An Extended Cox Prognostic Model of ER/PR+ and HER2− Breast Cancer Patients

YIQUN XIE (xieyiqun@aliyun.com)
Huangpu Branch, Shanghai Ninth people's Hospital, Shanghai JiaoTong University School of Medcine
https://orcid.org/0000-0002-3722-8215

WENTING CUI
Huangpu Branch, Shanghai Ninth People's Hospital, Shanghai JiaoTong University School of Medicine, P.R. China

YANG LIU
School of Statistics, East China Normal University, P.R. China

XIZHOU LI (lixizhou721@126.com)
Department of General surgery, Changhai Hospital, Second Military Medical University, P.R. China.

Research

Keywords: Breast cancer, Estrogen receptor, Progesterone receptor, Human epidermal growth factor receptor 2, Clinicopathological prognostic factor, Prognostic model

Posted Date: June 12th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-33841/v1

Abstract

Background: The purpose of this study was to explore a new ER/PR+ and HER2− breast cancer prognostic model called the extended Cox prognostic model by us for determining the cut-off values for multiple continuous prognostic factors and their interaction via the new modeling idea and variable selection method.

Methods: A total of 335 patients with ER/PR+, HER2- breast cancer were enrolled for the final analysis. The primary endpoint was breast cancer-specific mortality (BCSM). The prognostic factors (histological grade, histological type, stage, T, N, lymphovascular invasion, P53, Ki67, ER, PR, age) were included in this study. The four continuous valuables (Ki67, ER, PR, age) were partitioned into a series of binary variables which all were fitted in the multivariate Cox analysis. A smoothly clipped absolute deviation (SCAD) variable selection method was used. Model performance was expressed in discrimination and calibration.

Results: We developed an extended Cox model with a time threshold at 164 weeks (more than 3 years) post-operation. We found that the cut-off values for PR, Ki67 and age were 20%, 60% and 41-55 years respectively. There was interaction between age and PR for the patients with age≥41years and PR≥20% after 164 weeks post-operation. The patients with age≥41years and PR≥20% after 164 weeks post-operation had relatively higher mortality than before 164 weeks postoperatively.

Conclusions: Our study would offer the guidance in the prognosis for the patients with ER/PR+ and HER2− breast cancer in China. The new idea would be one of ideas for modeling and determining the cut-off values of prognostic factors in future.

0. Introduction

Breast cancer is a malignant carcinoma with the highest occurrence among Chinese women. In China, Shanghai with the increase on breast cancer, the studies on the prognostic models of breast cancer is especially becoming to be needed in the past years [1]. Recently the prognostic gene signatures (Oncotype DX, Mammaprint, and so on) are taken more seriously, however the current prognostic gene signatures are not ready to be used in the clinic practice due to a plethora of concerns in the cost and technology, regardless of first- or second-generation gene signatures [2]. Furthermore, recently the prognostic value of the classic clinicopathologic variables is taken seriously once more [3–5]. And some evidence indicates that the clinicopathologic variables models are excellent surrogates for the prognostic gene signatures. Hence the classic clinicopathologic variables models are highly valued in China due to the feasibility in clinic practice. From 2006 to present, the National Comprehensive Cancer Network (NCCN) had classified invasive breast cancer into four subtypes [6]. The molecular subtype of breast cancer was a classification method similar to intrinsic subtype and was more suitable for current clinical practice in China, also serving as an independent prognostic factor [7, 8]. Among the four molecular subtypes of breast cancer, ER and/or PR + and HER2- (estrogen receptor and/or progesterone receptor positive and human epidermal growth factor receptor 2 negative) occurred most commonly and

accounted for approximately 60% of breast cancer patients [8, 9]. There was more urgent demands and wider impacts to explore the improved prognostic model for patients with ER/PR + and HER2 − breast cancers based upon classic clinicopathologic variables to meet the unique clinic needs in China.

The current classic prognostic algorithms (PREDICT, Adjuvant! Online, and Nottingham Prognostic Index) are far from ideal [10–12]. These models were often based upon the data sets from non-Chinese or non-Asian patients. Specifically, it was assumed that these prognostic factors, including Ki67 (a nuclear marker of cell proliferation), estrogen receptor (ER), progesterone receptor (PR) and Age, were continuous factors, or it was assumed that the cut-off values of prognostic factors, including Ki67, estrogen receptor (ER), progesterone receptor (PR) and Age, were determined merely based on univariate analysis, experience or speculation. In addition, existing models ignored the interaction effect between the prognostic factors. Therefore, the current models showed poor accuracy and were not suitable to the clinic practice in China. It is critical to develop novel improved prognostic algorithm to analyze the clinic data from China.

For this reason, in this study we selected 335 patients with ER/PR + and HER2 − breast cancer to explore a new ER/PR + and HER2 − breast cancer prognostic model using classical clinicopathologic variables, called the extended Cox prognostic model by us. The cut-off values for multiple continuous prognostic factors were determined and the interaction effect between the factors was elucidated via the new modeling idea and variable selection method.

1. Patients And Methods

1.1 Study population

All patients with invasive unilateral breast cancer admitted to the Department of Breast Surgery at Huangpu Branch, Shanghai Ninth people's Hospital, Shanghai JiaoTong University School of Medicine from January 2009 to December 2009 were evaluated. Information obtained from medical records included age at diagnosis, number of lymph nodes sampled and number of positive lymph nodes (categorised as 0, 1 to 3, 4 to 9, and 10 + nodes positive [4]), lymphovascular invasion (categorised as positive or negative), tumour size (categorised as < 21 mm, 21 to 50 mm, 50 + mm[4]), histological grade (categorised as I, II, III [4]), pathological type, protein 53 (p53) status, proliferating cell nuclear antigen Ki67 (Ki67) status, ER status, PR status, HER2 status, information on local therapy (wide local excision, mastectomy, radiotherapy), and type of adjuvant systemic therapy (chemotherapy, endocrine therapy, both). The patients with any one of the following conditions were excluded from the analyses, including mucinous carcinoma, cribriform carcinoma, tubular carcinoma[6], or incomplete information, or who received chemotherapy or radiation before operation, or who did not undergo surgery, or who did not complete local treatment (local excision without radiotherapy), or no axillary lymph node dissection, or who did not complete adjuvant systemic therapy (chemotherapy and endocrine therapy), leaving 568 individuals. Variables for each patient included age, TNM status, T stage, N stage, pathological subtype,

histology grade, lymphovascular invasion, p53 status, Ki67 status, ER status, PR status, vital status, and survival time.

