CANCER INNOVATION

ORIGINAL ARTICLE

Prognostic nomograms for young breast cancer: A retrospective study based on the SEER and METABRIC databases

Yongxin Li[1] | Xinlong Tao[1] | Yinyin Ye[1] | Yuyao Tang[1] | Zhengbo Xu[2] | Yaming Tian[3] | Zhen Liu[1] | Jiuda Zhao[1] (ORCID)

[1]Breast Disease Diagnosis and Treatment Center of Affiliated Hospital of Qinghai University & Affiliated Cancer Hospital of Qinghai University, Xining, Qinghai, China

[2]Qinghai University, Xining, Qinghai, China

[3]Department of Imaging, Affiliated Hospital of Qinghai University, Xining, Qinghai, China

Correspondence
Zhen Liu and Jiuda Zhao, Breast Disease Diagnosis and Treatment Center of Affiliated Hospital of Qinghai University & Affiliated Cancer Hospital of Qinghai University, Xining 810000, Qinghai, China.
Email: qhdxlz@163.com and jiudazhao@126.com

Abstract

Background: Young breast cancer (YBC) is a subset of breast cancer that is often more aggressive, but less is known about its prognosis. In this study, we aimed to generate nomograms to predict the overall survival (OS) and breast cancer-specific survival (BCSS) of YBC patients.

Methods: Data of women diagnosed with YBC between 2010 and 2020 were obtained from the Surveillance, Epidemiology, and End Results (SEER) database. The patients were randomly allocated into a training cohort ($n = 15,227$) and internal validation cohort ($n = 6,526$) at a 7:3 ratio. With the Cox regression models, significant prognostic factors were identified and used to construct 3-, 5-, and 10-year nomograms of OS and BCSS. Data from the Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) database were used as an external validation cohort ($n = 90$).

Results: We constructed nomograms incorporating 10 prognostic factors for OS and BCSS. These nomograms demonstrated strong predictive accuracy for OS and BCSS in the training cohort, with C-indexes of 0.806 and 0.813, respectively. The calibration curves verified that the nomograms have good prediction accuracy. Decision curve analysis demonstrated their practical clinical value for predicting YBC patient survival rates. Additionally, we provided dynamic nomograms to improve the operability of the results. The risk stratification ability assessment also showed that the OS

Abbreviations: AJCC, American Joint Committee on Cancer; ANOVA, analysis of variance; BC, breast cancer; BCSS, breast cancer–specific survival; CI, confidence interval; C-index, concordance index; DCA, decision curve analysis; ER, estrogen receptor; HER2, human epidermal growth factor receptor 2; HR, hazard ratio; IDC, invasive ductal carcinoma; ILC, invasive lobular carcinoma; METABRIC, Molecular Taxonomy of Breast Cancer International Consortium; OS, overall survival; PR, progesterone receptor; ROC, receiver operating characteristic; SD, standard deviation; SEER, the Surveillance, Epidemiology, and End Results; TNM, tumor-node-metastasis; YBC, young breast cancer.

Yongxin Li, Xinlong Tao, and Yinyin Ye contributed equally to this work and shared the co-first authorship.

Zhen Liu and Jiuda Zhao shared the co-corresponding authorship.

---

and BCSS rates of the low-risk group were significantly better than those of the high-risk group.

Conclusions: Here, we generated and validated more comprehensive and accurate OS and BCSS nomograms than models previously developed for YBC. These nomograms can help clinicians evaluate patient prognosis and make clinical decisions.

KEYWORDS

breast cancer-specific survival, nomogram, overall survival, prognostic model, young breast cancer

1 | INTRODUCTION

Breast cancer (BC) ranks as the second most prevalent cancer type and fourth leading cause of cancer-related deaths worldwide [1]. Many factors affect BC patient prognosis, including pathological differentiation, stage, histological type, treatment, and age [2–5]. Additionally, young age is an independent adverse factor for BC patient survival, with younger patients usually having worse prognoses than older patients [6, 7]. Young breast cancer (YBC) is generally defined as BC in an individual less than 40 years old [7–9]. Although YBC is less common, it is characterized by more aggressive behavior and a poorer prognosis [6, 10, 11].

Numerous factors contribute to the relatively poor prognosis of YBC patients, while many questions remain unanswered. Two studies have developed prognostic models for YBC. However, these models exhibit certain limitations, such as a small number of predicted variables, short forecasted lifespan, single endpoint, and lack of dynamic nomograms [12, 13]. Furthermore, the low incidence of YBC results in fewer patients being included in most studies, limiting the precision and comprehensiveness of prognosis and treatment research for these individuals. The American Joint Committee on Cancer (AJCC) tumor-node-metastasis (TNM) staging system is commonly used to evaluate BC prognosis. However, because of the specific challenges associated with YBC, this system may not provide a complete and thorough prognosis for this subset of cancer patients. Thus, there is a critical need to develop more accurate and comprehensive prognostic models that can effectively guide clinicians to make informed decisions regarding YBC treatment.

In this study, we generated and validated a nomogram to predict YBC patient prognosis using data from the Surveillance, Epidemiology, and End Results (SEER) and Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) databases. Nomograms are extensively used for estimating the probability of death or recurrence for individual patients by integrating key prognostic factors [4, 14]. This prognostic model potentially possesses strong risk stratification capabilities, which are essential for predicting YBC patient survival and guiding the selection of optimal treatments to extend their lives. In addition, we created dynamic nomograms to facilitate personalized prognosis prediction for YBC patients.

