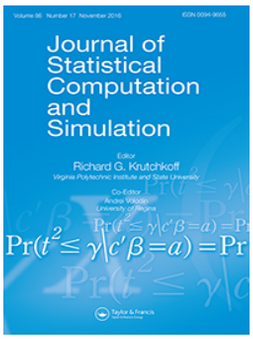

Taylor & Francis
Taylor & Francis Group

Journal of Statistical Computation and Simulation

ISSN: 0094-9655 (Print) 1563-5163 (Online) Journal homepage: www.tandfonline.com/journals/gscs20

Cumulative/dynamic ROC curve estimation

Pablo Martínez-Camblor, Gustavo F. Bayón & Sonia Pérez-Fernández

To cite this article: Pablo Martínez-Camblor, Gustavo F. Bayón & Sonia Pérez-Fernández (2016) Cumulative/dynamic ROC curve estimation, Journal of Statistical Computation and Simulation, 86:17, 3582-3594, DOI: 10.1080/00949655.2016.1175442

To link to this article: https://doi.org/10.1080/00949655.2016.1175442

JOURNAL OF STATISTICAL COMPUTATION AND SIMULATION, 2016
VOL. 86, NO. 17, 3582–3594
http://dx.doi.org/10.1080/00949655.2016.1175442

Taylor & Francis
Taylor & Francis Group

Cumulative/dynamic ROC curve estimation

Pablo Martínez-Camblor[a,b], Gustavo F. Bayón[c] and Sonia Pérez-Fernández[d]

[a]Hospital Universitario Central de Asturias (HUCA), Asturies, Spain; [b]Universidad Autonoma de Chile, Santiago, Chile; [c]Instituto Universitario de Oncología del Principado de Asturias (IUOPA), Asturies, Spain; [d]Departamento de Estadística, Universidad de Oviedo, Asturies, Spain

ABSTRACT

Receiver operating-characteristic (ROC) curve is a popular graphical method frequently used in order to study the diagnostic capacity of continuous (bio)markers. When the considered outcome is a time-dependent variable, the direct generalization is known as *cumulative/dynamic* ROC curve. For a fixed point of time, *t*, one subject is allocated into the *positive* group if the event happens before *t* and into the *negative* group if the event is not happened at *t*. The presence of *censored* subject, which can not be directly assigned into a group, is the main handicap of this approach. The proposed cumulative/dynamic ROC curve estimator assigns a probability to belong to the negative (positive) group to the subjects censored previously to *t*. The performance of the resulting estimator is studied from Monte Carlo simulations. Some real-world applications are reported. Results suggest that the new estimators provide a good approximation to the real cumulative/dynamic ROC curve.

ARTICLE HISTORY
Received 5 December 2015
Accepted 4 April 2016

KEYWORDS
Cumulative/dynamic ROC curve; censored data; Cox regression model; sensitivity; specificity

1. Introduction

The receiver operating-characteristic (ROC) curve [1] is a popular graphical method which, given a diagnostic (bio)marker, displays the false-positive rate (i.e. the inability of the marker to recognize a normal subject, without the studied characteristic, as normal) against the true-positive rate (i.e. the ability of the marker to detect the characteristic of interest, frequently one disease) for all possible thresholds. In addition, the area under the ROC curve (AUC) is frequently used as diagnostic accuracy index. Particularly, in models where the dependent variable is dichotomous, it can be read as a goodness-of-fit index, for instance, in logistic regression models. In the last decades, both ROC curve and the AUC have received great attention in the specialized literature. There exists a large number of papers which deal with both theoretical and practical aspects of the ROC curve and related problems (see, for instance, Martínez-Camblor [2] for a recent review). The general ROC curve comparison (the manuscripts of Moise et al.,[3] Venkatraman and Begg,[4] Venkatraman,[5] Bandos et al.[6] Braun and Alonzo,[7] Martínez-Camblor et al.[8] Krzanowski and Hand [9] and Martínez-Camblor et al. [10] deal with different aspects of this problem from different approaches), ROC curve regression (Cai [11] and Rodríguez-Álvarez et al. [12]), ROC curve for time-dependent events (Heagerty and Zheng,[13] Wolf et al.,[14] among others), meta-analysis of ROC curves (Rutter and Gatsonis,[15] Martínez-Camblor [16]) or the study and estimation of its associated cut-off point (Yousef et al. [17] and Martínez-Camblor [18]) are among the main focuses of interest. Of course, there also exist a

CONTACT Pablo Martínez-Camblor pablomc@ficyt.es Departamento de Estadística, Facultad de Ciencias, calle Calvo Sotelo s/n, 33006 Oviedo, Asturies, Spain

number of softwares which perform calculus and figures, just as example, the `R package`: `pROC`, `ROCR` or `rocplus`, among much others, are freely available in the `CRAN` (www.r-project.org).

The case where the studied characteristic (disease) is a time-dependent event has also been considered from different approaches (see Blanche et al. [19] for a review of this topic). The most direct ROC curve generalization is to reduce the time-dependent event to a dichotomous variable for each particular fixed moment t. Hence, assuming, without loss of generality, that larger values of the marker, X, are associated with higher probabilities of the event, the (cumulative) sensitivity, or true-positive rate, and the (dynamic) specificity, or true-negative rate, are, respectively, defined by

$$S_E^{\mathbb{C}}(x,t) = \mathcal{P}\{X > x | \, T \leq t\}, \tag{1}$$

$$S_P^{\mathbb{D}}(x,t) = \mathcal{P}\{X \leq x | \, T > t\}, \tag{2}$$

where T denotes the time variable. The resulting ROC curve, based on the above *cumulative* sensitivity and *dynamic* specificity, is known as *cumulative/dynamic* (C/D) ROC curve,[4] $\mathcal{R}_t^{\mathbb{C}/\mathbb{D}}(\cdot)$. In this approach, all subjects will be used at any fixed time t. The ith individual is considered as positive (case) if $t_i \leq t$, and as negative (control) if $t_i > t$. The cumulative/dynamic ROC curve is appropriate if the predictive model is built by using markers measured at baseline and the researcher is interested in the prognostic properties for a particular (or a small number of) time. With complete information, empirical estimators can be directly defined by,

$$\hat{S}_E^{\mathbb{C}}(x,t) = \#\{x_i > x \wedge t_i \leq t\}/\#\{t_i \leq t\}, \tag{3}$$

$$\hat{S}_P^{\mathbb{D}}(x,t) = \#\{x_i \leq x \wedge t_i > t\}/\#\{t_i > t\}. \tag{4}$$

Of course, censored data are the main handicap in order to estimate $\mathcal{R}_t^{\mathbb{C}/\mathbb{D}}$. Subjects failing before t or with a follow-up longer than t are directly allocated into the positive and the negative group, respectively. However, it is not clear what to do with those subjects censored before t; this is the case of subject D in Figure 1.

