同意の取得，細胞の維持・管理にはヒト・金・モノ・時間の経営資源が必要となる。

細胞自律的な病態と非細胞自律的な病態（cell autonomousとnon-cell autonomous）

*In vivo*の解析ができないことがiPS細胞モデルの欠点でもある。従来のiPS細胞を用いた解析では，単一の細胞種に着目した細胞自律的な解析が主体であった。しかし，多種類の細胞が組み合わさって一つの個体が生み出されることを考慮すると，それぞれの疾患も一種類の細胞のみの問題で発症しているとは限らない。例えば，運動ニューロン病と考えられてきた筋萎縮性側索硬化症（ALS）では，グリア細胞がその病態に重要な役割を果たしていることが明らかになってきている[11]。このような非細胞自律的な病態解析では，個体として解析できる動物モデルのほうが，iPS細胞による*in vitro*モデルに比べて一歩秀でているといえる。筆者らのグループでは患者，あるいは健常者由来のiPS細胞からさまざまな細胞を誘導して共培養することで，複数の細胞種の相互作用に基づく非細胞自律的な病態に注目した解析を進めている。最近では多能性幹細胞から三次元的に分化誘導する脳オルガノイド培養法[12]や，眼球の三次元培養法[13]など，器官としての分化誘導法が構築されつつある。したがって，単一の細胞モデルから共培養系による組織としてのモデル，あるいは器官としての疾患モデルへと発展させることで，iPS細胞の欠点である「*in vivo*の解析ができない」問題点を克服し，本当の意味での「疾患モデル」へと進化できる可能性がある。

どのような疾患が解析対象となるのか？

このようなiPS細胞の特性を理解したうえで，解析対象の疾患を選択していく必要がある。まず解析対象となるのは，やはり原因遺伝子がすでに同定されている，遺伝性疾患である。孤発性疾患は，同じ名称の疾患であっても，さまざまな原因が混在する「ヘテロな疾患群」であることが多く，多様な解析結果が得られ，解析に困難を伴うことが予想される。一方，単一の遺伝子異常に起因する遺伝性疾患であれば，そのレスキューやゲノム編集による遺伝子治療も可能であり，解釈しやすい結果が得られる可能性が高い。ただし先に述べたように，原因遺伝子の同定されている遺伝性疾患はごく一部であり，多くの場合が孤発性であることを考えると，遺伝性疾患で明らかにされた病態メカニズムが，どのくらい孤発性疾患にもあてはまるのか，原因遺伝子の同定されている疾患の解析結果を，どのように孤発性疾患にあてはめていくのかということが重要な課題となる。現在，神経系疾患のなかで解析対象として特に着目されているアルツハイマー病（AD），パーキンソン病（PD），ALSなどは，このような問題のよい例であるといえる。

一方，エピジェネティックスの異常や環境因子が影響する疾患については，解析に注意を要する。特に，発生段階での生体内での「リプログラミング」過程の異常に起因する疾患では，*in vitro*でのリプログラミングにより作製されたiPS細胞では，そのままでは疾患を再現するのは難しい。また環境因子が影響する場合は，疾患感受性が高いという前提で，*in vitro*でいかにその環境を再現するかが重要となる。ミトコンドリア疾患などでみられるヘテロプラスミーにも注意しなくてはならない。これらを考慮すると，iPS細胞を用いた解析に適した疾患，適さない疾患があり，創薬への応用を慎重に検討する必要があると同時に，疾患ごとに達成度は変わってくるだろうと予想される。

疾患特異的iPS細胞を用いた薬剤スクリーニング

疾患特異的iPS細胞を用いた最大の魅力の一つが，ヒト細胞，患者由来細胞を用いた，薬効・毒性評価や薬剤スクリーニングへの応用である（**図1**）。iPS細胞の応用により，従来の動物実験の問題点や，臨床試験におけるコスト，時間を大幅に軽減できる可能性がある。また，iPS細胞技術を用いた既存薬のスクリーニングを行うことがで

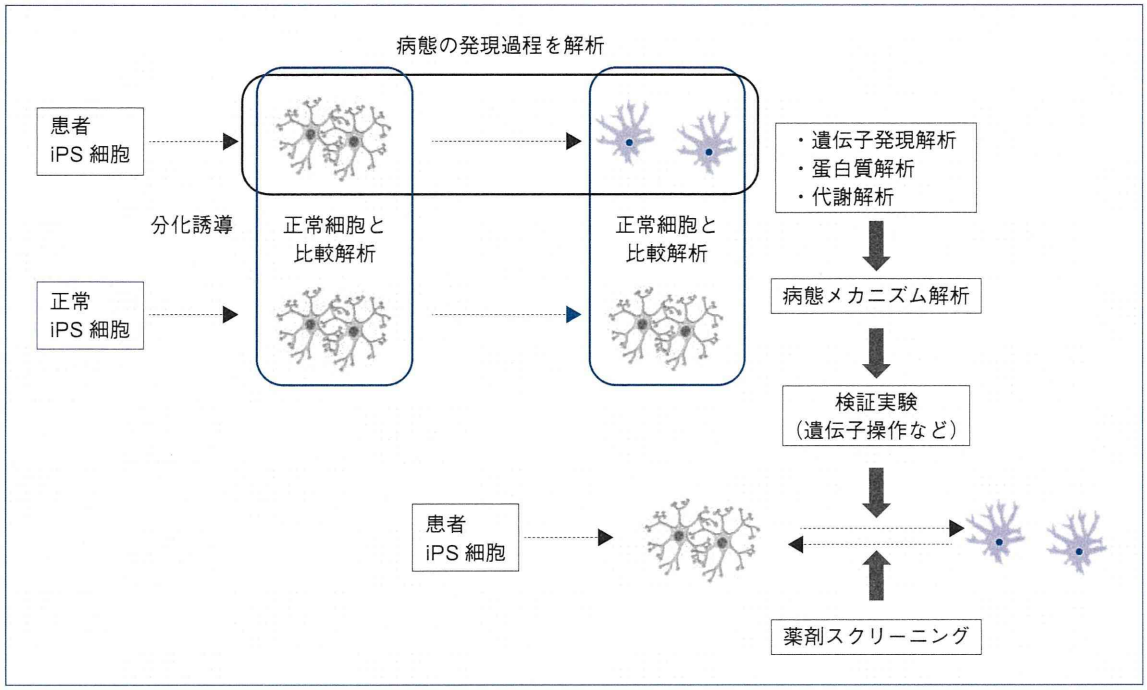

図1　疾患特異的iPS細胞を用いた病態解析と創薬

きれば，さらに迅速に臨床応用にいたる可能性がある（いわゆるドラッグリポジショニング）。このような疾患特異的iPS細胞を用いた応用例としては，ALS患者由来iPS細胞から誘導した運動ニューロンを用いて複数の薬剤の効果を検討し，表現型を改善しうる薬剤を見出した報告や[14]，家族性自律神経失調症のiPS細胞由来神経堤細胞を用いて，原因遺伝子である*IKBKAP*の正常型フォームの発現上昇を指標に6,912のコンパウンドをスクリーニングし，有効な薬剤を見出した報告などが特筆される[15]。

