Black Sea Journal of Health Science
doi: 10.19127/bshealthscience.1107599

Open Access Journal
e-ISSN: 2619 – 9041

Research Article
Volume 5 - Issue 3: 411-416 / September 2022

TIME-DEPENDENT RECEIVER OPERATING CHARACTERISTIC ANALYSIS AND APPLICATIONS IN THE FIELD OF MEDICINE

Ceren EFE SAYIN[1]*, İlker ÜNAL[2]

[1]*Kahramanmaraş Sütçü imam University, Faculty of Medicine, Department of Biostatistics and Medical Informatics, 46040, Kahramanmaraş, Türkiye*
[2]*Çukurova University, Faculty of Medicine, Department of Biostatistics and Medical Informatics, 01250, Adana, Türkiye*

Abstract: When there is a time-dependency between the biomarker and the event of interest (death, disease, relapse etc.), classical receiver operating characteristic (ROC) analysis may not be able to estimate the true performance of the biomarker. For such cases, time-dependent ROC, an extended version of the standard ROC, is developed. In this study, the aim is to demonstrate applications of this modified ROC method on medical datasets and find out if it should be preferred over classical ROC for time-dependent situations. Comparison between classical ROC and Kaplan-Meier (KM) estimator, which is one of the two time-dependent ROC estimators, has been made using datasets in this study. Nearest Neighbor Estimator (NNE), the alternative of KM estimator, is also applied on all datasets. Then the findings of these two approaches are compared. It is concluded that time-dependent ROC method is superior to the standard ROC analysis. In general, the closer to the event time, the higher performance is observed. Especially, biomarkers measured at last 12 or 6 months before the event are determined to be better at classification than the earlier measurements. Though in all applications KM and NNE yielded very similar results, the latter is considered to be more appropriate to evaluate the performance of a biomarker when a time dependent data is also censored. Results of this study show that time-dependent ROC analysis performs more accurately when there is a time dependency between the biomarker and the event of interest; therefore, it is recommended.

Keywords: Time-dependent ROC, Diagnostic tests, Biomarker, Kaplan-Meier, Nearest neighbor estimator

***Corresponding author:** Kahramanmaraş Sütçü imam University, Faculty of Medicine, Department of Biostatistics and Medical Informatics, 46040, Kahramanmaraş, Türkiye
E mail: cerenefe.stat@gmail.com (C. EFE SAYIN)
Ceren EFE SAYIN https://orcid.org/0000-0001-9506-9219
İlker ÜNAL https://orcid.org/0000-0002-9485-3295

Received: April 22, 2022
Accepted: May 10, 2022
Published: September 01, 2022
Cite as: Efe Sayın C, Ünal İ. 2022. Time-dependent ROC analysis and applications in the field of medicine. BSJ Health Sci, 5(3): 411-416.

1. Introduction

Purpose of a diagnostic test is to detect and sometimes predict a certain disease. For example, biomarkers such as PSA or CA-125 concentration in blood serum provides diagnosis for cancer before its clinical onset. Time between the diagnosis and the clinical onset of the disease is called "lead time". The earlier the diagnosis, the better are the chances for the patient (Pepe, 1997).

Receiver operating characteristic (ROC) analysis is a graphical approach for comparison of two empirical distributions (Ünal, 2018). It is also commonly used for evaluating the accuracy of a continuous diagnostic test or a biomarker (Ünal, 2010). Standard ROC analysis considers only the present status of the individual (healthy or diseased). However, the simple two-class status might not always be the case. In some prospective cohort studies, status of the individual may shift from one class to another during follow-up. Or disease may not occur at the same time as the biomarker measurement (Zheng et al., 2006). Such characteristics of dataset indicate a time-dependency between the biomarker and the event of interest. At this point, a potential problem arises since the classical ROC approach might be insufficient for time-dependent cases.

