SORT 41 (1) January-June 2017, 73-92

Comparison of two discrimination indexes in the categorisation of continuous predictors in time-to-event studies

Irantzu Barrio*[,1,5], María Xosé Rodríguez-Álvarez[2,6,7], Luis Meira-Machado[3]
Cristobal Esteban[4,5] and Inmaculada Arostegui[1,5,6]

Abstract

The Cox proportional hazards model is the most widely used survival prediction model for analysing time-to-event data. To measure the discrimination ability of a survival model the concordance probability index is widely used. In this work we studied and compared the performance of two different estimators of the concordance probability when a continuous predictor variable is categorised in a Cox proportional hazards regression model. In particular, we compared the c-index and the concordance probability estimator. We evaluated the empirical performance of both estimators through simulations. To categorise the predictor variable we propose a methodology which considers the maximal discrimination attained for the categorical variable. We applied this methodology to a cohort of patients with chronic obstructive pulmonary disease, in particular, we categorised the predictor variable forced expiratory volume in one second in percentage.

MSC: 62N01, 62P10.

Keywords: Categorisation, prediction models, cutpoint, Cox model.

* Corresponding author: E-mail: irantzu.barrio@ehu.eus, Tel.: +34-946012504.
Address: Departamento de Matemática Aplicada, Estadística e Investigación Operativa. Facultad de Ciencia y Tecnología. Universidad del País Vasco UPV/EHU. Barrio Sarriena s/n. 48940 Leioa.

[1] Departamento de Matemática Aplicada, Estadística e Investigación Operativa. Universidad del País Vasco UPV/EHU.

[2] Departamento de Estadística e Investigación Operativa and Biomedical Research Centre (CINBIO). Universidade de Vigo.

[3] Centre of Mathematics and Department of Mathematics and Applications. University of Minho, Portugal.

[4] Servicio de Neumología, Hospital Galdakao-Usansolo.

[5] Red de Investigación en Servicios de Salud en Enfermedades Crónicas (REDISSEC).

[6] BCAM-Basque Center for Applied Mathematics.

[7] IKERBASQUE, Basque Foundation for Science, Bilbao, Spain.
Received: May 2016
Accepted: December 2016

1. Introduction

In the medical field, prediction models have been gaining importance as a support for decision-making, whereby the increased knowledge of potential predictors helps the decision-making process. When the interest relies on predicting the survival time of patients with a certain disease, survival prediction models are commonly used. Particularly, the Cox model (Cox and Oakes, 1984) is the one most widely used in the medical field (Steyerberg et al., 2013).

The development of prediction models may require assumptions about the relationship between the covariates and the response variable. For instance, a common practice in medical research is to categorise continuous predictor variables when a linear relationship does not hold (Turner et al., 2010; Barrio et al., 2016).

The selection of an optimal cutpoint for prognosis purposes has been largely discussed in the literature. For instance, Faraggi and Simon (1996) proposed a cross validation approach to select the cutpoint to classify patients into two risk groups based on the minimisation of the significance level of the logrank test proposed by Lausen and Schumacher (1996). Later, Sima and Gönen (2013) proposed the maximisation of the concordance probability (Gönen and Heller, 2005) as the criterion to dichotomise a continuous predictor. In addition, Liu and Jin (2015) and Rota et al. (2015) have recently proposed non parametric methods to select a time-dependent optimal cutpoint to classify individuals as diseased or disease-free at a given time point t.

However, the aim of this work differs from those presented above. Our goal is to categorise the predictor variable into any possible number of categories to be incorporated in a prediction model, whereas when looking for a unique cutpoint the goal is to classify a patient as diseased or disease-free at a certain time point. This work was motivated in the context of the Stable-COPD study (Esteban et al., 2014) where a model was developed to predict five-year survival in patients with a stable chronic obstructive pulmonary disease (COPD). Clinical researchers aimed to use a categorised version of the predictor variable forced expiratory volume in one second in percentage ($FEV_{1\%}$) in a multiple survival model, but there was no agreement regarding the selection of the optimal cutpoints.

In the context where the outcome of interest takes only two possible values, a proposal has been done to categorise a continuous predictor variable in a logistic regression model by maximising the area under the receiver operating characteristic (ROC) curve (AUC)(Barrio et al., 2016). In fact, the proposal consists of the categorisation of a continuous predictor variable such that the discriminative ability of the prediction model for the categorised variable is maximised. In the context of survival regression models, as far as we know, no proposal has been done to categorise continuous predictor variables. In this paper, we propose to categorise a continuous predictor variable in a Cox proportional hazards regression model as an extension of the work proposed by Barrio et al. (2016) and based in part on the work done for a single cutpoint by Sima and Gönen (2013). However, the challenge is how to measure the discriminative ability of

a survival model. Established concepts for binary outcomes have been commonly used by researchers, yet a standard approach has not emerged (Pepe et al., 2008; Schmid and Potapov, 2012). A commonly used parameter is the concordance probability, a global measure which has been defined differently in the literature (Liu and Jin, 2015). In this paper we studied and compared the performance of two different discrimination ability estimators named the c-index (Harrell et al., 1982) and the concordance probability estimator (CPE, Gönen and Heller (2005)) as the parameters to maximise in the categorisation process. Therefore, the goal of this paper is to compare the performance of the CPE and c-index as concordance probability estimators to maximise in the location of optimal cutpoints to categorise continuous predictors in a Cox proportional hazards regression model.

The rest of the paper has been organised as follows. Section 2 outlines the method proposed for categorising continuous variables in a Cox proportional hazards regression model. In Section 3, the performance of the proposed methodology is investigated through simulations. Section 4 provides a description of the Stable-COPD study of stable patients with COPD and the application of the proposed methodology to this data set. Finally, the main conclusions of our paper and some practical recommendations are deferred to Section 5.

2. Methods

This section describes the proposed methodology to categorise a continuous predictor variable in a Cox proportional hazards regression model. We begin by introducing the needed notation and background in Section 2.1 and Section 2.2 and we describe the approach to categorise a continuous predictor variable in Section 2.3.

2.1. Notation and preliminaries

Let T be a non-negative random variable representing the time until the event of interest occurs. As usual, we assume that these event times might be subject to univariate right-censoring denoted by C, which we assume to be independent of T. Let $\boldsymbol{Z} = [Z_1 \dots Z_p]^\intercal$ be a set of time invariant predictor variables in which we may be interested in terms of studying their relationship with the survival time T.

The most widely used survival regression model is the semiparametric Cox proportional hazards model (Cox, 1972), where the hazard function for T in a time t given the covariate vector $\boldsymbol{Z}$ is expressed as,

$$h(t|\boldsymbol{Z}) = h_0(t)\exp(\boldsymbol{Z}^\intercal\boldsymbol{\beta}) \tag{1}$$

where $h_0(t)$ is the baseline hazard function and $\boldsymbol{\beta}$ is the regression coefficients vector.

2.2. Discriminative ability of a prediction model

In general, it is common to measure the discriminative ability of a prediction model by the concordance probability (Gönen and Heller, 2005). In a setting where the outcome is time-to-event, the concordance probability is usually defined as (Pencina and D'Agostino, 2004)

$$\mathfrak{C} = P(\widetilde{T}_l > \widetilde{T}_m | T_l > T_m), \tag{2}$$

where l and m denote two independent individuals, T_j is the actual survival time of subject j, and $\widetilde{T}_j$ is the predicted survival time provided by the survival prediction model under evaluation.