しかし，このような薬剤スクリーニングを推進していくためには，いくつもの重要なハードルを乗り越える必要がある。まず，いかにして表現型を捉えるか。疾患特異的iPS細胞から誘導した細胞では，わずかな表現型の差しか観察されないことも多く，このわずかな差を確かな差として捉えるための高感度の検出システムと，十分な再現性（十分なサンプル数）を得ることが必要不可欠である。そのため，簡便なiPS細胞の樹立法，分化誘導法，評価法の構築や，自動化によるハイスループットスクリーニングも重要である。また，iPS細胞創薬を加速度的に進行させるためには，iPS細胞リソースの効率的な活用が重要である。そのためには，疾患特異的iPS細胞バンクの充実は必要不可欠であり，わが国でも，京都大学iPS細胞研究所（CiRA）を中心に事業が推進されている。今後，さらにiPS細胞から分化した細胞（神経幹細胞などの分化の途中段階の細胞）の細胞バンクを構築することができれば，iPS細胞創薬にさらに拍車がかかると考えられる。最終的なiPS細胞創薬の実現には，企業の参画をスムーズに進めていくことが課題である。海外では，アカデミアと企業が共同してiPS細胞創薬に向けたシステムを構築しつつあり，特にアカデミアでは遂行しにくい実用化研究をベンチャー企業などがうまく取りもつこと，また製薬企業がもつ莫大なシーズ，ライブラリーを活用できる環境を構築することが肝要である。そのためには，成果・知的財産の取り扱いについても十分に議論する必要がある[16]。

おわりに

以上を総合して,「iPS細胞創薬は10年以内に実現するか?」という問いに答えるとしたら,「iPS細胞創薬」は技術的には実現すると思われるが,その時期や実用化は「疾患によって異なる」と答えたいと思う。

REFERENCES

1) Merkle FT, Eggan K : Modeling human disease with pluripotent stem cells : from genome association to function. *Cell Stem Cell* 12 : 656-668, 2013
2) Sasaki E, Suemizu H, Shimada A, et al : Generation of transgenic non-human primates with germline transmission. *Nature* 459 : 523-527, 2009
3) Takahashi K, Tanabe K, Ohnuki M, et al : Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* 131 : 861-872, 2007
4) Chambers SM, Fasano CA, Papapetrou EP, et al : Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling. *Nat Biotechnol* 27 : 275-280, 2009
5) Numasawa-Kuroiwa Y, Okada Y, Shibata S, et al : Involvement of ER Stress in Dysmyelination of Pelizaeus-Merzbacher Disease with PLP1 Missense Mutations Shown by iPSC-Derived Oligodendrocytes. *Stem Cell Reports* 2 : 648-661, 2014
6) Osafune K, Caron L, Borowiak M, et al : Marked differences in differentiation propensity among human embryonic stem cell lines. *Nat Biotechnol* 26 : 313-315, 2008
7) Hu BY, Weick JP, Yu J, et al : Neural differentiation of human induced pluripotent stem cells follows developmental principles but with variable potency. *Proc Natl Acad Sci U S A* 107 : 4335-4340, 2010
8) Brons IG, Smithers LE, Trotter MW, et al : Derivation of pluripotent epiblast stem cells from mammalian embryos. *Nature* 448 : 191-195, 2007
9) Tesar PJ, Chenoweth JG, Brook FA, et al : New cell lines from mouse epiblast share defining features with human embryonic stem cells. *Nature* 448 : 196-199, 2007
10) Miller JD, Ganat YM, Kishinevsky S, et al : Human iPSC-based modeling of late-onset disease via progerin-induced aging. *Cell Stem Cell* 13 : 691-705, 2013
11) Yamanaka K, Chun SJ, Boillee S, et al : Astrocytes as determinants of disease progression in inherited amyotrophic lateral sclerosis. *Nat Neurosci* 11 : 251-253, 2008
12) Lancaster MA, Renner M, Martin CA, et al : Cerebral organoids model human brain development and microcephaly. *Nature* 501 : 373-379, 2013
13) Kadoshima T, Sakaguchi H, Nakano T, et al : Self-organization of axial polarity, inside-out layer pattern, and species-specific progenitor dynamics in human ES cell-derived neocortex. *Proc Natl Acad Sci U S A* 110 : 20284-20289, 2013
14) Egawa N, Kitaoka S, Tsukita K, et al : Drug Screening for ALS Using Patient-Specific Induced Pluripotent Stem Cells. *Sci Transl Med* 4 : 145ra104, 2012
15) Lee G, Ramirez CN, Kim H, et al : Large-scale screening using familial dysautonomia induced pluripotent stem cells identifies compounds that rescue IKBKAP expression. *Nat Biotechnol* 30 : 1244-1248, 2012
16) 中辻憲夫, 淺田　孝, 仙石慎太郎 : ヒト多能性幹細胞(ES/iPS細胞)の医療・創薬応用を目指す技術と政策－グローバルビューと海外動向－. 再生医療 12 : 316-341, 2013

MAIN PAPER

Pharmaceutical Statistics

(wileyonlinelibrary.com) DOI: 10.1002/pst.1630

Published online 16 July 2014 in Wiley Online Library

Assessing the prediction accuracy of cure in the Cox proportional hazards cure model: an application to breast cancer data

Junichi Asano,[a]* Akihiro Hirakawa,[b] and Chikuma Hamada[c]

A cure rate model is a survival model incorporating the cure rate with the assumption that the population contains both uncured and cured individuals. It is a powerful statistical tool for prognostic studies, especially in cancer. The cure rate is important for making treatment decisions in clinical practice. The proportional hazards (PH) cure model can predict the cure rate for each patient. This contains a logistic regression component for the cure rate and a Cox regression component to estimate the hazard for uncured patients. A measure for quantifying the predictive accuracy of the cure rate estimated by the Cox PH cure model is required, as there has been a lack of previous research in this area. We used the Cox PH cure model for the breast cancer data; however, the area under the receiver operating characteristic curve (AUC) could not be estimated because many patients were censored. In this study, we used imputation-based AUCs to assess the predictive accuracy of the cure rate from the PH cure model. We examined the precision of these AUCs using simulation studies. The results demonstrated that the imputation-based AUCs were estimable and their biases were negligibly small in many cases, although ordinary AUC could not be estimated. Additionally, we introduced the bias-correction method of imputation-based AUCs and found that the bias-corrected estimate successfully compensated the overestimation in the simulation studies. We also illustrated the estimation of the imputation-based AUCs using breast cancer data.

