J.Natn.Sci.Foundation Sri Lanka 2018 46 (1): 11 - 21
DOI: http://dx.doi.org/10.4038/jnsfsr.v46i1.8261

RESEARCH ARTICLE

Adaptation of the weighted Kaplan-Meier method to time-dependent ROC curves

Deniz Sigirli[1*], Ilker Ercan[1], Ozkan Balcin[2] & Ekrem Kaya[2]
[1] *Department of Biostatistics, Faculty of Medicine, Uludag University, Bursa, Turkey.*
[2] *Department of General Surgery, Faculty of Medicine, Uludag University, Bursa, Turkey.*

Revised: 10 March 2017; Accepted: 17 August 2017

Abstract: This study was aimed at adapting the weighted Kaplan-Meier method to time-dependent ROC curve analysis. The performances of these two time-dependent ROC curve methods were compared, in which the Kaplan-Meier estimator and weighted Kaplan-Meier estimator were used. An application was presented for pancreatic cancer patients to evaluate the prognostic ability of the CA19-9 antigen. A simulation study was performed for different scenarios to see the performance of the proposed method. In all situations, it is observed that the AUC values that were obtained by the weighted time-dependent ROC (WTDR) curves more closely approximated the real AUC values than the classical time-dependent ROC (TDR) curve method and has got smaller mean square error rates.

Keywords: Censored data, time-dependent ROC curves, weighted Kaplan-Meier.

INTRODUCTION

In classical ROC analysis, disease status is treated as a fixed, stable feature of the experimental unit. However, disease status can change over time, and those individuals who are not diseased can develop the disease during the study period. There can be a certain time lag between the time that the diagnostic test is conducted and the onset of the disease. How well a diagnostic test result, having been measured at the beginning of the study, can discriminate between diseased and undiseased individuals at a [0, t] follow-up time, is the question that must be addressed in such situations.

In literature there are several discussions of time-dependent ROC curves (Etzioni *et al*., 1999; Heagerty *et al*., 2000; Slate & Turnbull, 2000; Cai *et al*., 2003; 2006; Heagerty & Zheng, 2005; Chambless & Diao, 2006; Uno *et al*., 2007; Hung & Chiang, 2010a; Martínez-Camblor *et al*., 2016). Cumulative/dynamic, incident/static and incident/dynamic estimators for time-dependent sensitivity/specificity and related ROC curves were defined by Heagerty and Zheng (2005) and also discussed by Cai *et al*. (2006) and Pepe *et al*. (2008). Sensitivity can be estimated using cases that are stratified according to the time at which the event occurs (incident sensitivity) or using all cases identified up to time t (cumulative sensitivity). Specificity can be estimated using all individuals who are not cases at time t (dynamic specificity) or using only those individuals who are event free through a fixed follow-up period (static specificity). The incident/static definition of sensitivity/specificity was considered by several authors (Etzioni *et al*., 1999; Slate & Turnbull, 2000; Heagerty & Zheng, 2005; Cai *et al*., 2006), and two methods were proposed by Etzioni *et al*. (1999). The first one calculates the ROC curve given estimates of the longitudinal model parameters by utilising random-effects models to capture the correlation between within-subject measurements. The second one is based on estimating the ROC curve at any time of interest, by setting the time covariate to a specific value. Slate and Turnbull (2000) gave definitions of sensitivity, specificity and related ROC curve for

* Corresponding author (sigirli@uludag.edu.tr; https://orcid.org/0000-0002-4006-3263)

the longitudinal biomarker data by focusing on fully Bayesian hierarchical models and the latent disease process models. Cai *et al*. (2006) used generalised linear models to estimate time-dependent ROC (TDR) curve with incident sensitivity in censored data. They modelled the dependence in time by using vectors of polynomial or spline basis functions.

The present study was focused on the cumulative/dynamic definition where each individual plays the role of control for times $t \leq X$, and then contributes as a case for times $t > X$, where X is the failure time. Time-dependent ROC curves related to cumulative sensitivity and dynamic specificity has recently been used by a number of authors (Heagerty *et al*., 2000; Chambless & Diao, 2006; Uno *et al*., 2007; Song & Zhou, 2008; Hung & Chiang, 2010a; Wolf *et al*., 2011; Blanche *et al*., 2013a; Li *et al*., 2015; Li, 2016; Martínez-Camblor *et al*., 2016; Rodríguez-Álvarez *et al*., 2016). Heagerty *et al*. (2000) proposed estimators based on the cumulative distribution function of the biomarker and Kaplan-Meier estimator of the survival function. Since this method does not satisfy the condition of monotonicity for the ROC curve, they also proposed an estimator based on the nearest neighbour estimator of Akritas (Akritas *et al*., 1994), which estimates the bivariate distribution of the marker and the failure time using kernel smoothing techniques. Two methods were proposed by Chambless and Diao (2006). The first one is a recursive calculation over the ordered times of events, analogous to the Kaplan-Meier approach to survival function estimation. This method does not guarantee the monotonicity. The second approach uses a regression survival model to estimate the conditional survival probability of the outcome at time t given the marker Y.

Song and Zhou (2008) defined the covariate specific time-dependent ROC curve using both the cumulative and incident sensitivity. Uno *et al*. (2007) and Hung and Chiang (2010a) have proposed estimators based on inverse probability of censoring weighting (IPCW). Blanche *et al*. (2013a) proposed a conditional IPWC method, which is the modified version of IPCW, to obtain a nonparametric estimator robust to marker-dependent censoring. They gave a detailed review of the time-dependent ROC curve estimators proposed in literature and compared the properties. Wolf *et al*. (2011) introduced a method for calculating sensitivity and specificity for censored data based on the Nelson-Aalen estimator and they used isotonic regression to achieve monotonicity for the ROC curve. Li *et al*. (2015) proposed a weighting method to estimate cumulative/dynamic time-dependent sensitivity/specificity and related ROC curve nonparametrically by using uniform kernel, which has connections to the methods in Heagerty *et al*. (2000) and Chambless and Diao (2006). Martínez-Camblor *et al.* (2016) proposed a methodology, which assigns a probability of belonging to a group by using proportional hazard Cox regression model and Kaplan–Meier estimator for cumulative/dynamic ROC curve estimation. Rodríguez-Álvarez *et al*. (2016) proposed nonparametric regression estimators Nadaraya-Watson kernel weights for cumulative/dynamic ROC curves in the presence of covariate-dependent censoring. Also, Li (2016) studied estimation of cumulative/dynamic time-dependent ROC and area under the ROC curve (AUC) for left-truncated and right-censored data.

