Henry Ford Health System
Henry Ford Health System Scholarly Commons

Public Health Sciences Articles | Public Health Sciences

1-28-2021

The ROC of Cox proportional hazards cure models with application in cancer studies

Yilong Zhang

Xiaoxia Han

Yongzhao Shao

Follow this and additional works at: https://scholarlycommons.henryford.com/publichealthsciences_articles

Lifetime Data Analysis
https://doi.org/10.1007/s10985-021-09516-6

The ROC of Cox proportional hazards cure models with application in cancer studies

Yilong Zhang[1] · Xiaoxia Han[2] · Yongzhao Shao[3]

Received: 12 June 2019 / Accepted: 13 January 2021
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC part of Springer Nature 2021

Abstract

With recent advancement in cancer screening and treatment, many patients with cancers are identified at early stage and clinically cured. Importantly, uncured patients should be treated timely before the cancer progresses to advanced stages for which therapeutic options are rather limited. It is also crucial to identify uncured subjects among patients with early-stage cancers for clinical trials to develop effective adjuvant therapies. Thus, it is of interest to develop statistical predictive models with as high accuracy as possible in predicting the latent cure status. The receiver operating characteristic curve (ROC) and the area under the ROC curve (AUC) are among the most widely used statistical metrics for assessing predictive accuracy or discriminatory power for a dichotomous outcome (cured/uncured). Yet the conventional AUC cannot be directly used due to incompletely observed cure status. In this article, we proposed new estimates of the ROC curve and its AUC for predicting latent cure status in Cox proportional hazards (PH) cure models and transformation cure models. We developed explicit formulas to estimate sensitivity, specificity, the ROC and its AUC without requiring to know the patient cure status. We also developed EM type estimates to approximate sensitivity, specificity, ROC and AUC conditional on observed data. Numerical studies were used to assess their finite-sample performance of the proposed methods. Both methods are consistent and have similar efficiency as shown in our numerical studies. A melanoma dataset was used to demonstrate the utility of the proposed estimates of the ROC curve for the latent cure status. We also have developed an `R` package called `evacure` to efficiently compute the proposed estimates.

Keywords Mixture cure models · Predictive accuracy · Latent cure status · Area under the ROC curve · Sensitivity

Yilong Zhang and Xiaoxia Han have contributed equally to this work.

Extended author information available on the last page of the article

Published online: 28 January 2021

1 Introduction

With recent advancement in cancer screening, diagnosis and treatment, a large portion of melanoma and other cancers are identified at early stage, and many patients are clinically cured and will never experience recurrence, metastasis or death due to the primary cancer (Othus et al. 2012; Peng and Taylor 2013; Aravanis et al. 2017). Importantly, among patients with early-stage cancers, uncured patients should be identified early and treated timely before their cancers progress to advanced stage for which therapeutic options are generally limited. However, there does not exist a simple clinical tool that can timely identify which patient is cured with certainty. Thus, in precision medicine for melanoma and other cancers, it is of interest to develop accurate statistical predictive models to infer the latent cure status. In a population consisting of a mixture of latent cured and uncured patients, the conventional survival models including Cox proportional hazards (PH) models are not capable of identifying patients with high versus low risk of being uncured (Peng and Dear 2000; Sy and Taylor 2000). As an alternative method, mixture cure models have been developed to simultaneously model the cure status and the survival time for a mixture of latent cured and uncured patients in a large body of literature (Farewell 1982; Kuk and Chen 1992; Peng and Dear 2000; Peng and Taylor 2013). The mixture cure models can be used to predict cure status and to evaluate long-term treatment effects of adjuvant therapies (e.g. immunotherapies) for uncured patients in the presence of latent cured ones (Peng and Dear 2000; Peng and Taylor 2013; Jiang et al. 2017; Zhang and Shao 2018). They have been applied to the analysis of survival time for different cancer types, including breast cancer (Pocock et al. 1982; Farewell 1986; Broët et al. 2006; Brown et al. 2008; Yilmaz et al. 2013), multiple myeloma (Crowley et al. 2010), prostate cancer (Kim et al. 2009), colon cancer (Lambert et al. 2010) and melanoma (Kirkwood et al. 1996; Chen et al. 1999; Andersson et al. 2014).

Despite the merits of mixture cure models, they are still underused in clinical practice (Othus et al. 2012). One of the reasons is the existence of major knowledge gaps on evaluating predictive accuracy of cure status. Recently, a great progress has been made by Jiang et al. (2017) who proposed a method to assess the prediction accuracy of cure status for mixture cure models based on an extension of the Brier score (Brier 1950). Importantly, sensitivity and specificity as well as ROC curves are more widely applicable (e.g. in case-control and other settings) than other measures. The AUCs under ROC curves provide a useful overall measure of discriminatory power of binary classifiers. Thus, the ROC curves and AUCs are among the most widely used discriminatory measures of classifiers and predictive scores for binary outcomes. Unfortunately, formulas for sensitivity, specificity, ROC curves and AUCs are not directly applicable in the context of incompletely observed cure status. If we can observe a representative random sample with known cure status, we can use the likelihood score to estimate both sensitivity and specificity. By Neyman–Pearson lemma, the likelihood score can be used to form an optimal ROC curve in the sense of maximizing sensitivity for any given specificity at all the points on the ROC curve (McIntosh and Pepe 2002). In particular, Asano et al. (2014) proposed some naive estimators for sensitivity and specificity for Cox PH cure models by assuming the existence of a known 'cured time', after which all censored patients are considered

cured. However, they did not provide a practical guide on selection of a suitable 'cured time'. In cancer precision medicine and many other applications, the exact 'cured time' is usually unknown and variable from patient to patient. An improper selection of a cutoff point for 'cured time' can introduce serious biases. In addition to this conceptual shortcomings, the method proposed by Asano et al. (2014) generally need further correction for biases and is also computational intensive. To overcome all these challenges, in this paper, we proposed consistent estimates of sensitivity and specificity as well as ROC curves using incompletely observed cure status in the context of Cox PH cure models and other transformation cure models. In turn, the area under the ROC curve (AUC) or partial AUC can also be estimated consistently which provides easy to use and intuitive measures of discriminatory power.

The paper was organized as follows. In Sect. 2, we introduced ROC and its AUC and other prognostic accuracy metrics for cured status within the framework of Cox PH cure models. We developed explicit formulas to estimate sensitivity, specificity, ROC and AUC (and partial AUC). Then, we generalized the results to transformation cure models (Lu and Ying 2004). We also developed an EM type approach to approximate sensitivity, specificity, ROC and AUC conditional on observed data. Both methods are consistent and have similar efficiency as in our numerical studies. The extensive simulation studies to examine the performance of our proposed estimators were reported in Sect. 3. In Sect. 4, we used a melanoma dataset to illustrate the utility of the proposed prognostic metrics. We presented a brief discussion and concluded this paper in Sect. 5. An `R` package `evacure` is available to estimate the proposed metrics.

2 Method

Cox PH cure models explicitly model the overall population as a mixture of cured and uncured patients (Sy and Taylor 2000; Peng and Dear 2000; Peng and Taylor 2013). Let Y denote latent uncure status (uncured $Y = 1$ and cured $Y = 0$). The probability of being uncured $\pi(Z) = Pr(Y = 1 \mid Z)$ is typically modeled with a logistic regression,

$$Pr(Y = 1|Z) = \pi(b^T Z) = \frac{\exp(b^T Z)}{1 + \exp(b^T Z)}, \tag{1}$$

where b is a p-dimensional unknown parameter vector of log odds ratios and Z is the observed baseline covariate vector associated with uncure status. Also, given a q-dimensional covariate vector X, the survival time of uncured patients, $S_u(t \mid X) = S(t \mid Y = 1, X)$ is modeled with Cox PH regression which is defined as

$$S_u(t \mid X) = S_0(t)^{\exp(\beta^T X)}, \tag{2}$$

where $S_0(t)$ denotes the unknown baseline survival function of uncured subjects with $X = 0$ and β denotes the q-dimensional unknown parameter vector. The p-dimensional and q-dimensional covariates X and Z associated with uncure status and survival time, respectively, can share same components. Under the Cox PH cure model, the survival function of the survival time T of a randomly selected patient is a sum of

two parts, which is

$$S_T(t \mid X, Z) = \pi(Z) S_u(t \mid X) + 1 - \pi(Z), \tag{3}$$

where $\pi(Z) = Pr(Y = 1 \mid Z)$ is modeled in (1) and the survival function $S_u(t \mid X)$ is modeled in (2).

