Published in final edited form as:

Stat Med. 2020 August 15; 39(18): 2437–2446. doi:10.1002/sim.8547.

Predictive accuracy of markers or risk scores for interval censored survival data

Yuan Wu¹, Xiaofei Wang¹, Jiaying Lin¹, Beilin Jia², Kouros Owzar¹

¹Department of Biostatistics and Bioinformatics, Duke University, Durham, North Carolina

²Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina

Abstract

Methods for the evaluation of the predictive accuracy of biomarkers with respect to survival outcomes subject to right censoring have been discussed extensively in the literature. In cancer and other diseases, survival outcomes are commonly subject to interval censoring by design or due to the follow up schema. In this article, we present an estimator for the area under the time-dependent receiver operating characteristic (ROC) curve for interval censored data based on a non-parametric sieve maximum likelihood approach. We establish the asymptotic properties of the proposed estimator and illustrate its finite-sample properties using a simulation study. The application of our method is illustrated using data from a cancer clinical study. An open-source R package to implement the proposed method is available on Comprehensive R Archive Network.

Keywords

area under ROC curve; interval censoring; joint distribution; Sieve estimation

1 | INTRODUCTION

The receiver operator curve (ROC)¹ serves as an established and widely used tool for visual assessment of the sensitivity and specificity of diagnostic tests, and of risk and prediction scores derived from machine learning applications.² While originally developed for binary outcomes, the ROC concept has been extended to the evaluation of the predictive accuracy of tests and markers with respect to time-to-event outcomes.³ In this context, a time-dependent analog of the ROC is constructed on the basis of corresponding time-dependent analogs of sensitivity and specificity.

Correspondence Yuan Wu, Department of Biostatistics and Bioinformatics, Duke University, 2424 Erwin Rd, Durham, NC 27705 USA. yuan.wu@duke.edu.

DATA ACCESSIBILITY

The data for the analysis of CALGB 30801 is from Alliance. Restrictions apply to the availability of this data set, which was used under license for CALGB 30801. This data set is available from the corresponding author with the permission of Alliance.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

There exists a rich body of literature for estimation of the time-dependent ROC when the time-to-event outcome is subject to a right-censoring mechanism.³⁻⁵ The application of these methods is limited in many cancer studies as the corresponding time-to-event outcomes of interest are invariably subject to interval-censoring mechanisms by virtue of clinical practice or study design. To be more specific, we consider Cancer and Leukemia Group B (CALGB) 30801, a randomized double-blinded phase III study evaluating the role of selective cyclooxygenase-2 (COX-2) inhibition in patients with advanced non-small cell lung cancer.⁶ The primary endpoint for this study was progression-free survival (PFS). Patients were monitored until the first confirmed assessment of tumor progression. The actual time of this event is not observable as it is realized between two consecutive assessments dates. Similarly, toxicity events for this and most cancer studies are not reported in real time but rather by dosing cycles. Neutropenia, defined as an abnormally low count of neutrophils, is a common toxicity associated with many chemotherapy agents, including gemcitabine, pemetrexed, or carboplatin, the agents used in CALGB 30801. A high-grade neutropenia event is defined when the absolute neutrophil count falls below 1000 cells per microliter of blood. As neutrophil counts are typically measured right before the chemotherapy dose is administered, the time of this event, when the neutrophil count crosses the critical threshold, is not observable. What are observed are the date of the first cycle at which the patient's count as observed to be below the threshold, along with the date of the previous cycle when the count was recorded to be above the threshold. Consequently, the actual time of toxicity is not observable and effectively interval censored between the dates of two consecutive drug cycles.

A naïve approach for estimating the ROC in the presence of an interval-censored mechanism is to impute the event time using for example the midpoint or the right end of the last observed time interval. While this approach is convenient, in the sense that it allows for re-purposing methods developed for right-censored data, it is biased. Li and Ma⁷ proposed a nonparametric approach for estimating the ROC and area under ROC curve (AUC) in the presence of interval censoring. To estimate the curve at time say $t > 0$, they exclude data from any patient whose last observed time interval contains t . The authors formally quantify the loss of information using a fraction and point out that as this fraction increases, the accuracy of the estimator decreases, and the variance is inflated. Jacqmin-Gadda et al⁸ proposed two approaches to estimate the time-dependent ROC and AUC in the context of semi-competing time-to-event outcomes subject to interval censoring. Their first approach is a fully model based extension of the well-known Cox regression type illness-death model for modeling a marker effect on the two competing risks. Their second approach also uses results from the illness-death model to impute the probability of subjects becoming diseased before time t under interval censoring. These two approaches could be potentially applied for interval censored single event case. But obviously the misspecification of the illness-death model is very likely to introduce estimation bias in the single event case.