Under the Cox proportional hazards regression model (1), Pencina and D'Agostino (2004) showed that the concordance probability for the predicted survival times defined in equation (2) is equivalent to the concordance probability for the predicted probability of survival and thus equivalent to the concordance probability defined in terms of the linear predictor of the Cox proportional hazards model given in (1), i.e.,

$$\mathfrak{C} = P(S(t|\mathbf{Z}_l) > S(t|\mathbf{Z}_m) | T_l > T_m) = P(\eta_l < \eta_m | T_l > T_m), \quad \forall t \tag{3}$$

where $S(t|\mathbf{Z}_j) = P(T_j \geq t|\mathbf{Z}_j)$ and $\eta_j = \mathbf{Z}_j^{\mathrm{T}}\boldsymbol{\beta}$. If the concordance probability takes a value of 0.5 then the resulting model provides non informative predictions whereas models predicting better than chance should result in values of $\mathfrak{C}$ lying in the interval $(0.5, 1)$.

From now on, let us denote as $\{\mathbf{z}_i, y_i, \delta_i\}_{i=1}^N$ a sample of size N, where $\mathbf{z}_i$ represents the observed value of the predictor variables for subject i, y_i represents the observed follow-up time for subject i, being the minimum between the censoring (c_i) and the event (t_i) times, i.e. $y_i = min(t_i, c_i)$, and δ_i represents whether subject i is an event ($\delta_i = 1$) or is censored ($\delta_i = 0$). Thus, $\delta_i = I(t_i \leq c_i)$.

In the presence of right censoring, it is difficult to estimate the concordance probability because a problem arises with the comparison of predicted and observed survival times. Harrell et al. (1982) proposed an estimator for the concordance probability called the c-index which is defined as "the proportion of all pairs of patients for which we could determine the ordering of survival times such that the predictions are concordant". More specifically, Harrell et al. (1982) classified the pairs of individuals as *usable* or *unusable*. A pair of individuals is considered unusable in two different situations. One, when both individuals had the event at the same time and, two, if the following time for the individual without the event was shorter than the time until the event for the individual having the event. Thus, the *c*-index estimator proposed by Harrell et al. (1982) is the proportion of usable individual pairs in which the estimated survival times and the observed survival times are concordant and is computed by forming all pairs of observed data where the individual with the shorter follow-up time is an event. Specifically, the *c*-index estimator proposed by Harrell et al. (1982) for model (1) would have the following expression

$$c = \frac{\sum\sum_{i<j} \{I(y_i < y_j)I(\widehat{\eta}_i > \widehat{\eta}_j)\delta_i + I(y_j < y_i)I(\widehat{\eta}_j > \widehat{\eta}_i)\delta_j\}}{\sum\sum_{i<j} \{I(y_i < y_j)\delta_i + I(y_j < y_i)\delta_j\}}, \tag{4}$$

where $\widehat{\eta}_j = \mathbf{z}_j^{\top}\widehat{\boldsymbol{\beta}}$, being $\widehat{\boldsymbol{\beta}}$ the vector of the estimated regression coefficients of the Cox proportional hazards regression model.

Even though it is widely used in practice, as pointed out by Gönen and Heller (2005), the c-index estimator proposed by Harrell et al. (1982) is biased and the bias increases with the censoring rate. Hence, Gönen and Heller (2005) proposed an alternative estimator called the concordance probability estimator (CPE), which under the proportional hazards assumption is a consistent estimator of the concordance probability. This estimator is defined as

$$CPE = \frac{2}{N(N-1)} \sum\sum_{i<j} \left\{ \frac{I(\widehat{\eta}_i > \widehat{\eta}_j)}{1 + e^{\widehat{\eta}_j - \widehat{\eta}_i}} + \frac{I(\widehat{\eta}_j > \widehat{\eta}_i)}{1 + e^{\widehat{\eta}_i - \widehat{\eta}_j}} \right\}. \tag{5}$$

Although it has been usually overlooked in the literature, we would like to note that the definition of concordance probability given by Gönen and Heller (2005) (see equation (1) in that paper), differs from that defined on equation (3). In fact, the CPE given in (5) represents an estimator of $P(T_l > T_m | \eta_l < \eta_m)$. Hence the c-index and the CPE estimate, in general, different quantities.

Different estimators have been proposed in the literature to estimate the concordance probability (Schmid and Potapov, 2012). In this paper, we focused on the c-index and the CPE for two main reasons. First, Schmid and Potapov (2012) carried out a comparison of different discrimination indexes and none of the estimators proved to be stable in all scenarios. In addition, previous work has been done on the comparison of these two estimators in the selection of an optimal cutpoint in a Cox proportional hazards regression model and we intended to extend the research done by Sima and Gönen (2013) to the categorisation of a continuous predictor variable in a multiple Cox proportional hazards prediction model.

2.3. Selection of optimal cutpoints in Cox proportional regression models

Let X be a continuous predictor variable which we want to categorise in a Cox proportional hazards regression model considering the presence of other p predictors, $Z_1, \ldots, Z_p$. Our proposal is to categorise X in such a way that the best multiple predictive survival model is obtained, considering the maximal concordance probability achieved. Specifically, given k the number of cutpoints set for categorising X in $k+1$ intervals, let us denote X_{cat_k} the corresponding categorised variable taking values from 0 to k, and $\boldsymbol{x}_k = [x_1 \ldots x_k]^{\top}$ the vector of k cutpoints which maximises the discriminative ability of the Cox proportional hazards regression model in equation (6):

$$h(t|Z_1,\ldots,Z_p,X_{cat_k}) = h_0(t)e^{\sum_{r=1}^{p}\alpha_r Z_r+\sum_{q=1}^{k}\beta_q 1_{\{X_{cat_k}=q\}}}. \quad (6)$$

Note that in this expression the linear predictor η is in fact $\sum_{r=1}^{p}\alpha_r Z_r+\sum_{q=1}^{k}\beta_q 1_{\{X_{cat_k}=q\}}$.

To estimate the vector of the cutpoints of X that maximises the concordance probability, we propose to make use of the algorithms *AddFor* and *Genetic* proposed by Barrio et al. (2016). The former looks sequentially for the k cutpoints whereas the later looks for the vector of the optimal cutpoints using genetic algorithms. This implies that the *Genetic* algorithm is computationally more expensive than the *AddFor*. Nevertheless, it has been proven to perform better specially when two cutpoints are looked for (Barrio et al., 2016). For this reason, in this paper we have limited to the use of the *Genetic* algorithm. In addition, the optimal number of cutpoints can be selected by means of a bootstrap confidence interval for the difference of the bias-corrected concordance probability as proposed by Barrio et al. (2016) and extended here to the Cox proportional hazards setting. Detailed information regarding this approach can be seen in the Supplementary Material.

Note that our approach can be easily applied also to the univariate Cox proportional hazards model in which no other predictors $\boldsymbol{Z}$ are present. However, in this case there will be many ties on the linear predictor and hence the expressions given in equations (4) and (5) need to be modified accordingly (see Appendix for further details).

3. Simulation study

In this section we present a simulation study conducted to analyse the empirical performance of the methodology proposed in Section 2. We report here the results obtained in this study and compare the performance of the two concordance probability estimators considered. The simulation study is explained in detail below.

