Biometrika (2018), **103**, 1, *pp*. 1–16
Printed in Great Britain Advance Access publication on 31 July 2018

Assessing cure status prediction from survival data using receiver operating characteristic curves

BY M. AMICO, I. VAN KEILEGOM AND B. HAN

Research Center for Operations Research and Statistics, KU Leuven, Naamsestaat, 69 Leuven, Belgium

mailis.amico@kuleuven.be ingrid.vankeilegom@kuleuven.be bo.han1949@gmail.com

SUMMARY

Survival analysis relies on the hypothesis that, if the follow-up will be long enough, the event of interest will eventually be observed for all observations. This assumption, however, is often not realistic. In fact, when interest lies in the time until a relapse from a cancer or the time until the occurrence of a certain disease, for example, a fraction of the patients may never experience the event of interest. The survival data then contain a cure fraction or long-term survivors usually associated with infinite survival times. A common approach to model and analyze this type of data consists in using cure models. Two types of information can therefore be obtained: the survival at a given time and the cure status, both possibly modelled as a function of the covariates. The cure status is often of interest for medical practitioners and one is usually interested in predicting it based on markers. The receiver operating characteristic, ROC, curves are one way to evaluate these predicting performances. However, the classical ROC curve method is not appropriate since the cure status is partially unobserved due to the presence of censoring in survival data. In this research, we propose a ROC curve estimator aiming to evaluate the cured/non-cured status classification performance from cure survival data. This estimator, which handles the presence of censoring, decomposes sensitivity and specificity by means of the definition of conditional probability, and estimates these two quantities by means of weighted empirical distribution functions. The mixture cure model is used to calculate the weights. Based on simulations, we demonstrate the good performance of the proposed method and compare it with the classical ROC curve nonparametric estimator that would be obtained if the cure status was fully observed. We also compare our proposed method with the ROC curves of Heagerty et al. (2000) for classical survival analysis. Finally, we illustrate the methodology on a breast cancer data set.

Some key words: Area under the curve; Cure model; ROC curve; Sensitivity; Specificity; Survival analysis.

1. INTRODUCTION

A fundamental assumption of survival analysis is that all subjects under study will eventually experience the event of interest. In some contexts, however, it may happen that a fraction of the subjects never experience this particular event. Indeed, when interest lies in the time until a woman gets pregnant, for example, some women will never have a child. Likewise, when one is interested in the time until a patient relapses from a cancer, some of them may never experience a relapse. In both of these examples, the assumption stated above does not hold and it seems reasonable to consider that the survival data do not only contain susceptible observations, but

that they rather are a combination of two types of subjects: those who experience the event and those who do not, these latter subjects being considered as long-term survivors or as cured subjects.

A common difficulty when working with survival data is the presence of right censoring, meaning that only a lower bound of the survival time is observed. As a consequence, the exact event time is only observed for some observations and the remaining individuals are censored. In the presence of a cure fraction, cured subjects are always censored since they never experience the event of interest. As a result, the censored fraction contains both cured and uncured observations and the cure status is then unobserved. In order to take such feature into account, classical survival analysis has been extended to cure models. Initially introduced by the works of Boag (1949) and Berkson & Gager (1952), the literature on cure models is mainly composed of two classes of models, namely, the mixture cure model, introduced by Farewell (1977) and Farewell (1982), and the promotion time cure model proposed by Yakovlev et al. (1996).

Cure models consider two quantities: the survival and the cure status and the literature on cure models mainly focused on modelling the effect of covariates on these two quantities (see Amico & Van Keilegom (2018) for a detailed literature review on that topic). However, very little has been done on evaluating the performance of predicting these two outcomes based on cure survival data, even though good predictions are essential for practitioners. Indeed, when there exists a possible cure fraction, we can think of situations where one would be interested in predicting who is cured and who is not based on marker(s) in order to determine if a treatment is necessary to prevent a cancer relapse. Likewise, being able to correctly predict the survival probability of an uncured patient after a certain time by taking into account the presence of cured subjects in the data is also important. A first contribution to that topic is due to Yu et al. (2008) who propose to validate individual prediction for patients with prostate cancer performed based on a joint longitudinal survival-cure model. Recently, Beyene et al. (2019) investigate the accuracy of time-dependent event prediction, extending to cure survival data the results that have been previously obtained for classical survival analysis by Heagerty et al. (2000), Heagerty & Zheng (2005), Chambless & Diao (2006), Blanche et al. (2013), Li et al. (2018) among others. Zhang & Shao (2018) propose a concordance measure, in the spirit of the c-index proposed by Harrell et al. (1982) and Harrell et al. (1984), to assess the prediction accuracy of the overall survival for uncured patients by taking into account the presence of a cure fraction. They extend the work of Göner & Heller (2005) for the Cox (1972) proportional hazards model. For the cure status, on the contrary, nothing has been done to the best of our knowledge, while it is an important issue.

To evaluate if a classifier M, corresponding to a single variable or a combination of variables, classifies correctly a set of subjects into two classes, called cases and controls, one usually considers jointly two quantities: the sensitivity which corresponds to the proportion of subjects classified as cases when they are effectively a case, and the specificity, that is, the proportion of subjects classified as a control when they effectively belong to the control class. When the classifier M is measured on a continuous scale, it has to be dichotomized in order to perform a binary classification. Let us suppose that the classes are represented by the binary variable D, such that $D = 1$ for a case, and $D = 0$ for a control and let us consider that a subject i is classified as a case when its classifier M_i is such that $M_i > k$, for some threshold k. As k can take on several values, there exist several possible sensitivities and specificities. To summarize all the information, one usually considers a receiver operating characteristic, ROC, curve, described, for example, in Pepe (2003) and Krzanowski & Hand (2009), which represents graphically all

possible combinations of the sensitivity, and one minus the specificity:

$$Se(k) = \text{pr}(M > k \mid D = 1), \tag{1}$$
$$1 - Sp(k) = \text{pr}(M > k \mid D = 0), \tag{2}$$

that can be obtained from all possible dichotomized versions of M, based on the value of the threshold k. It plots the sensitivity against one minus the specificity for all possible values of $k \in \mathbb{R}$ and its equation is $ROC(u) = Se\left\{(1 - Sp)^{-1}(u)\right\}$, $0 < u < 1$, where u is an index. The ROC curve is a monotone increasing function in the quadrant $(0, 1) \times (0, 1)$ and its position in the quadrant indicates the ability of the classifier M to discriminate between the two classes. A perfect classifier is such that $\text{pr}(M > k \mid D = 1) = 1$ and $\text{pr}(M > k \mid D = 0) = 0$ for some k. In that case all observations are perfectly classified. Graphically, it corresponds to a point of coordinate $(0, 1)$. Conversely, an uninformative classifier is such that $\text{pr}(M > k \mid D = 1) = \text{pr}(M > k \mid D = 0)$, for all k. In this situation, the distribution of M is the same in the two classes and the ROC curve is equal to the bisector. Alongside the ROC curve, one usually computes the area under the curve $AUC = \int_0^1 ROC(u)\, du$, which summarizes into one single value the performance of M. An area under the curve equal to 1 corresponds to a perfect classifier, while an area under the curve equal to 0.5 is obtained for an uninformative classifier.