2 | MATERIALS AND METHODS

2.1 | Patients

The study data were sourced from the SEER and METABRIC databases. The SEER database represents the largest publicly accessible repository of cancer patient information in the United States. SEER*Stat version 8.4.3, created by the National Cancer Institute, was used for data extraction [15]. We collected data from patients diagnosed with YBC from 2010 to 2020. The inclusion criteria were as follows: (1) female patients who were initially diagnosed with primary BC; (2) under the age of 40; (3) detailed information was available for TNM stage, histological typing, and pathological differentiation; (4) the estrogen receptor (ER) status, progesterone receptor (PR) status, and human epidermal growth factor receptor 2 (HER2) status were determined; and (5) details about surgery, chemotherapy, and radiation treatments were available. The exclusion criteria were as follows: (1) patients who were lost to follow-up or were followed for less than 1 month and (2) patients diagnosed by autopsy or death certificate. We then randomly assigned eligible patients to the training cohort or internal validation cohort at a 7:3 ratio to create and validate the nomogram, respectively.

To further validate the accuracy of the nomogram, patients diagnosed with YBC in the METABRIC database were assembled for external validation. Participants in this validation group were selected using the same inclusion and exclusion criteria as those in the initial training cohort.

2.2 | Variables

The variables examined in this study included age at diagnosis (<40 years), race (black, white, other), laterality (left, right, other), pathological differentiation (grade I, II, III/IV), TNM stage (I, II, III, IV), histological type (IDC, IDL, other), ER status, PR status, HER2 status, and treatment-related information (surgery, radiotherapy, and chemotherapy). Overall survival (OS) was the primary endpoint of the study, while breast cancer-specific survival (BCSS) was the secondary endpoint. OS was defined as the duration from the initial diagnosis of BC to death from any cause. BCSS was defined as the duration from the date of diagnosis until death related to BC.

2.3 | Statistical analysis

Descriptive analysis of the patient characteristics in the SEER and METABRIC databases was conducted, with data for continuous variables presented as the mean ± standard deviation (SD). For the assessment of differences between three groups, analysis of variance (ANOVA) was performed. Data for categorical variables were presented as percentages and evaluated using Pearson's Chi-squared test. Cox regression analyses were used to determine the survival variables in YBC patients, expressed as hazard ratios (HRs) with 95% confidence intervals (CIs). Using the Cox regression analysis results, we constructed nomograms to predict the 3-, 5-, and 10- year OS and BCSS probabilities.

To evaluate the predictive performance of the nomogram, we introduced internal and external validation cohorts to test its performance. The concordance index (C-index) and receiver operating characteristic (ROC) curves were used to evaluate the predictive performance and accuracy of the nomogram. Additionally, calibration curves were used to determine the discriminability of the nomogram, while decision curve analysis (DCA) was used to evaluate the clinical utility of the nomogram [16–18]. Risk stratification was performed on the nomogram, which was divided into low- and high-risk groups. Kaplan–Meier curve analysis and log-rank tests were used to examine the prognostic effect of YBC patients.

All statistical analyses and data visualization were performed using R (version 4.3.3). A two-sided p value < 0.05 was considered statistically significant.

3 | RESULTS

3.1 | Patient baseline characteristics

Using the screening criteria, we included 21,753 eligible patients with YBC, including the training cohort ($n = 15,227$) and internal validation cohort ($n = 6526$). A detailed data screening flow chart is provided in Figure 1. In addition, patients with YBC from the
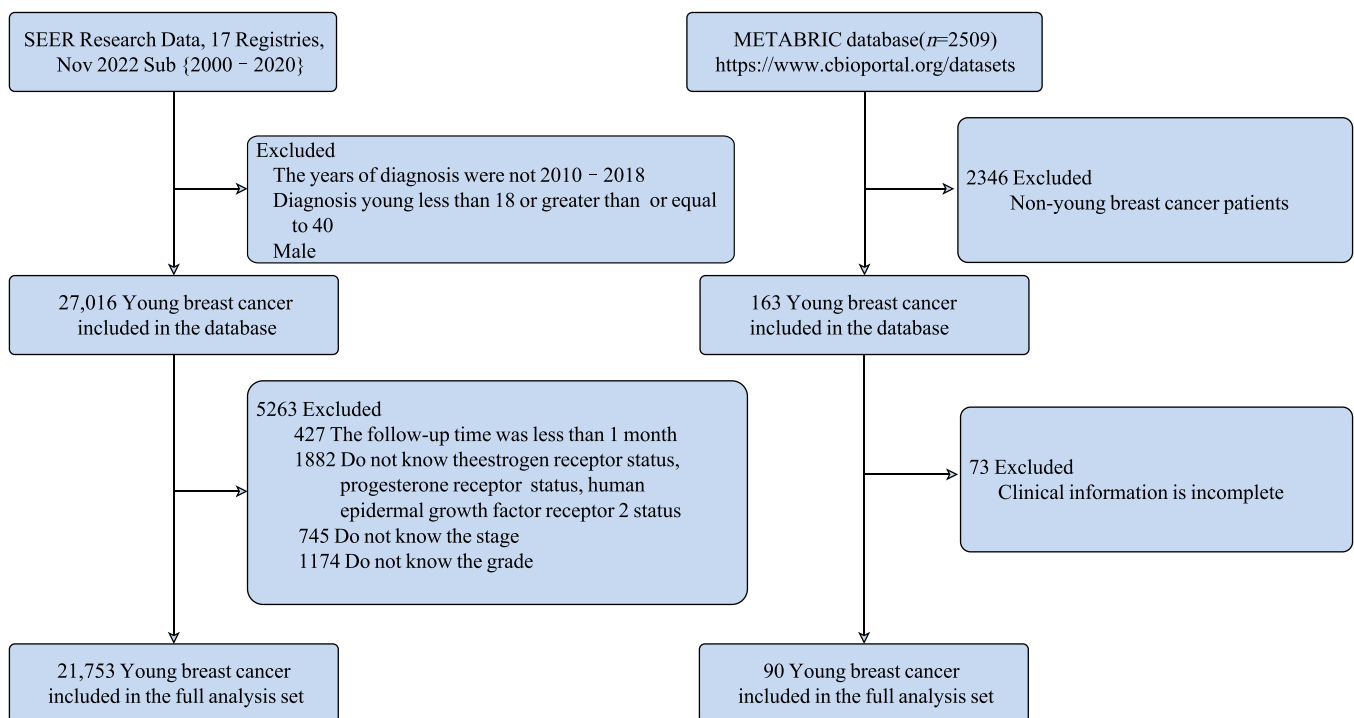