Perhaps, in this case, the first temptation is to remove these subjects from the analysis (this method is labelled as the *naive* method in [19] and we have called it *direct* method). However, removing these subjects can produce bias if the biomarker values and the censoring time are not independent. The second idea is to define an estimator based on the traditional Kaplan–Meier survivor function. However, this estimator, proposed by Heagerty et al.,[20] has serious drawbacks; (i) it can take values greater than 1, (ii) it can drive to non-monotone sensitivity or specificity functions. Alternatively, in the same paper, the authors also proposed a $\mathcal{R}_t^{\mathbb{C}/\mathbb{D}}$ estimator based on the nearest neighbour estimator (KNN) for bivariate distributions under random censoring proposed by Akritas [21]; main handicap of this estimator is that the researcher must define a smoothing parameter. Uno et al. [22] and Hung and Chiang [23] separately proposed to correct the *direct* estimator by weighting the observations kept in the subsample of uncensored subjects before time t by their probability of being kept in the subsample, that is their probability of being uncensored; from this approach the specificity estimator is

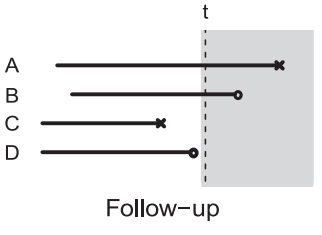

Figure 1. Schematic situation; A and B are allocated in the negative group, C in the positive group while D is not directly assigned.

the same that the one used by the *direct* method. Most recently, Wolf et al. [14] proposed an estimator based on the Nelson–Aalen cumulative incidence curve (CIC). The use of the Nelson–Aalen estimator avoids the point (i) but, as the authors recognize, additional procedures, such as isotonic regression, must be used in order to avoid the point (ii).

Other possible time-dependent generalizations for the sensitivity, S_E, and the specificity, S_P, have been proposed (see Cai et al. [24] for a recent approach). Particularly, Heagerty and Zheng [13] considered the *incident* sensitivity defined by Etzioni et al. [25] as

$$S_E^{\mathbb{I}}(x,t) = \mathcal{P}\{X > x | \, T = t\},$$

to introduce the *incident/dynamic* (I/D) ROC curve, $\mathcal{R}_t^{\mathbb{I}/\mathbb{D}}(p) = S_E^{\mathbb{I}}([1 - S_P^{\mathbb{D}}]^{-1}(t,p), t)$ $(0 \leq p \leq 1)$ where $[1 - S_P^{\mathbb{D}}]^{-1}(t,p) = \inf\{x : [1 - S_P^{\mathbb{D}}](x,t) \leq p\}$. From this approach, the ith subject is considered as control for $t_i > t$ and plays the role of case when $t_i = t$. The I/D ROC curve approach has clear links with the risk functions and, therefore, also with survival hazard models. In addition, the incident sensitivity allows a direct generalization to the case where the considered marker, X, is also a time-dependent variable $(= X(t))$. However, it is difficult to interpret without the proportional hazard assumption.

In this paper, the authors are concerned with the C/D ROC curve estimation. With this goal, the *undefined subjects* (those censored before t) are treated as *mixed subjects*. They are not completely assigned to a group; but the probability of to be or not be in the group is considered. Rest of the paper is organized asfollows ; in Section 2 the proposed estimator is described and its performance is evaluated, via Monte Carlo simulations, in Section 3. Section 4 is devoted to real application; particularly, the relationship between the forced ventilatory volume 1 sec. (FEV_1) and mortality in chronic obstructive pulmonary disease (COPD) patients is studied by using the *COllaborative COhorts to assess Multicomponents Indices of COPD in Spain* (COCOMICS) dataset (see [26]). In Section 5 we present our conclusions. Finally, as appendix, we provide a set of `R functions` which are useful in order to handle, in practice, the proposed methodology.

2. Cumulative/dynamic ROC curve estimation

Conventionally, let $\{y_i\}_{1\leq i\leq N} = \{z_i, \delta_i, x_i\}_{1\leq i\leq N}$ be an independent random sample (with size N) where for $1 \leq i \leq N$, z_i stands for the observed time; $z_i = min\{t_i, \, c_i\}$, with t_i being the time to event and c_i the censoring time, δ_i is the status (δ_i takes the value 1 if $z_i = t_i$ and 0 if $z_i = c_i$) and x_i stands for the (bio)marker value. From the Bayes theorem, the definitions of sensitivity and specificity given in (1) and (2) are equivalent to

$$S_E^{\mathbb{C}}(x,t) = \frac{\mathcal{P}\{X > x \wedge T \leq t\}}{\mathcal{P}\{T \leq t\}} = \frac{\int \mathcal{P}\{X > x \wedge T \leq t | y\} \, \mathrm{d}\boldsymbol{F}_Y}{\int \mathcal{P}\{T \leq t | y\} \, \mathrm{d}\boldsymbol{F}_Y}, \tag{5}$$

$$S_P^{\mathbb{D}}(x,t) = \frac{\mathcal{P}\{X \leq x \wedge T > t\}}{\mathcal{P}\{T > t\}} = \frac{\int \mathcal{P}\{X \leq x \wedge T > t | y\} \, \mathrm{d}\boldsymbol{F}_Y}{\int \mathcal{P}\{T > t | y\} \, \mathrm{d}\boldsymbol{F}_Y}, \tag{6}$$

where y is the observed value ($\{z, \delta, x\}$) and $\boldsymbol{F}_Y$ its cumulative distribution function . Note that, in the sampling context, for $i \in \{1, \ldots N\}$, $\mathcal{P}\{X > x \wedge T \leq t | y_i\} = \mathcal{P}\{T \leq t | y_i\} \cdot I_{(x,\infty)}(x_i)$ ($I_A(x)$ is the indicator function; takes the value 1 if $x \in A$ and 0 otherwise). Then, if $\hat{P}_i$ $(= \hat{P}_i(N))$ is an adequate estimator for $\mathcal{P}\{T > t | y_i\}$, the empirical estimators for the sensitivity and the specificity can be written as,

$$\hat{S}_E^{\mathbb{C}}(x,t) = \frac{\sum_{i=1}^N (1 - \hat{P}_i) \cdot I_{(x,\infty)}(x_i)}{\sum_{i=1}^N (1 - \hat{P}_i)}, \tag{7}$$

$$\hat{S}_P^{\mathbb{D}}(x,t) = \frac{\sum_{i=1}^N \hat{P}_i \cdot I_{(-\infty,x]}(x_i)}{\sum_{i=1}^N \hat{P}_i}. \tag{8}$$

Obviously, $z_i > t$ implies $\hat{P}_i = 1$ and an event previous to t implies $\hat{P}_i = 0$; therefore, for complete information, these estimators are the usual empirical ones. In addition, the consistency of the above estimators is direct if the estimator of $\mathcal{P}\{T > t|y_i\}$ $(1 \leq i \leq N)$ has appealing conditions, particularly, that $|\hat{P}_i - \mathcal{P}\{T > t|y_i\}| \to_N 0$ in probability. On the other hand, it is obvious that the resulting C/D ROC curve overcomes the drawbacks reported by the Kaplan–Meier-based C/D ROC curve: for a fixed point t, it is monotone and always takes values below or equal to 1.