Addressing the need for an analysis method for aforementioned conditions, Etzioni et al. (1999) proposed time-dependent ROC analysis. Based on their paper, other researchers continued working on this area. Especially in recent years, popularity of time-dependent ROC has increased. In order not to leave out this time-dependent nature of the disease onset, "time-dependent" sensitivity, specificity and ROC curve concepts are put forward and have taken place in several studies (Zheng and Heagerty, 2004; Antolini et al., 2005; Heagerty and Zheng, 2005; Zheng and Heagerty, 2007).

Time-dependent ROC analysis is suitable for both longitudinal and survival data. In this paper, the main focus is to demonstrate applications of this modified ROC method on survival data while providing a brief information about its theory, and find out if it should be preferred over classical ROC for time-dependent situations.

2. Materials and Methods

Event of interest in survival data is usually death or recurrence of a disease. Leisenring et al. (1997) and Balasubramanian and Lagakos (2001) defined time-dependent sensitivity using the "incident cases and long-term controls" approach for selection of cases and controls in survival data. Test result is measured at a

baseline ($t = 0$) and test's capacity of predicting the event happening on a time t in the future is investigated. In other words, the question is how well the test predicts people that *will* have the disease in a certain $[0, t]$ follow-up interval.

There are two time-dependent ROC curve estimators in the literature: Kaplan-Meier (KM) estimator and Nearest-Neighbor estimator (NNE). A brief information on the theory of these two methods is provided in the following section.

2.1. Kaplan-Meier Estimator (KM)

Sensitivity and specificity are expressed using Bayes' theorem as Equation 1:

$$P\{X > c|D(t) = 1\} = \frac{\{1 - S(t|X > c)\}P\{X > c\}}{\{1 - S(t)\}}$$
$$P\{X \leq c \mid D(t) = 0\} = \frac{S(t|X \leq c)P\{X \leq c\}}{S(t)} \quad (1)$$

where $S(t) = P(T > t)$ is the survival function and $S(t \mid X > c)$ is the conditional form of $S(t)$.

KM estimator is a commonly used nonparametric estimator of $S(t)$, proposed by Kaplan and Meier (1958) and expressed as given in Equation 2:

$$\hat{S}_{KM}(t) = \prod_{s \in T_n, s \leq t} \left\{1 - \frac{\sum_j 1(Z_j = s)\delta_i}{\sum_j 1(Z_j \geq s)}\right\} \quad (2)$$

where T_n be defined as unique values of Z_i when $\delta_i = 1$ (δ_i being the censoring indicator).

In order to estimate the survival function, KM estimator uses all the information in the data, including censored observations. A simple estimator for sensitivity and specificity at time t may be calculated as below (Equations 3 and 4), by combining empirical distribution function of biomarker X and KM estimator (where $\hat{F}_X(c) = \sum 1(X_i \leq c)/n$):

$$\hat{P}_{KM}\{X > c|D(t) = 1\} = \frac{\{1 - \hat{S}_{KM}(t|X > c)\}\{1 - \hat{F}_X(c)\}}{\{1 - \hat{S}_{KM}(t)\}} \quad (3)$$

$$\hat{P}_{KM}\{X \leq c|D(t) = 0\} = \frac{\hat{S}_{KM}(t|X \leq c)\hat{F}_X(c)}{\hat{S}_{KM}(t)} \quad (4)$$

One complication of this estimator is that it does not assure the monotonicity of the TPR and FPR; and therefore the ROC curve. More specifically, although in definition $P\{X > c|D(t) = 1\} \geq P\{X > c'|D(t) = 1\}$ when $c' > c$; the monotonicity might be distorted by the nature of Bayes theorem and KM method. In addition, the conditional KM estimator $\hat{S}_{KM}(t|X > c)$ assumes that the censoring process is independent of biomarker, which is another complication of the ROC estimation based on KM method (Heagerty el al., 2000).