In this article, we propose a non-parametric approach for estimating the time-dependent ROC and AUC when the outcome is subject to an interval-censoring mechanism. Our approach is summarized as follows. Let T denote the time of the event of interest and M denote a quantitative marker whose predictive performance with respect to T is to be

assessed on the basis of the time-dependent ROC. T is subject to interval censoring and M is assumed to be observable.

We first adopt a sieve spline approach to estimate the joint distribution of (T, M) and the corresponding marginal distribution of M . That is, the joint and marginal functions are restricted in spline function classes (sub-sets of nonparametric function classes) for their maximum likelihood estimation (MLE). The resulting estimates are then used to produce plug-in estimates of the time-dependent sensitivity and specificity functions, which are in turn used to produce a plug-in estimator of the time-dependent ROC function.

The article is organized as follows. In the next section, we outline the technical considerations for our proposed method. Thereafter, we illustrate its finite-sample operating characteristics using a simulation study. Finally, we present an analysis assessing the performance of COX2 and Prostaglandin E metabolite (pGEM) as predictive biomarkers for PFS in advanced non-small cell lung cancer based on data from CALGB 30801 and conclude the article with a discussion. The theoretical results are developed in Web Appendix 1. Specifically, we show that for each t in the support of the censoring time, this plug-in ROC estimator is uniformly consistent on the support of the continuous marker, and for each t , the corresponding AUC estimator is consistent. An open-source R⁹ extension package, *intcensROC*,¹⁰ to implement the proposed method is available through the Comprehensive R Archive Network (CRAN). The scripts to replicate the results from the simulation study using this package are included as online supplementary material.

2 | METHODS

2.1 | Sieve estimators for the ROC curve and the corresponding AUC

In this section, we outline a spline-based sieve MLE approach for estimation of the joint distribution of the event time T and marker M . Once this estimate is obtained, we construct plug-in estimators for the ROC curve and the AUC at time $t > 0$ based on the following definitions as given in Reference ³:

$$ROC_t(p) = TP_t\{FP_t^{-1}(p)\} \text{ and } AUC_t = \int_0^1 ROC_t(p)dp, \quad (1)$$

where $TP_t(\cdot)$ and $FP_t(\cdot)$ denote the true positive and false positive functions evaluated at time t and are defined as

$$TP_t(m) = \Pr(M > m \mid T \leq t) = \frac{F_1(t) - F(t, m)}{F_1(t)}$$

and

$$FP_t(m) = \Pr(M > m \mid T > t) = \frac{1 - F_1(t) - F_2(m) + F(t, m)}{1 - F_1(t)},$$

where $F(\cdot, \cdot)$, $F_1(\cdot)$, and $F_2(\cdot)$ denote the joint distribution function of (T, M) , the marginal distribution function of T , and the marginal distribution of M , respectively. If we assume that M has a finite support of $[0, \tau_M]$, τ_M is a positive finite constant, then $F_1(t) = F(t, \tau_M)$. Hence, we have

$$TP_t(m) = \frac{F(t, \tau_M) - F(t, m)}{F(t, \tau_M)} \quad \text{and} \quad FP_t(m) = \frac{1 - F(t, \tau_M) - F_2(m) + F(t, m)}{1 - F(t, \tau_M)}.$$

It is supposed that the event time T is interval censored by observation times U and V and that the marker M is observable. (U, V) is assumed to be independent of (T, M) . What is observed for patient $i \in \{1, \dots, n\}$ is the sextuple $(u_i, v_i, m_i, \delta_i^{(1)}, \delta_i^{(2)}, \delta_i^{(3)})$ where u_i and v_i are the observation times, m_i is the observed marker value, and $\delta_i^{(1)} = 1_{[t_i \leq u_i]}$, $\delta_i^{(2)} = 1_{[u_i < t_i \leq v_i]}$, and $\delta_i^{(3)} = 1_{[t_i > v_i]}$ are the event indicators for left, interval, and right censoring, respectively. In these event definitions, t_i denotes the latent event time for patient i . Note that U and V could be two random observation times or result from a group censoring mechanism. The reader is referred to Reference ¹¹ for a detailed account on interval censoring.

By virtue of the independence assumption between (U, V) and (T, M) , the likelihood is reduced to

$$\Pi_i^n = 1 \left\{ \frac{\partial F(u_i, m_i)}{\partial m} \right\}^{\delta_i^{(1)}} \left\{ \frac{\partial F(v_i, m_i)}{\partial m} - \frac{\partial F(u_i, m_i)}{\partial m} \right\}^{\delta_i^{(2)}} \left\{ \frac{dF_2(m_i)}{dm} - \frac{\partial F(v_i, m_i)}{\partial m} \right\}^{\delta_i^{(3)}}. \quad (2)$$