1.2 Categorize the patients into ER/ PR + and HER2- subtype of breast cancer

All formalin-fixed paraffin-embedded (FFPE) tumor blocks were collected at the time of surgery prior to adjuvant therapy and were stored at room temperature. Approval was obtained from Institutional Review Boards (IRB). Tumor sections of 4 to 10 μm were cut. One section was stained by hematoxylin/eosin (HE) to confirm the presence of invasive carcinoma, and other sections were used for molecular analyses by two independent pathologists. ER, PR and HER2 were assayed by immunohistochemistry （IHC） and evaluated according to standard criteria[9, 13, 14].

The criteria for evaluating ER and PR in breast cancer cells by IHC [13]: ER or PR was positive if the cell nuclei showed brown color. In one section, five high-power regions were selected randomly. The patient was assigned to subtype ER/PR + if the percentage of positive cells was ≥ 1% in these regions,. The criteria for evaluating HER2 by IHC: the patients were categorized into four subtypes: 0, 1+, 2+, or 3+ [9, 14]. Subtypes HER2 IHC 0 and 1 + were HER2 negative [HER2 (-)]. Subtype HER2 IHC 2 + was equivocal in HER2 status. Finally subtype HER2 IHC 3 + was HER2 positive. Tumor cells of subtype HER2 IHC 2 + were further analyzed by fluorescence in situ hybridization (FISH). According to the average copy number of HER2 protein or the HER2/chromosome 17 centromere (CEP17) ratio, the FISH results were considered positive, equivocal, or negative [9]. Patients with equivocal FISH result were excluded from the study. Among the 568 patients, the HER2 status of 185 individuals was detected as HER2 IHC (++) [9, 14]. Tissue microarrays (TMAs) were constructed from the tissue core with 1.5 mm in diameter to detect HER2 status of these 185 patients using FISH methods. Finally, the 335 individuals of 568 patients were determined as ER/PR + and HER2 − type, a study population.

1.3 Treatment for the patients.

All patients underwent modified radical mastectomy/breast-conserving surgery and adjuvant chemotherapy. They were treated with four to six cycles of CEF (cyclophosphamide, epirubicin, and fluorouracil) chemotherapy, or four cycles of CEF followed by four cycles of T (docetaxel) chemotherapy, or four to six cycles of TEC (docetaxel, epirubicin, and cyclophosphamide) chemotherapy after surgery. If necessary, patients would receive postoperative radiotherapy followed with endocrine therapy, but not trastuzumab treatment.

1.4 Follow-up

The primary endpoint of this study was breast cancer-specific mortality (BCSM), which was determined by following up the survival of patients over certain time period. From the first day after surgery till the death or the end of the study (September 15 2016), we made telephone calls and/or outpatient visits every 3 months to follow up the survival status of the patients and the cause of death. The follow-up

records were double-checked with that from Department of Cancer Control & Prevention, Shanghai Municipal Center for Disease Control & Prevention.

1.5 Statistical analysis

The extended Cox prognostic model was developed in all eligible patients as follows: First, to determine the cut-off values for each continuous valuables (Ki67, ER, PR, age), these four continuous valuables were partitioned into a series of binary variables. Second, all variables were fitted in the multivariate Cox analysis. A SCAD (smoothly clipped absolute deviation) variable selection method [15] was used to build up a Cox prognostic model to determine the independent variables, the cut-off values and interaction effect between different factors. Finally, during developing the model, we reckoned that the model could be divided into two parts by a certain time point. We built up a new model, named as extended Cox prognostic model, with a time threshold at 164 weeks (more than 3 years) based upon the smallest Akaike information criterion (AIC) value.

We evaluated the predictive accuracy of extended Cox prognostic model based upon the parameters of discrimination and calibration. For the model discrimination, the receiver-operator-characteristic (ROC) curves were plotted for the data at 1-year, 3-year and 5-year post operation [16], respectively. And we also calculated the areas under the receiver-operator-characteristic (ROC) curves (AUC). Model calibration was assessed by a simplified goodness-of-fit (GOF) method [17]. We compared the number of deaths observed and calculated at 5-year post-operation. We grouped the risk scores into 5 sets and then calculated the GOF statistics for each set. This provides a goodness of fit Chi-square test. All analyses were conducted using R software version 3.3.2.

2. Results

2.1 Patient characteristics

All of 335 patients were females. The median age of all patients was 53 years old with a range of 26–89 years old. All clinical and pathological data of the patients was complete, including age, tumor size, axillary lymph node status, lymphovascular invasion status, histological grade, pathological type, ER/PR/HER2/Ki67/P53 status, surgery and postoperative adjuvant treatment (specified in chemotherapy, targeted therapy, radiotherapy, and endocrine therapy). Among 335 patients, 270 cases had nonspecific invasive breast cancer, 15 cases had invasive lobular carcinomas, 14 cases had mixed carcinomas, 15 cases had micropapillary carcinomas, 5 cases had invasive papillary carcinomas, 5 cases had neuroendocrine carcinomas, 1 case had metaplastic carcinoma, and 10 cases had intraductal and microinvasive carcinomas. Among 335 patients, 22 (6.6%), 267 (79.7%), and 46 (13.7%) had histologic grade I, II, and III tumors, respectively [4]. According to 2017 American Joint Committee on Cancer(AJCC)staging system[4, 5], 206 (61.5%), 126 (37.6%), 3 (0.9%), and 0 patients had stage T1, T2, T3, and T4, respectively, while 98 (29.3%), 41 (12.2%), and 16 (4.8%) patients had stage N1, N2, and N3, respectively. Of the patients, 130 (38.8%) were in stage I, 150 (44.8%) in stage II, and 55 (16.4%) in stage

III. No lymph node metastasis was found in 180 patients (53.7%), while vascular tumor emboli were present in 115 patients (34.3%).

