Original Article

Construction and validation of a nomogram for predicting the prognosis of breast cancer patients who received adjuvant therapy: an analysis based on the SEER database

Qingliang Jiang[1#^], Xianglin Liu[2#], Yuting Wu[2], Jiaqi Du[2], Yingying Rao[2], Jiayu Li[2], Hengyu Li[2^]

[1]Department of Hepatobiliary Surgery, Eastern Hepatobiliary Surgery Hospital, Naval Medical University, Shanghai, China; [2]Department of Breast and Thyroid Surgery, Changhai Hospital, Naval Medical University, Shanghai, China

Contributions: (I) Conception and design: H Li, Q Jiang, X Liu; (II) Administrative support: None; (III) Provision of study materials or patients: None; (IV) Collection and assembly of data: J Du, Y Wu, Y Rao, J Li; (V) Data analysis and interpretation: X Liu, Q Jiang; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

[#]These authors contributed equally to this work as co-first authors.

Correspondence to: Hengyu Li, MD, PhD. Department of Breast and Thyroid Surgery, Changhai Hospital, Naval Medical University, 168 Changhai Road, Yangpu District, Shanghai 200433, China. Email: drlhy@foxmail.com.

Background: Breast cancer is the most common malignant tumor in women globally. Despite advances in primary treatment, the role of adjuvant therapy in reducing recurrence and improving survival is critical; however, there is a notable lack of tailored prognostic models for patients receiving adjuvant therapy. This study used the Surveillance, Epidemiology, and End Results (SEER) database to develop a prognostic nomogram for breast cancer patients receiving adjuvant therapy.

Methods: The data of breast cancer patients who received adjuvant therapy after surgery in 2014–2015 were extracted from the SEER database. Univariate Cox regression identified significant prognostic variables that were further refined by least absolute shrinkage and selection operator (LASSO) regression and cross-validation analyses. These variables were incorporated into a multivariate Cox regression analysis to establish the predictive model. This model was visualized and validated using various statistical measures.

Results: A total of 54,960 patients were included in the study, with 38,472 in the training set and 16,488 in the validation set. Age, sex, race, marital status, grade, tumor (T) stage, lymph node (N) stage, subtype, and radiotherapy were found to be significant independent risk factors of 1-, 3-, and 5-year overall survival (OS). The receiver operating characteristic curve area for 1-, 3-, and 5-year OS was >0.76 in both sets. The consistency index values were 0.768 and 0.763 for the training and validation sets, respectively. The calibration curves showed good fit, and the nomogram exhibited substantial clinical utility.

Conclusions: Incorporating various significant factors, the constructed nomogram was able to effectively predict the prognosis of breast cancer patients who received adjuvant therapy. This nomogram extends understandings of complex prognosis scenarios. In addition, it could enhance personalized treatment plans and assist in patient counseling.

Keywords: Breast cancer; adjuvant therapy; prognosis; Surveillance, Epidemiology, and End Results database (SEER database)

Submitted Jan 02, 2024. Accepted for publication May 30, 2024. Published online Jun 27, 2024.
doi: 10.21037/gs-23-537
View this article at: https://dx.doi.org/10.21037/gs-23-537

^ ORCID: Qingliang Jiang, 0000-0003-1793-8305; Hengyu Li, 0000-0002-1007-6493.

Introduction

Breast cancer is one of the most common malignant tumors in women worldwide, and it also has one of the highest mortality rates (1). According to the latest global cancer burden data released by the International Agency on Cancer Research (IARC) of the World Health Organization (WHO) (2), the number of new cases of breast cancer reached 2.26 million in 2020, making breast cancer the most common cancer worldwide for the first time. Breast cancer accounted for 11.7% of all new cancer cases in 2020 (2). Breast cancer is also the fifth leading cause of cancer-related death worldwide, claiming 685,000 lives in 2020 (2). As a populous country, the number of breast cancer patients in China is high (3). Among Chinese women, breast cancer ranked first in terms of incidence (19.9%) and fourth in terms of mortality (9.9%) among all cancers in 2020 (2). The high incidence of breast cancer and the risk of recurrence, which lasts between 10 and 32 years (4), also make postoperative treatment one of the most important aspects of breast cancer treatment.

In the postoperative treatment of breast cancer, adjuvant therapy (e.g., chemotherapy and radiotherapy) can effectively kill residual cancer cells, reduce the risk of recurrence, and improve the survival rate of patients (5-8). Adjuvant therapy has become an important means of breast cancer treatment; however, uncertainty as to its treatment effects remains. Factors such as individual differences, tumor characteristics, and treatment regimens may affect patient prognosis. Therefore, it is necessary to study the prognostic risk factors of patients receiving adjuvant therapy, particularly in the absence of existing models specifically designed for this patient cohort.

To better investigate the prognostic risk factors of breast cancer patients treated with adjuvant therapy after surgery, data from the Surveillance, Epidemiology, and End Results (SEER) database were used in this study. The SEER database collects the clinical, demographic, and treatment information of breast cancer patients from 18 regions in the United States and consolidates it into a centralized database (9). The application of this large database has promoted extensive breast cancer-related research and provided important support for the early prediction, treatment, and management of breast cancer. This study aimed to develop and validate a new multivariable prediction model using the SEER database to improve the accuracy of prognosis and treatment outcomes for breast cancer patients undergoing adjuvant therapy. This model is expected to enable more personalized prediction, taking into account the diversity of patient characteristics and treatment responses. We present this article in accordance with the TRIPOD reporting checklist (available at https://gs.amegroups.com/article/view/10.21037/gs-23-537/rc).

Highlight box

Key findings

- The study developed a prognostic nomogram for breast cancer patients receiving adjuvant therapy, using data from the Surveillance, Epidemiology, and End Results (SEER) database.
- The significant independent risk factors for overall survival (OS) included age, sex, race, marital status, grade, tumor (T) stage, lymph node (N) stage, subtype, and radiotherapy.
- The model demonstrated high accuracy and utility in predicting 1-, 3-, and 5-year OS as validated by various statistical measures.

What is known and what is new?

- The prognosis of breast cancer patients post-adjuvant therapy is influenced by various factors; previous models have used clinical and biological data for prognosis prediction.
- This study developed a comprehensive nomogram incorporating a wide range of factors, including demographic, clinical, and treatment-related variables. It uniquely applied least absolute shrinkage and selection operator regression for the variable selection, enhancing the performance and interpretability of the model.

What is the implication, and what should change now?

- The developed nomogram enables the more accurate and personalized prediction of the survival of breast cancer patients undergoing adjuvant therapy, and could potentially guide clinical decision making.
- Clinicians should consider incorporating this nomogram into their practice to enable better prognostic assessment and tailored patient counseling. Additionally, the healthcare system should acknowledge the value of integrating comprehensive data analysis tools in enhancing cancer care.