Keywords: cancer prognosis; Cox proportional hazards cure model; logistic regression; area under the receiver operating characteristic curve; imputation

1. INTRODUCTION

In cancer studies, some patients with long-term censored recurrence-free periods may be considered cured, whereas others may eventually have a recurrence. For example, a subset of primary breast cancer patients treated with neoadjuvant chemotherapy (NAC) was reported to achieve long-term disease-free survival (DFS) [1]. These patients did not experience recurrence, metastases, or death during the long-term follow-up study (e.g., for over 10 years) and were therefore considered to be clinically 'cured'. Examples for melanoma and prostate cancer data are shown in [2].

In cancer prognostic studies involving patients with long-term censored survival, the clinical variables associated with cure need to be identified, and the cure probability should be predicted for each patient. Determining this cure rate is important for making treatment decisions in clinical practice. In such cases, the proportional hazards (PH) cure model can be useful [3–6]. The PH cure model contains a logistic regression component for the cure rate and a Cox regression component for simultaneously estimating the hazard for uncured patients. Therefore, a measure for quantifying the predictive accuracy of the cure rate estimated by the PH cure model is required. In addition, a measure for evaluating the predictive accuracy of the survival data is required. In the ordinary PH model, the *c*-statistic is a well-known measure [7,8] and have been widely studied [9–14]. In the cure model, however, the PH assumption dose not generally hold more often for uncured patients than any other non-PH assumptions because the hazard for these patients is zero during the trial; therefore, the *c*-statistic cannot be interpreted in the cure model. Thus, although an alternative index to the *c*-statistic is an unresolved research problem in this field, we focus on measuring the predictive accuracy of the cure rate in the cure model based on the breast cancer data analysis presented later.

In general regression modeling strategies, several measures for quantifying the predictive accuracy have been proposed: (i) an explained variation measure quantifying the proportion of variation explained by a covariate [15]; (ii) a calibration measure quantifying the agreement between observed outcomes and predictions [16]; and (iii) a discrimination measure that shows how well the model can discriminate between those with and those without the outcome. The area under the receiver operating characteristic curve (AUC) is a well-known measure and has been widely used [17]. However, a measure for quantifying the predictive

[a] *Biostatistics Group, Center for Product Evaluation, Pharmaceuticals and Medical Devices Agency, Tokyo 100-0013, Japan*

[b] *Biostatistics Section, Center for Advanced Medicine and Clinical Research, Nagoya University Graduate School of Medicine, Nagoya 466-8560, Japan*

[c] *Department of Management Science, Graduate School of Engineering, Tokyo University of Science, Tokyo 162-8601, Japan*

**Correspondence to: Junichi Asano, Biostatistics Group, Center for Product Evaluation, Pharmaceuticals and Medical Devices Agency, 3-3-2 Kasumigaseki, Chiyoda-ku, Tokyo 100-0013, Japan.*
E-mail: asano-junichi@pmda.go.jp

accuracy of the cure rate estimated by the PH cure model has not been adequately studied.

In terms of using the cure model, we believe that a key component in the assessment of predictive accuracy is the ability to distinguish between those subjects who will never experience a recurrence (cured) from those who will eventually experience a recurrence (not cured). A discrimination measure also needs to be easy to interpret from a clinical viewpoint. We applied our method to a dataset of breast cancer patients receiving NAC at the National Cancer Center Hospital, Tokyo, Japan, between May 1995 and July 2007 [18,19]. We used the Cox PH cure model; however, the ordinary AUC could not be estimated because many patients were censored and their cure status was treated as 'missing'.

In this paper, we address the difficulty of AUC calculation in the presence of missing data by utilizing the method proposed by Alonzo and Pepe [20]. Specifically, we apply the imputation-based AUCs developed by them in order to assess the predictive accuracy of the cure rate estimated by the PH cure model. We conduct simulation studies to examine the precision of the imputation-based AUCs. We also examine the usefulness of these AUCs using the breast cancer data and discuss its performance.

2. METHODS

2.1. PH cure model

Let Y be an indicator that an individual will eventually ($Y = 1$) or never ($Y = 0$) experience the event, with probability $p = \Pr(Y = 1)$. Let T denote the time to the event, defined only when $Y = 1$, with probability density function $f(t|Y = 1)$ and survival function $S(t|Y = 1)$. For a censored individual, Y is not observed. We assume an independent, noninformative, random censoring model, and that censoring is statistically independent of Y. In the PH cure model, the probability density function of T can be written as follows:

$$f(t) = pf_1(t|Y = 1) + (1 - p)f_2(t|Y = 0) = pf_1(t|Y = 1). \quad (1)$$

The cumulative distribution function is defined as $F(t) = p\int_0^t f_1(u|Y = 1)du$, and therefore, the survival function $S(t) = 1 - F(t)$ can be expressed as follows:

$$S(t) = (1 - p) + pS(t|Y = 1). \quad (2)$$

The survival functions for the overall patient population and uncured patients for $t \to \infty$ are $S(t) \to 1-p$ and $S(t|Y = 1) \to 0$, respectively. For the probability of cure $1 - p$, numerous studies [3–6] assume a logistic regression:

$$\Pr(Y = 0 \mid \boldsymbol{x}) = 1 - p(\boldsymbol{x}) = \exp\left(\beta_0 + \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{x}\right) \Big/ \left\{1 + \exp\left(\beta_0 + \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{x}\right)\right\}, \quad (3)$$

where β_0 is the intercept, $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_U)^{\mathrm{T}}$ is the vector of regression coefficient, and $\boldsymbol{x} = (x_1, \cdots, x_U)^{\mathrm{T}}$ is the vector of exploratory variable.

Next, the Cox regression model is assumed for time t [21]:

$$\lambda(t|Y = 1, \boldsymbol{x}) = \lambda_0(t|Y = 1)\exp\left(\boldsymbol{\gamma}^{\mathrm{T}}\boldsymbol{x}\right), \quad (4)$$

where $\lambda_0(t|Y = 1)$ is the baseline hazard function for uncured patients and $\boldsymbol{\gamma} = (\gamma_1, \cdots, \gamma_U)^{\mathrm{T}}$ is the vector of regression coefficient. The cumulative hazard function for uncured patients is defined using Equation (4) as follows: $\Lambda(t|Y = 1, x) = \Lambda_0(t|Y = 1)\exp(\boldsymbol{\gamma}^{\mathrm{T}}\boldsymbol{x})$. Therefore, the cumulative baseline hazard function can be written as follows: $\Lambda_0(t|Y = 1) = \int_0^t \lambda_0(u|Y = 1)du$. The survival function for uncured patients is $S(t|Y = 1, \boldsymbol{x}) = S_0(t|Y = 1)^{\exp(\boldsymbol{\gamma}^{\mathrm{T}}\boldsymbol{x})}$.