2.2 Prognostic factors

According to whether the prognostic factor was categorical or continuous, Table I listed the prognostic factors included in this study. To determine the cut-off values for continuous variables (Ki67, ER, PR, Age), we transformed each continuous factor (Ki67, ER, PR, Age) into a series of binary variables using each observed value as the partition point. Table II provides the possible partition points for these four continuous factors. Each partition point converted this continuous variable into one binary variable. Taken Ki67 as an example, point 60% converted Ki67 into one binary variable named as “Ki67 status (60%)”, which would be one if the immunochemical status of Ki67 is no less than 60% and zero otherwise. Besides, it was mandatory that each sample size of each binary variable was not less than 10. In addition, according to clinical practice, the age factor was also transformed into two other categorical variables sets. Set one assigned 1: age < 40 years; 2: 40 ≤ age < 60 years; 3: age ≥ 60 years. Set two categorized 1: age < 35 years; 2: 35 ≤ age < 65 years; 3: age ≥ 65 years. For each patient case, 97 variables were determined, including categorical factors (histological grade, histological type, stage, T, N, lymphovascular invasion), continuous variables (Ki67, ER, PR, Age) and a series of binary variables. Based on previous clinic study and clinical practice [18], we introduced four potential interaction terms of T and N, ER and PR, age and ER, age and PR. So we determined 1730 variables finally with the addition of interaction terms.

Categorical factors	histological grade, histological type, stage, T, N, lymphovascular invasion, P53
Continuous factors	Ki67, ER, PR, age

Table I
Original prognostic factors included in this study

Prognostic factor	Partition points
Ki67 (%)	1, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80
ER (%)	10, 15, 20, 30, 40, 50, 60, 65, 70, 75, 80, 85, 90, 95, 98
PR (%)	5, 10, 15, 20, 25, 30, 40, 50, 60, 65, 70, 80, 85, 90, 95
age (years)	37, 38, 39, …… , to 75, 77

Table II
Partition points of continuous factors Ki67, ER, PR and Age

On September 15th 2016, 311 of the 335 patients were still being followed, 24 were lost from follow-up (including one non-cancer death). Among the 311 patients, 28 patients died from breast cancer. The median follow-up duration was 370 weeks with the censoring rate 91.6%. The longest survival time of death case was 341 weeks. The survival probability at 341 weeks (about six and one half years) of breast cancer patients was 91.2% (Fig. 1).

2.4 Prognostic model development

2.4.1 Cox proportional hazards model without interaction terms

We first introduced 97 variables without interaction items into the model, and a SCAD variable selection method [15] was used to develop the Cox proportional hazard model (named as Model 1).

$$\mathrm{h}(t) = h_0(t)\exp\{1.38x_1 + 1.01x_2 + 1.643x_3 - 0.949x_4 + 1.04x_5\} \text{ Eq. (1)}$$

The prognostic factors, Beta-coefficients, hazard ratios for prognostic factors, and the P values of Z tests are provided in Table III.

Variable	Prognostic factor	Coefficient	HR(95%CI)	p-value
X_1	Histological grade (I, II, III)	1.380	3.976(1.851,8.843)	<0.001
x_2	N status (0, 1, 2, 3)	1.010	2.746(1.959,3.849)	<0.001
X_3	Ki67 status (60%)	1.643	5.172(2.288,11.691)	<0.001
X_4	PR status (20%)	-0.949	0.387(0.181,0.826)	0.014
X_5	Age (55years)	1.040	2.828(1.285,6.222)	0.01
HR = hazard ratio, CI = confidence interval.				
*Ki67 status (60%): a binary variable which is one if Ki67 is no less than 60% and zero otherwise.				
*PR status (20%): a binary variable which is one if PR is no less than 20% and zero otherwise.				
*Age (55years): a binary variable which is one if age is no less than 55 years and zero otherwise.				

Table III
Prognostic factors, coefficients, hazard ratios, and P values of the Z tests in Model 1

The multivariable Cox Regression analysis for BCSM without interaction terms, Model 1, showed that histological grade, N status, Ki67 status, PR status and age were statistically significant to predict patient survival. The cut-off value for Ki67 was 60%, the cut-off value for PR was 20%, and the cut-off value for age was 55 years old.

nteraction terms

Secondly, 1730 variables included interaction terms were used in the model to evaluate the interaction effect between prognostic factors. A SCAD variable selection method [15] was also used to develop the Cox proportional hazard model (named as Model 2).

$$\mathrm{h}(t) = h_0(t)\exp\left\{1.392x_1 + 0.995x_2 + 1.595x_3 + 1.202x_5 - 1.157x_6\right\} \quad \text{Eq. (2)}$$

The prognostic factors, Beta-coefficients, hazard ratios for prognostic factors, and the P values of Z tests are provided in Table IV.

Variable	Prognostic factor	Coefficient	HR(95%CI)	p-value
X_1	Histological grade (I, II, III)	1.392	4.024(1.865,8.684)	<0.001
X_2	N status (0, 1, 2, 3)	0.995	2.706(1.929,3.795)	<0.001
X_3	Ki67 Status (60%)	1.595	4.930(2.181,11.143)	<0.001
X_5	Age (55years)	1.202	3.325(1.496,7.389)	0.003
X_6	PR - age (Age 41 >=years, PR >=20%)	-1.157	0.315(0.148,0.669)	0.003

HR = hazard ratio, CI = confidence interval.