FIGURE 1 The detailed data screening flow chart.

METABRIC database ($n = 90$) served as the external validation cohort. For pathological features, the main pathological type in both databases was invasive ductal carcinoma (SEER database: 85.5%, METABRIC database: 93.3%). There was a significant difference in pathological differentiation ($p < 0.001$), with high-grade tumors being more common in YBC patients (SEER database: 54.8%, METABRIC database: 81.1%). The detailed demographic and psychological characteristics of the patients are shown in Table 1.

TABLE 1 Characteristics of the study cohorts.

Characteristic	Training cohort $n = 15{,}227$	Internal validation cohort $n = 6526$	External validation cohort $n = 90$	p value
Age (years)	34.72 ± 3.81	34.70 ± 3.82	35.46 ± 3.66	0.200[a]
Race				<0.001[b]
Black	2192 (14.4)	939 (14.4)	0 (0.0)	
White[c]	10,688 (70.2)	4645 (71.2)	90 (100.0)	
Others	2347 (15.4)	942 (14.4)	0 (0.0)	
Laterality				<0.001[b]
Left	7570 (49.7)	3280 (50.3)	37 (41.1)	
Right	7646 (50.2)	3245 (49.7)	50 (55.6)	
Others	11 (0.1)	1 (0.0)	3 (3.3)	
Grade				<0.001[b]
I	1259 (8.3)	548 (8.4)	0 (0.0)	
II	5624 (36.9)	2409 (36.9)	17 (18.9)	
III/IV	8344 (54.8)	3569 (54.7)	73 (81.1)	
Histology				0.200[b]
IDC	12,983 (85.3)	5608 (85.9)	84 (93.3)	
ILC	413 (2.7)	166 (2.5)	1 (1.1)	
Others	1831 (12.0)	752 (11.5)	5 (5.6)	
Stage				0.003[b]
I	5236 (34.4)	2227 (34.1)	23 (25.6)	
II	6431 (42.2)	2856 (43.8)	53 (58.9)	
III	2651 (17.4)	1057 (16.2)	14 (15.6)	
IV	909 (6.0)	386 (5.9)	0 (0.0)	
ER status				<0.001[b]
Positive	11,247 (73.9)	4797 (73.5)	36 (40.0)	
Negative	3980 (26.1)	1729 (26.5)	54 (60.0)	
PR status				<0.001[b]
Positive	9876 (64.9)	4205 (64.4)	28 (31.1)	
Negative	5351 (35.1)	2321 (35.6)	62 (68.9)	
HER2 status				0.700[b]
Positive	3687 (24.2)	1595 (24.4)	25 (27.8)	
Negative	11,540 (75.8)	4931 (75.6)	65 (72.2)	
Chemotherapy				0.200[b]
Yes	11,785 (77.4)	5043 (77.3)	62 (68.9)	
No/Unknown	3442 (22.6)	1483 (22.7)	28 (31.1)	

TABLE 1 (Continued)

Characteristic	Training cohort $n = 15{,}227$	Internal validation cohort $n = 6526$	External validation cohort $n = 90$	p value
Radiation				0.001[b]
Yes	8092 (53.1)	3453 (52.9)	65 (72.2)	
No/Unknown	7135 (46.9)	3073 (47.1)	25 (27.8)	
Surgery				0.200[b]
Yes	14,195 (93.2)	6100 (93.5)	88 (97.8)	
No/Unknown	1032 (6.8)	426 (6.5)	2 (2.2)	

Note: Data are presented as mean ± SD, or *n* (%).

Abbreviations: ER, estrogen receptor; HER2, human epidermal growth factor receptor 2; IDC, invasive ductal carcinoma; ILC, invasive lobular carcinoma; PR, progesterone receptor.

[a]One-way ANOVA.

[b]Pearson's Chi-squared.

[c]The default race in external validation is white.

3.2 | Factors associated with OS and BCSS

Following the univariate analysis, the factors we screened were further tested using a multivariate analysis. The multivariate analysis results suggested that race, histological type, TNM stage, pathological differentiation, ER status, PR status, HER2 status, and surgery were all independently confirmed to be correlated with OS and BCSS. In addition, because chemotherapy and radiotherapy can impact patient prognosis [19–22], we included both when constructing the prognostic model. The detailed results of the univariate and multivariate analyses are presented in Table 2.

