Time Profile of Time-Dependent Area Under the ROC Curve for Survival Data

J. Lambert[1,2,*], R. Porcher[2] and S. Chevret[1,2]

[1]*Equipe de recherche Epidémiologie Clinique et STatistiques pour la Recherche en sAnté (ECSTRA), Inserm et Université Paris 7, France*

[2]*UMR-1153, Centre Recherche en Epidémiologie et bioStatistiques Sorbonne Paris Cité CRESS, direction: Philippe Ravaud, France*

Abstract: In the setting of survival analysis, the time-dependent area under the receiver operating characteristic curve (AUC) has been proposed as a discrimination measure of interest. In contrast with the diagnostic setting, the definitions of time-dependent sensitivity and specificity are required. This paper evaluates the time-dependent profile of the resulting AUC(t), which has not been previously assessed. We show that, even when the effect of a binary biomarker on the hazard rate is constant, the value of AUC(t) varies over time according to the prevalence of the marker. The Time-profile of the continuous biomarker is illustrated with numerical integration, and data on several prognostic factors in AML are examined.

Keywords: Survival analysis, Prognostic models, Time-dependent AUC, Proportional hazards models.

1. INTRODUCTION

Prospective cohort studies are commonly conducted to assess and compare the prognostic value of several biomarkers in a population of interest, using a right-censored endpoint. In such a survival setting, there has been a growing interest in predictive accuracy measures [1], ranging from measures of the proportion of variation explained by the covariates, which extend the R^2 to survival data [2,3], to reclassification measures that focus on the incremental value of biomarkers compared to pre-existing risk scores [4]. Concept of discrimination encompasses measures such as sensitivity, specificity, Receiver Operating Characteristic (ROC) curve and area under the ROC curve and is related to how a biomarker or a model can distinguish low from high risk patients, in a direct extension of the diagnostic settings where we aim at distinguishing between patients with and without the disease. Calibration and discrimination are the main components for describing predictive accuracy and discrimination appears to be of primary interest since, if a model have poor discrimination, no calibration can correct the model. On the contrary, a model with good discrimination can be recalibrated [5].

Recently, Heagerty and colleagues [6] introduced the concepts of time-dependent sensitivity and specificity, and the area under the resulting time-dependent ROC curve (AUC) to apply discrimination measures to the setting of prognostic biomarkers. On the model of cases and control in the diagnostic setting, they defined several types of time-dependent case and time-dependent control. Briefly, cases can be defined as “cumulative” or “incident” and controls can be defined as “dynamic” or “static”, resulting in three different types of discrimination measures. In addition to the choice of discrimination measure, because the outcome status is time-dependent, discrimination measures such as sensitivity, specificity, ROC curves and their AUC values are functions of time as well. Thus, when evaluating the predictive accuracy of any biomarker for distinguishing *t*-year survivors, both the choice the discrimination measure and the choice of *t* may influence the result. The use of varying values of *t* to evaluate the predictive accuracy of biomarkers is not rare, notably in the biomedical litterature where the time-dependent AUC is increasingly used and plotted over time. For example, in a recent work, Maisel *et al*. [7] reported the prognostic value of baseline copeptin and cardiac troponin based on trial data from 1,967 patients with chest pain. They compared the time-dependent AUC values for survival up to 180 days, at several time points, to conclude that copeptin was of prime interest for short-term (<30 days) mortality prediction and that cardiac troponin was stronger for long-term (>60 days) outcome prediction. This suggests that the prognostic impact of these biomarkers is time-variant. Time varying effect of a covariate is a well-known cause of non-proportional hazards, which can be handled through the useof time-dependent coefficients in the classical semi-parametric Cox proportional hazards model [8]. Intuitively, one could expect time-dependent AUC to vary in the same

*Address correspondence to this author at the équipe Epidemiologie Clinique et STatistique pour la Recherche en sAnté (ECSTRA), Service de Biostatistique et Information Médicale, Hôpital Saint Louis, 1 av Claude Vellefaux, 75010 Paris, France; Tel: +33142385297; Fax: +33142499745; E-mail: jerome.lambert@univ-paris-diderot.fr

E-ISSN: 1929-6029/15

direction as the time-varying coefficients. Nevertheless, it appears that time-varyingeffect in the not the onlysource of variation across time of the accuracy measures. The aim of this work was to study the time profile of the time-dependent AUC, in the specific case of a time-fixed effect of a biomarker. In Section 2, we present several definitions of the time-dependent AUC in the standard setting of a baseline biomarker and univariate survival data. Section 3 presents real data from a recent clinical study conducted with 278 patients with acute myeloblasticleukaemia, which further motivated this work and data from an ancient trial on primary biliary cirrhosis which is often used to illustrate methodological issues in survival analysis [9]. Section 4 presents some properties of the time-dependent AUC as a function of time in the case of a binary prognostic biomarker and exponential survival times, and Section 5 assesses the expected time-profile in the more complex case of a continuous biomarker and Weibull failure times using numerical integration. Finally, some discussion is provided in Section 6.

2. TIME-DEPENDENT AUC

In the diagnostic setting, the outcome *D* is a binary variable (*D*=1 for cases and *D*=0 for controls), and the ROC curve for a continuous diagnostic marker *Z* plots the true positive rate or sensitivity, *TPR*(*c*)=Pr(*Z*>*c*|*D*=1) against the false positive rate, or one minus the specificity, *FPR*(*c*)=Pr(*Z*>*c*|*D*=0) for all possible threshold values *c*. The area under the ROC curve, or AUC, is then defined as:

$$AUC = \int_{-\infty}^{+\infty} TPR(c)\frac{\partial FPR(c)}{\partial c} \quad (1)$$

A simple interpretation of AUC is the probability that the diagnostic test will correctly order a random couple (*i*,*j*) of case and control, which, assuming that higher values of the diagnostic test are associated with higher probabilities of the disease, is:

$$AUC = P(Z_i > Z_j \,|\, D_i = 1, D_j = 0) \quad (2)$$

To extend the diagnostic to the prognostic setting, one must account for the feature of survival data. Let *T* denote the time to some event of interest. We consider the practical situation of a time-independent biomarker assessed at baseline, denoted by *Z*, and we assume that higher values of *Z* are associated with shorter event times. The survival time *T* can be seen as a time-dependent binary outcome *via* the counting process representation $N_i^*(t)$.