* PR - age (Age 41 >=years, PR >=20%) : a binary variable, an interaction item between age and PR, which is one if age is no less than 41years and PR is no less than 20% and zero otherwise

Table IV
Prognostic factors, coefficients, hazard ratios, and P values of the Z tests in Model 2

Model 2 showed that there may be an interaction effect between PR (20%) and age (41 years). But PR (20%) and age (41 years) were not statistically significant to predict patient survival and the cut-off for age was 41 years old, which was different to Model. These indicated that Model 2 may not be self-sufficient.

2.4.3 Extended Cox prognostic model

In order to further optimize the two above Cox models, we first included all variables in model 1 and model 2, which were histological grade, N status, Ki67 status (60%), PR status (20%), age (55 years), age(41 years) and PR-age (age ≥ 41years, PR ≥ 20), into the Cox model. The outcome indicated PR status (20%)□age (41 years old) and PR-age was not statistically significant. Furthermore, taking previous clinic experience into account, we were aware that the interaction between PR and age may vary according to the different time periods after surgery and that the model was maybe divided into two parts by a certain time point. To determine the optimal time point, we considered all of both the death and

Cox model 3 (named as “extended Cox prognostic model”)

with a time threshold at 164 weeks (more than 3 years) after surgery based upon the smallest Akaike information criterion (AIC) value.

$$h(t) = \begin{cases} h_0(t)\exp\{A\}, & t \le 164 weeks \\ h_0(t)\exp\{A + 1.594x_8\}, & t > 164 weeks \end{cases} \quad \text{Eq. (3)}$$

$$A = 1.329x_1 + 0.972x_2 + 1.65x_3 - 1.802x_4 + 1.378x_5 - 1.563x_7$$

.

The prognostic factors, Beta-coefficients, hazard ratios for prognostic factors, and the P values of Z tests are provided in Table V and Table VI.

Variable	Prognostic factor	coefficient	HR(95%CI)	p-value
X_1	Histological grade (I, II, III)	1.329	3.777(1.739–8.204)	<0.001
X_2	N status (0, 1, 2, 3)	0.972	2.644(1.872–3.734)	<0.001
X_3	Ki67 Status (60%)	1.650	5.208(2.311,11.738)	<0.001
X_4	PR Status (20%)	-1.802	0.165(0.050–0.548)	0.003
X_5	Age (55years)	1.378	3.968(1.548,10.171)	0.004
X_7	Age (41years)	-1.563	0.210(0.053–0.829)	0.026
X_8	PR - Age after 164 weeks (Age 41 >=years, PR >=20%)	1.594	4.922(1.039,23.325)	0.045

HR = hazard ratio, CI = confidence interval.

*Age (41years): a binary variable which is one if age is no less than 41 years and zero otherwise.

*PR-age after 164 weeks (Age ≥41years, PR ≥20%): a binary variable, an interaction item between age and PR status after 164 weeks after surgery, which is one if age is no less than 41 years old and PR is no less than 20% and zero otherwise.

Table V
Prognostic factors, coefficients, hazard ratios, and P values of the Z tests in the extended Cox model

Prognostic factor	Coefficient	HR	95% CI	P-value
Histological grade (I, II, III)	1.33	3.78	1.74 to 8.20	<0.001
N status (0, 1, 2, 3)	0.97	2.64	1.87 to 3.73	<0.001
Ki67 status				
low	0.00	1.00		
high	1.65	5.21	2.31 to 11.74	<0.001
PR status				
low	0.00	1.00		
high	-1.80	0.17	0.05 to 0.55	0.003
age*				
young	0.00	1.00		
middle-aged	-1.56	0.21	0.05 to 0.83	0.025
elderly	-0.18	0.83	0.25 to 2.72	0.761
PR-age after 164 weeks*				
0	0.00	1.00		
1	1.59	4.92	1.04 to 23.33	0.045

HR = hazard ratio, CI = confidence interval.

* As to the age factor, the patients were divided into three groups: youth group (<41 years old), middle-aged group (41–55 years old) and elderly group (55 and above years old).

* means interaction item between PR status and age after 164 weeks postoperatively.

Table VI
Coefficients, hazard ratios (95% CI) and P-values from the prognostic model

The extended Cox prognostic model determined the cut-off values for multiple continuous prognostic factors and the interaction effect between the factors. Firstly, the cut-off values of the prognostic factors were determined by Cox prognostic model with SCAD variable selection method [15]. Among 1730 predictors, histological grade and N status were considered to be categorical factors. Whereas, the model showed Ki67, PR and age were also categorical factors and the prognostic model automatically determined their reasonable cut-off values. For Ki67 expression, a cut-off value at 60% was selected to distinguish between low-expression (< 60%) and high-expression (≥ 60%). For PR expression, a cut-off value at 20% was selected to distinguish between low-expression (< 20%) and high-expression (≥ 20%). As to the age factor, we categorized the patients into three groups: the old group with at least 55 years old

when received surgery, the young group with less than 41 years old and middle-aged group between 41 to 55 years old. Secondly, for the interaction effect, after 164 weeks (more than 3 years) after surgery, there was interaction between age and PR. The patients with age ≥ 41 years and PR ≥ 20% after 164 weeks (more than 3 years) after surgery had relatively higher mortality than before 164 weeks postoperatively.

