Communications for Statistical Applications and Methods
2024, Vol. 31, No. 4, 365–375

https://doi.org/10.29220/CSAM.2024.31.4.365
Print ISSN 2287-7843 / Online ISSN 2383-4757

Estimation of the time-dependent AUC for cure rate model with covariate dependent censoring

Yang-Jin Kim[1,a]

[a]Department of Statistics, Sookmyung Women's University, Korea

Abstract

Diverse methods to evaluate the prediction model of a time to event have been proposed in the context of right censored data where all subjects are subject to be susceptible. A time-dependent AUC (area under curve) measures the predictive ability of a marker based on case group and control one which are varying over time. When a substantial portion of subjects are event-free, a population consists of a susceptible group and a cured one. An uncertain curability of censored subjects makes it difficult to define both case group and control one. In this paper, our goal is to propose a time-dependent AUC for a cure rate model when a censoring distribution is related with covariates. A class of inverse probability of censoring weighted (IPCW) AUC estimators is proposed to adjust the possible sampling bias. We evaluate the finite sample performance of the suggested methods with diverse simulation schemes and the application to the melanoma dataset is presented to compare with other methods.

Keywords: cure rate model, discrimination, IPCW, mixture model, prediction accuracy, time-dependent ROC

1. Introduction

In prognostic studies, it happens a substantial portion of patients can be event-free, which is denoted as a cured group or a risk-free group. For evaluating the effect of covariates both on the cure rate and on the failure time of susceptible (uncured) patients, several models have been proposed. Among them, the mixture model is expressed of a logistic model for the cure rate and a proportional hazard regression model for susceptible patients (Kuk and Chen, 1992; Maller and Zhou, 1996). In the context of a cure rate model, two issues related with a predictive accuracy have been considered. The first one is to predict who is cured and the second is to predict the survival probabilities of uncured subjects based on the markers. Both issues can be dealt by extending the classical discriminative accuracy measures such as the ROC curve and C-index.

The ROC curve has been the most frequently applied measure by providing both a graph and an AUC value. There are two probabilities to construct the curve; a sensitivity is defined as the probability of having a higher marker value among a case group (true positive rate; TPR) and a specificity is defined as the probability of having a lower marker value among a control one (true negative rate; TNR), respectively. These probabilities have been changed according to the threshold value of a marker and are displayed as the ROC curve where plots sensitivity against one minus specificity over

This research is supported by Korean research foundation (NRF-2020R1A2C1A01100755).

[1]Corresponding author: Department of Statistics, Sookmyung Women's University, chengpa-ro 47 gil, Yongsan-Gu, Seoul 04310, Korea. E-mail: yjin@sookmyung.ac.kr

Published 31 July 2024 / journal homepage: http://csam.or.kr

all possible thresholds. The predictive performance of a marker can be evaluated with the AUC (area under curve) where a higher AUC value indicates a better performance.

For survival data, the time to event as the response variable has been observed during follow-up and changed over time which results in a time-dependent AUC denoted by AUC(t). Suppose that M denotes the continuous marker to evaluate the predictive accuracy where the marker can be a risk factor, a combination of several risk factors or a risk score derived from working prediction models. Without the loss of generality, a higher value of M is assumed to indicate a higher chance of experiencing the event and of bringing an early event time. Heagerty and Zhang (2005) proposed several types of time dependent sensitivity, specificity and the corresponding time-dependent ROC curves. Among them, we consider a cumulative sensitivity $Se^C(c,t)$ and a dynamic specificity $Sp^D(c,t)$ defined as follows,

$$Se^C(t,c) = \Pr(M > c \mid T \le t) \tag{1.1}$$

$$Sp^D(t,c) = \Pr(M \le c \mid T > t), \tag{1.2}$$

the corresponding ROC curve is given by $\mathrm{ROC}^{C/D}(t) = Se^C[(1 - Sp^D(p,t))^{-1}, t]$, $p \in [0,1]$ and the resulting $\mathrm{AUC}^{C/D}(t)$ is defined as

$$\mathrm{AUC}^{C/D}(t) = \Pr\left(M_i > M_j \mid T_i \le t, T_j > t\right), \quad i \ne j,$$

which is interpreted as the probability that for two randomly chosen subjects, one experiencing the event prior to t has the greater marker value compared to the other one free from the event at t. This definition is more relevant in clinical studies to discriminate between subjects experiencing the event and those event-free prior to the specific time (Pepe, 2003; Kamarudin *et al.*, 2017).

For a cure rate model, most estimators of the AUC have been focused on the cure probability of a prediction model. Asano *et al.* (2014) proposed two estimators of AUC by incorporating the full imputation and the mean score imputation for unknown cure status as the extension of Alonzo and Pepe (2005)'s method and Asano and Hirakawa (2017) also suggested the C-index with different weights for three groups (cure, uncured and censored subjects). Recently, several time-dependent AUC estimators have been proposed for evaluating a survival probability of susceptible group. Beyene *et al.* (2019) applied a nonparametric estimator of AUC proposed by Li *et al.* (2018) to cure model and Wang and Wang (2020) considered to implement a smoothing technique into the conditional survival functions of two main cure rate models such as mixture model and bounded cumulative hazard model (Yakovlev *et al.*, 1993).