Methods

General information

This study was based on the clinical data of 18 SEER cancer registries. A total of 54,960 patients with breast cancer who received adjuvant therapy from 2014 to 2015 were selected from the SEER database as the research subjects. These patients were randomly allocated to the training set and validation set at a 2:1 ratio, resulting in 38,472 patients in the training set and 16,488 patients in the validation set. Demographic data, clinical indicators, and prognostic

follow-up information were collected. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

Inclusion and exclusion criteria

To be eligible for inclusion in this study, the patients had to meet the following inclusion criteria: (I) have been diagnosed with breast cancer from 2014 to 2015; (II) have received adjuvant therapy; and (III) have a pathological confirmation of the diagnosis. Patients were excluded from the study if they met any of the following exclusion criteria: (I) had a lack of detailed treatment information; (II) had missing clinical data or prognostic information; and/or (III) had distant metastasis at the time of diagnosis.

Statistical analyses

Overall survival (OS) was defined as the time from the date of cancer diagnosis to the date of death from any cause. If a patient was still alive at the last follow-up, their data were censored at that point. In this study, the follow-up data collection continued until the last update of the SEER database, which was in 2020. Thus, the data comprised the most recent data available at the time of download. This date represents the cut-off date for all patient follow-up information, allowing for a uniform assessment and analysis of the study subjects.

SEER* Stat Version 8.4.0.1 software (RRID:SCR_003293) was used to collect the data, and RStudio 4.2.2 software (RRID:SCR_000432) was used to analyze and process the data. The count data are expressed as the number of cases or percentage, and the Chi-square test was used for comparisons between groups. Univariate Cox regression was used to analyze the factors influencing the prognosis of breast cancer patients who received adjuvant therapy. The significant variables of the univariate analysis were screened by the least absolute shrinkage and selection operator (LASSO) regression and cross-validation analyses. LASSO regression was also used to identify the optimal predictor variables and avoid a certain degree of overfitting. In the cross-validation analysis, the selected lambda-1se value was the lambda of the simplest model obtained within one variance of the lambda-minimum value. This value established a model with good performance and a minimum number of independent variables, enabling the best predictor to be identified. The variables screened again were included in the multivariate Cox regression analysis to determine the significant independent risk factors for OS (P<0.05). Finally, the independent risk factors were used to draw a nomogram to predict the 1-, 3-, and 5-year OS of the patients, and a regression model was established.

Internal verification was used as the verification method. The concordance index (C-index) plot of the training set and the validation set over time was drawn. The calibration curve (using the 500 bootstrap automatic sampling method), receiver operating characteristic (ROC) curve, and decision curve analysis (DCA) results were drawn to verify the reliability and practicability of the model. We calculated the OS score of each patient based on the nomogram model, and used RStudio 4.2.2 to calculate the optimal cut-off value of the total score. Based on the optimal cut-off value, patients were divided into low- and high-risk groups. The Kaplan-Meier (K-M) method and log-rank test were used to compare survival differences between the risk groups. A P value <0.05 was considered statistically significant.

Results

Patient characteristics

A total of 54,960 breast cancer patients were included in this study, including 38,472 in the training set and 16,488 in the validation set. According to professional knowledge and the literature review, the selected research factors were age, race, marital status, sex, grade (tumor histological grade), tumor (T) stage, lymph node (N) stage [derived from the American Joint Committee on Cancer (AJCC), 6th edition], primary site, laterality, subtype, radiotherapy, chemotherapy, and tumor location. The baseline characteristics of the patients in the training and validation sets are shown in *Table 1*.

Analysis of prognostic factors

A total of 13 variables were included in the univariate Cox regression analysis. The results showed that all the variables, except laterality [hazard ratio =0.97, 95% confidence interval (CI): 0.90–1.04, P=0.39], were prognostic factors for breast cancer patients who received adjuvant therapy (P<0.05) (*Table 2* and *Figure 1A*). To avoid a certain degree of overfitting, a LASSO regression analysis (*Figure 1B*) and cross-validation analysis (*Figure 1C*) were performed of the above 12 variables. The constructed model was optimal when lambda was lambda.1se (lambda =0.00580). At this point, the primary site, chemotherapy, and tumor location were excluded from the LASSO regression. The subsequent

Table 1 Baseline characteristics of patients who received adjuvant therapy for breast cancer

Variables	Internal validation cohort (N=16,488)	Training cohort (N=38,472)	Overall (N=54,960)	P value
Age (years)				0.21
<50	2,931 (17.8)	6,678 (17.4)	9,609 (17.5)	
50–59	3,961 (24.0)	9,459 (24.6)	13,420 (24.4)	
60–69	5,072 (30.8)	11,983 (31.1)	17,055 (31.0)	
70–79	3,418 (20.7)	7,725 (20.1)	11,143 (20.3)	
≥80	1,106 (6.7)	2,627 (6.8)	3,733 (6.8)	
Sex				0.20
Female	16,381 (99.4)	38,182 (99.2)	54,563 (99.3)	
Male	107 (0.6)	290 (0.8)	397 (0.7)	
Race				0.72
Black	1,539 (9.3)	3,507 (9.1)	5,046 (9.2)	
White	13,270 (80.5)	31,033 (80.7)	44,303 (80.6)	
Other	1,679 (10.2)	3,932 (10.2)	5,611 (10.2)	
Marital status				0.54
Single	6,497 (39.4)	15,050 (39.1)	21,547 (39.2)	
Married	9,991 (60.6)	23,422 (60.9)	33,413 (60.8)	
Grade				0.14
I	4,431 (26.9)	10,374 (27.0)	14,805 (26.9)	
II	7,823 (47.4)	17,966 (46.7)	25,789 (46.9)	
III	4,219 (25.6)	10,110 (26.3)	14,329 (26.1)	
IV	15 (0.1)	22 (0.1)	37 (0.1)	
Stage T				0.84
T1	11,373 (69.0)	26,524 (68.9)	37,897 (69.0)	
T2	4,498 (27.3)	10,509 (27.3)	15,007 (27.3)	
T3	516 (3.1)	1,226 (3.2)	1,742 (3.2)	
T4	101 (0.6)	213 (0.6)	314 (0.6)	
Stage N				0.42
N0	12,467 (75.6)	12,467 (75.6)	41,503 (75.5)	
N1	3,228 (19.6)	7,462 (19.4)	10,690 (19.5)	
N2	558 (3.4)	1,370 (3.6)	1,928 (3.5)	
N3	235 (1.4)	604 (1.6)	839 (1.5)	

Table 1 (*continued*)

Table 1 (*continued*)