Several extensions of the cross-sectional discriminative measures have been proposed [8], which differ in how the time-dependent "cases" (events) and "controls" (non-events) are defined. Time-dependent cases can be defined either as *incident* cases, corresponding to patients who failed at *t* (*T*=*t*) or *cumulative* cases, corresponding to patients who failed by *t* (*T*≤*t*). We subsequently define *incident* true positive rate $(TPR^I(c,t) = P(Z > c|T = t)$, and *cumulative* true positive rate $(TPR^C(c,t) = P(Z > c|T \le t)$. Time-dependent controls are mostly defined as *dynamic*, corresponding to patients still free of the event at *t* (*T*>*t*) and thus defining a *dynamic* false positive rate $(FPR^D(c,t) = P(Z > c|T > t)$. Combining these definitions, the two commonly used time-dependent AUCs, namely *cumulative/dynamic* and *incident/ dynamic* AUC(t), are defined as follows:

$$AUC^{C,D}(t) = P(Z_i > Z_j | T_i \le t, T_j > t) \quad (3)$$

$$AUC^{I,D}(t) = P(Z_i > Z_j | T_i = t, T_j > t) \quad (4)$$

Due to the nature of time-to-event data, we only observe $X=T\wedge C$, where *C* represents an independent censoring time. Several estimators of the $AUC^{C,D}(t)$ and the $AUC^{I,D}(t)$have been proposed to cope with censoring. For cumulative/dynamic AUC(t), estimators can be based on estimates of the conditional survival function $P(T > t|Z = z)$, using either the Cox model to derive the estimate $\hat{S}_n(t|z)$ [10] or a nearest neighbour estimator for the bivariate distribution function of (*Z*,*T*) [11], or based on the inverse probability of censoring weighted (IPCW) estimator [12,13]. For incident/dynamic AUC(t), semi-parametric methods using a Cox model have been proposed to estimate incident sensitivity and dynamic specificity, while standard numerical integration techniques are used to compute an estimate for $AUC^{I,D}(t)$. Non-parametric methods have also been proposed [14]. Most of these estimators have been made available in standard softwares such as R [15] with the timeROC, risksetROC, survivalROC and survAUC packages [16,17,18,19].

3. ILLUSTRATING EXAMPLE

We used prospectively recorded data to assess the prognostic value of biomolecular markers in acute myeloblasticleukaemia (AML). This cohort study was nested into a randomised controlled trial that showed the event-free survival and overall survival benefit of gemtuzumabozogamicin (GO) during induction and consolidation in AML patients aged 50-70 years (ALFA-

Table 1: ALFA-0701 Study: Estimated Hazards Ratio (HR) of Death with Ninety-Five Percent Confidence Intervals (95%CI) for Potential Predictors

Variable	HR	95%CI	p-value	Prevalence
SNP-A lesions	2.52	(1.68-3.77)	<0.001	54%
NPM1 Wild	1.39	(0.95-2.03)	0.089	66%
DNMT3A mutated	2.21	(1.31-3.72)	0.003	34%

0701 study) [20]. We used three biomarkers based on analysis of karyotype by single nucleotide polymorphism array (lesions of SNP-A) or molecular analysis (DNMT3A mutation, or NPM1 mutation), all measured at baseline as potential predictors of death. Among the 278 patients enrolled in the study, 131 died during follow-up, with 1-year and 2-years overall survival rates estimated at 71.3% (95%CI: 66.1% – 76.8%) and 48.4% (95%CI: 42.1% – 55.7%), respectively and median survival of 22.2 months. Table **1** reports the estimated hazards ratios of death (data truncated at 2 years) of these biomarkers, which exhibit varying strengths of association with binary covariates from 1.4 (NPM1 wild type) to 2.5 (SNP-A lesions).

We further assessed the discriminative value of these biomarkers, first based on the $AUC^{C,D}(t)$ computed for increasing values of time using the timeROC R package as illustrated in Figure **1** (upper panel). There were some surprising findings. First, at some specific time points, there was no direct relationship between the value of the $AUC^{C,D}(t)$ and the estimated HR for binary covariates. For instance, while the estimated HR of SNP-A was higher than that of

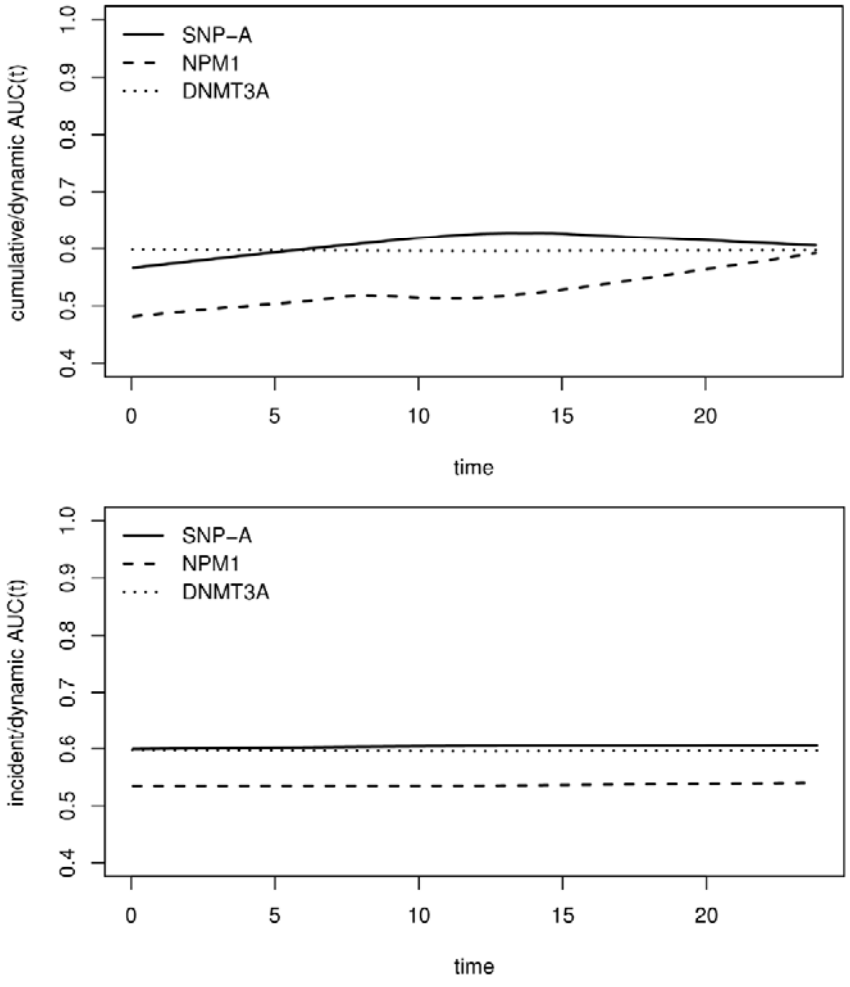

Figure 1: ALFA-0701 study. Time dependent AUC of biomarkers for discrimination of mortality over time. Curves smoothed by a locally weighted polynomial regression. Upper panel refers to $AUC^{C,D}(t)$ and lower panel to and $AUC^{I,D}(t)$time profiles.

DNMT3A, the $AUC^{C,D}(t)$ associated with DNMT3A was sometimes higher than that of SNP-A. Second, the time profiles of these $AUC^{C,D}(t)$ were not all the same, with some exhibiting an increase over time (such as that of NPM1) and others having different trends over time (e.g., SNP-A).