The area under the ROC curve (AUC) was defined as the probability that the marker value of a randomly selected case exceeds the marker value of a randomly selected control (Hanley & McNeil, 1982; Pepe, 2003). The extension of AUC to the time-dependent setting has been discussed by many authors (Heagerty & Zheng, 2005; Chambless & Diao, 2006; Chiang & Hung, 2010; Hung & Chiang, 2010b; Cai *et al*., 2011; Viallon & Latouche, 2011; Schmid *et al*., 2015; Lambert & Chevret, 2016). The construction in their works was to estimate AUC (t) for each t time. Blanche *et al*. (2013b) presented different estimators of the time-dependent AUC for univariate survival data, longitudinal setting and competing events setting. The time-dependent ROC curve can be calculated by plotting the sensitivity *versus* 1-specificity for a range of cut-off values and the AUC can be calculated using standard numerical integration methods such as trapezoid rule (Campbell, 1994; Chambless & Diao, 2006).

The term borrowing strength is typically used in Bayesian statistics and generally references an attempt to improve precision by using additional information from allied sources.

The weighted likelihood (WL) function has been designed to incorporate information from populations that are relevant, but not of prime inferential interest to the study population (Hu, 1994). The WL function suggested by Hu and Zidek (2002) is based on the result of James and Stein (1961) insofar as, in terms of the sum of the mean-square-error-of-estimation criterion, the sample averages could be improved upon by borrowing information from the other samples – the so-called Stein paradox (James & Stein, 1961; Hu & Zidek, 2002). Similar to the James-Stein estimator, a WL estimator that facilitates drawing inferences on one sample by using additional information from different populations was suggested by Hu and Zidek (2002).

Wang (2001), Wang *et al*. (2004), and Wang and Zidek (2005) used cross-validation procedure for adaptively choosing the weights and gave the analytical forms of the adaptive weights when the WL estimation is a linear combination of the maximum likelihood estimations. They estimated the WL whereby the data are regarded as samples from *m* populations, and proposed adaptive weights, which were allowed to depend on the data. If the distribution function of the first group is F_1, the weighted empirical distribution function to estimate it can be given as follows:

$$\hat{F}_\lambda(x) = \sum_{i=1}^{m} \lambda_i \hat{F}_i(x), \qquad ...(1)$$

with $\sum_{i=1}^{m} \lambda_i = 1$ and $\lambda_i \geq 0$, where $\hat{F}_i(x)$ is the empirical distribution function related to the i^{th} population and $X_{i1}, \ldots, X_{in_i}$ is the sample of n_i individuals, drawn from the i^{th} population. Plante (2008a) showed that WL can be derived from the entropy maximisation principle using the weighted empirical distribution function given above, and suggested minimum averaged mean squared error (MAMSE) weights. Plante (2008b; 2009) used MAMSE weights for right-censored data and proposed adaptively weighted Kaplan-Meier estimates as non-parametric estimators for lifetime data, which borrows strength from *m* different populations to draw inferences for just one population of interest, that has a similar distribution to other *m-1* populations.

This article is aimed at using a weighted Kaplan-Meier estimator to obtain time-dependent ROC curves, which could handle right-censored data in determining the discrimination ability of a marker.

METHODOLOGY

Let X_{ij} be the death time for the j^{th} individual in the i^{th} population, V_{ij} be the censoring time for the j^{th} individual of the i^{th} population, and $Z_{ij} = \min(X_{ij}, V_{ij})$ be the follow-up time; thus, if $\delta_{ij} = 1(V_{ij} \geq X_{ij})$, then (Z_{ij}, δ_{ij}) is observed for $i = 1, \ldots, m; j = 1, \ldots, n_i$. The Kaplan-Meier estimate of the probability of survival beyond time *t*, which is a non-parametric estimator of the survival function *S*(*t*), can be written as below for the i^{th} population (Kaplan & Meier, 1958):

$$\widehat{KM}_i(t) = \prod_{0 \leq s \leq t} \left\{1 - \frac{dN_i(s)}{Y_i(s)}\right\}, \qquad ...(2)$$

where $N_i(s) = \sum_{j=1}^{n_i} 1(Z_{ij} \leq s, \delta_{ij} = 1)$, $dN_i(s) = N_i(s) - N_i(s^-)$, $Y_i(s) = \sum_{j=1}^{n_i} 1(Z_{ij} \geq s)$. An optimisation problem in which the optimal weights can be obtained, which minimises the objective function given below under the constraints $\sum_{i=1}^{m} \lambda_i = 1$ and $\lambda_i \geq 0$ was defined by Plante (2009):

$$\mathrm{P}(\boldsymbol{\lambda}) = \int_0^U \left[\left\{\hat{F}_1(t) - \sum_{i=1}^{m} \lambda_i \hat{F}_i(t)\right\}^2 + \sum_{i=1}^{m} \lambda_i{}^2 \widehat{var}\{\hat{F}_i(t)\}\right] d\hat{F}_1(t), \qquad ...(3)$$

where $\hat{F}_i(t)$ is $1 - \widehat{KM}_i(t)$ and the weights are chosen to minimise $\mathrm{P}(\boldsymbol{\lambda})$. In the objective function the squared difference was required to be minimised so that weights that make $\hat{F}_i(t)$ close to $\hat{F}_1(t)$ should be selected. U is the upper limit, which is set smaller than the largest follow-up time. An algorithm for obtaining these optimal weights, which can be noted as w_i, was also proposed (Plante, 2008b; 2009). So the weighted Kaplan-Meier estimate for the probability of survival beyond time *t* can be given as

$$\hat{S}_{WKM}(t) = 1 - \sum_{i=1}^{m} w_i \hat{F}_i(t). \qquad ...(4)$$

Various approaches have been proposed that can be used when the output variable of interest is an event that can take place at any time after the diagnostic test has been administered. Heagerty *et al.* (2000) proposed a ROC curve estimator based on the Kaplan-Meier function that can be used when the disease onset time is censored. They gave time-dependent sensitivity and specificity using Bayes' theorem (Heagerty *et al*., 2000). In these sensitivity and specificity definitions, conditional survival functions are estimated on different subsamples when *c* varies. Let Y_{ij} be a continuous diagnostic test result measured on the j^{th} individual of the i^{th} population $(i = 1, \ldots, m; j = 1, \ldots, n_i)$ and c is the cut-off value for the marker values $[c \in (-\infty, +\infty)]$. For large values of *c*, the sample size for $Y > c$ may be small for getting the conditional Kaplan-Meier estimate. In this paper, it is aimed at increasing the sample size by using Plante's method (2009). The weighted conditional Kaplan-Meier estimate will typically be smoother since steps can occur at the times of failure from all the populations. By using the weighted Kaplan-Meier estimator instead of the survival function $S(t)$ and by using the sample distribution function of *Y*, the sensitivity and specificity can be written again as follows, respectively in equation (5) and equation (6);