In this paper, we propose to develop a ROC curve approach in order to evaluate the accuracy of a single covariate or a combination of covariates to predict the cure status based on cure survival data. Since the cure status is missing for censored observations, classical ROC curve approaches, which rely on the knowledge of the classes of the observations, can not be directly implemented in this context. An important issue to address is then how to handle the latency of the cure status. Our proposal is presented in §2 alongside some important points related to the estimation of the sensitivity and the specificity. In §3, some asymptotic properties are presented, followed in §4 by the investigation of the finite sample performance of the proposed method in the case of a known classifier through an extensive simulation study. §5 illustrates the practical use of our proposal on a breast cancer dataset, while §6 concludes with some final remarks and discussion. Finally, the Supplementary Material contains the theoretical development of our estimators, the proofs of the asymptotic properties derived in §3, the simulations for the case of an unknown classifier, as well as further finite sample results.

2. Methodology

2.1. *Infeasible estimators*

Let us consider a non-negative random variable denoted by T, which represents the survival time, with survival function $S(t) = \text{pr}(T > t)$, and let us assume that there exists a cure fraction. To further define this situation, let us consider that a cured subject is such that $T = \infty$, in order to represent the fact that the event never happens. A popular model for cure survival data is the mixture cure model (Farewell, 1982), which assumes that the population of interest is a mixture of a cured and an uncured sub-population, and which models the survival function for the entire population as a mixture model:

$$S_{pop}(t \mid x, z) = 1 - p(x) + p(x)S_u(t \mid z), \qquad t \geq 0, \tag{3}$$

where $p(x) = \text{pr}(T < \infty \mid X = x)$ is the probability of being uncured, referred to as the incidence, with X a vector of covariates, and $S_u(t \mid z) = \text{pr}(T > t \mid T < \infty, Z = z)$ is the conditional survival function for uncured observations, referred to as the latency, with Z another vector of covariates that may share some or all components or be completely different from

X. In what follows, we assume that T is subject to random right censoring and that instead of observing T, we rather observe the follow-up time $Y = \min(T, C)$ and the censoring indicator $\Delta = I(T \leq C)$, where C denotes the censoring time that is supposed to be independent of T given X and Z, and where $I(\cdot)$ is the indicator function. We further consider that we have a random sample of n independent and identically distributed observations $(Y_i, \Delta_i, X_i, Z_i),\ i = 1, \ldots, n$, having the same distribution as (Y, Δ, X, Z).

To evaluate the predictive accuracy of M for the cure status $D = I(T = \infty)$, we assume that $M = \gamma_0 + \gamma^T X$, where γ is a vector of parameters associated with X and γ_0 is an intercept term. We further allow X to be unidimensional or multidimensional. In the latter case, the vector of parameters $(\gamma_0, \gamma)^T$ can be known (in which case M is a known score such as a genetic score, for example), or it can be unknown and needs to be estimated.

A simple and common nonparametric method to estimate a ROC curve consists in estimating the sensitivity and the specificity by their empirical distribution functions:

$$\check{S}e(k) = 1 - \frac{1}{\check{N}_1} \sum_{i=1}^{n} \check{W}_{i1} I(M_i \leq k), \tag{4}$$

$$\check{S}p(k) = \frac{1}{\check{N}_0} \sum_{i=1}^{n} \check{W}_{i0} I(M_i \leq k), \tag{5}$$

where $\check{W}_{i1} = I(D_i = 1)$, $\check{W}_{i0} = I(D_i = 0)$, $\check{N}_1 = \sum_{i=1}^{n} \check{W}_{i1}$ and $\check{N}_0 = n - \check{N}_1$. The ROC curve estimator takes therefore the form of a step function with jumps at each M_i. When working with cure survival data, however, these estimators cannot be used as the cure status is unobserved. A way around this difficulty is not to consider two but three types of subjects following the cure threshold proposed by Taylor (1995). This proposal consists in considering as cured an observation with a censored follow-up time greater than the last uncensored follow-up time, denoted by τ. This rule makes reasonably sense when there is a clear evidence for the presence of a cure fraction. In such a context, we consider the existence of two sub-populations, and it is reasonable to consider that, when the follow-up period is sufficiently long and when it goes well after the last uncensored event time τ, observations with a censored follow-up time greater than most event times can be categorized as cured. Based on this rule, it is therefore possible to distinguish three types of observations. In fact, an uncensored subject experiences the event. It then belongs to the non-cured population with certainty, that is, $D = 0$. Based on the cure threshold, censored observations can be separated into two groups, those with a follow-up time $Y > \tau$, for whom $D = 1$, and those with a follow-up time $Y \leq \tau$. For this latter case, a probability, $\text{pr}(D = 1 \mid X, Z, C, T > C)$, replaces the unobserved cure status. It follows that estimators for the sensitivity and the specificity are the weighted empirical distribution functions:

$$\tilde{S}e(k) = 1 - \frac{1}{\tilde{N}_1} \sum_{i=1}^{n} \tilde{W}_{i1}\ I(M_i \leq k), \tag{6}$$

$$\tilde{S}p(k) = \frac{1}{\tilde{N}_0} \sum_{i=1}^{n} \tilde{W}_{i0}\ I(M_i \leq k), \tag{7}$$

where $\tilde{W}_{i1} = (1 - \Delta_i)\ \text{pr}(D = 1 \mid X = X_i, Z = Z_i, C = C_i, T > C_i)$, $\tilde{W}_{i0} = 1 - \tilde{W}_{i1}$, $\tilde{N}_1 = \sum_{i=1}^{n} \tilde{W}_{i1}$, and $\tilde{N}_0 = n - \tilde{N}_1$. Furthermore, when the cure threshold is assumed, $\tilde{W}_{i1}$ can further be written as $\tilde{W}_{i1} = (1 - \Delta_i)\{I(Y_i > \tau) + I(Y_i \leq \tau)\ \text{pr}(D = 1 \mid X = X_i, Z = Z_i, C =$

$C_i, T > C_i)\}$. An infeasible estimator for the ROC curve is then

$$\tilde{ROC}(u) = \tilde{Se}\{(1 - \tilde{Sp})^{-1}(u)\},\ 0 < u < 1. \tag{8}$$

This estimator is a monotone increasing function of u and is invariant to strictly increasing transformations of M, which are both required properties of ROC curves as described by Pepe (2003). The corresponding estimator for the area under the curve is

$$\tilde{AUC} = \frac{1}{\tilde{N}_0 \tilde{N}_1} \sum_{i=1}^{n} \sum_{j=1}^{n} I(M_j > M_i) \tilde{W}_{j1} \tilde{W}_{i0}. \tag{9}$$

The development of these estimators relies on the decomposition of the sensitivity, the specificity and the area under the curve based on the definition of conditional probability. The theoretical elements can be found in Section 1 of the Supplementary Material.

These estimators consider a random design. However, they can also be applied when the design is fixed, with obvious adaptations in the notations.