As discussed in Reference ¹², a purely nonparametric MLE approach for optimizing Equation (2) is both computationally and theoretically challenging. To optimize the likelihood, we propose to use spline-based sieve approach. In addition to assuming that $M \in [0, \tau_M]$, suppose that $T \in [0, \tau_T]$ for a positive constant τ_T . On these bounded intervals, we construct two sets of B-splines of order l : $\{B_j^{(1), l}(t)\}_{j=1}^{p_n}$ with knot sequence $\tilde{\xi}$ as

$$\tilde{\xi} = \left\{ (\xi_j)_{j=1}^{p_n+l} : 0 = \xi_1 = \dots = \xi_l < \xi_{l+1} < \dots < \xi_{p_n} < \xi_{p_n+1} = \dots = \xi_{p_n+l} = \tau_T \right\},$$

and $\{B_k^{(2), l}(m)\}_{k=1}^{q_n}$ with the knot sequence $\tilde{\eta}$ as

$$\tilde{\eta} = \left\{ (\eta_k)_{k=1}^{q_n+l} : 0 = \eta_1 = \dots = \eta_l < \eta_{l+1} < \dots < \eta_{q_n} < \eta_{q_n+1} = \dots = \eta_{q_n+l} = \tau_M \right\},$$

where p_n and q_n are both positive integers dependent on the sample size n . Let

$$F_n(t, m) = \sum_{j=1}^{p_n} \sum_{k=1}^{q_n} \alpha_{j,k} B_j^{(1),l}(t) B_k^{(2),l}(m) \quad (3)$$

and

$$F_{n,2}(m) = \sum_{k=1}^{q_n} \beta_k B_k^{(2),l}(m), \quad (4)$$

be the joint and marginal distribution functions for (T, M) restricted to classes of spline functions. As F_n and $F_{n,2}$ are distribution functions, for any $t \in [0, \tau_T]$ and any $m \in [0, \tau_M]$, we have that $F_n(0, m) = F_n(t, 0) = F_{n,2}(0) = 0$. Lemma 1 in Web Appendix 1 establishes that the spline constraints in (2.3) in Reference ¹² can be reduced to

$$\begin{aligned} \alpha_{j,1} &= 0 \text{ for } j = 1, \dots, p_n, \\ \alpha_{1,k} &= 0 \text{ for } k = 2, \dots, q_n, \\ (\alpha_{j+1,k+1} - \alpha_{j+1,k}) - (\alpha_{j,k+1} - \alpha_{j,k}) &\geq 0 \\ &\text{for } j = 1, \dots, p_n - 1, k = 1, \dots, q_n - 1, \\ \beta_1 &= 0, \\ (\beta_{k+1} - \beta_k) - (\alpha_{p_n,k+1} - \alpha_{p_n,k}) &\geq 0 \text{ for } k = 1, \dots, q_n - 1, \\ \beta_{q_n} &\leq 1. \end{aligned} \quad (5)$$

Substitute $F_n(\cdot, \cdot) = R(\cdot, \cdot)$ and $F_{n,2}(\cdot) = F_2(\cdot)$ in Equation (2), by Equations (3) and (4), we obtain the following spline-based log likelihood function

$$\begin{aligned} \bar{l}_n(\alpha, \beta) &= \sum_{i=1}^n \left[\delta_i^{(1)} \log \frac{\partial \sum_{j=1}^{p_n} \sum_{k=1}^{q_n} \alpha_{j,k} B_j^{(1),l}(u_i) B_k^{(2),l}(m_i)}{\partial m} \right. \\ &\quad + \delta_i^{(2)} \log \frac{\partial \left\{ \sum_{j=1}^{p_n} \sum_{k=1}^{q_n} \alpha_{j,k} B_j^{(1),l}(v_i) B_k^{(2),l}(m_i) - \sum_{j=1}^{p_n} \sum_{k=1}^{q_n} \alpha_{j,k} B_j^{(1),l}(u_i) B_k^{(2),l}(m_i) \right\}}{\partial m} \\ &\quad + \delta_i^{(3)} \log \frac{\partial \left\{ \sum_{k=1}^{q_n} \beta_k B_k^{(2),l}(m_i) - \sum_{j=1}^{p_n} \sum_{k=1}^{q_n} \alpha_{j,k} B_j^{(1),l}(v_i) B_k^{(2),l}(m_i) \right\}}{\partial m} \\ &\quad \left. \right], \end{aligned} \quad (6)$$

where $\alpha = \{\alpha_{j,k}\}_{j=1, \dots, p_n, k=1, \dots, q_n}$ and $\beta = \{\beta_k\}_{k=1, \dots, q_n}$. In the proposed sieve MLE approach, nonparametric distribution functions are restricted to classes of spline functions for their estimation. This is equivalent to finding the maximizer $(\hat{\alpha}, \hat{\beta})$ for (6) subject to the constraints in Equation (5). By plugging $(\hat{\alpha}, \hat{\beta})$ into Equations (3) and (4), we obtain the sieve MLE for $(F_0(\cdot, \cdot), F_{0,2}(\cdot))$, the true distribution functions for (T, M) .