$$sensitivity(c,t) = P(Y_1 > c|D_1(t) = 1) = \frac{\{1 - \hat{S}_{WKM}(t|Y > c)\}\{1 - \hat{F}_{Y_1}(c)\}}{\{1 - \hat{S}_{WKM}(t)\}} \quad ...(5)$$

$$specificity(c,t) = P(Y_1 \leq c|D_1(t) = 0) = \frac{\hat{S}_{WKM}(t|Y \leq c)\,\hat{F}_{Y_1}(c)}{\hat{S}_{WKM}(t)} \quad ...(6)$$

where, $D_1(t)$ denotes failure status at any time t for the 1st population with $D_1(t) = 1$ indicating that the subject has had an event prior to time t; $\hat{S}_{WKM}(t)$ is the weighted Kaplan-Meier estimator for the 1st population, calculated by using data from m populations as $\hat{S}_{WKM}(t) = \sum_{i=1}^{m} w_i S_i(t)$; $\hat{S}_{WKM}(t|Y > c)$ is the conditional Kaplan-Meier estimator for the 1st population calculated by using data from $Y_i > c$ subsets of m populations for $i = 1, \ldots, m$ as $\hat{S}_{WKM}(t|Y > c) = \sum_{i=1}^{m} w_i S_i(t|Y_i > c)$; and $\hat{F}_{Y_1}(c) = \sum_{i=1}^{m} w_i \hat{F}_{Y_i}(c)$ where $\hat{F}_{Y_i}(c) = \frac{1}{n_i}\sum_{j=1}^{n_i} 1(Y_{ij} \leq c)$. Here $\hat{S}_{WKM}(t)$ is the proposed estimator for the 1st population's survival function, $S_1(t)$, which is the parameter of interest. So the weights are chosen to minimise the difference between $\hat{S}_{WKM}(t)$ and $S_1(t)$. Likewise, $\hat{F}_{Y_1}(c)$ is an estimator for the first population's cumulative distribution function. The steps of the algorithm proposed by Plante *et al*. (2008a; 2009), which is used to calculate weights, had been conducted as to never give 0 weight to the 1st population since it is the population of interest.

Simulation methodology and scenarios

It is aimed at comparing the AUC values obtained from the time-dependent ROC curve using the Kaplan-Meier function, with the AUC values obtained by using the weighted Kaplan-Meier function. For the number of populations m = 2, a variety of sample sizes (n_1-n_2: 25 – 50, 25 – 100, 50 – 100, 50 – 250, 100 – 250 and 100 – 500), death times and continuous diagnostic test results were generated from $[\log(X), Y] \sim N(\mu_1 = 0, \mu_2 = 0, \sigma_1 = 1, \sigma_2 = 1, \rho)$ by taking the correlation between the marker and the log(time) as $\rho = -0.7$ and $\rho = -0.8$. Per convention, ρ was taken to be negative so that the higher marker value indicates a smaller event time. Independent censoring times were generated from censored normal distribution as $[\log(V)] \sim N(\mu = 0, \sigma = 1)$ for fixed censoring rates. Censoring rates (c_1-c_2) were taken as 40 – 30 %, 40 – 50 %, 40 – 70 %, 60 – 30 %, 60 – 50 %, and 60 – 70 %. The ROC curve, which uses the weighted Kaplan-Meier function was calculated for the 1st group using the measurements of two groups, and the ROC curve that uses the Kaplan-Meier function was calculated for the 1st group using the measurements of the 1st group. Simulation strategy used by Heagerty and Zheng (2005) was implemented. To create the 'true' ROC curve, false positive (FP) values were fixed to be 0.01, 0.02,…, 0.99 and true positive (TP) values were estimated at these FP rates. The (FP, TP) pairs were estimated and then the TP rate corresponding to the given FP rate was interpolated for a given simulation. AUC values were calculated by the trapezoidal rule using these TP and FP pairs and then by averaging the AUC values over the number of simulations to get an estimate of the true AUC. TP, FP and AUC values were calculated using the survivalROC 1.0.3 package for the method which uses Kaplan-Meier estimator (Heagerty & Saha-Chaudhuri, 2013). ROC curves and AUC values were calculated by using the Kaplan-Meier and weighted Kaplan-Meier estimators for the 1st sample. One thousand repetitions were performed for each scenario.

Table 1: AUC values obtained for $\rho = -0.7$ for different censoring rates and sample sizes

c_1 = 0.40, c_2 = 0.30	WTDR				TDR			
n_1-n_2	Mean	SEM	SD	MSE	Mean	SEM	SD	MSE
25-50	0.80207	0.00361	0.11416	0.01591	0.78198	0.00523	0.16532	0.03276
25-100	0.81149	0.00330	0.10435	0.01284	0.78191	0.00526	0.16632	0.03310
50-100	0.83075	0.00234	0.07408	0.00611	0.81100	0.00384	0.12128	0.01670
50-250	0.83840	0.00196	0.06206	0.00415	0.81324	0.00382	0.12086	0.01641
100-250	0.84567	0.00145	0.04599	0.00222	0.83734	0.00226	0.07136	0.00543
100-500	0.84889	0.00125	0.03965	0.00162	0.83973	0.00225	0.07101	0.00530
500-500	0.85539	0.00074	0.02333	0.00054	0.85508	0.00096	0.03022	0.00091
500-1000	0.85407	0.00065	0.02058	0.00043	0.85302	0.00096	0.03026	0.00092
1000-1000	0.85599	0.00051	0.01611	0.00026	0.85551	0.00068	0.02162	0.00047