3.3 | Nomogram construction

From the Cox regression models, we constructed nomograms for predicting the 3-, 5-, and 10-year OS and BCSS in patients with YBC (Figures 2a and 3a). To predict patient OS and BCSS using these nomograms, we assigned scores to each variable with different variables accounting for different scores. By adding the scores of each variable, the total score of each patient could be obtained, which corresponds to their OS and BCSS rates. TNM stage accounted for the most weight in the nomogram, followed by HER2 status, pathological differentiation, and histological type, among other factors.

In addition, we created dynamic nomograms to facilitate a clinician's use and personalized prediction of patient prognosis (Figures 2b and 3b). The 10 predicted values can be selected in the left interface according to different patients, followed by clicking the "Predict" button. This finally leads to the survival probability and 95% CIs being derived in the right interface. Details are available on the dynamic nomogram website for OS (https://1947195299lyx.shinyapps.io/YBC-OS/) and on the dynamic nomogram website for BCSS (https://1947195299lyx.shinyapps.io/DynNomapp/).

3.4 | Calibration and validation of the nomogram

In the training cohort, the C-indexes of the OS and BCSS nomograms were 0.806 and 0.813, respectively. The area under the ROC curve (AUC) values for the OS nomogram at 3-, 5-, and 10-years were 0.854, 0.823, and 0.818, respectively. Similarly, for the BCSS nomogram, the AUC values at these time points were 0.856, 0.827, and 0.823, respectively, as shown in Figure 4. The results of the validation cohorts are detailed in Figure S1. These findings indicated that both prognostic models demonstrate good performance and accuracy.

The calibration curves in the training cohort (Figure 5) and validation cohorts (Figure S2) suggested that the OS and BCSS nomograms have good discriminability. In addition, the DCA curves in the training cohort (Figure 6) showed that both nomograms had a significantly positive net benefit for the risk of death, demonstrating their value in predicting patient survival in real clinical practice. Additionally, the DCA curves in the validation cohorts (Figure S3) showed good predictive power in the medium risk threshold range.

3.5 | Risk stratification ability assessment of the nomogram

Finally, we calculated the median score of the OS (0.87) and BCSS (0.84) nomograms in the training cohort, then

TABLE 2 Univariate and multivariate analysis of OS and BCSS in the training cohort.

Variable	n (%)	OS (univariable) HR (95% CI, p value)	OS (multivariable) HR (95% CI, p value)	BCSS (univariable) HR (95% CI, p value)	BCSS (multivariable) HR (95% CI, p value)
Race					
Black	3131 (14.4)				
White	15,333 (70.5)	0.59 (0.54–0.64, <0.001)	0.73 (0.66–0.79, <0.001)	0.60 (0.54–0.66, <0.001)	0.73 (0.67–0.81, <0.001)
Others	3289 (15.1)	0.45 (0.39–0.51, <0.001)	0.58 (0.51–0.67, <0.001)	0.45 (0.39–0.52, <0.001)	0.59 (0.51–0.68, <0.001)
Laterality					
Left	10,850 (49.9)				
Right	10,891 (50.1)	0.92 (0.86–0.99, 0.032)	0.95 (0.89–1.03, 0.196)	0.94 (0.87–1.02, 0.148)	0.98 (0.91–1.06, 0.587)
Others	12 (0.1)	2.84 (1.18–6.83, 0.020)	0.91 (0.38–2.21, 0.842)	2.73 (1.02–7.28, 0.045)	0.82 (0.31–2.20, 0.696)
Histology					
IDC	18,591 (85.5)				
ILC	579 (2.7)	1.27 (1.03–1.57, 0.024)	1.38 (1.12–1.71, 0.003)	1.31 (1.05–1.63, 0.015)	1.45 (1.16–1.82, 0.001)
Others	2583 (11.9)	0.92 (0.82–1.03, 0.160)	0.95 (0.85–1.07, 0.391)	0.95 (0.84–1.07, 0.359)	0.97 (0.86–1.10, 0.638)
Stage					
I	7463 (34.3)				
II	9287 (42.7)	2.64 (2.29–3.05, <0.001)	2.16 (1.86–2.51, <0.001)	2.90 (2.47–3.41, <0.001)	2.34 (1.98–2.76, <0.001)
III	3708 (17.0)	7.88 (6.83–9.10, <0.001)	6.81 (5.86–7.93, <0.001)	9.21 (7.84–10.82, <0.001)	7.85 (6.63–9.30, <0.001)
IV	1295 (6.0)	25.29 (21.82–29.32, <0.001)	19.21 (16.25–22.72, <0.001)	30.20 (25.61–35.61, <0.001)	23.26 (19.33–27.98, <0.001)
Grade					
I	1807 (8.3)				
II	8033 (36.9)	2.38 (1.90–2.99, <0.001)	1.61 (1.28–2.03, <0.001)	2.53 (1.97–3.25, <0.001)	1.70 (1.31–2.19, <0.001)
III/IV	11,913 (54.8)	3.87 (3.10–4.84, <0.001)	2.15 (1.70–2.71, <0.001)	4.22 (3.30–5.39, <0.001)	2.35 (1.82–3.03, <0.001)
ER status					
Negative	5709 (26.2)				
Positive	16,044 (73.8)	0.54 (0.50–0.58, <0.001)	0.69 (0.62–0.77, <0.001)	0.54 (0.50–0.59, <0.001)	0.70 (0.62–0.79, <0.001)

TABLE 2 (Continued)