We denote the observed data for patients i ($i = 1, \cdots, n$) by $(t_i, \delta_i, \boldsymbol{x}_i)$, where t_i is the observed event or censoring time, $\delta_i = 1$ if t_i is uncensored, and $\delta_i = 0$ otherwise. To clearly explain the PH cure model, we suppose that common U variables are included in the logistic and Cox PH models and that variable selection is performed (in practice, different variables can be included in each model). We denote the k distinct event times by $t_{(1)} < t_{(2)} < \cdots < t_{(k)}$. It follows that if $\delta_i = 1$, then $y_i = 1$, and if $\delta_i = 0$, then y_i is not observed. Here, we use the EM algorithm to estimate the parameters $(\beta_0, \boldsymbol{\beta}, \boldsymbol{\gamma}, \Lambda_0)$ with the method developed by Sy and Taylor [6], and obtain the maximum likelihood estimates $(\hat{\beta}_0, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}, \hat{\Lambda}_0)$ [6]. The complete-data full likelihood function of the PH cure model is as follows:

$$\begin{aligned} L(\beta_0, \boldsymbol{\beta}, \boldsymbol{\gamma}, \Lambda_0; \boldsymbol{y}, \boldsymbol{x}) &= \prod_{i=1}^{n} p^{y_i}(\boldsymbol{x}_i)\{1 - p(\boldsymbol{x}_i)\}^{1-y_i} \\ &\quad \cdot \prod_{i=1}^{n} \left\{\lambda_0(t_i|Y = 1)\exp\left(\boldsymbol{\gamma}^{\mathrm{T}}\boldsymbol{x}_i\right)\right\}^{\delta_i y_i} \\ &\quad \times \exp\left\{-y_i\Lambda_0(t_i|Y = 1)\exp\left(\boldsymbol{\gamma}^{\mathrm{T}}\boldsymbol{x}_i\right)\right\} \\ &= L_1(\beta_0, \boldsymbol{\beta}; \boldsymbol{y}, \boldsymbol{x})L_2(\boldsymbol{\gamma}, \Lambda_0; \boldsymbol{y}, \boldsymbol{x}), \end{aligned} \quad (5)$$

where $\boldsymbol{y} = (y_1, \cdots, y_n)^{\mathrm{T}}$ [6].

2.2. Ordinary AUC

Using the estimates of $(\beta_0, \boldsymbol{\beta})$, we calculate a cure rate $\hat{p}_c(\boldsymbol{x}) = 1 - \hat{p}(\boldsymbol{x}) = \exp\left(\hat{\beta}_0 + \hat{\boldsymbol{\beta}}^{\mathrm{T}}\boldsymbol{x}\right) / \left\{1 + \exp\left(\hat{\beta}_0 + \hat{\boldsymbol{\beta}}^{\mathrm{T}}\boldsymbol{x}\right)\right\}$. Next, let τ be the period that determines the clinical 'cure'. It follows that (i) if $t_i > \tau$, then $y_i = 0$; (ii) if $t_i \leqslant \tau$ and $\delta_i = 1$, then $y_i = 1$; and (iii) if $t_i \leqslant \tau$ and $\delta_i = 0$, then y_i is missing. Let V be an indicator of the following: if $y_i = 0$ or 1, then $v_i = 1$, and if y_i is missing, then $v_i = 0$. We define n_1 as the sum of the number of patients with $v_i = 1$ and $y_i = 1$, and further n_0 as the number of patients with $v_i = 1$ and $y_i = 0$. The ordinal AUC only includes data from those patients with a cured status ($y_i = 0$ and 1).

The true positive rate (TPR) and false positive rate (FPR) can be estimated by using

$$\widehat{TPR}(c) = \frac{\sum_{i=1}^{n} I\left(\hat{p}_c(\boldsymbol{x}_i) \leqslant c\right) \cdot v_i \cdot y_i}{\sum_{i=1}^{n} \cdot v_i \cdot y_i}, \quad (6)$$

$$\widehat{FPR}(c) = \frac{\sum_{i=1}^{n} I\left(\hat{p}_c(\boldsymbol{x}_i) \leqslant c\right) \cdot v_i \cdot (1 - y_i)}{\sum_{i=1}^{n} \cdot v_i \cdot (1 - y_i)}, \quad (7)$$

where c is the cut-off point ($0 \leqslant c \leqslant 1$). We then estimate the ordinary AUC by using the nonparametric method [22].

$$\widehat{AUC} = \frac{\sum_{i=1}^{n_0}\sum_{j=1}^{n_1} \Psi\left\{\hat{p}_c(\boldsymbol{x}_i), \hat{p}_c(\boldsymbol{x}_j)\right\}}{n_0 \cdot n_1}, \quad (8)$$

where Ψ is a function of two variables: $\Psi\left\{\hat{p}_c(\boldsymbol{x}_i), \hat{p}_c(\boldsymbol{x}_j)\right\} = 1$ if $\hat{p}_c(\boldsymbol{x}_i) > \hat{p}_c(\boldsymbol{x}_j)$, $\Psi\left\{\hat{p}_c(\boldsymbol{x}_i), \hat{p}_c(\boldsymbol{x}_j)\right\} = 1/2$ if $\hat{p}_c(\boldsymbol{x}_i) = \hat{p}_c(\boldsymbol{x}_j)$,

$\Psi\{\hat{p}_c(\boldsymbol{x}_i), \hat{p}_c(\boldsymbol{x}_j)\} = 0$ if $\hat{p}_c(\boldsymbol{x}_i) < \hat{p}_c(\boldsymbol{x}_j)$. This estimate of the AUC is equivalent to that estimated by using the trapezoidal method [23].

