated sensitivity, specificity, and predictive values.

Table II displays cut-point estimates for specified survival rates and times using the bootstrap method, as well as estimated survival time quantiles and survival rates for given times. Also listed are the sensitivities, specificities, and the predictive values for the selected cut-points. The estimated cut-point of serum bilirubin (mg/dl) decreased from 2.33 to 1.40 mg/dl with increasing time for survival rates decreasing from 0.9 to 0.4. The associated sensitivity estimates were in the range of 0.6811–0.7435, and the speci-

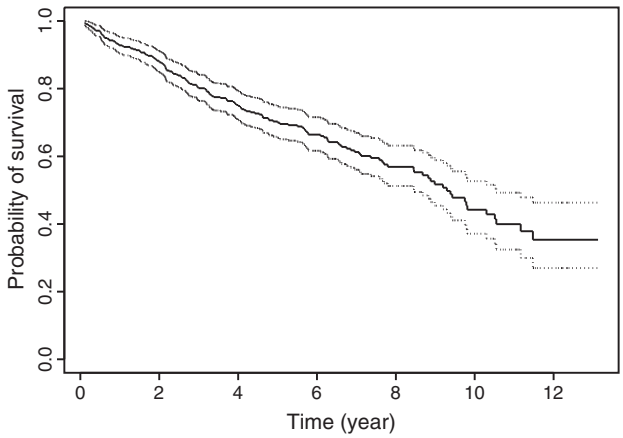

Figure 1. The estimated survival curve with 95% confidence band for of 418 patients with primary biliary cirrhosis.

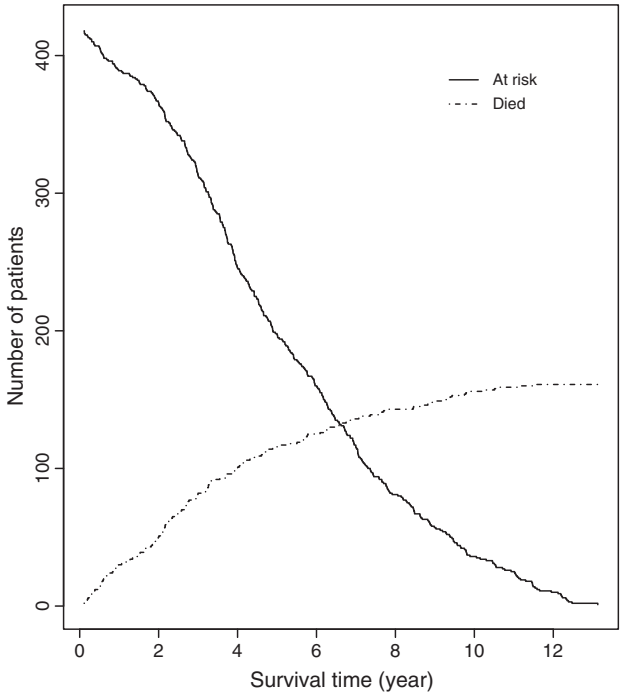

Figure 2. Number of patients at risk and cumulative number of deaths over time.

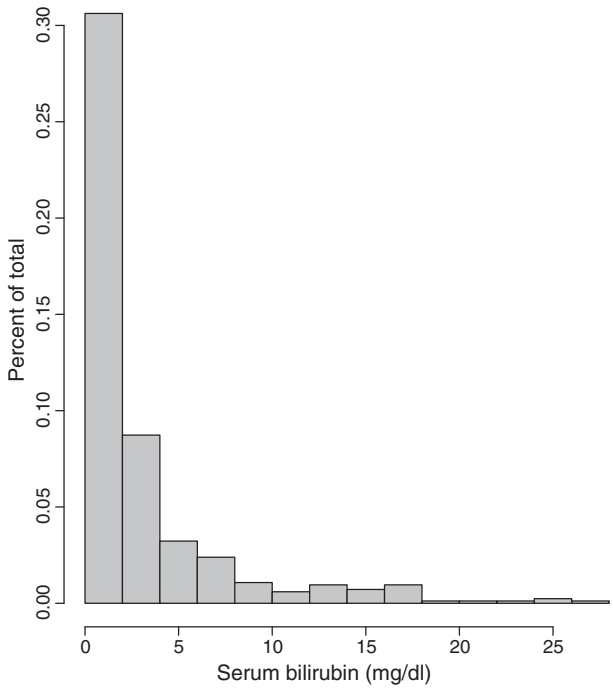

Figure 3. Histogram of baseline serum bilirubin.

ficity estimates were in the range of 0.6851–0.7650, suggesting that these cut-points could be useful. The NPVs were all above 0.61, indicating that the patients with serum bilirubin below the cut-points may have good probabilities to survive beyond the expected time. The PPVs of the cut-points were above 0.61 for the survival rates of 0.6 or lower, suggesting that fair predictions of death for those with serum bilirubin above the cut-points worked only for longer time periods.