Using the fact that the sieve MLE of the joint distribution of the event time and the marker is consistent, in Web appendix 1, we prove that the proposed ROC and AUC estimators are consistent.

2.2 | Computing the sieve MLE

Given that the B-spline based sieve MLE approach for Equation (6) involves complicated constraints (5). Similar to the approach used in Reference ¹², we propose to use I-splines and its derivatives to simplify the computation.

Let I_j^{l-1} and M_j^{l-1} denote the j th I-spline and M-spline bases, respectively, as defined in References ¹⁴ and ¹³, where $M_j^{l-1}(t) = \frac{dI_j^{l-1}(t)}{dt}$. ¹² showed that $I_j^{l-1}(t) = \sum_{h=j+1}^{p_n} B_h^l(t)$ for $j=0, \dots, p_n-1$. Note that for notational convenience, we let the I-spline index start from 0 instead of from 1. Here, both B_j^l and I_j^{l-1} are of degree $l-1$, but M_j^{l-1} is of degree $l-2$. Some basic algebra reveal that $F_n(\cdot, \cdot)$ and $F_{n,2}(\cdot)$ given by Equations (3) and (4) with constraints (5) are equivalent to

$$F_n(t, m) = \sum_{j=1}^{p_n-1} \sum_{k=1}^{q_n-1} \gamma_{j,k} I_j^{(1),l-1}(t) I_k^{(2),l-1}(m), \quad (7)$$

and

$$F_{n,2}(m) = \sum_{k=1}^{q_n-1} \left\{ \sum_{j=1}^{p_n-1} \gamma_{j,k} + \omega_k \right\} I_k^{(2),l-1}(m), \quad (8)$$

subject to the constraints

$$\begin{aligned} &\gamma_{j,k} \geq 0 \text{ for } j=1, \dots, p_n-1, k=1, \dots, q_n-1, \\ &\omega_k \geq 0, k=1, \dots, q_n-1, \\ &\sum_{j=1}^{p_n-1} \sum_{k=1}^{q_n-1} \gamma_{j,k} + \sum_{k=1}^{q_n-1} \omega_k \leq 1, \end{aligned} \quad (9)$$

where in Equations (7) and (8) $I_j^{(1),l-1}(t) = \sum_{h=j+1}^{p_n} B_h^{(1),l}(t)$ for $j=0, 1, \dots, p_n-1$ and $I_k^{(2),l-1}(t) = \sum_{h=k+1}^{p_n} B_h^{(2),l}(t)$ for $k=0, 1, \dots, q_n-1$.

By Equations (7), (8), and $M_j^{l-1}(t) = \frac{dI_j^{l-1}(t)}{dt}$, we rewrite the B-spline-based log likelihood in Equation (6) as

$$\begin{aligned}
\bar{l}_n(\gamma, \omega) = & \sum_{i=1}^n \left[\delta_i^{(1)} \log \left\{ \sum_{j=1}^{p_n-1} \sum_{k=1}^{q_n-1} \gamma_{j,k} I_j^{(1),l-1}(u_i) M_k^{(2),l-1}(m_i) \right\} \right. \\
& + \delta_i^{(2)} \log \left\{ \sum_{j=1}^{p_n-1} \sum_{k=1}^{q_n-1} \gamma_{j,k} I_j^{(1),l-1}(v_i) M_k^{(2),l-1}(m_i) \right. \\
& - \sum_{j=1}^{p_n-1} \sum_{k=1}^{q_n-1} \gamma_{j,k} I_j^{(1),l-1}(u_i) M_k^{(2),l-1}(m_i) \left. \right\} \\
& + \delta_i^{(3)} \log \left\{ \sum_{k=1}^{q_n-1} \left(\sum_{j=1}^{p_n-1} \gamma_{j,k} + \omega_k \right) M_k^{(2),l}(m_i) \right. \\
& \left. \left. - \sum_{j=1}^{p_n-1} \sum_{k=1}^{q_n-1} \gamma_{j,k} I_j^{(1),l-1}(v_i) M_k^{(2),l-1}(m_i) \right\} \right],
\end{aligned} \tag{10}$$

where $\gamma = \{\gamma_{j,k}\}_{j=1, \dots, p_n-1, k=1, \dots, q_n-1}$, $\omega = \{\omega_k\}_{k=1, \dots, q_n-1}$.