Variables	Internal validation cohort (N=16,488)	Training cohort (N=38,472)	Overall (N=54,960)	P value
Primary site				0.63
Axillary tail and overlapping lesion	4,433 (26.9)	10,263 (26.7)	14,696 (26.7)	
Lower	2,461 (14.9)	5,870 (15.3)	8,331 (15.2)	
Nipple and areola and center	921 (5.6)	2,079 (5.4)	3,000 (5.5)	
Upper	8,673 (52.6)	20,260 (52.7)	28,933 (52.6)	
Laterality				0.90
Left—origin of primary	8,297 (50.3)	19,383 (50.4)	27,680 (50.4)	
Right—origin of primary	8,191 (49.7)	19,089 (49.6)	27,280 (49.6)	
Breast subtype				0.11
$HR^-/HER2^-$	1,219 (7.4)	2,766 (7.2)	3,985 (7.3)	
$HR^-/HER2^+$	412 (2.5)	1,040 (2.7)	1,452 (2.6)	
$HR^+/HER2^-$	13,444 (81.5)	31,177 (81.0)	44,621 (81.2)	
$HR^+/HER2^+$	1,413 (8.6)	3,489 (9.1)	4,902 (8.9)	
Chemotherapy				0.73
Non-chemotherapy	10,122 (61.4)	23,555 (61.2)	33,677 (61.3)	
Chemotherapy	6,366 (38.6)	14,917 (38.8)	21,283 (38.7)	
Radiation				0.68
Non-radiation	6,238 (37.8)	14,483 (37.6)	20,721 (37.7)	
Radiation	10,250 (62.2)	23,989 (62.4)	34,239 (62.3)	
Location of the tumor				0.89
Localized	12,269 (74.4)	28,652 (74.5)	40,921 (74.5)	
Direct extension or regional	4,219 (25.6)	9,820 (25.5)	14,039 (25.5)	
Status*				0.89
Alive	15,120 (91.7)	35,295 (91.7)	50,415 (91.7)	
Dead	1,368 (8.3)	3,177 (8.3)	4,545 (8.3)	

Data are expressed as n (%). *, the survival status variable is based on the latest update of the SEER database (November 2022 Submission). HR, hormone receptor; HER2, human epidermal growth factor receptor 2.

multivariate Cox regression analysis of the remaining nine variables confirmed that each was a significant independent predictor of OS (P<0.05, *Figure 1D*).

Construction and validation of the prognostic prediction nomogram

Based on the results of the analysis, the prognostic factors with significant differences and clinical significance in the Cox proportional hazards regression model (i.e., age, sex, race, marital status, grade, T stage, N stage, subtype, and radiotherapy) were included in the nomogram. RStudio software was used to construct the nomogram (*Figure 2*).

As a visual model, a nomogram specifies the scoring standard according to the regression coefficient of all the independent variables, and gives each level of each

Table 2 Cox regression analysis of the factors affecting the prognosis of breast cancer patients who received adjuvant therapy

Variables	Univariate analysis			Multivariate analysis		
	HR	95% CI	P value	HR	95% CI	P value
Age (years)						
<50	Ref			Ref		
50–59	1.14	0.98–1.32	0.10	1.30	1.11–1.51	0.001
60–69	1.54	1.34–1.77	<0.001	1.91	1.66–2.20	<0.001
70–79	3.09	2.70–3.54	<0.001	3.80	3.31–4.35	<0.001
≥80	8.42	7.33–9.67	<0.001	8.41	7.29–9.69	<0.001
Sex						
Female	Ref			Ref		
Male	2.46	1.89–3.21	<0.001	1.81	1.39–2.37	<0.001
Race						
Black	Ref			Ref		
White	0.67	0.61–0.75	<0.001	0.75	0.68–0.84	<0.001
Other	0.45	0.38–0.53	<0.001	0.59	0.50–0.70	<0.001
Marital status						
Single	Ref			Ref		
Married	0.51	0.47–0.55	<0.001	0.71	0.66–0.77	<0.001
Grade						
I	Ref			Ref		
II	1.39	1.26–1.54	<0.001	1.17	1.06–1.30	0.002
III	2.37	2.15–2.62	<0.001	1.77	1.58–1.98	<0.001
IV	0.78	0.11–5.55	0.80	0.73	0.10–5.22	0.76
Stage T						
T1	Ref			Ref		
T2	2.10	1.95–2.26	<0.001	1.48	1.37–1.60	<0.001
T3	3.49	3.05–4.01	<0.001	2.18	1.88–2.53	<0.001
T4	6.91	5.46–8.74	<0.001	2.28	1.79–2.91	<0.001
Stage N						
N0	Ref			Ref		
N1	1.60	1.47–1.74	<0.001	1.53	1.41–1.67	<0.001
N2	3.02	2.65–3.44	<0.001	2.36	2.05–2.71	<0.001
N3	5.00	4.28–5.84	<0.001	3.50	2.96–4.13	<0.001

Table 2 (*continued*)

Table 2 (*continued*)

Variables	Univariate analysis			Multivariate analysis		
	HR	95% CI	P value	HR	95% CI	P value
Primary site*						
Axillary tail and overlapping lesion	Ref			–	–	–
Lower	1.03	0.92–1.15	0.56	–	–	–
Nipple and areola and center	1.47	1.28–1.70	<0.001	–	–	–
Upper	0.96	0.88–1.04	0.35	–	–	–
Laterality						
Left—origin of primary	Ref			–	–	–
Right—origin of primary	0.97	0.90–1.04	0.39	–	–	–
Breast subtype						
$HR^-/HER2^-$	Ref			Ref		
$HR^-/HER2^+$	0.78	0.64–0.96	0.02	0.68	0.55–0.83	<0.001
$HR^+/HER2^-$	0.50	0.45–0.56	<0.001	0.60	0.53–0.68	<0.001
$HR^+/HER2^+$	0.54	0.46–0.63	<0.001	0.51	0.44–0.60	<0.001
Chemotherapy						
Non-chemotherapy	Ref			–	–	–
Chemotherapy	1.10	1.02–1.18	0.009	–	–	–
Radiation						
Non-radiation	Ref			Ref		
Radiation	0.57	0.53–0.61	<0.001	0.64	0.60–0.69	<0.001
Location of the tumor						
Localized	Ref			–	–	–
Direct extension or regional	2.08	1.93–2.23	<0.001	–	–	–

*, see the International Classification of Diseases for Oncology, Third Edition (ICD-O-3). HR, hazard ratio; CI, confidence interval; HR, hormone receptor; HER2, human epidermal growth factor receptor 2; Ref, reference.

independent variable a score, which is mainly composed of the variable name and the scale line, and the corresponding line segment of each variable is marked with the scale information, representing the value range of the variable. The length of the tick indicates the contribution of the factor to the outcome event. The score in the figure is the single item score, indicating the corresponding single item score of each variable at different values. The total score represents the total score of the corresponding individual scores after the values of all the variables are added. Based on the total score, a vertical line can be drawn to obtain the 1-, 3-, and 5-year OS of the patient.