When computing the $AUC^{I,D}(t)$ using the risksetROC R package, time profiles appeared somewhat different (Figure **1**, lower pannel), with rather constant values over time ranked in agreement with increasing hazards ratios.

In all cases, there was no statistical evidence of non-proportionality of hazards based on weighted Schoenfeld residuals [21], with p-values ranging from 0.22 to 0.78 according to the prognostic marker.

The second example is based on data from the Mayo Clinic trial in primary biliary cirrhosis (PBC) conducted between 1974 and 1984. A total of 424 PBC patients, referred to Mayo Clinic, were followed until either death or liver transplantation (median follow-up of seven years). Patients 'age and presence of ascites at enrollment in the study were both predictors of event-free survival (Table **2**), with no evidence for a time-varying effect (p-value for non-proportionality of hazards: 0.123 and 0.3). Time profiles of cumulative/dynamic and incident/dynamic AUC(t) were both decreasing for ascites, whereas age exhibited a rather constant incident/dyamic AUC(t) and a decreasing then increasing cumulative/dynamic AUC(t)

We thus attempted to obtain further insights in the time profiles of the AUC(t).

4. THE TIME-DEPENDENT AUC FOR A BINARY MARKER AND AN EXPONENTIAL SURVIVAL TIME

We first analytically derived the cumulative/dynamic AUC and the incident/dynamic AUC in the particular case of an exponentially distributed survival time and a binary marker. We assume that *Z* is distributed according to a Bernoulli distribution of probability *p* and *T* according to an exponential distribution of conditional mean $(T|Z=z)=(\lambda_0 \exp(\beta Z))^{-1}$. Let θ=exp(β) be the hazard ratio associated with *Z*. To account for tied *Z*s, previous equations (3) and (4) have to be modified as follows:

$$AUC^{C,D}(t)=P(Z_i>Z_j|T_i\le t,T_j>t)+\frac{1}{2}P(Z_i>Z_j|T_i\le t,T_j>t) \quad (5)$$

$$AUC^{I,D}(t)=P(Z_i>Z_j|T_i=t,T_j>t)+\frac{1}{2}P(Z_i=Z_j|T_i=t,T_j>t) \quad (6)$$

Using Bayes formula, $AUC^{C,D}(t)$ and $AUC^{I,D}(t)$ can then be expressed as functions of *t*, *p*, λ_0 and θ as follows:

$$AUC^{C,D}(t)=\frac{1}{2}\left[1+\frac{p(1-p)(e^{-\lambda_0 t}-e^{-\lambda_0\theta t})}{\left\{1-\left((1-p)e^{-\lambda_0 t}+pe^{-\lambda_0\theta t}\right)\right\}\times\left\{(1-p)e^{-\lambda_0 t}+pe^{-\lambda_0\theta t}\right\}}\right] \quad (7)$$

$$AUC^{I,D}(t)=\frac{1}{2}\left[1+\frac{p(1-p)(\theta-1)e^{-\lambda_0 t(1+\theta)t}}{\left\{(1-p)e^{-\lambda_0 t}+p\theta e^{-\lambda_0\theta t}\right\}\times\left\{(1-p)e^{-\lambda_0 t}+pe^{-\lambda_0\theta t}\right\}}\right] \quad (8)$$

According to the hazard ratio θ and the prevalence *p* of the marker, various time profiles of $AUC^{C,D}(t)$ and $AUC^{I,D}(t)$ can be found, as illustrated in Figure **2**, and this despite the hazard ratio associated with the marker being constant. Indeed, $AUC^{C,D}(t)$ and $AUC^{I,D}(t)$are constant only when the biomarker has no prognostic value (θ=1), with a value of 0.5, as expected.

When *t* tends to ∞, both $AUC^{C,D}(t)$ and $AUC^{I,D}(t)$tend to a finite value independently of the hazard ratio θ. This limit is related to the prevalence of the biomarker for $AUC^{C,D}(t)$and is always equal to 0.5 for $AUC^{I,D}(t)$ (see 7.1 for demonstration).

$$\lim_{t\to\infty} AUC^{C,D}(t)=0.5+0.5p$$

$$\lim_{t\to\infty} AUC^{I,D}(t)=0.5$$

Moreover, the higher the hazard ratio, the sooner the limit is reached. When *t*=0, $AUC^{C,D}(0)$ is

Table 2: Primary Biliary Cirrhosis Study: Estimated Hazards Ratio (HR) of Death with Ninety-Five Percent Confidence Intervals (95%CI) for Potential Predictors (Age was Standardized)

Variable	HR	95%CI	p-value	Prevalence
Ascites	6.63	(4.20-10.46)	<0.001	8%
Age	1.27	(1.10-1.48)	0.002	-