Now, the proposed sieve MLE problem is equivalent to finding the maximizer $(\hat{\gamma}, \hat{\omega})$ for Equation (10) subject to the simpler set of constraints (9). The optimization can be efficiently implemented using the generalized gradient projection algorithm.^{12,15,16}

The spline knot sequence for the event time component is chosen based on the observed times $\{(u_i, v_i)\}_{i=1}^n$. Specifically, we first let $\mathcal{O} = \left\{ u_i \delta_i^{(1)} + \frac{u_i + v_i}{2} \delta_i^{(2)} + v_i \delta_i^{(3)} \right\}_{i=1}^n$, that is, each member of $\mathcal{O}$ equals u_i for left censoring, $(u_i + v_i)/2$ for interval censoring and v_i for right censoring. Then we let the number of the interior knots be $\lceil n^{1/3} \rceil$ (the closest integer to $n^{1/3}$), and put interior knots at the quantiles of $\mathcal{O}$. In the marker direction, the knot sequence can be directly chosen based on the quantiles of $\{m_i\}_{i=1}^n$.

As we have pointed in Section 2.1, once we have the sieve MLE estimates, the plug-in spline estimators for ROC and AUC are readily obtained (see Equation (1)). For statistical inference for the AUC, we propose to use the BCa method¹⁷ for computing bootstrap confidence intervals.

3 | SIMULATION STUDY

We evaluate the finite-sample operating characteristics of our method on the basis of the following simulation study. We assume that event time T follows an exponential distribution with hazard rate $\lambda > 0$ and that the marker M follows a beta distribution with density

$f_2(m) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} m^{\alpha-1} (1-m)^{\beta-1}$. In these illustrations, T and M are either independent of each other, jointly generated by a Clayton copula,¹⁸ with parameter $\mu > 1$

$$F(t, m) = \Pr(T \leq t, M \leq m) = \left\{ F_1(t)^{\mu-1} + F_2(m)^{\mu-1} - 1 \right\}^{1/(\mu-1)},$$

where $F_1(\cdot)$ and $F_2(\cdot)$, respectively, represent the marginal distribution functions of T and M . We quantify the dependence between T and M using Kendall's τ .¹⁹ For the Clayton's copula, larger values of the dependence parameter μ imply stronger association. More specifically, μ is related to τ through $\tau = \frac{\mu-1}{\mu+1}$.¹⁸

The number of assessments, K_c , is assumed to follow a geometric distribution with parameter $v > 0$. The distance between two contiguous assessment times, L_c , is assumed to be fixed. For a given right censoring rate $\rho \in (0, 1)$, the parameter v is calibrated so that $\rho = \Pr(T > L_c K_c)$. Uniform noise, distributed over the interval $(-L_c/6, L_c/6)$, is added to each assessment time to account for patient noncompliance. Based on the relationship between event time and the actual assessment times, we can get the actual values for u_i , v_i , $\delta_i^{(1)}$, $\delta_i^{(2)}$, and $\delta_i^{(3)}$ in the likelihood function (10). Note that $\delta_i^{(1)} = 1$ (left censoring) implies $u_i = L_c$, $\delta_i^{(2)} = 1$ (interval censoring) implies u_i and v_i are two consecutive observation times with $v_i - u_i = L_c$ and $\delta_i^{(3)} = 1$ (right censoring) implies v_i is the last assessment time.

For estimation, we consider spline basis functions of order $l = 3$, that is, we use quadratic and M-spline basis functions, cubic I-spline basis functions throughout the simulation as mentioned in Section 2.2. The knot sequence for the splines is chosen as described in Section 2.2. T is generated from an exponential distribution with hazard rate $\lambda = \log(2)/30$. M is generated from a beta distribution with $\alpha = 2.35$ and $\beta = 1.87$, and then M is scaled from 0 to 10. We consider $\tau = 0$, $\tau = 0.2$, and $\tau = 0.6$ to represent independence, weak association, and strong association and consider right censoring rates of $\rho = 0.3$ and $\rho = 0.5$ to represent low versus high levels of right censoring. We choose $L_c = 6$ for assessment times. The ROC and the AUC are estimated at times $t = 12$ and 28. To know how early 12 and 28 are among the generated observation times, it is reasonable to evaluate the quantile of the censored observations by $\Pr(\mathcal{O} \leq t)$ (the probability of elements in $\mathcal{O}$ less than t) with

$$\mathcal{O} = \left\{ u_i \delta_i^{(1)} + \frac{u_i + v_i}{2} \delta_i^{(2)} + v_i \delta_i^{(3)} \right\}_{i=1}^n \text{ as in Section 2.2. We have } \Pr(\mathcal{O} \leq 12) = 0.32 \text{ and } 0.40,$$

respectively, for right censoring rates 0.3 and 0.5; $\Pr(\mathcal{O} \leq 28) = 0.62$ and 0.74, respectively, for right censoring rates 0.3 and 0.5. We consider sample sizes of $n = 100$ and 300. Coverage probabilities, at the nominal two-sided 95% level, are assessed by calculating confidence intervals using the BCa method on the basis of $B = 1000$ bootstrap replicates. Each illustration is based on $N = 1000$ simulation replicates. We note that the putative parameter values for the distributions are chosen to mimic those from CALGB 30801.