2.2. *Feasible estimators*

The probability $\text{pr}(D = 1 \mid X, Z, C, T > C)$ is involved in the infeasible estimators (6) and (7) of the sensitivity and the specificity, as well as in the infeasible estimator (9) of the area under the curve. It is therefore necessary to estimate this quantity in order to obtain estimators that can be used in practice. Based on the definition of conditional probability, this probability can be written as

$$\text{pr}(D = 1 \mid X, Z, C, T > C) = \frac{\text{pr}(T = \infty \mid X, Z, C)}{\text{pr}(T > C \mid X, Z, C)} = \frac{\text{pr}(T = \infty \mid X, Z)}{\text{pr}(T > C \mid X, Z, C)},$$

since T and C are independent given X and Z. Since we suppose that the data come from the mixture cure model (3), it can be further written as

$$\frac{\text{pr}(T = \infty \mid X, Z)}{\text{pr}(T > C \mid X, Z, C)} = \frac{1 - p(X)}{1 - p(X) + p(X) S_u(C \mid Z)}. \tag{10}$$

The literature on cure models offers various modelling approaches for the mixture cure model (3). The most common one is the logistic/Cox mixture cure model proposed by Kuk & Chen (1992), and further studied by Sy & Taylor (2000) and Peng & Dear (2000). This proposal assumes a logistic model for p, that is $p(x) = \exp(\gamma_0 + \gamma^T x)/\{1 + \exp(\gamma_0 + \gamma^T x)\}$ and considers a Cox model for S_u, where $S_u(t \mid z) = S_0(t)^{\exp(\beta^T z)}$, with $S_0(t) = \text{pr}(T > t \mid T < \infty, Z = 0)$, a baseline conditional survival function which remains totally unspecified, and β a vector of parameters associated with Z. A drawback of this model, however, is that the estimator for $\text{pr}(D = 1 \mid X, Z, C, T > C)$ relies on a parametric assumption for p which may not be fulfilled by the data. An alternative model is the single-index/Cox mixture cure model proposed by Amico et al. (2019), which assumes a single-index structure for p, that is $p(x) = g(\gamma^T x)$, where g is a smooth unknown function, with a Cox model for S_u. This single-index/Cox cure model assumes a less restrictive model for p and it may therefore be more appropriate. Both approaches are considered and their respective finite sample performances are compared in §4. The estimators for $\tilde{W}_{i0}$ and $\tilde{W}_{i1}$ are

$$\hat{W}_{i1} = I(Y_i > \tau) + (1 - \Delta_i)\, I(Y_i \leq \tau)\, \frac{1 - \hat{p}(X_i)}{\{1 - \hat{p}(X_i)\} + \hat{p}(X_i) \hat{S}_u(Y_i \mid Z_i)}$$

$$\hat{W}_{i0} = 1 - \hat{W}_{i1},$$

and are obtained by either a logistic/Cox cure model or a single-index/Cox cure model. The feasible estimators of Se, Sp, ROC and AUC are now

$$\hat{Se}(k) = 1 - \frac{1}{\hat{N}_1} \sum_{i=1}^{n} \hat{W}_{i1}\, I(M_i \leq k), \tag{11}$$

$$\hat{Sp}(k) = \frac{1}{\hat{N}_0} \sum_{i=1}^{n} \hat{W}_{i0}\, I(M_i \leq k), \tag{12}$$

$$\hat{ROC}(u) = \hat{Se}\{(1 - \hat{Sp})^{-1}(u)\}, \qquad 0 < u < 1, \tag{13}$$

$$\hat{AUC} = \frac{1}{\hat{N}_0 \hat{N}_1} \sum_{i=1}^{n} \sum_{j=1}^{n} I(M_j > M_i) \hat{W}_{j1} \hat{W}_{i0}, \tag{14}$$

where $\hat{N}_1 = \sum_{i=1}^{n} \hat{W}_{i1}$ and $\hat{N}_0 = n - \hat{N}_1$.

Both X and Z enter in the computation of $\hat{W}_0$ and $\hat{W}_1$, while M only relies on X. For the choice of the covariates to include in X, we consider those included in M. A more delicate question concerns the choice of the covariates to consider for Z. When M only contains one covariate, and when there is only one covariate available in the data, it is easy to assume that $X = Z$. If there are several covariates in the data, or when M is a combination of covariates, on the contrary, the choice of Z will depend on the knowledge of the topic of the analysis, and on which covariates are thought to influence the survival of uncured subjects. In such contexts, Z can be partially or fully identical to X, or completely different from X. However, we are not free of misspecification. The influence of a misspecification of this vector on the estimation of the ROC curve is therefore investigated through simulations in §4.

3. ASYMPTOTIC THEORY

In this section we will develop the limiting distribution of the proposed estimators of the sensitivity, the specificity, the ROC curve and the area under the curve given in equations (11), (12), (13) and (14). In the previous section these estimators were constructed either based on a logistic/Cox mixture cure model or on a single-index/Cox mixture cure model. However, asymptotic theory for the estimation of these models has only been developed so far under the logistic/Cox model (Lu, 2008), and so we restrict attention in this section to the latter model. The proofs of the results of this section can be found in Section 2 of the Supplementary Material.

THEOREM 1. *Assume that conditions 1–4 in Lu (2008) are satisfied and that the logistic/Cox mixture cure model is valid. Then,*

$$\hat{Se}(k) - Se(k) = n^{-1} \sum_{i=1}^{n} \eta_{Se}(X_i, Z_i, Y_i, \Delta_i, k) + R_{n,Se}(k)$$

$$\hat{Sp}(k) - Sp(k) = n^{-1} \sum_{i=1}^{n} \eta_{Sp}(X_i, Z_i, Y_i, \Delta_i, k) + R_{n,Sp}(k),$$

where $\sup_k |R_{n,Se}(k)| = o_{pr}(n^{-1/2})$, $\sup_k |R_{n,Sp}(k)| = o_{pr}(n^{-1/2})$, *and* $\eta_{Se}(x, z, y, \delta, k)$ *and* $\eta_{Sp}(x, z, y, \delta, k)$ *are defined in the Supplementary Material.*

Moreover, the process $n^{1/2}\{\hat{Se}(k) - Se(k)\}$, $k \in \mathbb{R}$, *converges weakly to a Gaussian process* $Z_{Se}(k)$ *with zero mean and covariance function given by*

$$cov\{Z_{Se}(k_1), Z_{Se}(k_2)\} = E\big\{\eta_{Se}(X, Z, Y, \Delta, k_1)\, \eta_{Se}(X, Z, Y, \Delta, k_2)\big\},$$

and the process $n^{1/2}\{\hat{Sp}(k) - Sp(k)\},\ k \in \mathbb{R}$, *converges weakly to a Gaussian process* $Z_{Sp}(k)$ *with zero mean and covariance function given by*

$$cov\{Z_{Sp}(k_1), Z_{Sp}(k_2)\} = E\big\{\eta_{Sp}(X, Z, Y, \Delta, k_1)\ \eta_{Sp}(X, Z, Y, \Delta, k_2)\big\}.$$

As a corollary to the above result we now state the limiting distribution of the estimator $\hat{ROC}(u)$ defined in (13) and of the estimator $\hat{AUC}_\delta$, given by

$$\hat{AUC}_\delta = \int_\delta^{1-\delta} \hat{ROC}(u)\, du.$$

For technical reasons we need to restrict the integration to the internal $[\delta, 1-\delta]$, for some small $\delta > 0$, which can however be made arbitrarily close to the interval $[0, 1]$. The corresponding theoretical area under the curve is denoted by $AUC_\delta = \int_\delta^{1-\delta} ROC(u)\, du$.