Let C denote the censoring time and T the survival time. Let $\tilde{T} = \min(T, C)$ denotes the observed time and $\delta = I(T \leq C)$ is the censoring indicator. It is assumed that given covariates X_i and Z_i for the ith subject, C_i and T_i are independent. The observed dataset $O = \{O_i = (\tilde{T}_i, \delta_i, X_i, Z_i), i = 1, \ldots N\}$ are assumed to be from independent subjects. As is well known, the expectation-maximization (EM) algorithms can be conveniently used to estimate the unknown parameters $\{\beta, b, S_0(t)\}$ (Peng and Dear 2000; Sy and Taylor 2000; Wang et al. 2017). The consistency and asymptotically normality of the maximum likelihood estimates (MLEs) for β and b are established by Fang et al. (2005) under some general regularity conditions.

In the above Cox PH cure model, we can treat the log odds $l = b^T Z$ in Eq. (1) as a risk score for predicting the latent uncure status. Given a cutoff point c, the sensitivity, specificity in terms of the incompletely observed uncure status Y can be defined as

$$\text{sen}(c) = Pr(l > c | Y = 1); \quad \text{spe}(c) = Pr(l < c | Y = 0). \tag{4}$$

The commonly used receiver operating characteristic (ROC) curve is a plot of sensitivity against one minus specificity for all possible cutoff points of a classifier:

$$\text{ROC} = \{(1\text{-spe}(c), \text{sen}(c)), \ c \in \Re\}. \tag{5}$$

The area under the ROC curve (AUC) is an important tool to quantify the overall predictive performance of a classifier (Hanley and McNeil 1982). Partial AUCs are also useful for comparison of ROC curves in many cases.

2.1 A substitution approach

We first provide an approach directly based on the output of the widely used EM algorithm for fitting the Cox PH mixture cure models. Intuitively, the uncure probability, $w_i = Pr(Y_i = 1 \mid O)$ is a reasonable surrogate for the incompletely observed uncure status Y_i. It is estimated in order to proceed the E-step in the EM algorithm while fitting the Cox PH cure model. We propose to estimate sensitivity, specificity and AUC by replacing the unknown uncure status with the estimate of w_i. For the rest of the paper, we referred this method as the substitution approach.

Without loss of generality, we denote the baseline survival function $S_0(\cdot) = S_0(\cdot, \nu)$ where ν is the parameter vector for the baseline survival function. Note that ν can be finite-dimensional for a parametric Cox PH model (Farewell 1986; Peng et al. 1998; Peng and Carriere 2002) or infinite-dimensional for a semiparametric Cox PH model (Peng and Dear 2000; Sy and Taylor 2000). Let model parameters be $\theta = (\beta, b, \nu)$,

and the estimate of w_i can be written as (Wang et al. 2017)

$$w_i(\hat{\theta}) = \delta_i + (1-\delta_i)\frac{\pi(\hat{b}^T Z_i) S_0(t, \hat{\nu})^{\exp(\hat{\beta}^T X_i)}}{1 - \pi(\hat{b}^T Z_i) + \pi(\hat{b}^T Z_i) S_0(t, \hat{\nu})^{\exp(\hat{\beta}^T X_i)}}. \tag{6}$$

Therefore, the estimators of sensitivity and specificity can be written as

$$\widehat{\text{sen}}_n(c, \hat{\theta}) = \frac{\sum_{i=1}^n w_i(\hat{\theta}) I(\hat{b}^T Z_i > c)}{\sum_{i=1}^n w_i(\hat{\theta})}, \tag{7}$$

$$\widehat{\text{spe}}_n(c, \hat{\theta}) = \frac{\sum_{i=1}^n (1 - w_i(\hat{\theta})) I(\hat{b}^T Z_i < c)}{\sum_{i=1}^n (1 - w_i(\hat{\theta}))}, \tag{8}$$

where c is the cutoff point. It is worth noting that Y_i is known for subjects who have encountered the events (i.e. $w_i(\hat{\theta}) = \delta_i = Y_i = 1$). Only censored subjects with unknown uncure status Y_i need to be imputed.

By letting the cutoff point c vary, the ROC curve $\text{ROC} = \{(1\text{-spe}(c), \text{sen}(c)),\ c \in \Re\}$ can also be estimated accordingly, by

$$\widehat{\text{ROC}}_n = \left\{ \left(1 - \widehat{\text{spe}}_n(c, \hat{\theta}),\ \widehat{\text{sen}}_n(c, \hat{\theta})\right),\ c \in \Re \right\}. \tag{9}$$

For two randomly selected independent subjects with observed data O_1 and O_2, the AUC definition is well-known to be equivalent to $AUC = Pr(l_1 > l_2 \mid Y_1 = 1;\ Y_2 = 0)$. Thus, AUC can be estimated by

$$\widehat{\text{AUC}}_n(\hat{\theta}) = \frac{\sum_{i \neq j} I(\hat{b}^T Z_i > \hat{b}^T Z_j) w_i(\hat{\theta})(1 - w_j(\hat{\theta}))}{\sum_{i \neq j} w_i(\hat{\theta})(1 - w_j(\hat{\theta}))}. \tag{10}$$

Similarly, using θ to replace $\hat{\theta}$ in the above formulas, we can obtain $\widehat{\text{sen}}_n(c, \theta)$, $\widehat{\text{spe}}_n(c, \theta)$, and $\widehat{\text{AUC}}_n(\theta)$ as functions of θ from the above three Eqs. (7), (8), and (10), respectively.

Next, we investigate the consistency and asymptotic normality of the sensitivity, specificity and AUC estimates under parametric Cox PH cure models that assume ν is in some finite-dimensional set. Under the assumptions (A1)–(A4) in the "Appendix I", the estimators $\widehat{\text{sen}}_n(c, \hat{\theta})$, $\widehat{\text{spe}}_n(c, \hat{\theta})$ and $\widehat{\text{AUC}}_n(\hat{\theta})$ are consistent to $\text{sen}(c)$, $\text{spe}(c)$ and AUC, respectively, and these estimators are asymptotically normal. Recall $O_i = (\tilde{T}_i, \delta_i, X_i, Z_i),\ i = 1, \ldots, n$ denote the observed data that are independent copies of $O = (\tilde{T}, \delta, X, Z)$.