The established nomogram underwent internal validation. The C-index values of the training set and the validation set calculated by RStudio were 0.768 (95% CI: 0.760–0.771) and 0.763 (95% CI: 0.750–0.776) respectively, and the ROC curves were plotted. The areas under the curve (AUCs) of the ROC curves at 1, 3, and 5 years were

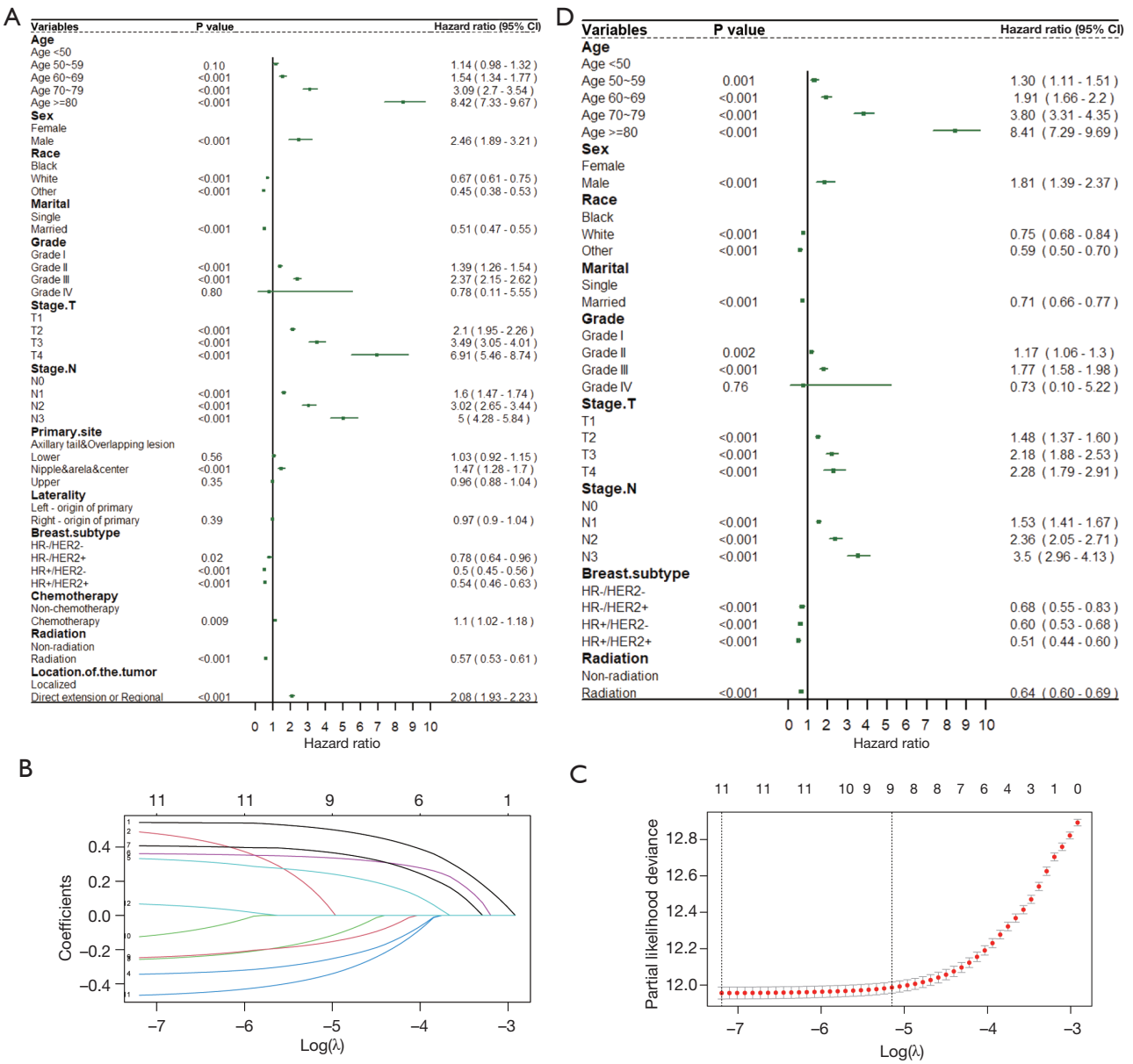

Figure 1 Analysis of prognostic factors. (A) Forest plot for the univariate Cox regression analysis. (B) LASSO regression analysis of selected variables in the training set. (C) Cross-validation analysis of selected variables in the training set. (D) Forest plot for the multivariate Cox regression analysis. CI, confidence interval; HR, hormone receptor; HER2, human epidermal growth factor receptor 2.

0.785, 0.793, and 0.775, respectively, in the training set (*Figure 3A*), and 0.798, 0.781, and 0.767, respectively, in the validation set (*Figure 3B*). In addition, the time C-index curves and 1-, 3-, and 5-year calibration curves of the training set and the validation set were plotted. The C-index curves were all >0.7 (*Figure 3C*), indicating that the model had high accuracy. The calibration curves (*Figure 3D*) of the two groups were close to the ideal reference line of 45 degrees, and the predicted values and actual results were well fitted, indicating that the survival prediction rates at 1, 3, and 5 years predicted by the model were in good agreement with the actual survival prediction rates. Finally, a clinical DCA of the training set and the validation set at 1, 3, and 5 years (*Figure 3E*) was conducted to check the clinical utility of the model, and the results showed that the nomogram had a good net benefit for clinical utility.