Variable	n (%)	OS (univariable) HR (95% CI, p value)	OS (multivariable) HR (95% CI, p value)	BCSS (univariable) HR (95% CI, p value)	BCSS (multivariable) HR (95% CI, p value)
PR status					
Negative	7672 (35.3)				
Positive	14,081 (64.7)	0.55 (0.51–0.60, <0.001)	0.78 (0.70–0.87, <0.001)	0.56 (0.51–0.60, <0.001)	0.78 (0.69–0.88, <0.001)
HER2 status					
Negative	16,471 (75.7)				
Positive	5282 (24.3)	0.70 (0.64–0.76, <0.001)	0.46 (0.42–0.50, <0.001)	0.67 (0.61–0.74, <0.001)	0.43 (0.39–0.48, <0.001)
Chemotherapy					
No/Unknown	4925 (22.6)				
Yes	16,828 (77.4)	1.70 (1.52–1.89, <0.001)	0.92 (0.82–1.03, 0.146)	1.78 (1.59–2.00, <0.001)	0.93 (0.82–1.05, 0.256)
Radiation					
No/Unknown	10,208 (46.9)				
Yes	11,545 (53.1)	1.07 (1.00–1.15, 0.061)	1.01 (0.93–1.09, 0.887)	1.11 (1.03–1.20, 0.007)	1.03 (0.94–1.12, 0.502)
Surgery					
No/Unknown	1458 (6.7)				
Yes	20,295 (93.3)	0.18 (0.17–0.20, <0.001)	0.50 (0.45–0.56, <0.001)	0.18 (0.16–0.20, <0.001)	0.51 (0.45–0.57, <0.001)

Abbreviations: BCSS, breast cancer-specific survival; CI, confidence interval; ER, estrogen receptor; HR, hazard ratio; HER2, human epidermal growth factor receptor 2; IDC, invasive ductal carcinoma; ILC, invasive lobular carcinoma; OS, overall survival; PR, progesterone receptor.
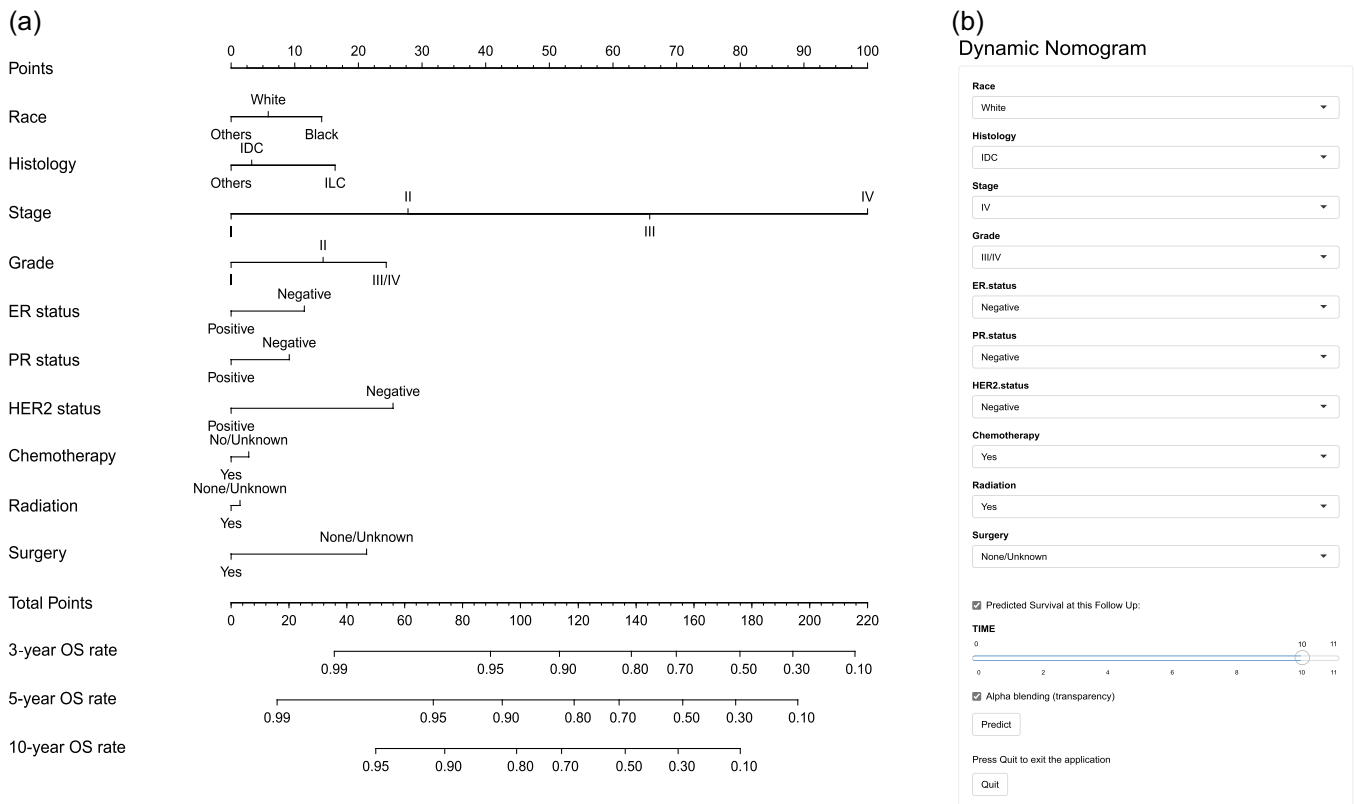
FIGURE 2 Nomogram for predicting OS in patients with young breast cancer. (a) Nomogram for predicting 3-, 5-, and 10-year OS in patients with young breast cancer. (b) Dynamic nomogram for predicting OS in patients with young breast cancer. ER, estrogen receptor; HER2, human epidermal growth factor receptor 2; IDC, invasive ductal carcinoma; ILC, invasive lobular carcinoma; OS, overall survival; PR, progesterone receptor.

evenly divided the patients into low-risk and high-risk groups relative to this median score. From the constructed Kaplan–Meier survival curves, we observed a significant difference between the low-risk and high-risk groups. In both the training and validation cohorts, the low-risk groups for OS (Figure 7a–c) and BCSS (Figure 7d–f) had better outcomes than the corresponding high-risk groups.