Different procedures can be used in order to estimate the above probability in those subjects which are not absolutely defined. We proposed a semi-parametric one; by using a proportional hazard Cox regression model; and a non-parametric proposal; by using directly the Kaplan–Meier estimator. From the proportional hazard Cox regression model, we can estimate the hazard function $\lambda(t) = \lambda_0(t)\exp\{\beta \cdot X\}$, this quantity allows to compute $\hat{P}_i = \hat{S}(t|X = x_i)/\hat{S}(z_i|X = x_i)$, where $\hat{S}$ is the survival function estimated from the Cox regression model. Unfortunately, due to X is usually a continuous variable, for the Kaplan–Meier method we cannot directly estimate the probability for one particular value $X = x_i$ and this condition is replaced by $X \leq x_i$; then we select values satisfying $X \leq x_i$ and compute the Kaplan–Meier estimator to obtain $\hat{P}_i = \hat{S}_{KM}(t)/\hat{S}_{KM}(z_i)$, where $\hat{S}_{KM}$ is the survival function estimated by the Kaplan–Meier method referred to those subjects satisfying $X \leq x_i$.

Indeterminate or *mixed subjects* play a fundamental role. Note that an absolute uncertainty about the situation of the ith subject at time t $(\hat{P}_i = 1/2)$ implies that, when this subject is allocated into a group, we will commit an error with a probability of 1/2, with independence of the group in which it will be allocated; in these cases, the capacity diagnostic of the studied biomarker will be limited.

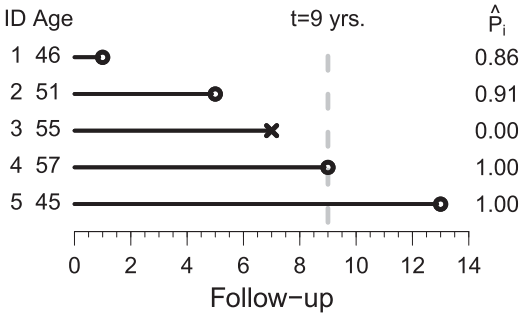

Figure 2. Schematic situation for IDs 1–5 of the kidney transplantation data.

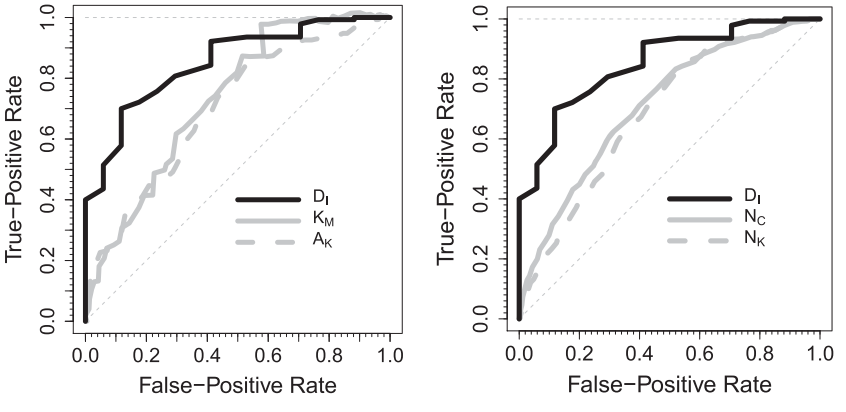

Figure 3. At left, direct ROC curve estimation (D_I; removing the *mixed subjects*), Kaplan–Meier (K_M) and KNN (A_K; with span= $0.01 \cdot N^{-1/5}$) based cumulative/dynamic ROC curve estimations. At right, D_I (as references) and proposed estimations based on both Cox regression (N_C) and Kaplan–Meier (N_K).

In order to illustrate the problem, we have considered a real problem. Particularly, we have considered a dataset which contains information about 863 ($= N$) kidney transplant patients. These data, free available within the `R package KMsurv`, have been previously used with the same goal by Wolf et al.[14] Information about time to death and age of patients (in this case used as a mortality marker) were collected. Interested reader is referred to Klein and Moeschberger [27] for more details about the data. Figure 2 depicts the situation for the first five subjects when the considered time is 9 years: the patient with ID = 3 died the seventh year, hence it is within the positive group; patients 4 and 5 were still alive at the ninth year, therefore, they are within the negative group. However, the

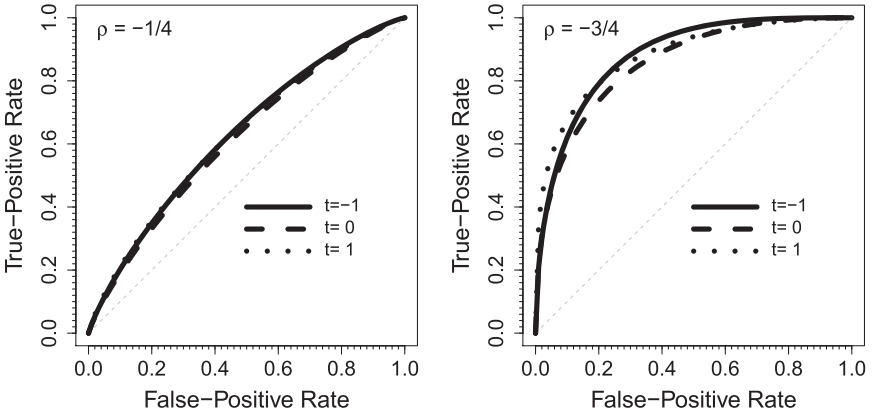

Figure 4. Real ROC curves from where the samples were drawn.