All computations were performed in (64 bit) R 3.2.3 (R Core Team, 2016) and a workstation equipped with 24GB of RAM, an Intel Xeon E5620 processor (2.40 Ghz), and Windows 7 operating system. Specifically, the `genoud` function of the `rgenoud` (Mebane and Sekhon, 2011) package was used to compute the genetic algorithms, the `cph` function of the `rms` package (Harrell, 2015) was used for the estimation of the Cox proportional hazards regression model and the c-index, and the `phcpe2` function of the package `CPE` (Mo et al., 2012) was used to estimate the CPE.

3.1. Scenarios and set-up

To simulate the data we assumed that X is a continuous predictor variable normally distributed with mean $\mu = 0$ and standard deviation $\sigma = 2$ and Z a continuous predictor variable normally distributed with mean $\mu = 1$ and standard deviation $\sigma = 1$. Considering the theoretical optimal cutpoints, $c_1, c_2, \ldots, c_k$, we built a categorical variable, X_{cat_k},

such that $X_{cat_k} = 0$ if $X \leq c_1$, $X_{cat_k} = 1$ if $c_1 < X \leq c_2, \ldots,$ and $X_{cat_k} = k$ if $X > c_k$. Survival times T were generated assuming a Weibull baseline hazard function such that the Cox proportional hazards model is satisfied (Meira-Machado and Faria, 2014). Specifically,

$$T = \left(\frac{-log(1-U)}{\lambda e^{\beta_1 1_{\{X_{cat_k}=1\}} + \ldots + \beta_k 1_{\{X_{cat_k}=k\}} + \alpha Z}} \right)^{1/\gamma}, \tag{7}$$

were U follows a uniform distribution on the interval $(0,1)$, and λ and γ denote the scale and shape parameters of the Weibull distribution respectively. An independent uniform censoring time C was generated, according to the uniform model $U(0,\tau)$, and the event indicator δ was defined as $I(T \leq C)$. The parameter τ was chosen to obtain censoring percentages of about 20%, 50% and 70%. Simulations were performed for sample sizes of $N = 500$ and $N = 1000$. In all cases, $R = 500$ replicates of simulated data were performed.

Several settings were considered in this simulation study, which are summarised in Table 1. First of all, we considered $k = 1, 2$ and 3 as the number of cutpoints. In Scenario I, $k = 1$ was considered with three different alternatives for the theoretical cutpoint a) centred on the distribution of X, i.e., $c_1 = 0$; b) shifted to the high risk area, $c_1 = 1.5$; and c) shifted to the low risk area, $c_1 = -1.5$. In Scenario II we considered two theoretical cutpoints $c_1 = -1$ and $c_2 = 1$. Finally three theoretical cutpoints $c_1 = -1.5, c_2 = 0$ and $c_3 = 1.5$ where considered in Scenario III. In the later scenario, we also considered two different settings, one in which a monotonic increase risk relationship was considered (IIIa) and the other for a non-monotonic risk relationship (IIIb).

The performance of each of the concordance probability estimators considered was evaluated by means of the bias and mean square error (MSE) of the estimated optimal cutpoints for each iteration as follows:

$$MSE_s = \frac{1}{k} \sum_{d=1}^{k} (x_{sd} - c_d)^2$$

where x_{sd} is the estimated d^{th} optimal cutpoint in the simulation s and c_d is the d^{th} theoretical cutpoint.

3.2. Results

Given the large number of proposed scenarios and different conclusions obtained, we begin by summarising the main findings.

Simulation results suggest that, in general, both indexes performed similarly in terms of the mean square error when it comes to low censoring rates (20%). However, for large censoring rates (70%), the c-index performed better than the CPE in all scenarios considered. As could have been expected, in all cases the bias and MSE decrease as the sample size increase.

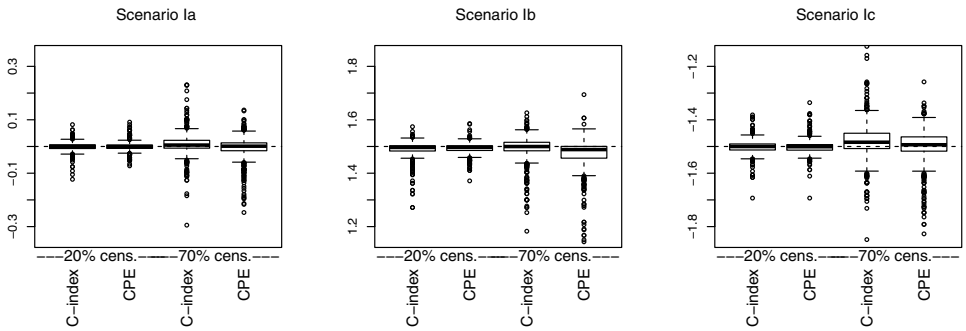


Figure 1: *Boxplot of the estimated optimal cutpoints based on 500 simulated data sets, $N = 500$ sample size and one theoretical cutpoint. Results are shown for censoring rates of* 20% *and* 70% *and c-index and CPE discriminative ability estimators. From left to right: (a) theoretical cutpoint, $c = 0$; (b) theoretical cutpoint, $c = 1.5$; and (c) theoretical cutpoint, $c = -1.5$. The theoretical cutpoint is represented with a dashed line.*

Table 1: *Description of the different scenarios considered for the simulation study. Weibull baseline hazard function with shape γ and scale λ. Uniform censoring $C\,U(0,\tau)$.*

Scenario	**Theoretical cutpoints**	**Parameters**	**Censorship (τ)**		
			20%	**50%**	**70%**
Ia	0	$\gamma = 1, \lambda = 0.1$ $\beta_1 = 2.5, \alpha = 1$	11	1.6	0.5
Ib	1.5	$\gamma = 1, \lambda = 0.1$ $\beta_1 = 2.5, \alpha = 1$	19	3.5	1.15
Ic	-1.5	$\gamma = 1, \lambda = 0.1$ $\beta_1 = 2.5, \alpha = 1$	4.75	0.75	0.27
II	-1 & 1	$\gamma = 1, \lambda = 0.1$ $\beta_1 = 1.5, \beta_2 = 2.5$ $\alpha = 1$	8.5	1.5	0.5
IIIa	-1.5 & 0 & 1.5	$\gamma = 1, \lambda = 0.1$ $\beta_1 = 1.5, \beta_2 = 2.5$ $\beta_3 = 3.5, \alpha = 1$	5.25	0.85	0.27
IIIb	-1.5 & 0 & 1.5	$\gamma = 1, \lambda = 0.1$ $\beta_1 = 1.5, \beta_2 = -1$ $\beta_3 = 1.5, \alpha = 1$	21	3.5	1.15

Let us turn now to a more detailed discussion of the results of this study. Figure 1 depicts the boxplot of the estimated optimal cutpoints over 500 simulated data sets, for the c-index and CPE estimators and a sample size of $N = 500$ and censoring rates of 20% and 70% for Scenarios Ia, Ib and Ic, where a single optimal cutpoint is searched for.