2.2. Nearest Neighbor Estimator (NNE)

An ROC curve estimator is obtained using the binary distribution function $F(c, t) = P(X \leq c, T \leq t)$ of Akritas (1994), or its equivalent $S(c, t) = P(X > c, T > t)$ as cited in Heagerty et al (2000). This estimator depends on the expression $S(c, t) = \int_c^{\infty} S(t|X = s)dF_X(s)$ and calculated as given in Equation 5:

$$\hat{S}_{\lambda_n}(c, t) = \frac{1}{n}\sum_i \hat{S}_{\lambda_n}(t|X = X_i)1(X_i > c) \quad (5)$$

where $F_X(s)$ is the distibution function of X and $\hat{S}_{\lambda_n}(t|X = X_i)$ is an estimator for the conditional survival function specified by a smoothing parameter λ_n.

When X is not a categorical variable and there is not sufficient number of observations contained in each X_i, smoothing becomes necessary for the estimation of $\hat{S}_{\lambda_n}(t|X = X_i)$. It is performed by the means of a kernel function $K_{\lambda_n}(X_j, X_i)$ depending on λ_n and after weighting of the KM estimator, $\hat{S}_{\lambda_n}(t|X = X_i)$ is attained as given in Equation 6:

$$\hat{S}_{\lambda_n}(t|X = X_i) = \prod_{s \in T_n, s \leq t} \left\{1 - \frac{\sum_j K_{\lambda_n}(X_j, X_i)1(Z_j = s)\delta_i}{\sum_j K_{\lambda_n}(X_j, X_i)1(Z_j \geq s)}\right\} \quad (6)$$

Akritas (1994) prefers $K_{\lambda_n}(X_j, X_i) = 1\{-\lambda_n < \hat{F}_X(X_i) - \hat{F}_X(X_j) < \lambda_n\}$ as a kernel function, the nearest neighbor to point (0,1). The percentage of observations contained in each neighborhood is represented by $2\lambda_n \epsilon (0,1)$. Even if the selection of other kernels is not unorthodox, using the nearest neighbor would prevent the ROC estimations from the effect of monotone transformations on the biomarker.

Sensitivity and specificity calculated (Equation 7) with the NNE approach is as follows (where $\hat{S}_{\lambda_n}(t) = \hat{S}_{\lambda_n}(-\infty, t)$):

$$\hat{P}_{\lambda_n}\{X > c \mid D(t) = 1\} = \frac{1 - \hat{F}_X(c) - \hat{S}_{\lambda_n}(c, t)}{1 - \hat{S}_{\lambda_n}(t)}$$
$$\hat{P}_{\lambda_n}\{X \leq c \mid D(t) = 0\} = 1 - \frac{\hat{S}_{\lambda_n}(c, t)}{\hat{S}_{\lambda_n}(t)} \quad (7)$$

As opposed to KM estimator, the equations above fulfill the condition that the ROC curve is monotone increasing. Additionally, that NNE allows a censoring process dependent on the biomarker X is another advantage of the method.

2.3. Data and Software

Two methods (KM, NNE) considering the time effect when evaluating the performance of diagnostic tests are examined in this study. Regarding results are compared to the ones obtained using classical ROC analysis which is independent of time. Additionally, a more specific comparison between KM and NNE methods is performed.

Datasets used in the applications are as follows and all contain a biomarker, a survival time and a censoring variable:

- Open datasets named "Melano" and "Paquid" from the "timeROC" package found in R program (Blanche et al., 2013).
- A part of Hodgkin Lymphoma data from a multicenter study conducted by Paydas et al. (2021) in Oncology Department in Cukurova University, Adana, Türkiye.

2.3.1. Melano dataset

This dataset contains information obtained from 205 malignant melanoma patients who had radical surgery in 1962-1977 in University of Odense, Denmark. Considering survival time after the operation, death is the event of interest and tumor thickness (1/100 mm) is the biomarker. All patients were followed until the end of 1977, and 134 survived while 71 died, 14 of which is not cancer related (which are excluded from the study). Applications are carried on the remaining 191 patients.