In this study, our interest is to suggest the time-dependent AUC estimator based on the inverse probability censoring weight (IPCW) technique, when a censoring is related with a covariate. The rest of this article is organized as follows. In Section 2, we introduce the notations and propose three types of time-dependent AUCs. In Section 3, the finite sample performance of suggested methods is evaluated through simulation studies. Application of the suggested methods to a melanoma dataset is presented in Section 4 and several discussions are given in Section 5.

2. Time-dependent AUC using IPCW

In the context of survival data, the time to event is not always observed due to the censoring related with diverse observation schemes. Furthermore, in the presence of nonsusceptible (cured) patients, the time to event is denoted as $T = UT^* + (1-U)\infty$, where U is an indicator that equals 1 if the subject

is susceptible (not cured) and 0 if cured (insusceptible) and T^* denotes an event time of a susceptible subject. Given a covariate vector Z, let $\pi(Z) = \Pr(U = 1|Z)$ be a susceptible probability modeled with Z. Let C denote a random censoring time with a survival function $G(c) = \Pr(C \geq c)$ and the censoring time is assumed to be independent of T^* conditional on the covariate vector Z. Then observable data is denoted as $(\tilde{T}, \delta, Z)$, where $\tilde{T} = \min(T, C)$ and $\delta = I(T < C)$. When $\delta = 1$, the individual has experienced an event, thus $U = 1$. When $\delta = 0$, however, the information of U is missing. Therefore, a population survival function $S(t|Z)$ is expressed as

$$S\ (t \mid Z) = \pi(Z)\tilde{S}\ (t \mid Z, U = 1) + 1 - \pi(Z), \tag{2.1}$$

where $\tilde{S}(t|Z, U = 1)$ denotes the conditional survival function of a susceptible group. As $t \to \infty$, $\tilde{S}(\infty|Z, U = 1) = 0$, but $S(\infty|Z) = 1 - \pi(Z)$. Therefore, ignoring a cure fraction would result in the biased inference of the survival function.

For modelling the susceptible rate of a subject i, a logistic regression model is implemented for estimating the effect of the covariate vector $X_i = (1, Z_i)$

$$\pi_i = \Pr\left(U_i = 1 \mid X_i\right) = \frac{\exp(X_i^{'}\gamma)}{1 + \exp(X_i^{'}\gamma)}. \tag{2.2}$$

Under a PH model assumption, the conditional hazard model of a susceptible group is written as

$$\lambda\left(t \mid Z_i, U_i = 1\right) = \lambda_0(t)\exp\left(Z_i^{'}\beta\right), \tag{2.3}$$

then $\tilde{S}_i(t|Z_i, U_i = 1) = \exp(-\Lambda(t|Z_i, U_i = 1)) = \exp(-\Lambda_0(t)e^{Z_i^{'}\beta})$. For estimating $\theta = (\gamma, \beta, \lambda_0)$, the EM algorithm is implemented to recover the unknown event status in a mixture model (Lam *et al.*, 2008; Sy and Taylor, 2000; Kim and Jhun, 2008).

For evaluating the prediction model of a cure rate data, a risk score $M_i = Z_i^{'}\hat{\beta}$ is utilized as a marker. To reflect the susceptibility of a censored subject on cumulative sensitivity and dynamic specificity in (1.1) and (1.2), several methods have been proposed.

Beyene *et al.* (2019) considered a missing status of a censored subject $\tilde{T}_i$ and implemented the probability of experiencing an event until $t > \tilde{T}_i$ when a subject i is censored at $\tilde{T}_i$ which is expressed as

$$B_i(t) = 1 - P\left(T > t \mid T > \tilde{T}_i\right), \quad t > \tilde{T}_i,$$

is estimated

$$\hat{B}_i(t) = \delta_i I\left(\tilde{T}_i \leq t\right) + (1 - \delta_i)\left(1 - \frac{\hat{S}\left(t \mid Z_i\right)}{\hat{S}\left(\tilde{T}_i \mid Z_i\right)}\right) I\left(\tilde{T}_i \leq t\right),$$

where the survival function $\hat{S}(t|Z) = \hat{\pi}(Z)\hat{\tilde{S}}(t|Z, U = 1) + (1 - \hat{\pi}(Z))$ is estimated from the cure rate model (2.1). Then a time-dependent cumulative sensitivity, dynamic specificity and corresponding $\mathrm{AUC}_B(t)$ are estimated as follows,

$$Se_B(c, t) = \frac{\sum_{i=1}^{n} I(M_i > c)\hat{B}_i(t)}{\sum_{i=1}^{n} \hat{B}_i(t)}, \quad Sp_B(c, t) = \frac{\sum_{i=1}^{n} I(M_i \leq c)(1 - \hat{B}_i(t))}{\sum_{i=1}^{n}(1 - \hat{B}_i(t))}, \tag{2.4}$$

$$\mathrm{AUC}_B(t) = \frac{\sum_{i=1}^n \sum_{j=1}^n \hat{B}_i(t)(1-\hat{B}_j(t))M_{ij}}{\sum_{i=1}^n \sum_{j=1}^n \hat{B}_i(t)(1-\hat{B}_j(t))}, \quad M_{ij} = I\left(M_i > M_j\right). \tag{2.5}$$

For a same problem, Wang and Wang (2020) directly implemented the estimated survival function as follows,

$$Se_W(c,t) = \frac{\sum_{i=1}^n I(M_i > c)(1-\hat{S}_i(t \mid Z_i))}{\sum_{i=1}^n (1-\hat{S}_i(t \mid Z_i))}, \quad Sp_W(c,t) = \frac{\sum_{i=1}^n I(M_i \le c)\hat{S}_i(t \mid Z_i)}{\sum_{i=1}^n \hat{S}_i(t \mid Z_i)},$$

$$\mathrm{AUC}_W(t) = \frac{\sum_{i=1}^n \sum_{j=1}^n (1-\hat{S}_i(t \mid Z_i))\hat{S}_j(t \mid Z_j)M_{ij}}{\sum_{i=1}^n \sum_{j=1}^n (1-\hat{S}_i(t \mid Z_i))\hat{S}_j(t \mid Z_j)}, \tag{2.6}$$

where they applied a smoothing technique to obtain $\mathrm{AUC}_W(t)$.

