Am J Cancer Res 2023;13(11):5065-5081
www.ajcr.us /ISSN:2156-6976/ajcr0150743

Original Article

Development and validation of a prognostic nomogram for breast cancer patients who underwent chemoradiotherapy and surgery: a retrospective cohort study based on the SEER database and two Chinese cohorts

Huan Wang[1], Guang-Fa Xia[1], Zi-Ran Zhang[1], Xi Luo[2], Juan-Ying Zhu[1*], Hui-Ke Wang[2*]

[1]Maternity and Child Health Care Affiliated Hospital, Jiaxing University, Jiaxing 314000, Zhejiang, China; [2]Department of Oncology, The First Affiliated Hospital of Zhengzhou University, Zhengzhou 450052, Henan, China. [*]Equal contributors.

Received April 16, 2023; Accepted October 4, 2023; Epub November 15, 2023; Published November 30, 2023

Abstract: There is no strong evidence indicating the optimal treatment for breast cancer (BC) and no specific prognostic model. The aim of this study was to establish nomograms to predict the overall survival (OS) of BC patients receiving chemoradiotherapy and surgery, thereby quantifying survival benefits and improving patient management. A total of 1877 patients with primary nonmetastatic BC who received chemoradiotherapy and surgery from 2010 to 2019 were identified from the Surveillance, Epidemiology and End Results (SEER) database as the training cohort, 804 as the internal validation cohort, and 796 patients from the First Affiliated Hospital of Zhengzhou University (n=324) and Jiaxing Maternal and Child Health Hospital (n=472) as the external validation cohort. Least absolute shrinkage and selection operator (LASSO), univariate, and multivariate Cox regression analyses were performed in the training cohort to determine independent prognostic factors for BC, and a nomogram was constructed to predict 3-year, 5-year, and 8-year OS. The final model incorporated 7 factors that significantly affect OS: race, location, positive regional nodes, T stage, N stage, subtype, and grade. The calibration curves showed good consistency between the predicted survival and actual outcomes. Time-dependent receiver operating characteristic (ROC) curves and the time-dependent area under the curve (AUC) confirmed that the accuracy and clinical usefulness of the constructed nomograms were favorable. Decision curve analysis (DCA) and net reclassification improvement (NRI) also demonstrated that this nomogram was more suitable for clinical use than the 7th American Joint Committee on Cancer (AJCC) tumor node metastasis (TNM) staging system and the previous prediction model. In the training cohort and the internal validation cohort, the concordance indices (C-index) of the nomogram for predicting OS (0.723 and 0.649, respectively) were greater than those of the 7th AJCC TNM staging system and the previous prediction model. In addition, based on Kaplan-Meier (K-M) survival curves, the survival differences among different risk stratifications were statistically significant, indicating that our risk model was accurate. In this study, we determined independent prognostic factors for OS in patients with primary nonmetastatic BC treated with chemoradiotherapy and surgery. A new and accurate nomogram for predicting 3-, 5-, and 8-year OS in this patient population was developed and validated for potential clinical applicability.

Keywords: Breast cancer, chemoradiotherapy, nomogram, SEER program, risk classification system, overall survival

Introduction

Breast cancer (BC) is one of the most common malignancies in women, accounting for approximately 30% of new cancers occurring in women [1-3], and its mortality rate (15%) is the second highest among all malignancies in women [4]. It has been reported that since the middle of the last century, the incidence of BC has increased at a rate of 0.5% per year worldwide. In 2020, there were more than 2.3 million new cases of BC and 685,000 BC-related deaths worldwide,

and the number of new cases is expected to increase to more than 3 million by 2040, with the number of deaths reaching 1 million annually. In the United States, the 5-year survival rates for localized, regional, and distant metastatic BC are 99%, 85%, and 27%, respectively [5].

The management of BC is multidisciplinary [5]. Factors affecting the prognosis of BC patients include age, race, ethnicity, body weight, tumor and disease characteristics, response to treatment, etc. [5], and BC can also be classified into different subtypes according to hormone receptor positivity, human epidermal growth factor receptor type 2 (HER2) status, and tumor cell proliferation activity. TNM staging is performed according to tumor size, local aggressiveness, and lymph node and distant metastasis status [6]. Neoadjuvant and adjuvant therapies are recommended for each subtype and stage. For the vast majority of stage I-III patients with operable early BC, the main treatment mode is surgery, and the standard of combined adjuvant therapy is surgery + continuous chemotherapy (CT) + radiotherapy (RT) [7-9], but their optimal integration is still controversial [7, 9, 10]. At present, the efficacy of treatment for stage IV BC patients is still poor, but surgery, chemotherapy, radiotherapy, immunotherapy and targeted therapy have been widely used. Among them, the role of RT, commonly used in palliative care for BC patients, was highlighted by the National Cancer Database study. With the development of imaging methods, RT can be used as a definitive treatment for some stage 0-III patients who explicitly refuse surgery, providing more accurate nonsurgical treatment than previously possible [11, 12]. Preclinical studies have shown that RT has an antitumor effect by initiating and activating cytotoxic T cells, activating dendritic cells and type I interferon-dependent immune responses, and upregulating PD-L1 on bone marine-derived suppressor cells [13-17]. Therefore, choosing the best treatment for each patient, accounting for their wishes, may become more important in the future.

Interestingly, with early screening and improved treatment for BC, BC mortality has decreased by 42% over the past 30 years [5]. Therefore, the early identification of BC patients and accurate risk stratification of BC are of great importance in clinical practice and can help doctors adopt more active individualized treatments, especially for high-risk patients. Currently, there are many prognostic models based on traditional survival analysis methods, such as TNM staging, but they often ignore conflicting events. In addition, while independent prognostic factors associated with OS have been identified in BC, there is currently no universally accepted scoring system to predict long-term OS in this population. Nomograms are a simple and multivariable visualization tool for quantifying and predicting risk and prognosis in the field of oncology [18-22], and these models are based on key variables such as traditional clinicopathological features to more accurately estimate individual survival to aid clinical decision-making and promote the development of precision medicine. Nomograms for predicting the prognosis of BC patients have been developed and verified [23, 24].

For these reasons, we used the SEER database, which contains much multicenter and high-quality study data, and retrospective data from two Chinese hospitals to identify independent prognostic factors associated with OS in this subpopulation and developed a new nomogram for 3-, 5-, and 8-year OS prediction. The nomogram was compared with the traditional 7th AJCC TNM stage and the previous prediction model to provide some reference for identifying high-risk patients and making individualized treatment decisions in clinical practice.

Material and methods

Data source and patient selection

First, we extracted and screened the clinical data of BC patients from the SEER public database from 2010 to 2019, gradually identified patients who met the inclusion criteria, and finally included 2681 BC patients in the study, including 1877 patients as the training cohort and 804 patients as the internal validation cohort. In addition, a total of 796 BC patients treated at the First Affiliated Hospital of Zhengzhou University and Jiaxing Maternal and Child Health Hospital from January 2009 to August 2022 were retrospectively analyzed for external verification.

In both the training cohort and the validation cohort, the inclusion criteria were as follows: (1)

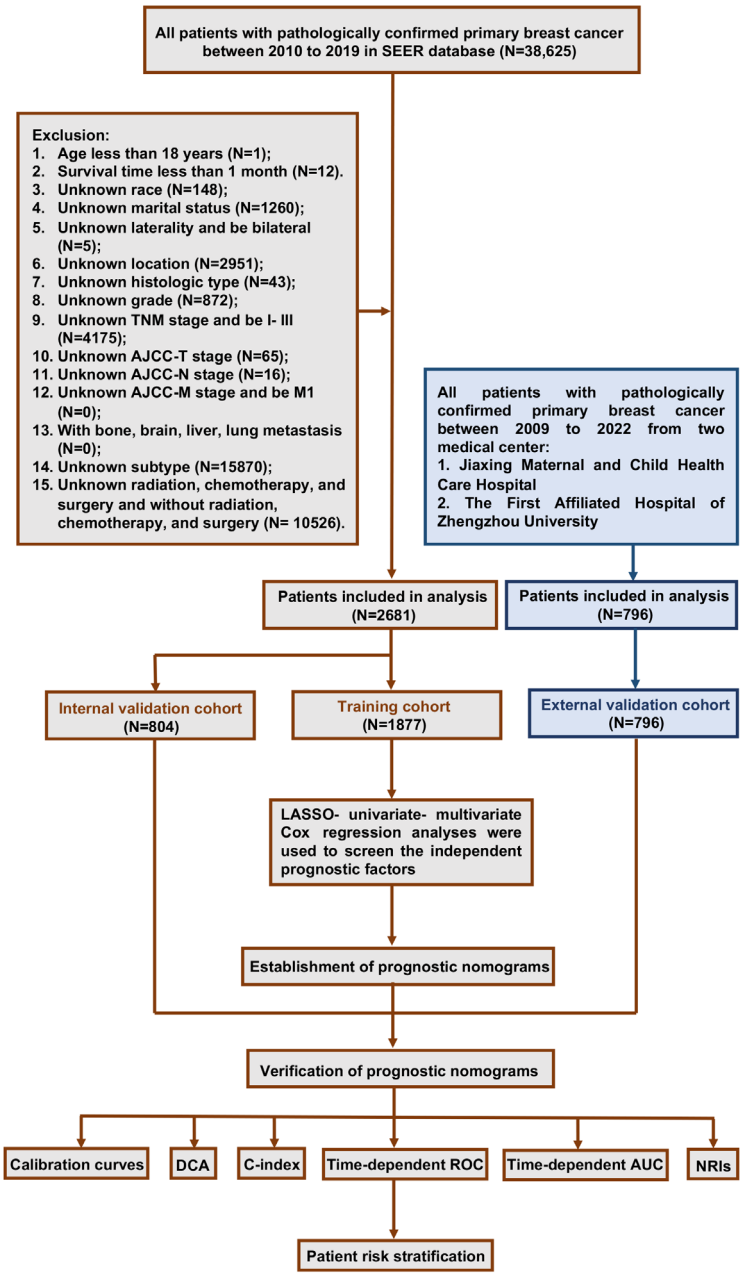

Figure 1. Flow chart of the design. SEER, Surveillance, Epidemiology and End Results; AJCC, American Joint Committee on Cancer; LASSO, least absolute shrinkage and selection operator; DCA, decision curve analysis; ROC, receiver operating characteristic; AUC, area under the curve; C-index, concordance index; NRI, net reclassification improvement; TNM, tumor node metastasis.