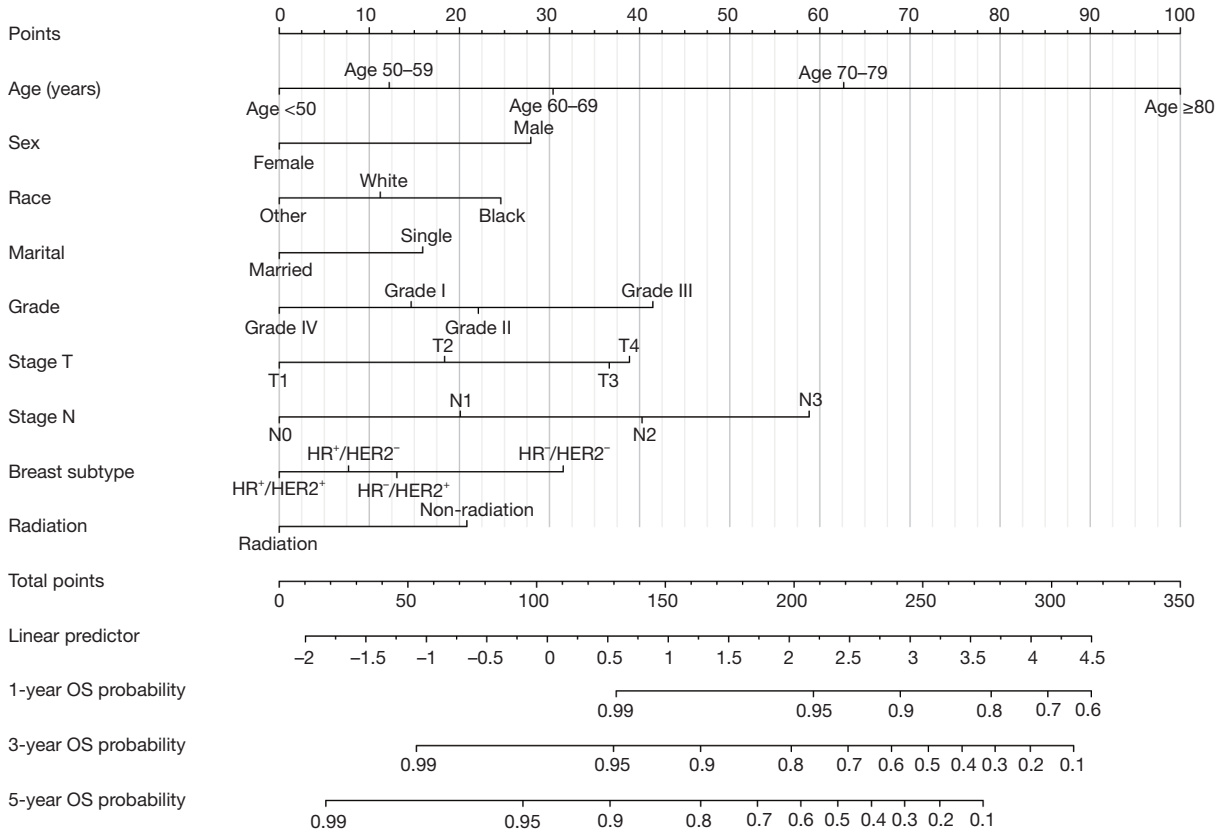

Figure 2 Nomogram for predicting the prognosis of breast cancer patients treated with adjuvant therapy after surgery. HR, hormone receptor; HER2, human epidermal growth factor receptor 2; OS, overall survival.

Risk stratification

Using the nomogram model, the total OS score of each patient was calculated. The best cut-off value of the total OS score of the patient was obtained by RStudio software, and the risk stratification was performed. The best cut-off value of the total OS score of the patients in the training set was 152.94. Based on the best cut-off value, the patients were divided into the low-risk group (total score <152.94) and the high-risk group (total score >152.94). The optimal cut-off value of the total OS score of the patients in the validation set was 129.34. Based on the optimal cut-off value, the patients were divided into the low-risk group (total score <129.34) and high-risk group (total score >129.34). The K-M method and log-rank test were used to compare the survival differences between the risk groups and to plot the K-M curves for both the training and validation sets (*Figure 4*). The results showed that the 1-, 3-, and 5-year OS rates of the low-risk group in the training set were 99.5%, 97.3%, and 94.3%, respectively, and those of the high-risk group were 96.6%, 83.8%, and 70.5%, respectively. In the validation cohort, the 1-, 3-, and 5-year OS rates of the low-risk patients were 99.6%, 97.4%, and 94.5%, respectively, and those of high-risk patients were 97.1%, 84.0%, and 71.7%, respectively. There were significant statistical differences between the risk groups in both the training and validation sets (P<0.001), and the prognosis of the low-risk group was significantly better than that of the high-risk group.

Discussion

Adjuvant therapy refers to the selection of appropriate chemotherapy, endocrine therapy, molecular targeted therapy, immunotherapy, and radiotherapy according to the clinical stage, molecular subtype, gene expression classification, and other factors after surgical resection of the tumor to eliminate possible residual cancer cells in