Table 2: *Simulations results when one theoretical optimal cutpoints was sought for (Scenarios Ia, Ib and Ic), censorship of 20%, 50% and 70% and the* Genetic *algorithm. Mean, standard deviation, median, bias and mean MSE for the estimated cutpoints are reported when CPE or c-index concordance probability estimators are used as the maximisation criteria.*

Sample size	**Cens.**	**theoretical cutpoint**	**Cutpoint Estimation**							
			CPE				**c-index**			
			Mean (sd)	**Median**	**Bias**	**MSE**	**Mean (sd)**	**Median**	**Bias**	**MSE**
Scenario Ia										
$N = 500$	20%	0	0.000 (0.018)	0.000	0.000	0.000	−0.001 (0.019)	0.000	−0.001	0.000
	50%	0	−0.001 (0.027)	0.000	−0.001	0.001	0.000 (0.025)	0.001	0.000	0.001
	70%	0	−0.007 (0.047)	0.001	−0.007	0.002	0.008 (0.046)	0.005	0.008	0.002
$N = 1000$	20%	0	−0.002 (0.009)	0.000	−0.002	0.000	−0.001 (0.010)	0.000	−0.001	0.000
	50%	0	−0.003 (0.012)	0.000	−0.003	0.000	−0.001 (0.015)	0.000	−0.001	0.000
	70%	0	−0.006 (0.031)	0.000	−0.006	0.001	0.003 (0.026)	0.003	0.003	0.001
Scenario Ib										
$N = 500$	20%	1.5	1.493 (0.021)	1.497	−0.007	0.000	1.489 (0.032)	1.497	−0.011	0.001
	50%	1.5	1.490 (0.030)	1.495	−0.010	0.001	1.490 (0.037)	1.498	−0.010	0.001
	70%	1.5	1.470 (0.061)	1.488	−0.030	0.005	1.491 (0.051)	1.500	−0.009	0.003
$N = 1000$	20%	1.5	1.498 (0.013)	1.499	−0.002	0.000	1.496 (0.016)	1.499	−0.004	0.000
	50%	1.5	1.496 (0.015)	1.499	−0.004	0.000	1.497 (0.018)	1.500	−0.003	0.000
	70%	1.5	1.483 (0.031)	1.492	−0.017	0.001	1.496 (0.023)	1.500	−0.004	0.001
Scenario Ic										
$N = 500$	20%	−1.5	−1.501 (0.028)	−1.501	−0.001	0.001	−1.502 (0.029)	−1.500	−0.002	0.001
	50%	−1.5	−1.500 (0.042)	−1.498	0.000	0.002	−1.491 (0.053)	−1.494	0.009	0.003
	70%	−1.5	−1.508 (0.096)	−1.494	−0.008	0.009	−1.478 (0.087)	−1.484	0.022	0.008
$N = 1000$	20%	−1.5	−1.499 (0.015)	−1.500	0.001	0.000	−1.500 (0.015)	−1.500	0.000	0.000
	50%	−1.5	−1.498 (0.021)	−1.498	0.002	0.000	−1.496 (0.025)	−1.498	0.004	0.001
	70%	−1.5	−1.498 (0.047)	−1.495	0.002	0.002	−1.488 (0.048)	−1.492	0.012	0.002

Table 3: *Simulations results when two theoretical optimal cutpoints were sought for (Scenario II), censorship of 20%, 50% and 70% and the* Genetic *algorithm. Mean, standard deviation, median, bias and mean MSE for the estimated cutpoints are reported when CPE or c-index concordance probability estimators are used as the maximisation criteria.*

Sample size	**Cens.**	**theoretical cutpoint**	**Cutpoint Estimation**							
			CPE				**c-index**			
			Mean (sd)	**Median**	**Bias**	**MSE**	**Mean (sd)**	**Median**	**Bias**	**MSE**
Scenario II										
N=500	20%	−1	−1.007 (0.049)	−1.002	−0.007	0.006	−1.006 (0.059)	−1.000	−0.006	0.011
		1	0.984 (0.098)	0.992	−0.016		0.979 (0.132)	0.996	−0.021	
	50%	−1	−1.010 (0.096)	−1.000	−0.010	0.017	−1.000 (0.118)	−0.993	0.000	0.028
		1	0.969 (0.155)	0.989	−0.031		0.969 (0.203)	0.999	−0.031	
	70%	−1	−1.200 (0.463)	−1.021	−0.200	0.440	−0.997 (0.269)	−0.983	0.003	0.085
		1	0.677 (0.723)	0.957	−0.323		0.998 (0.313)	1.005	−0.002	
N=1000	20%	−1	−1.002 (0.023)	−1.001	−0.002	0.001	−1.002 (0.026)	−0.999	−0.002	0.002
		1	0.991 (0.040)	0.995	−0.009		0.994 (0.062)	0.997	−0.006	
	50%	−1	−1.003 (0.046)	−0.999	−0.003	0.003	−1.004 (0.054)	−0.997	−0.004	0.004
		1	0.987 (0.059)	0.994	−0.013		0.991 (0.075)	0.998	−0.009	
	70%	−1	−1.030 (0.157)	−0.998	−0.030	0.059	−0.997 (0.104)	−0.993	0.003	0.015
		1	0.928 (0.296)	0.982	−0.072		1.000 (0.138)	1.000	0.000	

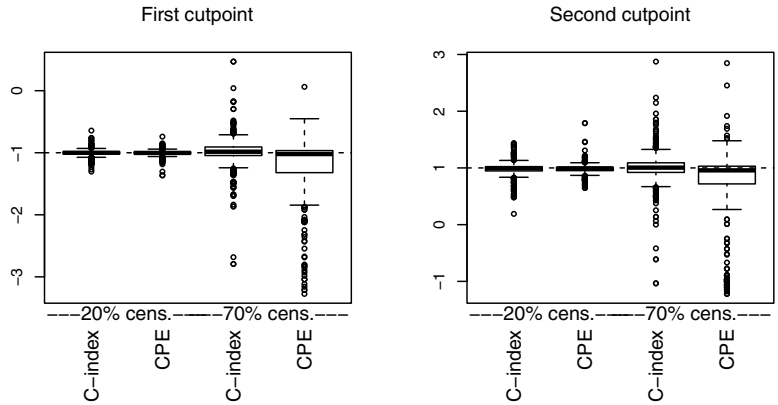


Figure 2: *Boxplot of the estimated optimal cutpoints based on 500 simulated data sets, $N = 500$ sample size and two theoretical cutpoints. Results are shown for censoring rates of* 20% *and* 70% *and c-index and CPE discriminative ability estimators. The theoretical cutpoint is represented with a dashed line.*

Numerical results for these scenarios are given in Table 2. As can be seen, our approach performed satisfactorily regardless of the location of the theoretical cutpoint, with, as said before, the c-index performing slightly better for high censoring rates. However, this can not be considered a general rule. Simulations studies conducted in a univariate setting showed that neither the CPE nor the c-index performed satisfactorily, especially when the optimal cutpoint is non centred. These results are presented and discussed in detail in the Supplementary Material (Table B1 and Figure B1).