In general survival data, a right censoring causes a biased sampling when a censoring distribution is related with a certain subpopulation which is sometimes modelled with a vector of covariates. Inverse probability of censoring weighting (IPCW) technique has been originally proposed to adjust for dependent censoring (Robins, 1993; Robins and Finkelstein, 2000). Under a competing risk data, it has been adopted to reflect the effect of the subpopulation with competing event (Fine and Gray, 1999) and also applied to the discriminative measures such as C-index (Uno *et al.*, 2011) and AUC(t) (Blanche *et al.*, 2013).

In this paper, we propose a class of IPCW estimators of time-dependent AUC(t) when a censoring distribution is related with covariates. Set $W_i(\tilde{T}_i) = 1/\hat{G}(\tilde{T}_i)$, where $\hat{G}$ denotes the estimated censoring survival function obtained from either a Kaplan-Meier estimator or regression models given a covariate vector Z_i.

The first estimator is to incorporate IPCW into Beyene's method (2.4) and (2.5) as follows,

$$Se_{BW}(c,t) = \frac{\sum_{i=1}^n I(M_i > c)\hat{B}_i(t)\hat{W}_i(T_i)}{\sum_{i=1}^n \hat{B}_i(t)\hat{W}_i(\tilde{T}_i)},$$

$$Sp_{BW}(c,t) = \frac{\sum_{i=1}^n I(M_i \le c)(1-\hat{B}_i(t))\hat{W}_i(t)}{\sum_{i=1}^n (1-\hat{B}_i(t))\hat{W}_i(t)},$$

$$\mathrm{AUC}_{BW}(t) = \frac{\sum_{i=1}^n \sum_{j=1}^n \hat{B}_i(t)(1-\hat{B}_j(t))M_{ij}\hat{W}_i(\tilde{T}_i)\hat{W}_j(t)}{\sum_{i=1}^n \sum_{j=1}^n \hat{B}_i(t)(1-\hat{B}_j(t))\hat{W}_i(\tilde{T}_i)\hat{W}_j(t)}. \tag{2.7}$$

Blanche *et al.* (2013) proposed the IPCW estimators of $ROC(t)$ under competing risk data and explained the role of weights on both case and two types of control. As the second estimator, we extend their idea to a cure rate model. For the cumulative sensitivity (1.1), the IPCW is incorporated into the case group who has experienced the event until t.

$$Se_{CW}(c,t) = \frac{\sum_{i=1}^n I(M_i > c)I(\tilde{T}_i \le t, \delta_i = 1)W_i(\tilde{T}_i)}{\sum_{i=1}^n I(\tilde{T}_i \le t, \delta_i = 1)W_i(\tilde{T}_i)}.$$

Table 1: Bias(sd) of AUC(t) at $C \sim \exp(\theta_c), \theta_c = 0.5$

(Cure, Cen)	n	t	AUC_{Uno}	AUC_B	AUC_W	AUC_{BW}	$\text{AUC}_{CW,1}$	$\text{AUC}_{CW,2}$
(0.15, 0.35)	200	$t_{(0.15)}$	0.0001 (0.053)	0.0001 (0.052)	0.0001 (0.028)	0.0001 (0.052)	0.020 (0.057)	0.0001 (0.053)
		$t_{(0.30)}$	0.001 (0.041)	0.001 (0.040)	0.003 (0.029)	0.002 (0.040)	0.035 (0.046)	0.001 (0.041)
	400	$t_{(0.15)}$	0.0001 (0.037)	0.0001 (0.036)	0.0002 (0.022)	0.0001 (0.036)	0.031 (0.040)	0.0001 (0.037)
		$t_{(0.30)}$	0.002 (0.028)	0.001 (0.028)	0.001 (0.022)	0.001 (0.027)	0.032 (0.032)	0.002 (0.028)
(0.30, 0.55)	200	$t_{(0.15)}$	0.001 (0.050)	0.001 (0.052)	0.0031 (0.037)	0.001 (0.052)	0.0581 (0.059)	0.0001 (0.053)
		$t_{(0.30)}$	0.001 (0.040)	0.001 (0.043)	0.003 (0.038)	0.002 (0.043)	0.071 (0.047)	0.000 (0.045)
	400	$t_{(0.15)}$	0.001 (0.038)	0.0001 (0.037)	0.001 (0.022)	0.0001 (0.037)	0.054 (0.044)	0.001 (0.038)
		$t_{(0.30)}$	0.002 (0.032)	0.0001 (0.032)	0.0001 (0.021)	0.0001 (0.032)	0.0617 (0.040)	0.001 (0.029)
(0.50, 0.70)	200	$t_{(0.15)}$	0.008 (0.049)	0.011 (0.057)	0.017 (0.048)	0.011 (0.058)	0.112 (0.062)	0.010 (0.058)
		$t_{(0.30)}$	0.005 (0.038)	0.003 (0.055)	0.000 (0.052)	0.000 (0.055)	0.097 (0.049)	0.002 (0.056)
	400	$t_{(0.15)}$	0.012 (0.038)	0.0011 (0.032)	0.0008 (0.024)	0.0006 (0.031)	0.101 (0.041)	0.0010 (0.033)
		$t_{(0.30)}$	0.003 (0.029)	0.003 (0.027)	0.005 (0.024)	0.005 (0.027)	0.104 (0.034)	0.003 (0.028)

cure: cure rate; cp: censoring rate;