COROLLARY 1. *Assume that conditions 1–4 in Lu (2008) are satisfied and that the logistic/Cox mixture cure model is valid. Assume in addition that* $\inf_{k_1 \le k \le k_2} Sp'(k) > 0$, *where* $k_1 = (1-Sp)^{-1}(\delta)$ *and* $k_2 = (1-Sp)^{-1}(1-\delta)$ *for some* $\delta > 0$, *and that the functions* Se *and* Sp *are twice continuously differentiable on* $[k_1, k_2]$. *Then,*

$$\hat{ROC}(u) - ROC(u) = n^{-1} \sum_{i=1}^{n} \eta_{ROC}(X_i, Z_i, Y_i, \Delta_i, u) + R_{n,ROC}(u),$$

where $\sup_{\delta \le u \le 1-\delta} |R_{n,ROC}(u)| = o_{pr}(n^{-1/2})$, *and*

$$\eta_{ROC}(x, z, y, \delta, u) = \eta_{Se}\big\{x, z, y, \delta, (1-Sp)^{-1}(u)\big\} + \frac{Se'\{(1-Sp)^{-1}(u)\}}{(1-Sp)'\{(1-Sp)^{-1}(u)\}} \eta_{Sp}\big\{x, z, y, \delta, (1-Sp)^{-1}(u)\big\}.$$

Moreover, the process $n^{1/2}\{\hat{ROC}(u) - ROC(u)\}$, *with* $u \in [\delta, 1-\delta]$, *converges weakly to a Gaussian process* $Z_{ROC}(u)$ *with zero mean and covariance function*

$$cov\{Z_{ROC}(u_1), Z_{ROC}(u_2)\} = E\big\{\eta_{ROC}(X, Z, Y, \Delta, u_1)\ \eta_{ROC}(X, Z, Y, \Delta, u_2)\big\},$$

and

$$n^{1/2}(\hat{AUC}_\delta - AUC_\delta) \xrightarrow{d} N(0, \sigma^2_{AUC}),$$

where

$$\sigma^2_{AUC} = \int_\delta^{1-\delta} \int_\delta^{1-\delta} E\{Z_{ROC}(u_1) Z_{ROC}(u_2)\}\, du_1\, du_2.$$

4. FINITE SAMPLE PERFORMANCE

4.1. *Some preliminaries*

In this section, an extensive simulation study is performed in order to evaluate the finite sample performance of the estimator (13) of the ROC curve. Two versions of this estimator are considered: when a logistic/Cox mixture cure model is assumed for W_0 and W_1, and when a single-index/Cox mixture cure model is assumed for W_0 and W_1. These models are estimated assuming the method proposed by Sy & Taylor (2000) based on the expectation-maximization algorithm and according to the maximum likelihood approach described in Amico et al. (2019), respectively. Both the case of known and unknown M are investigated, and for both of them, the

following points are analyzed. First, we are interested in the general performance of the proposed estimators of the sensitivity and the specificity. Particular interest lies in the effect of censoring and of an incorrect specification of the vector Z. Then, other points include a misspecification of the model for S_u and a non-logistic model for the cure proportion p.

To assess the performance of our proposed method, we consider two infeasible competing estimators of the ROC curve: the first one assumes that the cure status is known and is obtained from the estimators (4) and (5) for the sensitivity and the specificity, and the second one uses the true weights and corresponds to the estimator (8) combined with (10), based on the true values of p and S_u. Furthermore, as suggested by a referee, in the case of a known classifier we also compare our proposal with the method proposed by Heagerty et al. (2000), which was developed for 'classical' survival data without taking a possible cure fraction into account. Instead of having a single binary outcome, the survival time is considered as a time-varying binary outcome and ROC curves can be computed at each time point. In our context, we are interested in the predicting performance of markers on the probability of being cured, so we can think of using this approach by computing a ROC curve at the cure threshold.

Note that several methods have been proposed in the literature for survival data, these methods standing out by the definition of sensitivity and specificity they assumed: incident or cumulative sensitivity and static or dynamic specificity (see Heagerty & Zheng (2005) for a detailed definition of these different types of sensitivity and specificity). We have chosen Heagerty et al. (2000)'s method because the definition of sensitivity and specificity was the same as the one we assume (cumulative sensitivity and dynamic specificity). The method is based on the following idea: first they rewrite the sensitivity and specificity based on Bayes theorem. Then, they estimate the unknown quantities using nonparametric estimators (namely a Kaplan-Meier estimator or an estimator coming from nearest neighbor estimation of the bivariate distribution). So contrary to our method the latter method does not make any model assumption. It does not only differ in the assumptions on the underlying model, but also in the way the estimators are constructed. We refer to Heagerty et al. (2000) for more details. Since survival models without cure fraction depend on covariates only through the survival function itself, only the vector X has been used for the computation of the ROC curve based on this method.

In conclusion, five estimators are compared in this simulation study: the proposed estimators based on either the logistic/Cox model or the single-index/Cox model, the infeasible estimators based on the true cure status or on the true weights, and the estimator of Heagerty et al. (2000).

4.2. *Data generating process*

Within this section, we assume that the data are generated from the mixture cure model (3). The data generating process is as follows. First, the incidence is considered. The first step consists in generating the uncure probability p according to the model $p(x) = g(\gamma^T x)$, where $g(\cdot)$ is a link function. Primary interest lies in the logistic link function, that is, $g(a) = \exp(\gamma_0 + a)/\{1 + \exp(\gamma_0 + a)\}$, which gives the logistic regression model. However, other link functions can also be assumed. The second step consists in generating, for given x, the uncure status $1 - D$ from a Bernoulli distribution with parameter equal to $p(x)$. Next, the latency is generated. We consider two models for the survival function of the uncured observations. The first model is a Gompertz model with survival function $S(t \mid z) = S_0(t)^{\exp(\beta^T z)}$, $S_0(t) = \exp[-\theta\alpha^{-1}\{\exp(\alpha t) - 1\}]$, $\theta = 0.5$ and $\alpha = 0.03$. The second model is an accelerated failure time model assuming a log-logistic distribution for T and with survival function $S_u(t \mid z) = \left[1 + \lambda \left\{t/\exp(\beta^T z)\right\}^{\kappa}\right]^{-1}$ where $\lambda = 0.05$ and $\kappa = 2.5$. Contrarily to the Gompertz model, the accelerated failure time model does not respect the proportional hazards property. Since $\hat{W}_0$ and $\hat{W}_1$ are obtained from a mixture cure model assuming a Cox model for the

latency, this allows us to verify whether a model misspecification of S_u affects the ROC curve estimate. Next, we generate the censoring time from an uniform distribution on $[U_{min}, U_{max}]$ that is independent of T, X and Z. We further truncate the survival times of the susceptible observations at $U_{max} - 1$ so that the support of C is larger than the support of the susceptible T's. We finally generate the follow-up time $Y = \min(T, C)$ and the censoring indicator $\Delta = I(T \leq C)$.

4.3. *Known classifier*

First, we consider the case where the classifier takes the form of a single variable or of a known one-dimensional score denoted by X. Note that when $\dim(X) = 1$, the single-index model reduces to a non-parametric model. We assume three different scenarios for the incidence. The first two scenarios assume a logistic regression model for $p(x)$ corresponding to different discriminations between the cured and the uncured sub-populations as shown in Figure 1 (a) and (b). Scenario 1 is such that $X \sim N(2, 2.5)$, $\gamma_0 = 0$, $\gamma_1 = 1$ and $AUC = 0.9016$. This scenario corresponds to a a good discrimination with a cure proportion equal to 25.6%. Scenario 2 assumes that $X \sim N(1.2, 1)$, $\gamma_0 = 0$, $\gamma_1 = 1$ and $AUC = 0.7374$. This scenario is associated with a moderate separation between the two sub-populations, and the proportion of cured subjects is equal to 26.9%. The third scenario assumes a non-logistic model for $p(x)$ with a non-monotone shape in order to evaluate the performance of the two estimators in such a case. The link function is given by $g(a) = [\sin\{(3/2)\ \pi\ a\} + 1]/2$. Its characteristics are as follows: $X \sim \text{Unif}(0, 1)$, $\gamma_1 = 1$, and $AUC = 0.8124$, corresponding to a good separation between cured and uncured sub-populations as shown in Figure 1 (c). The cure proportion equals 39.4%.