patients with BC confirmed by histopathological examination according to the malignant behavior coded by ICD-O-3/WHO 2008; (2) patients with BC as the first tumor without multiple primary malignancies and without metastases of other sites; (3) patients with TNM stage I-III who received chemoradiotherapy and surgery; (4) patients with complete follow-up data on OS and a survival time over 1 month; and (5) patients with complete case information. The detailed flow chart of the study population selection process is shown in **Figure 1**. The study was carried out in accordance with the Declaration of Helsinki and approved by the Ethics Committee of Jiaxing Maternal and Child Health Hospital.

Endpoints and study variables

The following variables were selected as potential prognostic factors at the time of diagnosis: age, sex, race, marital status, laterality, location, regional nodes examined, positive regional nodes, histologic type, T stage, N stage, TNM stage, grade, subtype, and tumor size.

We used the 7th TNM staging system to define T, N, M, and TNM staging in patients with clinical stages. Tumor locations were divided into the upper-outer quadrant, up-per-inner quadrant, lower-outer quadrant, lower-inner quadrant and others. Histologic type was classified into infiltrating ductal carcinoma (IDC) and others. The histologic types were divided into well differentiated (Grade I), moderately differentiated (Grade II), poorly differentiated (Grade III) and undifferentiated (Grade IV). Subtypes were divided into HR-/HER2- (triple-negative), HR-/HER2+ (Her2-enriched), HR+/HER2- (luminal A) and HR+/HER2+ (luminal B). Tumor size was assessed according to maximum tumor diameter and was classified as less than 2 cm or greater than 2 cm. The outcome of the study was OS, defined as the time between the date of diagnosis and the date of death from any

cause or the last follow-up. In the training cohort, the median OS time was 73.0 months (range 5.0~119.0 months); the median OS time in the internal validation cohort was 74.0 months (range 6.0~119.0 months); and the median OS time in the external validation cohort was 86.2 months (range 9.7~398.8 months). All patients in the validation cohort were followed up by telephone or in-hospital visits. For the validation cohort, the follow-up period ended on February 17, 2023.

Statistical analysis

The clinical characteristics and prognostic data of patients in the training cohort were extracted from large-scale SEER cancer registry data by SEER*Stat software (version 8.3.5). SPSS 26.0 software and RStudio 4.2.2 were used for statistical analyses and to graph the data. To compare variables between the training and validation cohorts and compare categorical variables, Pearson Chi-square tests were used. $P<0.05$ was considered statistically significant, and all tests were two-sided.

The training cohort was used to develop the nomogram, while the validation cohort was used to verify the model efficacy. LASSO, univariate, and multivariate Cox regression analyses were performed in the training cohort to identify independent prognostic factors associated with OS in BC patients, and hazard ratios (HRs) and 95% confidence intervals (CIs) were calculated. The variables selected by LASSO and those with $P<0.10$ in univariate Cox analysis were included in the multivariate analysis for mixed, forward, backward and stepwise regression analyses. The variable combination with the smallest Akaike information criterion (AIC) value was selected by analysis of variance (ANOVA), and the variables with $P<0.05$ were eligible to be used to generate the nomograms. Each variable was converted to a corresponding 0-100 scale, and then the points assigned for each variable were summed to obtain the total score, which was finally converted to predict the OS of BC patients at 3, 5, and 8 years. Finally, according to the median risk score of the model, patients in the training cohort and the validation cohort were assigned to a high-risk group or a low-risk group. K-M curves and log-rank tests were used for survival analyses to explore the differences in survival rates among BC patients with different risk classifications. Time-dependent ROC curves, time-dependent AUCs and calibration curves (1000 samples, 45-degree line as the best model) were used to evaluate the prediction accuracy and discrimination of the nomogram. DCA was used to test the clinical benefit and application value of the nomogram. We also calculated the C-index and NRI and compared the nomogram with the 7th AJCC TNM staging system and 5 previous prediction models to further demonstrate the superiority of our model in terms of clinical benefit and practicability. If the C-index and AUC values were between 0.5 and 0.6, between 0.6 and 0.7, or greater than 0.8, the model's predictive performance was considered poor, average, or good, respectively. An NRI >0 indicated a positive improvement, meaning that the prediction ability of the new model was improved compared with that of the old model. An NRI <0 indicated a negative change, indicating a decline in the predictive power of the new model.

Results

Patient characteristics

Table 1 shows the demographic and clinicopathological characteristics of the training cohort (n=1877), internal validation cohort (n=804) and external validation cohort (n=796). Interestingly, except for regional node positivity, there was no significant difference between patients in the training cohort and the internal validation cohort, indicating a uniform distribution between them ($P\geq0.05$). Moreover, there were no significant differences in the distribution of sex or laterality between the training cohort and the external validation cohort. Although the remaining variables were different between these two cohorts ($P<0.05$), overall, half of the patients had tumors in the upper quadrant, and more than 60% of the patients were married, had more than 3 regional nodes examined, had a tumor size greater than 2 cm and had a positive ER status. The histologic type of BC in more than 80% of patients was IDC, more than 80% of patients had stage T1+T2 or N0+N1 disease, and the grade was concentrated in stages II and III. However, it is worth noting that positive regional lymph nodes were detected in 88.1% of patients in the external validation cohort and less than 54.6% in

Table 1. Demographics and clinicopathologic characteristics of the cohorts with BC

Variables	Training cohort N=1877 (%)	Internal validation cohort N=804 (%)	External validation cohort N=796 (%)	Total population N=3477 (%)	*P**-value	*P*#-value
Age (y)					0.558	<0.001
<55	1013 (54.0)	424 (53.6)	515 (64.7)	1952 (56.1)		
≥55	864 (46.0)	380 (47.3)	281 (35.3)	1525 (43.9)		
Sex					0.179	0.319
Female	1867 (99.5)	796 (99.0)	794 (99.7)	3457 (99.4)		
Male	10 (0.5)	8 (1.0)	2 (0.3)	20 (0.6)		
Race					0.810	<0.001
Black	881 (46.9)	382 (47.5)	0 (0.0)	1263 (36.3)		
White	918 (48.9)	385 (47.9)	0 (0.0)	1303 (37.5)		
Others	78 (4.2)	37 (4.6)	796 (100.0)	911 (26.2)		
Marital status					0.353	<0.001
Married	1059 (56.4)	438 (54.5)	691 (86.8)	2188 (62.9)		
Unmarried	818 (43.6)	366 (45.5)	105 (13.2)	1289 (37.1)		
Laterality					0.226	0.182
Left	935 (49.8)	421 (52.4)	419 (52.6)	1775 (51.0)		
Right	942 (50.2)	383 (47.6)	377 (47.4)	1702 (49.0)		
Location					0.906	0.023
Upper-outer quadrant	746 (39.7)	332 (41.3)	319 (40.1)	1397 (40.2)		
Upper-inner quadrant	255 (13.6)	108 (13.4)	102 (12.8)	465 (13.4)		
Lower-outer quadrant	172 (9.2)	75 (9.3)	56 (7.0)	303 (8.7)		
Lower-inner quadrant	112 (6.0)	42 (5.2)	31 (3.9)	185 (5.3)		
Others	592 (31.5)	247 (30.7)	288 (36.2)	1127 (32.4)		
Regional nodes examined					0.929	<0.001
<3	477 (25.4)	203 (25.2)	331 (41.6)	1011 (29.1)		
≥3	1400 (74.6)	601 (74.8)	465 (58.4)	2466 (70.9)		
Regional nodes positive					0.048	<0.001
No	828 (44.1)	388 (48.3)	701 (88.1)	1917 (55.1)		
Yes	1049 (55.9)	416 (51.7)	95 (11.9)	1560 (44.9)		
Histologic Type					0.347	<0.001
Infiltrating duct carcinoma	1543 (82.2)	673 (83.7)	602 (80.2)	2818 (81.0)		
Others	334 (17.8)	131 (16.3)	194 (24.4)	659 (19.0)		
T stage					0.228	<0.001
T1	750 (40.0)	318 (39.6)	509 (63.9)	1577 (45.4)		
T2	817 (43.5)	365 (45.4)	265 (33.3)	1447 (41.6)		
T3	227 (12.1)	78 (9.7)	18 (2.3)	323 (9.3)		
T4	83 (4.4)	43 (5.3)	4 (0.5)	130 (3.7)		
N stage					0.051	<0.001
N0	786 (41.9)	371 (46.1)	582 (73.1)	1739 (50.0)		
N1	765 (40.8)	302 (37.6)	140 (17.6)	1207 (34.7)		
N2	227 (12.1)	79 (9.8)	53 (6.7)	359 (10.3)		
N3	99 (5.3)	52 (6.5)	21 (2.6)	172 (4.9)		
TNM stage					0.669	<0.001
I	478 (25.4)	206 (25.6)	418 (52.5)	1102 (31.7)		
II	921 (49.1)	406 (50.5)	297 (37.3)	1624 (46.7)		
III	478 (25.5)	192 (23.9)	81 (10.2)	751 (21.6)		
Grade					0.611	<0.001
Grade I	132 (7.0)	54 (6.7)	21 (2.6)	207 (6.0)		
Grade II	681 (36.3)	292 (36.3)	479 (60.2)	1452 (41.8)		
Grade III	1060 (56.5)	458 (57.0)	296 (37.2)	1814 (52.2)		
Grade IV	4 (0.2)	0 (0.0)	0 (0.0)	4 (0.1)		
ER status					0.149	<0.001
Negative	559 (29.8)	262 (32.6)	462 (58.0)	1283 (36.9)		
Positive	1318 (70.2)	542 (67.4)	334 (42.0)	2194 (63.1)		