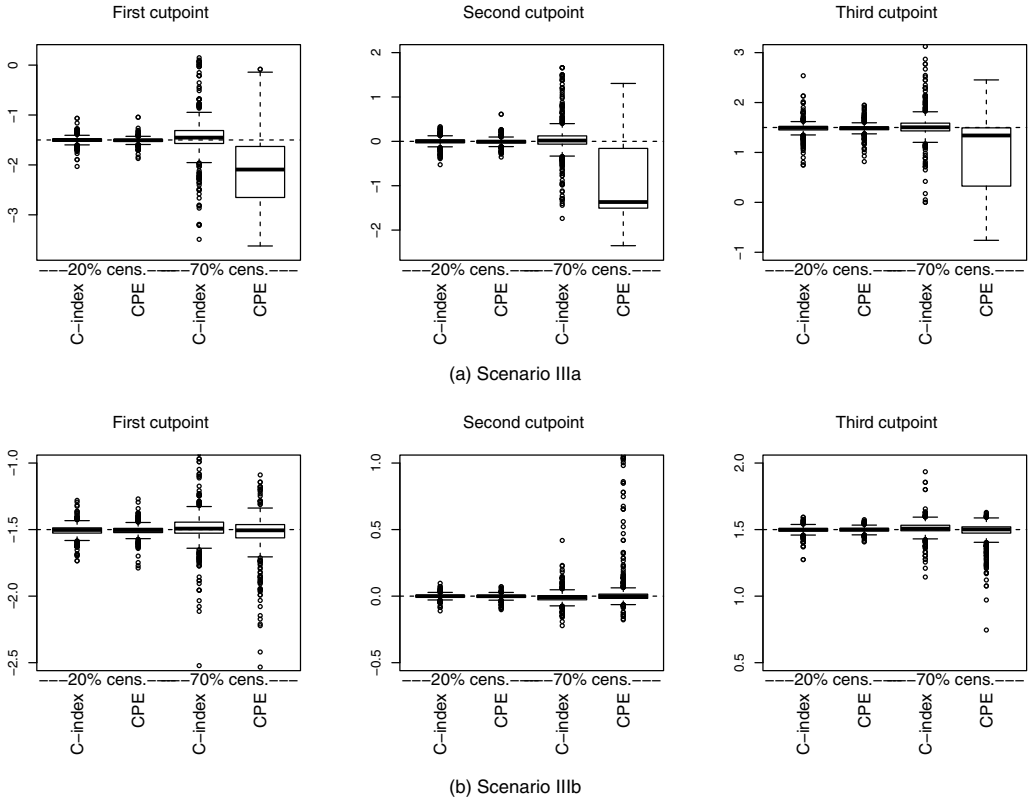
Figure 2 depicts the boxplots of the estimated optimal cutpoints for Scenario II, where two optimal cutpoints are sought for. Numerical results are reported in Table 3. Once again, the c-index outperformed the CPE when high censoring rates were considered. Nevertheless, for censoring rates below 50% both estimators performed satisfactorily.

Finally, Figure 3 depicts the boxplots of the estimated optimal cutpoints for Scenarios IIIa and IIIb, where three optimal cutpoints are sought for a monotonic increasing and non-monotonic risk relationship, respectively. Numerical results are reported in Table 4. These results suggest that the method performed satisfactorily regardless of the risk relationship considered. Nevertheless, for high censoring rates, the CPE performed better when a non-monotonic risk relationship was considered.

We must note that when more than one cutpoint is searched for, the estimated cutpoints have been ordered from the smallest to the largest to classify them as "first", "second" or "third" cutpoints. This may cause an incorrect classification whenever the estimated smallest cutpoint corresponds to the theoretical "second" cutpoint for example.

Table 4: *Simulations results when three theoretical optimal cutpoints were looked for with a monotonic increasing and non-monotonic relationship with the outcome (Scenarios IIIa and IIIb) and censorship of 20%, 50% and 70%. Mean, standard deviation, median, bias and mean squared error (MSE) for the estimated cutpoints over 500 simulated data sets are reported when CPE or C-index concordance probability estimators are used as the maximisation criteria.*

Sample size	Cens.	theoretical cutpoint	Cutpoint Estimation							
			CPE				c-index			
			Mean (sd)	Median	Bias	MSE	Mean (sd)	Median	Bias	MSE
Scenario IIIa										
$N = 500$	20%	−1,5	−1.507 (0.066)	−1.505	−0.007	0.008	−1.501 (0.078)	−1.499	−0.001	0.012
		0	−0.006 (0.087)	−0.006	−0.006		−0.002 (0.096)	0.002	−0.002	
		1,5	1.483 (0.107)	1.489	−0.017		1.484 (0.141)	1.495	−0.016	
	50%	−1,5	−1.568 (0.269)	−1.501	−0.068	0.075	−1.486 (0.177)	−1.488	0.014	0.032
		0	−0.078 (0.313)	−0.007	−0.078		−0.006 (0.174)	0.002	−0.006	
		1,5	1.458 (0.211)	1.483	−0.042		1.490 (0.181)	1.495	−0.010	
	70%	−1,5	−2.119 (0.675)	−2.093	−0.619	0.983	−1.442 (0.476)	−1.453	0.058	0.175
		0	−0.951 (0.744)	−1.368	−0.951		0.045 (0.450)	0.018	0.045	
		1,5	1.006 (0.641)	1.340	−0.494		1.523 (0.305)	1.504	0.023	
$N = 1000$	20%	−1,5	−1.499 (0.029)	−1.500	0.001	0.002	−1.499 (0.036)	−1.500	0.001	0.002
		0	−0.003 (0.040)	−0.003	−0.003		−0.004 (0.045)	−0.001	−0.004	
		1,5	1.492 (0.056)	1.498	−0.008		1.487 (0.058)	1.497	−0.013	
	50%	−1,5	−1.500 (0.072)	−1.497	0.000	0.005	−1.490 (0.084)	−1.494	0.010	0.006
		0	−0.005 (0.064)	−0.003	−0.005		0.002 (0.072)	0.001	0.002	
		1,5	1.487 (0.070)	1.496	−0.013		1.491 (0.069)	1.499	−0.009	
	70%	−1,5	−1.739 (0.469)	−1.525	−0.239	0.340	−1.455 (0.211)	−1.479	0.045	0.035
		0	−0.354 (0.635)	−0.025	−0.354		0.027 (0.188)	0.008	0.027	
		1,5	1.335 (0.436)	1.486	−0.165		1.512 (0.154)	1.500	0.012	
Scenario IIIb										
$N = 500$	20%	−1,5	−1.506 (0.047)	−1.503	−0.006	0.001	−1.508 (0.052)	−1.503	−0.008	0.001
		0	−0.002 (0.019)	−0.001	−0.002		−0.001 (0.018)	0.000	−0.001	
		1,5	1.499 (0.021)	1.500	−0.001		1.496 (0.029)	1.499	−0.004	
	50%	−1,5	−1.513 (0.070)	−1.505	−0.013	0.002	−1.506 (0.086)	−1.500	−0.006	0.003
		0	−0.002 (0.025)	−0.001	−0.002		−0.005 (0.027)	−0.002	−0.005	
		1,5	1.497 (0.033)	1.500	−0.003		1.499 (0.043)	1.502	−0.001	
	70%	−1,5	−1.469 (0.410)	−1.506	0.031	0.073	−1.495 (0.176)	−1.493	0.005	0.013
		0	0.050 (0.198)	−0.001	0.050		−0.010 (0.054)	−0.007	−0.010	
		1,5	1.481 (0.087)	1.501	−0.019		1.508 (0.067)	1.507	0.008	
$N = 1000$	20%	−1,5	−1.504 (0.023)	−1.502	−0.004	0.000	−1.502 (0.024)	−1.501	−0.002	0.000
		0	0.000 (0.010)	0.000	0.000		0.001 (0.010)	0.000	0.001	
		1,5	1.501 (0.012)	1.501	0.001		1.499 (0.013)	1.500	−0.001	
	50%	−1,5	−1.504 (0.034)	−1.502	−0.004	0.001	−1.501 (0.039)	−1.500	−0.001	0.001
		0	0.002 (0.013)	0.000	0.002		0.000 (0.016)	0.000	0.000	
		1,5	1.499 (0.020)	1.500	−0.001		1.500 (0.019)	1.501	0.000	
	70%	−1,5	−1.514 (0.071)	−1.502	−0.014	0.003	−1.504 (0.080)	−1.498	−0.004	0.003
		0	0.003 (0.033)	0.000	0.003		−0.003 (0.023)	−0.003	−0.003	
		1,5	1.494 (0.039)	1.501	−0.006		1.500 (0.036)	1.503	0.000	