In our study, we found that the hazard ratio for BCSM increased 2.78 times with the histological grade increasing one level. If N statuses increased one level, the hazard ratio for BCSM increased 1.64 times. The patients with high Ki67 expression had the hazard ratio for BCSM 4.21 times higher than the patients with low Ki67 expression. Within 164 weeks post-operation, the hazard ratio for BCSM of the patients with low PR expression was 4.88 times higher than that of the patients with high PR expression. Within 164 weeks (more than 3 years) post-operation, the patients aged < 41 had the highest hazard ratio for BCSM, followed by the patients aged ≥ 55, while the patients aged 41 to 55 showed the lowest hazard ratio for BCSM. After 164 weeks (more than 3 years) post-operation, there was interaction effect between age and PR for the patients with age ≥ 41 years and PR ≥ 20%. The hazard ratio for BCSM of the patients with age ≥ 41 years and PR ≥ 20% elevated after 164 weeks (more than 3 years) post-operation. For the patients with high PR expression, the age was positively correlated with the mortality after 164 weeks (more than 3 years) post-operation. For patients with high PR expression after 164 weeks (more than 3 years) post-operation, the patients aged 41 to 55 had nearly the same hazard ratio for BCSM as those aged < 41, while the hazard ratio for BCSM of patients aged ≥ 55 years was 3.09 times higher than that of patients aged < 41 years.

2.5 Model discrimination and calibration

As expected, this extended Cox prognostic model showed good discrimination. Figure 2 shows the receiver operator characteristic (ROC) curves for the breast cancer-specific mortality (BCSM) at 1-year, 3-year and 5-year postoperatively. The area under the ROC curves (AUC) for the breast cancer-specific mortality (BCSM) at 1-year, 3-year and 5-year postoperatively are shown in Table VII. AUC values for the breast cancer-specific mortality (BCSM) at 1-year, 3-year and 5-year postoperatively are all larger than 0.80. AUC value for the breast cancer-specific mortality (BCSM) at 3-year post-operation was as high as 0.94, with 95% CI from 0.89 to 0.99. So our extended Cox prognostic model showed good discrimination.

	AUC	95% CI
1-year	0.85	0.78 to 0.93
3-year	0.94	0.89 to 0.99
5-year	0.81	0.72 to 0.90
CI = confidence interval.		

Table VII
The AUC and 95% CI for our extended Cox prognostic model at 1-year, 3-year and 5-year post-operation.

The extended Cox prognostic model was also well calibrated by the goodness-of-fit (GOF) test [17]. We grouped the risk scores into 5 groups and then calculated the GOF statistic 0.54, with P-value 0.76 (> 0.05). These indicated that our extended Cox prognostic model fit was good.

3. Discussion

In this study, we explored an extended Cox model for the prognosis of ER/PR + and HER2 − breast cancer, with calculating the cut-off points of prognostic factors and their interaction. The cut-off values of Ki67, PR and age and the interaction between the age and PR status were generated from model calculation. The model was well calibrated and provided a high degree of discrimination.

Here, we found that the prognosis of patients was associated with histological grade, N, Ki67, PR statuses, and age. And we found that the cut-off values for PR, Ki67 and age were 20%, 60% and 41–55 years respectively. It was important to point out that the cut-off values of prognostic factors, including Ki67, PR and Age, was determined only based on our extended Cox prognostic model (a multivariable analysis), and not based on univariate analysis, experience or speculation. It's different from previous studies [10–12, 19, 20]. The prognosis of patients was associated with histological grade, N, Ki67 and PR statuses, and age, which was consistent with previous reports [18]. Histological grade and N status had a linear effect on the hazard ratio. The cut-off value (20%) for PR was consistent with the St. Gallen consensus of 2013 [21, 22], which indicated the high fidelity of our model. We reckoned that the breast cancer with low PR expression was probably a different intrinsic subtype of breast cancer, Luminal B subtype. Our prognostic model determined 60% as the cut-off value for Ki67 status, which was much higher than those in other studies (e.g., 14% and 20%) [18, 21–24]. We reckoned that this situation was because of a lack of a standardised procedure for Ki67 assessment and the controversial Ki67 assay interpretation [25], and it was also because of determining the cut-off value for Ki67 by the ROC method, an univariate analysis method in previous study [19, 23]. So the prognostic value of Ki67 index in breast cancer was to be further explored. The cut-off value (41 and 55 years) for age was consistent with the age range of perimenopausal and menopause Chinese women though it was different from those of other studies [18, 26]. Hence our models showed high consistence to the current available golden standard, such as PR factor. The values are well correlated to the physiological conditions of our patients, such as age factor. And it was the most important that our model generates the cut-off value based upon algorithm calculation without empirical biases or univariate analysis.

In our study, algorithmic analysis by the extended Cox prognostic model showed that there was interaction among age and PR for the patients with age $\geq$ 41 years and PR $\geq$ 20% after 164 weeks (more than 3 years) post-operation. The interaction among age and PR made the patients with age $\geq$ 41 years and PR $\geq$ 20% relatively higher mortality after 164 weeks (more than 3 years) post-operation. We found that the older the patients with ER/PR+, HER2−, PR $\geq$ 20% were, the lower survival exceeding 164 weeks post surgery they would have. The existence of the interaction between age and PR status had been previously reported [18]. We reckoned this may be related to multiple factors, including the subtype of the health of the patients after 164 weeks (more than

3 years) after surgery. The tumor cells in the different subtypes of breast cancer had different growth characteristics under the different sex hormone levels. Luminal A subtype of breast cancer was more likely to recur and metastasize after 164 weeks (more than 3 years) after surgery in menopause woman. So the drug selection in adjuvant endocrine therapy for the perimenopausal or menopause patients with ER/PR+, HER2−, PR ≥ 20% needed careful consideration. And the endocrine therapy should last long enough. Obviously our study about the interaction was deeper and clearer.

Our studies showed the advantages in the following three aspects. 1. We adopted a new idea to determine the cut-off values only based on our extended Cox prognostic model, a multivariable analysis. The cut-off values for the multiple prognostic factors including age/PR /Ki67 status were determined, which were statistically significant for prognosis. 2. During developing the model, we introduced the potential interactions and reckoned that the model could be divided into two parts by a certain time point. An extended Cox model with a time threshold at 164 weeks (more than 3 years) post-operation was built up based upon statistical analysis. We found that there was interaction between age and PR after 164 weeks post-operation. The hazard ratio for BCSM of the patients with age ≥ 41 years and PR ≥ 20% elevated after 164 weeks (more than 3 years) post-operation. 3. Our study was derived from Chinese clinical data, which could be most relevant model towards Chinese clinical practice. Due to the gene similarity, the model could also apply to the prognosis of breast cancer in other Asian female population. The cut-off value (41 and 55 years) for age was consistent with the age range of perimenopausal and menopause Chinese women, though it was different from those of other studies.