Subtype					0.337	<0.001
HR-/HER2- (Triple-Negative)	388 (20.7)	175 (21.8)	292 (36.7)	855 (24.6)		
HR-/HER2+ (HER2-enriched)	129 (6.9)	69 (8.6)	177 (22.2)	375 (10.8)		
HR+/HER2- (Luminal A)	994 (53.0)	415 (51.6)	41 (5.2)	1450 (41.7)		
HR+/HER2+ (Luminal B)	366 (19.5)	145 (18.0)	286 (35.9)	797 (22.9)		
Tumor size (cm)					0.235	<0.001
<2	680 (36.2)	272 (33.8)	398 (50.0)	1350 (38.8)		
≥2	1197 (63.8)	532 (66.2)	398 (50.0)	2127 (61.2)		

Data are shown as N (%). Grade I, well differentiated; Grade II, moderately differentiated; Grade III, poorly differentiated; Grade IV, undifferentiated; anaplastic. Abbreviations: BC, breast cancer; TNM, tumor node metastasis; HER2, human epidermal growth factor receptor type 2. P^{*} refers to *P*-value in the comparison between training cohort and internal validation cohort. $P^{\#}$ refers to *P*-value in the comparison between training cohort and external validation cohort.

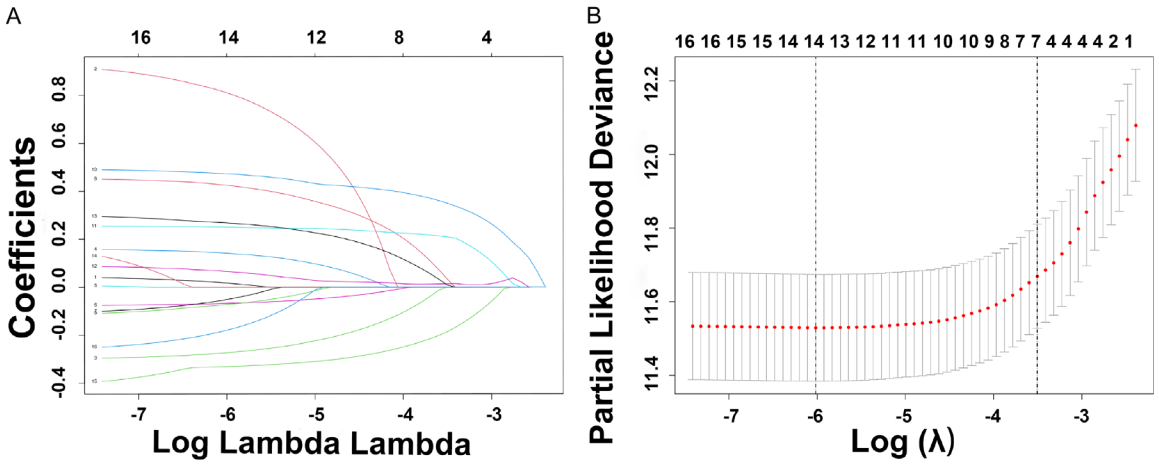

Figure 2. Selection of the prognostic factors using LASSO Cox regression. A. For clinicopathological features, LASSO coefficient profiles are plotted vs. log λ sequences. The dotted vertical line shows the nonzero coefficients, where 7 nonzero coefficients are included. B. Tuning parameter λ based on minimum criteria in the LASSO regression. The partial likelihood binomial deviance is plotted against log λ. By using the minimum standard error of the minimum criteria, dotted vertical lines are set at the optimal values log λ, where factors are selected. LASSO, least absolute shrinkage and selection operator.

the SEER cohort. Other significant differences were race, where the SEER cohort was overwhelmingly white, and all patients in the external validation cohort were Asian (100%). In addition, 52.6% of patients in the SEER cohort had luminal A cancer, while the main subtypes in the external validation cohort were triple-negative and luminal B.

Screening for predictive factors

Based on the included variables, LASSO analysis was first used (**Figure 2**) to identify 14 indicators related to OS, which were input into the univariate Cox regression model. In addition, multivariate Cox regression analysis was performed using variables with $P<0.10$ in univariate Cox regression analysis, and the influence of confounding variables was eliminated to further determine independent prognostic factors for BC patients (**Table 2**). The independent risk factors included grade ($P=0.005$), race ($P=0.004$), positive regional nodes ($P=0.003$), location ($P=0.024$), AJCC-T stage ($P<0.001$), AJCC-N stage ($P=0.015$), and subtype ($P<0.001$). In addition, K-M curves showed that the above indices were significantly correlated with OS ($P<0.05$) (**Figure 3**). The independent predictors above, including location, achieved AUC ranges of greater than 0.5 for predicting 10-year OS indicating reasonable estimates (**Figure 4**).

Development and Validation of the nomograms

Here, we set up two nomograms. **Figure 5A** was generated according to the 7th AJCC TNM staging system and contains the T, N, and M stages. **Figure 5B** shows a complex model containing 7

Table 2. Univariate and multivariate Cox analyses on variables for the prediction of OS of BC patients

Variables	Univariate analysis			Multivariate analysis		
	HR	95% CI	*P* value	HR	95% CI	*P* value
Age	1.008	0.804~1.264	0.943	~	~	~
Grade	1.488	1.222~1.812	<0.001	1.355	1.097~1.672	0.005
Sex	2.237	0.833~5.999	0.110	~	~	~
Race	0.692	0.563~0.850	<0.001	0.740	0.603~0.907	0.004
Marital status	1.312	1.047~1.643	0.018	1.184	0.942~1.489	0.148
Regional nodes examined	1.399	1.058~1.849	0.018	0.896	0.663~1.212	0.477
Regional nodes positive	2.183	1.701~2.803	<0.001	1.62	1.174~2.236	0.003
Histologic	0.853	0.625~1.164	0.317	~	~	~
Location	0.941	0.881~1.006	0.074	0.926	0.866~0.990	0.024
T stage	1.838	1.631~2.070	<0.001	1.644	1.371~1.971	<0.001
N stage	1.627	1.449~1.828	<0.001	1.282	1.048~1.567	0.015
TNM stage	2.188	1.848~2.586	<0.001	1.091	0.780~1.527	0.609
Subtype	0.712	0.641~0.791	<0.001	0.709	0.632~0.795	<0.001
Tumor size (cm)	1.927	1.485~2.503	<0.001	0.761	0.556~1.040	0.087

Abbreviations: BC, breast cancer; OS, overall survival; AJCC, American Joint Committee on Cancer; TNM, tumor node metastasis.

statistically significant variables based on multivariate analysis of the training cohort. Both nomograms were used to predict OS probabilities at 3, 5, and 8 years. Simply put, each factor and subtype were assigned a score on the scoring table, and by adding the scores together and locating the corresponding position on the bottom scale, we calculated the probability of 3-, 5-, and 8-year OS. Next, the nomogram was verified for predictive accuracy in the training and verification cohorts. We tested 1000 bootstrap samples for the training cohort, internal validation cohort and external validation cohort, and the calibration graphs showed that the nomogram performed well in these cohorts (**Figure 6**).

Comparison of different models

To further compare the clinical application prospects and predictive performance of the nomogram with that of the 7th AJCC TNM staging system, we plotted DCA curves (**Figure 7**) and ROC curves (**Figure 8**). In the training cohort, the DCA curve showed that our model had a significantly better performance. In addition, in these cohorts, our model showed higher AUCs for all ROC curves, indicating superior prognostic accuracy over the 7th AJCC TNM staging system. In the training cohort, internal validation cohort and external validation cohort, the AUCs of the nomogram in predicting 3-, 5- and 8-year OS were 0.732, 0.761, and 0.715; 0.683, 0.644, and 0.676; and 0.869, 0.845, and 0.742 respectively. Similarly, these values were 0.640, 0.675, and 0.651; 0.625, 0.617, and 0.657; and 0.947, 0.890, and 0.807, respectively, for the 7th AJCC TNM staging system (**Figure 8**).