Table 1. Mean $\pm$ standard deviation of $0.01 \cdot \sqrt{N} \cdot \int_0^1 |\hat{\mathcal{R}}(p) - \mathcal{R}(p)| \mathrm{d}p$, where $\mathcal{R}$ is the real C/D ROC curve and $\hat{\mathcal{R}}$ is its estimation, computed from 5000 Monte Carlo iterations for $\tau = 0$.

N	ρ	%C	$\log(t)$	K_M	A_K	D_I	N_C	N_K
100	$-\frac{1}{4}$	20	−1	0.834 ± 0.376	0.833 ± 0.399	0.827 ± 0.365	0.828 ± 0.373	0.830 ± 0.375
			0	0.613 ± 0.271	0.614 ± 0.289	0.625 ± 0.294	0.587 ± 0.266	0.591 ± 0.267
			1	0.934 ± 0.463	0.920 ± 0.460	0.887 ± 0.443	0.767 ± 0.392	0.767 ± 0.415
100	$-\frac{3}{4}$	20	−1	0.476 ± 0.214	0.474 ± 0.248	0.471 ± 0.214	0.468 ± 0.214	0.471 ± 0.214
			0	0.424 ± 0.182	0.429 ± 0.212	0.410 ± 0.179	0.390 ± 0.173	0.399 ± 0.181
			1	0.591 ± 0.320	0.623 ± 0.362	0.486 ± 0.283	0.411 ± 0.182	0.476 ± 0.176
100	$-\frac{1}{4}$	50	−1	0.868 ± 0.391	0.866 ± 0.420	0.853 ± 0.374	0.801 ± 0.368	0.835 ± 0.385
			0	0.744 ± 0.338	0.745 ± 0.358	0.750 ± 0.350	0.585 ± 0.299	0.614 ± 0.310
			1	1.519 ± 0.866	1.092 ± 0.542	1.123 ± 0.453	0.813 ± 0.483	0.956 ± 0.646
100	$-\frac{3}{4}$	50	−1	0.532 ± 0.231	0.492 ± 0.250	0.489 ± 0.216	0.448 ± 0.211	0.481 ± 0.221
			0	0.570 ± 0.243	0.533 ± 0.264	0.487 ± 0.215	0.365 ± 0.171	0.466 ± 0.243
			1	1.511 ± 1.065	1.051 ± 0.667	1.118 ± 1.009	0.391 ± 0.181	0.761 ± 0.490
200	$-\frac{1}{4}$	20	−1	0.821 ± 0.356	0.807 ± 0.381	0.831 ± 0.387	0.816 ± 0.354	0.818 ± 0.356
			0	0.618 ± 0.267	0.612 ± 0.286	0.630 ± 0.297	0.591 ± 0.260	0.595 ± 0.261
			1	0.950 ± 0.442	0.952 ± 0.472	0.997 ± 0.496	0.784 ± 0.371	0.784 ± 0.398
200	$-\frac{3}{4}$	20	−1	0.484 ± 0.201	0.499 ± 0.247	0.481 ± 0.201	0.478 ± 0.201	0.481 ± 0.201
			0	0.435 ± 0.181	0.435 ± 0.208	0.424 ± 0.177	0.400 ± 0.171	0.412 ± 0.179
			1	0.596 ± 0.258	0.618 ± 0.328	0.539 ± 0.245	0.428 ± 0.181	0.532 ± 0.285
200	$-\frac{1}{4}$	50	−1	0.866 ± 0.380	0.856 ± 0.407	0.879 ± 0.412	0.801 ± 0.358	0.835 ± 0.374
			0	0.749 ± 0.334	0.750 ± 0.359	0.784 ± 0.379	0.589 ± 0.300	0.621 ± 0.313
			1	1.599 ± 0.882	1.284 ± 0.638	1.453 ± 0.703	0.859 ± 0.509	0.995 ± 0.671
200	$-\frac{3}{4}$	50	−1	0.535 ± 0.224	0.525 ± 0.264	0.498 ± 0.209	0.456 ± 0.213	0.494 ± 0.219
			0	0.581 ± 0.238	0.529 ± 0.247	0.536 ± 0.223	0.375 ± 0.170	0.412 ± 0.263
			1	1.278 ± 0.755	1.177 ± 0.689	1.173 ± 0.642	0.457 ± 0.220	1.042 ± 0.517

Table 2. Mean $\pm$ standard deviation of $0.01 \cdot \sqrt{N} \cdot \int_0^1 |\hat{\mathcal{R}}(p) - \mathcal{R}(p)|dt$, where $\mathcal{R}$ is the real C/D ROC curve and $\hat{\mathcal{R}}$ is its estimation, computed from 5000 Monte Carlo iterations for $\tau = \frac{1}{4}$.

N	ρ	%C	$\log(t)$	K_M	A_K	D_I	N_C	N_K
100	$-\frac{1}{4}$	20	-1	0.828 ± 0.375	0.829 ± 0.402	0.829 ± 0.370	0.825 ± 0.375	0.830 ± 0.376
			0	0.601 ± 0.270	0.614 ± 0.294	0.614 ± 0.301	0.588 ± 0.269	0.604 ± 0.277
			1	0.867 ± 0.421	0.913 ± 0.446	0.914 ± 0.407	0.754 ± 0.384	0.823 ± 0.424
100	$-\frac{3}{4}$	20	-1	0.468 ± 0.207	0.465 ± 0.238	0.470 ± 0.207	0.466 ± 0.205	0.468 ± 0.205
			0	0.411 ± 0.179	0.421 ± 0.196	0.415 ± 0.182	0.389 ± 0.167	0.399 ± 0.169
			1	0.532 ± 0.291	0.536 ± 0.281	0.522 ± 0.334	0.398 ± 0.169	0.412 ± 0.208
100	$-\frac{1}{4}$	50	-1	0.830 ± 0.382	0.872 ± 0.420	0.845 ± 0.364	0.803 ± 0.372	0.855 ± 0.396
			0	0.657 ± 0.394	0.748 ± 0.354	0.741 ± 0.339	0.582 ± 0.299	0.672 ± 0.345
			1	1.253 ± 0.698	1.054 ± 0.495	1.078 ± 0.409	0.779 ± 0.462	1.123 ± 0.657
100	$-\frac{3}{4}$	50	-1	0.509 ± 0.235	0.484 ± 0.239	0.494 ± 0.225	0.450 ± 0.206	0.482 ± 0.205
			0	0.563 ± 0.257	0.499 ± 0.223	0.493 ± 0.235	0.368 ± 0.168	0.421 ± 0.189
			1	1.271 ± 0.909	0.698 ± 0.388	1.240 ± 1.041	0.343 ± 0.162	0.539 ± 0.369
200	$-\frac{1}{4}$	20	-1	0.822 ± 0.359	0.818 ± 0.384	0.832 ± 0.381	0.820 ± 0.359	0.824 ± 0.361
			0	0.605 ± 0.266	0.614 ± 0.289	0.638 ± 0.305	0.595 ± 0.266	0.612 ± 0.275
			1	0.870 ± 0.402	0.938 ± 0.452	1.029 ± 0.510	0.758 ± 0.374	0.836 ± 0.422
200	$-\frac{3}{4}$	20	-1	0.474 ± 0.203	0.493 ± 0.246	0.476 ± 0.204	0.472 ± 0.202	0.474 ± 0.201
			0	0.430 ± 0.188	0.430 ± 0.202	0.433 ± 0.189	0.406 ± 0.173	0.418 ± 0.176
			1	0.561 ± 0.264	0.547 ± 0.258	0.559 ± 0.288	0.423 ± 0.174	0.433 ± 0.202
200	$-\frac{1}{4}$	50	-1	0.831 ± 0.371	0.859 ± 0.405	0.881 ± 0.410	0.803 ± 0.364	0.855 ± 0.392
			0	0.674 ± 0.307	0.753 ± 0.364	0.822 ± 0.406	0.593 ± 0.305	0.692 ± 0.356
			1	1.294 ± 0.691	1.222 ± 0.570	1.360 ± 0.619	0.818 ± 0.493	1.217 ± 0.721
200	$-\frac{3}{4}$	50	-1	0.535 ± 0.230	0.497 ± 0.228	0.499 ± 0.206	0.451 ± 0.188	0.488 ± 0.192
			0	0.633 ± 0.274	0.515 ± 0.213	0.518 ± 0.221	0.385 ± 0.172	0.444 ± 0.193
			1	1.129 ± 0.678	0.786 ± 0.414	0.931 ± 0.797	0.391 ± 0.186	0.625 ± 0.442