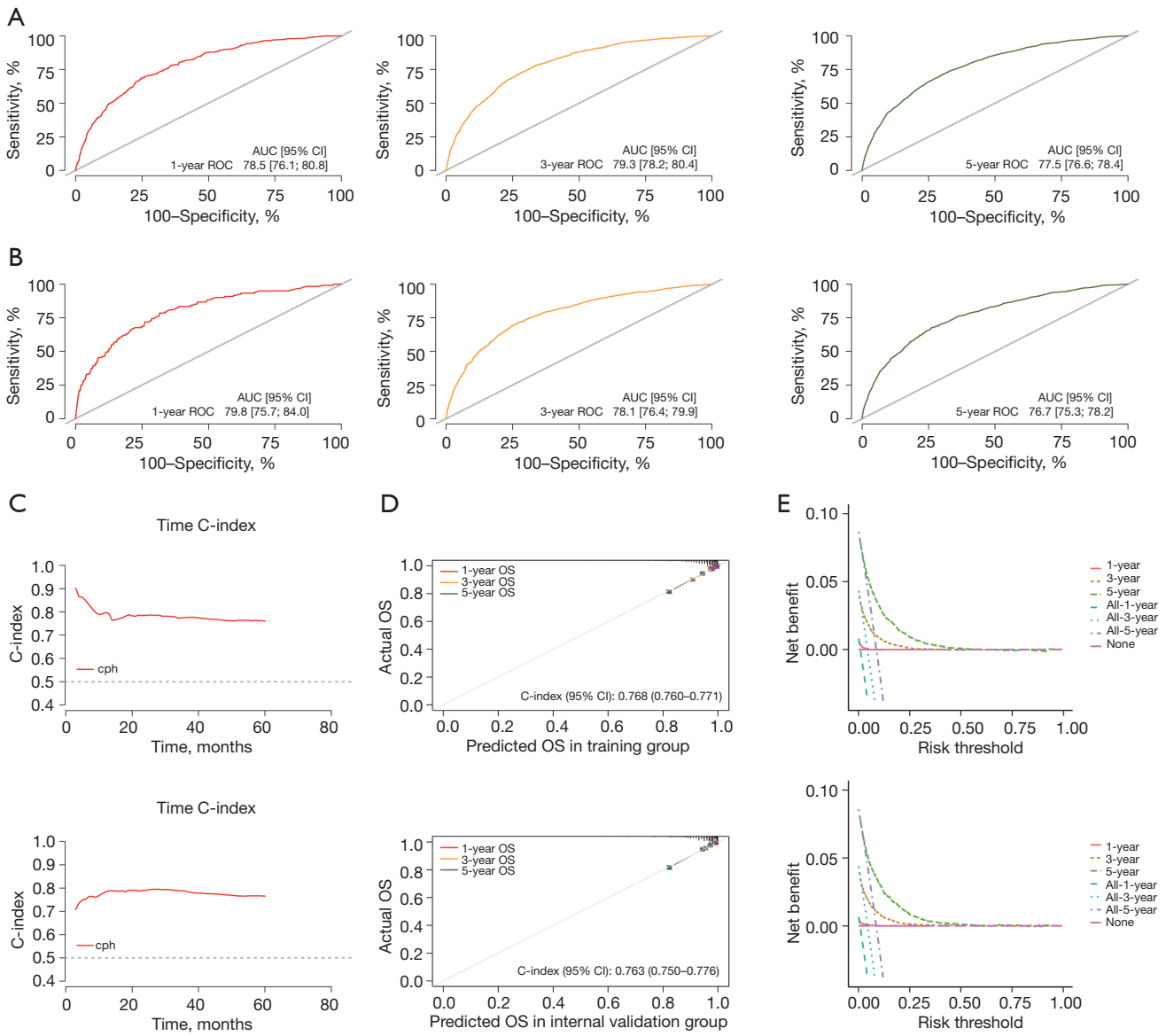

Figure 3 Validation of the prognostic prediction nomogram. (A) ROC curves at 1, 3, and 5 years for the training set. (B) ROC curves at 1, 3, and 5 years for the validation set. (C) C-index values over time for the training and validation sets. (D) Calibration curves for the training and validation sets. (E) DCA curves of the training and validation sets. AUC, area under the curve; ROC, receiver operating characteristic; C-index, concordance index; cph, Cox proportional hazards; OS, overall survival; CI, confidence interval; DCA, decision curve analysis.

the body, reduce the risk of recurrence or metastasis, and improve the survival rate and quality of life of patients (10). Adjuvant therapy for breast cancer is an important part of the comprehensive treatment of breast cancer. In recent years, with the continuous development of new drugs and the continuous application of new technologies, adjuvant therapy for breast cancer has progressed rapidly, providing more choices and improving patient prognosis.

The indication, regimen, and timing of adjuvant therapy for breast cancer should be selected according to individual evaluations. At present, a variety of guidelines and consensus have been established to provide clinical reference. To more accurately evaluate the prognosis of breast cancer patients and the benefit of adjuvant therapy, some prognostic prediction models based on gene expression profiles or immunohistochemistry have emerged in recent years,

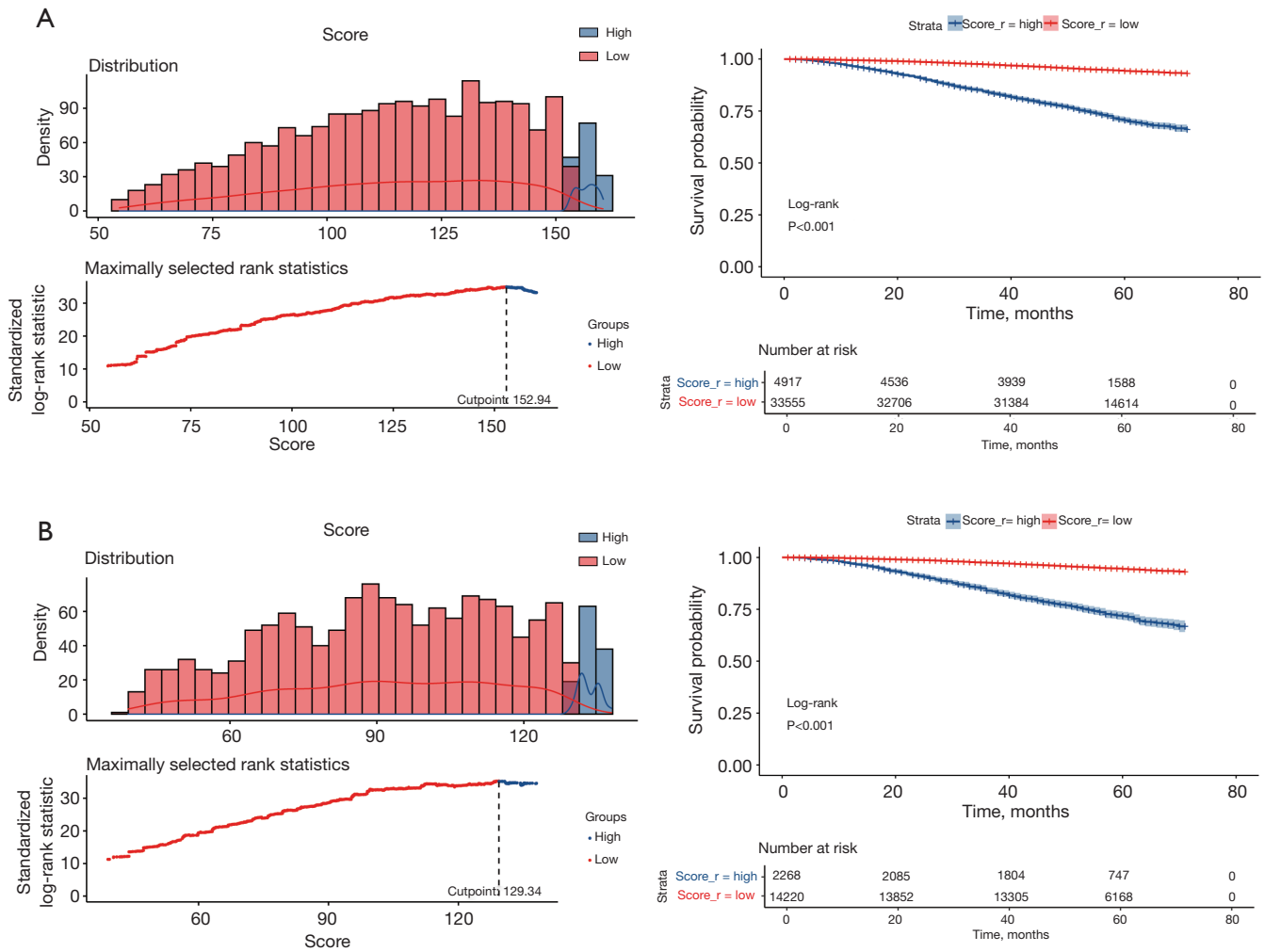

Figure 4 High- and low-risk stratification and Kaplan-Meier curves for the training (A) and validation (B) sets.

such as PAM50 molecular typing (11), the oncotype DX recurrence score system, and the MammaPrint evaluation system (12). These models can help physicians and patients to develop more reasonable adjuvant treatment plans and avoid overtreatment or undertreatment. However, there has been no systematic prognostic study and nomogram construction for breast cancer patients receiving adjuvant therapy. Studies have confirmed some prognostic factors for the survival of breast cancer patients, including age, race, marital status, sex, grade (tumor histological grade), and subtype (13-16). Drawing on clinical practice and the National Comprehensive Cancer Network (NCCN) guidelines (17), this study included all factors accessible to the SEER database in the screening analysis.