In conclusion, using the new modeling idea and statistical method, an extended Cox prognostic model for the prognosis of ER/PR + and HER2 − breast cancer was explored with calculating the cut-off points of prognostic factors and their interaction. The new idea and statistical method in our study was different from previous studies, especially the study about the cut-off value of age and its interaction. Moreover, the result of our study would offer the guidance in the prognosis and treatment for the patients with ER/PR + and HER2 − breast cancer in China. The new idea used in our study would be one of ideas for determining the cut-off value of prognostic factors in future. And the prognostic model could be divided into two parts by a certain time point, which would be a new idea of developing a prognostic model.

Abbreviations

Estrogen receptor and/or progesterone receptor positive and human epidermal growth factor receptor 2 negative (ER and/or PR+ and HER2-), Breast cancer-specific mortality (BCSM), smoothly clipped absolute deviation (SCAD).

Declarations

Acknowledgements

We grandly thank the Department of Cancer Control & Prevention, Shanghai Municipal Center for Disease Control & Prevention. And we would like to thank Dr. Fangwei Shao of Zhejiang University-University of Illinois at Urbana-Champaign Institute, Zhejiang University, Haining, China.

Funding

This study was supported by the Shanghai Science and Technology Committee (STCSM) (No. 14411972400), Huangpu District Shanghai Municipal Health Commission (No. 2019BJ04), and Huangpu District Shanghai Municipal Health Commission (No. 2019GG05).

Availability of data and materials

All data generated or analyzed during this study are included in this published article.

Author's contributions

YX, XL, designed and supervised the project. YX, TH analyzed the results and drafted the manuscript. YX, WC financed the project. LY, YX conducted the computational analysis. XL, TH, and WC performed the experiments. YX, XL and TH prepared the figures and tables. All authors have read and approved the content of the manuscript.

Ethics approval and consent to participate

All protocols involving the use of humans were approved by the Ethics Committee of Huangpu Branch, Shanghai Ninth people's Hospital, Shanghai JiaoTong University School of Medicine, China. Written informed consent was obtained from all subjects for the use of their tissue in this study.

Consent for publication

Not applicable

Competing interests

The authors declare that they have no competing interests

References

1. Fan L, Strasser-Weippl K, Li JJ, St Louis J, Finkelstein DM, Yu KD, Chen WQ, Shao ZM, Goss PE: **Breast cancer in China.** *Lancet Oncol* 2014, **15:**e279-289.
2. Ribnikar D, Cardoso F: **Tailoring Chemotherapy in Early-Stage Breast Cancer: Based on Tumor Biology or Tumor Burden?** *Am Soc Clin Oncol Educ Book* 2016, **35:**e31-38.
3. Cardoso F, van't Veer LJ, Bogaerts J, Slaets L, Viale G, Delaloge S, Pierga JY, Brain E, Causeret S, DeLorenzi M, et al: **70-Gene Signature as an Aid to Treatment Decisions in Early-Stage Breast Cancer.** *N Engl J Med* 2016, **375:**717-729.
4. Hortobagyi GN, Connolly JL, D'Orsi CJ, Edge SB, Mittendorf EA, Rugo HS, Solin LJ, Weaver DL, Winchester DJ, Giuliano A (Eds.): **Breast. In: AJCC Cancer Staging Manual.8th ed.** New York: Springer International Publishing; 2017.
5. Giuliano AE, Connolly JL, Edge SB, Mittendorf EA, Rugo HS, Solin LJ, Weaver DL, Winchester DJ, Hortobagyi GN: **Breast Cancer-Major changes in the American Joint Committee on Cancer eighth edition cancer staging manual.** *CA Cancer J Clin* 2017, **67:**290-303.
6. Koh WJ, Greer BE, Abu-Rustum NR, Campos SM, Cho KR, Chon HS, Chu C, Cohn D, Crispens MA, Dizon DS, et al: **Vulvar Cancer, Version 1.2017, NCCN Clinical Practice Guidelines in Oncology.** *J Natl Compr Canc Netw* 2017, **15:**92-120.
7. Carey LA, Perou CM, Livasy CA, Dressler LG, Cowan D, Conway K, Karaca G, Troester MA, Tse CK, Edmiston S, et al: **Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study.** *Jama* 2006, **295:**2492-2502.
8. Chen XS, Ma CD, Wu JY, Yang WT, Lu HF, Wu J, Lu JS, Shao ZM, Shen ZZ, Shen KW: **Molecular subtype approximated by quantitative estrogen receptor, progesterone receptor and Her2 can predict the prognosis of breast cancer.** *Tumori* 2010, **96:**103-110.
9. Li J, Chen Z, Su K, Zeng J: **Clinicopathological classification and traditional prognostic indicators of breast cancer.** *Int J Clin Exp Pathol* 2015, **8:**8500-8505.
10. Haybittle JL, Blamey RW, Elston CW, Johnson J, Doyle PJ, Campbell FC, Nicholson RI, Griffiths K: **A prognostic index in primary breast cancer.** *Br J Cancer* 1982, **45:**361-366.
11. Olivotto IA, Bajdik CD, Ravdin PM, Speers CH, Coldman AJ, Norris BD, Davis GJ, Chia SK, Gelmon KA: **Population-based validation of the prognostic model ADJUVANT! for early breast cancer.** *J Clin Oncol* 2005, **23:**2716-2725.
12. Wishart GC, Azzato EM, Greenberg DC, Rashbass J, Kearins O, Lawrence G, Caldas C, Pharoah PD: **PREDICT: a new UK prognostic model that predicts survival following surgery for invasive breast cancer.** *Breast Cancer Res* 2010, **12:**R1.
13. Hammond ME, Hayes DF, Dowsett M, Allred DC, Hagerty KL, Badve S, Fitzgibbons PL, Francis G, Goldstein NS, Hayes M, et al: **American Society of Clinical Oncology/College Of American Pathologists guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer.** *J Clin Oncol* 2010, **28:**2784-2795.
14. Wolff AC, Hammond ME, Hicks DG, Dowsett M, McShane LM, Allison KH, Allred DC, Bartlett JM, **…ndations for human epidermal growth factor receptor 2**

testing in breast cancer: American Society of Clinical Oncology/College of American Pathologists clinical practice guideline update. *Arch Pathol Lab Med* 2014, **138:**241-256.