Based on the training cohort, the C-index of the nomogram was 0.723, the 95% CI was 0.694~0.752, the C-index of the 7th AJCC TNM staging system was 0.652, and the 95% CI was 0.623~0.681. In the internal validation cohort and external validation cohort, the C-indices of the above 2 models were 0.649 and 0.623 and 0.757 and 0.806, respectively (**Table 3**). In addition, whether in the training cohort, internal validation cohort or external validation cohort, compared with the nomogram in this paper, the NRI of the 7th AJCC TNM staging system and the previous prediction model were all less than or equal to 0 (**Table 4**). Overall, compared with the 7th AJCC TNM staging system and the previous prediction model, the new model we constructed had a better overall prediction performance (**Figure 9**).

Nomogram performance in risk group stratification

To facilitate personalized management, it is necessary to classify patients according to

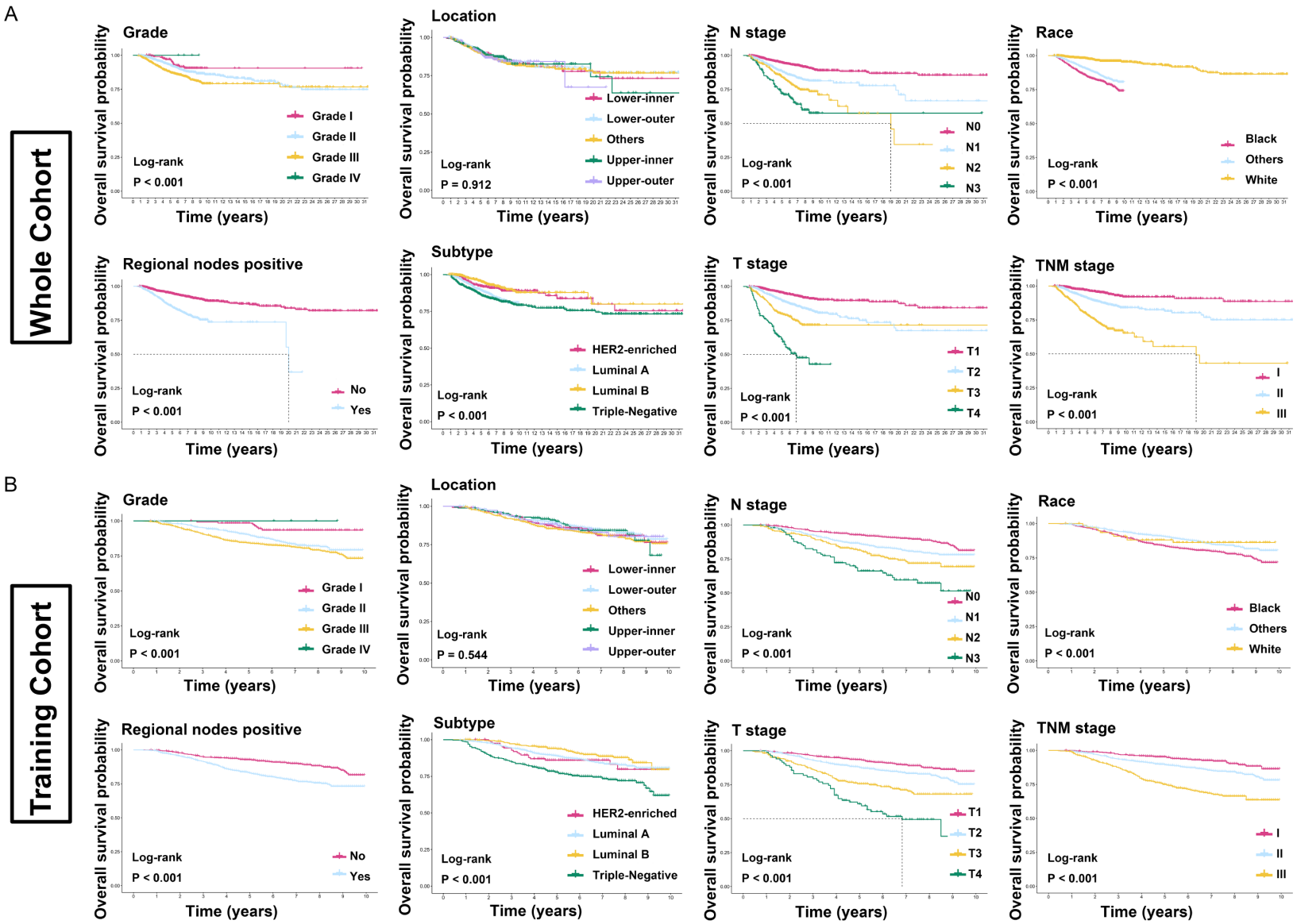

Figure 3. K-M curves for risk stratification (A) in the whole cohort and (B) in the training cohort. K-M, Kaplan-Meier.

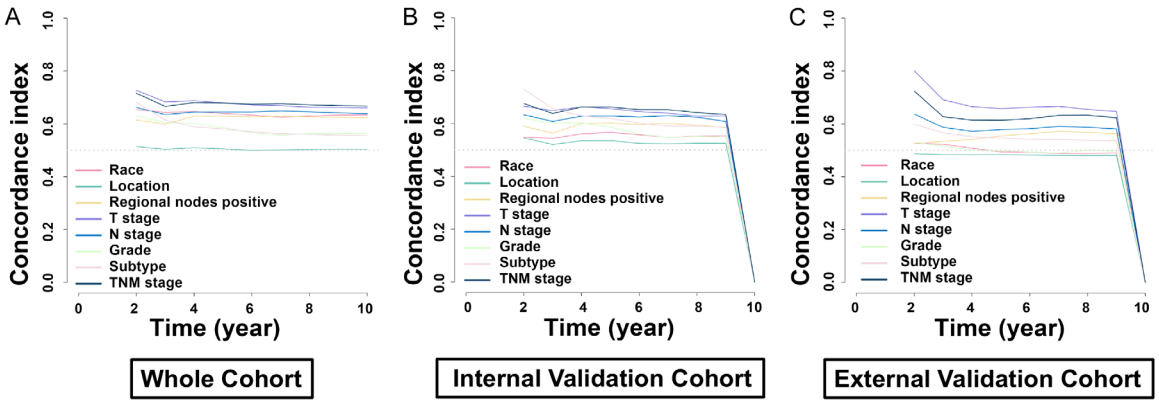

Figure 4. The time-dependent AUC curves of grade, location, T stage, race, regional nodes positive, subtype, N stage, and TNM stage for OS (A) in the whole cohort, (B) in the internal validation cohort, and (C) in the external validation cohort. AUC, area under the curve; OS, overall survival.

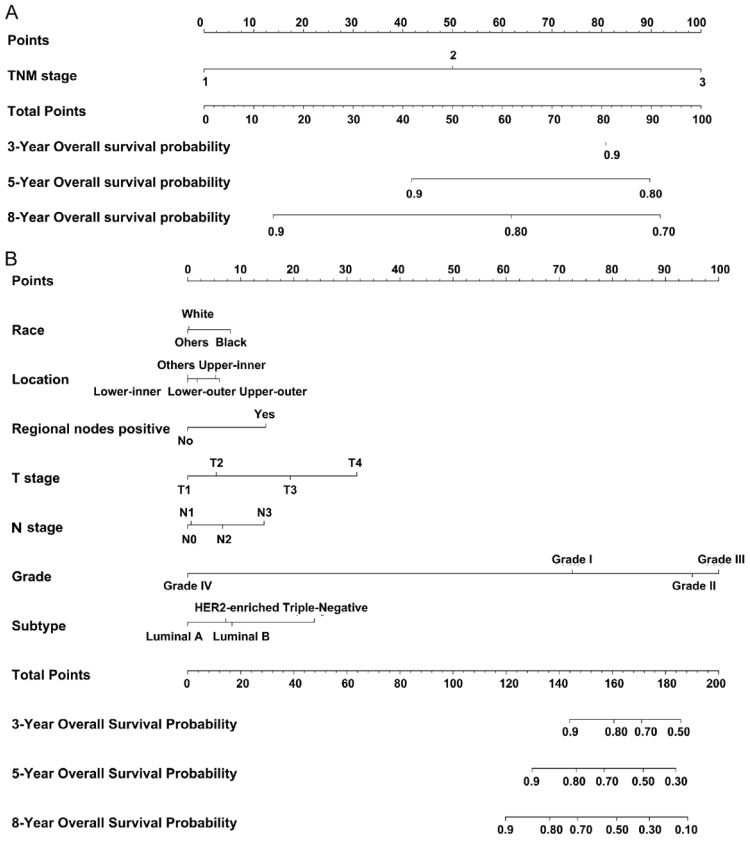

Figure 5. Two nomograms used to predict 3-, 5- and 8-year OS rates of BC. A. Model 1 nomogram was established according to tumor AJCC-TNM stage. B. Model 2 nomogram was established according to race, location, regional nodes positive, subtype, T stage, N stage, and grade. For each variable, the value of an individual is placed on the axis, and a line is drawn upward to determine how many points exist for each variable. The survival axis is drawn below the total points axis, which is then used to determine the 3-, 5-, and 8-year OS rates. AJCC, American Joint Committee on Cancer; BC, breast cancer; OS, overall survival.

their mortality risk. Each patient in the cohort was given a risk score based on the nomogram model, and the median risk score was set as the threshold value. We then evaluated the performance of our risk model by plotting the K-M survival curves of OS for the training and validation cohorts (**Figure 10**). The significant difference in survival seen between the high-risk and low-risk groups indicated that our risk model was accurate (P<0.001).