2.3.2. Paquid dataset

This dataset consists of the records belonging to 2561 patients participated in a prospective cohort study in southwestern France in 1988. There are two different time covariates as "time until the onset of Alzheimer's" and "time until death before the onset of Alzheimer's". Since the event of interest in the applications is the onset of Alzheimer's disease, those who died without it are removed from the data. Analyses are performed on the remaining 1927 participants, out of which 449 had the disease at the end. Time after registration until the onset of the disease is considered as the time covariate. Two test scores are credited as biomarkers: DSST (Digit Symbol Substitution Score Test) and MMSE (Mini Mental State Examination). Due to the fact that lower values of DSST and MMSE indicate disease existence, reciprocal values of the scores are used so as to obtain ROC curves in the upper diagonal.

2.3.3. Hodgkin lymphoma (HL) dataset

Out of 364 HL patients that this data contains, 88 had died from HL while 276 survived. A score with seven factors (age, sex, stage, hemoglobin, albumin, lymphocyte count and white cell count) developed for diagnosis of HL is called IPS7 (IPS: International Prognostic Score). A reduced version of IPS with 3 factors (IPS3) alternative to IPS7 was later proposed (Diefenbach et al., 2015). Both IPS3 and IPS7 are evaluated as biomarker in the applications of this paper. Event of interest is death, and the survival time is the time covariate.

The applications in this study are carried out using R 3.2.2 (Vienna, Austria, 2015) and SPSS v20 (Armonk, New York U.S.A, 2011). R libraries are: "survivalROC" (Heagerty and Saha-Chaudhuri, 2013) for plotting the time-dependent ROC curves and comparison of KM and NNE approaches; "timeROC" (Blanche et al., 2013) for the standard error calculations regarding the ROC curves at specific time points.

3. Results

3.1. Time-Dependent ROC vs Classical ROC

3.1.1. Melano dataset

Data consist of 191 malignant melanoma patients. Tumor thickness along with the post-operation survival statuses and times of the patients are recorded in 15 years. Diagnostic performance of tumor thickness is of interest. Since a period of 15 years is too long for diagnostic purposes, data of the last 6 years to event (death) are used in time-dependent ROC analysis applications of this study (Table 1).

Table 1. Area under the curve (AUC), standard errors (SE) and 95% CIs obtained from Melano dataset

Time to Event	AUC	SE	95% CI
6 years	0.73	0.05	0.63 - 0.81
Classical ROC	0.75	0.04	0.68 - 0.82
5 years	0.77	0.04	0.67 - 0.84
4 years	0.80	0.04	0.73 - 0.88
3 years	0.81	0.04	0.73 - 0.89
2 years	0.88	0.03	0.82 - 0.95
1 year	0.91	0.03	0.85 - 0.97

Findings in Table 1 show that prediction of mortality annually improved as getting closer to the event. Especially in the last two years' results are in favor of that time-dependent ROC analysis foresees death 17% (2 years to the event) and 21% (1 year to the event) more accurately than the classical ROC approach.

3.1.2. Paquid dataset

There are two test scores used as the biomarkers for Alzheimer's diagnosis in this dataset: DSST (Digit Symbol Substitution Score Test) and MMSE (Mini Mental State Examination). The event of interest is the onset of Alzheimer's and follow-up time is 12 years. Even 10 years before the event, time dependent AUC estimations still perform more than 10% better, when compared to the AUC estimation of the classical ROC for both DSST and MMSE (Table 2). Their ability to predict the event has increased at each time point, reaching the highest level in the end. Although they seem to behave similarly along the way, DSST has higher AUC estimations at all measurement times; and therefore it is found more successful than MMSE at classification (Table 2).

3.1.3. Hodgkin lymphoma (HL) dataset

Characteristics and measurements of 364 HL patients were recorded in this dataset, with an almost 19-year follow-up which is a very long time for a biomarker to predict the disease. Thus, only the last 6 years have been taken into account in this study. Diagnostic performance of IPS7 and IPS3 in estimating mortality is examined at each time point. Area under the curve (AUC) of the classical ROC and time-dependent AUCs are presented in Table 3.