Continued -

- continued from page 14

$c_1 = 0.40, c_2 = 0.50$		WTDR				TDR		
n_1-n_2	Mean	SEM	SD	MSE	Mean	SEM	SD	MSE
25-50	0.79752	0.00375	0.11850	0.01738	0.78491	0.00527	0.16671	0.03279
25-100	0.81655	0.00336	0.10618	0.01279	0.79324	0.00529	0.16721	0.03186
50-100	0.82790	0.00241	0.07623	0.00658	0.81368	0.00382	0.12085	0.01635
50-250	0.83654	0.00200	0.06328	0.00441	0.81648	0.00375	0.11853	0.01490
100-250	0.84526	0.00153	0.04827	0.00243	0.84006	0.00227	0.07191	0.00541
100-500	0.84796	0.00130	0.04116	0.00181	0.83955	0.00228	0.07212	0.00636
500-500	0.85520	0.00078	0.02472	0.00061	0.85524	0.00096	0.03031	0.00092
500-1000	0.85396	0.00069	0.02192	0.00048	0.85275	0.00095	0.03003	0.00091
1000-1000	0.85588	0.00053	0.01689	0.00029	0.85552	0.00068	0.02148	0.00046
$c_1 = 0.40, c_2 = 0.70$		WTDR				TDR		
n_1-n_2	Mean	SEM	SD	MSE	Mean	SEM	SD	MSE
25-50	0.77819	0.00413	0.13062	0.02600	0.77763	0.00530	0.16753	0.03415
25-100	0.80576	0.00355	0.11230	0.01511	0.79267	0.00516	0.16320	0.03060
50-100	0.81739	0.00271	0.08569	0.00881	0.81342	0.00375	0.11856	0.01584
50-250	0.83122	0.00235	0.07432	0.00686	0.82158	0.00911	0.28817	0.08413
100-250	0.84011	0.00165	0.05206	0.00295	0.83600	0.00242	0.07658	0.00625
100-500	0.84795	0.00148	0.04677	0.00225	0.83955	0.00228	0.07206	0.00618
500-500	0.85467	0.00084	0.02652	0.00070	0.85525	0.00097	0.03052	0.00093
500-1000	0.85326	0.00076	0.02395	0.00058	0.85267	0.00095	0.03016	0.00092
1000-1000	0.85588	0.00058	0.01833	0.00034	0.85541	0.00069	0.02181	0.00048
$c_1 = 0.60, c_2 = 0.30$		WTDR				TDR		
n_1-n_2	Mean	SEM	SD	MSE	Mean	SEM	SD	MSE
25-50	0.77513	0.00453	0.14324	0.02701	0.73191	0.00648	0.20492	0.05731
25-100	0.78605	0.00411	0.13009	0.02177	0.73552	0.00634	0.20044	0.05461
50-100	0.81470	0.00316	0.09982	0.01165	0.76596	0.00529	0.16714	0.03599
50-250	0.82074	0.00010	0.08347	0.00819	0.76194	0.00017	0.16528	0.03610
100-250	0.83619	0.00186	0.05885	0.00385	0.80481	0.00351	0.11106	0.01493
100-500	0.84118	0.00165	0.05229	0.00295	0.80628	0.00328	0.10368	0.01319
500-500	0.85377	0.00087	0.02736	0.00075	0.85197	0.00121	0.03811	0.00147
500-1000	0.85319	0.00075	0.02374	0.00057	0.84912	0.00119	0.03761	0.00146
1000-1000	0.85556	0.00060	0.01895	0.00036	0.85441	0.00085	0.02702	0.00073
$c_1 = 0.60, c_2 = 0.50$		WTDR				TDR		
n_1-n_2	Mean	SEM	SD	MSE	Mean	SEM	SD	MSE
25-50	0.76811	0.00488	0.15426	0.03147	0.73798	0.00633	0.20030	0.05397
25-100	0.77560	0.00449	0.14204	0.02659	0.73138	0.00649	0.20517	0.05754
50-100	0.80122	0.00314	0.09922	0.01282	0.75650	0.00515	0.16289	0.03637
50-250	0.81418	0.00290	0.09185	0.01016	0.75916	0.00512	0.16185	0.03552
100-250	0.83260	0.00212	0.06707	0.00503	0.80500	0.00355	0.11218	0.01516
100-500	0.84014	0.00180	0.05685	0.00347	0.81483	0.00362	0.11436	0.01567
500-500	0.85128	0.00101	0.03195	0.00104	0.85093	0.00152	0.03895	0.00154
500-1000	0.85287	0.00080	0.02535	0.00065	0.84910	0.00119	0.03768	0.00147
1000-1000	0.85529	0.00063	0.02003	0.00040	0.85440	0.00085	0.02698	0.00073
$c_1 = 0.60, c_2 = 0.70$		WTDR				TDR		
n_1-n_2	Mean	SEM	SD	MSE	Mean	SEM	SD	MSE
25-50	0.75523	0.00497	0.15707	0.03477	0.74033	0.00628	0.19874	0.05280
25-100	0.77152	0.00461	0.14568	0.02831	0.73832	0.00635	0.20077	0.05408
50-100	0.78799	0.00356	0.11266	0.01728	0.76157	0.00522	0.16513	0.03613
50-250	0.81289	0.00306	0.09675	0.01120	0.76370	0.00531	0.16788	0.03664
100-250	0.82924	0.00207	0.06560	0.00501	0.80407	0.00344	0.10868	0.01448
100-500	0.84018	0.00180	0.05692	0.00349	0.81524	0.00335	0.10585	0.01284
500-500	0.85199	0.00099	0.03136	0.00100	0.85197	0.00121	0.03811	0.00147
500-1000	0.85178	0.00088	0.02776	0.00079	0.84911	0.00120	0.03787	0.00148
1000-1000	0.85491	0.00068	0.02153	0.00046	0.85442	0.00085	0.02681	0.00072

Table 2: AUC values obtained for $\rho = -0.8$ for different censoring rates and sample sizes