As information on median survival time is often used for disease prognosis, when a single cut-point is required, we recommend using a cut-point estimated for the median survival time. The estimated median

Table II. Cut-point estimates on serum bilirubin for prognosis of primary biliary cirrhosis.

Survival		Bootstrapping ($n_b = 500$)		Sensitivity	Specificity	Predictive values	
Rate	Time (95% CI)	$\hat{X}_b(r)$ $\left(\text{se}\hat{X}_b(r)\right)$	95% CI_b	Sen $\left(\hat{X}_b(r)\right)$	Spe $\left(\hat{X}_b(r)\right)$	PPV $\left(\hat{X}_b(r)\right)$	NPV $\left(\hat{X}_b(r)\right)$
0.9	1.68 (0.98, 2.11)	2.331 (0.414)	1.521, 3.142	0.6962	0.6851	0.2074	0.9352
0.8	3.15 (2.48, 3.89)	2.139 (0.219)	1.710, 2.568	0.7343	0.6990	0.3970	0.8843
0.7	5.01 (4.00, 6.22)	1.897 (0.203)	1.498, 2.295	0.7435	0.7045	0.5369	0.8439
0.6	7.12 (6.09, 8.68)	1.748 (0.224)	1.309, 2.187	0.6853	0.6884	0.6175	0.7411
0.5	9.36 (7.70, 10.47)	1.482 (0.256)	0.980, 1.983	0.6839	0.7185	0.7198	0.6835
0.4*	11.12 (9.39, 13.13)	1.390 (0.230)	0.939, 1.841	0.6811	0.7650	0.8112	0.6170
Time	Rate (95% CI)	$\hat{X}_b(t)$ $\left(\text{se}\hat{X}_b(t)\right)$		Sen $\left(\hat{X}_b(t)\right)$	Spe $\left(\hat{X}_b(t)\right)$		
3 years	0.801 (0.764, 0.841)	2.140 (0.216)	1.716, 2.563	0.7333	0.6982	0.3924	0.8856
5 years	0.699 (0.654, 0.748)	1.896 (0.204)	1.497, 2.295	0.7435	0.7045	0.5369	0.8439

CI, confidence interval; PPV, positive predictive value; NPV, negative predictive value; se, standard error.
*Excluding 101 datasets that had least survival rate estimates above 0.4.

survival time in the study sample was 9.36 years, and the estimated cut-point on serum bilirubin for median survival time was 1.48 mg/dl with 95% CI = (0.98, 1.98) mg/dl. The specificity of 0.684 suggested that patients who survived for 9.36 or more years are likely to have serum bilirubin below the cut-point, while sensitivity of 0.716 implied that patients who died before 9.36 years tend to have higher levels of serum bilirubin. The PPV of 0.72 quantified the risk of death by the median survival time for the patients with serum bilirubin above the cut-point; while the NPV of 0.68 estimated probability of survival beyond the time for the patients with serum bilirubin at or below the cut-point. Also useful is the 5-year survival rate and the related cut-point. At the survival time of 5 years, the survival rate estimate was very close to 0.7, and the related cut-point estimate was 1.90 mg/dl with 95% CI = (1.50, 2.29) mg/dl. The associated sensitivity and specificity both were above 0.7. The NPV was about 0.84, while the PPV was low (0.54), indicating that the cut-point was useful in predicting survival beyond 5 years but not good enough in predicting risk of death within 5 years.

5. Discussion

Instead of assuming a single cut-point model, we considered a more realistic scenario where a continuous prognostic biomarker has a relationship with survival time. To select a survival time-related cut-point on the biomarker with censored data, we used an objective function based on the extension of a selection criterion using concordance probability for binary classification, which incorporated time-dependent sensitivity and specificity. Using a nearest neighbor estimation method to estimate time-dependent sensitivity and specificity with censored data, we evaluated the objective function and searched for the optimal cut-point. The non-parametric procedure seemed to work well in selecting survival time-related cut-points and yielded satisfactory results in a simulation study.