Proposition 1 *Under the condition (A1)–(A4) in the "Appendix I", denote* $\sigma^2_{sen,\theta} = n \cdot var(\widehat{sen}_n(c, \hat{\theta}))$, $\sigma^2_{spe,\theta} = n \cdot var(\widehat{spe}_n(c, \hat{\theta}))$ *and* $\sigma^2_{AUC} = n \cdot var(\widehat{AUC}_n(\hat{\theta}))$, *in*

distribution we have

$$n^{1/2}\{\widehat{sen}_n(c,\hat{\theta}) - sen(c)\}/\sigma_{sen,\theta} \to \mathcal{N}(0,1),$$
$$n^{1/2}\{\widehat{spe}_n(c,\hat{\theta}) - spe(c)\}/\sigma_{spe,\theta} \to \mathcal{N}(0,1),$$
$$n^{1/2}\{\widehat{AUC}_n(\hat{\theta}) - AUC\}/\sigma_{AUC} \to \mathcal{N}(0,1).$$

For parametric Cox PH cure model (Farewell 1986; Peng et al. 1998), condition (A1) holds by noticing that $\hat{\theta}$ is a regular asymptotically linear (RAL) estimator of θ (Tsiatis 2007). Condition (A2) ensures both cured and uncured subjects have positive prevalence in the population. Conditions (A3) and (A4) are technical condition for local linear expansion of the estimators. A similar condition was used in Gönen and Heller (2005) and Zhang and Shao (2018). The proof of Proposition 1 and the plug-in estimator of variances are available in "Appendix I". In practice, given the explicit formulas, the variance and confidence interval of $\widehat{\text{sen}}_n(c,\hat{\theta})$, $\widehat{\text{spe}}_n(c,\hat{\theta})$ and $\widehat{\text{AUC}}_n(\hat{\theta})$ can be estimated empirically using bootstrap. Clearly, the pointwise consistency of ROC curve can be deduced from the above proposition. Moreover, our proposed approach can be used to compare predictive accuracy of two models based on two sets of possibly correlated covariates, e.g., one contains an extra new biomarker and the other only contains an existing set of covariates. The comparison can be made based on comparison of the two ROC curves and the difference of the two estimated AUCs or partial AUC. Consistent estimates of the difference between the two AUCs or partial AUC can be obtained from the above proposition. Importantly, given the proposed explicit formulas for AUCs, it is convenient and computationally efficient to construct 95% confidence intervals for the differences of the two AUCs using bootstrap (Liu and Jin 2009; Uno et al. 2011; Liu et al. 2016; Han et al. 2017). The numerical performance of our proposed estimators in finite sample sizes for Cox PH cure model was investigated in Sect. 3.

We further consider the estimation of sensitivity, specificity and AUC beyond Cox PH cure models as the proportional hazards assumption does not always hold for uncured patients. For example, in recent immunology studies (Gandhi et al. 2018), non-proportional hazards scenarios were observed. Transformation mixture cure models include many widely used non-proportional hazards models (Lu and Ying 2004). Specifically, transformation mixture cure models assume the failure time of uncured patients follows a transformation model

$$H(T) = -X\beta + \epsilon. \tag{11}$$

Therefore, for the transformation mixture cure models, the survival function of uncured patients, $S_u(t \mid X)$ can be written as

$$S_u(t \mid X) = Pr(\epsilon > H(t) + \beta^T X). \tag{12}$$

$H(\cdot)$ is a monotone increasing function and ϵ is the error term with a continuous distribution. If $H(\cdot)$ is unknown and ϵ has the extreme value distribution with $S_\epsilon(t) = \exp\{-\exp(t)\}$, then the model (12) is the Cox PH model that is equivalent to the

formula (2). If $H(\cdot)$ is unknown and ϵ has the logistic distribution with $S_\epsilon(t) = 1/\{1+\exp(t)\}$, then the model (12) is the proportional odds model. If $H(\cdot) = \log(\cdot)$, the above model (12) is the accelerated failure time (AFT) model.

For the transformation mixture cure models, the estimator of unknown uncured status, w_i can be written as

$$w_i(\hat{\theta}) = \delta_i + (1-\delta_i)\frac{\pi(\hat{b}^T Z_i)\hat{S}_u(t \mid X)}{1-\pi(\hat{b}^T Z_i)+\pi(\hat{b}^T Z_i)\hat{S}_u(t \mid X)}, \tag{13}$$

where $\hat{S}_u(t \mid X)$ is the estimator of $S_u(t \mid X)$ under specific model assumption of $H(\cdot)$ and distribution of ϵ. Therefore, $\widehat{\text{sen}}_n(c,\hat{\theta})$, $\widehat{\text{spe}}_n(c,\hat{\theta})$ and $\widehat{\text{AUC}}_n(\hat{\theta})$ can be estimated by using Eqs. (7), (8) and (10). The consistent and asymptotically normal estimates for the transformation model (11) can be obtained based on martingale estimating equations (Chen et al. 2002; Lu and Ying 2004) maximum likelihood estimates (Zeng and Lin 2007), or the maximum rank correlation estimates with self-induced smoothing (Zhang et al. 2018). For parametric transformation mixture cure model, when the conditions (A1)–(A4) hold, following the similar argument for Cox PH cure model, by Proposition 1, $\widehat{\text{sen}}_n(c,\hat{\theta})$, $\widehat{\text{spe}}_n(c,\hat{\theta})$ and $\widehat{\text{AUC}}_n(\hat{\theta})$ are consistent and asymptotic normal.

2.2 An EM type approach

In the preceding subsection, we have introduced consistent estimators of sensitivity, specificity, ROC curve as well as the AUC under the ROC in evaluating the predictive performance of the log-odds based predictive score $l = b^T Z$ for the latent cure status. The proposed approach is applicable in the presence of missing cure status Y due to the common data censoring in the Cox PH cure models. The proposed substitution approach in Sect. 2.1 estimates have explicit formulas directly using the output of the widely used EM algorithm for fitting the Cox PH mixture cure models and thus they are computational efficiently.

As suggested by the Associate Editor, alternative approaches should also be discussed and compared with the preceding substitution approach. In this subsection, we investigated an EM type approach that directly approximates the expectation of sensitivity, specificity, ROC curve as well as the AUC (and partial AUC) conditional on observed data. Specifically, we proved the consistency of the new estimates in "Appendix II", provided an EM type algorithm to calculate the estimates, and compared their performance with the substitution approach-based estimators introduced in Sect. 2.1. For brevity, we focus on exposition of estimating sensitivity because the basic idea and method for the approximate of sensitivity, specificity, ROC, and AUC are all the same.

Given a cutoff point c, the sensitivity $\text{sen}(c)$ as defined in Eq. (4) is

$$\text{sen}(c) = Pr(l > c|Y=1) = Pr(b^T Z > c|Y=1).$$

In the classical case where we have a training set with completely observed $\{(Z_i, Y_i), i = 1, \ldots, n\}$, given $n_1 = \sum Y_i$, i.e. the number of $\{Y_i = 1\}$, the sensitivity $\text{sen}(c)$ can be estimated using the sample proportion

$$\text{sen}_n(c) = \frac{\sum_{i=1}^n I(l_i > c, Y_i = 1)}{n_1} = \frac{\sum_{i=1}^n I(b^T Z_i > c) Y_i}{\sum_{i=1}^n Y_i}. \tag{14}$$

As is well known, given n_1, the sample proportion $\text{sen}_n(c)$ is an UMVUE for estimating the population proportion $\text{sen}(c)$ with i.i.d. observations $\{(Z_i, Y_i), i = 1, \ldots, n\}$. When we have incomplete observations on Y_i's in $\{(Z_i, Y_i), i = 1, \ldots, n\}$ due to censoring, it is natural to consider estimating $\text{sen}(c)$ using

$$\widetilde{\text{sen}}_n(c) = E(\text{sen}_n(c)|O), \tag{15}$$

where O is the sigma-algebra generated by the observed data.

$Y_i \mid O$ has a Bernoulli distribution with probability $w_i = P(Y_i = 1 \mid O_i)$ and w_i can be approximated by Eq. (6). For unobserved Y_i's, we can directly draw samples from the Bernoulli distribution to approximate $\widetilde{\text{sen}}_n(c) = E(\text{sen}_n(c)|O)$. The EM type estimate of $\widetilde{\text{sen}}_n(c)$ is worth of investigation as it can be potentially advantageous in terms of having smaller variance and mean squared error (MSE) compared to the newly proposed substitution approach. In "Appendix II", we show that the EM type estimate $\widetilde{\text{sen}}_n(c)$ is approximately equal to the substitution approach-based estimate $\widehat{\text{sen}}_n(c)$ in Eq. (7). Therefore, by Slutsky's lemma, both $\widetilde{\text{sen}}_n(c)$ and $\widehat{\text{sen}}_n(c)$ are consistent because the consistency of $\widehat{\text{sen}}_n(c)$ has already been established in the preceding subsection.