2.3. Imputation-based AUCs for the Cox PH cure model

In this section, we propose to use two methods for estimating the AUC for the Cox PH cure model by utilizing the method proposed by Alonzo and Pepe [20]. One is the full imputation (FI), which replaces the y_i of the ordinary AUC with $1 - \hat{p}_c(\boldsymbol{x}_i)$, and subsequently calculates the TPR and FPR as follows:

$$\widehat{TPR}_{FI}(c) = \frac{\sum_{i=1}^{n} I\left(\hat{p}_c(\boldsymbol{x}_i) \leq c\right) \cdot \left\{1 - \hat{p}_c(\boldsymbol{x}_i)\right\}}{\sum_{i=1}^{n} \cdot \left\{1 - \hat{p}_c(\boldsymbol{x}_i)\right\}}, \tag{9}$$

$$\widehat{FPR}_{FI}(c) = \frac{\sum_{i=1}^{n} I\left(\hat{p}_c(\boldsymbol{x}_i) \leq c\right) \cdot \hat{p}_c(\boldsymbol{x}_i)}{\sum_{i=1}^{n} \cdot \hat{p}_c(\boldsymbol{x}_i)}. \tag{10}$$

The other method is the mean score imputation (MSI) method, which replaces the y_i of the ordinary AUC with $1 - \hat{p}_c(\boldsymbol{x}_i)$ for only censored patients, and subsequently calculates the TPR and FPR as follows:

$$\widehat{TPR}_{MSI}(c) = \frac{\sum_{i=1}^{n} I\left(\hat{p}_c(\boldsymbol{x}_i) \leq c\right) \cdot \left[v_i \cdot y_i + (1 - v_i) \cdot \left\{1 - \hat{p}_c(\boldsymbol{x}_i)\right\}\right]}{\sum_{i=1}^{n} \cdot \left[v_i \cdot y_i + (1 - v_i) \cdot \left\{1 - \hat{p}_c(\boldsymbol{x}_i)\right\}\right]}, \tag{11}$$

$$\widehat{FPR}_{MSI}(c) = \frac{\sum_{i=1}^{n} I\left(\hat{p}_c(\boldsymbol{x}_i) \leq c\right) \cdot \left\{v_i \cdot (1 - y_i) + (1 - v_i) \cdot \hat{p}_c(\boldsymbol{x}_i)\right\}}{\sum_{i=1}^{n} \cdot \left\{v_i \cdot (1 - y_i) + (1 - v_i) \cdot \hat{p}_c(\boldsymbol{x}_i)\right\}}. \tag{12}$$

Using the trapezoidal method, we estimate the AUCs for the FI and MSI methods, respectively.

2.4. Bias-correction using the bootstrap method

In the simulation studies shown in the next section, the proposed AUCs uniformly and slightly overestimate the true AUC. In this section, we introduce the bias-corrected estimators of the proposed AUCs using the bootstrap method presented by Copas and Corbett [24]. Specifically, we first estimated the logistic regression model, $\hat{p}_c^*(\boldsymbol{x}_i)$, from the bootstrap sample, and calculated the proposed AUC by using the $\hat{p}_c^*(\boldsymbol{x}_i)$ and the bootstrap sample, termed $\text{AUC}_{\text{bootstrap}}$. Next, we estimated the logistic regression model, $\hat{p}_c(\boldsymbol{x}_i)$, from the original sample. We then calculated the proposed AUC by using the $\hat{p}_c(\boldsymbol{x}_i)$, and the bootstrap sample, termed $\text{AUC}^*_{\text{bootstrap}}$. We repeated this step 200 times and obtained the mean difference (Δ) of $\text{AUC}_{\text{bootstrap}} - \text{AUC}^*_{\text{bootstrap}}$. Finally, we calculated the proposed AUC using the original sample, termed $\text{AUC}_{\text{original}}$, and used the $\text{AUC}_{\text{original}} - \Delta$ as the bias-corrected estimator of the proposed AUC. We also examined the accuracy of the bias-corrected estimates of the proposed AUCs using simulation; details are provided in the next section.

3. SIMULATION STUDIES

We evaluated the performance of the ordinary, FI-based, and MSI-based AUCs through the simulation studies. We assumed the PH cure model for a time-to-event outcome t_e as $f(t_e) = pf_1(t_e|Y = 1)$. In this model, we also assumed the logistic regression for the cure rate as $1 - p = \exp\left(\beta_0 + \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{x}\right) / \left\{1 + \exp\left(\beta_0 + \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{x}\right)\right\}$, and the Cox regression hazard for uncured patients as $\lambda(t_e|Y = 1) = \lambda_0(t_e|Y = 1)\exp\left(\boldsymbol{\gamma}^{\mathrm{T}}\boldsymbol{x}\right)$. Thus, the probability density function of t_e is given by the following:

$$f(t_e) = \left[1 - \exp\left(\beta_0 + \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{x}\right) \Big/ \left\{1 + \exp\left(\beta_0 + \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{x}\right)\right\}\right] \times \exp\left(\boldsymbol{\gamma}^{\mathrm{T}}\boldsymbol{x}\right) \exp(-t_e)^{\exp(\boldsymbol{\gamma}^{\mathrm{T}}\boldsymbol{x})}, \tag{13}$$

Table I. True coefficients of $\beta_1, \beta_2, \gamma_1$, and γ_2 under four scenarios ($\beta_0 = -1.1$ throughout).

	AUC = 0.6		AUC = 0.8	
Scenario	(β_1, γ_1)	(β_2, γ_2)	(β_1, γ_1)	(β_2, γ_2)
1	(−0.255, 0.255)	(0.255, −0.255)	(−0.895, 0.895)	(0.895, −0.895)
2	(−0.360, 0.000)	(0.000, −0.360)	(−2.000, 0.000)	(0.000, −2.000)
3	(0.000, 0.360)	(0.360, 0.000)	(0.000, 1.950)	(1.950, 0.0000)
4	(−0.260, 0.000)	(0.260, 0.000)	(−0.980, 0.000)	(0.980, 0.000)

Table II. Success rate of AUC calculation for the ordinary AUC (%).

		AUC = 0.6				AUC = 0.8			
λ_c	N	1	2	3	4	1	2	3	4
0.4	100	38	37	37	37	43	58	55	42
	400	84	85	84	84	90	96	96	89
0.7	100	2	2	2	2	3	4	4	3
	400	11	11	10	11	13	17	17	12

where $\lambda_0(t_e|Y = 1)$ is set to 1. Using $f(t_e)$, $\beta_0, \boldsymbol{\beta}, \boldsymbol{\gamma}$ and $\boldsymbol{x}$, we obtained the t_e for patient i as follows:

$$t_{e,i} = \frac{1}{-\exp\left(\boldsymbol{\gamma}^{\mathrm{T}}\boldsymbol{x}_i\right)} \times \log\left[\frac{S_{e,i} - \exp\left(\beta_0 + \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{x}_i\right) / \left\{1 + \exp\left(\beta_0 + \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{x}_i\right)\right\}}{1 - \exp\left(\beta_0 + \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{x}_i\right) / \left\{1 + \exp\left(\beta_0 + \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{x}_i\right)\right\}}\right], \quad (14)$$