4 | DISCUSSION

YBC is a distinct subgroup within BC. Because of its relatively low incidence, there is a paucity of research and established treatment protocols for YBC [23, 24]. Therefore, our study evaluated YBC patient prognosis by constructing and validating 3-, 5-, and 10-year nomograms of OS and BCSS using data from the SEER and METABRIC databases. Both the C-indexes and AUC values demonstrate the good performance of our prognostic model. The calibration curves indicate an agreement between the predicted probability and actual observed probability. Furthermore, the DCA curves demonstrate the practical clinical value of this model for predicting patient survival. The risk stratification results

showed that the OS and BCSS rates in the low-risk group were significantly higher than those in the high-risk group. In addition, we constructed dynamic nomograms to assist clinicians with understanding patient prognosis and make informed and personalized treatment decisions. In the Results section, we also presented free online links to these dynamic nomograms for the convenience of clinicians.

Here, we identified race, histological type, TNM stage, pathological differentiation, ER status, PR status, HER2 status, and surgery as independent prognostic factors for OS and BCSS using the Cox regression model. Previous studies have also shown correlations between these factors and BC prognosis [25–29]. Because chemotherapy and radiotherapy have been shown in clinical practice and related studies to directly impact prognosis, we also included both treatment methods when constructing the nomograms [19–22]. Although the inclusion of these two variables in the nomograms had a small contribution, it helped make our nomogram predictions more detailed. TNM stage was given the most weight in the OS and BCSS nomograms. The poor prognosis of YBC patients may be caused by a more advanced tumor stage, more aggressive biology, and
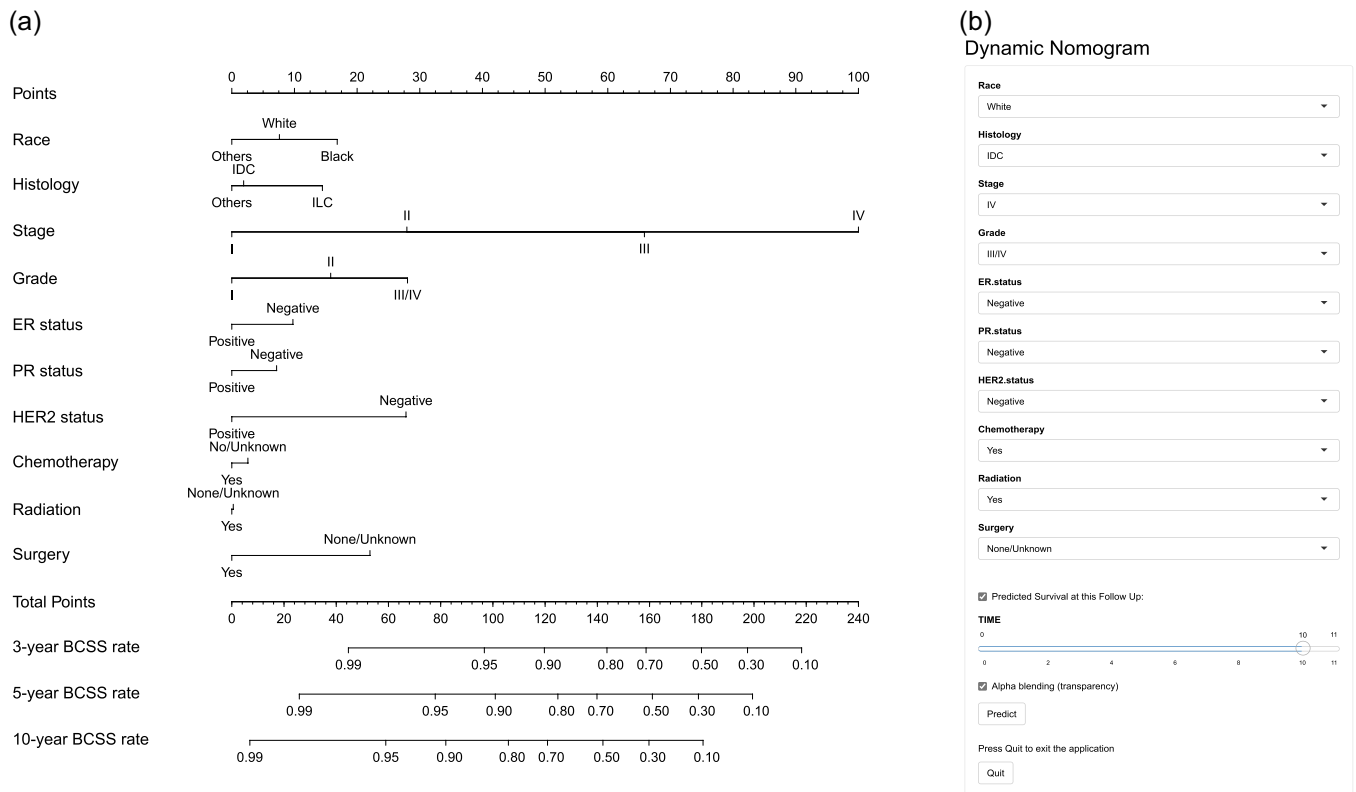
FIGURE 3 Nomogram for predicting BCSS in patients with young breast cancer. (a) Nomogram for predicting 3-, 5-, and 10-year BCSS in patients with young breast cancer. (b) Dynamic nomogram for predicting BCSS in patients with young breast cancer. BCSS, breast cancer-specific survival; ER, estrogen receptor; HER2, human epidermal growth factor receptor 2; IDC, invasive ductal carcinoma; ILC, invasive lobular carcinoma; PR, progesterone receptor.
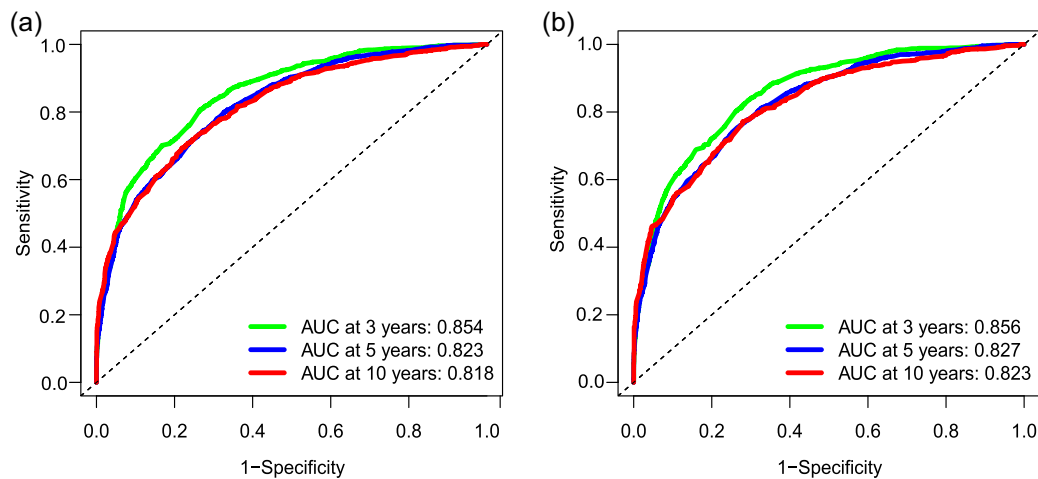