Discussion

We developed and constructed a nomogram with clinical value based on data from the SEER database and included all available factors affecting the prognosis of BC. Finally, race, location, regional node positivity, subtype, AJCC-T stage, AJCC-N stage, and grade were identified as independent prognostic factors for OS. Notably, we used data from the Chinese population for external validation, and the calibration curve, C-index, time-dependent ROC curve, time-dependent AUC, DCA, IDI, NRI and other aspects showed

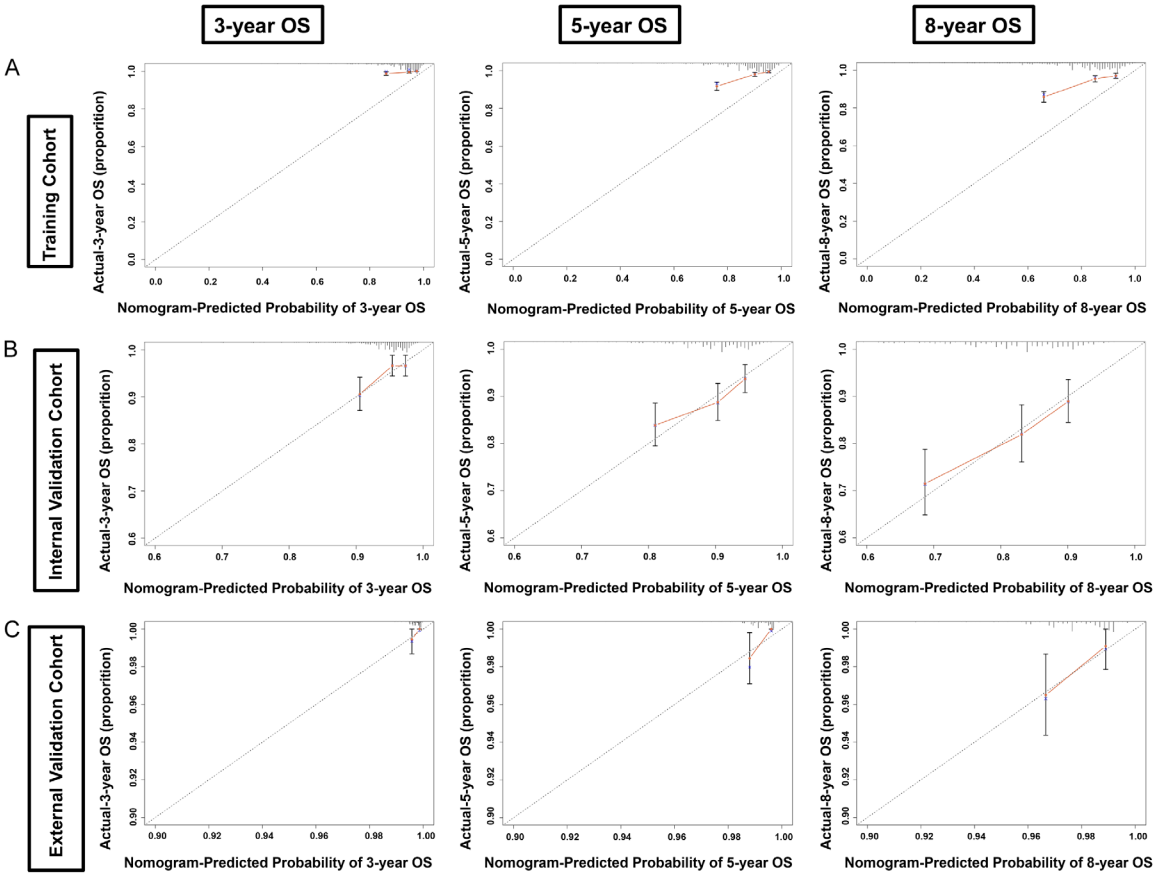

Figure 6. Assessment of nomogram used to predict 3-, 5-, and 8-year OS rates in the patients with BC in the (A) training cohort, (B) internal validation cohort, and (C) external validation cohort. The X-axis represents the model-predicted survival, and the Y-axis represents actual survival. The bar represents 95% CI as measured by K-M analysis, and the dotted line represents the ideal reference line. BC, breast cancer; OS, overall survival; K-M, Kaplan-Meier; CI, confidence interval.

that our model performed well in the validation cohort. It was proven that our nomogram has better prediction and discrimination ability than the 7th AJCC TNM staging system and the previous prediction model. The BC patients were categorized in high- and low-risk subgroups to provide comprehensive guidance for clinical practice.

The AJCC-TNM staging system is the most classic risk stratification and treatment strategy selection system for cancer patients [25, 26] and is commonly used to evaluate the prognosis of BC patients and determine treatment strategies [27, 28]. Our study showed that the HRs for T stage, N stage and TNM stage, and AJCC 7th edition in multivariate analysis were 1.644 (1.371~1.971) (P<0.001), 1.282 (1.048~1.567) (P=0.015), and 1.091 (0.780~1.527) (P=0.609), respectively. The T, N, and M stages of tumors can affect the prognosis of BC. With progression in TNM stage, the prognosis of patients deteriorates significantly, and higher stages are negatively correlated with the survival time of patients [29]. However, this most widely used system ignores many factors that have been shown to be highly correlated with OS and does not accurately identify survival differences between cancer subtypes, which was one of the motivations for this study.

In terms of social demographic data, marital status, age and race were found to be independent risk factors for BC patients. Studies have shown that marriage can reduce the risk of death by 25%. Unmarried patients often lack care and support from spouses, leading to increased psychological stress, and are more likely to suffer from chronic psychological distress, bad living habits, endocrine system dis-

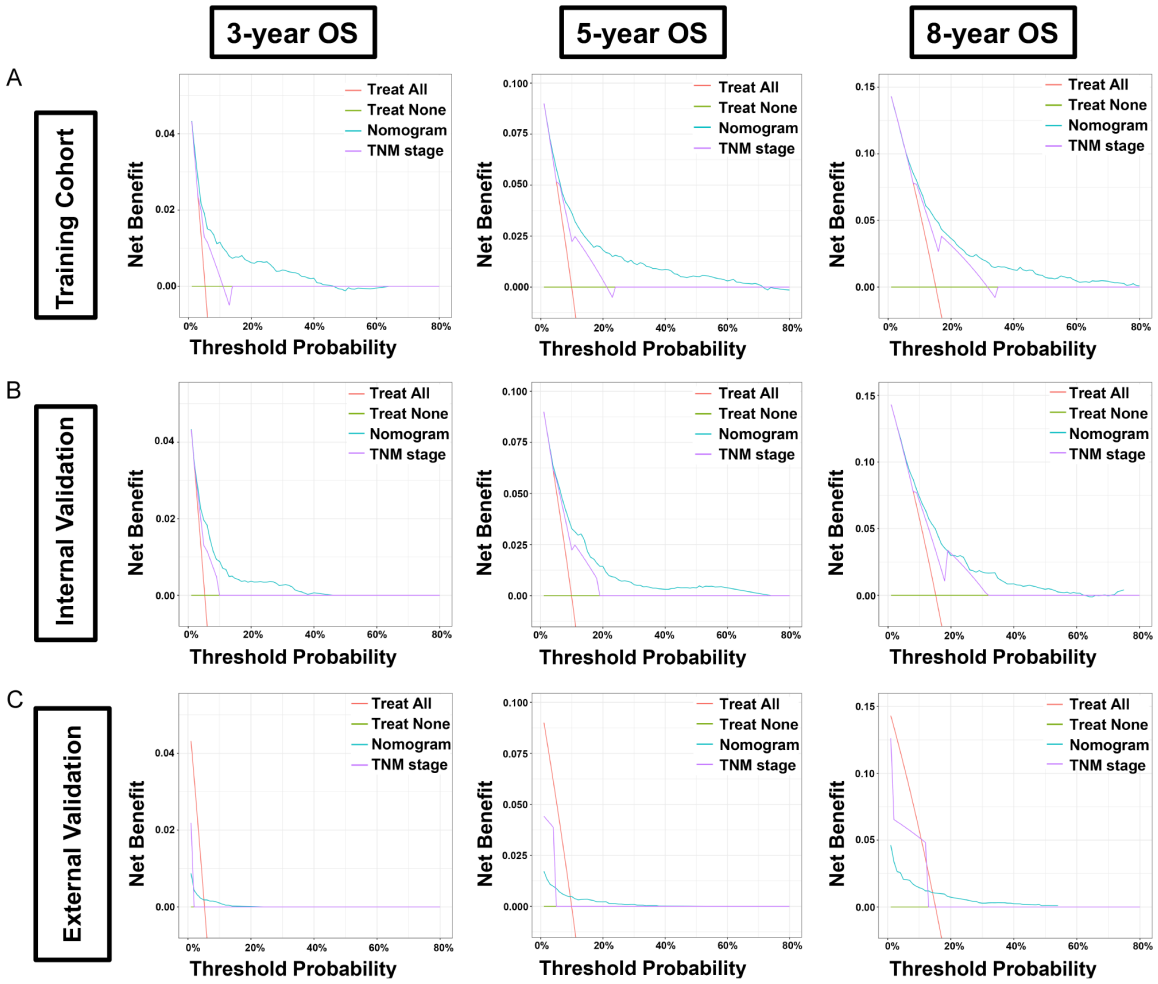

Figure 7. The DCA curves were plotted based on 3-, 5-, and 8-year OS benefits in the (A) training cohort, (B) internal validation cohort, and (C) external validation cohort. DCA, decision curve analysis; OS, overall survival.