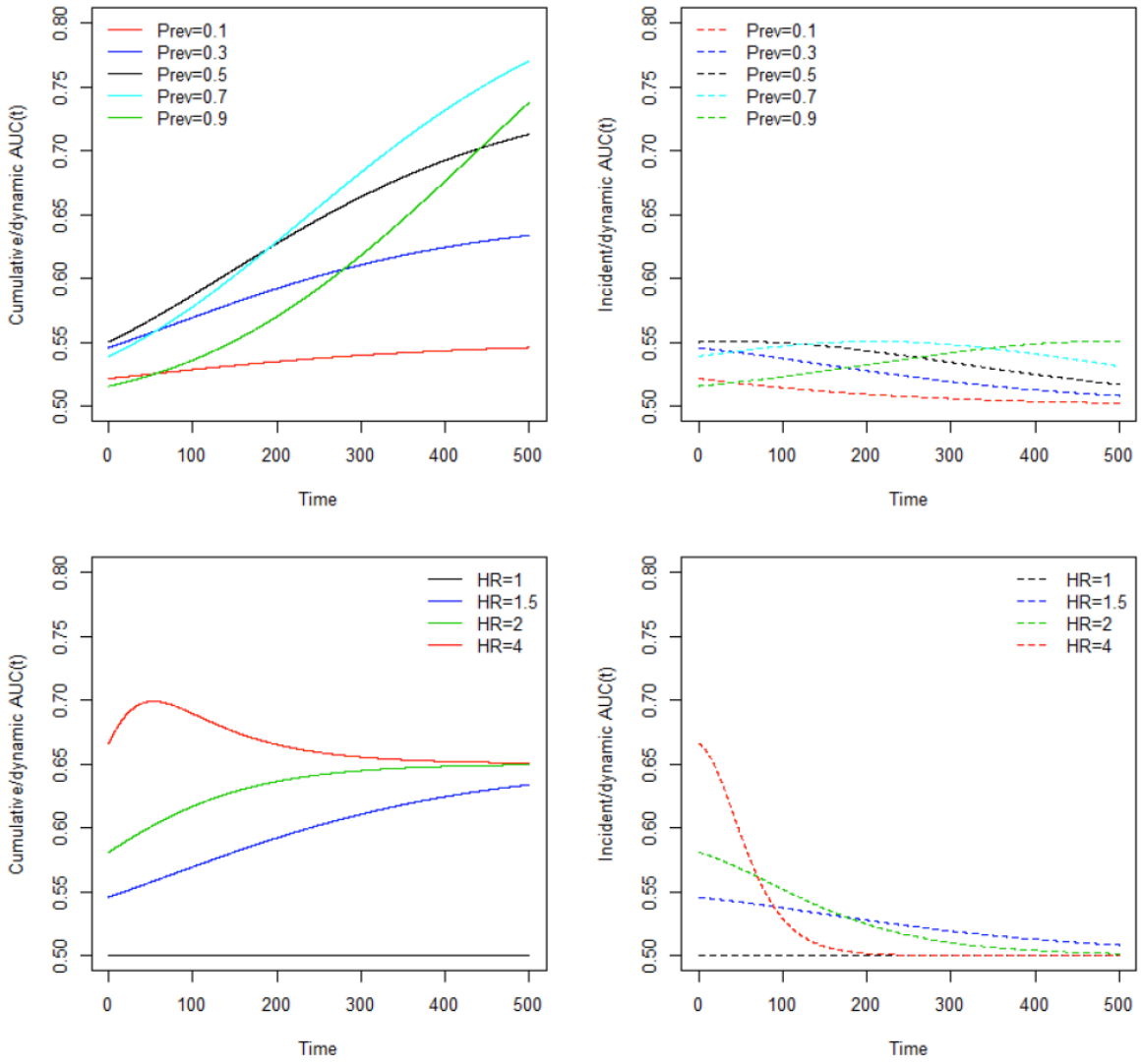

Figure 2: Profiles of $AUC^{C,D}(t)$ (plain lines) and $AUC^{I,D}(t)$ (dashed lines) for varying *p* with fixed β=log(1.5) and $\lambda_0 = 0.01$ (upper panel) and varying β with fixed *p*=0.3 and $\lambda_0 = 0.01$ (lower panel).

indeterminate while $AUC^{I,D}(t)$ is related to both the hazard ratio and the prevalence, as follows:

$$AUC^{C,D}(0) = 0.5 + \frac{1}{2}\frac{P(1-P)(\theta-1)}{(1-P)+P\theta}$$

$AUC^{C,D}(t)$ appears to either monotonically increases if $1<\theta\leq 2$ or increases up to a maximum and then decreases if $\theta>2$, but $AUC^{I,D}(t)$ always has a maximum (see 7.2 for demonstration in the specific case where p=0.5).

Finally, except in the case where the biomarker has no prognostic value (where $AUC^{I,D}(t) = 0.5 \forall t), AUC^{C,D}(t)$ is always higher than $AUC^{I,D}(t)$ which is a general result independent of the distribution of the biomarker and the failure times (see 7.3 for demonstration).

During the first 24 months of the alfa-study, an increasing profile of $AUC^{C,D}(t)$ is expected, whereas, $AUC^{C,D}(t)$ for NPM should be rather constant. Discrepancies between expected and observed time-profiles could be due to an insufficient sample size and/or to the polynomial smoothing. Observed $AUC^{I,D}(t)$ are closer to their expected profiles (Figure **3**).

5. TIME-DEPENDENT AUC FOR A CONTINOUS BIOMARKER

The profiles of $AUC^{C,D}(t)$ and $AUC^{I,D}(t)$ in more complex situations when the marker *Z* is continuous were investigated using numerical integration. Given the marginal distribution of marker and survival distribution of time conditional on the marker, theoretical values OF $AUC^{C,D}(t)$ and $AUC^{I,D}(t)$ can be obtained by:

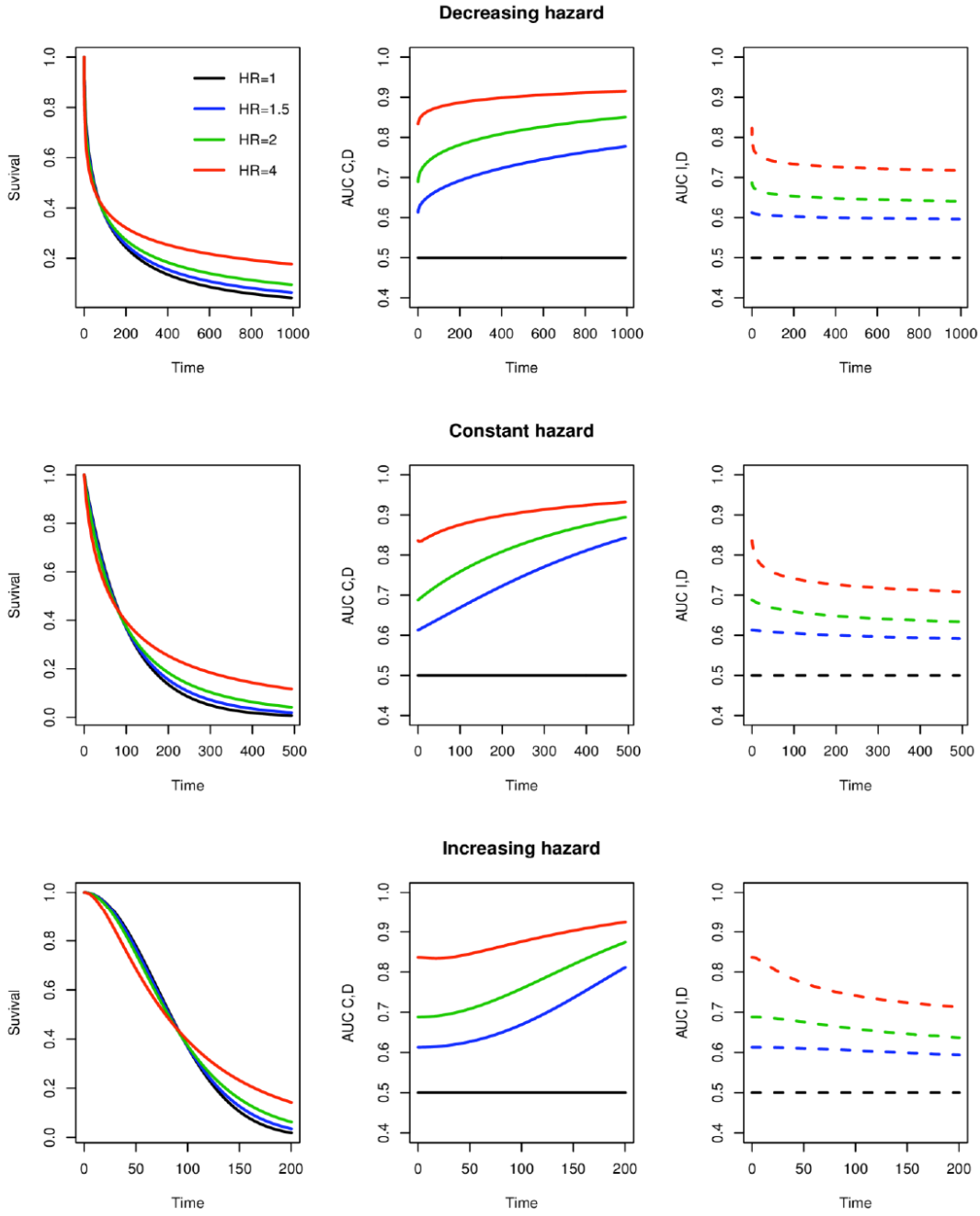

Figure 3: Time profiles of $AUC^{C,D}(t)$ (plain line) and $AUC^{I,D}(t)$(dashed line) for different distributions of the failure times and with hazards ratio of 1 (black line), 1.5 (blue line), 2(green line) and 4 (red line). Left column displays marginal survival functions for the 4 hazard ratios, in the 3 cases of Weibull failure times with decreasing (upper plots), constant (middle plots) and increasing (lower plots) hazards. Central and right columns of the figure display the resulting cumulative/dynamic and incident/dynamic AUCs

$$AUC^{C,D}(t) = \int_{-\infty}^{\infty}\int_{C}^{\infty} \frac{F(t;Z=z)[1-F(t,Z=c)]}{[1-F(t)]F(t)} g(z)g(c)dzdc \quad (9)$$

$$AUC^{I,D}(t) = \int_{-\infty}^{\infty}\int_{C}^{\infty} \frac{f(t;Z=z)[1-F(t,Z=c)]}{[1-F(t)]f(t)} g(z)g(c)dzdc \quad (10)$$

Failure times were generated using the Weibull distribution with a decreasing hazard (shape=0.5), constant hazard (shape=1, i.e. the exponential distribution) or increasing hazard (shape=2) and data were not censored. Marker Z was normally distributed, with mean 0 and variance 1 ($Z \sim \mathcal{N}(0,1)$). As illustrated for a binary biomarker, we assessed no (HR=1), mild (HR=1.5), moderate (HR=2) and high (HR=4) prognostic value of the marker. $AUC^{C,D}(t)$ and $AUC^{I,D}(t)$ were computed by numerical integration of formula (9) and (10).

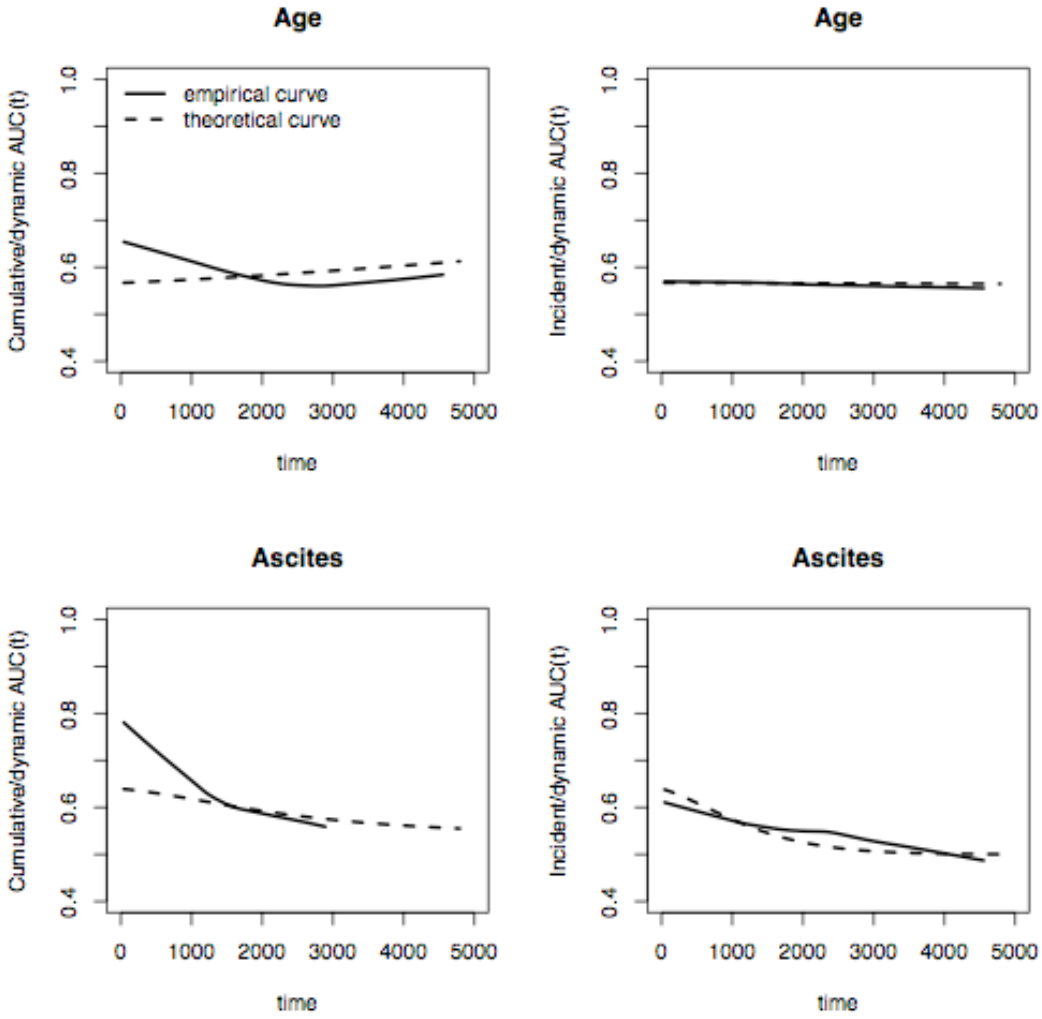

Figure 4: PBC study. $AUC^{C,D}(t)$ (left panel) and $AUC^{I,D}(t)$ (right panel) time profiles for age (upper panel) and ascites (lower panel). Plain lines are empirical curves smoothed by a locally weighted polynomial regression, and dashed lines are theoretical profiles from corresponding Weibull conditional distributions.

Figure **4** illustrates that, whatever the distribution of survival times, the $AUC^{C,D}(t)$ monotonically increased with time, contrary to the $AUC^{I,D}(t)$ which monotonically decreased, though in a somewhat narrower range. As previously demonstrated for binary biomarkers, the values of the $AUC^{C,D}(t)$ are greater than the values of the $AUC^{I,D}(t)$. In contrast to the binary setting, clear-cut crossing of the $AUC^{I,D}(t)$ was no longer observed, and the higher the HR, the higher were the values of both $AUC^{C,D}(t)$ and $AUC^{I,D}(t)$. Nevertheless, choosing two different time points for two biomarkers of distinct prognostic values may achieve a difference in the AUC(t) favoring the biomarker with the lowest prognostic value. For instance, in the case of a Weibull failure times distribution with increasing hazard, the value of $AUC^{C,D}(t)$ was 0.701 at time 40 for a biomarker with mild prognostic value (HR=2) and 0.752 at time 160 for a biomarker with moderate prognostic value (HR=1.5).