15. Fan J, Li R: **Variable selection for Cox's proportional hazards model and frailty model.** *The Annals of Statistics* 2002, **30:**74-99.
16. Heagerty PJ, Lumley T, Pepe MS: **Time-dependent ROC curves for censored survival data and a diagnostic marker.** *Biometrics* 2000, **56:**337-344.
17. May S, Hosmer DW: **A simplified method of calculating an overall goodness-of-fit test for the Cox proportional hazards model.** *Lifetime Data Anal* 1998, **4:**109-120.
18. Yao N, Song Z, Wang X, Yang S, Song H: **Prognostic Impact of Progesterone Receptor Status in Chinese Estrogen Receptor Positive Invasive Breast Cancer Patients.** *J Breast Cancer* 2017, **20:**160-169.
19. Zhu X, Chen L, Huang B, Wang Y, Ji L, Wu J, Di G, Liu G, Yu K, Shao Z, Wang Z: **The prognostic and predictive potential of Ki-67 in triple-negative breast cancer.** *Sci Rep* 2020, **10:**225.
20. Qin T, Zeng YD, Lu Q, Zhang X, Qin GE, Zheng Q, Xu F, Peng R, Yuan Z, Wang S: **Nomogram Model of LNR Predicts Survival in Premenopausal Patients with Node-positive Luminal Breast Cancer.** *Anticancer Res* 2017, **37:**4575-4586.
21. Goldhirsch A, Winer EP, Coates AS, Gelber RD, Piccart-Gebhart M, Thurlimann B, Senn HJ: **Personalizing the treatment of women with early breast cancer: highlights of the St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2013.** *Ann Oncol* 2013, **24:**2206-2223.
22. Prat A, Cheang MC, Martin M, Parker JS, Carrasco E, Caballero R, Tyldesley S, Gelmon K, Bernard PS, Nielsen TO, Perou CM: **Prognostic significance of progesterone receptor-positive tumor cells within immunohistochemically defined luminal A breast cancer.** *J Clin Oncol* 2013, **31:**203-209.
23. Cheang MC, Chia SK, Voduc D, Gao D, Leung S, Snider J, Watson M, Davies S, Bernard PS, Parker JS, et al: **Ki67 index, HER2 status, and prognosis of patients with luminal B breast cancer.** *J Natl Cancer Inst* 2009, **101:**736-750.
24. Nishimura R, Osako T, Okumura Y, Hayashi M, Toyozumi Y, Arima N: **Ki-67 as a prognostic marker according to breast cancer subtype and a predictor of recurrence time in primary breast cancer.** *Exp Ther Med* 2010, **1:**747-754.
25. Penault-Llorca F, Radosevic-Robin N: **Ki67 assessment in breast cancer: an update.** *Pathology* 2017, **49:**166-171.
26. Liedtke C, Rody A, Gluz O, Baumann K, Beyer D, Kohls EB, Lausen K, Hanker L, Holtrich U, Becker S, Karn T: **The prognostic impact of age in different molecular subtypes of breast cancer.** *Breast Cancer Res Treat* 2015, **152:**667-673.