Table 2. Area under the curve (AUC), standard errors (SE) and 95% CIs obtained from two biomarkers of Paquid dataset

	DSST			MMSE		
Time to Event	AUC	SE	95% CI	AUC	SE	95% CI
Classical ROC	0.69	0.01	0.66 - 0.71	0.63	0.02	0.60 - 0.66
12 years	0.74	0.02	0.71 - 0.77	0.66	0.02	0.63 - 0.69
11 years	0.77	0.02	0.72 - 0.78	0.69	0.02	0.65 - 0.72
10 years	0.79	0.02	0.74 - 0.80	0.71	0.02	0.67 - 0.74
9 years	0.79	0.02	0.75 - 0.81	0.71	0.02	0.66 - 0.73
8 years	0.79	0.02	0.74 - 0.81	0.71	0.02	0.67 - 0.74
7 years	0.80	0.02	0.75 - 0.82	0.73	0.02	0.68 - 0.76
6 years	0.81	0.02	0.76 - 0.84	0.75	0.02	0.70 - 0.79
5 years	0.81	0.02	0.77 - 0.84	0.74	0.02	0.69 - 0.79
4 years	0.83	0.02	0.77 - 0.86	0.74	0.03	0.69 - 0.80
3 years	0.83	0.02	0.77 - 0.87	0.77	0.03	0.70 - 0.82
2 years	0.88	0.03	0.81 - 0.93	0.82	0.04	0.75 - 0.89
1 year	0.91	0.04	0.83 - 0.98	0.82	0.07	0.68 - 0.97

Table 3. Area under the curve (AUC), standard errors (SE) and 95% CIs obtained from two biomarkers of HL dataset

	IPS7			IPS3		
Time to Event	AUC	SE	95% CI	AUC	SE	95% CI
Classical ROC	0.62	0.04	0.54 - 0.69	0.65	0.04	0.58 - 0.72
6 years	0.62	0.04	0.54 - 0.70	0.64	0.04	0.57 - 0.71
5 years	0.61	0.04	0.55 - 0.70	0.65	0.04	0.58 - 0.72
4 years	0.62	0.04	0.56 - 0.72	0.67	0.04	0.60 - 0.74
3 years	0.64	0.04	0.56 - 0.73	0.69	0.04	0.62 - 0.77
2 years	0.64	0.05	0.56 - 0.74	0.70	0.04	0.63 - 0.78
1.5 years	0.67	0.05	0.58 - 0.77	0.74	0.04	0.66 - 0.82
1 year	0.66	0.05	0.56 - 0.77	0.72	0.05	0.63 - 0.81
6 months	0.75	0.06	0.63 - 0.87	0.81	0.05	0.71 - 0.91

For both IPS7 and IPS3, AUCs of earlier time points are found to be similar or close to the classical ROC value, while a general increase can be mentioned when getting closer to event of death. Maximum AUC value is 0.75 for IPS7 whereas IPS3 reaches 0.81 level, 6 months before the event. Only the last 6 months' measurements have been significantly different from the rest of the time points as well as the classical ROC for both biomarkers. In other words, IPS7 and IPS3 at 6 months to the event predict mortality, respectively 21% and 24% better than classical ROC.

3.2. Kaplan-Meier Estimator (KM) vs Nearest Neighbor Estimator (NNE)

After establishing time-dependent ROC analysis is a better way when there is a time dependency between the biomarker and event of interest, two time-dependent AUC estimators KM and NNE are compared to find out if one is superior to the other.

In Table 4, KM and NNE AUCs again fall into one another's 95% confidence intervals, though KM values are obtained higher than NNE values.

Table 5 demonstrates that KM yielded higher values than NNE for both DSST and MMSE biomarkers. An additional interpretation of this table is that DSST is a better biomarker than MMSE when the two is compared in each method.