6. DISCUSSION

Discrimination measures for survival outcomes have led to several extensions of the classically defined AUC, widely used in the diagnostic setting. Three time-dependent measures have been proposed by Heagerty and al. [5] according to how cases (events) an controls (non-events) are defined: Cumulative/Dynamic ($AUC^{C,D}(t)$), Incident/Dynamic ($AUC^{I,D}(t)$) and Incident/Static $AUC^{I,S}(t)$. The first two definitions have been more extensively studied and are increasingly used in the medical literature. Nevertheless, no clear guidance has emerged in the statistical literature regarding the choice of these definitions for cases and controls [22]. Arguments favoring $AUC^{C,D}(t)$ are that cumulative/dynamic definitions of cases and controls, respectively, are more natural and intuitive, and thus can be seen as direct extensions of those in the diagnostic setting. Notably, they may appear more

appropriate for clinical decision-making, such as enrolment in clinical trials. On the other hand, by using $AUC^{I,D}(t)$, the incident true positive rate and the dynamic false positive rate parallel the multiple contributions that a subject can make to the likelihood function. Moreover, $AUC^{I,D}(t)$ can be averaged over the entire the follow-up period to obtain a time-independent summary measure that is directly related to a global concordance measure, whereas no such meaningful averaging have been proposed for $AUC^{C,D}(t)$.

However, the interpretation of trends of AUC values over time has never been studied in a theoretical framework, particularly in the case of a fixed association where one could imagine that discrimination would be constant across time. In this paper, using analytical expression of both $AUC^{C,D}(t)$ and $AUC^{I,D}(t)$ in the simple case of a binary marker and exponential survival times or numerical integration for more complex cases with a continuous marker, we highlighted several interesting properties. First, at a given timepoint, the value of $AUC^{C,D}(t)$ of a biomarker is alwas higher than that of the $AUC^{I,D}(t)$. Second, we showed that in case of a continuous biomarker, $AUC^{C,D}(t)$ mostly increased while $AUC^{I,D}(t)$ decreased over time. Given these potential differences, the definition of the chosen AUC(t) should be clearly reported when applied in the medical literature, and the choice of the time point of interest should be carefully justified.

Third, in the context of binary time-fixed outcomes, the discriminatory measure has been directly related to the measure of association, and a recent study has shown that the relation between the AUC and the odds ratio depends on the distribution of the explanatory variable [23]. Similarly, in the survival setting, using a binary biomarker and exponential failure times, we showed that the $AUC^{C,D}(t)$ and $AUC^{I,D}(t)$ depended not only on the effect of the biomarker on the hazard of death but also on its prevalence. This explains at least partially the results observed in the ALFA-0701 study.

Finally, it should be underlined that both empirical $AUC^{C,D}(t)$ and $AUC^{I,D}(t)$ can be quite different from their model-based theoretical values. This is illustrated bye time-profile of discriminative values of two predictors in the PBC study, namely age and presence of ascites (Figure **4**). While empirical cumulative/dynamic AUC(t) of age exhibit a decreasing then increasing profile, the profile of the corresponding Weibull model is increasing. These discrepancies seem milder for incident/dynamic AUC(t). Such differences could be due to the fact that smoothing does not perform very well, and that early and late estimation of AUC(t), corresponding to few event, are unreliable. Similarly to what has been stated in the context of proportion or in the survival setting, model-based measures might be preferable to empirical non parametric estimations [24,25].

Thus, based on these previous points, we warned against the use of discrimination measures such as AUC(t) when assessing and comparing the predictive value of biomarkers. This warning is in agreement with previous reports in the context of logistic regression, where testing for improvement in AUC has been shown to be equivalent and less powerful than testing whether the new predictor variable is significantly different from zero in multivariable regression models [26,27,28]. Otherwise, as an illustration of the discriminatory performance of a biomarker using a single measure, the use of time-independent discriminatory measures, such as the global concordance index proposed by Harrell and modified by Uno [29,30] the concordance probability estimate proposed by Gonen [31], or the integration of time-dependent AUC [5,32,33] may be interesting alternatives.

FUNDING

This work was supported by the Institut national de la santé et de la recherche médicale (Inserm).

R code is available upon request.

7. APPENDIX

7.1. Limits when *t* Tends to ∞

Equation (7) can be rewritten as:

$$AUC^{C,D}(t) = \frac{1}{2}\left\{1 + p(1-p)\frac{e^{-\lambda ambda\ 0} - e^{-\theta l_0 t}}{(1-v(t)) \times v(t)}\right\}$$

with

$$v(t) = (1-p)^{-\lambda_0 t} + p^{-\theta\lambda_0 t} \quad (2)$$

when t tends to ∞ the limit of $(1–v(t))$ is equal to 1 and

$$\frac{e^{-l_0} - e^{-l_0 t}}{v(t)} = \frac{1 - e^{-(\theta-1)l_0 t}}{1 + p(e^{-(\theta-1)l_0 t} - 1)}$$

with the limits of the numerator and denominator being equal to 1 and $(1–p)$, respectively. Thus, $\lim_{t\to\infty} AUC^{C,D}(t) = 0.5(1+p)$

Equation (8) can be rewritten as:

$$AUC^{C,D}(t) = 0.5 + \frac{p(1-p)(\theta-1)l_0}{2}F(t)$$

with

$$F(t) = \frac{e^{(1-\theta)l_0 t}}{p(1-p)l_0 e^{-(\theta-1)l_0 t} + p^2\theta l_0 e^{-2(\theta-1)l_0 t} + (1-p)^2 l_0 + p(1-p)\theta l_0 e^{-(\theta-1)l_0 t}}$$

Because θ>1, when *t* tends to ∞ the limits of the numerator and denominator of *F* are equal to 0 and $(1-p)^2\lambda_0$, respectively. Thus, $\lim_{t\to\infty} AUC^{I,D}(t) = 0.5$