The procedure is flexible with choices of different weight functions and sets of H(t) or G(r) for the objective function using locally time-averaged concordance probability. Although there are many options for the weight function, we used a constant weight for computational simplicity. We used the se-based interval for H(t) and G(r) because it can be interpreted as a CI for $r(t)$ or $T(r)$. The wider the interval, the more observed time points within the interval would be used for the objective function, which would not only increase the computational cost but, in some cases, would also introduce noise. For a wider interval, using a weight function that has more weights on the points around the center helps to reduce the impact of the points closer to the bounds.

In the simulation study, we used $n_b = 100$ for the bootstrap procedure to save computational time. To better estimate the variance of the cut-point estimator, however, it is recommended to use a large n_b. In the application we used $n_b = 500$.

Using a prognostic biomarker to predict disease prognosis, an informative profile should include the estimates of the survival time quantile for a given r or the estimated survival rate for a given t in a certain range. It should include the related cut-point estimates and their CI estimates. It would be informative as well to provide sensitivities, specificities, and predictive values associated with the selected cut-points. If

the cut-point estimates have similar values over a wide range of survival times or survival rates, the use of a single cut-point would be effective.

There are some limitations to the method. First, the estimated cut-point for a higher survival rate may have larger variation as shown in the simulation study. Secondly, using censored data, the cut-point for a low survival rate or a long survival time may not always be estimable.

A useful survival time-related cut-point should correspond to a clinically meaningful dichotomization of survival time (or rate) for 'good' versus 'poor' prognosis. As median survival time is often used for disease prognosis, the cut-point for a survival rate of 0.5 would be of clinical interest. To estimate it, investigators should make efforts to follow patients for a sufficient amount of time and reduce dropout rates so that the median survival time can be estimated precisely. When the cut-point is only estimable for survival rates above 0.5, a profile with a narrow range of survival rates would still be useful if some clinically meaningful survival rate is in the range. Similarly, if a survival time (for example, 5 years) is of clinical interest, then to estimate the cut-point for the time, we would better collect data far beyond that time point.

Acknowledgements

We thank the reviewers and associate editor for their helpful comments to improve the manuscript.

References

1. Zhou XH, Obuchowski NA, McClish DK. *Statistical Methods in Diagnostic Medicine*. Wiley: New York, 2002.
2. Pepe MS. *The Statistical Evaluation of Medical Tests for Classification and Prediction*. Oxford University Press: New York, 2003.
3. Youden WJ. Index for rating diagnostic tests. *Cancer* 1950; **3**:32–35.
4. Fluss R, Faraggi D, Reiser B. Estimation of the Youden index and its associated cutoff point. *Biometrical Journal* 2005; **47**(4):458–472.
5. Perkins NJ, Schisterman EF. The inconsistency of "optimal" cut-points using two ROC based criteria. *American Journal of Epidemiology* 2006; **163**(7):670–675.
6. Liu X. Classification accuracy and cut-point selection. *Statistics in Medicine* 2012; **31**:2676–2686.
7. Heagerty PJ, Lumley T, Pepe MS. Time-dependent ROC curves for censored survival data and a diagnostic marker. *Biometrics* 2000; **56**:337–344.
8. Gönen M, Sima C. Optimal cutpoint estimation with censored data. *Journal of Statistical Theory and Practice* 2013; **7**: 345–359.
9. Harrell FE, Lee KL, Calife RM, Pryor DB, Rosati RA. Regression modeling strategies for improved prognostic prediction. *Statistics in Medicine* 1984; **3**(2):143–152.
10. Harrell FE, Lee KL, Mark DB. Tutorial in biostatistics multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in Medicine* 1996; **15**(4):361–387.
11. Pencina MJ, D'Agostino RB. Overall C as a measure of discrimination in survival analysis: model specific population value and confidence interval estimation. *Statistics in Medicine* 2004; **23**:2109–2123.
12. Gönen M, Heller G. Concordance probability and discriminative power in proportional hazards regression. *Biometrika* 2005; **92**:965–970.
13. Liu X, Jin Z, Graziano JH. Comparing paired biomarkers in predicting quantitative health outcome subject to random censoring. *Statistical Methods in Medical Research* 2012. DOI: 10.1177/0962280212460434.
14. Akritas MG. Nearest neighbor estimation of a bivariate distribution under random censoring. *Annals of Statistics* 1994; **22**:1299–1327.
15. Heagerty PJ. Time-dependent ROC curve estimation from censored survival data, 2014. http://cran.r-project.org/web/packages/survivalROC/survivalROC.pdf.
16. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association* 1958; **53**:457–481.
17. Efron B, Tibshirani RJ. *An Introduction to the Bootstrap*. Chapman & Hall: New York, 1993.
18. Fleming TR, Harrington D. *Counting Processes and Survival Analysis*. Wiley: New York, 1991.