$c_1 = 0.40, c_2 = 0.30$			WTDR				TDR	
n_1-n_2	Mean	SEM	SD	MSE	Mean	SEM	SD	MSE
25-50	0.84973	0.00328	0.10387	0.01414	0.83168	0.00495	0.15646	0.02980
25-100	0.85961	0.00267	0.08456	0.00918	0.83462	0.00469	0.14826	0.02688
50-100	0.88167	0.00196	0.06183	0.00435	0.85954	0.00356	0.11252	0.01469
50-250	0.88223	0.00174	0.05502	0.00353	0.85544	0.00353	0.11153	0.01486
100-250	0.89367	0.00117	0.03704	0.00149	0.88590	0.00194	0.06150	0.00413
100-500	0.89747	0.00105	0.03312	0.00115	0.88550	0.00192	0.06064	0.00386
500-500	0.90309	0.00059	0.01866	0.00035	0.90223	0.00077	0.02427	0.00060
500-1000	0.90344	0.00054	0.01701	0.00029	0.90150	0.00076	0.02411	0.00059
1000-1000	0.90438	0.00039	0.01227	0.00015	0.90393	0.00050	0.01578	0.00025
$c_1 = 0.40, c_2 = 0.50$			WTDR				TDR	
n_1-n_2	Mean	SEM	SD	MSE	Mean	SEM	SD	MSE
25-50	0.84852	0.00332	0.10508	0.01419	0.83496	0.00488	0.15436	0.02867
25-100	0.85616	0.00282	0.08908	0.01029	0.83376	0.00484	0.15316	0.02847
50-100	0.87354	0.00211	0.06665	0.00541	0.86126	0.00331	0.10477	0.01286
50-250	0.87998	0.00178	0.05623	0.00377	0.85864	0.00337	0.10641	0.01344
100-250	0.89125	0.00125	0.03937	0.00173	0.88474	0.00192	0.06086	0.00410
100-500	0.89607	0.00119	0.03772	0.00150	0.88677	0.00188	0.05947	0.00386
500-500	0.90262	0.00062	0.01974	0.00039	0.90221	0.00077	0.02442	0.00060
500-1000	0.90394	0.00054	0.01703	0.00029	0.90252	0.00074	0.02329	0.00055
1000-1000	0.90403	0.00042	0.01334	0.00018	0.90345	0.00051	0.01613	0.00026
$c_1 = 0.40, c_2 = 0.70$			WTDR				TDR	
n_1-n_2	Mean	SEM	SD	MSE	Mean	SEM	SD	MSE
25-50	0.83973	0.00340	0.10752	0.01578	0.83010	0.00486	0.15353	0.02912
25-100	0.84548	0.00336	0.10617	0.01477	0.83741	0.00469	0.14841	0.02654
50-100	0.86063	0.00256	0.08105	0.00851	0.85814	0.00345	0.10911	0.01406
50-250	0.87896	0.00186	0.05887	0.00413	0.86227	0.00326	0.10313	0.01243
100-250	0.88855	0.00137	0.04319	0.00213	0.88421	0.00186	0.05895	0.00389
100-500	0.89135	0.00122	0.03856	0.00166	0.88444	0.00185	0.05843	0.00382
500-500	0.90183	0.00067	0.02112	0.00045	0.90222	0.00077	0.02437	0.00060
500-1000	0.90225	0.00060	0.01885	0.00036	0.90120	0.00075	0.02377	0.00058
1000-1000	0.90430	0.00046	0.01452	0.00021	0.90391	0.00052	0.01660	0.00028
$c_1 = 0.60, c_2 = 0.30$			WTDR				TDR	
n_1-n_2	Mean	SEM	SD	MSE	Mean	SEM	SD	MSE
25-50	0.80253	0.00443	0.14007	0.03005	0.76039	0.00643	0.20348	0.06220
25-100	0.83423	0.00426	0.13484	0.02314	0.77237	0.00656	0.20758	0.06056
50-100	0.85680	0.00287	0.09074	0.01052	0.81170	0.00513	0.16217	0.03493
50-250	0.86569	0.00243	0.07688	0.00743	0.81060	0.00508	0.16057	0.03462
100-250	0.88467	0.00156	0.04946	0.00284	0.85629	0.00303	0.09594	0.01154
100-500	0.88797	0.00154	0.04867	0.00265	0.85366	0.00343	0.10838	0.01434
500-500	0.90202	0.00069	0.02181	0.00048	0.89931	0.00096	0.03037	0.00095
500-1000	0.90197	0.00063	0.02006	0.00041	0.89795	0.00098	0.03093	0.00100
1000-1000	0.90356	0.00051	0.01612	0.00026	0.90193	0.00069	0.02168	0.00048
$c_1 = 0.60, c_2 = 0.50$			WTDR				TDR	
n_1-n_2	Mean	SEM	SD	MSE	Mean	SEM	SD	MSE
25-50	0.79308	0.00481	0.15195	0.03556	0.76148	0.00643	0.20320	0.06183
25-100	0.81490	0.00448	0.14163	0.02813	0.76236	0.00648	0.20479	0.06222
50-100	0.85437	0.00355	0.11221	0.01512	0.81986	0.00520	0.16437	0.03419
50-250	0.86280	0.00311	0.09838	0.01143	0.81877	0.00512	0.16202	0.03361
100-250	0.87963	0.00159	0.05026	0.00315	0.85438	0.00293	0.09275	0.01113

Continued -

- continued from page 16

$c_1 = 0.60, c_2 = 0.50$		WTDR				TDR		
n_1-n_2	Mean	SEM	SD	MSE	Mean	SEM	SD	MSE
100-500	0.88618	0.00156	0.04920	0.00276	0.85619	0.00303	0.09568	0.01150
500-500	0.90122	0.00073	0.02301	0.00054	0.89931	0.00096	0.03038	0.00095
500-1000	0.90197	0.00063	0.02006	0.00041	0.89795	0.00098	0.03093	0.00100
1000-1000	0.90214	0.00052	0.01638	0.00027	0.90094	0.00069	0.02174	0.00049
$c_1 = 0.60, c_2 = 0.70$		WTDR				TDR		
n_1-n_2	Mean	SEM	SD	MSE	Mean	SEM	SD	MSE
25-50	0.79266	0.00496	0.15669	0.03709	0.76334	0.00642	0.20290	0.06112
25-100	0.81356	0.00456	0.14423	0.02909	0.77114	0.00644	0.20357	0.05925
50-100	0.84163	0.00341	0.10791	0.01562	0.81065	0.0051	0.16132	0.03485
50-250	0.85728	0.00307	0.09705	0.01166	0.81300	0.00512	0.16202	0.03464
100-250	0.87249	0.00186	0.05888	0.00450	0.85417	0.00296	0.09375	0.01134
100-500	0.88230	0.00169	0.05333	0.00334	0.85406	0.00295	0.09341	0.01128
500-500	0.90001	0.00079	0.02484	0.00064	0.89931	0.00096	0.03044	0.00096
500-1000	0.90184	0.00063	0.02006	0.00041	0.89795	0.00098	0.03093	0.00100
1000-1000	0.90354	0.00055	0.01725	0.00030	0.90266	0.00068	0.02155	0.00047