The steps to approximate $\widehat{\text{sen}}_n(c)$ are summarized below:

- Step 1: Calculate $\hat{\theta}$ from the EM-algorithm for Cox PH cure model.
- Step 2: Calculate $w_i(\hat{\theta})$ using Eq. (6).
- Step 3: For unknown Y_i (when $\delta_i = 0$), draw sample from the Bernoulli distribution with probability $w_i(\hat{\theta})$ denoted as $Y_i^{(m)}$.
- Step 4: For various cutoff point c, estimate the sensitivity, specificity, and ROC using sample proportion based on $\{Y_i^{(m)}, \ \hat{b}^T Z_i\}$. (For simplicity, we denote $Y_i^{(m)} = Y_i$ for observed Y_i.)

$$\text{sen}_n^{(m)}(c) = \frac{\sum_{i=1}^n Y_i^{(m)} I(\{b^{(m)}\}^T Z_i > c\})}{\sum_{i=1}^n Y_i^{(m)}},$$

$$\text{spe}_n^{(m)}(c) = \frac{\sum_{i=1}^n (1 - Y_i^{(m)}) I(\{b^{(m)}\}^T Z_i < c\})}{\sum_{i=1}^n (1 - Y_i^{(m)})},$$

$$\text{ROC}_n^{(m)} = \left\{ (1 - \text{spe}_n^{(m)}(c), \ \ \text{sen}_n^{(m)}(c)) : c \in \Re \right\}.$$

- Step 5: Repeat steps 3 and 4 M times for a large integer M. Then obtain the sample mean over the M iterations as estimates of sensitivity, specificity and ROC as below

$$\widetilde{\text{sen}}_n(c) = \frac{1}{M}\sum_{m=1}^{M} \text{sen}_n^{(m)}(c), \tag{16}$$

$$\widetilde{\text{spe}}_n(c) = \frac{1}{M}\sum_{m=1}^{M} \text{spe}_n^{(m)}(c), \tag{17}$$

$$\widetilde{\text{ROC}}_n = \left\{(1 - \text{spe}_n(c), \quad \text{sen}_n(c)) : c \in \Re\right\}, \tag{18}$$

Once we have the EM type estimates of the ROC, the AUC and partial AUC can also be obtained numerically in a straightforward manner. Bootstrap can be used to obtain the standard deviation of the corresponding estimates. However, the iterative sampling and numerical evaluation of conditional expectation are needed within each Bootstrap samples, which is computationally intensive compared with the substitution approach in Sect. 2.1.

3 Simulation studies

We conducted extensive simulation studies to verify the validity of the proposed estimators for prognostic metrics using Weibull mixture cure models (Farewell 1986) with various cured and censoring proportions. We also conducted extensive simulation studies using the EM type approach in Sect. 2.2 and compared its performance to the substitution approach in Sect. 2.1. We considered regression setting with a continuous covariate generated from a standard normal distribution and a binary covariate generated from a Bernoulli distribution with a 'success' probability 0.5. Both covariates had effect on cure probability and event time among uncured patients. Censoring time was generated from uniform distribution $U(0, u)$, where constant u was chosen to obtain the desirable censoring proportion p_c. The probability of being uncured was generated from a logistic model and event time for uncured patients was generated from a Weibull model. We conducted simulation studies with three scenarios (S1-S3) using different censoring rate p_c and cure rate p_{cu} as shown in Tables 1 and 2. The estimates of sensitivity, specificity, AUC were based on sample size $n = 200$ with 1000 simulations with $B = 200$ bootstrap samples. Given the true AUC, we determined the corresponding coefficients $(b_1, b_2, \beta_1, \beta_2)$ that satisfied the targeted setting for censoring proportion and cure rate. For all simulation configurations, semiparametric Cox PH cure models were used to analyze the data and prognostic metrics were calculated based on the fitted models. For each simulated training data, we also simulated a test dataset to estimate the sensitivity, specificity, and AUC using the semiparametric Cox PH cure model estimated from the training set.

Tables 1, 2 and 3 summarize simulation results for three scenarios (S1–S3). Table 1 presents simulation results for the substitution approach in Sect. 2.1 based on Eqs. (7), (8) and (10). Table 2, presents the simulation results of the EM type approach in Sect. 2.2 based on Eqs. (16), (17), and area under the estimated ROC (18) with $M = 500$. In Tables 1 and 2, under scenario S1, with censoring proportion $p_c = 60\%$ and cured proportion $p_{cu} = 50\%$ (i.e. 10% of the uncured patients are censored), $\widehat{\text{AUC}}_n(\hat{\theta})$

was relatively close to true value when sample size $n = 200$. We also observed that the estimates of AUC were comparable in training and test data sets. When we increased the censoring proportion in uncured patients to 30% by either decreasing the cured proportion p_{cu} to 30% in S2 or increasing the censoring proportion p_c to 80% in S3, the standard deviation of $\widehat{\text{AUC}}_n(\hat{\theta})$ became larger. The estimates of sensitivity and specificity with different cutoff points were also close to the true value in all three scenarios for both training and test data sets. The difference in variance for both substitution approach and EM type approach turned out to be extremely small as shown in Tables 1 and 2.

4 Application to an NYU IMCG melanoma dataset

In this section, we applied the proposed methods to a cohort of melanoma patients enrolled in the New York University Interdisciplinary Melanoma Cooperative Group (NYU IMCG) Registry (Wich et al. 2009; Qian et al. 2018) between 2002 and 2009 with follow-up until 2013. The covariates included age at diagnosis, primary tumor thickness (mm), primary tumor ulceration (present vs absent), primary tumor mitotic index (present vs absent), and primary tumor anatomic site (extremity vs axial/head and neck). There were 1164 patients with available data from this cohort as previously described (Cymerman et al. 2016). To focus on early stage cancer, we included 960 stage I and II patients for analyses. The response variable was recurrence free survival. The median follow up time was 3.5 years and overall censoring proportion was 87.2%.

We fitted a Cox PH mixture model to take latent cured patients into consideration. The model estimates, 95% confidence intervals, and p values were summarized in Table 4. Primary tumor ulceration status (Yes/No) was statistically significant in the logistic regression component for predicting uncure status. The probability of being uncured was higher among patients whose primary tumor had ulceration (OR=2.79, P=0.010). Primary tumor thickness and primary tumor mitotic index were statistically significant in Cox PH component among uncured subjects. Shorter survival was expected for patients with thicker tumor (HR=1.19, P=0.008), and patients with mitotic index in their primary tumor (HR=1.92, P=0.043). Compared with the ordinary Cox PH model, the Cox PH mixture cure model allowed us to extract the cure information and identify primary tumor ulceration as factors associated with cure status. This was largely in agreement of the finding using a dataset of 205 stage I melanoma patients at Plastic Surgery Unit in Odense (Zhang 2016) and section 3.3 of Zhang and Shao (2018).