Figures

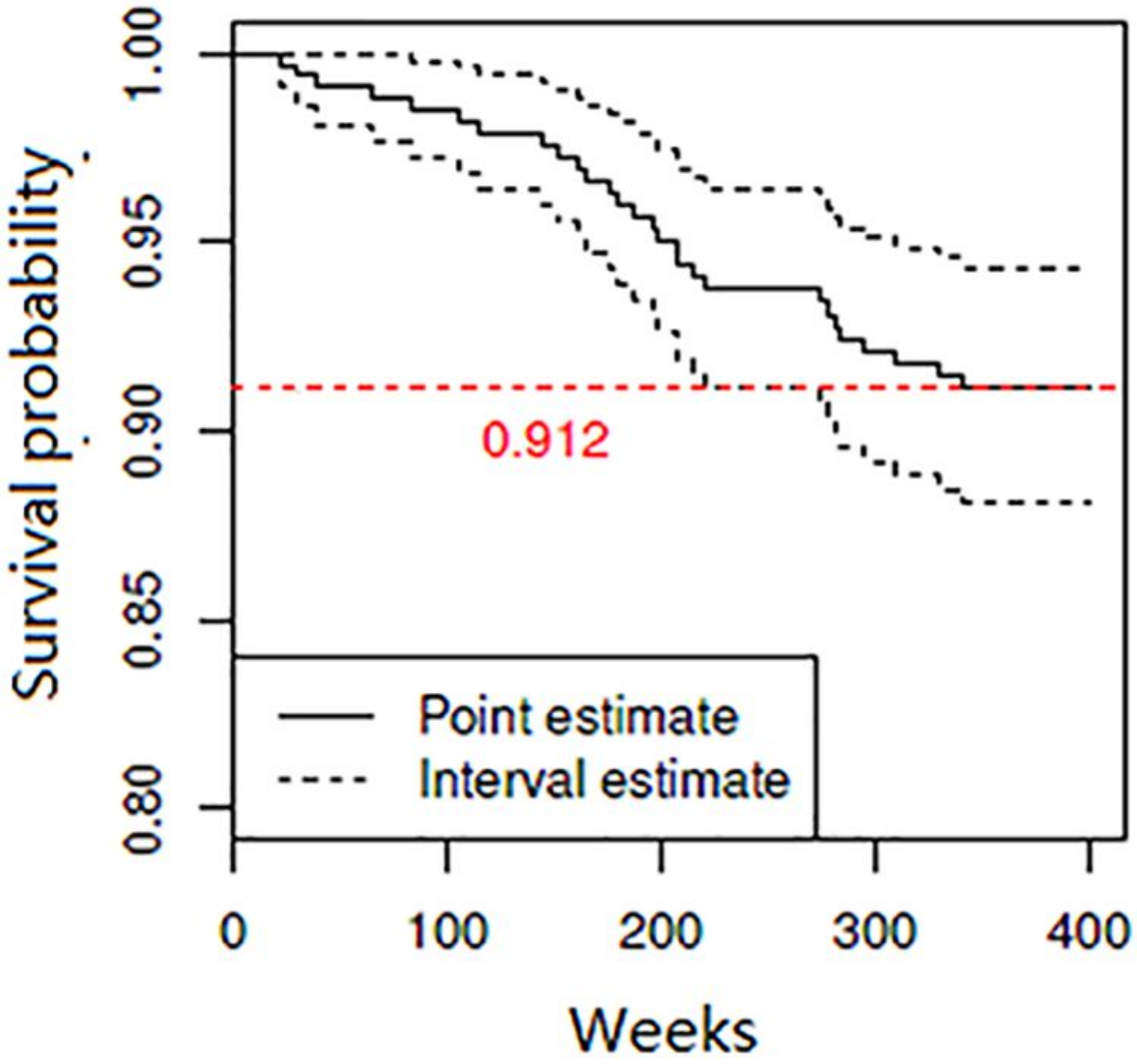

Figure 1

Kaplan-Meier survival probability curves.

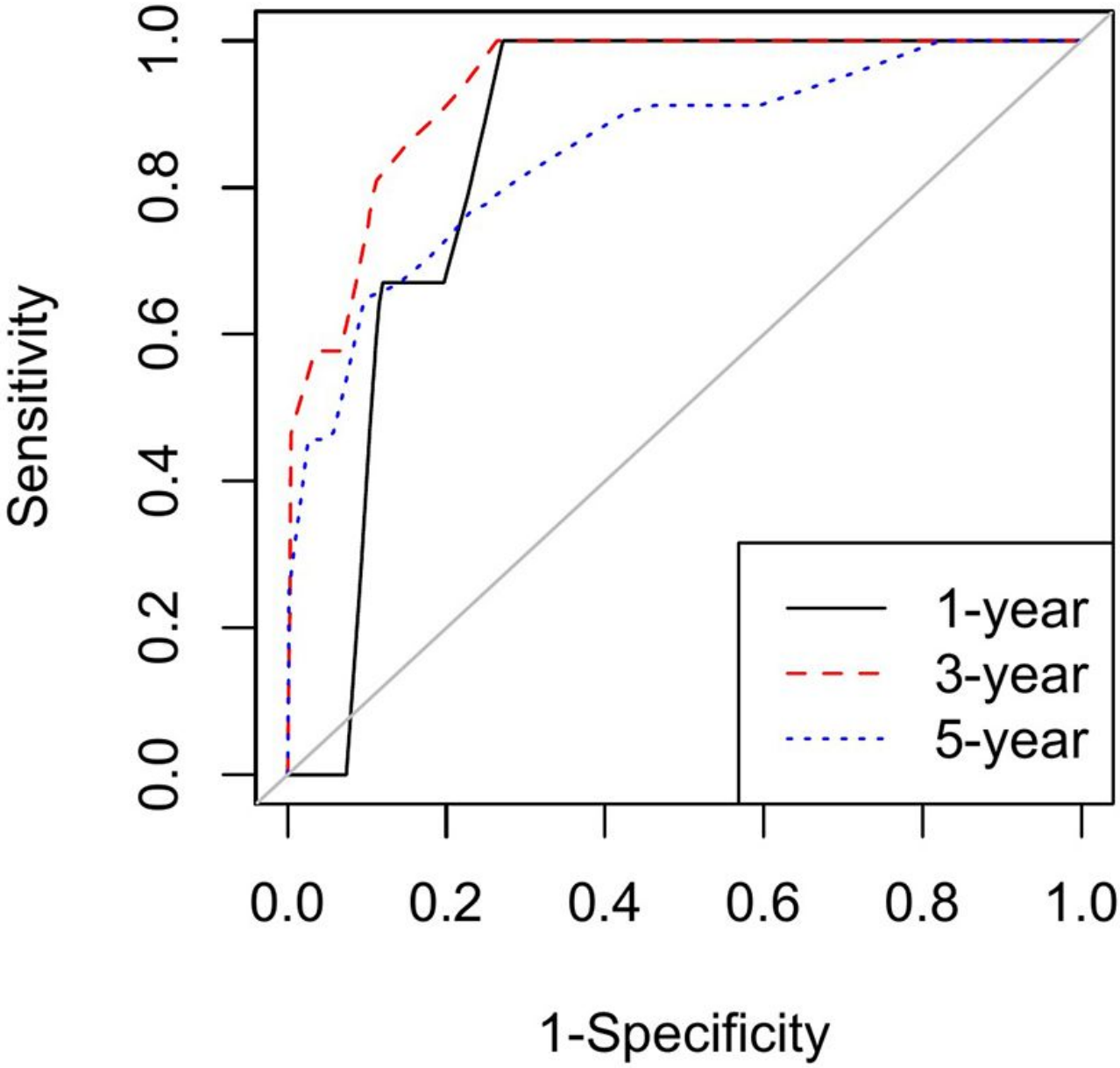

Figure 2

ROC curves for our extended Cox prognostic model at 1-year, 3-year and 5-year post-operation.