orders and declines in immune system function [30, 31], ultimately accelerating tumor growth and patient death [32]. Age is generally regarded as an important reference factor affecting the occurrence and development of tumors, their biological characteristics and the prognosis of BC. Studies have found that the incidence of BC increases with age, doubling roughly every 10 years until the increase in BC begins to slow during menopause [33]. In terms of prognosis, it is generally accepted that older patients have higher mortality, which may be related to the fact that they tend to have more underlying conditions, such as hypertension, hyperlipidemia, heart disease, cerebrovascular disease, and diabetes [34], which can lead to poor tolerance and compliance with treatment. However, in terms of molecular pathology, young BC patients (<35 years of age) were found to have worse pathology and classification (with larger tumors, positive lymph nodes, higher histological grade, nonluminal disease, and higher Ki-67 expression), worse prognosis, and higher rates of metastasis and recurrence [35]. In addition, our study suggested that black women have a worse prognosis than white women, which may be related to black patients often having poorer economic conditions and living environments and less access to medical resources and surgical treatment [1, 36-40].

BC patients were grouped as having luminal A, luminal B, HER2-enriched and triple-negative subtypes of cancer based on three key therapeutic targets, ER, PR and HER2. Consistent with epidemiological studies, our study found significant differences in the prognoses of BC

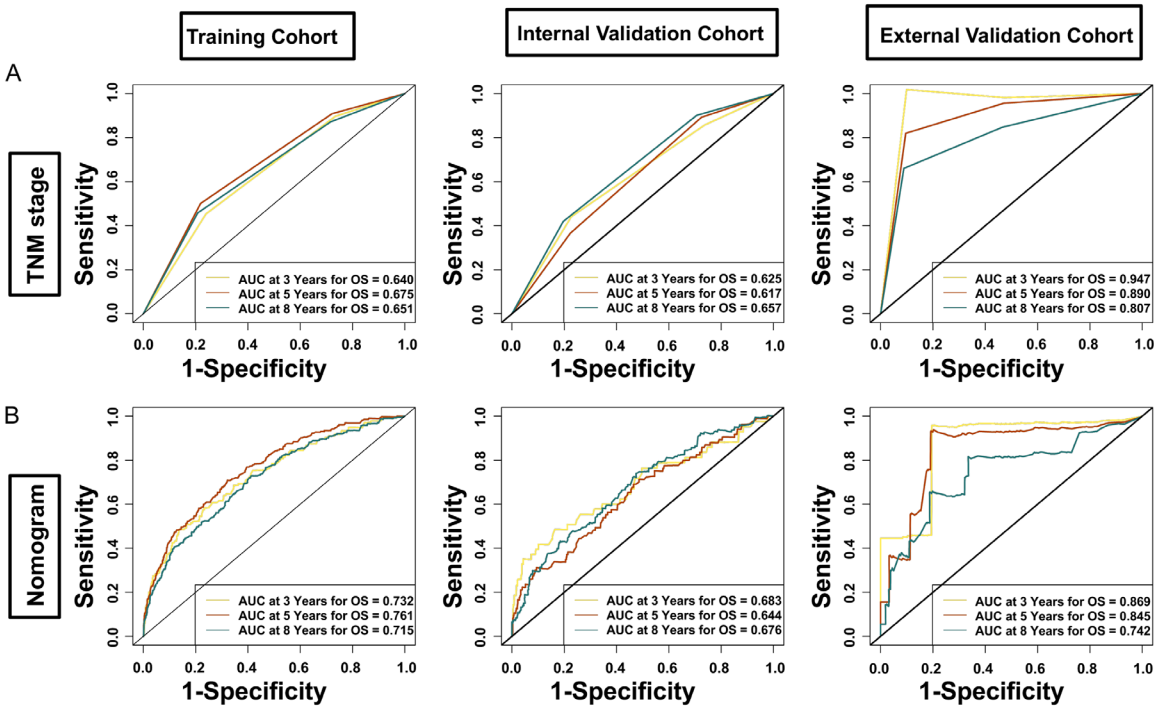

Figure 8. ROC curves of (A) TNM stage and (B) nomogram in the training cohort, internal validation cohort, and external validation cohort for predicting 3-, 5-, and 8-year OS rates in the patients with BC. OS, overall survival; BC, breast cancer; ROC, receiver operating characteristic.

Table 3. The C-Index the nomograms for predicting BC patients' OS in the training cohort, internal validation cohort and external validation cohort

Cohort	TNM stage		Nomogram	
	C-index	95% CI	C-index	95% CI
Training cohort	0.652	0.623~0.681	0.723	0.694~0.752
Internal validation cohort	0.623	0.678~0.668	0.649	0.594~0.704
External validation cohort	0.806	0.722~0.890	0.757	0.655~0.859

Abbreviations: BC, breast cancer; OS, overall survival; AJCC, American Joint Committee on Cancer; TNM, tumor node metastasis; C-index, concordance index.

Table 4. The predictive performance (NRI) of different models for predicting BC patients' OS in the training cohort, internal validation cohort and external validation cohort

Cohort	TNM stage			A previous prediction model [44]		
	Outcome	NRI	95% CI	Outcome	NRI	95% CI
Training cohort	3-year OS	-0.211	-0.296~-0.111	3-year OS	-0.048	-0.149~-0.283
	5-year OS	-0.201	-0.313~-0.096	5-year OS	-0.173	-0.286~-0.072
	8-year OS	-0.210	-0.306~-0.144	8-year OS	-0.170	-0.299~-0.055
Internal validation cohort	3-year OS	-0.216	-0.445~0.001	3-year OS	0.020	-0.262~0.082
	5-year OS	-0.194	-0.326~-0.034	5-year OS	0.005	-0.171~0.083
	8-year OS	-0.182	-0.375~-0.004	8-year OS	-0.024	-0.226~0.085
External validation cohort	3-year OS	0.000	0.000~0.003	3-year OS	0.000	-0.004~0.003
	5-year OS	0.000	-0.486~0.008	5-year OS	0.000	-0.407~0.008
	8-year OS	-0.097	-0.601~0.010	8-year OS	-0.097	-0.601~0.010

Abbreviations: BC, breast cancer; OS, overall survival; AJCC, American Joint Committee on Cancer; TNM, tumor node metastasis; NRI, net reclassification improvement.

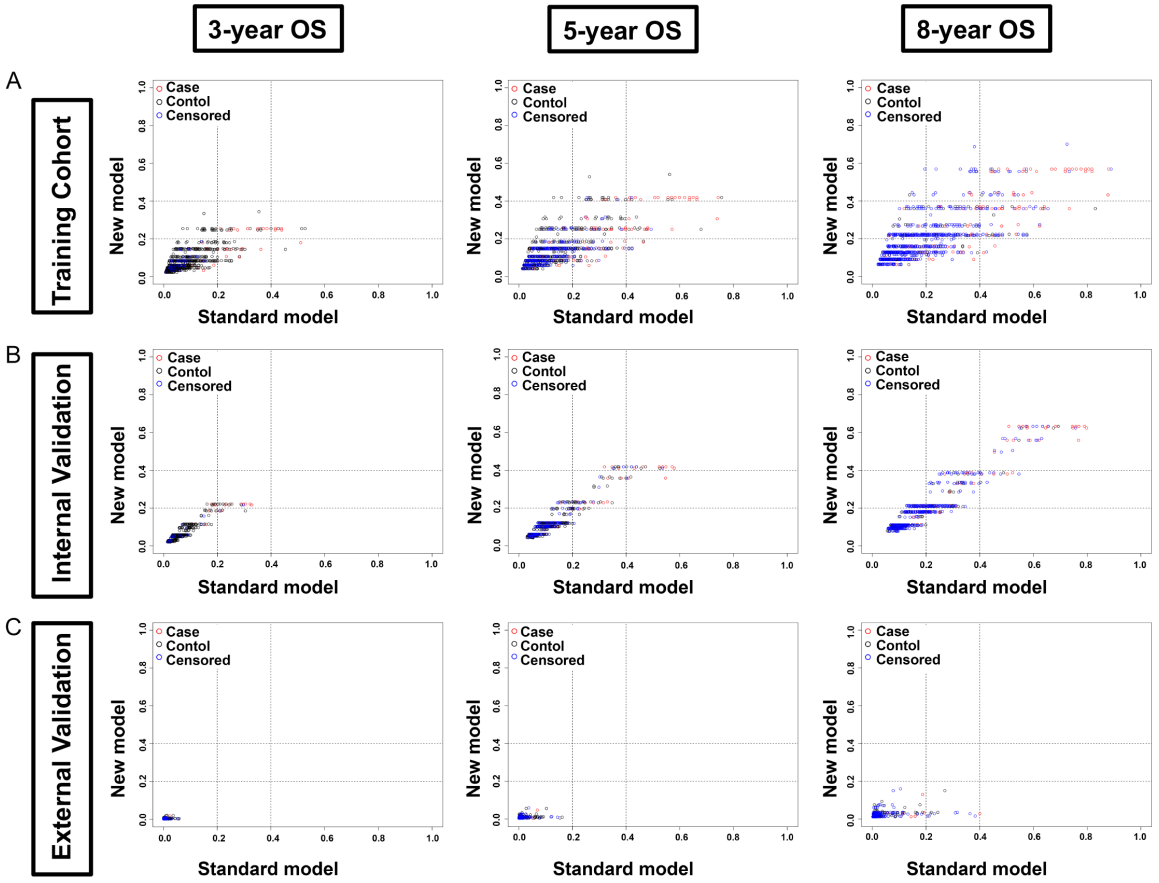

Figure 9. The NRI curves of the previous prediction model were plotted based on 3-, 5-, 8-year OS benefits in the (A) training cohort, (B) internal validation cohort, and (C) external validation cohort in the patients with BC. NRI, net reclassification improvement; OS, overall survival; BC, breast cancer.