For the control group of a dynamic specificity in (1.2), two versions are presented. The first version of control group, the event-free subjects at t have weighted with both a susceptible proportion π_i and $W_i(t)$ in order to reflect the chance of experiencing the event and the time to event is certain to be greater than t. The second version expands the control by including the subjects censored before t which are weighted with the conditional survival probability. Therefore, the two versions of dynamic specificity $Sp_{CW,1}(c,t)$ and $Sp_{CW,2}(c,t)$ are estimated by

$$Sp_{CW,1}(c,t) = \frac{\sum_{i=1}^{n} I(M_i \le c) I(\tilde{T}_i > t) \hat{W}_i(t) \hat{\pi}_i}{\sum_{i=1}^{n} I(\tilde{T}_i > t) \hat{W}_i(t) \hat{\pi}_i},$$

$$Sp_{CW,2}(c,t) = \frac{\sum_{i=1}^{n} I(M_i \le c)(I(\tilde{T}_i > t) + (\hat{S}(t)/\hat{S}(\tilde{T}_i)) I(\tilde{T}_i < t, \delta_i = 0)) \hat{W}_i(t)}{\sum_{i=1}^{n} (I(\tilde{T}_i > t) + (\hat{S}(t)/\hat{S}(\tilde{T}_i)) I(\tilde{T}_i < t, \delta_i = 0)) \hat{W}_i(t)}.$$

Then the corresponding time-dependent AUC(t)s are estimated by

$$\text{AUC}_{CW,1}(t) = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} M_{ij} I(\tilde{T}_i \le t, \delta_i = 1) I(\tilde{T}_j > t) \hat{W}_i(T_i) \hat{W}_j(t) \hat{\pi}_j}{\sum_{i=1}^{n} \sum_{j=1}^{n} I(\tilde{T}_i \le t, \delta_i = 1) I(\tilde{T}_j > t) \hat{W}_i(T_i) \hat{W}_j(t) \hat{\pi}_j}, \tag{2.8}$$

$$\text{AUC}_{CW,2}(t) = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} M_{ij} I(\tilde{T}_i \le t, \delta_i = 1)(I(\tilde{T}_i > t) + (\hat{S}(t)/\hat{S}(\tilde{T}_i)) I(\tilde{T}_i < t, \delta_i = 0)) \hat{W}_i(\tilde{T}_i) \hat{W}_j(t)}{\sum_{i=1}^{n} \sum_{j=1}^{n} I(\tilde{T}_i \le t, \delta_i = 1)(I(\tilde{T}_j > t) + (S(t)/\hat{S}(\tilde{T}_j)) I(\tilde{T}_j < t, \delta_i = 0)) \hat{W}_i(\tilde{T}_i) \hat{W}_j(t)}. \tag{2.9}$$

Table 2: Bias(sd) of AUC(t) at $c \sim \exp(0.5\exp(1.0Z))$

(Cure,cen)	n	t	AUC_{Uno}	AUC_B	AUC_W	AUC_{BW}	$\text{AUC}_{CW,1}$	$\text{AUC}_{CW,2}$
(0.15, 0.35)	200	$t_{(0.15)}$	0.020	0.012	0.010	0.012	0.013	0.013
			(0.049)	(0.047)	(0.031)	(0.048)	(0.055)	(0.050)
		$t_{(0.30)}$	0.037	0.020	0.014	0.016	0.004	0.001
			(0.042)	(0.041)	(0.032)	(0.043)	(0.049)	(0.045)
	400	$t_{(0.15)}$	0.021	0.032	0.012	0.012	0.014	0.001
			(0.038)	(0.037)	(0.022)	(0.039)	(0.042)	(0.039)
		$t_{(0.30)}$	0.036	0.020	0.019	0.016	0.005	0.005
			(0.029)	(0.029)	(0.023)	(0.04)	(0.041)	(0.040)
(0.30, 0.55)	200	$t_{(0.15)}$	0.034	0.019	0.013	0.017	0.028	0.001
			(0.052)	(0.050)	(0.037)	(0.052)	(0.059)	(0.051)
		$t_{(0.30)}$	0.056	0.028	0.025	0.020	0.008	0.008
			(0.042)	(0.042)	(0.037)	(0.050)	(0.056)	(0.053)
	400	$t_{(0.15)}$	0.033	0.018	0.016	0.015	0.032	0.003
			(0.037)	(0.035)	(0.024)	(0.038)	(0.043)	(0.039)
		$t_{(0.30)}$	0.057	0.031	0.030	0.019	0.013	0.000
			(0.031)	(0.029)	(0.024)	(0.053)	(0.061)	(0.060)
(0.50, 0.70)	200	$t_{(0.15)}$	0.039	0.017	0.012	0.015	0.056	0.005
			(0.053)	(0.062)	(0.054)	(0.065)	(0.070)	(0.067)
		$t_{(0.30)}$	0.070	0.030	0.027	0.012	0.030	0.002
			(0.041)	(0.058)	(0.057)	(0.075)	(0.082)	(0.077)
	400	$t_{(0.15)}$	0.036	0.018	0.013	0.018	0.061	0.003
			(0.034)	(0.035)	(0.028)	(0.036)	(0.044)	(0.036)
		$t_{(0.30)}$	0.067	0.032	0.030	0.012	0.038	0.002
			(0.029)	(0.028)	(0.020)	(0.057)	(0.070)	(0.061)

Cure: cure rate; cen: censoring rate;

For the variance estimation, the bootstrap samples are generated and the confidence intervals are obtained from their standard deviations.