real situation for patients 1 and 2 is unknown. The probability that those patients are still alive nine years after the follow-up can be estimated by using the Kaplan–Meier estimator. Particularly, for the first subject, we must make the estimation considering only the subjects with age smaller than 46 resulting the estimation $\hat{P}_1 = 0.88/0.95\ (= 0.86)$. Similarly, for the second subject, considering only the subjects with age smaller than 51, it is obtained the value $\hat{P}_2 = 0.91$; in the proposed model, both are considered as *mixed subjects*.

Figure 3 depicts different ROC curve estimations. At left, in order to make the usual ROC curve (D_I); the unclassified subjects, called *mixed*, were removed and a total of 140 positives and 17 negatives were finally considered. Both the Kaplan–Meier (K_M) and the KNN (A_K) based estimations for $\mathcal{R}_t^{\mathbb{C}/\mathbb{D}}$ ($t = 9$ years.) are also depicted (computed by using the `R package`: `survivalROC`). Observed differences between the A_K and the CIC estimations were negligible when the `span` parameter (related with the KNN method) is close to zero. At right, the same usual ROC curve (plotted as references); and the C/D ROC curve estimations by using the method proposed above based on both the proportional hazard Cox regression (N_C) and the non-parametric Kaplan–Meier estimator (N_K).

3. Simulation study

In order to study the practical behaviour of the proposed methodology, a Monte Carlo simulation study was carried out. Similarly to Heatherty and Zheng,[13] the joint distribution of $\{\log(Time), Marker\}$ is a standard bivariate normal distribution with correlation coefficient ρ (cases $-\frac{1}{4}$ and $-\frac{3}{4}$ were studied). In addition, two different sample sizes ($N = 100,\ 200$) were explored. The considered distribution for the censoring time, $\log(C)$, was also normal with standard deviation 1 and mean 0 ($\mathcal{P}\{C < T\} = \frac{1}{2}$) and 1.19 ($\mathcal{P}\{C < T\} = \frac{1}{5}$). The censoring time was drawn independently of the time to event, but not of (bio)marker values, which are related in such a way that $\mathbb{E}[\log(C) \cdot M] =$

τ (cases $\tau = 0$, $\frac{1}{4}$ were considered). C/D ROC curve estimations for $\log(Time) = -1,\ 0,\ 1$ were computed based on the Kaplan–Meier estimator (K_M), the KNN method with $\texttt{span} = 0.01 \cdot N^{-1/5}$ (A_K), the *direct method*, which removes the subjects with incomplete information, called *mixed subjects* in this paper (D_I), and by using the proposed estimator based on both the Cox regression (N_C) and Kaplan–Meier estimator (N_K). Method proposed by Uno et al. [22] and Hung and Chiang [23] has not been considered because as is quoted by Blanche et al. [19] its results are similar to KNN. Figure 4 depicts the theoretical ROC curves from where the samples were drawn.

Table 1 shows the mean $\pm$ standard deviation for $0.01 \cdot \sqrt{N} \cdot \int_0^1 |\hat{\mathcal{R}}(p) - \mathcal{R}(p)| \mathrm{d}p$, where $\mathcal{R}$ is the real C/D ROC curve and $\hat{\mathcal{R}}$ is its estimation when $\tau = 0$ (censoring time and (bio)marker values are drawn independently). Five studied methods performed similarly. Proposed N_C statistics obtained better results than N_K (Note that in the simulated scenario it is satisfied the proportional hazard assumption.) and it was the best of all in most cases. As expected, the observed differences between the proposed method and the other ones were clearer when the censored percentages were larger, particularly, for % C= 50 and $\log(t) = 1$, the proposed methodology achieved the best results.

Table 2 is the same as Table 1 but considering that $\tau = \frac{1}{4}$. Observed results were similar to the previous obtained ones. It seems that five considered estimators are robust respect to this correlation configuration between the (bio)marker and censoring time.