7.2. Monotonicity in the Case where *p*=1–*p*=0.5

7.2.1. AUC^{C,D}(t)

Let $f(t) = {}^{-\lambda_0 t} - {}^{-\theta\lambda_0 t}$ and $g(t) = {}^{-\lambda_0 t} + {}^{-\theta\lambda_0 t}$

$$AUC^{C,D}(t) = \frac{1}{2}\left[1 + \frac{f(t)}{g(t)\{2-g(t)\}}\right]$$

$$\frac{dAUC^{C,D}(t)}{dt} = \frac{1}{2}\frac{f'(t)g(t)(2-g(t)) - f(t)g'(t)(2-2g(t))}{\{g(t)(2-g(t))\}^2}$$

$$= \frac{1}{2}\frac{D(t)}{\{g(t)(2-g(t))\}^2}$$

with

$$D(t) = 4(\theta-1)\lambda_0{}^{-(\theta+1)\lambda_0 t} + (2\lambda_0 - 3\theta\lambda_0)^{-(2+\theta)\lambda_0 t} + (3-2\theta)\lambda_0{}^{-(1+2\theta)\lambda_0 t} - \lambda_0{}^{-3\lambda_0 t} + \theta\lambda_0{}^{-3\theta\lambda_0 t}$$

When *t*=0, $D(t) = 2\lambda_0 > 0$

When $t\to\infty$, *D*(*t*) is the same sign as $4(\theta-1)\lambda_0 e^{-\lambda_0(1+\theta)t} - \lambda_0{}^{-3\lambda_0 t}$. It is thus negative for θ+1>3 i.e. θ>2 and positive for θ≤2

7.2.2. AUC^{I,D}(t)

Let $f(t) = {}^{-(\theta+1)\lambda_0 t}$ and $g = (\lambda_0{}^{-\lambda_0 t} + \theta\lambda_0{}^{-\theta\lambda_0 t})({}^{-\lambda_0 t} + {}^{-\theta\lambda_0 t})$

$$AUC^{I,D}(t) = \frac{1}{2}\left\{1 + \frac{(\theta-1)l_0 f(t)}{g(t)}\right\}$$

$$\frac{dAUC^{I,D}(t)}{dt} = \frac{(\theta-1)l_0}{2}\frac{f'(t)g(t) - f((t)g'(t)}{g^2(t)} = \frac{(\theta-1)l_0}{2}\frac{D(t)}{g^2(t)}$$

with

$$D(t) = \theta\lambda_0(\theta-1)\lambda_0{}^{-(1+3\theta)\lambda_0 t} + \lambda_0(1-\theta)\lambda_0{}^{-(3+\theta)\lambda_0 t}$$

When *t*=0 $D(t) = \{(\theta-1)\lambda_0\}^2 > 0$ When $t\to\infty$, *D*(*t*) is the same sign as $\lambda_0(1-\theta)\lambda_0{}^{-(3+\theta)\lambda_0 t}$ which is <0

7.3. Relation between AUC^{C,D}(t) and AUC^{I,D}(t)

First, AUC^{C,D}(t) and AUC^{I,D}(t) can be written as a function of the corresponding definitions of time-dependent True Positive Rate (*TPR*(*t*)) and False Positive Rate (*FPR*(*t*)) Cumulative TPR is defined as $TPR^C(c,t) = \Pr(X > c | T \le t)$, Incident TPR is defined as $TPR^I(c,t) = \Pr(X > c | T = t)$ and Dynamic FPR is defined as $FPR^D c,(t) = \Pr(X > c | T > t)$.

$$AUC^{C,D}(t) = \int_{-\infty}^{\infty} TPR^C(c,t)d[FPR^D(c,t)]$$

$$AUC^{I,D}(t) = \int_{-\infty}^{\infty} TPR^I(c,t)d[FPR^D(c,t)]$$

Second, TPR^C(c,t) can be expressed knowing TPR^I(c,t) and the distribution of failure times *f*(*t*):

$$TPR^C(c,t) = \frac{\int_0^t TPR^I(c,u)f(u)du}{\int_0^t f(u)du}$$

Assuming that *AUC*>0.5, if $t_i < t_j$ we can write that $Pr(X > c \,|\, T = t_i) \ge Pr(x > c \,|\, T = t_j)$

Hence, on [0,t]:

$$TPR^C(c,u) \ge TPR^I(c,u)$$

$$\int_0^t TPR^C(c,u)f(u)du \ge \int_0^t TPR^I(c,u)f(u)du$$

$$\frac{\int_0^t TPR^I(c,u)f(u)du}{\int_0^t f(u)du} \ge TPR^I(c,t)$$

$$TPR^C(c,t) \ge TPR^I(c,t)$$

Thus, AUC^{C,D}(t) ≥ AUC^{I,D}(t)

7.4. Weibull Failure Times and Binary Biomarker

As for exponential failure times, $AUC^{C,D}(t)$ and $AUC^{I,D}(t)$ can be expressed as a function of *p*, λ_0, β and *t* as follows, with $\lambda_1 = \lambda_0 exp(\beta)$:

$$AUC^{C,D}(t) = \frac{1}{2}\left[1 + \frac{p(1-p)(e^{-(l_0 t)^a} - e^{-(l_1 t)^a})}{\{p(1-e^{-(l_1 t)^a}) + (1-p)(1-e^{-(l_0 t)^a})\} \times \{(1-p)e^{-(l_0 t)^a} + pe^{-(l_1 t)^a}\}}\right] \quad \text{(A.1)}$$

$$AUC^{I,D}(t) = \frac{1}{2}\left[1 + \frac{p(1-p)(e^{-(l_1 t)^a} e^{-(l_0 t)^a})}{\{pl_1^a e^{-(l_1 t)^a} + (1-p)l_0^a e^{-(l_0 t)^a}\} \times \{(1-p)e^{-(l_0 t)^a} + pe^{-(l_1 t)^a}\}}\right] \quad \text{(A.2)}$$

REFERENCES

[1] Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, Pencina MJ, Kattan MW. Assessing the performance of prediction models: a framework for some traditional and novel measures. Epidemiology 2010; 21(1): 128. http://dx.doi.org/10.1097/EDE.0b013e3181c30fb2

[2] Schemper M, Henderson R. Predictive accuracy and explained variation in Cox regression. Biometrics 2000; 56(1): 249-55. http://dx.doi.org/10.1111/j.0006-341X.2000.00249.x

[3] O'Quigley J, Xu R. Explained variation in proportional hazards regression. In: Crowley J, Ed. Handbook of Statistics in Clinical Oncology, Marcel Dekker, New York 2001; pp. 397-409.