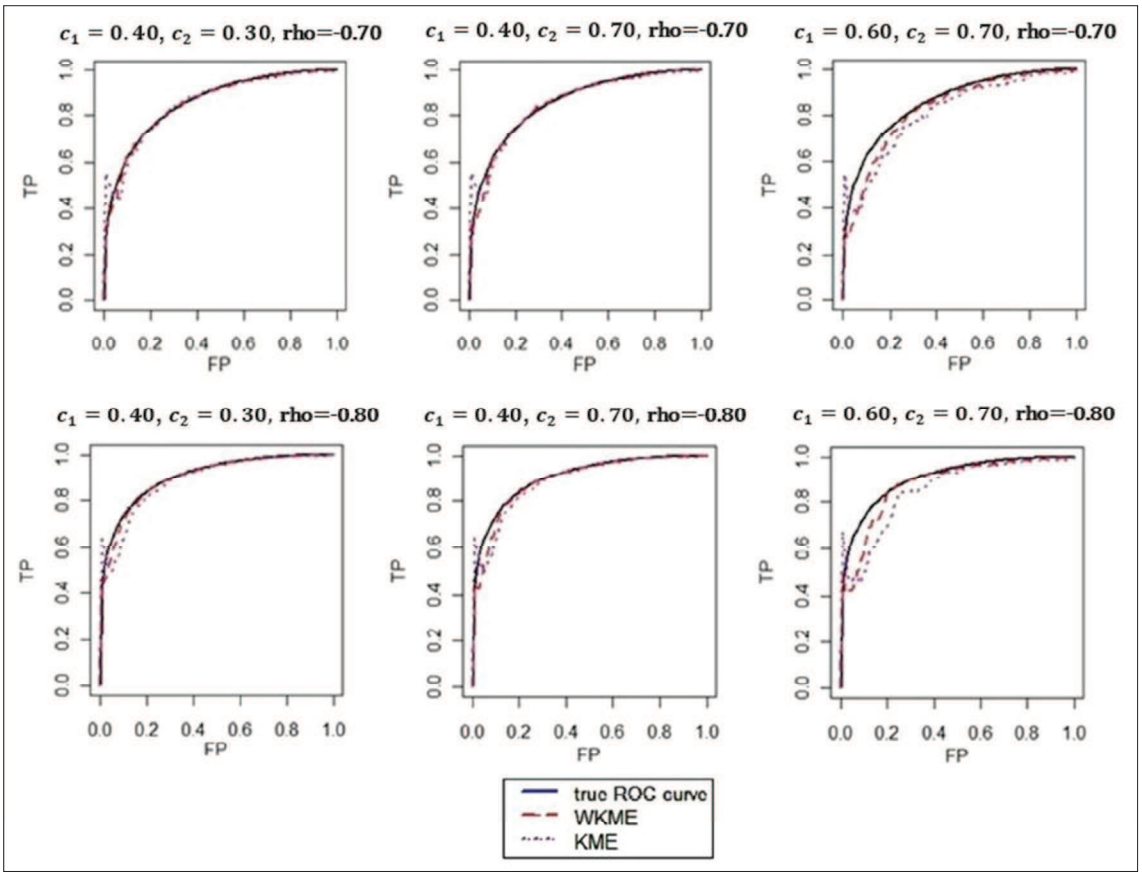

Figure 1: Cumulative/dynamic ROC curves for different censoring rates and different rho values (WKME: time-dependent ROC curves which use weighted Kaplan-Meier estimation, KME: time-dependent ROC curves which use Kaplan-Meier estimation, c_1: censoring rate for group 1, c_2: censoring rate for group 2).

Simulation results

Mean, standard deviation (SD), standard error of mean (SEM) and mean square error (MSE) for the estimate of AUC(t), for log(t) = 1 and for the data generated with different censoring rates and different sample sizes were calculated. The simulation results for $\rho = -0.7$ are given in Table 1 and for $\rho = -0.8$ in Table 2. The real AUC values were 0.855 for $\rho = -0.7$ and 0.905 for $\rho = -0.8$.

ROC curve plots for $n_1 = 100$, $n_2 = 250$; $\rho_1 = -0.7$, $\rho_2 = -0.8$; $c_1 - c_2 = 0.40 - 0.30$, $c_1 - c_2 = 0.40 - 0.70$ and $c_1 - c_2 = 0.60 - 0.70$ are given in Figure 1.

A real application

Patients

Data from 24 patients with carcinoma of the ampulla of Vater (mean ± standard deviation age: 59.92 ± 13.55; 11 males and 13 females) and 30 patients with adenocarcinoma of the pancreatic head (mean ± standard deviation age: 58.10 ± 13.22; 22 males and 8 females) were examined retrospectively (Uludag University ethical committee no: 2016-2/34). The Kaplan-Meier mean survival times were 24.32 months [standard error: 2.86; 95 % CI: (18.71 – 29.94)] for the ampullary cancer patients and 24.38 months [standard error: 4.30; 95 % CI: (15.95 – 32.81)] for the patients with adenocarcinoma of the pancreatic head.

Ampullary cancer (i.e., carcinoma of the ampulla of Vater) is a fairly rare pancreatic cancer that starts at the location where the bile duct and the pancreatic duct meet and empty into the duodenum (the ampulla of Vater). It has been aimed at assessing the value of the preoperative plasma CA 19-9 level in predicting the mortality of ampullary cancer patients and to examine the proper cut-off points for the CA 19-9 level by using weighted time-dependent ROC analysis. AUC values and cut-off points for CA 19-9 for the ampullary cancer dataset (which has smaller sample size) were estimated by borrowing strength from the second dataset of pancreatic-head cancer patients.

Data analysis

To assess the performance of CA 19-9 across the study period, AUC(t) values were calculated for different t values with weighted time-dependent ROC (WTDR) and time-dependent ROC (TDR) methods. Bootstrapped variance and 95 % CIs were calculated for the AUC values from 500 bootstrap repetitions of the dataset. The null hypothesis that the AUC did not differ from 0.5 was tested. Cut-off values were determined by means of producing the highest Youden J index for significant AUC (t) values. Analyses were performed using R 3.3.0 software (R Core Team, 2013).

RESULTS AND DISCUSSION

By using the WTDR method, CA 19-9 was found to be a significant marker between the 1st and 35th months. At early times, the cut-off for the CA 19-9 level was 192, but after one and a half years, this cutoff moved to 138. However, by using the TDR method, CA 19-9 was found to be significant only from the 17th to the 26th months. The sensitivity, specificity and cut-off values exhibited variations according to different time points for CA 19-9 (Tables 3 and 4).