where $S_{e,i}$ is a uniform random number between 0 and 1. The probability density function of time-to-censor t_c is $f_c(t_c) = \lambda_c \exp(-\lambda_c t_c)$. Given λ_c, we obtained t_c for a patient i:

$$t_{c,i} = -\frac{1}{\lambda_c} \log S_{c,i}, \quad (15)$$

where $S_{c,i}$ is a uniform random number between 0 and 1, and λ_c is set to 0.4 for mild censoring and to 0.7 for a heavy proportion of censoring. Both time-to-event $t_{e,i}$ and time-to-censor $t_{c,i}$ were generated for a patient; if $t_{e,i} \leqslant t_{c,i}$, then the time t_i was set to $t_{e,i}$; otherwise, the time t_i was set to $t_{c,i}$. We considered two variables $\boldsymbol{x} = (x_1, x_2)$, which were independently generated from a normal distribution with mean vector $\mathbf{0}$, and covariance matrix with variance $\mathbf{1}$, respectively. For generating y_i and v_i, we assumed the following: if $t_i > 10$, then $y_i = 0$ and $v_i = 1$; if $t_i \leqslant 10$ and $\delta_i = 1$, then $y_i = 1$ and $v_i = 1$; and if $t_i \leqslant 10$ and $\delta_i = 0$, then y_i is a missing value and $v_i = 0$.

We performed simulation studies with four scenarios, as shown in Table I. Given the true AUC of 0.6 (or 0.8), we determined the corresponding coefficients $(\beta_1, \beta_2, \gamma_1, \gamma_2)$ that satisfied each target AUC. In scenario 1, x_1 and x_2 influenced both the cure rate and the hazard for uncured patients, respectively. In scenario 2, x_1 affected the cure rate, while x_2 influenced the hazard for uncured patients. In scenario 3, x_1 influenced the hazard for uncured patients, while x_2 affected the cure rate. In scenario 4, x_1 and x_2 influenced only the cure rate, respectively. The num-

Table III. Estimates of AUCs based on the FI and MSI methods.

			AUC = 0.6				AUC = 0.8			
			Scenario				Scenario			
λ_c	n	Method	1	2	3	4	1	2	3	4
0.4	100	Ordinal	0.70	0.61	0.64	0.68	0.85	0.78	0.77	0.81
		FI	0.67	0.63	0.63	0.67	0.79	0.84	0.83	0.83
		MSI	0.69	0.64	0.64	0.69	0.84	0.85	0.84	0.84
	400	Ordinal	0.63	0.59	0.60	0.61	0.84	0.79	0.79	0.80
		FI	0.61	0.60	0.60	0.62	0.78	0.85	0.85	0.81
		MSI	0.65	0.62	0.62	0.63	0.85	0.87	0.86	0.84
0.7	100	Ordinal	0.75	0.62	0.67	0.71	0.81	0.70	0.73	0.78
		FI	0.68	0.63	0.63	0.67	0.80	0.84	0.83	0.81
		MSI	0.70	0.64	0.64	0.69	0.84	0.84	0.84	0.83
	400	Ordinal	0.69	0.56	0.66	0.63	0.88	0.76	0.80	0.83
		FI	0.62	0.60	0.60	0.61	0.79	0.85	0.85	0.80
		MSI	0.66	0.61	0.61	0.63	0.85	0.86	0.86	0.82

Table IV. Estimates of bias-corrected AUCs based on the FI and MSI methods.

			AUC = 0.6				AUC = 0.8			
			Scenario				Scenario			
λ_c	n	Method	1	2	3	4	1	2	3	4
0.4	100	FI	0.63	0.61	0.61	0.65	0.74	0.79	0.78	0.79
		MSI	0.67	0.63	0.63	0.67	0.79	0.80	0.80	0.81
	400	FI	0.58	0.59	0.60	0.61	0.75	0.80	0.80	0.79
		MSI	0.62	0.60	0.60	0.62	0.80	0.82	0.82	0.82
0.7	100	FI	0.64	0.61	0.61	0.64	0.75	0.78	0.78	0.77
		MSI	0.67	0.62	0.62	0.67	0.79	0.80	0.80	0.79
	400	FI	0.57	0.59	0.60	0.61	0.73	0.76	0.76	0.75
		MSI	0.57	0.58	0.58	0.58	0.70	0.73	0.72	0.73

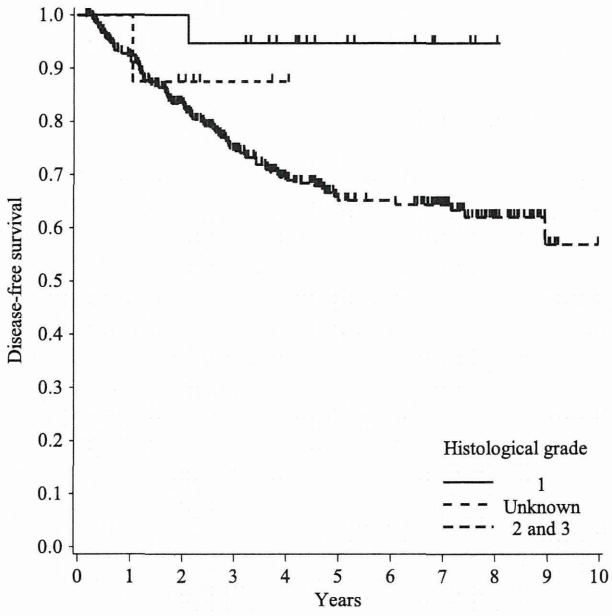

Figure 1. Kaplan–Meier curve for DFS in patients with histological grades of 1, unknown, and 2 and 3 respectively. Short vertical lines indicate censored data points.

ber of patients (N) was set to 100 or 400. The simulations were performed 1000 times for each setting. For the simulations, we programmed the SAS code accordingly using SAS/IML (version 9.3; SAS Institute Inc., Cary, NC, USA).

We investigated the success rates, the proportion of times that we could calculate the AUC for 1000 simulations, for the ordinary AUC calculation (Table II) and calculated the average values of the AUC estimates based on the FI and MSI methods (Table III). The success rates for the ordinary AUC calculations were 35–50% for $\lambda_c = 0.4$ and $N = 100$, 85–95% for $\lambda_c = 0.4$ and $N = 400$, 5% for $\lambda_c = 0.7$ and $N = 100$, and 10% for $\lambda_c = 0.7$ and $N = 400$. As expected, the success rates for FI-based and MSI-based AUC calculations were 100% for all scenarios. We investigated the average bias that is the difference between the true AUC and estimates of FI-based and MSI-based AUCs in 1000 simulations. As shown in Table III, we found the maximum of bias (0.1) when $N = 100$, while the bias was negligibly small when $N = 400$. The bias for the MSI-based AUC was slightly larger than that for the FI-based AUC in all scenarios.