4. Real-data application: the COCOMICS study

The proposed methodology is applied to study the capacity of FEV_1 to predict mortality in COPD patients. With this goal, we consider the data of the (COCOMICS study). This dataset included 11 Spanish cohorts with a total of 3633 patients out of a total of 15,878.17 people per year. The interested reader is referred to Soriano et al. [28] and to Marin et al. [26] for complete information about the

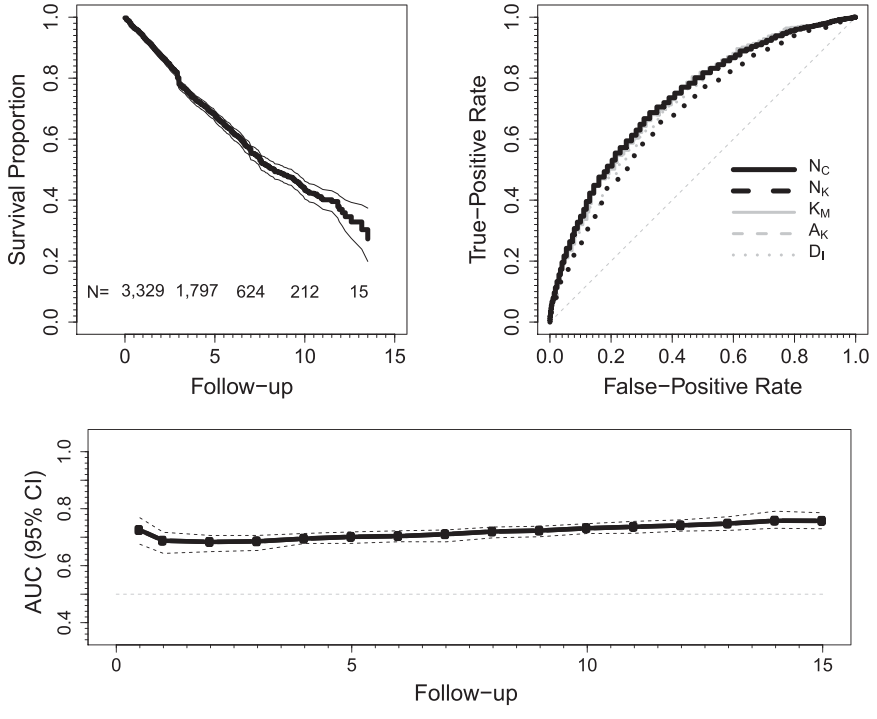

Figure 5. Kaplan–Meier survival curve estimation, top-left; C/D ROC curve estimation at ten years by using the five considered methods, top-right; AUC evolution with a 95% confidence interval, down.

data. Figure 5 (up-left) depicts the Kaplan–Meier estimation with a 95% confidence interval for the COCOMICS data and the number of patients at risk at 1, 4, 7, 10 and 13 years.

In this dataset, differences between the five considered C/D ROC curve estimations at 10 years were really small (see Figure 5, top-right). The proposed estimator based on Kaplan–Meier showed itself more conservative than the other ones. At this point, 1212 subjects were classified as positives and 198 were allocated into the negative group.

The prediction capacity of the (bio)marker (measured from the AUC) was high along the follow-up time. It was about 0.73 at 6 months to decrease until 0.68 at 2 years (lowest value); after that, the diagnostic capacity increases until 0.76 at 14 years. The integrated AUC along the follow-up time was 0.72. Figure 5 (down) depicts the AUC evolution with a 95% confidence interval (based on bootstrap replications).

5. Conclusions

In this paper the authors deal with the cumulative/dynamic ROC curve estimation in the presence of right-censored data. The proposed methodology assigns a probability of belonging to the negative group (respectively to the positive group) to those patients whose real status at considered point remains unknown (those censored before the considered point). This probability can be estimated from different methods; in this paper a semi-parametric (based on the Cox regression) and a non-parametric (based on the Kaplan–Meier) ones are considered. However, other methods can be used with this goal, even methods adapted to particular situations, ranging from competing-risk or multi-state contexts, to the non-monotone relationship ROC curve generalization.[29] The proposed estimator avoids the drawbacks of the previously existing ones; (i) it is monotone and always ranges between 0 and 1, (ii) it does not depend on smooth parameters. In the real-data example, the five considered estimators performed similarly, but simulation suggests that both N_C and N_K perform well and they are always better than the previous ones when the observed percentage of censorship is high. When some correlation between the (bio)marker and the censoring time is added, obtained results do not change substantially.

It should be noted that the direct method, removing the *undefined* subjects, usually obtains good results. However, do not use these individuals is not a good methodological practice; we are removing a particular subset of individuals which can have some interesting properties. For instance, if there exists some kind of relationship between the considered marker and the censoring time, the results will be strongly biased.

Perhaps, the main flaws in the presented methodology is the lack of a rigorous study of its theoretical properties. However, the proposed estimator has two different parts; the first one is the traditional ROC curve for complete data, and the second one is to assign, to each mixed subject, the probability that the event has not happened before the fixed point t. This probability depends on several factors which are not previously clear. At this point, it is good to remark again the good results provided by the simulation study. Note that, this approach is absolutely different to the *weighting* approach proposed by Uno et al. [22] and Hung and Chiang.[23]

Finally, as appendix, we provide some `R functions` which compute the proposed methodology. The authors expect that a suitable `R package` will be soon available on the `CRAN`.

Acknowledgments

The authors are grateful to the COCOMICS group for the permission to use the data.

Disclosure Statement

No potential conflict of interest was reported by the authors.

Funding

This paper was financial supported from Grants MTM2014-55966-P and MTM2015-63971-P of the Spanish Ministerio de Economía y Competitividad and by Grant FC-15-GRUPIN14-101 from the Principado de Asturias.