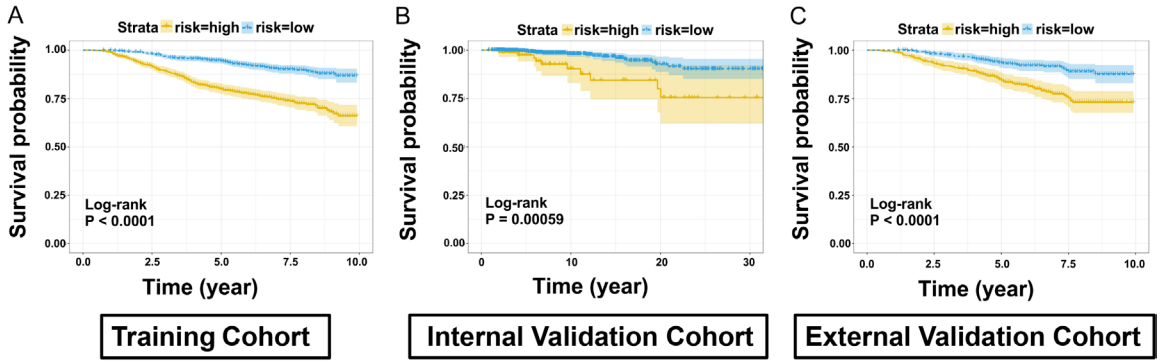

Figure 10. K-M OS curves plotted according to the median stratification of nomogram score in the (A) training cohort, (B) internal validation cohort, and (C) external validation cohort. K-M, Kaplan-Meier; OS, overall survival.

patients with different molecular subtypes, among which the triple-negative molecular subtype had a poor prognosis. There is evidence for an interaction between the ER and HER2 pathways. The ER pathway can be used as a bypass activation mechanism of downstream signals and may activate the HER2 pathway [41]. The activated HER2 signaling pathway will further promote the activity of the ER pathway, ultimately leading to impaired endocrine thera-

py responses and possibly altering the tumor response to HER2-targeted therapy [42]. In addition, some neoadjuvant clinical studies targeting anti-HER2 have shown a lower pathologic complete response rate (pCR) in patients with the luminal B subtype than in patients with the luminal A subtype [43].

This study included high-quality SEER data in the training cohort and a multicenter Chinese population as the validation cohort to ensure accuracy in identifying independent risk factors affecting OS in BC patients. To our knowledge, this study established a competitive risk model for predicting OS in BC patients and verified the accuracy of the model. However, there are still potential limitations to the study. First, this study is a retrospective study, and selection bias may lead to bias in the results. Second, the SEER database lacks specific information about systemic therapy and some important molecular factors, such as HER2-targeted therapy, endocrine therapy, immunotherapy and treatment response; programmed death ligand 1 (PD-L1) expression or microsatellite status; tumor markers, family history, menstrual history, fertility status, and weight. If more comprehensive patient information can be included, the prediction accuracy of the model can be improved; however, with further improvement of the database, these problems can be further solved. Finally, large prospective studies or high-quality randomized controlled trials are needed for further validation.

Conclusions

In conclusion, we developed and validated a competitive nomogram and risk classification system to reliably predict OS at 3, 5, and 8 years in BC patients, with superior prognostic value compared to the 7th AJCC TNM staging system alone and compared to a previously published prediction model. This nomogram can improve prognosis assessment, better predict individual survival, guide follow-up management strategies, and assist in facilitating individualized therapy. The created nomogram has promising clinical application prospects.

Acknowledgements

We acknowledge the work of doctors in these individuals (Drs. Huan Wang, Guang-Fa Xia, Zi-Ran Zhang, and Juan-Ying Zhu). This work was supported by the Jiaxing Science and Technology Plan Project (No. 2022AD30078); and the Natural Science Foundation of Henan Province (No. 222300420557).

Informed consent was obtained from all subjects involved in the study.

Disclosure of conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Abbreviations

BC, breast cancer; OS, overall survival; SEER, Surveillance, Epidemiology and End Results; LASSO, Least absolute shrinkage and selection operator; ROC, receiver operating characteristic; AUC, area under the curve; DCA, decision curve analysis; C-index, concordance index; NRI, net reclassification improvement; AJCC, American Joint Committee on Cancer; TNM, tumor node metastasis; K-M, Kaplan-Meier; HER2, human epidermal growth factor receptor type 2; CT, chemotherapy; RT, radiotherapy; IDC, infiltrating duct carcinoma; HR, hazard ratio; CI, confidence interval; AIC, akaike information criterion; ANOVA, analysis of variance; pCR, pathologic complete response; PD-L1, programmed death ligand 1.

Address correspondence to: Hui-Ke Wang, Department of Oncology, The First Affiliated Hospital of Zhengzhou University, Zhengzhou 450052, Henan, China. Tel: +86-15083086069; E-mail: 862367541@qq.com; Juan-Ying Zhu, Maternity and Child Health Care Affiliated Hospital, Jiaxing University, Jiaxing 314000, Zhejiang, China. Tel: +86-13736881201; E-mail: 2720597818@qq.com

References

[1] Giaquinto AN, Sung H, Miller KD, Kramer JL, Newman LA, Minihan A, Jemal A and Siegel RL. Breast cancer statistics, 2022. CA Cancer J Clin 2022; 72: 524-541.

[2] Ren X, Cui H, Wu J, Zhou R, Wang N, Liu D, Xie X, Zhang H, Liu D, Ma X, Dang C, Kang H and Lin S. Identification of a combined apoptosis and hypoxia gene signature for predicting prognosis and immune infiltration in breast cancer. Cancer Med 2022; 11: 3886-3901.

[3] Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A and Bray F. Global can-

cer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2021; 71: 209-249.
[4] Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA and Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018; 68: 394-424.
[5] Siegel RL, Miller KD, Wagle NS and Jemal A. Cancer statistics, 2023. CA Cancer J Clin 2023; 73: 17-48.
[6] Waks AG and Winer EP. Breast cancer treatment: a review. JAMA 2019; 321: 288-300.
[7] Huang J, Barbera L, Brouwers M, Browman G and Mackillop WJ. Does delay in starting treatment affect the outcomes of radiotherapy? A systematic review. J Clin Oncol 2003; 21: 555-563.
[8] Nielsen HM and Overgaard M. Survey of a Cochrane review concerning sequencing of adjuvant chemotherapy and postoperative radiotherapy in breast cancer. Ugeskr Laeger 2007; 169: 3101-3104.
[9] McGee SF, Mazzarello S, Caudrelier JM, Lima MAG, Hutton B, Sienkiewicz M, Stober C, Fernandes R, Ibrahim MFK, Vandermeer L, Hilton J, Shorr R, Fergusson D and Clemons M. Optimal sequence of adjuvant endocrine and radiation therapy in early-stage breast cancer - a systematic review. Cancer Treat Rev 2018; 69: 132-142.
[10] Heller DR, Zhuo H, Zhang Y, Parikh N, Fusi S, Alperovich M, Lannin DR, Higgins SA, Avraham T and Killelea BK. Surgical outcomes of mastectomy with immediate autologous reconstruction followed by radiation. Ann Surg Oncol 2021; 28: 2169-2179.
[11] Radak M, Lafta HY and Fallahi H. Machine learning and deep learning techniques for breast cancer diagnosis and classification: a comprehensive review of medical imaging studies. J Cancer Res Clin Oncol 2023; 149: 10473-10491.
[12] Muradali D, Fletcher GG, Cordeiro E, Fienberg S, George R, Kulkarni S, Seely JM, Shaheen R and Eisen A. Preoperative breast magnetic resonance imaging: an ontario health (cancer care ontario) clinical practice guideline. Curr Oncol 2023; 30: 6255-6270.
[13] Lim JY, Gerber SA, Murphy SP and Lord EM. Type I interferons induced by radiation therapy mediate recruitment and effector function of CD8(+) T cells. Cancer Immunol Immunother 2014; 63: 259-271.
[14] Burnette BC, Liang H, Lee Y, Chlewicki L, Khodarev NN, Weichselbaum RR, Fu YX and Auh SL. The efficacy of radiotherapy relies upon induction of type i interferon-dependent innate and adaptive immunity. Cancer Res 2011; 71: 2488-2496.
[15] Deng L, Liang H, Xu M, Yang X, Burnette B, Arina A, Li XD, Mauceri H, Beckett M, Darga T, Huang X, Gajewski TF, Chen ZJ, Fu YX and Weichselbaum RR. STING-dependent cytosolic DNA sensing promotes radiation-induced type I interferon-dependent antitumor immunity in immunogenic tumors. Immunity 2014; 41: 843-852.
[16] Woo SR, Fuertes MB, Corrales L, Spranger S, Furdyna MJ, Leung MY, Duggan R, Wang Y, Barber GN, Fitzgerald KA, Alegre ML and Gajewski TF. STING-dependent cytosolic DNA sensing mediates innate immune recognition of immunogenic tumors. Immunity 2014; 41: 830-842.
[17] Cai H, Li L, Slavik K, Huang J, Yin T, Hédelin L, Xiang Z, Yang Y, Li X, Chen Y, Wei Z, Deng H, Chen D, Jiao R, Martins N, Meignin C, Kranzusch P and Imler JL. A novel virus-induced cyclic dinucleotide, 2'3'-c-di-GMP, mediates STING-dependent antiviral immunity in Drosophila. bioRxiv 2023; [Epub ahead of print].
[18] Jin C, Cao J, Cai Y, Wang L, Liu K, Shen W and Hu J. A nomogram for predicting the risk of invasive pulmonary adenocarcinoma for patients with solitary peripheral subsolid nodules. J Thorac Cardiovasc Surg 2017; 153: 462-469, e461.
[19] Veale D, Miles S, Bramley S, Muir G and Hodsoll J. Am I normal? A systematic review and construction of nomograms for flaccid and erect penis length and circumference in up to 15,521 men. BJU Int 2015; 115: 978-986.
[20] Wu J, Qiu J, Jiang W, Qiu J, Zhang L, Zhao R and Yu C. Development and validation of a nomogram predicting the probability of type a aortic dissection at a diameter below 55 mm: a retrospective cohort study. Int J Surg 2018; 60: 266-272.
[21] Balachandran VP, Gonen M, Smith JJ and DeMatteo RP. Nomograms in oncology: more than meets the eye. Lancet Oncol 2015; 16: e173-e180.
[22] Wu H, Ding P, Wu J, Sun C, Guo H, Chen S, Lowe S, Yang P, Tian Y, Liu Y and Zhao Q. A new online dynamic nomogram: construction and validation of a predictive model for distant metastasis risk and prognosis in patients with gastrointestinal stromal tumors. J Gastrointest Surg 2023; 27: 1429-1444.
[23] Hwang HW, Jung H, Hyeon J, Park YH, Ahn JS, Im YH, Nam SJ, Kim SW, Lee JE, Yu JH, Lee SK, Choi M, Cho SY and Cho EY. A nomogram to predict pathologic complete response (pCR) and the value of tumor-infiltrating lymphocytes (TILs) for prediction of response to neoadju-