Additionally, we calculated the average values of the bias-corrected AUC estimates using the FI and MSI methods (Table IV). We found that the bias-corrected estimates of the proposed AUCs successfully compensated the overestimation

Table V. Results of the PH cure models for the breast cancer data using stepwise variable selection ($\alpha_{in} = \alpha_{out} = 0.15$).

Variables	Odds ratio (95% CI)	p-value	Hazard ratio (95% CI)	p-value
Hormone receptor status		<0.001		0.018
Negative	1		1	
Positive	0.31(0.18, 0.51)		2.83(1.20, 6.69)	
Endocrine therapy		Not selected		<0.001
No			1	
Yes			0.06(0.02, 0.22)	
Age		Not selected		Not selected
⩾35				
<35				
Clinical stage		Not selected		0.007
IIA/IIB/IIIA			1	
IIIB/IIIC			1.90 (1.19, 3.01)	
Histological grade		<0.001		0.082
1	1		1	
Unknown	0.38 (0.07, 2.17)		3.72(0.23,61.14)	
2 and 3	0.65 (0.24, 1.78)		8.26(1.13, 60.19)	
HER2		0.098		0.005
Negative	1		1	
Positive	0.65(0.39, 1.08)		1.87 (1.21, 2.88)	
Clinical response		Not selected		<0.001
SD/PD			1	
PR/CR			0.40 (0.24, 0.67)	
Number of lymph node metastases		<0.001		0.019
0	1		1	
1 to 3	0.54 (0.33, 0.91)		1.62 (0.88, 3.00)	
⩾4	0.12 (0.06, 0.22)		2.95 (1.68, 5.19)	

CI, confidence interval; HER2, human epidermal growth factor receptor 2; SD, stable disease; PD, progressive disease; CR, complete response; PR, partial response.

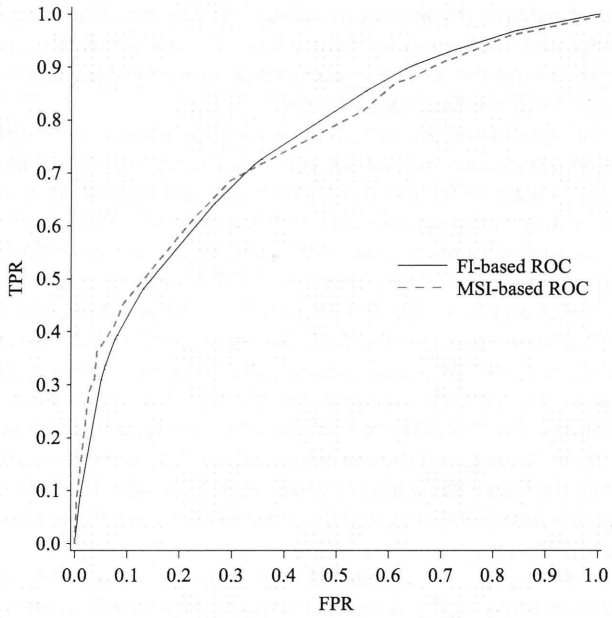

Figure 2. ROC curves for the FI and MSI methods.

in most cases, but over-corrections were observed in the MSI methods when $\lambda_C = 0.7$ and AUC $= 0.8$. We also show the ordinary AUC estimates for reference, although the ordinal AUC could not necessarily be calculated in each simulation because many patients were censored and their cure status was treated as 'missing', as shown in Table II.

4. APPLICATION TO BREAST CANCER DATA

We applied our method to a dataset of 368 breast cancer patients treated with NAC. For example, Figure 1 shows the Kaplan–Meier curves for DFS in patients with histological grades of 1 and 2–3; the respective 8-year DFS rates were 94.5% and 62.1%, respectively. The patients with a histological grade of 1 who had not dropped out were considered clinically 'cured'. This indicates that a cure model would fit the data well. We, therefore, applied the Cox PH cure model to the breast cancer data and calculated the FI-based and MSI-based AUCs. The following variables at baseline were assessed in the model: hormone receptor status (Positive versus Negative), endocrine therapy (Yes versus No), age (<35 years versus ⩾35 years), clinical stage at diagnosis (IIA, IIB, or IIIA versus IIIB or IIIC), histological grading (Unknown versus 1 versus 2 or 3), human epidermal growth factor 2 status (Positive versus Negative), clinical response to NAC (complete or partial response versus stable (SD) or progressive disease), and the number of lymph-node metastases (0 versus 1–3 versus ⩾4). We estimated the parameters with an EM algorithm, using the method developed by Sy and Taylor [6]. Table V shows the results of the analysis using the PH cure model with stepwise variable selection ($\alpha_{in} = \alpha_{out} = 0.15$) [19]. Using a 10-year as period that determines the clinical 'cure', we estimated the ordinary, FI-based, and MSI-based AUCs. Figure 2 shows the ROC curves for the FI and MSI methods. The estimates of the AUC were 0.756 for the FI-based method and 0.755 for the MSI-based method; however, the ordinary AUC could not be calculated because many patients were censored and their cure status was treated as 'missing'. Additionally, the estimates of the bias-corrected AUC were 0.705 for the FI-based method and 0.703 for the MSI-based method.

5. DISCUSSION

In this study, we proposed to use the FI-based and MSI-based AUCs in order to assess the predictive accuracy of cure rates estimated by the PH cure model. We have shown that the FI-based and MSI-based AUC can be estimated irrespective of the amount of censored data (i.e., missing data), although the ordinary AUC cannot. In addition, simulation studies showed that the precision of the estimates was satisfactory and the FI-based AUC was superior to the MSI-based one in many cases. We would recommend the use of the FI-based AUC for assessing the predictive accuracy of cure rates from the Cox PH cure model. In the application to the breast cancer data, the estimates of the FI-based and MSI-based AUCs were both 0.75, and those values could be interpreted as the moderate accuracy [25].