References

[1] Green DM, Swets JA. Signal detection theory and psychophysics. New York: Wiley; 1966.
[2] Martínez-Camblor P. Area under the ROC curve comparison in the presence of missing data. J Korean Statist Soc. 2013;42(4):431–442.
[3] Moise A, Clement B, Raissis M. A test for crossing receiver operating characteristic (ROC) curves. Commun Stat –Theory Methods. 1988;17:1985–2003.
[4] Venkatraman ES, Begg CB. A distribution-free procedure for comparing receiver operating characteristic curves from a paired experiment. Biometrika. 1996;83(4):835–848.
[5] Venkatraman ES. A permutation test to compare receiver operating characteristic curves. Biometrics. 2000;56:1134–1138.
[6] Bandos AI, Rockette HE, Gur D. A permutation test sensitive to differences in areas for comparing ROC curves from paired design. Stat Med. 2005;24:2873–2893.
[7] Braun T, Alonzo TA. A modified sign test for comparing paired ROC curves. Biostatistics. 2008;9:364–372.
[8] Martínez-Camblor P, Carleos C, Corral N. Powerful nonparametric statistics to compare *k*-independent ROC curves. J Appl Stat. 2011;38(7):1317–1332.
[9] Krzanowski WJ, Hand DJ. Testing the difference between two Kolmogorov–Smirnov values in the context of receiver operating characteristic curves. J Appl Stat. 2011;38(3):437–450.
[10] Martínez-Camblor P, Carleos C, Corral N. General nonparametric ROC curves comparison. J Korean Statist Soc. 2013;42:71–81.
[11] Cai T. Semi-parametric ROC regression analysis with placement values. Biostatistics. 2004;5(1):45–60.
[12] Rodríguez-Álvarez MX, Tahoces PG, Cadarso-Suárez C, Lado MJ. Comparative study of ROC regression techniques-Applications for the computed-aided diagnostic system in breast cancer detection. Comput Stat Data Anal. 2011;55(1):888–902.
[13] Heagerty P, Zheng Y. Survival model predictive accuracy and ROC curves. Biometrics. 2005;61(1):92–105.
[14] Wolf P, Schmidt G, Ulm K. The use of ROC for defining the validity of the prognostic index in censored data. Stat Probab Lett. 2011;81:783–791.
[15] Rutter CM, Gatsonis CA. A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. Stat Med. 2001;20:2865–2884.
[16] Martínez-Camblor P. Fully non-parametric ROC curve estimation for random-effects meta-analysis. Statist Methods Med Res. 2014, doi:10.1177/0962280214537047.
[17] Yousef WA, Kundu S, Wagner RF. Nonparametric estimation of the threshold at an operating point on the ROC curve. Comput Stat Data Anal. 2009;33(12):4370–4383.
[18] Martínez-Camblor P. Nonparametric cut-off point estimation for diagnostic decisions with weighted errors. Rev Colombiana Estadística. 2011;34(1):133–146.
[19] Blanche P, Dartigues JF, Jacqmin-Gadda1 H. Review and comparison of ROC curve estimators for a time-dependent outcome with marker-dependent censoring. Biom J. 2013;5:687–704.
[20] Heagerty PJ, Lumley T, Pepe MS. Time-dependent ROC curves for censored survival data and a diagnostic marker. Biometrics. 2000;56:337–344.
[21] Akritas MG. Nearest neighbor estimation of a bivariate distribution under random censoring. Ann Stat. 1994;22(3):1299–1327.
[22] Uno H, Cai T, Tian L, Wei LJ. Evaluating prediction rules for t-Year survivors with censored regression models. J Am Stat Assoc. 2007;102(478):527–537.
[23] Hung H, Chiang CT. Estimation methods for time-dependent AUC models with survival data. Can J Stat. 2010;38(1):8–26.
[24] Cai T, Pepe MS, Zheng Y, Lumley T, Jenny NS. The sensitivity and specificity of markers for event times. Biostatistics. 2006;7(2):182–197.
[25] Etzioni R, Pepe MS, Longton G, Hu C, Goodman G. Incorporating the time dimension in receiver operating characteristic curves: a case study of prostate cancer. Med Decis Mak. 1999;19:242–251.
[26] Marin JM, Alfageme I, Almagro P, et al. Multicomponent indices to predict mortality in COPD: the COCOMICS study. Eur Respiratory J. 2013;42:323–332.
[27] Klein JP, Moeschberger ML. Survival analysis: techniques for censored and truncated data. New York: Springer; 2003.
[28] Soriano J, Alfageme I, Almagro P, Casanova C, et al. Distribution and prognostic validity of the new global initiative for chronic obstructive lung disease grading classification. Chest. 2013;143(3):694–702.

[29] Martínez-Camblor P, Corral N, Rey C, Pascual J, Cernuda-Morollón E. ROC curve generalization for non-monotone relationships. Stat Methods Med Res. 2014, doi:10.1177/0962280214541095.