Age is an important factor affecting patient prognosis in many cancers, including breast cancer. Previous studies have shown that middle-aged patients have better OS and breast cancer-specific survival than younger and older patients (18,19). This may be related to younger or older women having more aggressive or difficult-to-treat types of breast cancer. This study focused on middle-aged and elderly patients; therefore, detailed stratification was only performed for patients aged 50 to 80 years. The results showed that age stratification had a statistically significant effect on the prognosis of breast cancer patients who received adjuvant therapy, especially those aged 70–79 years (hazard ratio =3.80, 95% CI: 3.31–4.35) and ≥80 years (hazard ratio =8.41, 95% CI: 7.29–9.69). Breast cancer mortality increases with age (20) and increases significantly in women aged >70 years (21). Therefore, for breast cancer

patients aged >70 years, a variety of factors should be comprehensively considered to prolong OS and improve prognosis in those receiving adjuvant therapy, including the characteristics of the tumor, survival expectancy, geriatric assessment results, treatment goals, preferences, and values (22).

The molecular classification of breast cancer based on hormone receptor (HR) and human epidermal growth factor receptor 2 (HER2) status serves as a critical prognostic and therapeutic indicator (23). The most prevalent subtype, $HR^+/HER2^-$, generally forecasts a favorable prognosis; however, it may exhibit endocrine resistance. The $HR^+/HER2^+$ subtype, which is characterized by high proliferation and aggressiveness, responds well to targeted therapies. Conversely, $HR^-/HER2^+$ breast cancer, which relies heavily on the HER2 signaling pathway, shows a robust response to anti-HER2 treatments. Effective targeted treatments are currently lacking for the $HR^-/HER2^-$ subtype, which is noted for its heterogeneity and having the poorest prognosis. Among these subtypes, patients with $HR^+/HER2^+$ have the best prognosis (hazard ratio =0.51, 95% CI: 0.44–0.60), while those with $HR^-/HER2^-$ face the most challenging outcomes. This conclusion is consistent with findings from previous studies of treated (24) and metastatic (25,26) breast cancer patients.

Many previous studies have confirmed that the selection of individualized adjuvant therapy according to different clinical stages, pathological types, molecular markers, and other factors can effectively reduce the risk of recurrence and metastasis, and improve the survival rate and quality of life of breast cancer patients (27). In view of the information accessible via the SEER database, we only selected radiotherapy and chemotherapy for the analysis and construction of the prognostic models.

As early as the 1980s, adjuvant chemotherapy was found to have positive effects on the survival of breast cancer patients (28), but it has always been difficult to identify those who might benefit from this therapy, which has limited clinical benefits, especially in terms of long-term survival. In this study, adjuvant radiotherapy demonstrated a significant improvement in prognosis (hazard ratio =0.64, 95% CI: 0.60–0.69); however, adjuvant chemotherapy did not significantly alter the prognosis of the patients. This allowed adjuvant chemotherapy to be excluded from the final prediction model construction, along with the primary site and location of the tumor, after the LASSO regression and cross-validation analyses.

To help guide decisions about adjuvant chemotherapy and further improve the clinical benefits, several multi-gene detection tools have been developed and used in clinical practice. The results of multi-gene detection [such as 21 genes (29,30) or 70 genes (31,32)] can be combined with the patient's age and menopausal status to make a comprehensive decision about the indications of adjuvant chemotherapy.

With the development of medical technology, the treatment of cancer will inevitably move towards individualization and precision. In adjuvant targeted therapy for breast cancer, the most appropriate drugs and regimens can be selected based on different molecular markers and patient characteristics to avoid overtreatment or undertreatment. It can also be combined with other treatment methods (e.g., chemotherapy, radiotherapy, and endocrine therapy) to enhance the effect of comprehensive treatment and overcome the drug resistance or inefficiency that may arise from single treatments.

T-DXd, the "star drug" in targeted therapy, is a novel antibody-conjugated drug targeting the HER2 receptor, which has the characteristics of high efficiency, broad spectrum, and penetration of the blood-brain barrier. It can also target some cancer cells that are resistant or ineffective to other HER2-targeting drugs. At present, in the latest Destiny-breast04 study (33), unprecedented clinical benefits have also been shown for people with low HER2 expression. The subsequent Destiny-breast05 trial (NCT04622319) aims to compare the efficacy and safety of T-DXd and T-DM1 as adjuvant therapy for patients with HER2-positive early breast cancer, which is expected to significantly improve the long-term prognosis of patients with breast cancer after surgery.

We developed a prognostic nomogram based on multiple clinical and biological factors, including age, sex, race, marital status, grade, T stage, N stage, subtype, and radiotherapy. Notably, LASSO regression and cross-validation analyses were used to screen the prognostic factors. Compared with traditional screening methods, these methods have obvious advantages in feature selection, model generalization ability evaluation, and model selection, which can help improve the performance and interpretation ability of prediction models. To the best of our knowledge, this is a very rare nomogram for this population, and this model showed good prognostic performance for breast cancer patients receiving adjuvant therapy. In this model, the C-index values of the nomogram prediction model in the training set and validation set were 0.768 and 0.763, respectively. The AUCs of the ROC curves at 1, 3, and

5 years were 0.785, 0.793, and 0.775, respectively, in the training set, and 0.798, 0.781, and 0.767, respectively, in the validation set, indicating that the model had good predictive ability. The calibration curves showed that the actual probabilities of the 1-, 3-, and 5-year OS of breast cancer patients treated with adjuvant therapy were closely aligned with the predicted probabilities. This prediction model could help clinicians to identify breast cancer patients with a poor prognosis in adjuvant therapy. Through close postoperative monitoring, individualized adjuvant therapy, and the timely adjustment of treatments as necessary, the quality of life of these patients could be improved.

The limitations of this study are its retrospective design and lack of external validation. Due to the limitations of the SEER database, our study lacked any evaluation of ultrasound image features, the Ki67 index, gene mutations, and other factors that may affect prognosis (e.g., obesity, alcohol consumption, and smoking). Therefore, there are still some areas for future studies to be improved, such as expanding the sample size and scope, increasing external validation, optimizing the design and application of the nomogram model, and exploring other prognostic factors and mechanisms.

Conclusions

In short, being aged ≥80 years, male, black, and single, and having a higher grade (III), higher T stage (T4), higher N stage (N3), and $HR^-/HER2^-$, and not receiving radiotherapy were associated with a poor prognosis in breast cancer patients treated with adjuvant therapy. Being aged ≥80 years was the most significant prognostic factor. Based on data from the SEER database, we successfully constructed a prognostic prediction nomogram for breast cancer patients who received adjuvant therapy, and the nomogram was shown to have a good ability to evaluate the 1-, 3-, and 5-year OS of the patients. This model is of great significance, as it may assist clinicians to identify breast cancer patients with a poor prognosis after adjuvant therapy in a timely manner and make further clinical decisions.

Acknowledgments

Funding: None.