We further evaluated the prognostic accuracy of the fitted model. AUC for predicting cure status is estimated to be 0.72 (95% CI 0.69–0.81). It is known that predicting cure status among patients with early stage melanoma after removing primary tumor is quite challenging. Nevertheless, this example illustrates that we can assess the prognostic utility of log-thickness and ulceration without the need to know who is surely cured. The predictive accuracy of such prognostic models might be further improved by including other biomarkers such as levels of circulating microRNA (Friedman et al. 2012) and ctDNA (Chang et al. 2016), clinicopathologic characteristics (Cymerman

Table 1 Estimates of sensitivity, specificity, and AUC in Cox PH cure model using the substitution approach in Sect. 2.1

Estimators	True values	Estimates	Empirical SD	Bootstrap SD	RMSE	Estimates from test data
S1: $p_c = 60\%$, $p_{cu} = 50\%$, $n = 200$						
AUC	0.866	0.869	0.029	0.030	0.029	0.868
$sen_{25}\%$	0.952	0.953	0.021	0.021	0.021	0.952
$sen_{50}\%$	0.781	0.778	0.035	0.037	0.035	0.766
$sen_{75}\%$	0.452	0.445	0.029	0.036	0.029	0.428
$spe_{25}\%$	0.456	0.459	0.034	0.042	0.034	0.475
$spe_{50}\%$	0.783	0.793	0.041	0.043	0.042	0.803
$spe_{70}\%$	0.954	0.959	0.020	0.020	0.020	0.957
S2: $p_c = 60\%$, $p_{cu} = 30\%$, $n = 200$						
AUC	0.868	0.873	0.037	0.038	0.037	0.870
$sen_{25}\%$	0.895	0.897	0.023	0.023	0.023	0.892
$sen_{50}\%$	0.654	0.649	0.026	0.029	0.026	0.638
$sen_{75}\%$	0.341	0.333	0.016	0.025	0.017	0.327
$spe_{25}\%$	0.616	0.625	0.070	0.079	0.071	0.635
$spe_{50}\%$	0.886	0.894	0.049	0.049	0.050	0.891
$spe_{75}\%$	0.981	0.980	0.014	0.014	0.014	0.980
S3: $p_c = 80\%$, $p_{cu} = 50\%$, $n = 200$						
AUC	0.866	0.87	0.043	0.041	0.043	0.866
$sen_{25}\%$	0.952	0.952	0.023	0.022	0.022	0.951
$sen_{50}\%$	0.779	0.776	0.039	0.038	0.039	0.765
$sen_{75}\%$	0.452	0.44	0.032	0.035	0.034	0.434
$spe_{25}\%$	0.458	0.465	0.054	0.056	0.054	0.468
$spe_{50}\%$	0.783	0.798	0.062	0.060	0.064	0.796
$spe_{75}\%$	0.955	0.957	0.027	0.024	0.027	0.957

pc censoring rate; *pcu* cure rate, *n* sample size

Table 2 Estimates of sensitivity, specificity, and AUC in Cox PH cure model using the EM type approach in Sect. 2.2

Estimators	True values	Estimates	Empirical SD	Bootstrap SD	RMSE	Estimates from test data
S1: $p_c = 60\%$, $p_{cu} = 50\%$, $n = 200$						
AUC	0.866	0.869	0.029	0.030	0.029	0.868
$sen_{25}\%$	0.952	0.953	0.021	0.021	0.021	0.952
$sen_{50}\%$	0.781	0.783	0.035	0.037	0.035	0.771
$sen_{75}\%$	0.452	0.453	0.029	0.036	0.029	0.436
$spe_{25}\%$	0.456	0.460	0.034	0.042	0.034	0.476
$spe_{50}\%$	0.783	0.789	0.041	0.043	0.041	0.799
$spe_{70}\%$	0.954	0.957	0.021	0.020	0.021	0.956
S2: $p_c = 60\%$, $p_{cu} = 30\%$, $n = 200$						
AUC	0.868	0.873	0.037	0.038	0.037	0.870
$sen_{25}\%$	0.895	0.897	0.023	0.024	0.023	0.892
$sen_{50}\%$	0.654	0.654	0.026	0.029	0.026	0.644
$sen_{75}\%$	0.341	0.34	0.016	0.025	0.016	0.334
$spe_{25}\%$	0.616	0.627	0.071	0.079	0.071	0.636
$spe_{50}\%$	0.886	0.892	0.050	0.049	0.050	0.889
$spe_{75}\%$	0.981	0.98	0.014	0.015	0.014	0.980
S3: $p_c = 80\%$, $p_{cu} = 50\%$, $n = 200$						
AUC	0.866	0.870	0.043	0.041	0.043	0.866
$sen_{25}\%$	0.952	0.952	0.023	0.022	0.023	0.951
$sen_{50}\%$	0.779	0.781	0.038	0.038	0.038	0.770
$sen_{75}\%.$	0.452	0.449	0.032	0.035	0.032	0.442
$spe_{25}\%$	0.458	0.466	0.054	0.057	0.054	0.469
$spe_{50}\%$	0.783	0.794	0.062	0.060	0.063	0.792
$spe_{75}\%$	0.955	0.956	0.028	0.024	0.028	0.955

p_c censoring rate; p_{cu} cure rate, n sample size

Table 3 Simulation results for model parameters in Cox PH cure model

Scenario	p_c	p_{cu}	$\hat{b}_1$	$\hat{b}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$
S1	60	50	1.06	− 1.03	− 1.02	1.02
S2	60	30	1.09	− 1.08	− 1.02	1.02
S3	80	50	1.09	− 1.08	− 1.04	1.03

p_c censoring rate; p_{cu} cure rate. True value: $b_1 = 1, b_2 = -1, \beta_1 = -1, \beta_2 = 1$

Table 4 Cox PH cure model regression results for the NYU melanoma data

Variables	Logistic regression OR (95% CI)	P	Cox PH HR (95% CI)	P
Age, years	1.01 (0.99, 1.03)	0.483	1.02 (0.99, 1.05)	0.164
Primary tumor thickness, mm	1.21 (0.97, 1.51)	0.089	1.19 (1.05, 1.35)	0.008
Primary tumor ulceration	2.79 (1.28, 6.08)	0.010	1.65 (0.79, 3.44)	0.179
Primary tumor mitotic index	1.33 (0.67, 2.61)	0.413	1.92 (1.02, 3.62)	0.043
Primary tumor anatomic site	0.95 (0.60, 1.49)	0.811	0.82 (0.50, 1.36)	0.447

et al. 2016), and treatment profiles (Sun et al. 2016). If clinicians want to design a trial to investigate new adjuvant therapies for these patients with early stage melanoma, simple random sampling may require a larger number of subjects to detect new treatment efficacy as there are non-negligible cured patients in this cohort who are not informative for treatment efficacy. By estimating each patient's cure probability after surgery, clinicians can oversample patients with high uncure probability and undersample patients with low uncured probability to reduce the sample size of a trial to evaluate new treatment efficacy. Thus, precious trial resources can be allocated to patients with high uncure probability, and the trial may require less patients to achieve the same or even higher statistical power. In addition, this can reduce the likelihood of exposing cured patients to toxicity and drug-related adverse effects.

5 Discussion

The rapid advancement of cancer screening and treatments result in an increased proportion of patients being cured. Although mixture cure models are increasingly discussed in applied statistical literature, they are still underused in clinical applications. One reason is that there is no rigorously defined metrics available to assess the accuracy for such predictive models.

In this paper, we have filled in this knowledge gap by developing consistent and asymptotic normal estimates for sensitivity, specificity and AUC to assess predictive accuracy of cure status for Cox PH cure mode and other mixture cure models. Note that we have obtained pointwise consistent estimates of the ROC curve, thus we can also consistently estimate partial AUCs in addition to the overall AUC. Moreover, our proposed approach can also be used to compare predictive accuracy of different

sets of possibly correlated covariates, which is crucial for personalized medicine. The comparison can be made based on the difference of their estimated AUCs. Consistent estimates of the difference between the two AUCs can be obtained using our results and it is straightforward to construct 95% confidence interval for the differences of the two AUCs using bootstrap (Liu and Jin 2009; Uno et al. 2011; Liu et al. 2016; Han et al. 2017). We also extend the results to transformation mixture cure models. Simulation studies showed that both the substitution approached-based estimators and the EM type estimators performed well in finite sample size in Cox PH cure models. The substitution approach-based estimators are computational efficient compared with the EM type estimators. It is not clear whether there exists a situation where the EM type approach $\widetilde{\text{sen}}_n(c)$ meaningfully outperform the substitution approach-based estimators $\widehat{\text{sen}}_n(c)$ with non-negligible margin of variance estimation.