3. Simulation

In this section, the performance of the suggested estimators is evaluated with three situations; (i) light censoring; 35% (cure-rate: 15%), (ii) medium censoring; 55% (cure-rate: 30%) and (iii) heavy censoring; 70%(cure-rate: 50%), respectively. The difference of these censoring rates is inclined to the amount of cure rates. For reflecting the effect of a covariate on cure rate, a failure time and a censoring distribution, a covriate Z is generated from a normal distribution $N(0, 1)$. A cure status $U = \{0, 1\}$ is generated based on $\Pr(U = 1) = (\exp(\gamma_0 + \gamma_1 Z))/(1 + \exp(\gamma_0 + \gamma_1 Z))$, where γ_0 is selected to get a suitable cure rate and $\gamma_1 = -1.0$. For a subject with $U = 1$, generate a failure time T^* from a hazard function $\lambda(t|U = 1) = \lambda_0(t)\exp(\beta Z)$, where a baseline hazard function is assumed to follow a Weibull distribution and $\beta = 0.5$ is assigned to represent the effect of covariate on failure time. Let the marker define as $M_i = Z_i'\hat{\beta}$ using the regression coefficient estimated at (2.3).

To compare the performance of several IPCW estimators of AUC(t), a censoring time is generated with two scenarios. (i) Covariate independent censoring: $\lambda_c = \theta_c$ where $\theta_c = 0.5$ and (ii) to reflect the effect of covariate on the censoring, $\lambda_c = 0.5\exp(1.0Z)$. Then the observable time is composed of $(\tilde{T}, \delta, X)$, where $\tilde{T} = \min(T^*, C)$ and $\delta = I(T^* < C)$. For a cured subject with $U = 0$, set $\tilde{T} = C$ and $\delta = 0$.

300 datasets are generated with two sample sizes $n = 200$ and $n = 400$. Table 1 and Table 2 show the biases and standard deviations of six estimators ($\text{AUC}_{Uno}(t)$, $\text{AUC}_B(t)$ in (2.5), $\text{AUC}_W(t)$ in (2.6), $\text{AUC}_{BW}(t)$ in (2.7), $\text{AUC}_{CW,1}(t)$ in (2.8) and $\text{AUC}_{CW,2}(t)$ in (2.9)) at two percentile points $(t_{(0.15)}, t_{(0.30)})$.

Table 3: Coverage probability of AUC(t) at $n = 200$ and (cure,cen) = (40%, 60%)

				Covariate independent censoring				
		$t_{(0.15)}$				$t_{(0.30)}$		
Method	Est	SD	SE	CP	Est	SD	SE	CP
AUC_{Uno}	0.706	0.051	0.050	0.929	0.722	0.042	0.040	0.948
AUC_B	0.707	0.054	0.051	0.935	0.723	0.048	0.043	0.967
AUC_W	0.702	0.040	0.031	0.962	0.722	0.041	0.031	0.967
AUC^*_{BW}	0.706	0.054	0.047	0.935	0.721	0.047	0.036	0.967
$AUC^*_{CW,1}$	0.643	0.059	0.060	0.801	0.659	0.049	0.047	0.775
$AUC^*_{CW,2}$	0.701	0.056	0.052	0.917	0.719	0.049	0.041	0.961
				Covariate dependent censoring				
		$t_{(0.15)}$				$t_{(0.30)}$		
Method	Est	SD	SE	CP	Est	SD	SE	CP
AUC_{Uno}	0.720	0.055	0.051	0.890	0.75	0.044	0.042	0.680
AUC_B	0.708	0.056	0.057	0.910	0.727	0.056	0.051	0.830
AUC_W	0.705	0.049	0.045	0.860	0.723	0.052	0.047	0.830
AUC_{BW}	0.707	0.056	0.060	0.900	0.718	0.062	0.064	0.900
$AUC_{CW,1}$	0.660	0.057	0.065	0.940	0.692	0.059	0.068	0.970
$AUC_{CW,2}$	0.689	0.056	0.061	0.930	0.693	0.058	0.064	0.930

Here $AUC_{Uno}(t)$(Uno *et al.*, 2016) is also presented to show the effect of ignoring the cure rate but reflecting the IPCW and obtained from the R package SurvAUC.