Table 6 presents that very similar results are obtained from KM and NNE methods for IPS7 and IPS3 at all-time points. In addition, when the two biomarkers are compared using KM and NNE approaches, it can be concluded that one score is not better than the other; in fact, they are statistically the same.

Table 4. Comparison of KM and NNE methods using Melano dataset

Time to Event	KM-AUC (95% CI)	NNE-AUC (95% CI)
6 years	0.73 (0.63 - 0.81)	0.69 (0.62 - 0.79)
5 years	0.77 (0.67 - 0.84)	0.72 (0.65 - 0.82)
4 years	0.80 (0.73 - 0.88)	0.77 (0.71 - 0.87)
3 years	0.81 (0.73 - 0.89)	0.77 (0.68 - 0.87)
2 years	0.88 (0.82 - 0.95)	0.86 (0.78 - 0.94)
1 year	0.90 (0.85 - 0.97)	0.90 (0.83 - 0.98)

KM= Kaplan-Meier estimator, NNE= Nearest Neighbor estimator, AUC= area under the curve

Table 5. Comparison of KM and NNE methods using two biomarkers of Paquid dataset

	DSST		MMSE	
Time to Event	KM-AUC (95% CI)	NNE-AUC (95% CI)	KM-AUC (95% CI)	NNE-AUC (95% CI)
12 years	0.74 (0.71 - 0.77)	0.71 (0.71 - 0.74)	0.66 (0.63 - 0.69)	0.64 (0.61 - 0.68)
11 years	0.77 (0.72 - 0.78)	0.73 (0.70 - 0.76)	0.69 (0.65 - 0.72)	0.67 (0.64 - 0.71)
10 years	0.79 (0.74 - 0.80)	0.75 (0.72 - 0.78)	0.71 (0.67 - 0.74)	0.69 (0.65 - 0.72)
9 years	0.79 (0.75 - 0.81)	0.75 (0.72 - 0.79)	0.71 (0.66 - 0.73)	0.69 (0.65 - 0.72)
8 years	0.79 (0.74 - 0.81)	0.76 (0.72 - 0.79)	0.71 (0.67 - 0.74)	0.69 (0.65 - 0.73)
7 years	0.80 (0.75 - 0.82)	0.77 (0.73 - 0.80)	0.73 (0.68 - 0.76)	0.71 (0.67 - 0.75)
6 years	0.81 (0.76 - 0.84)	0.78 (0.74 - 0.81)	0.75 (0.70 - 0.79)	0.73 (0.69 - 0.78)
5 years	0.81 (0.77 - 0.84)	0.78 (0.75 - 0.82)	0.74 (0.69 - 0.79)	0.73 (0.68 - 0.78)
4 years	0.83 (0.77 - 0.86)	0.79 (0.74 - 0.84)	0.74 (0.69 - 0.80)	0.73 (0.68 - 0.79)
3 years	0.83 (0.77 - 0.87)	0.80 (0.74 - 0.85)	0.77 (0.70 - 0.82)	0.75 (0.69 - 0.82)
2 years	0.88 (0.81 - 0.93)	0.86 (0.79 - 0.92)	0.82 (0.75 - 0.89)	0.81 (0.73 - 0.88)
1 year	0.91 (0.83 - 0.98)	0.91 (0.87 - 0.98)	0.82 (0.68 - 0.97)	0.81 (0.66 - 0.98)

KM= Kaplan-Meier estimator, NNE= Nearest Neighbor estimator, AUC= area under the curve