For each simulation scenario, the proposed AUC estimation results are compared to those using the approach presented by Li and Ma⁷ with applying the semiparametric approach as they suggested for estimating the standard error. The R scripts to reproduce the simulation results are included as Web Appendix 2.

The empirical censoring rates based on size 10 000 and repetition 1000 are presented in Table 1. The relative bias (re-Bias), standard deviation, and coverage probability, at the nominal two-sided 95% level, for estimation of AUC are shown in Table 1. For the Scenarios, we have considered the performance of our method, in terms of both re-Bias and

coverage probability, improves as the sample size is increased. In contrast, Li and Ma's method performs worse in presence of moderate dependence and is not valid in presence of strong dependence. Specifically, when Kendall's $\tau = 0$, our proposed method (Sieve) has similar performance to that by Li-Ma. However, when Kendall's $\tau = 0.2$, our method performs considerably better having smaller re-Bias and more reliable coverage probability; and when Kendall's $\tau = 0.6$, the Li-Ma method is no longer valid in the sense that with increased sample size the re-Bias is not reduced. This phenomenon may be caused by the fact that Li-Ma has to ignore some data information for their estimation, which is valid only for independent case and becomes more biased for the case with stronger underlying association. Additionally, we observe that the trend in terms of the estimation bias for different association parameters is consistent with both time points and both right censoring rates.

Figure 1 presents the results for estimating the ROC curve, at time $t = 12$ under a right censoring rate of 0.5 for $\tau = 0, 0.2$, or 0.6 , and $n = 100$ or 300 . The estimation becomes more accurate when the size is increased from 100 to 300. Figure 1 shows that the proposed method performs the best when $\tau = 0.2$. This is consistent with the results in Table 1 where the smallest re-Bias is observed when $\tau = 0.2$.

4 | ANALYSIS OF CALGB 30801

We applied our AUC estimator for analysis of CALGB 30801 data.⁶ CALGB 30801 is a randomized phase III double blind trial that evaluated the effect of selective cyclooxygenase-2 (COX-2) inhibitor in advanced non-small cell lung cancer. Three hundred twelve (312) eligible patients were randomized in equal ratio to standard chemotherapy plus COX-2 inhibitor and standard chemotherapy plus placebo. The hypothesis is that the patients with a high COX-2 level would benefit more from chemotherapy plus COX-2 inhibitor than chemotherapy plus placebo, but the trial failed to show difference between two treatment arms for the patients when COX-2 level is high as well as all patients regardless of the COX-2 level. COX-2 was measured on tissues collected from biopsy and it is often considered a more accurate measure for COX-2 expression level and was used as stratification factor for patient randomization, while Prostaglandin E metabolite (pGEM) was a biomarker for COX-2 expression level measured from urine samples. As there is suspicion that the COX-2 assay for tissues was contaminated during the trial, the investigators are interested to characterize the prediction accuracy of COX-2 and pGEM for PFS. The tissues and urine samples for measuring COX-2 and pGEM were collected at baseline. The measures for COX-2 were available prior to randomization and the level of COX-2 was one of the stratification factors for randomization. The measures of pGEM were available after randomization by analyzing the specimens collected at baseline. PFS is defined as the time from randomization to progression. As the progression of the tumors was monitored on the basis of radiologic assessments once every 6 weeks during protocol therapy and every 8 weeks in the first two years and then once every 24 weeks in the next 3 years, PFS is a time-to-event endpoint subject to interval censoring.

The CALGB 30801 data include interval censored progression free survival and two markers measured at baseline (COX-2 and pGEM) for 312 patients. As there is no significant

treatment effect for the patients with high COX-2 levels as well as for all patients, we included all 312 patients in the biomarker analysis for prediction accuracy. The median observation time is 18.8 weeks. We also produce the Kaplan-Meier plots (Figure 2) for both COX-2 and pGEM markers, the markers are transferred into a binary factor “low” and “high” regarding to the median marker levels.

Table 2 presents the AUC estimation ($\hat{AUC}$) using both the proposed method (sieve) and the approach presented by Li and Ma (Li-Ma)⁷ at 5 time points (6, 12, 18, 24, and 30 weeks) with markers COX-2 and pGEM. As for simulation studies we adopted BCa method for sieve and the semiparametric bootstrap as suggested in Reference⁷ both with 1000 resamples to generate each 95% confidence interval (95% *CI*). It is shown that there is no big difference in estimation results between two method and there is no evidence that either marker is associated with the event time. These results are consistent with the simulation results with $\tau = 0$, when both methods perform similarly.

5 | FINAL REMARKS

We have proposed spline based plug-in estimators for time-dependent ROC curve and its corresponding AUC measure based on interval censored time to event data and a continuous marker. Our simulation results confirm our theoretical result showing consistency of our method. For finite sample sizes, our proposed method performs considerably better than the method proposed by Li and Ma⁷ when the event time and the marker are moderately associated. Our results suggest that the method of Li and Ma is not applicable under strong dependence. The results also suggest the BCa bootstrapping confidence interval¹⁷ can be used for statistical inference on our proposed AUC estimator when the sample size is large.