(a) Scenario IIIa

(b) Scenario IIIb

Figure 3: *Boxplot of the estimated optimal cutpoints based on 500 simulated data sets, $N = 500$ sample size, three theoretical cutpoints, monotonic increasing and non-monotonic relationship. Results are shown for censoring rates of* 20% *and* 70% *and c-index and CPE discriminative ability estimators. From top to bottom: (a) Monotonic increasing relationship (Scenario IIIa); (b) non-monotonic relationship (Scenario IIIb). Theoretical cutpoints are represented with a dashed line.*

4. Application to the Stable-COPD study

Chronic obstructive pulmonary disease (COPD) is one of the most common chronic diseases, its prevalence is expected to increase over the next few decades (Buist et al., 2008), and is a leading cause of death in developed countries. Patients being treated for COPD at five outpatient respiratory clinics affiliated with the Hospital Galdakao-Usansolo in Biscay were recruited in the Stable-COPD study (Esteban et al., 2014). Patients were consecutively included in the study if they had been diagnosed with COPD for at least six months and had been receiving medical care at one of the hospital respiratory outpatient facilities for at least six months. Their COPD had to be stable for six weeks before enrolment. Patients were followed for up to five years. In total, information for 543 patients was obtained of which the 96.13% were men, the mean age was

of 68.32 and the 30.76% died in a 5-years period for which the mean survival time was of 2.77 years. The main selected variables collected in this study included sociodemographic variables, forced expiratory volume in one second in percentage ($FEV_{1\%}$), body mass index (BMI), dyspnea measured with the modified scale of the Medical Research Council (mMRC, Fletcher et al. (1959)) and the walking distance among others. A brief description of the main variables used in this paper is given in Table 5. One of the main goals of this study was to develop prediction models for patients with stable COPD.

Table 5: *A description of the selected variables from the Stable-COPD study ($N = 543$).*

Variable	**Mean (sd)**	**Range**
Age	68.32 (8.32)	33 - 86
$FEV_{1\%}$	55 (13.31)	18 - 105
BMI	28.28 (4.43)	16.38 - 44.04
Time until event (days)	1574.89 (483.43)	23 - 2045
Walking distance	408.89 (92.43)	46 - 644
Sex$^{\ddagger} - Men$	522 (96.13 %)	
Dyspnoea‡		
1	69 (12.71 %)	
2	264 (48.62 %)	
3	166 (30.57 %)	
4	23 (4.24 %)	
5	21 (3.87 %)	
5-year mortality$^{\ddagger} - Yes$	167 (30.76 %)	

‡Categorical variables are shown as absolute and relative frequencies

Table 6: *Airflow obstruction level measured by $FEV_{1\%}$ based on the different cutpoints used in the literature to categorise the continuous $FEV_{1\%}$ variable.*

Criteria	**Mild**	**Moderate**	**Severe**	**Very Severe**
GOLD	≥ 80	$[50-80)$	$[30-50)$	< 30
BODE	≥ 65	$[50-65)$	$(35-50)$	≤ 35
HADO	> 65	$[50-65]$	$[35-50)$	< 35
ADO		≥ 65	$(35-65)$	≤ 35
DOSE		≥ 50	$[30-50)$	< 30
COCOMICS	≥ 70	$(55-70)$	$(35-55]$	≤ 35

An important predictor for COPD mortality or hospitalisation is $FEV_{1\%}$, which is commonly used by clinicians to diagnose and measure the severity of the disease (Vestbo et al., 2013). Recently, several scores have been proposed which include a categorised version of $FEV_{1\%}$ among the predictor variables. The most commonly used scores are the original BODE index (Celli et al., 2004), HADO index (Esteban et al., 2006), ADO index (Puhan et al., 2009), and DOSE (Jones et al., 2009). Although all prediction scores are based on prediction models which use a categorised version of the predictor variable $FEV_{1\%}$, not all of them use the same cutpoints (see Table 6). To date, the most

widely-used cutpoints are the ones proposed by the Global Obstructive Lung Disease (GOLD) guidelines (mild ≥ 80, moderate 50-79, severe 30-49 and very severe < 30, Rabe et al. (2007)). More recently, Almagro et al. (2014) have proposed a new categorisation of $FEV_{1\%}$ to predict five-year survival in COPD patients. This research was framed within the Collaborative Cohorts to Assess Multicomponent Indices of COPD in Spain (COCOMICS) study.

Hence, and taking all this into account, three factors motivated us to look for the best categorisation of the variable $FEV_{1\%}$ as part of the development of the prediction survival model in the Stable-COPD study. First of all, this variable is an important predictor of survival in COPD patients. Since other prediction models and especially clinical guidelines use a categorised version of this variable, the clinicians involved in the study considered it was necessary to include a categorised version of this variable in the prediction model. Second, recent research shows the importance of seeking optimal cutpoints for this variable (Almagro et al., 2014). Third, as indicated above, to date there are no unified criteria on how to categorise the variable $FEV_{1\%}$.

We looked for the best categorisation of the predictor variable $FEV_{1\%}$ in a multivariate setting, taking into account the effect of age and dyspnoea, which are seen as important predictors for the severity of patients with stable COPD (Bestall et al., 1999). In fact, these variables together with a categorisation of $FEV_{1\%}$ are the ones used in the ADO index (Puhan et al., 2009), which turned out to be the best multivariate score to predict 5-year mortality based on the c-index (Marin et al., 2013). The censoring rate in our data set was 66.6%. Considering the results obtained in the simulation study, the c-index concordance probability estimator was used to select the optimal cutpoints since it appeared to perform better under this scenario. To select the optimal number of cutpoints we considered the bootstrap confidence interval for the bias-corrected c-index using $B = 200$ bootstrap replicates. In this data set, the proportional hazards assumption was verified (Grambsch and Therneau, 1994).

In a first stage we looked for $k = 3$ cutpoints and compared them with $k = 2$ cutpoints, which are also the number of cutpoints used in the categorisation of $FEV_{1\%}$ in the ADO index. Using the c-index estimator and the *Genetic* algorithm we obtained that the optimal cutpoints were $(29.32, 50.69)$ and $(29.90, 49.95, 50.54)$ when we looked for $k = 2$ and $k = 3$ number of cutpoints, respectively. In this case, the optimal cutpoints obtained when the CPE was used as the concordance probability estimator were almost the same, being $(29.79, 50.63)$ for $k = 2$ and $(29.69, 49.37, 50.82)$ for $k = 3$. When we compared $k = 2$ versus $k = 3$ number of cutpoints, we obtained a 95% bootstrap CI $(-0.005, 0.015)$ for the difference bias-corrected c-index. Consequently, the optimal number of cutpoints considering the multivariate setting would be $k = 2$, resulting in mild-moderate $(> 50\%)$, severe $([30\% - 50\%])$ and very severe $(< 30\%)$ categories. Note that the estimated cutpoints matched up with those used in the DOSE index (Jones et al., 2009) and those proposed in the GOLD guidelines (Rabe et al., 2007). The estimated cutpoint which separated the categories severe from very severe, differed slightly from the one used in the BODE, HADO and ADO indexes i.e., 35, which was iden-

tified by the American Thoracic Society (Celli et al., 2004). For illustration purposes, we would like to indicate that the bias corrected c-index for the ADO categorisation proposal was 0.701, lower than the 0.717 obtained using our approach.