Table 6. Comparison of KM and NNE methods using two biomarkers of HL dataset

	IPS7		IPS3	
Time to Event	KM-AUC (95% CI)	NNE-AUC (95% CI)	KM-AUC (95% CI)	NNE-AUC (95% CI)
6 years	0.61 (0.54 - 0.70)	0.61 (0.53 - 0.68)	0.64 (0.57 - 0.71)	0.64 (0.55 - 0.71)
5 years	0.61 (0.55 - 0.70)	0.60 (0.52 - 0.68)	0.65 (0.58 - 0.72)	0.65 (0.56 - 0.72)
4 years	0.62 (0.56 - 0.72)	0.62 (0.54 - 0.70)	0.67 (0.60 - 0.74)	0.67 (0.57 - 0.74)
3 years	0.64 (0.56 - 0.73)	0.64 (0.55 - 0.72)	0.69 (0.62 - 0.77)	0.69 (0.59 - 0.77)
2 years	0.64 (0.56 - 0.74)	0.64 (0.54 - 0.73)	0.70 (0.63 - 0.78)	0.70 (0.60 - 0.77)
1.5 years	0.67 (0.58 - 0.77)	0.66 (0.56 - 0.75)	0.74 (0.66 - 0.82)	0.74 (0.64 - 0.82)
1 year	0.66 (0.56 - 0.77)	0.66 (0.55 - 0.76)	0.72 (0.63 - 0.81)	0.72 (0.61 - 0.81)
6 months	0.75 (0.63 - 0.87)	0.75 (0.62 - 0.86)	0.81 (0.71 - 0.91)	0.81 (0.68 - 0.92)

KM= Kaplan-Meier estimator, NNE= Nearest Neighbor estimator, AUC= area under the curve

4. Discussion

In this study, the most common time-dependent ROC methods (KM, NNE) are examined in theory and application. Their performances are compared both to each other and to classical ROC analysis. When biomarkers measured at different time-points are compared, (except for some situations) AUC estimations are found to show monotone increase similar to the literature. In the real-data application detailed in Martinez-Camblor et al. (2016), AUC estimations corresponding to biomarker values obtained in the beginning of the follow-up descended for a while, and then started to escalate as the time of failure became closer. This fluctuation might be due to the loss of diagnostic ability of a biomarker in long follow-up periods.

Though not sufficient number of studies comparing classical and time-dependent ROC exist in the literature, paper of Chambless and Diao (2006) is one of the few in this regard. Writers explained that AUC values obtained from these two ROC methods cannot be the same. Results of this study have turned out to be supportive to their claim. When KM and NNE methods are compared to each other at time t, in all datasets they yielded similar results

parallel to Blanche et al. (2013). Although KM method results in higher estimations indicating higher performance, it has a limitation that does not allow the censoring process to be dependent on the biomarker. On the other hand, NNE method does not possess such restrictions. Even if its AUC estimations are lower, the smoother curves are obtained with NNE; and thus the well-known feature of the ROC curve being a monotone increasing function is satisfied. Hence, when deciding which to prefer, choosing NNE option would be more advantageous.

5. Conclusion

The insufficiency of standard ROC analysis, when time is a parameter that cannot be set aside, is proven with examples in this study. Time-dependent ROC analysis for time-dependent situations is recommended as a substitution. KM and NNE as time-dependent estimators gave mostly similar results; yet, one must consider the limitations of KM while choosing between the two.

Author Contributions

Concept: C.E.S. (50%) and İ.Ü. (50%), Design: C.E.S. (50%) and İ.Ü. (50%), Supervision: C.E.S. (50%) and İ.Ü. (50%), Data collection and/or processing: C.E.S. (50%) and İ.Ü. (50%), Data analysis and/or interpretation: C.E.S. (80%) and İ.Ü. (20%), Literature search: C.E.S. (50%) and İ.Ü. (50%), Writing: C.E.S. (50%) and İ.Ü. (50%), Critical review: C.E.S. (50%) and İ.Ü. (50%), Submission and revision C.E.S. (50%) and İ.Ü. (50%). All authors reviewed and approved final version of the manuscript.

Conflict of Interest

The authors declared that there is no conflict of interest.

Ethical Approval/Informed Consent

Ethics committee approval is not required for this study because used data are open for public usage.

Acknowledgements

This manuscript is a short summary of MSc thesis. A part of this study was published at 18th national and 1st international biostatistics congress.