The proposed method can be extended to the competing risk or multi-marker frameworks.

Some observations from the simulation results shown in Table 1 and Figure 1 bear discussion.

When estimating AUC at time 12 (a relatively early time), the relative bias under a right censoring rate of 0.3 may be larger than that under a right censoring rate of 0.5. In Figure 3, we illustrate the distributions of current status times under light and heavy right censoring. We observe that under light right censoring, the distribution of the current status times is skewed to the left away from the early time points. As the knots are assigned based on quantiles of the censoring times, the performance of sieve estimation at relatively early times may be worse under light right censoring than that at later time points.

We found that when the association between the event time and the marker starts from independence and becomes stronger, the estimation bias changes from positive to negative. This may be related to the bias for estimating the joint distribution. The coverage probability of our method, as approximated using BCa bootstrap confidence intervals, is working well for large sample sizes but not for small sample sizes (ie, $n = 100$). This may be due to the fact that asymptotic normality for our method may not hold.

A user-friendly R⁹ package to implement our proposed method has already been released on CRAN (<https://CRAN.R-project.org/package=intcensROC>).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

The research was supported in part by award number P01CA142538 from the National Cancer Institute. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of Health. The authors thank the associator editor and referees for their helpful comments.

Funding information

National Cancer Institute, Grant/Award Number: P01CA142538

REFERENCES

1. Zweig MH, Campbell G. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clin Chem*. 1993;39:561–577. [PubMed: 8472349]
2. Spackman KA. Signal detection theory: valuable tools for evaluating inductive learning Paper presented at: Proceedings of the 6th International Workshop on Machine Learning; 1989:160–163; San Mateo, CA.
3. Heagerty PJ, Lumley T, Pepe MS. Time dependent ROC curves for censored survival data and a diagnostic marker. *Biometrics*. 2000;56:337–344. [PubMed: 10877287]
4. Heagerty PJ, Zheng Y. Survival model predictive accuracy and ROC curves. *Biometrics*. 2005;61:92–105. [PubMed: 15737082]
5. Saha-Chaudhuri P, Heagerty PJ. Non-parametric estimation of a time-dependent predictive accuracy curve. *Biostatistics*. 2013;14:42–59. [PubMed: 22734044]
6. Edelman MJ, Wang X, Hodgson L, Cheney RT, Baggstrom MQ, Thomas SP. Phase III randomized, placebo-controlled, double-blind trial of celecoxib in addition to standard chemotherapy for advanced non-small-cell lung cancer with cyclooxygenase-2 overexpression: CALGB 30801 (Alliance). *J Clin Oncol*. 2017;35:2184–2192. [PubMed: 28489511]
7. Li J, Ma S. Time-dependent roc analysis under diverse censoring patterns. *Stat Med*. 2011;30:1266–1277. [PubMed: 21538452]
8. Jacqmin-Gadda H, Blanche P, Chary E, Touraine C, Dartigues J. Receiver operating characteristic curve estimation for time to event with semicompeting risks and interval censoring. *Stat Methods Med Res*. 2016;25:2750–2766. [PubMed: 24803510]
9. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2018.
10. Lin J, Wu Y, Wang X, Owzar K. IntcensROC: fast spline function based constrained maximum likelihood estimator for AUC estimation of interval censored survival data. R package version 0.1.1; 2018.
11. Sun J The Statistical Analysis of Interval-Censored Failure Time Data. New York, NY: Springer-Verlag; 2006.
12. Wu Y, Zhang Y. Partially monotone tensor spline estimation of the joint distribution function with bivariate current status data. *Ann Stat*. 2012;40:1609–1636.
13. Schumaker L Spline Function: Basic Theory. New York, NY: John Wiley; 1981.
14. Ramsay JO. Monotone regression splines in action. *Stat Sci*. 1988;3:425–441.
15. Jamshidian M On algorithms for restricted maximum likelihood estimation. *Comput Stat Data Anal*. 2004;45:137–157.
16. Zhang Y, Hua L, Huang J. A spline-based semiparametric maximum likelihood estimation for the cox model with interval-censored data. *Scandinav J Stat*. 2010;37:338–354.

17. Thomas DJ, Bradley E. Bootstrap confidence intervals. *Stat Sci.* 1996;11:189–212.
18. Nelsen RB. *An Introduction to Copulas*. 2nd ed. New York, NY: Springer; 2006.
19. Kendall M A new measure of rank correlation. *Biometrika.* 1938;30:81–89.

FIGURE 1.