5. Discussion

Categorisation of a continuous predictor variable is a commonly used strategy in biomedical research (Turner et al., 2010), where decisions are usually based on the risk classification of patients. To the best of our knowledge, up to now, no approaches have been proposed in the literature for the categorisation of a continuous predictor variable in a multiple Cox proportional hazards regression model. In this paper, we have proposed and validated by means of simulations a methodology to categorise a continuous predictor variable by maximising the concordance probability of the final model for the categorised variable.

Although the objective is different, several methods have been proposed in the literature to select optimal cutpoints (a unique cutpoint) for the prognosis of a disease (Faraggi and Simon, 1996; Sima and Gönen, 2013). In that context, the aim is to select the best cutpoint to dichotomise a variable and classify individuals as diseased or disease-free based on that cutpoint. Sima and Gönen (2013) proposed the maximal discrimination as a method to dichotomise a continuous predictor. They compared the maximisation of the discrimination indexes CPE and c-index together with the maximisation of the log-rank, Wald and partial likelihood ratio statistics for the location of one optimal cutpoint.

Our proposal is different to Sima and Gönen's proposal in one main aspect. Our goal is to categorise a continuous predictor variable to be used in a Cox proportional hazards regression model, considering any possible number of cutpoints. In fact, the most common scores used to predict mortality in COPD patients, such as BODE or ADO, use categorised versions (with more than two categories) of continuous predictors (Celli et al., 2004; Puhan et al., 2009). Furthermore, the methodology that we propose considers the effect that other predictor or confounding variables may have on the selection of the optimal cutpoints. Finally, our proposal allows to select the optimal number of cutpoints to categorise the predictor variable using a bootstrap confidence interval for the difference of the bias-corrected concordance probability estimators.

This proposal is an extension of the methodology proposed by Barrio et al. (2016) for the logistic regression setting. However, in time-to-event studies different estimators of the concordance probability have been proposed. In this paper we have studied and compared the performance of two estimators: the c-index and the CPE, in order to evaluate their performance in the categorisation of a continuous predictor variable in a Cox proportional hazards regression model.

The finite sample performance of the concordance probability estimators considered, i.e., c-index and CPE, was investigated through simulations. Results indicate that both concordance probability estimators performed satisfactorily in a multiple Cox regression model for any number of cutpoints and low-moderate censoring rates ($\leq 50\%$). When the censoring rate considered was high ($> 70\%$), the c-index appeared to outperform the CPE in all the scenarios considered. Additionally, the simulation results for three cutpoints showed that optimal cutpoints can be obtained regardless of the relationship of the latent continuous variable and the outcome. However, when we looked for a unique cutpoint in a univariate Cox proportional hazards regression model, results differed depending on the location of the theoretical cutpoint (results shown in the Supplementary Material). In fact, we observe that for a unique and not centred cutpoint, neither CPE nor c-index performed satisfactorily. Depending on whether the theoretical cutpoint was positively or negatively migrated from the centre of the distribution, smaller bias and MSE values were obtained for CPE or c-index. We must take into a account that when a univariate model is considered and the predictor variable takes only two possible values, there are many ties on the estimated survival probabilities and hence it may have an impact on the estimated concordance probability. Consequently, based on the simulation results obtained, we give the following recommendations for use in practice. For low-moderate censoring rates ($\leq 50\%$), either the c-index or the CPE can be used as maximisation criteria to obtain optimal cutpoints. However, for high censoring rates we recommend the c-index as the concordance probability estimator to maximise. Finally, we do not recommend the use of this proposal for dichotomisation in a univariate model.

Although we tried to evaluate many different scenarios, we could not address all possible real world settings and hence the conclusions we got can be extended only to those situations that were defined in the simulation study. In the scenarios we simulated we considered true optimal cutpoints in order to be able to compare the estimated cutpoints with those theoretical ones. Nevertheless, in practice neither the location or the number of cutpoints are known. We are aware that in theory the optimal number of cutpoints for the categorisation of a continuous variable does not exist, since above all the possible number of cutpoints, the best option would be the continuous variable. However, in clinical practice categorical versions of the continuous variables can be preferred without it always being clear which is the best number of categories to be used. For those situations we provided a proposal to decide among different number of cutpoints based on the bootstrap confidence interval (Barrio et al., 2016) which has been extended to the Cox proportional hazards regression model (see Supplementary Material). Although further research is needed to provide accurate methods for the selection of the optimal number of cutpoints, the results suggest that, when using the c-index, the number of cutpoints can be selected based on the bootstrap CI for the difference of the bias corrected estimated concordance probability.

In this paper we have not considered time-dependent discriminative ability measures as a parameter for selecting optimal cutpoints. Note that the concordance probability index is a global measure that does not take into account the time at which the prediction

of the event is desired. This implies that the optimal cutpoints are considered to be the same whatever the time of interest is. However, this may not be necessarily true. To overcome this problem, we are currently working on the application of time-dependent discrimination measures (Heagerty and Zheng, 2005; Pepe et al., 2008) in the search for time-dependent optimal cutpoints.

When we applied the proposed methodology to the Stable-COPD study, we saw that the cutpoints obtained to categorise the predictor variable $FEV_{1\%}$ corresponded to cutpoints previously used in the literature, obtaining clinically valid optimal cutpoints.

To summarise, we have compared the performance of two concordance probability estimators as the maximisation criteria to obtain optimal cutpoints to categorise continuous predictor variables in a Cox proportional hazards regression model. By means of simulations we have seen that the methodology proposed for categorising continuous predictors in a Cox proportional hazards regression model provides the optimal location and number of the cutpoints. Additionally, we have implemented this methodology into an R function which leads to easy use of this methodology in practice.

Funding

This study was supported by grants IT620-13 from the Departamento de Educación, Política Lingüística y Cultura del Gobierno Vasco, MTM2013-40941-P, MTM2014-55966-P and MTM2016-74931-P from the Ministerio de Economía y Competitividad and FEDER and the Agrupamento INBIOMED from DXPCTSUG-FEDER unha maneira de facer Europa (2012/273). María Xosé Rodríguez-Álvarez acknowledges financial support for Severo Ochoa Program SEV-2013-0323 and Basque Government BERC Program 2014-2017. Luís Meira-Machado acknowledges financial support for Portuguese Funds through FCT-"Fundação para a Ciência e a Tecnologia", within Project UID/MAT/00013/2013. The collection of the COPD data used for this study was supported in part by grants PI020510 from the Fondo de Investigación Sanitaria and 200111002, 2005111008, from the Departamento de Salud del Gobierno Vasco and the Research Committee Hospital Galdakao-Usansolo.

Acknowledgement

We would like to thank the referees for their constructive comments and suggestions which led to substantial improvement of this paper.

Conflict of interest

The authors declare that there are no conflicts of interest.

References

Almagro, P., Martinez-Camblor, P., Soriano, J., Marin, J., Alfageme, I., Casanova, C., Esteban, C., Soler-Cataluña, J., De-Torres, J., and Celli, B. (2014). Finding the best thresholds of FEV1 and dyspnea to predict 5-year survival in COPD patients: the COCOMICS study. *PLoS One*, 9:e89866.