We have investigated the properties of the proposed AUCs. The consistency of the FI-based and MSI-based AUCs have been demonstrated by Alonzo and Pepe [20], although these AUCs seem to be overestimated. Because the proposed AUCs are based on those of Alonzo and Pepe [20], we would expect to see similar properties in the two proposed AUCs. Indeed, the simulation studies showed that the proposed AUCs were overestimated. We performed simulation studies where the number of patients was set to 1000 with end-of-study censoring. The proposed AUCs were identical to their true values when the sample size was as large as 1000 (data not shown). To correct the overestimation of the proposed AUC estimates, we also introduced the bias-corrected estimators of the proposed AUCs using the bootstrap method. Simulation studies demonstrated that the bias-corrected estimates successfully compensated the overestimation in most cases, but over-corrections were observed in the MSI-based AUC when $\lambda_C = 0.7$ and AUC $= 0.8$. When applied to breast cancer data, the original estimates of the FI-based and MSI-based AUCs were both 0.75, while the corresponding bias-corrected estimates were both 0.70.

As cancer treatments are rapidly developing, therapies that result in an increased proportion of patients being long-term survivors (i.e., cured patients) are likely to be encountered more and more frequently. Although our study focused on breast cancer, the cure rate model could be useful for other types of cancer. Therapies for a number of cancer types are believed to induce a cure among some groups of patients. In such cases, the use of the Cox PH cure model is preferable to the standard Cox PH model. The use of the Cox PH cure model in cancer prognostic studies has been recommended in recent studies [2,26–28]. Therefore, further research into regression modeling strategies using the Cox PH cure model is warranted. We recently developed a stepwise variable selection method for the Cox PH cure model [19]. The assessment of model fit, collinearity, identification of overly influential observations, and the validation of the fitted model should be evaluated in more detail in future studies.

Acknowledgements

The views expressed in this article are those of the authors and do not necessarily reflect the official views of Pharmaceuticals and Medical Devices Agency.

REFERENCES

[1] Rastogi P, Anderson SJ, Bear HD, Geyer CE, Kahlenberg MS, Robidoux A, Margolese RG, Hoehn JL, Vogel VG, Dakhil SR, Tamkus D, King KM, Pajon ER, Wright MJ, Robert J, Paik S, Mamounas EP, Wolmark N. Preoperative chemotherapy: updates of national surgical adjuvant breast and bowel project protocols B-18 and B-27. *Journal of Clinical Oncology* 2008; **26**:778–785.

[2] Chen MH, Kim S. Cure rate models with applications to melanoma and prostate cancer data. In *Design and analysis of clinical trials with time-to-event endpoints, Chapter 15*, Peace KE (ed.) Chapman & Hall/CRC: FL; 2009; pp. 349–370.

[3] Farewell VT. The use of mixture models for the analysis of survival data with long-term survivors. *Biometrics* 1982; **38**:1041–1046.

[4] Kuk AYC, Chen C. A mixture model combining logistic regression with proportional hazards regression. *Biometrika* 1992; **79**:531–541.

[5] Peng Y, Dear KB. A nonparametric mixture model for cure rate estimation. *Biometrics* 2000; **56**:237–243.

[6] Sy JP, Taylor JMG. Estimation in a Cox proportional hazards cure model. *Biometrics* 2000; **56**:227–236.

[7] Harrell FE, Califf RM, Pryor DB, Lee KL, Rosati RA. Evaluating the yield of medical tests. *Journal of the American Medical Association* 1982; **247**:2543–2546.

[8] Harrell FE, Lee KL, Mark DB. Tutorial in biostatistics multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in Medicine* 1996; **15**:361–387.

[9] Pencina MJ, D'Agostino RB. Overall C as a measure of discrimination in survival analysis: model specific population value and confidence interval estimation. *Statistics in Medicine* 2004; **23**:2109–2123.

[10] Gönen M, Heller G. Concordance probability and discriminatory power in proportional hazards regression. *Biometrika* 2005; **92**(4):965–970.

[11] Antolini L, Boracchi P, Biganzoli E. A time-dependent discrimination index for survival data. *Statistics in Medicine* 2005; **24**:3927–3944.

[12] Heagerty PJ, Zheng Y. Survival model predictive accuracy and ROC curves. *Biometrics* 2005; **61**:92–105.

[13] Chambless LE, Diao G. Estimation of time-dependent area under the ROC curve for long-term risk prediction. *Statistics in Medicine* 2006; **25**:3474–3486.

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

[15] Negelkerke NJD. A note a general definition of the coefficient of determination. *Biometrika* 1991; **78**:691–692.

[16] Hosmer DW. *Applied logistic regression* (2nd edn). Wiley: New York, 2000.

[17] Steyerberg EW. *Clinical prediction models; a practical approach to development, validation, and updating*. Springer: New York, 2009.

[18] Hirata T, Shimizu C, Yonemori K, Hirakawa A, Kouno T, Tamura K, Ando M, Katsumata N, Fujiwara Y. Change in the hormone receptor status following administration of neoadjuvant chemotherapy and its impact on the long-term outcome in patients with primary breast cancer. *British Journal of Cancer* 2009; **101**:1529–1536.

[19] Asano J, Hirakawa A, Hamada C. A stepwise variable selection for a Cox proportional hazards cure model with application to breast cancer data. *Japanese Journal of Biometrics* 2013; **34**:21–34.

[20] Alonzo TA, Pepe MS. Assessing accuracy of a continuous screening test in the presence of verification bias. *Journal of the Royal Statistics Society, Series C* 2005; **54**:173–190.

[21] Cox DR. Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society, Series B* 1972; **34**:187–220.

[22] Hanley JA, Therneau TM. The meaning and use of the area under the receiver operating characteristic (ROC) curve. *Radiology* 1982; **143**:29–36.

[23] Zhou XH, Obuchowski NA, McClish DK. *Statistical methods in diagnostic medicine* (2nd edn). Wiley: New York, 2011.

[24] Copas JB, Corbett P. Overestimation of the receiver operating characteristic curve for logistic regression. *Biometrika* 2002; **89**(2):315–331.

[25] Swets JA. Measuring the accuracy of diagnostic systems. *Science* 1988; **240**:1285–1293.

[26] Othus M, Barlogie B, LeBlanc ML, Crowley JJ. Cure models as a useful statistical tool for analyzing survival. *Clinical Cancer Research* 2012; **18**:3731–3736.

[27] Gardin C, Chevret S, Pautas C, Turlure P, Raffoux E, Thomas X, Quesnel B, de Revel T, de Botton S, Gachard N, Renneville A, Boissel N, Preudhomme C, Terre C, Fenaux P, Bordessoule D, Celli-Lebras K, Castaigne S, Dombret H. Superior long-term outcome with idarubicin compared with high-dose daunorubicin in patients with acute myeloid leukemia age 50 years and older. *Journal of Clinical Oncology* 2013; **31**:321–327.

[28] Shah A, Andersson TML, Rachet B, Bjorkholm M, Lambert PC. Survival and cure of acute myeloid leukemia in England, 1971-2006: a population-based study. *British Journal of Haematology* 2013; **162**:509–516.