In this paper, we used a real dataset, an early stage melanoma cohort, to demonstrate the utility of our proposed method in cancer studies. Our method can also be used for other types of cancer. Therapies for a number of cancer types, including breast cancer (Asano et al. 2014), prostate cancer (Chen and Kim 2009), melanoma (Weber et al. 2017), gliomas (Sun et al. 2019), colon cancer (Sargent et al. 2009) and myeloid leukemia (Othus et al. 2012) are believed to induce a non-neglectable proportion of cured patients. In such cases, Cox PH cure models are preferable to the conventional Cox PH models, as they take the unobservable cured patients into consideration. By identifying uncured patients, clinicians may assign adjuvant therapy and these uncured patients may survive longer with appropriate treatments or follow-up surgeries.

An increased use of cure models for clinical trial design is foreseen due to the increasing patients have been cured after surgery. Patricia Bernardo and Ibrahim (2000) proposed group sequential designs for cure models with early stopping in favor of the null hypothesis. Psioda and Ibrahim (2018) developed a general Bayesian clinical trial design with a cured fraction using previous completed clinical trial information. Predicting cure probability after surgery is crucial for cost-effective study design. Clinicians may want to design clinical trials to investigate novel therapeutics strategies for uncured patients with cancer. With accurate estimation of each patient's cure probability after surgery, clinicians can oversample patients with high uncure probability and undersample patients with low uncured probability to increase the power for evaluating new treatment efficacy. Furthermore, it is desirable to evaluate efficacy for different subgroups of uncured patients (i.e., with different sets of characteristics or covariates) without requiring the existence of a gold standard to identify which patients is surely cured. This is important for precision medicine or personalized medicine.

A supplementary R package `evacure` is available to estimate the sensitivity, specificity, and AUC with bootstrap confidence interval for Cox PH cure models and transformation mixture cure model at github https://github.com/elong0527/evacure. Our new metrics and R package `evacure` will facilitate the usage of mixture cure models in clinical practice.

Acknowledgements This research was supported by the National Institutes of Health Grants P50CA225450, P30CA16087, P30AG008051 and P30AG066512. The authors would like to thank Dr. Iman Osman and NYU IMCG for the melanoma dataset used in the manuscript. The authors would like to thank the reviewer and associate editor for constructive suggestions that lead to improved quality of the paper.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Appendix I

We prove Proposition 1 in Appendix I. First, we list the four assumptions required to prove the Proposition 1.

- (A1) There exists a square-integrable random vector $U(O)$ such that we have $E\{U(O)\} = 0$ and $E\{UU^T\}$ is nonsingular, and

$$n^{1/2}(\hat{\theta} - \theta) = n^{-1/2} \sum_{i=1}^{n} U(O_i) + o_p(1). \tag{19}$$

- (A2) A positive proportion of the sample is uncured, i.e. $\sum_{i=1}^{n} \pi_i(b)/n \to_P c_0$ as $n \to \infty$, where $0 < c_0 < 1$ is a constant.
- (A3) Asymptotic linearity of $\widehat{\text{sen}}_n(c, \hat{\theta})$ and $\widehat{\text{spe}}_n(c, \hat{\theta})$: in a neighborhood of true parameter θ, $\partial E\left\{\widehat{\text{sen}}_n(c, \theta)\right\}/\partial\theta$ and $\partial E\left\{\widehat{\text{spe}}_n(c, \theta)\right\}/\partial\theta$ exist, and

$$\widehat{\text{sen}}_n(c, \hat{\theta}) = \widehat{\text{sen}}_n(c, \theta) + D_1(c, \theta)^T(\hat{\theta} - \theta) + o_p(\hat{\theta} - \theta), \tag{20}$$

$$\widehat{\text{spe}}_n(c, \hat{\theta}) = \widehat{\text{spe}}_n(c, \theta) + D_2(c, \theta)^T(\hat{\theta} - \theta) + o_p(\hat{\theta} - \theta), \tag{21}$$

  where $D_1(c, \theta) = \lim_{n\to\infty} \partial E\left\{\widehat{\text{sen}}_n(c, \theta)\right\}/\partial\theta$ is assumed to exist and we have similarly $D_2(c, \theta) = \lim_{n\to\infty} \partial E\left\{\widehat{\text{spe}}_n(c, \theta)\right\}/\partial\theta$.
- (A4) Asymptotic linearity of $\widehat{\text{AUC}}_n(\hat{\theta})$: in a neighborhood of true parameter θ

$$\widehat{\text{AUC}}_n(\hat{\theta}) = \widehat{\text{AUC}}_n(\theta) + D(\theta)^T(\hat{\theta} - \theta) + o_p(\hat{\theta} - \theta), \tag{22}$$

  where $D(\theta) = \lim_{n\to\infty} \partial E\left\{\widehat{\text{AUC}}_n(\theta)\right\}/\partial\theta$.

Based on the condition (A1)–(A4), we prove the consistency and asymptotic normality of $\widehat{\text{AUC}}$ in Proposition 1. The proof for the sensitivity and specificity follow the same idea and omit here. Let $A_n(\theta)$ and $B_n(\theta)$ be

$$A_n(\theta) = \frac{2}{n(n-1)} \sum_{i<j} I(\hat{l}_i > \hat{l}_j)\hat{w}_i(1 - \hat{w}_j),$$

and

$$B_n(\theta) = \frac{2}{n(n-1)} \sum_{i<j} \hat{w}_i(1 - \hat{w}_j),$$

where $\theta = (\beta, b, \nu)$. It is clear that

$$\widehat{\mathrm{AUC}}_n(\hat{\theta}) = \frac{A_n(\hat{\theta})}{B_n(\hat{\theta})}.$$

Let $A = Pr(l_1 > l_2,\ Y_1 = 1,\ Y_2 = 0)$ and $B = Pr(Y_1 = 1,\ Y_2 = 0)$. The property of multivariate U-statistics implies that $(A_n(\theta), B_n(\theta))$ is an unbiased estimator of (A, B) and asymptotically normal (Götze 1987). Taylor's expansion of AUC can be written as

$$\begin{aligned}\widehat{\mathrm{AUC}}_n(\theta) &= \frac{A}{B} + \frac{1}{B}(A_n(\theta) - A) - \frac{A}{B^2}(B_n(\theta) - B) + o_p(A_n(\theta) - A) + o_p(B_n(\theta) - B) \\ &= \mathrm{AUC} + \frac{1}{B}(A_n(\theta) - A) - \frac{A}{B^2}(B_n(\theta) - B) + o_p(n^{-1/2}). \end{aligned} \tag{23}$$

By conditions (A1), (A4) and Central Limit Theorem, we have,

$$\begin{aligned}\widehat{\mathrm{AUC}}_n(\hat{\theta}) - \widehat{\mathrm{AUC}}_n(\theta) &= D(\theta)^T(\hat{\theta} - \theta) + o_p(\hat{\theta} - \theta) \\ &= \frac{1}{n} D^T(\theta) \sum_{i=1}^{n} U(O_i) + o_p(n^{-1/2}). \end{aligned} \tag{24}$$

By (23) and (24), we have

$$\widehat{\mathrm{AUC}}_n(\hat{\theta}) - \mathrm{AUC} = \frac{1}{n} D^T(\theta) \sum_{i=1}^{n} U(O_i) + \frac{1}{B}(A_n(\theta) - A) - \frac{A}{B^2}(B_n(\theta) - B) + o_p(n^{-1/2}).$$

By condition (A2) and unbiasedness of $(A_n(\theta), B_n(\theta))$, we have

$$n^{1/2} E\left(\widehat{\mathrm{AUC}}_n(\hat{\theta}) - \mathrm{AUC}\right) = o(1).$$

Denote

$$\begin{aligned} u_{ij}(\theta) =& D^T(\theta)\{U(O_i) + U(O_j)\} + \\ & \frac{1}{B} I(\hat{l}_i > \hat{l}_j)\hat{w}_i(1 - \hat{w}_j) + \frac{A}{B^2}\hat{w}_i(1 - \hat{w}_j). \end{aligned}$$

We have

$$\widehat{\mathrm{AUC}}_n(\hat{\theta}) - \mathrm{AUC} = \frac{2}{n(n-1)} \sum_{i<j} u_{ij}(\theta) + o_p(n^{-1/2}).$$