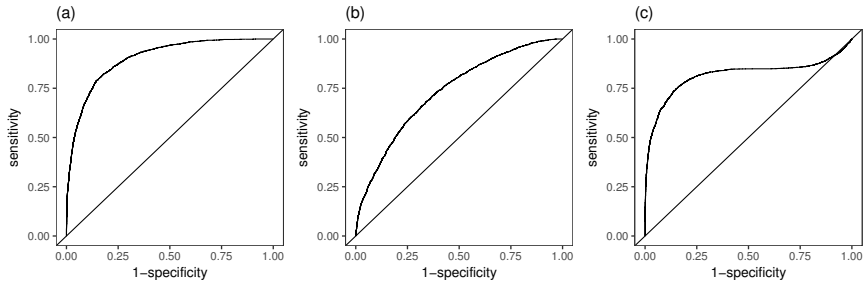

Fig. 1: True ROC curves when the classifier is known for (a) scenario 1, (b) scenario 2 and (c) scenario 3.

For the survival times, we consider for the Gompertz model two covariates, Z_1 and Z_2, that are independent, following a Bernoulli distribution with parameter equal to 0.6 and 0.2, respectively. The associated vector of parameters is $\beta = (1.5, -0.5)^T$. For the uniform distribution considered for the censoring time C, we assume that $U_{min} = 0$ and three different values are considered for U_{max} : 65, 25 and 10, corresponding to three different levels of censoring denoted by level 1, level 2 and level 3. For the accelerated failure time model, two independent covariates, Z_1 and Z_2, are considered, following a Bernoulli distribution with parameter equal to 0.6 and 0.3, respectively. The associated vector of parameters is $\beta = (0.7, -0.3)^T$. As for the Gompertz model, the censoring time is generated from a uniform distribution with $U_{min} = 0$ and with three different values for U_{max}. These values are chosen such that the proportion of censored observations with a follow-up time lower than or equal to τ is the same as for the Gompertz model, in order to allow comparison between the two models.

A total of five settings is considered. They correspond to the combination of scenario 1 and scenario 2 with the two models for the latency, as well as scenario 3 in combination with the

Gompertz model only, since this last scenario serves to assess the performance of our estimator assuming a logistic/Cox model when the logistic model is not satisfied. To assess the effect of a misspecification of Z, the estimators are further estimated assuming that $Z = X$, on scenario 1/Gompertz and scenario 2/Gompertz. Table 1 summarizes the characteristics of these settings.

Table 1: *Characteristics when the classifier is known: the cure rate, the upper bound U_{max} of the support of the censoring distribution, the probability p_U that finite values of T exceed U_{max} before truncating T, the censoring rate and the percentage of censored observations for which $Y \leq \tau$.*

		Latency type							
		Gompertz model				AFT model			
Incidence type	cure rate	U_{max}	p_U	censoring rate	$\Delta = 0$, $Y \leq \tau$	U_{max}	p_U	censoring rate	$\Delta = 0$, $Y \leq \tau$
Scenario 1	25.6%	65	0	27.0%	5.3%	345	0	27.0%	5.3%
	25.6%	25	0	29.1%	12.6%	121	0.0002	29.5%	12.6%
	25.6%	10	0.0027	34.0%	25.4%	44	0.0036	35.9%	25.5%
Scenario 2	26.9%	65	0	28.2%	5.3%	360	0	28.2%	5.3%
	26.9%	25	0	30.3%	13.0%	120	0.0003	30.8%	12.9%
	26.9%	10	0.0026	35.1%	25.9%	45	0.0034	37.1%	26.1%
Scenario 3	39.4%	65	0	40.5%	6.8%				
	39.4%	25	0	42.2%	16.2%				
	39.4%	10	0.0020	46.3%	31.4%				
	39.4%	5	0.0230	51.8%	43.1%				

For each setting we consider 500 datasets, and we carry out simulations for two sample sizes, namely $n = 250$ (in the main text) and $n = 500$ (in the Supplementary Material). Two criteria are considered to compare the five estimators, namely the L1 distance between the true and the estimated ROC curves and the area under the curve. The L1 distance is given by

$$L1 = V^{-1} \sum_{i=1}^{V} \mid \hat{ROC}(u_i) - ROC(u_i) \mid,$$

where $\hat{ROC}$ is one of the ROC curve estimates and ROC is the true ROC curve. It is computed over a grid of points $u_i = i/100$ for $i = 1, \ldots, V = 99$. For our estimators, both when a logistic/Cox and a single-index/Cox model are assumed, the area under the curve is given by

$$\hat{AUC} = \frac{1}{\hat{N}_0 \hat{N}_1} \sum_{i=1}^{n} \sum_{j=1}^{n} \Big[\{ I(M_j > M_i) + 0.5 \times I(M_j = M_i) \} \hat{W}_{j1} \hat{W}_{i0} \Big].$$

For the estimators based on the known cure status and the true weights, the formula is almost the same, but with different W_{i0}, W_{i1}, N_0 and N_1. For the estimator that is based on the known cure status, they are replaced by $\check{W}_{i0}$, $\check{W}_{i1}$, $\check{N}_0$ and $\check{N}_1$, while for the estimator based on the true weights, they are given by $\tilde{W}_{i0}$, $\tilde{W}_{i1}$, $\tilde{N}_0$ and $\tilde{N}_1$. Note that these formulas take into account possible ties in M with the added term $0.5 \times I(M_j = M_i)$. Finally, for the estimator of Heagerty et al. (2000) we restrict attention to the AUC, since the L1 distance is harder to compute for the latter method.

As can be seen in Figure 2, when the censoring rate is close to the cure rate and when everything is specified correctly, our proposals perform almost as well as the two infeasible competi-