Appendix: R functions

The following `R functions` make the computes and display plots for the proposed C/D ROC curve estimator. The notation used is direct.

```
assignProbability <- function(stime, status, marker, predict.time, method=c('Cox', 'KM'),
                              undefinedIndices) {
                                  A function which assigns the probability of belonging to the negative group
                                  to those patients censored before the considered point
                                  - stime: observed time of each subject
                                  - status: status of each subject (takes the value 0 if the
                                  subject is censored and 1 otherwise)
                                  - marker: (bio)marker values
                                  - predict.time: considered time point
                                  - method: procedure used in order to estimate the probability;
                                  'Cox' stands for proposed method based on Cox regression
                                  'KM' stands for proposed method based on Kaplan-Meier estimator
                                  - undefinedIndices: indices of mixed subjects
  method <- match.arg(method)
  results <- 1:length(undefinedIndices)
                                  vector results will contain the probabilities
                                  of mixed subjects

  Proposed method based on Cox regression

  if (method == 'Cox') {
    fit <- coxph(Surv(stime, status) ~ marker)
                                  fits a Cox proportional hazard regression model whose
                                  covariate is marker, i.e., h(t) = h_0(t) · e^{βX}
    md <- survfit(fit, newdata=data.frame(cbind(stime,status,marker)))
                                  survival curves from the previously fitted Cox model

    for (j in 1:length(undefinedIndices))
                                  for each mixed subject
    {
      f <- approxfun(c(min(md$time)-1, md$time, max(md$time)+1),
                     c(1, md$surv[, undefinedIndices[j]], 0))

                                  f constains a list of points which linearly interpolate given data points,
                                  so it is an approximation of the survival function estimated from the
                                  Cox regression model
      results[j] <- f(predict.time) / f(stime[undefinedIndices[j]])
                                  probability considered in mixed subjects based on
                                  Cox regression model

      if (is.na(results[j])) {
                                  if results[j] is a missing value, it takes the value 1, so the
                                  j-th mixed subject is assigned to the negative group
        results[j] <- 1
      }
    }

  Proposed method based on Kaplan-Meier estimator

  } else if (method == 'KM') {
    for (j in 1:length(undefinedIndices))
                                  for each mixed subject
    {
      idx <- which(marker <= marker[undefinedIndices[j]])

                                  idx constains the indices of those subjects whose marker is smaller than
                                  or equal to the marker of the considered mixed subject
      fit <- survfit(Surv(stime[idx], status[idx]) ~ 1)
                                  survival curve from Kaplan-Meier estimator using only the
                                  previous subjects
      f <- stepfun(fit$time, c(1, fit$surv))
                                  interpolating step function from survival curve above
      results[j] <- f(predict.time) / f(stime[undefinedIndices[j]])
                                  probability considered in mixed subjects based on
                                  Kaplan-Meier estimator

      if (is.na(results[j])) {
                                  if results[j] is a missing value, it takes the value 1, so the
                                  j-th mixed subject is assigned to the negative group
        results[j] <- 1
      }
```

```
    }
    return(results)                                   return probabilities of mixed subjects
}


singleCdroc <- function(stime, status, marker, predict.time, method=c('Cox', 'KM')) {
                                  A function which returns the estimated sensitivity, specificity, AUC and other aspects
                                  of cumulative/dynamic ROC curve estimation based on the proposed methodology
  method <- match.arg(method)
  positiveIndices <- which(stime <= predict.time & status == 1)
                                                  indices of those subjects whose event ocurrs before the considered
                                                                                  time point (predict.time)
  negativeIndices <- which(stime > predict.time)
                                                  indices of those subjects whose observed time is longer than predict.time
  undefinedIndices <- which(stime <= predict.time & status == 0)
                                                  indices of those subjects censored before predict.time (mixed subjects)
  undefinedProb <- NULL
  if (length(undefinedIndices) > 0) {                                  if there exists some mixed subjects
    undefinedProb <- assignProbability(stime, status, marker, predict.time, method,
                                       undefinedIndices)
                                                  undefinedProb is a vector containing the probability of belonging to
                                                                  the negative group of mixed subjects
  }
  cutPoints <- c(min(marker) - 1, unique(sort(marker)), max(marker) + 1)
                                  different thresholds considered will be the different values of marker on the data;
                                                  it will be also considered min(marker)-1 and max(marker)+1
  nSens <- length(positiveIndices) + sum(1 - undefinedProb)
                                  denominator of the estimator for the sensitivity considered in page 6, taking into
                                  account that "positive patients" have a probability of belonging to the negative
                                                                                  group equals zero
  nSpec <- length(negativeIndices) + sum(undefinedProb)
                                  denominator of the estimator for the specificity considered in page 6, taking into
                                  account that "negative patients" have a probability of belonging to the negative
                                                                                  group equals one
  sensitivity <- cutPoints                                  both sensitivity and specificity vectors will have
  specificity <- cutPoints                                  the same length than cutPoints vector
  for (i in 1:length(cutPoints))                            for each considered threshold for the (bio)marker
  {
  sensitivity[i] <- (sum(marker[positiveIndices] > cutPoints[i]) +
                       sum(1 - undefinedProb[marker[undefinedIndices] > cutPoints[i]])) / nSens
                                                  sensitivity estimation from expression in page 6
  specificity[i] <- (sum(marker[negativeIndices] <= cutPoints[i]) +
                       sum(undefinedProb[marker[undefinedIndices] <= cutPoints[i]])) / nSpec
                                                  specificity estimation from expression in page 6
  }
  rocFunction <- approxfun(1 - specificity, sensitivity)
                                  rocFunction contains a list of points which linearly interpolate given data
                                  points of the cumulative/dynamic ROC curve estimation
  auc <- integrate(rocFunction,0,1)
                                  AUC estimation by adaptive quadrature of rocFunction over the unit interval
  results <- list(TP=sensitivity, TN=specificity, undefinedProb=undefinedProb,
                  cutPoints=cutPoints, auc=auc$value, aucAbsError=auc$abs.error,
                  predict.time=predict.time, method=method)
                                  output:
                                  - TP: sensitivity estimation, true positive rate
```

```
                                                    - TN: specificity estimation, true negative rate
                                                    - undefinedProb: probability of belonging to the
                                                    negative group to those mixed subjects
                                                    - cutPoints: thresholds considered
                                                    - auc: AUC, the final estimate of the integral above
                                                    - aucAbsError: estimation of the modulus of
                                                    the absolute error from integral
                                                    estimation above
                                                    - predict.time: considered time point
                                                    - method: procedure used in order to estimate
                                                    the probability considered; it can takes
                                                    the value "Cox" or "KM"
    attr(results, 'class') <- 'cdroc'

    return(results)                                                    return outputs above
}


cdroc <- function(stime, status, marker, predict.time, method=c('Cox', 'KM'), ci=FALSE,
                  boot.n=100, conf.level=0.95, seed=2032) {

                                    A function which returns outputs of singleCdroc function above, including some aspects
                                    about a conf.level confidence interval (if ci is TRUE) for the area under the
                                    cumulative/dynamic ROC curve
                                    - boot.n: number of bootstrap replications considered
                                    - seed: seed considered (for reproducibility)

    method <- match.arg(method)

    if (ci) {                                    if the user is interested in calculating confidence interval
        set.seed(seed)                                                    specify the seed
        sampledIndices <- sapply(1:boot.n, function(xx) sample(length(stime), length(stime),
                                  replace=TRUE))

                                    sampledIndices is a matrix containing the indices of each bootstrap replication

        allResults <- lapply(1:boot.n, function(index, sampledIndices) {
                        currentIndices <- sampledIndices[, index]
                        singleCdroc(stime[currentIndices], status[currentIndices],
                                    marker[currentIndices], predict.time, method) },
                        sampledIndices=sampledIndices)
                                    apply the function singleCdroc to each bootstrap sample

        result <- singleCdroc(stime, status, marker, predict.time, method)
                                    apply the function singleCdroc to the main sample
        allAucs <- sapply(allResults, function(xx) xx$auc)
                                    vector allAucs contains AUC estimations of each bootstrap sample
        result$meanAuc <- mean(allAucs)                                    mean of AUC estimations
        result$ciAuc <- quantile(allAucs, c(1 - conf.level, conf.level))
                                    conf.level confidence interval for AUC
        result$ci <- TRUE
        result$boot.n <- boot.n
        result$conf.level <- conf.level
        result$seed <- seed
        result$aucs <- allAucs

        return(result)                                                    return outputs above

    } else {                                    if the user is not interested in calculating confidence interval

        return(singleCdroc(stime, status, marker, predict.time, method))
                                    only return outputs of singleCdroc function
    }
}
```

```
summary.cdroc <- function(obj) {
                                        A function which returns some aspects about obj (class 'cdroc')
      cat('cdroc object: \n')
      cat('Number of cut points:', length(obj$cutPoints), '\n')
      cat('Method:', obj$method, '\n')
      cat('predict.time:', obj$predict.time, '\n')
      cat('AUC:', obj$auc, '\n')
      if (!is.null(ci)) {
          cat('Bootstrap number of replicates:', obj$boot.n, '\n')
          cat('Bootstrap AUC:', obj$meanAuc, '\n')
          cat('Bootstrap AUC Confidence Interval:', obj$ciAuc, '\n')
          cat('Bootstrap AUC Confidence Level:', obj$conf.level, '\n')
          cat('Bootstrap seed:', obj$seed, '\n')
          }
}



print.cdroc <- function(obj) {
                                        A function which returns a summary of obj (class 'cdroc')
      summary(obj)
}



plot.cdroc <- function(obj, ...) {
 A function which plots obj (class 'cdroc') and the worst ROC curve possible (the diagonal from (0,0) to (1,1), whose AUC
                                                                                                        is 0.5)
                                        ...: arguments to be passed to methods, such as graphical parameters
      plot(1 - obj$TN, obj$TP, type='l', lwd=3, xlab='1 - specificity', ylab='sensitivity', xaxs='i',
            yaxs='i', ...)
      abline(0, 1, lty=2)
}
```