vant chemotherapy (NAC) in breast cancer patients. Breast Cancer Res Treat 2019; 173: 255-266.
[24] Yan M, Peng C, He D, Xu D and Yang C. A nomogram for enhancing the diagnostic effectiveness of solid breast BI-RADS 3-5 masses to determine malignancy based on imaging aspects of conventional ultrasonography and contrast-enhanced ultrasound. Clin Breast Cancer 2023; 23: 693-703.
[25] Fouad TM, Barrera AMG, Reuben JM, Lucci A, Woodward WA, Stauder MC, Lim B, DeSnyder SM, Arun B, Gildy B, Valero V, Hortobagyi GN and Ueno NT. Inflammatory breast cancer: a proposed conceptual shift in the UICC-AJCC TNM staging system. Lancet Oncol 2017; 18: e228-e232.
[26] Nam SH, Bae MR, Roh JL, Gong G, Cho KJ, Choi SH, Nam SY and Kim SY. A comparison of the 7th and 8th editions of the AJCC staging system in terms of predicting recurrence and survival in patients with papillary thyroid carcinoma. Oral Oncol 2018; 87: 158-164.
[27] Teichgraeber DC, Guirguis MS and Whitman GJ. Breast cancer staging: updates in the AJCC cancer staging manual, 8th edition, and current challenges for radiologists, from the AJR special series on cancer staging. AJR Am J Roentgenol 2021; 217: 278-290.
[28] O'Sullivan CC, Loprinzi CL and Haddad TC. Updates in the evaluation and management of breast cancer. Mayo Clin Proc 2018; 93: 794-807.
[29] Loibl S, Loirat D, Tolaney SM, Punie K, Oliveira M, Rugo HS, Bardia A, Hurvitz SA, Brufsky AM, Kalinsky K, Cortés J, O'Shaughnessy JA, Dieras V, Carey LA, Gianni L, Gharaibeh M, Preger L, Phan S, Chang L, Shi L and Piccart MJ. Health-related quality of life in the phase III ASCENT trial of sacituzumab govitecan versus standard chemotherapy in metastatic triple-negative breast cancer. Eur J Cancer 2023; 178: 23-33.
[30] Li Y, Wu G, Zhang Y, Yang W, Wang X, Duan L, Niu L, Chen J, Zhou W, Liu J, Fan D and Hong L. Effects of marital status on survival of retroperitoneal liposarcomas stratified by age and sex: a population-based study. Cancer Med 2023; 12: 1779-1790.
[31] Shrout MR, Renna ME, Madison AA, Alfano CM, Povoski SP, Lipari AM, Agnese DM, Farrar WB, Carson WE 3rd and Kiecolt-Glaser JK. Breast cancer survivors' satisfying marriages predict better psychological and physical health: a longitudinal comparison of satisfied, dissatisfied, and unmarried women. Psychooncology 2021; 30: 699-707.
[32] Ellis L, Canchola AJ, Spiegel D, Ladabaum U, Haile R and Gomez SL. Racial and ethnic disparities in cancer survival: the contribution of tumor, sociodemographic, institutional, and neighborhood characteristics. J Clin Oncol 2018; 36: 25-33.
[33] McPherson K, Steel CM and Dixon JM. ABC of breast diseases. Breast cancer--epidemiology, risk factors and genetics. BMJ 1994; 309: 1003-1006.
[34] Freedman RA, Keating NL, Lin NU, Winer EP, Vaz-Luis I, Lii J, Exman P and Barry WT. Breast cancer-specific survival by age: worse outcomes for the oldest patients. Cancer 2018; 124: 2184-2191.
[35] Li S, Wang X, Yang J, Lv M, Zhang X, Li C, Zhang L, Shen Y, Zhang X, Chen Z, Wang F, Wang X, Li D, Yi M and Yang J. Clinicopathological features and survival of early stage breast cancer in northwest China: a population-based retrospective study of 1287 patients. Thorac Cancer 2018; 9: 10-18.
[36] Ko NY, Hong S, Winn RA and Calip GS. Association of insurance status and racial disparities with the detection of early-stage breast cancer. JAMA Oncol 2020; 6: 385-392.
[37] Williams DR and Collins C. Racial residential segregation: a fundamental cause of racial disparities in health. Public Health Rep 2001; 116: 404-416.
[38] Finkelstein EA, Khavjou OA, Mobley LR, Haney DM and Will JC. Racial/ethnic disparities in coronary heart disease risk factors among WISEWOMAN enrollees. J Womens Health (Larchmt) 2004; 13: 503-518.
[39] Kressin NR and Petersen LA. Racial differences in the use of invasive cardiovascular procedures: review of the literature and prescription for future research. Ann Intern Med 2001; 135: 352-366.
[40] Berkman A, F Cole B, Ades PA, Dickey S, Higgins ST, Trentham-Dietz A, Sprague BL and Lakoski SG. Racial differences in breast cancer, cardiovascular disease, and all-cause mortality among women with ductal carcinoma in situ of the breast. Breast Cancer Res Treat 2014; 148: 407-413.
[41] Deluche E, Antoine A, Bachelot T, Lardy-Cleaud A, Dieras V, Brain E, Debled M, Jacot W, Mouret-Reynier MA, Goncalves A, Dalenc F, Patsouris A, Ferrero JM, Levy C, Lorgis V, Vanlemmens L, Lefeuvre-Plesse C, Mathoulin-Pelissier S, Petit T, Uwer L, Jouannaud C, Leheurteur M, Lacroix-Triki M, Courtinard C, Perol D, Robain M and Delaloge S. Contemporary outcomes of metastatic breast cancer among 22,000 women from the multicentre ESME cohort 2008-2016. Eur J Cancer 2020; 129: 60-70.
[42] DeMichele A, Yee D and Esserman L. Mechanisms of resistance to neoadjuvant chemotherapy in breast cancer. N Engl J Med 2017; 377: 2287-2289.

[43] Bidard FC, Kaklamani VG, Neven P, Streich G, Montero AJ, Forget F, Mouret-Reynier MA, Sohn JH, Taylor D, Harnden KK, Khong H, Kocsis J, Dalenc F, Dillon PM, Babu S, Waters S, Deleu I, García Sáenz JA, Bria E, Cazzaniga M, Lu J, Aftimos P, Cortés J, Liu S, Tonini G, Laurent D, Habboubi N, Conlan MG and Bardia A. Elacestrant (oral selective estrogen receptor degrader) versus standard endocrine therapy for estrogen receptor-positive, human epidermal growth factor receptor 2-negative advanced breast cancer: results from the randomized phase III EMERALD trial. J Clin Oncol 2022; 40: 3246-3256.

[44] Zhang D, Li L and Ma F. Prognosis stratification and postoperative radiation therapy utilization in adenoid cystic carcinoma of the breast. Breast 2022; 66: 40-48.