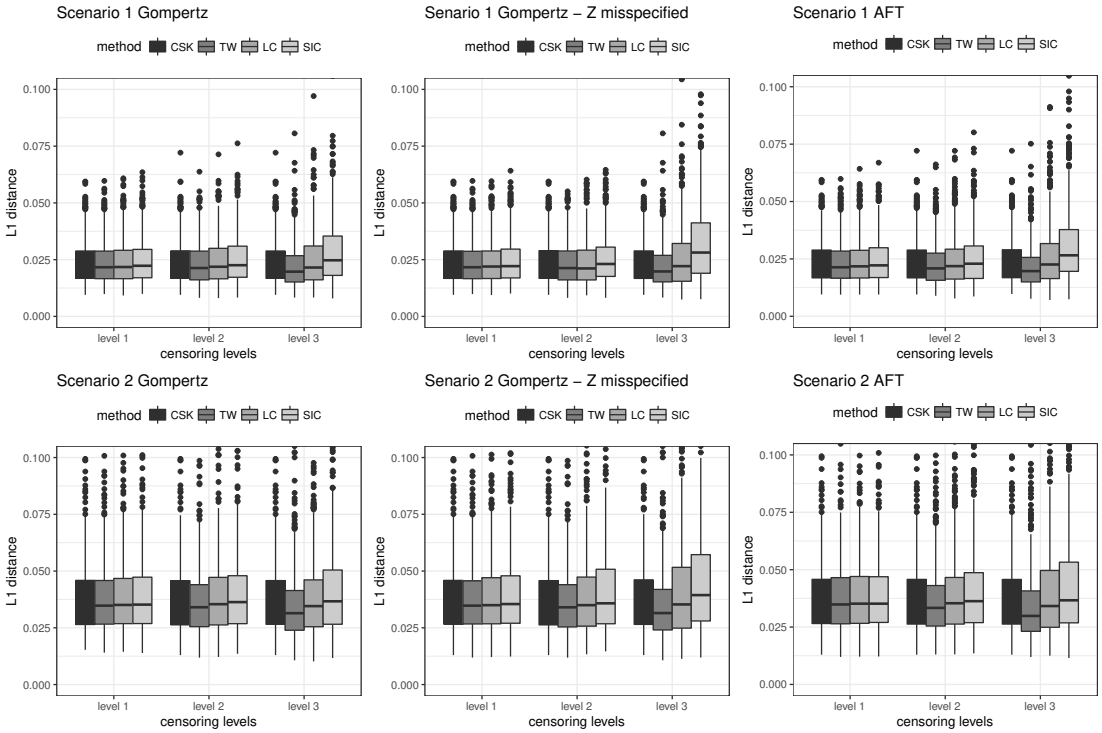

Fig. 2: Boxplots of the L1 distances when the classifier is known and the true link function is logistic for $n = 250$ (CSK, cure status known; TW, true weigths; LC, logistic/Cox model; SIC, single-index/Cox model).

tors in terms of L1 distance, whatever the model assumed for W_1. In such a case, very few censored observations are below τ, which are those with weight equal to $\text{pr}(D = 1 \mid X, Z, C, T > C)$. A larger censoring rate is associated with a higher L1 distance and a larger variance, particularly for our single-index/Cox approach under the third censoring level. When the censoring rate gets larger, fewer censored observations are located in the plateau, meaning that less censored observations are considered as cured, that is, with $W_{i1} = 1$. Furthermore, as shown by Amico et al. (2019), the single-index/Cox mixture cure model performs worse than the logistic/Cox cure model when the true model for the incidence is a logistic model and when the censoring rate increases as is the case for the third censoring level. Interestingly, our logistic/Cox model approach is close to the estimator based on the known cure status even for the third censoring level. Another interesting point is that the L1 distance for the estimator based on the true weights decreases slightly when the censoring rate increases. It seems that having more censored observations below τ produces better results when considering the true W_0 and W_1. In such a case, the sizes of the jumps are smaller and it seems that the ROC curve becomes 'smoother' and closer to the true curve. Nevertheless, this feature is not observed for our two methods. By comparing scenario 1 and scenario 2, we observe that the L1 distance is larger for scenario 2. Indeed, it is more difficult to correctly separate cured from uncured sub-populations based on this scenario since the discrimination is moderate. The discrimination between cured and uncured sub-populations seems therefore to have an influence on the performance of the estimators of the ROC curve. However, the general conclusions are the same for both scenarios. For the settings where Z is misspecified, we observe that when few censored observations are below τ, the L1 distance for

our proposals is only slightly higher in comparison with the two infeasible competitors, while when the number of censored observations below τ is larger as for the third censoring level, both our logistic/Cox and single-index/Cox estimators have higher L1 distance than when Z is correctly specified. Interestingly, for the second censoring level, the L1 distance for our logistic/Cox estimator seems to be comparable to the L1 distance of the estimator based on the known cure status for both scenarios while our single-index/Cox approach seems to already present some difficulties. However, we consider in these simulations the case where Z is completely misspecified, whereas it seems more likely to have only a partial misspecification of this vector of covariates in practical applications. We are therefore in an extreme case. When the survival times are generated according to an accelerated failure time model, our proposals show a higher increase in the L1 distance in comparison with when there is no misspecification, especially when the censoring rate increases. Our single-index/Cox estimator is still the least favourable estimator. However, a misspecification in Z affects the performance of our proposals more than a misspecification in the latency.

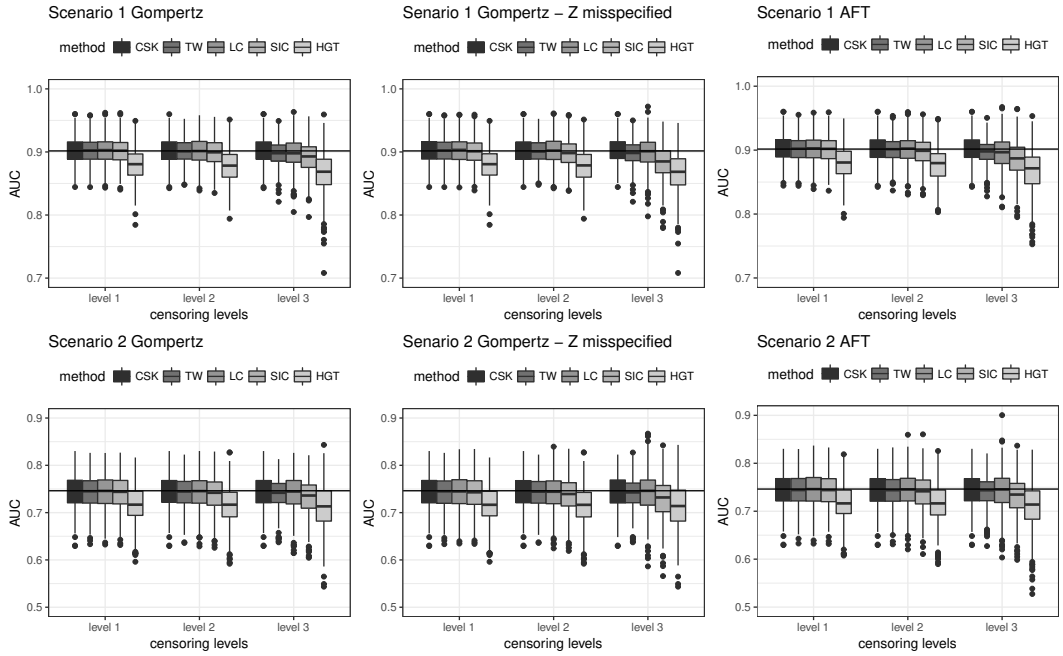

Fig. 3: Boxplots of the area under the curve when the classifier is known and the true link function is logistic for $n = 250$ (CSK, cure status known; TW, true weigths; LC, logistic/Cox model; SIC, single-index/Cox model; HGT, Heagerty et al. (2000)).

For the area under the curve, the same conclusions as for the L1 distances can be drawn, as shown in Figure 3. When the true incidence is logistic, the method based on a single-index/Cox model performs less good than for the logistic/Cox approach, especially for the third censoring rate. When a logistic/Cox model is assumed, on the contrary, our estimator is close to the estimated based on the known cure status for all censoring rates considered, but we observe more variability. Finally, the method of Heagerty et al. (2000) performs less good than all other estimators. This can probably be explained by the fact that this method is based on completely nonparametric estimators of the sensitivity and specificity. However, surprisingly, even when the model is misspecified (in the sense that the true model is AFT instead of Cox, or that Z is mis-

specified), the method of Heagerty et al. (2000) still behaves worse than its competitors that are based on wrong model assumptions.