In the present study, weighted time-dependent ROC curves that integrated additional data from different populations by using a weighted Kaplan-Meier estimator was presented. From the simulation studies, it was shown that ROC curves that were obtained by using weighted Kaplan-Meier method were closer to the real ROC curves and also the AUC values obtained by using WTDR produced MSE, SEM and SD values that were smaller than those of the TDR curves calculated from the Kaplan-Meier function for all sample sizes and censoring rates.

As expected, MSE values decreased as sample sizes increased; however, MSE values also decreased as the correlation between the marker value and the survival time increased. Better results were obtained for the situation where $\rho = -0.7$ than $\rho = -0.8$ for both of the methods. Additionally, as the censoring rates increased, it is observed that the MSE, SEM and SD values also increased, both for the AUC values obtained by using the weighted Kaplan-Meier function and for the AUC values obtained by using the classical Kaplan-Meier function. For the time-dependent ROC curves, which were obtained by using the weighted Kaplan-Meier function, the censoring rate of the first group (c_1) had a much larger effect than the censoring rate of the second group (c_2) on the increment of the MSE. However, for the same sample sizes and the same censoring rates for group 1, the weighted time-dependent ROC curves always yielded smaller MSE, SEM and SD values regardless of the value of the second group's censoring rate (c_2). Moreover, the differences between the MSE, SEM and SD values for the WTDR and the TDR methods became more apparent as c_1 increased from 0.40 to 0.60. In large sample sizes, an improvement was seen in the performances of both methods; however the MSE, SEM and SD values were still smaller for the weighted time-dependent ROC curves.

Table 3: AUCs and related p values for CA 19-9 in different time points obtained by WTDR

Time (month)	AUC	p value	Cut-off	Youden J	Sensitivity	Specificity
1	0.818	< 0.001	192	0.818	1.00	0.818
2 – 4	0.932	< 0.001	192	0.841	0.984	0.857
5 – 12	0.965	< 0.001	192	0.896	1.00	0.896
13	0.950	< 0.001	192	0.885	0.981	0.904
14 – 16	0.904	< 0.001	192	0.788	0.886	0.901
17 – 18	0.903	< 0.001	138.19	0.841	0.888	0.954
19 – 26	0.878	0.008	138.19	0.837	0.854	0.984
27 – 34	0.835	0.038	138.19	0.508	0.532	0.977
35	0.817	0.023	138.19	0.477	0.477	1.000

Table 4: AUCs and related p values for CA 19-9 in different time points obtained by TDR

Time (month)	AUC	p value	Cut-off	Youden J	Sensitivity	Specificity
1	0.761	0.054	-	-	-	-
2 – 4	0.610	0.248	-	-	-	-
5 – 16	0.767	0.071	-	-	-	-
17 – 26	0.837	0.012	192	0.697	0.697	1.00
27 – 33	0.722	0.057	-	-	-	-

Time-dependent ROC curves provide information about the time interval in which a diagnostic test or marker is most reliable and how reliable it is within that time interval. They also illuminate changes in the discriminative ability of the diagnostic test from the start of the study across the observation period. When there are additional data from similar populations, MAMSE weighted Kaplan-Meier estimator proposed by Plante (2009) generates smoother ROC curves. Plante (2009) suggested using bootstrapping techniques for obtaining standard errors and confidence intervals of the weighted Kaplan-Meier estimators. In the present study, a weighted Kaplan-Meier estimator has been used to generate time-dependent ROC curves and to estimate AUC values in pancreatic cancer patients with regard to CA 19-9, and p values and confidence intervals of AUC values were obtained for that real data. Differences in AUC values as well as cut-offs, sensitivity and specificity emerged for the two approaches. As the sample size increased for the WTDR curves, information on more time intervals has been obtained.

The usage of the weighted Kaplan-Meier function in ROC curves was investigated and it was applied to a real data example. The results were comparable to those from simulation studies. Time-dependent ROC curves using weighted Kaplan-Meier functions would be useful in practice, especially when sample sizes are small. For large censoring rates the sample size for the conditional survival functions may be small. Since the sample size has been increased by borrowing strength from the other population, the weighted Kaplan-Meier estimate gave smoother ROC curves.

REFERENCES

1. Akritas M. (1994). Nearest neighbor estimation of a bivariate distribution under random censoring. *The Annals of Statistics* **22**(3): 1299 – 1327. DOI: https://doi.org/10.1214/aos/1176325630
2. Blanche P., Dartigues J.F. & Jacqmin-Gadda H. (2013a). Review and comparison of ROC curve estimators for a time-dependent outcome with marker-dependent censoring. *Biometrical Journal* **55**(5): 687 – 704. DOI: https://doi.org/10.1002/bimj.201200045
3. Blanche P., Latouche A. & Viallon V. (2013b). Time-dependent AUC with right-censored data: a survey. *Risk Assessment and Evaluation of Predictions* **15**: 239 – 251. DOI: https://doi.org/10.1007/978-1-4614-8981-8_11
4. Cai T., Pepe M.S., Lumley T., Zheng Y. & Jenney N.S. (2003). The sensitivity and specificity of markers for event times. *UW Biostatistics Working Paper Series. Working Paper 188*. University of Washington, USA.
5. Cai T., Pepe M.S., Zheng Y., Lumley T. & Jenny N.S. (2006).The sensitivity and specificity of markers for event times. *Biostatistics* **7**(2): 182 – 197. DOI: https://doi.org/10.1093/biostatistics/kxi047
6. Cai T., Gerds T.A., Zheng Y. & Chen J. (2011). Robust prediction of t-year survival with data from multiple studies. *Biometrics* **67**(2): 436 – 444. DOI: https://doi.org/10.1111/j.1541-0420.2010.01462.x