Table 1 shows the biases(standard deviations) of suggested methods when a censoring distribution is independent of covariate. All estimates have similar results and seem to be unbiased. However, $AUC_{CW,1}(t)$ shows large biases all cases. It seems to be related with the definition of a control group. Implementing the weights to the subjects with $\tilde{T}_i > T$ seems to result in decreasing the size of the control group. Meanwhile, by augmenting the control group by including censored subjects, $AUC_{CW,2}(t)$ have smaller biases. For the standard deviation, the $AUC_W(t)$ based on smoothing technique has the smallest variation. Table 2 presents the simulation results at a covariate dependent censoring scenario. For non-IPCW estimator, $AUC_{Uno}(t)$ ignoring the cure rate has the largest biases increasing with a censoring rates. The biases at $t_{(0.3)}$ tend to be larger than ones at $t_{(0.15)}$. Among the suggested IPCW-based methods, $AUC_{CW,2}(t)$) has smallest biases at all situations and $AUC_{CW,1}(t)$ has smaller biases at low censoring rate but shows increasing biases. The IPCW-based AUCs at $t_{(0.30)}$ have smaller biases compared with those values at $t_{(0.15)}$ which have different result with non-IPCW ones. Also, according to standard deviation, the inclusion of weights brings the increment of variation of IPCW estimators.

Table 3 shows the coverage probabilities (CP) and the standard errors obtained using 50 bootstrap samples at $n = 200$ with a censoring rate 60% and a cure-rate 40%. In order to distinguish between IPCW estimators based on $\hat{G}(t)$ and $\hat{G}(t|Z)$, ($AUC^*_{BW}(t)$, $AUC^*_{CW,1}(t)$, $AUC^*_{CW,2}(t)$) represent the results obtained under the former case. Similar to the results of Table 1 and Table 2, under independent censoring scheme, $AUC_{CW,1}(t)$ has much smaller CP because of large biases. Under covariate dependent censoring, $AUC_{Uno}(t)$, $AUC_B(t)$ and $AUC_W(t)$ show undesirable results while $AUC_{CW,1}(t)$ and $AUC_{CW,2}(t)$ have coverage probabilities close to a nominal one.

4. Data analysis

We analyzed a malignant melanoma dataset which is available in the R package MASS. The dataset consists of 205 patients whose tumors were completely removed together with the skin within a distance of about 2.5cm around it at the operation. The study started in the period 1962–1977 and all patients have been followed for checking disease progression and survival until 1977. Among 205

Table 4: Summary of regression models of Melanoma dataset

	Cure model				Censoring distribution	
	Susceptible rate		Latency distribution		Cox PH	
Cov	Est(se)	p-value	Est(se)	p-value	Est(se)	p-value
Intercept	−2.33(0.929)	0.012				
Sex	0.288(0.582)	0.621	0.569(.532)	0.284	−0.041(0.178)	0.818
Log (thick)	0.042(0.095)	0.659	0.874(0.287)	0.002	−0.206(0.094)	0.028
Ulcer	1.323(0.641)	0.039	0.086(0.536)	0.872	0.196(0.197)	0.321
Age	0.019(0.017)	0.253	−0.008(0.013)	0.557	0.026(0.006)	<0.0001

Table 5: Estimation of AUC values and 95% CI at $t = (1, 4, 8)$ years of malignant melanoma patients

Method	$t = 1(\hat{S}(t) = 0.97)$		$t = 4(\hat{S}(t) = 0.82)$		$t = 8(\hat{S}(t) = 0.68)$	
	Est(se)	95% CI	Est(se)	95% CI	Est(se)	95% CI
AUC_{Uno}	0.904	(0.814,0.994)	0.824	(0.746,0.902)	0.772	(0.628,0.816)
AUC_B	0.868	(0.768,0.968)	0.812	(0.706,0.918)	0.737	(0.643,0.831)
AUC_W	0.789	(0.685,0.893)	0.780	(0.610,0.806)	0.731	(0.619,0.848)
covariate-independent censoring:$G(t)$						
AUC^*_{BW}	0.887	(0.785,989)	0.812	(0.710,0.914)	0.725	(0.615,0.835)
$\text{AUC}^*_{CW,1}$	0.839	(0.729,0.949)	0.755	(0.669,0.841)	0.641	(0.527,0.755)
$\text{AUC}^*_{CW,2}$	0.889	(0.787,0.991)	0.810	(0.706,0.914)	0.738	(0.634,0.842)
covariate-dependent censoring: $G(t\|Z)$						
AUC_{BW}	0.887	(0.773,1.000)	0.812	(0.675,0.949)	0.672	(0.428,0.915)
$\text{AUC}_{CW,1}$	0.838	(0.715,0.971)	0.755	(0.669,0.841)	0.583	(0.338,0.828)
$\text{AUC}_{CW,2}$	0.888	(0.774,1.000)	0.809	(0.674,0.944)	0.701	(0.485,0.917)

patients, only 57 patients died of melanoma, 14 one died from other causes and the remaining were alive. In this study, the death from other causes is regarded as a censoring (censoring rate 72%). The time scale is days since operation and four covariates Z such as sex (male = 1), age at operation and characteristics of the tumor such as tumor thickness (median = 1.94mm) and ulcer (1 = presence; 0 = absence). As the prediction model, we applied a PH cure model and R package smcure is used to estimate the parameters. According to Table 4, unlike Wang and Wang's result applying the additive model $\lambda(t|z) = \lambda_0(t) + \beta^{'} z$, ulcer is significant ($p$-value = 0.039) at the susceptibility and log (thickness) is significant in the hazard model (p-value = 0.002). At the regression model on censoring distribution, two covarites (log (thickness) and age) are significant under the PH model. That is, older patients with lower value of log (thickness) are likely to get censored.