By the standard theory of U-statistics, we have

$$n^{1/2}(\widehat{\mathrm{AUC}} - \mathrm{AUC})/\sigma_{AUC}^2 \to \mathcal{N}(0, 1),$$

in distribution, where

$$\sigma^2_{AUC} = \frac{2}{n(n-1)} \sum_{i<j,k \ j\neq k} \{u_{ij}(\theta) - \text{AUC}\}\{u_{ik}(\theta) - \text{AUC}\}.$$

Appendix II

We show that the EM type estimate $\widetilde{\text{sen}}_n(c)$ is consistent and approximately equal to the substitution approach-based estimate $\widehat{\text{sen}}_n(c)$ in Eq. (7). In the EM algorithm of the Cox PH cure models, the complete data log-likelihood is the sum of the log-likelihood of the logistic regression model and the log-likelihood of the Cox PH model. The missing cure status Y in the observed data is only involved in the logistic regression part. It is well known that the log-likelihood of the logistic regression model is a linear function of the disease status variable Y. Thus, the conditional mean of the complete data log-likelihood is a linear combination of $w = E(Y|O)$ which can be evaluated using Eq. (6) in the E-step of the EM algorithm. We have

$$E\left(\sum_{i=1}^{n} Y_i \Big| O\right) = \sum_{i=1}^{n} w_i. \tag{25}$$

Thus, both $\sum_{i=1}^n w_i$ and $[\sum_{i=1}^n w_i]^{-1}$ are O-measurable. Since $I(b^T Z_i > c)$ is also O-measurable, $E\left(I(b^T Z_i > c)Y_i \Big| O\right) = I(b^T Z_i > c)w_i$. Thus, we have

$$E\left(\sum_{i=1}^{n} I(b^T Z_i > c)Y_i \Big| O\right) = \sum_{i=1}^{n} I(b^T Z_i > c)w_i. \tag{26}$$

The proposed estimate $\widehat{\text{sen}}_n(c)$ of sen(c) in Eq. (7) of the above subsection is in the form

$$\widehat{\text{sen}}_n(c) = \frac{\sum I(b^T Z_i > c)w_i}{\sum w_i}.$$

Since $[\sum_{i=1}^n w_i]^{-1}$ is O-measurable, using Eq. (26), we have

$$\widehat{\text{sen}}_n(c) = \frac{E(\sum I(b^T Z_i > c)Y_i|O)}{\sum w_i} = E\left[\frac{\sum I(b^T Z_i > c)Y_i}{\sum w_i}\Big| O\right]. \tag{27}$$

Comparing terms of $\widehat{\text{sen}}_n(c)$ in Eq. (27) with the sample proportion sen$_n(c)$ in Eq. (14), we have

$$\widehat{\text{sen}}_n(c) = E\left[\frac{\sum Y_i}{\sum w_i}\text{sen}_n(c)\Big| O\right]. \tag{28}$$

From Eqs. (15) and (28), we obtain

$$\widehat{\text{sen}}_n(c) - \widetilde{\text{sen}}_n(c) = E\left[\left(\frac{\sum Y_i}{\sum w_i} - 1\right)\text{sen}_n(c)\Big| O\right]$$

By the Lyapunov Central Limit Theorem, we have

$$s_n := \sqrt{n}\left[\frac{\sum Y_i}{\sum w_i} - 1\right] \Rightarrow N(0, \Sigma_w), \tag{29}$$

where $\Sigma_w = \lim_{n\to\infty}[n^{-1}\sum_{i=1}^n w_i(1-w_i)]/[n^{-1}\sum_{i=1}^n w_i]^2$. That is, when n is large, we have $\frac{\sum Y_i}{\sum w_i} \approx 1$ and

$$\widehat{\text{sen}}_n(c) = E\left[\frac{\sum_{i=1}^n Y_i}{\sum w_i}\text{sen}_n(c)\Big| O\right] \approx E\left[\text{sen}_n(c)\Big| O\right] = \widetilde{\text{sen}}_n(c). \tag{30}$$

Or, equivalently, let $s_n = \sqrt{n}\left[\frac{\sum Y_i}{\sum w_i} - 1\right]$

$$\widehat{\text{sen}}_n(c) - \widetilde{\text{sen}}_n(c) = \frac{1}{\sqrt{n}}E\left[s_n \cdot \text{sen}_n(c)\Big| O\right]$$

Thus, $\widetilde{\text{sen}}_n(c)$ is approximately equal to the proposed EM-based estimate $\widehat{\text{sen}}_n(c)$. Therefore, by Slutsky's lemma, both are consistent because of the consistency of $\widehat{\text{sen}}_n(c)$ already established in the preceding subsection.

References

Andersson TML, Eriksson H, Hansson J, Månsson-Brahme E, Dickman PW, Eloranta S, Lambe M, Lambert PC (2014) Estimating the cure proportion of malignant melanoma, an alternative approach to assess long term survival: a population-based study. Cancer Epidemiol 38(1):93–99

Aravanis AM, Lee M, Klausner RD (2017) Next-generation sequencing of circulating tumor dna for early cancer detection. Cell 168(4):571–574

Asano J, Hirakawa A, Hamada C (2014) Assessing the prediction accuracy of cure in the Cox proportional hazards cure model: an application to breast cancer data. Pharmaceut Stat 13(6):357–363

Brier GW (1950) Verification of forecasts expressed in terms of probability. Mon Weather Rev 78(1):1–3

Broët P, Kuznetsov VA, Bergh J, Liu ET, Miller LD (2006) Identifying gene expression changes in breast cancer that distinguish early and late relapse among uncured patients. Bioinformatics 22(12):1477–1485

Brown M, Tsodikov A, Bauer KR, Parise CA, Caggiano V (2008) The role of human epidermal growth factor receptor 2 in the survival of women with estrogen and progesterone receptor-negative, invasive breast cancer: the California Cancer Registry, 1999–2004. Cancer 112(4):737–747

Chang G, Tadepalli J, Shao Y, Zhang Y, Osman I, Polsky D et al (2016) Sensitivity of plasma braf mutant and nras mutant cell-free dna assays to detect metastatic melanoma in patients with low recist scores and non-recist disease progression. Mol Oncol 10(1):157–165

Chen K, Jin Z, Ying Z (2002) Semiparametric analysis of transformation models with censored data. Biometrika 89(3):659–668

Chen MH, Kim S (2009) Cure rate models with application to melanoma and prostate cancer data. In: Peace KE (ed) Design and analysis of clinical trials with time-to-event endpoints. CRC press, pp 349–370