The average sieve estimated ROC curves for two sample sizes, $n = 100$ and 300 , with a right censoring rate of 0.5 evaluated at time $t = 12$: left panel for $\tau = 0$, middle panel for $\tau = 0.2$, and right panel for $\tau = 0.6$

FIGURE 2.

Kaplan-Meier plots based on the CALGB 30801 data for “high” and “low” marker level groups, where the data is split at the median of marker values for COX-2 and for pGEM

FIGURE 3.

The distributions of the current status times under heavy and light right censoring

TABLE 1

The comparison of the AUC estimation between the proposed sieve approach (sieve) and Li and Ma's approach (Li-Ma)⁷

$\tau = 0$											
Time	True AUC	Size	Censoring Rate			Sieve			Li-Ma		
			Left	Interval	Right	<i>re</i> -Bias	Std	95% CP	<i>re</i> -Bias	Std	95% CP
12	0.5000	100	0.13	0.57	0.30	0.1004	0.0385	0.918	0.1122	0.0417	0.946
		300				0.0608	0.0233	0.927	0.0619	0.0245	0.952
		100	0.13	0.37	0.50	0.1128	0.0416	0.915	0.1072	0.0381	0.960
		300				0.0632	0.0230	0.936	0.0605	0.0231	0.958
28	0.5000	100	0.13	0.57	0.30	0.0963	0.0357	0.916	0.1037	0.0381	0.942
		300				0.0565	0.0211	0.920	0.0556	0.0200	0.944
		100	0.13	0.37	0.50	0.0877	0.0346	0.952	0.1076	0.0388	0.910
		300				0.0624	0.0224	0.923	0.0646	0.0247	0.968
$\tau = 0.2$											
Time	True AUC	Size	Censoring Rate			Sieve			Li-Ma		
			Left	Interval	Right	<i>re</i> -Bias	Std	95% CP	<i>re</i> -Bias	Std	95% CP
12	0.6818	100	0.13	0.57	0.30	−0.0150	0.0594	0.845	−0.0441	0.0655	0.916
		300				−0.0051	0.0355	0.961	−0.0469	0.0384	0.900
		100	0.13	0.37	0.50	−0.0054	0.0637	0.878	−0.0956	0.0609	0.848
		300				−0.0015	0.0371	0.935	−0.1017	0.0389	0.654
28	0.6397	100	0.13	0.57	0.30	−0.0042	0.0537	0.846	−0.0302	0.0585	0.948
		300				−0.0024	0.0340	0.932	−0.0323	0.0333	0.950
		100	0.13	0.37	0.50	−0.0228	0.0568	0.840	−0.0655	0.0572	0.956
		300				−0.0027	0.0379	0.932	−0.0688	0.0377	0.896
$\tau = 0.6$											
Time	True AUC	Size	Censoring Rate			Sieve			Li-Ma		
			Left	Interval	Right	<i>re</i> -Bias	Std	95% CP	<i>re</i> -Bias	Std	95% CP
12	0.9473	100	0.13	0.57	0.30	−0.0353	0.0265	0.960	−0.1078	0.0539	0.438
		300				−0.0183	0.0155	0.961	−0.1083	0.0317	0.040
		100	0.13	0.37	0.50	−0.0370	0.0257	0.954	−0.1997	0.0572	0.034
		300				−0.0093	0.0152	0.968	−0.1950	0.0329	0.000
28	0.8948	100	0.13	0.57	0.30	−0.0128	0.0301	0.954	−0.0757	0.0418	0.816
		300				−0.0020	0.0191	0.930	−0.0766	0.0240	0.510
		100	0.13	0.37	0.50	−0.0533	0.0373	0.816	−0.1406	0.0519	0.276
		300				−0.0078	0.0203	0.953	−0.1431	0.0291	0.000

TABLE 2

The comparison of the AUC estimation between sieve and Li-Ma with markers COX-2 and pGEM at 5 time points based on the CALGB 30801 data

Weeks	COX-2				pGEM 1			
	Sieve		Li-Ma		Sieve		Li-Ma	
	$\widehat{AUC}$	95% CI	$\widehat{AUC}$	95% CI	$\widehat{AUC}$	95% CI	$\widehat{AUC}$	95% CI
6	0.594	[0.504, 0.680]	0.562	[0.486, 0.618]	0.509	[0.495, 0.591]	0.509	[0.455, 0.562]
12	0.512	[0.498, 0.602]	0.525	[0.486, 0.590]	0.507	[0.497, 0.647]	0.568	[0.525, 0.610]
18	0.546	[0.503, 0.612]	0.534	[0.489, 0.585]	0.514	[0.490, 0.660]	0.539	[0.502, 0.576]
24	0.531	[0.501, 0.598]	0.506	[0.494, 0.582]	0.510	[0.487, 0.717]	0.528	[0.493, 0.564]
30	0.509	[0.500, 0.593]	0.521	[0.500, 0.627]	0.547	[0.487, 0.717]	0.540	[0.503, 0.578]