7. Chambless L.E. & Diao G. (2006). Estimation of time-dependent area under the ROC curve for long-term risk prediction. *Statistics in Medicine* **25**(20): 3474 – 3486. DOI: https://doi.org/10.1002/sim.2299
8. Campbell G. (1994). Advances in statistical methodology for the evaluation of diagnostic and laboratory tests. *Statistics in Medicine* **13**: 499 – 508. DOI: https://doi.org/10.1002/sim.4780130513
9. Chiang C.T. & Hung H. (2010). Non-parametric estimation for time-dependent AUC. *Journal of Statistical Planning and Inference* **140**(5): 1162 – 1174. DOI: https://doi.org/10.1016/j.jspi.2009.10.012
10. Etzioni R., Pepe M., Longton G., Hu C. & Goodman G. (1999). Incorporating the time dimension in receiver operating characteristic curves: a case study of prostate cancer. *Medical Decision Making Journal* **19**(3): 242 – 251. DOI: https://doi.org/10.1177/0272989X9901900303
11. Hanley J.A. & McNeil B. (1982). The meaning and use of the area under the receiver operating characteristic (ROC) curve. *Radiology* **143**(1): 29 – 36. DOI: https://doi.org/10.1148/radiology.143.1.7063747
12. Heagerty P.J., Lumley T. & Pepe M.S. (2000). Time-dependent ROC curves for censored survival data and a diagnostic marker. *Biometrics* **56**(2): 337 – 344. DOI: https://doi.org/10.1111/j.0006-341X.2000.00337.x
13. Heagerty P.J. & Saha-Chaudhuri P. (2013). Time-dependent ROC curve estimation from censored survival data, *[Package 'survivalROC']. Version 1.0.3.*
14. Heagerty P.J. & Zheng Y. (2005). Survival model predictive accuracy and ROC curves. *Biometrics* **61**(1): 92 – 105. DOI: https://doi.org/10.1111/j.0006-341X.2005.030814.x
15. Hu F. (1994). Relevance weighted smoothing and a new bootstrap method. *Technical Report*. Department of Statistics, University of British Columbia, USA.
16. Hu F. & Zidek J.V. (2002). The weighted likelihood. *Canadian Journal of Statistics* **30**(3): 347 – 371. DOI: https://doi.org/10.2307/3316141
17. Hung H. & Chiang C.T. (2010a). Optimal composite markers for time-dependent receiver operating characteristic curves with censored survival data. *Scandinavian Journal of Statistics* **37**(4): 664 – 679. DOI: https://doi.org/10.1111/j.1467-9469.2009.00683.x
18. Hung H. & Chiang C.T. (2010b). Estimation methods for time-dependent AUC models with survival data. *Canadian Journal of Statistics* **38**(1): 8 – 26.
19. James W. & Stein C. (1961). Estimation with quadratic loss, *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, volume 1 (ed. J. Neyman), University of California Press, California, USA, pp. 361 – 379.
20. Kaplan E.L. & Meier P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association* **53**(282): 457 – 481. DOI: https://doi.org/10.1080/01621459.1958.10501452
21. Lambert J. & Chevret S. (2016). Summary measure of discrimination in survival models based on cumulative/dynamic time-dependent ROC curves. *Statistical Methods in Medical Research* **25**(5): 2088 – 2102. DOI: https://doi.org/10.1177/0962280213515571
22. Li L., Hu B. & Greene T. (2015). A simple method to estimate the time-dependent ROC curve under right censoring. *COBRA Preprint Series*, *Working Paper 114*. Available at *http: //biostats.bepress.com/cobra/art114*
23. Li S. (2016). Estimating time-dependent ROC curves using data under prevalent sampling. *Statistics in Medicine* **36**(8): 1285 – 1301. DOI: https://doi.org/10.1002/sim.7184
24. Martinez-Camblor P., Bayon G.F. & Perez-Fernandez S. (2016). Cumulative/dynamic ROC curve estimation. *Journal of Statistical Computation and Simulation* **86**(17): 3582 – 3594. DOI: https://doi.org/10.1080/00949655.2016.1175442
25. Pepe M.S. (2003). *The Statistical Evaluation of Medical Tests for Classification and Prediction*. Oxford University Press, UK.
26. Pepe M., Zheng Y., Jin Y., Huang Y., Parikh C. & Levy W. (2008). Evaluating the ROC performance of markers for future events. *Lifetime Data Analysis* **14**(1): 86 – 113. DOI: https://doi.org/10.1007/s10985-007-9073-x
27. Plante J.F. (2008a). Nonparametric adaptive likelihood weights. *Canadian Journal of Statistics* **36**(3): 443 – 461. DOI: https://doi.org/10.1002/cjs.5550360308
28. Plante J.F. (2008b). Calculation of minimum averaged mean squared error (MAMSE) weights *[The MAMSE Package]. Version 0.1-1.*
29. Plante J.F. (2009). About an adaptively weighted Kaplan-Meier estimate. *Lifetime Data Analysis* **15**(3): 295 – 315. DOI: https://doi.org/10.1007/s10985-009-9120-x
30. R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. *http://www.R-project.org/*
31. Rodríguez-Álvarez M.X., Meira-Machado L., Abu-Assi E. & Raposeiras-Roubín S. (2016). Nonparametric estimation of time-dependent ROC curves conditional on a continuous covariate. *Statistics in Medicine* **35**(7): 1090 – 1102. DOI: https://doi.org/10.1002/sim.6769
32. Schmid M., Kestler H.A. & Potapov S. (2015). On the validity of time-dependent AUC estimators. *Briefings in Bioinformatics* **16**(1): 153 – 168. DOI: https://doi.org/10.1093/bib/bbt059
33. Slate E.H. & Turnbull B.W. (2000). Statistical models for longitudinal biomarkers of disease onset. *Statistics in Medicine* **19**(4): 617 – 637.
34. Song X. & Zhou X. (2008). A semiparametric approach for the covariate-specific ROC curve with survival outcome. *Statistica Sinica* 1**8**: 947 – 965.
35. Uno H., Cai T., Tian L. & Wei L. (2007). Evaluating prediction rules for t-year survivors with censored regression models. *Journal of the American Statistical Association* **102**(478): 527 – 537. DOI: https://doi.org/10.1198/016214507000000149
36. Viallon V. & Latouche A. (2011). Discrimination measures for survival outcomes: connection between the AUC and the predictiveness curve. *Biometrical Journal* **53**(2): 217 – 236. DOI: https://doi.org/10.1002/bimj.201000153

37. Wang X. (2001). Maximum weighted likelihood estimation. *PhD thesis*, The University of British Columbia, Vancouver, Canada.
38. Wang X., Van Eeden C. & Zidek J.V. (2004). Asymptotic properties of maximum weighted likelihood estimators. *Journal of Statistical Planning and Inference* **119**: 37 – 54.
39. Wang X. & Zidek J.V. (2005). Selecting likelihood weights by cross-validation. *Annals of Statistics* **33**(2): 463 – 501.
DOI: https://doi.org/10.1214/009053604000001309
40. Wolf P., Schmidt G. & Ulma K. (2011). The use of ROC for defining the validity of the prognostic index in censored data. *Statistics and Probability Letters* **81**(7): 783 – 791.
DOI: https://doi.org/10.1016/j.spl.2011.02.021