[4] Pencina MJ, D'Agostino RB. Overall C as a measure of discrimination in survival analysis: model specific population value and confidence interval estimation. Statistics in Medicine 2004; 23(13): 2109-23. http://dx.doi.org/10.1002/sim.1802

[5] Harrell FE, Lee KL, Mark DB. Tutorial in biostatistics multivariable prognostic models: issues in developingmodels, evaluating assumptions and adequacy, and measuring and reducing errors. Statistics in Medicine 1996; 15: 361-387. http://dx.doi.org/10.1002/(SICI)1097-0258(19960229)15:4<361::AID-SIM168>3.0.CO;2-4

[6] Heagerty PJ, Zheng Y. Survival model predictive accuracy and ROC curves. Biometrics 2005; 61(1): 92-105. http://dx.doi.org/10.1111/j.0006-341X.2005.030814.x

[7] Maisel A, Mueller C, Neath SX, Christenson RH, Morgenthaler NG, McCord J, Nowak RM, Vilke G, Daniels LB, Hollander JE, Apple FS, Cannon C, Nagurney JT, Schreiber D, Defilippi C, Hogan C, Diercks DB, Stein JC, Headden G, Limkakeng ATJ, Anand I, Wu AH, Papassotiriou, J, Hartmann O, Ebmeyer S, Clopton P, Jaffe AS, Peacock WF. Copeptin Helps in the Early Detection of Patients With Acute Myocardial Infarction Primary Results of the CHOPIN Trial (Copeptin Helps in the early detection Of Patients with acute myocardial INfarction). J Am CollCardiol 2013; 62(2): 150-60. http://dx.doi.org/10.1016/j.jacc.2013.04.011

[8] Aalen O. Nonparametric inference for a family of counting process. The Annals of Statistics 1978; 6(4): 701-726. http://dx.doi.org/10.1214/aos/1176344247

[9] Therneau T, Grambsch P. Modeling Survival Data: Extending the Cox Model. Springer 2000. http://dx.doi.org/10.1007/978-1-4757-3294-8

[10] Chambless LE, Diao G. Estimation of time-dependent area under the roc curve for long-term risk prediction. Statistics in Medicine 2006; 25(20): 3474-3486. http://dx.doi.org/10.1002/sim.2299

[11] Heagerty PJ, Lumley T, Pepe MS. Time-dependent roc curves for censored survival data and a diagnosticmarker. Biometrics 2000; 56(2): 337-344. http://dx.doi.org/10.1111/j.0006-341X.2000.00337.x

[12] Hung H, Chiang C. Estimation methods for time-dependent auc models with survival data. The Canadianmarker. Biometrics 2000; 56(2): 337-344.

[13] Uno H, Cai T, Tian L, Wei LJ. Evaluating prediction rules for t-year survivors with censored regression models. Journal of the American Statistical Association 2007; 102(478). http://dx.doi.org/10.1198/016214507000000149

[14] Song X, Zhou X, Ma S. Nonparametric receiver operating characteristic-based evaluation for survival outcomes. Statistics in Medicine 2012; 31(23): 2660-2675. http://dx.doi.org/10.1002/sim.5386

[15] R Development Core Team: R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria 2012. R Foundation for Statistical Computing. ISBN 3-900051-07-0. http://www.R-project.org/

[16] Blanche P. timeROC: Time-dependent ROC Curve and AUC for Censored Survival Data 2012. R package version 0.1. http://CRAN.R-project.org/package=timeROC

[17] Heagerty PJ. Packaging by ParamitaSaha-Chaudhuri: risksetROC: Riskset ROC Curve Estimation from Censored Survival Data 2012. R package version 1.0.4. http://CRAN.R-project.org/package=risksetROC

[18] Heagerty PJ. Packaging by ParamitaSaha-Chaudhuri: survivalROC: Time-dependent ROC Curve Estimation from Censored Survival Data 2013. R package version 1.0.3. http://CRAN.R-project.org/package=survivalROC

[19] Potapov S, Adler W, Schmid M. survAUC: Estimators of Prediction Accuracy for Time-to-event Data 2011. R package version 1.0-4. http://CRAN.R-project.org/package=survAUC

[20] Castaigne S, Pautas C, Terré C, Raffoux E, Bordessoule D, Bastie J, Legrand O, Thomas X, Turlure P, Reman O, de Revel T, Gastaud L, de Dunzburg N, Contentin N, Henry E, Marolleau J, Aljijakli A, Rousselot P, Fenaux P, Preudhomme P, Chevret S, Dombret H. Effect of gemtuzumabozogamicin on survival of adult patients with de-novo acute myeloidleukaemia (alfa-0701): arandomised, open-label, phase 3 study. The Lancet 2012; 379(9825): 1508-1516. http://dx.doi.org/10.1016/S0140-6736(12)60485-1

[21] Grambsch PM, Therneau TM. Proportional hazards tests and diagnostics based on weighted residuals. Biometrika 1994; 81(3): 515-526. http://dx.doi.org/10.1093/biomet/81.3.515

[22] Blanche P, Latouche A, Viallon V. Time-dependent auc with right-censored data: a survey study. arXivpreprint arXiv:1210.6805 2012.

[23] Austin PC, Steyerberg EW. Interpreting the concordance statistic of a logistic regression model: relation tothe variance and odds ratio of a continuous explanatory variable. BMC Medical Research Methodology 2012; 12(1): 82. http://dx.doi.org/10.1186/1471-2288-12-82

[24] Agresti A. An introduction to categorical data analysis Wiley 1996; Vol. 135.

[25] May M, Sterne J, Egger M. Parametric survival models may be more accurate than Kaplan-Meier estimates. BMJ 2003; 326(7393): 822. http://dx.doi.org/10.1136/bmj.326.7393.822

[26] Vickers AJ, Cronin AM, Begg CB. One statistical test is sufficient for assessing new predictive markers. BMC Medical Research Methodology 2011; 11(1): 13. http://dx.doi.org/10.1186/1471-2288-11-13

[27] Pepe MS, Kerr KF, Longton G, Wang Z. Testing for improvement in prediction model performance. Statistics in medicine 2013; 32(9): 1467-1482. http://dx.doi.org/10.1002/sim.5727

[28] Austin PC, Steyerberg EW. Predictive accuracy of risk factors and markers: a simulation study of the effect ofnovel markers on different performance measures for logistic regression models. Statistics in Medicine 2013; 32(4): 661-672. http://dx.doi.org/10.1002/sim.5598

[29] Harrell FE, Lee KL, Mark DB. Tutorial in biostatistics multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Statistics in Medicine 1996; 15: 361-387. http://dx.doi.org/10.1002/(SICI)1097-0258(19960229)15:4<361::AID-SIM168>3.0.CO;2-4

[30] Uno H, Cai T, Pencina MJ, D'Agostino RB, Wei LJ. On the c-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. Statistics in Medicine 2011; 30(10): 1105-1117.

[31] Gönen M, Heller G. Concordance probability and discriminatory power in proportional hazards regression. Biometrika 2005; 92(4): 965-970. http://dx.doi.org/10.1093/biomet/92.4.965

[32] Li J, Ma S. Time-dependent roc analysis under diverse censoring patterns. Statistics in Medicine 2011; 30(11): 1266-1277.

[33] Lambert J, Chevret S. Summary measure of discrimination in survival models based on cumulative/dynamic time-dependent ROC curves. Stat Methods Med Res 2014; [Epub ahead of print]. http://dx.doi.org/10.1177/0962280213515571

Received on 18-11-2014 Accepted on 15-12-2014 Published on 27-01-2015

http://dx.doi.org/10.6000/1929-6029.2015.04.01.12