Table 5 shows the nine AUC(t) values estimated at $1, 4, 8$ years and 95% confidence intervals based on the standard errors obtained from 100 bootstrap samples. Here, the mark $M_i = Z_i^{'}\hat{\beta}$ is defined as the risk score calculated from the estimated latency distribution. The suggested IPCW estimators are presented with two versions according to the covariate-dependency on censoring distribution and $G(t)$ and $G(t|Z)$ give $(\text{AUC}^*_{BW}, \text{AUC}^*_{CW,1}, \text{AUC}^*_{CW,2})$ and $(\text{AUC}_{BW}, \text{AUC}_{CW,1}, \text{AUC}_{CW,2})$, respectively. Among unweighted AUC estimators, AUC_W has the smallest values at all times. AUC_{Uno} has higher values at all cases which is the same result as in the simulation. For AUC_B and its weight versions AUC^*_{BW} and AUC_{BW}, they have similar values at 1 year and 2 year but the weight versions have smaller values at 8 year. For IPCW estimators, comparing the results based on $G(t)$ and $G(t|Z)$, the AUC values at $t = 1$ and $t = 4$ year have almost same values but the covariate dependent AUC values at 8 year have smaller values and larger standard errors. This result is explained with a high censoring rate and uncommon weights W_i. $\text{AUC}_{CW,2}$ and AUC_{BW} show similar result but $\text{AUC}_{CW,1}$ has the smallest values at two censoring situations. According to simulation and data analysis, $\text{AUC}_{CW,1}$ is unsuitable to apply as the predictive measure. Figure 1 presents the ROC curves of six estimators

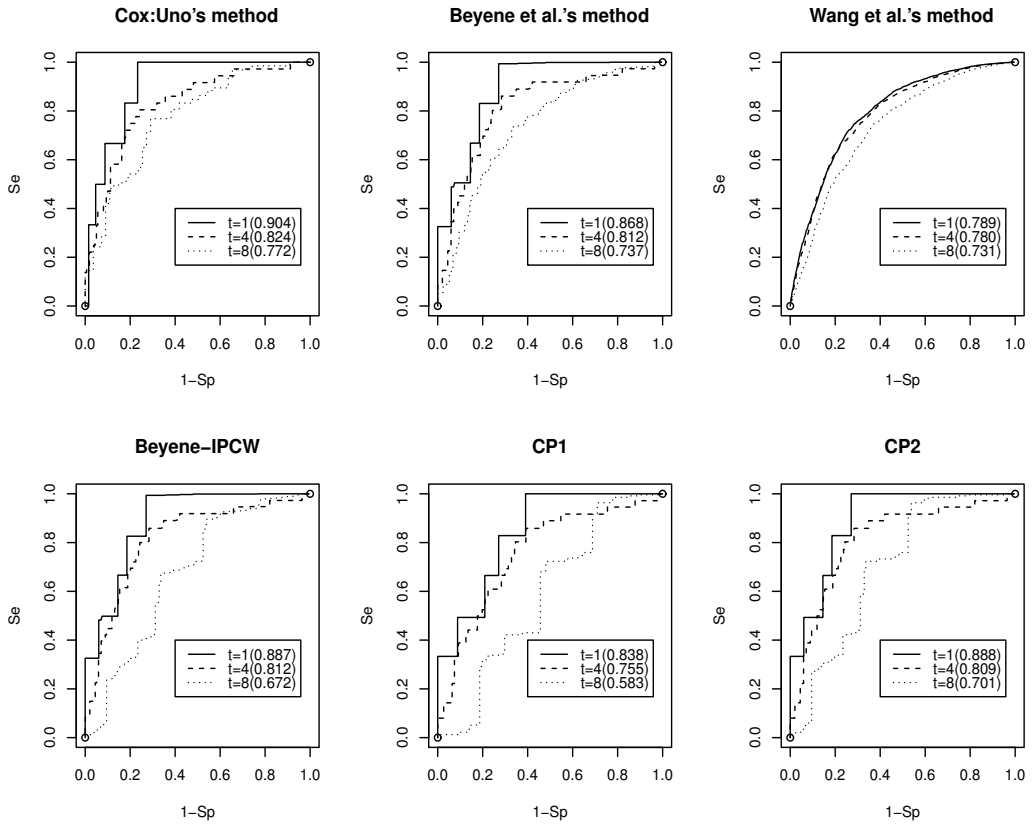

Figure 1: *ROC*(t) *curves (AUCs) of six estimators at* $t = 1, 4,$ *and* 8 *year.*

at $t = 1, 4$ and 8 year with only covariate dependent versions of IPCW.

5. Concluding remarks

In this paper, we applied the IPCW approach to estimate time-dependent AUC for cure rate models when a censoring distribution is related with covariates. Simulation results show that the proposed procedures work well for covariate dependent censoring and a large censoring rate. However, Uno's method $\mathrm{AUC}_{Uno}(t)$ for right censored data still works at covariate independent censoring but has largest biases at covariate dependent censoring. $\mathrm{AUC}_{W}(t)$ based on the smoothing technique has the smallest variation at all cases but shows large biases as censoring rate and sample size increases. Among the IPCW-version estimators, $\mathrm{AUC}_{CW,1}(t)$ shows a undesirable result at covariate independent censoring but has small bias only at the case with covariate dependent light censoring rate. The difference between $\mathrm{AUC}_{CW,1}(t)$ and $\mathrm{AUC}_{CW,2}(t)$ depends on the definition of the control group. At the former case $\mathrm{AUC}_{CW,1}(t)$, the only subjects with $\tilde{T} > t$ is included with weights which makes the influence of true susceptible group decrease thus brings the underestimated result. At the latter case $\mathrm{AUC}_{CW,2}(t)$, the censored subjects with $T_i < t$ is added to augment the control group with a weight $\Pr(T > t|T > \tilde{T}) = S(t)/S(\tilde{T})$. While it makes unbiased results at most scenarios, the implementation

of the estimated survival function causes the increment of variation.