Chen MH, Ibrahim JG, Sinha D (1999) A new bayesian model for survival data with a surviving fraction. J Am Stat Assoc 94(447):909–919
Couzin-Frankel J (2013) Cancer immunotherapy. Science 342(6165):1432–1433
Crowley J, Shaughnessy J, Bolejack V, Anaissie E, Van Rhee F, Barlogie B (2010) Cure fractions (CF) modeled from event-free survival and complete response duration plots in total therapy (TT) trials for newly diagnosed multiple myeloma (MM). J Clin Oncol 28(15_suppl):8119–8119
Cymerman RM, Shao Y, Wang K, Zhang Y, Murzaku EC, Penn LA, Osman I, Polsky D (2016) De novo vs nevus-associated melanomas: differences in associations with prognostic indicators and survival. JNCI J Natl Cancer Inst. https://doi.org/10.1093/jnci/djw121
Fang Hb, Li G, Sun J (2005) Maximum likelihood estimation in a semiparametric logistic/proportional-hazards mixture model. Scand J Stat 32(1):59–75
Farewell VT (1982) The use of mixture models for the analysis of survival data with long-term survivors. Biometrics 38:1041–1046
Farewell VT (1986) Mixture models in survival analysis: are they worth the risk? Can J Stat 14(3):257–262
Friedman E, Shang S, Hernando E, Shao Y, Osman I et al (2012) Serum microrna as biomarkers for recurrence in melanoma. J Transl Med 10(1):1–10
Gandhi L, Rodríguez-Abreu D, Gadgeel S, Esteban E, Felip E, De Angelis F, Domine M, Clingan P, Hochmair MJ, Powell SF et al (2018) Pembrolizumab plus chemotherapy in metastatic non-small-cell lung cancer. N Engl J Med 378(22):2078–2092
Gönen M, Heller G (2005) Concordance probability and discriminatory power in proportional hazards regression. Biometrika 92(4):965–970
Götze F (1987) Approximations for multivariate u-statistics. J Multivar Anal 22(2):212–229
Han X, Zhang Y, Shao Y (2017) On comparing 2 correlated C indices with censored survival data. Stat Med 36:4041–4049
Hanley JA, McNeil BJ (1982) The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 143(1):29–36
Jiang W, Sun H, Peng Y (2017) Prediction accuracy for the cure probabilities in mixture cure models. Stat Methods Med Res 26(5):2029–2041
Jin Z, Mesbah M (2014) Unidimensionality, agreement and concordance probability. In: Couallier V, Gerville-Réache L, Huber-Carol C, Limnios N, Mesbah M (eds) Statistical models and methods for reliability and survival analysis. Wiley, pp 3–19
Kim S, Xi Y, Chen MH (2009) A new latent cure rate marker model for survival data. Ann Appl Stat 3:1124–1146
Kirkwood JM, Strawderman MH, Ernstoff MS, Smith TJ, Borden EC, Blum RH (1996) Interferon alfa-2b adjuvant therapy of high-risk resected cutaneous melanoma: the eastern cooperative oncology group trial EST 1684. J Clin Oncol 14(1):7–17
Kuk AY, Chen CH (1992) A mixture model combining logistic regression with proportional hazards regression. Biometrika 79(3):531–541
Lambert P, Dickman P, Weston C, Thompson J (2010) Estimating the cure fraction in population-based cancer studies by using finite mixture models. J R Stat Soc Ser C (Appl Stat) 59(1):35–55
LeCam L (1986) Asymptotic methods in statistical decision theory. Spinger, New York
Liu X, Jin Z (2009) A non-parametric approach to scale reduction for uni-dimensional screening scales. Int J Biostat. https://doi.org/10.2202/1557-4679.1094
Liu X, Jin Z, Graziano JH (2016) Comparing paired biomarkers in predicting quantitative health outcome subject to random censoring. Stat Methods Med Res 25(1):447–457
Lu W, Ying Z (2004) On semiparametric transformation cure models. Biometrika 91(2):331–343
McIntosh M, Pepe M (2002) Combining several screening tests: optimality of the risk scorel. Biometrics 58:657–664
Oakes D (1999) Direct calculation of the information matrix via the EM algorithm. J R Stat Soc Ser B 61(2):479–482
Othus M, Barlogie B, LeBlanc ML, Crowley JJ (2012) Cure models as a useful statistical tool for analyzing survival. Clin Cancer Res 18(14):3731–3736
Patricia Bernardo M, Ibrahim JG (2000) Group sequential designs for cure rate models with early stopping in favour of the null hypothesis. Stat Med 19(22):3023–3035
Peng Y, Carriere K (2002) An empirical comparison of parametric and semiparametric cure models. Biometr J 44(8):1002–1014
Peng Y, Dear KB (2000) A nonparametric mixture model for cure rate estimation. Biometrics 56(1):237–243

Peng Y, Taylor JM (2013) Cure models. In: Handbook of survival analysis, p 113
Peng Y, Dear KB, Denham J et al (1998) A generalized f mixture model for cure rate estimation. Stat Med 17(8):813–830
Pocock SJ, Gore SM, Kerr GR (1982) Long term survival analysis: the curability of breast cancer. Stat Med 1(2):93–104
Psioda MA, Ibrahim JG (2018) Bayesian design of a survival trial with a cured fraction using historical data. Stat Med 37(26):3814–3831
Qian M, Ma MW, Fleming NH et al (2018) Clinicopathological characteristics at primary melanoma diagnosis as risk factors for brain metastasis. Melanoma Res 23(6):461–467
Robert C, Long GV, Brady B, Dutriaux C, Maio M, Mortier L, Hassel JC, Rutkowski P, McNeil C, Kalinka-Warzocha E et al (2015) Nivolumab in previously untreated melanoma without braf mutation. N Engl J Med 372(4):320–330
Sargent D, Sobrero A, Grothey A, O'Connell MJ, Buyse M, Andre T, Zheng Y, Green E, Labianca R, O'Callaghan C et al (2009) Evidence for cure by adjuvant therapy in colon cancer: observations based on individual patient data from 20,898 patients on 18 randomized trials. J Clin Oncol 27(6):872–877
Schlom J (2012) Therapeutic cancer vaccines: current status and moving forward. J Natl Cancer Inst 104(8):599–613
Sun X, Bao J, Shao Y (2016) Mathematical modeling of therapy-induced cancer drug resistance: connecting cancer mechanisms to population survival rates. Sci Rep 6:1–12
Sun X, Liu X, Xia M et al (2019) (2019) Multicellular gene network analysis identifies a macrophage-related gene signature predictive of therapeutic response and prognosis of gliomas. J Transl Med 17(1):159
Sy JP, Taylor JM (2000) Estimation in a Cox proportional hazards cure model. Biometrics 56(1):227–236
Tsiatis A (2007) Semiparametric theory and missing data. Springer, Berlin
Uno H, Cai T, Pencina MJ, D'Agostino RB, Wei L (2011) On the c-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. Stat Med 30(10):1105–1117
Wang A, Zhang Y, Shao Y (2017) On the likelihood of mixture cure models. Stat Probab Lett 131:51–55
Weber J, Mario Mandala, Michele Del Vecchio et al (2017) Adjuvant nivolumab versus ipilimumab in resected stage III or IV melanoma. N Engl J Med 377:1824–1835
Wich LG, Hamilton HK, Shapiro RL, Pavlick A, Berman RS, Polsky D, Goldberg JD, Hernando E, Manga P, Krogsgaard M et al (2009) Developing a multidisciplinary prospective melanoma biospecimen repository to advance translational research. Am J Transl Res 1(1):35
Wolchok JD, Kluger H, Callahan MK, Postow MA, Rizvi NA, Lesokhin AM, Segal NH, Ariyan CE, Gordon RA, Reed K et al (2013) Nivolumab plus ipilimumab in advanced melanoma. N Engl J Med 369(2):122–133
Yilmaz YE, Lawless JF, Andrulis IL, Bull SB (2013) Insights from mixture cure modeling of molecular markers for prognosis in breast cancer. J Clin Oncol 31:2047–2054
Zeng D, Lin D (2007) Maximum likelihood estimation in semiparametric regression models with censored data. J R Stat So Ser B (Stat Methodol) 69(4):507–564
Zhang Y (2016) Concordance probability in censored survival model and statistical methods for longitudinal microbiome data. PhD thesis, New York University
Zhang Y, Shao Y (2018) Concordance measure and discriminatory accuracy in transformation cure models. Biostatistics 19(1):14–26
Zhang Y, Jin Z, Shao Y, Ying Z (2018) Statistical inference on transformation models: a self-induced smoothing approach. J Nonparametr Stat 30(2):308–331

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Affiliations

Yilong Zhang[1] · Xiaoxia Han[2] · Yongzhao Shao[3]

✉ Yongzhao Shao
Yongzhao.Shao@nyulangone.org

[1] Department of Biostatistics and Research Decision Sciences, Merck & Co., Inc, Kenilworth, NJ, USA

[2] Department of Public Health Sciences, Henry Ford Health System, Detroit, MI, USA

[3] Departments of Population Health & Environmental Medicine, NYU Grossman School of Medicine, 180 Madison Ave, 4th Floor, Suite 455, New York, NY 10016, USA