References

Akritas M. 1994. Nearest neighbor estimation of a bivariate distribution under random censoring. Ann Stat, 22: 1299-1327.

Antolini L, Boracchi P, Biganzoli E. 2005. A time-dependent discrimination index for survival data. Stat Medic, 24: 3927-3944.

Balasubramanian R, Lagakos S. 2001. Estimation of the timing of perinatal transmission of HIV. Biometrics, 57: 1048-1058.

Blanche P, Dartigues JF, Jacqmin-Gadda H. 2013. Review and comparison of ROC curve estimators for a time-dependent outcome with marker-dependent censoring. Biometrical J, 5: 687-704.

Blanche P, Dartigues JF, Jacqmin-Gadda, H. 2013. Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks. Stat Medic, 32(30): 5381-5397.

Chambless L, Diao G. 2006. Estimation of time-dependent area under the ROC curve for long-term risk prediction. Stat Medic, 25: 3474-3486.

Diefenbach CS, Li H, Hong F, Gordon LI, Fisher RI, Bartlett NL. 2015. Evaluation of the international prognostic score (IPS-7) and a simpler prognostic score (IPS-3) for advanced Hodgkin lymphoma in the modern era. Br J Haematol, 171(4): 530-538.

Etzioni R, Pepe MS, Longton G, Chengcheng H, Goodman G. 1999. Incorporating the time dimension in receiver operating characteristic curves: a case study of prostate cancer. Medic Decision Making, 19: 242-251.

Heagerty PJ, Saha-Chaudhuri P. 2013. survivalROC: Time-dependent ROC curve estimation from censored survival data. R package version 1.0.3. URL: http://CRAN.R-project.org/package=survivalROC. (February 10, 2022).

Heagerty PJ, Lumley T, Pepe MS. 2000. Time-dependent ROC curves for censored survival data and a diagnostic marker. Biometrics, 56: 337-344.

Heagerty PJ, Zheng Y. 2005. Survival model predictive accuracy and ROC curves. Biometrics, 61: 92-105.

IBM Corp. 2011. IBM SPSS Statistics for Windows, Version 20.0. Armonk, NY: IBM Corp.

Kaplan E, Meier P. 1958. Nonparametric estimation from incomplete observations. J Amer Stat Assoc, 53: 457-481.

Leisenring W, Pepe MS, Longton G. 1997. A marginal regression modelling framework for evaluating medical diagnostic tests. Stat Medic, 16: 1263-1281.

Martinez-Camblor P, Bayon GF, Perez-Fernandez S. 2016. Cumulative/dynamic ROC curve estimation. J Stat Comput Simul, 86(17): 3582-3594.

Paydaş S, Lacin S, Doğan M, Barışta İ, Yıldız B, Seydaoğlu G. 2021. IPS-3 validation in 1012 cases with classical hodgkin lymphoma. Leukemia Res, 102: 106519.

Pepe MS. 1997. A regression modeling framework for ROC curves in medical diagnostic testing. Biometrika, 84: 595-608.

R Development Core Team. 2015. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL: http://www.R-project.org. (February 10, 2022).

Ünal İ. 2010. Tanı Standards and analytic techniques used in evaluation of diagnostic tests. PhD thesis, Çukurova University, Health Science Institute, Adana, Türkiye, pp: 161.

Ünal İ. 2018. Is the transformation useful to estimate the area under the ROC curve with skewed data? Cukurova Med J, 43: 141-147.

Zheng Y, Cai T, Feng Z. 2006. Application of the time-dependent ROC curves for prognostic accuracy with multiple biomarkers. Biometrics, 62: 279-287.

Zheng Y, Heagerty PJ. 2004. Semiparametric estimation of time-dependent ROC curves for longitudinal marker data. Biostatistics, 5(4): 615-632.

Zheng Y, Heagerty PJ. 2007. Prospective accuracy for longitudinal markers. Biometrics, 63: 332-341.