Footnote

Reporting Checklist: The authors have completed the TRIPOD reporting checklist. Available at https://gs.amegroups.com/article/view/10.21037/gs-23-537/rc

Peer Review File: Available at https://gs.amegroups.com/article/view/10.21037/gs-23-537/prf

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://gs.amegroups.com/article/view/10.21037/gs-23-537/coif). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

References

1. Siegel RL, Giaquinto AN, Jemal A. Cancer statistics, 2024. CA Cancer J Clin 2024;74:12-49.
2. Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin 2021;71:209-49.
3. National Health Commission of the People's Republic of China. China Health and Wellness Statistical Yearbook; 2022.
4. Fillon M. Breast cancer recurrence risk can remain for 10 to 32 years. CA Cancer J Clin 2022;72:197-9.
5. Radiotherapy to regional nodes in early breast cancer: an individual patient data meta-analysis of 14 324 women in 16 trials. Lancet 2023;402:1991-2003.
6. Early Breast Cancer Trialists' Collaborative Group (EBCTCG). Electronic address: bc.overview@ctsu.ox.ac; . Anthracycline-containing and taxane-containing

chemotherapy for early-stage operable breast cancer: a patient-level meta-analysis of 100 000 women from 86 randomised trials. Lancet 2023;401:1277-92.
7. Pan H, Gray R, Braybrooke J, et al. 20-Year Risks of Breast-Cancer Recurrence after Stopping Endocrine Therapy at 5 Years. N Engl J Med 2017;377:1836-46.
8. Li Y, Zhang H, Merkher Y, et al. Recent advances in therapeutic strategies for triple-negative breast cancer. J Hematol Oncol 2022;15:121.
9. Che WQ, Li YJ, Tsang CK, et al. How to use the Surveillance, Epidemiology, and End Results (SEER) data: research design and methodology. Mil Med Res 2023;10:50.
10. West HJ, Jin JO. Adjuvant Therapy. JAMA Oncol 2015;1:698.
11. Lænkholm AV, Jensen MB, Eriksen JO, et al. PAM50 Risk of Recurrence Score Predicts 10-Year Distant Recurrence in a Comprehensive Danish Cohort of Postmenopausal Women Allocated to 5 Years of Endocrine Therapy for Hormone Receptor-Positive Early Breast Cancer. J Clin Oncol 2018;36:735-40.
12. Andre F, Ismaila N, Allison KH, et al. Biomarkers for Adjuvant Endocrine and Chemotherapy in Early-Stage Breast Cancer: ASCO Guideline Update. J Clin Oncol 2022;40:1816-37.
13. Adami HO, Malker B, Holmberg L, et al. The relation between survival and age at diagnosis in breast cancer. N Engl J Med 1986;315:559-63.
14. Carey LA, Perou CM, Livasy CA, et al. Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. JAMA 2006;295:2492-502.
15. Zhu S, Lei C. Association between marital status and all-cause mortality of patients with metastatic breast cancer: a population-based study. Sci Rep 2023;13:9067.
16. Rakha EA, El-Sayed ME, Lee AH, et al. Prognostic significance of Nottingham histologic grade in invasive breast carcinoma. J Clin Oncol 2008;26:3153-8.
17. NCCN. The NCCN breast cancer clinical practice guidelines in oncology. Version 2.2024.
18. Chen HL, Zhou MQ, Tian W, et al. Effect of Age on Breast Cancer Patient Prognoses: A Population-Based Study Using the SEER 18 Database. PLoS One 2016;11:e0165409.
19. Brandt J, Garne JP, Tengrup I, et al. Age at diagnosis in relation to survival following breast cancer: a cohort study. World J Surg Oncol 2015;13:33.
20. Lei S, Zheng R, Zhang S, et al. Global patterns of breast cancer incidence and mortality: A population-based cancer registry data analysis from 2000 to 2020. Cancer Commun (Lond) 2021;41:1183-94.
21. Lima SM, Kehm RD, Terry MB. Global breast cancer incidence and mortality trends by region, age-groups, and fertility patterns. EClinicalMedicine 2021;38:100985.
22. Janeva S, Zhang C, Kovács A, et al. Adjuvant chemotherapy and survival in women aged 70 years and older with triple-negative breast cancer: a Swedish population-based propensity score-matched analysis. Lancet Healthy Longev 2020;1:e117-24.
23. Zhang X. Molecular Classification of Breast Cancer: Relevance and Challenges. Arch Pathol Lab Med 2023;147:46-51.
24. Yang SX, Polley EC. Systemic treatment and radiotherapy, breast cancer subtypes, and survival after long-term clinical follow-up. Breast Cancer Res Treat 2019;175:287-95.
25. Gong Y, Liu YR, Ji P, et al. Impact of molecular subtypes on metastatic breast cancer patients: a SEER population-based study. Sci Rep 2017;7:45411.
26. Hou L, Qiu M, Chen M, et al. The association between molecular type and prognosis of patients with stage IV breast cancer: an observational study based on SEER database. Gland Surg 2021;10:1889-98.
27. Sonnenblick A, Piccart M. Adjuvant systemic therapy in breast cancer: quo vadis? Ann Oncol 2015;26:1629-34.
28. Bonadonna G, Brusamolino E, Valagussa P, et al. Combination chemotherapy as an adjuvant treatment in operable breast cancer. N Engl J Med 1976;294:405-10.
29. Sparano JA, Gray RJ, Makower DF, et al. Adjuvant Chemotherapy Guided by a 21-Gene Expression Assay in Breast Cancer. N Engl J Med 2018;379:111-21.
30. Kalinsky K, Barlow WE, Gralow JR, et al. 21-Gene Assay to Inform Chemotherapy Benefit in Node-Positive Breast Cancer. N Engl J Med 2021;385:2336-47.
31. Piccart M, van 't Veer LJ, Poncet C, et al. 70-gene signature as an aid for treatment decisions in early breast cancer: updated results of the phase 3 randomised MINDACT trial with an exploratory analysis by age. Lancet Oncol 2021;22:476-88.
32. Lopes Cardozo JMN, Drukker CA, Rutgers EJT, et al. Outcome of Patients With an Ultralow-Risk 70-Gene Signature in the MINDACT Trial. J Clin Oncol

2022;40:1335-45.
33. Modi S, Jacot W, Yamashita T, et al. Trastuzumab Deruxtecan in Previously Treated HER2-Low Advanced Breast Cancer. N Engl J Med 2022;387:9-20.

(English Language Editor: L. Huleatt)

Cite this article as: Jiang Q, Liu X, Wu Y, Du J, Rao Y, Li J, Li H. Construction and validation of a nomogram for predicting the prognosis of breast cancer patients who received adjuvant therapy: an analysis based on the SEER database. Gland Surg 2024;13(6):927-941. doi: 10.21037/gs-23-537