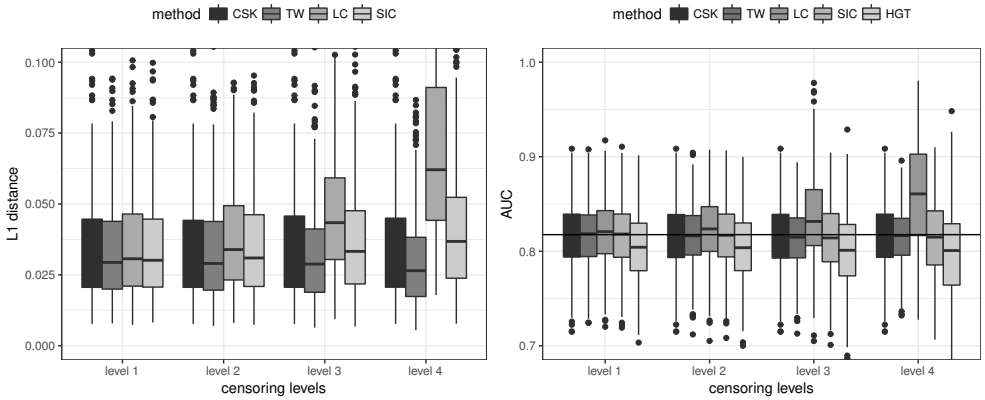

Fig. 4: Boxplots of the L1 distance and the area under the curve when the classifier is known and the true link function is not monotone for $n = 250$ (CSK, cure status known; TW, true weigths; LC, logistic/Cox model; SIC, single-index/Cox model; HGT, Heagerty et al. (2000)).

When the true incidence is not a logistic regression model, our logistic/Cox estimator has always higher L1 distances than for the single-index/Cox approach as can be seen in Figure 4. When the censoring rate gets larger, the difference between the logistic/Cox and the single-index/Cox approaches increases, and we observe that the single-index/Cox estimator outperforms the logistic/Cox one, especially for the third and the fourth censoring levels. It seems therefore that, when $\hat{p}(x)$ is inconsistent and when the proportion of censored observations below τ is large, our logistic/Cox estimator performs badly. Moreover, an interesting feature that was already observed from the previous settings and which is confirmed with the fourth censoring level here is that the more the censoring rate increases, the more the L1 distance of our method, for both the logistic/Cox and the single-index/Cox approach, increases since many more observations have $\hat{W}_0$ and $\hat{W}_1$ relying on $\text{pr}(D = 1 \mid X, Z, C, T > C)$ in that case. The censoring level is then crucial in the performance of our proposals. For the AUC, similar conclusions can be drawn. For the method of Heagerty et al. (2000) we notice again a somewhat less good behavior compared to its competitors that make use of correct model assumptions.

Finally, when $n = 500$, the same conclusions also apply for both the L1 distance and the area under the curve. The boxplots are in Section 4 of the Supplementary Material. As expected, the L1 distances are smaller and less variable than for $n = 250$. For the area under the curve, the variability is also lower.

4.4. *Unknown classifier*

As the conclusions for the case of an unknown classifier are very similar to those for the case of a known classifier, we do not give detailed simulation results here, and refer instead to Section 3 of the Supplementary Material for more details.

5. APPLICATION

In this section, we examine a data set from a breast cancer study, to illustrate the application of our methodology. The breast cancer data set is available from the Surveillance Epidemiology

and End Results (SEER) (www.seer.cancer.gov) database. Specifically, we are using the database 'Incidence - SEER 18 Regs Research Data' via SEER Stat. The data set consists of women diagnosed with breast cancer between 2004 and 2016, whose cancer state is classified as localized or regional in five arbitrary counties from the data set, namely those counties whose county code equals 133 (Georgia Greene County, Kentucky Letcher County and Iowa Monona County) or whose county code equals 205 (Georgia Mitchell County and Kentucky Rowan County). The failure time of interest is the time from diagnosis of breast cancer to death from breast cancer. Baseline covariates include tumor size, ER status (ER^+ vs ER^-) and PR status (PR^+ vs PR^-). Patients with unknown ER/PR status are excluded from the data set. We also excluded from the data set patients whose survival time was equal to zero, or for which the tumor size was unknown or not precisely known. The proposed methodology is capable of evaluating the performance of diagnostic medical tests on the prediction of the cure status of subjects. This breast cancer study aims to assess the predictive performances of these biomarkers.

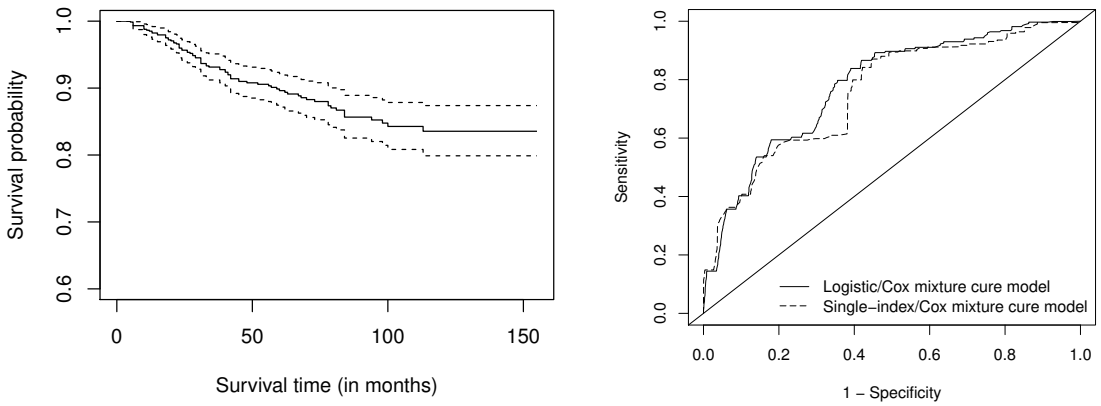

Fig. 5: Left panel: Kaplan-Meier survival curve for the breast cancer data set. Right panel: ROC curve estimates - solid curve: logistic/Cox mixture cure model, dashed curve: single-index/Cox mixture cure model.

The breast cancer data set consists of 753 patients, with a censoring rate of 89.5%. For this data set, tumor size ranges from 1 mm to 160 mm, and the proportion of ER^+ and PR^+ equals 78.5% and 69.1%, respectively. Figure 5 (left panel) shows the Kaplan-Meier survival curve for the breast cancer data set. The curve levels off at around 0.84 and a plateau is observed. There are 133 observations in the plateau. These elements are strong indications for the presence of a cure fraction in this data set.

First, a logistic/Cox mixture cure model is fitted with the three covariates considered in both parts of the model. To assess the predictive performance of the incidence modeling, the ROC curve developed in this paper is considered. The data set is split in a training set containing 502 observations, and in a test set containing 251 subjects. The split is made according to the $2/3 - 1/3$ rule described in Section 3 of the Supplementary Material. The logistic/Cox mixture cure model is estimated on the training set, and the ROC curve is computed on the test set.

Table 2 shows the results of fitting a logistic/Cox mixture cure model, including parameter estimates, standard errors and p-values for the breast cancer data set. At a 5% level, tumor size and PR status have significant impacts on incidence. Patients with larger tumor size and PR^-

Table 2: Parameter estimates, standard errors and p-values of the Wald test for the breast cancer data set based on a logistic/Cox mixture cure model.

	Incidence			Latency		
Covariates	Estimate	Std error	P-value	Estimate	Std error	P-value
Intercept	-1.791	0.345	< 0.001	–	–	–
Tumor size	0.028	0.007	< 0.001	0.003	0.007	0.663
ER status (ER^+ vs ER^-)	-0.284	0.461	0.538	-0.143	0.483	0.767
PR status (PR^+ vs PR^-)	-1.212	0.468	0.010	-0.490	0.491	0.318

breast cancer have a larger probability of being uncured. The effect of a unit (millimeter) increase in tumor size is to increase the odds of being uncured multiplicatively by $\exp(0.028)$, with other covariates being fixed. The odds of being uncured for patients with PR^+ status is $\exp(-1.212)$ times larger than that for patients classified as PR^- status.