At melanoma data analysis, nine AUC(t) values are similar at the 1 and 4 year and the suggested ones have smaller values than non-IPCW AUC as time increases. In particular, covariate-dependent versions have large variations which bring the wider confidence intervals.

As another discriminative measure, a concordance index or C-index is defined as the proportion of concordant pairs where a patient with an early event time is likely to have a higher marker. Asano and Hirakawa (2017) proposed the C-index reflecting the patients' cure status estimated the cure rate model. A time-dependent C-index $C(t)$ (Gerds *et al.*, 2013) will be considered to evaluate the prediction model of cure rate data.

As another interesting topic, dynamic prediction models have been studied with joint model and landmark approach when a cure rate model includes longitudinal covariates (Rizopulos *et al.*, 2017). A two-dimensional AUC(s, t) can utilize to evaluate a time-dependent marker $M(s)$ to predict the survival probability at time t where $s < t$.

References

Alonzo TA and Pepe MS (2005). Assessing accuracy of a continuous screening test in the presence of verification bias, *Journal of the Royal Statistical Society Series C: Applied Statistics*, **54**, 173–190.

Asano J, Hirakawa A, and Hamada C (2014). Assessing the prediction accuracy of cure in the Cox proportional hazards cure model: An application to breast cancer data, *Pharmaceutical Statistics,* **13**, 357–363.

Asano J and Hirakawa A (2017). Assessing the prediction accuracy of a cure model for censored survival data with long-term survivors: Application to breast cancer data, *Journal of Biopharmaceutical Statistics*, **27**, 918–932.

Beyene KM, Ghouch AE, and Oulhaj A (2019). On the validity of time-dependent AUC estimation in the presence of cure fraction, *Biometrical Journal*, **61**, 1430–1447.

Blache P, Dartigues J, and Jacqmin-Gadda H (2013). Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks, *Statistics in Medicine*, **32**, 5381–5397.

Fine JP and Gray RJ (1999). A proportional hazards model for the subdistribution of a competing risk, *Journal of the American Statistical Association*, **446**, 496–509.

Gerds TA, Kattan MW, Schumacher M, and Yu C (2013). Estimating a time-dependent concordance index for survival prediction models with covariate dependent censoring, *Statistics in Medicine*, **32**, 2173–2184.

Heagerty PJ and Zhang Y (2005). Survival model predictive accuracy and ROC curves, *Biometrics*, **61**, 92–105.

Kamarudin AN, Cox T, and Kolamunnage-Dona R (2017). Time-dependent ROC curve analysis in medical research: Current methods and applications, *BMC Medical Research Methodology*, **17**, 53.

Kim YJ and Jhun M (2008). Cure rate model for interval censored data, *Statistics in Medicine*, **27**, 3–14.

Kuk AYC and Chen C (1992). A mixture model combining logistic regression with proportional hazards regression, *Biometrika*, **79**, 531–541.

Li L, Greene T, and Hu B (2018). A simple method to estimate the time-dependent receiver operating characteristic curve and the area under the curve with right censored data, *Statistical Methods in*

Medical Research, **27**, 2264–2278.

Maller RA and Zhou X (1996). *Survival Analysis with Long-term Survivors*, Wiley, New York.

Pepe MS (2003). *The Statistical Evaluation of Medical Tests for Classification and Prediction*, Oxford University Press, USA.

Rizopoulos D, Molenberghs G, Emmanuel MEH, and Lesaffre E (2017). Dynamic predictions with time-dependent covariates in survival analysis using joint modeling and landmarking, *Biometrical Journal,* **59**, 1261–1276.

Robins JM (1993). Information recovery and bias adjustment in proportional hazards regression analysis of randomized trials using surrogate markers. In *Proceedings of the Biopharmaceutical Section, American Statistical Association*, Alexandria, Virginia, 24–33.

Robins JM and Finkelstein DM (2000). Correcting for noncompliance and dependent censoring in an AIDS clinical trial with inverse probability of censoring weighted (IPCW) log-rank tests, *Biometrics*, **56**, 779–788.

Sy JP and Taylor JM G (2000). Estimation in a Cox proportional hazards cure model, *Biometrics*, **56**, 227–236.

Uno H, Cai T, Pencina MJ, D'Agostino RB, and Wei LJ (2011). On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data, *Statistics in Medicine*, **30**, 1105–1117.

Wang Z and Wang X (2020). Evaluating the time-dependent predictive accuracy for event-to-time outcome with a cure fraction, *Pharmaceutical Statistics*, **19**, 955–974.

Yakovlev AY, Asselain B, Bardou VJ, Fourquet A, Hoang T, Rochefediere A, and Tsodikov AD (1993). A simple stochastic model of tumor recurrence and its application to data on premenopausal breast cancer, *Biometrie et Analyse de Donnees Spatio-temporelles*, **12**, 66–82.

Received October 23, 2023; Revised February 15, 2024; Accepted February 20, 2024