Barrio, I., Arostegui, I., Rodríguez-Álvarez, M. X., and Quintana, J. M. (2016). A new approach to categorising continuous variables in prediction models: Proposal and validation. *Statistical Methods in Medical Research*, in press.

Bestall, J. C., Paul, E. A., Garrod, R., Garnham, R., Jones, P. W., and Wedzicha, J. A. (1999). Usefulness of the medical research council (MRC) dyspnoea scale as a measure of disability in patients with chronic obstructive pulmonary disease. *Thorax*, 54, 581–586.

Buist, A. S., Vollmer, W. M., and McBurnie, M. A. (2008). Worldwide burden of COPD in high-and low-income countries. Part I. The Burden of Obstructive Lung Disease (BOLD) Initiative. *The International Journal of Tuberculosis and Lung Disease*, 12, 703–708.

Celli, B. R., Cote, C. G., Marin, J. M., Casanova, C., Montes de Oca, M., Mendez, R. A., Pinto Plata, V., and Cabral, H. J. (2004). The body-mass index, airflow obstruction, dyspnea, and exercise capacity index in chronic obstructive pulmonary disease. *New England Journal of Medicine*, 350, 1005–1012.

Cox, D. R. (1972). Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society, Series B*, 34, 187–220.

Cox, D. R. and Oakes, D. (1984). *Analysis of Survival Data*. CRC Press.

Esteban, C., Arostegui, I., Aburto, M., Moraza, J., Quintana, J. M., Aizpiri, S., Basualdo, L. V., and Capelastegui, A. (2014). Influence of changes in physical activity on frequency of hospitalization in chronic obstructive pulmonary disease. *Respirology*, 19, 330–338.

Esteban, C., Quintana, J. M., Aburto, M., Moraza, J., and Capelastegui, A. (2006). A simple score for assessing stable chronic obstructive pulmonary disease. *QJM - An International Journal of Medicine*, 99, 751–759.

Faraggi, D. and Simon, R. (1996). A simulation study of cross-validation for selecting an optmimal cutpoint in univariate survival analysis. *Statistics in Medicine*, 15, 2203–2213.

Fletcher, C. M., Elmes, P. C., Fairbairn, A. S., and Wood, C. H. (1959). The significance of respiratory symptoms and the diagnosis of chronic bronchitis in a working population. *British Medical Journal*, 2, 257.

Gönen, M. and Heller, G. (2005). Concordance probability and discriminatory power in proportional hazards regression. *Biometrika*, 92, 965–970.

Grambsch, P. M. and Therneau, T. M. (1994). Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika*, 81, 515–526.

Harrell, F. E. (2015). *rms: Regression Modeling Strategies*. R package version 4.3-0.

Harrell, F. E., Califf, R. M., Pryor, D. B., Lee, K. L., and Rosati, R. A. (1982). Evaluating the yield of medical tests. *JAMA: The Journal of the American Medical Association*, 247, 2543–2546.

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

Jones, R. C., Donaldson, G. C., Chavannes, N. H., Kida, K., Dickson-Spillmann, M., Harding, S., Wedzicha, J. A., Price, D., and Hyland, M. E. (2009). Derivation and validation of a composite index of severity in chronic obstructive pulmonary disease: the DOSE index. *American Journal of Respiratory and Critical Care Medicine*, 180, 1189–1195.

Lausen, B. and Schumacher, M. (1996). Evaluating the effect of optimized cutoff values in the assessment of prognostic factors. *Computational Statistics & Data Analysis*, 21, 307–326.

Liu, X. and Jin, Z. (2015). Optimal survival time-related cut-point with censored data. *Statistics in Medicine*, 34, 515–524.

Marin, J. M., Alfageme, I., Almagro, P., Casanova, C., Esteban, C., Soler-Cataluña, J. J., de Torres, J. P., Martínez-Camblor, P., Miravitlles, M., Celli, B. R., and Soriano, J. B. (2013). Multicomponent indices to predict survival in COPD: the COCOMICS study. *European Respiratory Journal*, 42, 323–332.

Mebane, W. R. and Sekhon, J. S. (2011). Genetic optimization using derivatives: the rgenoud package for R. *Journal of Statistical Software*, 42, 1–26.

Meira-Machado, L. and Faria, S. (2014). A simulation study comparing modeling approaches in an illness-death multi-state model. *Communications in Statistics-Simulation and Computation*, 43(5), 929–946.

Mo, Q., Gonen, M., and Heller, G. (2012). *CPE: Concordance Probability Estimates in Survival Analysis*. R package version 1.4.4.

Pencina, M. J. and D'Agostino, R. B. (2004). Overall c as a measure of discrimination in survival analysis: model specific population value and confidence interval estimation. *Statistics in medicine*, 23(13), 2109–2123.

Pepe, M. S., Zheng, Y., Jin, Y., Huang, Y., Parikh, C. R., and Levy, W. C. (2008). Evaluating the roc performance of markers for future events. *Lifetime data analysis*, 14(1), 86–113.

Puhan, M. A., Garcia-Aymerich, J., Frey, M., ter Riet, G., Antó, J. M., Agustí, A. G., Gómez, F. P., Rodríguez-Roisín, R., Moons, K. G., Kessels, A. G., and Held, U. (2009). Expansion of the prognostic assessment of patients with chronic obstructive pulmonary disease: the updated bode index and the ado index. *The Lancet*, 374, 704–711.

R Core Team (2016). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing.

Rabe, K. F., Hurd, S., Anzueto, A., Barnes, P. J., Buist, S. A., Calverley, P., Fukuchi, Y., Jenkins, C., Rodriguez-Roisin, R., van Weel, C., and Zielinski, J. (2007). Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: Gold executive summary. *American Journal of Respiratory and Critical Care Medicine*, 176, 532–555.

Rota, M., Antolini, L., and Valsecchi, M. G. (2015). Optimal cut-point definition in biomarkers: the case of censored failure time outcome. *BMC Medical Research Methodology*, 15, 24.

Schmid, M. and Potapov, S. (2012). A comparison of estimators to evaluate the discriminatory power of time-to-event models. *Statistics in Medicine*, 31, 2588–2609.

Sima, C. S. and Gönen, M. (2013). Optimal cutpoint estimation with censored data. *Journal of Statistical Theory and Practice*, 7, 345–359.

Steyerberg, E. W., Moons, K. G. M., van der Windt, D. A., Hayden, J. A., Perel, P., Schroter, S., Riley, R. D., Hemingway, H., Altman, D. G., and Group, P. (2013). Prognosis research strategy (PROGRESS) 3: prognostic model research. *PLoS Medicine*, 10, e1001381.

Turner, E., Dobson, J., and Pocock, J. (2010). Categorisation of continuous risk factors in epidemiological publications: a survey of current practice. *Epidemiologic Perspectives & Innovations*, 7, 9.

Vestbo, J., Hurd, S. S., Agustí, A. G., Jones, P. W., Vogelmeier, C., Anzueto, A., Barnes, P. J., Fabbri, L. M., Martinez, F. J., Nishimura, M., Stockley, R. A., Sin, D. D., and Rodriguez-Roisin, R. (2013). Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: Gold executive summary. *American Journal of Respiratory and Critical Care Medicine*, 187, 347–365.