To further investigate the predictive performance of these covariates on the cure probability, a ROC curve for the linear combination of these two covariates is computed. Both a logistic/Cox and a single-index/Cox mixture cure model are considered for W_0 and W_1. As can be seen in Figure 5 (right panel), the ROC curves for both approaches are close to each other, and the areas under the curves are almost equivalent. More specifically, the estimated AUCs are 0.778 for the curve based on a logistic/Cox mixture cure model and 0.755 for the one based on a single-index/Cox mixture cure model, respectively. Adopting the naive non-parametric bootstrap method, the estimated confidence interval for the AUC based on the 2.5% and 97.5% sample quantiles of bootstrap estimates is [0.676, 0.867] for the former, and [0.623, 0.848] for the latter. We can then conclude that tumor size and PR status are valuable predictors of the cure status for the considered breast cancer data set.

6. Concluding remarks

Throughout this article, we supposed that M is a linear combination of variables. However, it is possible to extend our proposal to the case where the classifier would be obtained from a different model. Further investigation would be necessary to assess the impact of such a situation on the computation of the weights, but our proposal is not restricted to the linear case. Furthermore, we have considered mixture cure models to compute $\hat{W}_0$ and $\hat{W}_1$. However, a promotion time cure model, such as the model proposed by Tsodikov (1998), could also be considered to estimate a ROC curve for cure status prediction from survival data.

Acknowledgement

The authors acknowledge support from the European Research Council (2016-2021, Horizon 2020 / ERC grant agreement No. 694409).

Supplementary material

The supplementary material available at *Biometrika* online includes the theoretical development of our estimators, the proofs of Theorem 1 and Corollary 1, the simulations for the case of an unknown classifier, as well as further finite sample results.

REFERENCES

AMICO, M. & VAN KEILEGOM, I. (2018). Cure models in survival analysis. *Annual Reviews of Statistics and Its Applications*, **5**, 311–342.

AMICO, M., VAN KEILEGOM, I. & LEGRAND, C. (2019). The single-index/Cox mixture cure model. *Biometrics*, **75**, 452-462.

BERKSON, J. & GAGE, R. P. (1952). Survival curve for cancer patients following treatment. *Journal of the American Statistical Association*, **47**, 501–515.

BEYENE, K. M., EL GHOUCH, A. & OULHAJ, A.(2019). On the validity of time-dependent AUC estimation in the presence of a cure fraction. *Biometrical Journal*, **61**, 1430-1447.

BLANCHE, P., DARTIGUES, J.-F., & JACQMIN-GADDA, H. (2013). Review and comparison of ROC curve estimators for a time-dependent outcome with marker-dependent censoring. *Biometrical Journal*, **55**, 687–704.

BOAG, J. W. (1949). Maximum likelihood estimates of the proportion of patients cured by cancer therapy. *Journal of the Royal Statistical Society* B, **11**, 15–53.

CHAMBLESS, L. E. & DIAO, G. (2006). Estimation of time-dependent area under the ROC curve for long-term risk prediction. *Statistics in Medicine*, **25**, 3474–3486.

COX, D. R. (1972). Regression models and life tables (with Discussion). *J. R. Statist. Soc.* B **34**, 187–220.

FAREWELL, V. T. (1977). A model for binary variable with time-censored observations. *Biometrika*, **64**, 43–46.

FAREWELL, V. T. (1982). The use of a mixture model for the analysis of survival data with long-term survivors. *Biometrics*, **38**, 1041–1046.

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

HARRELL, F. E., CALIFF, R. M., PRYOR, D. B., LEE, K. L. & ROSATI, R. A. (1982). Evaluating the yield of medical tests. *Journal of the American Medical Association*, **247**, 2543–2546.

HARRELL, F. E., LEE, K. L., CALIFF, R. M., PRYOR, D. B. & ROSATI, R. A. (1984). Regression modeling strategies for improved prognostic prediction. *Statistics in Medicine*, **3**, 143–152.

HEAGERTY, P. J., LUMLEY, T., & PEPE, M. S. (2000). Time-dependent ROC curves for censored survival data and a diagnostic marker. *Biometrics*, **56**, 337–344.

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

KRZANOWSKI, W. J. & HAND, D. J. (2009). *ROC Curves for Continuous Data Monographs on Statistics and Applied Probability*. Chapman & Hall/CRC, Boca Raton.

KUK, Y. C. & CHEN, C.-H. (1992). A mixture model combining logistic regression with proportional hazards regression. *Biometrika*, **79**, 531-541.

LAURIE, J. A., MOERTEL, C. G., FLEMING, T. R., WIEAND, H. S., LEIGH, J. E., RUBIN, J., MCCORMACK, G. W., GERSTNER, J. B., KROOK, J. E. & MALLIARD, J. (1989). Surgical adjuvant therapy of large-bowel carcinoma: An evaluation of levamisole and the combination of levamisole and fluorouracil: The North Central Cancer Treatment Group and the Mayo Clinic. *Journal of Clinical Oncology*, **7**, 1447-1456.

LI, L., GREENE, T. AND HU, B. (2018) A simple method to estimate the time-dependent receiver operating characteristic curve and the area under the curve with right censored data. *Statistical Methods in Medical Research*, **27**, 2264–2278.

LU, W. (2008). Maximum likelihood estimation in the proportional hazards cure model. *Annals of the Institute of Statistical Mathematics*, **60**, 545–574.

MOERTEL, C. G., FLEMING, T. R., MACDONALD, J. S., HALLER, D. G., LAURIE, J. A., GOODMAN, P. J., UNGERLEIDER, J. S., EMERSON, W. A., TORMEY, D. C., GLICK, J. H., VEEDER , M. H. & MAILLARD, J. A. (1990). Levamisole and fluorouracil for adjuvant therapy of resected colon carcinoma. *New England Journal of Medicine*, **332**, 352-358.

PENG, Y. & DEAR, K. B. G. (2000). A nonparametric mixture model for cure rate estimation. *Biometrics*, **56**, 237–243.

PEPE, M. S. (2003). *The Statistical Evaluation of Medical Tests for Classification and Prediction.* Oxford Statistical Science Series, Oxford University Press.

SY, J. P. & TAYLOR, J. M. G. (2000). Estimation in a Cox proportional hazards cure model. *Biometrics*, **56**, 227–236.

TAYLOR, J. M. G. (1995). Semi-parametric estimation in failure time mixture models. *Biometrics*, **51**, 899–907.

TSODIKOV, A. (1998). A proportional hazards model taking account of long-term survivors. *Biometrics*, **54**, 1508–1516.

YAKOVLEV, A. Y., TSODIKOV, A. D., & ASSELAIN, B. (1996). Stochastic models of tumor latency and their biostatistical applications. Vol. 1 of Mathematical Biology and Medicine, World Scientific, Singapore.

YU, M., TAYLOR,J. M. G. & SANDLER, H. (2008). Individual prediction in prostate cancer studies using a joint longitudinal survival-cure model. *Journal of the American Statistical Association*, **103**, 178–187.

ZHANG, Y. & SHAO, Y. (2018). Concordance measure and discriminatory accuracy in transformation cure models. *Biostatistics*, **19**, 14–26.

[*Received on* 2 *January* 2017. *Editorial decision on* 1 *April* 2017]