FIGURE 4 The ROC curves reflected the predictive performance of nomograms in patients with young breast cancer. (a) ROC curves for 3-, 5-, and 10-year OS of patients in the training cohort. (b) ROC curves for 3-, 5-, and 10-year BCSS of patients in the training cohort. BCSS, breast cancer-specific survival; OS, overall survival; ROC, receiver operating characteristic.

poorer genomic characteristics [30–33], which also confirms the importance of the disease stage in the nomograms. Despite the TNM staging system being widely used in clinical practice for BC, it has certain limitations for evaluating the condition and prognosis of patients. This system needs to be combined with other factors when conducting a comprehensive evaluation to better guide clinical treatment approaches. Therefore, our prognostic models developed in this study compensate for the shortcomings of the TNM staging system and
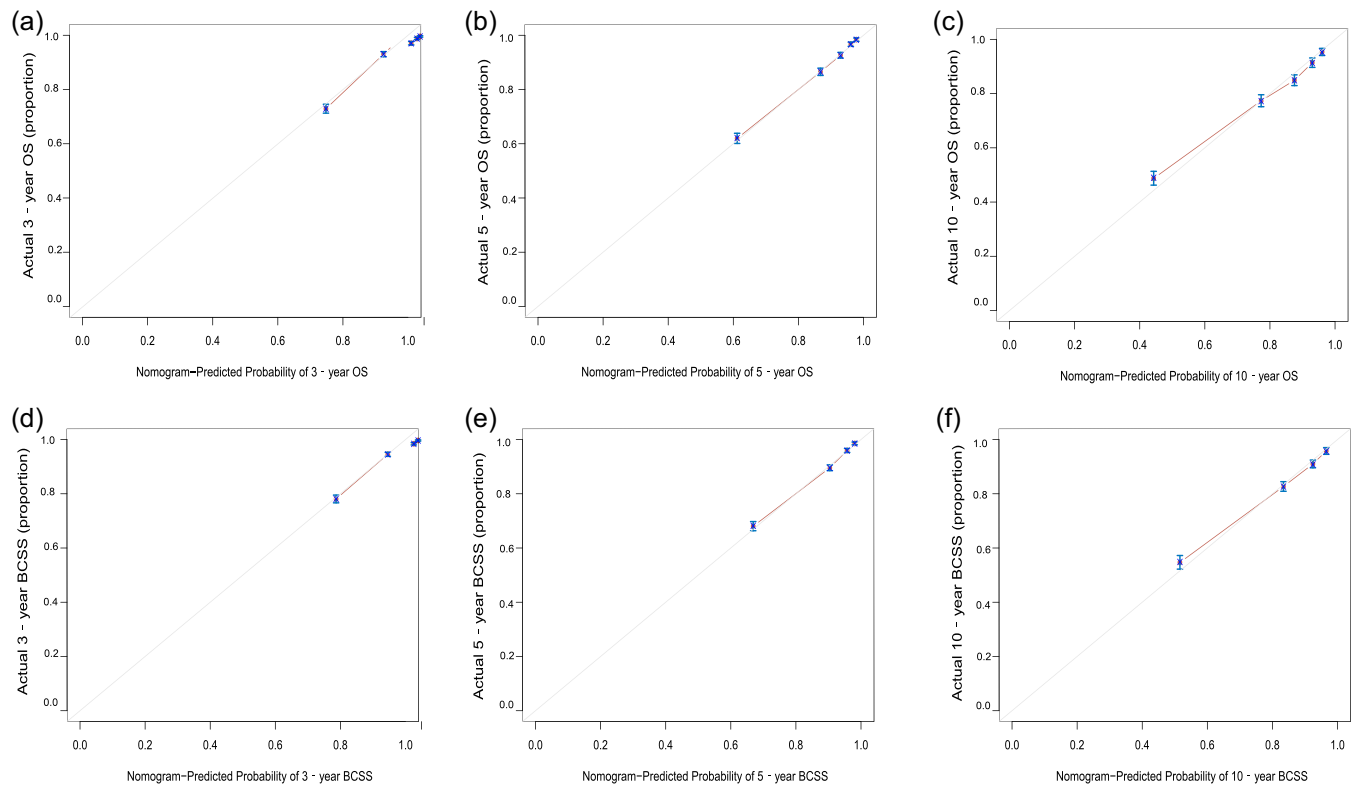
FIGURE 5 The calibration curves of the nomograms for predicting OS and BCSS in the training cohort. (a–c) Calibration curves for 3-, 5-, and 10-year OS of patients. (d–f) Calibration curves for 3-, 5-, and 10-year BCSS of patients. BCSS, breast cancer-specific survival; OS, overall survival.
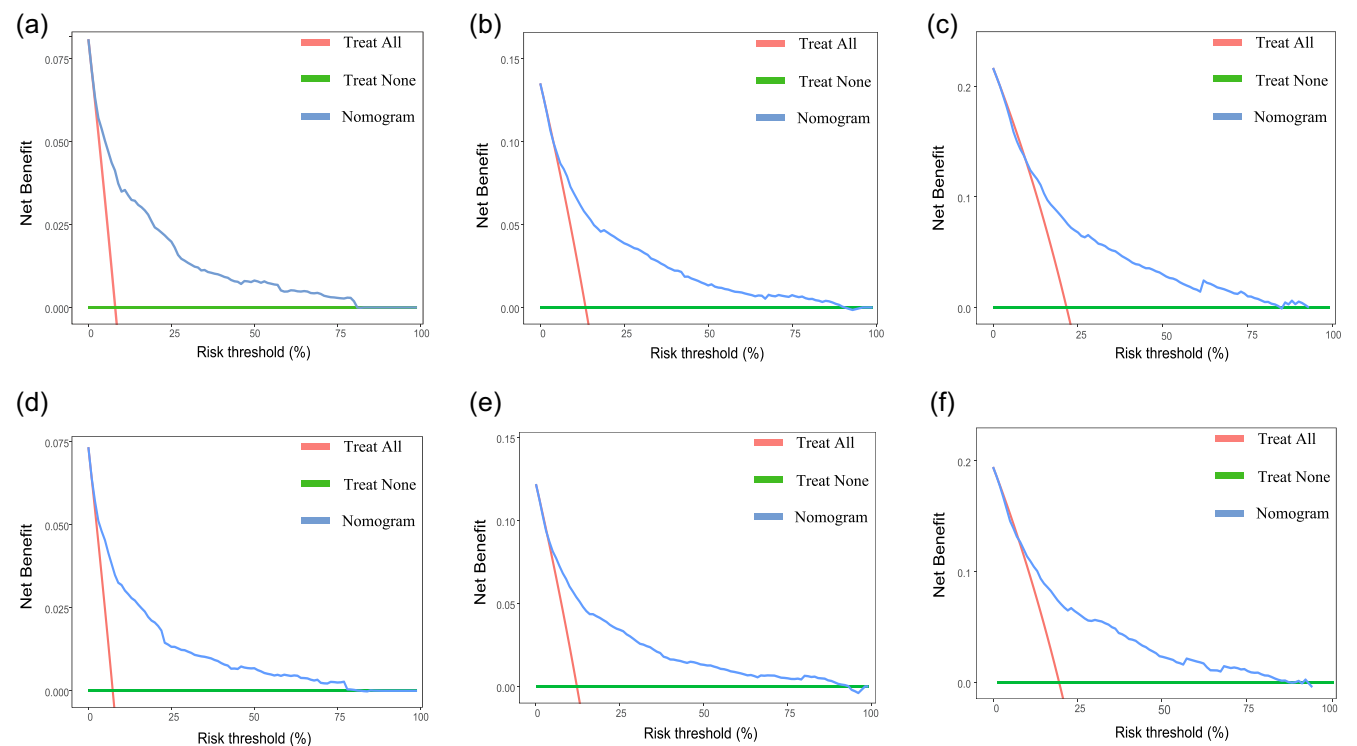

FIGURE 6 The DCA curves of the nomograms for predicting OS and BCSS in the training cohort. (a–c) DCA curves for 3-, 5-, and 10-year OS of patients. (d–f) DCA curves for 3-, 5-, and 10-year BCSS of patients. BCSS, breast cancer-specific survival; DCA, decision curve analysis; OS, overall survival.
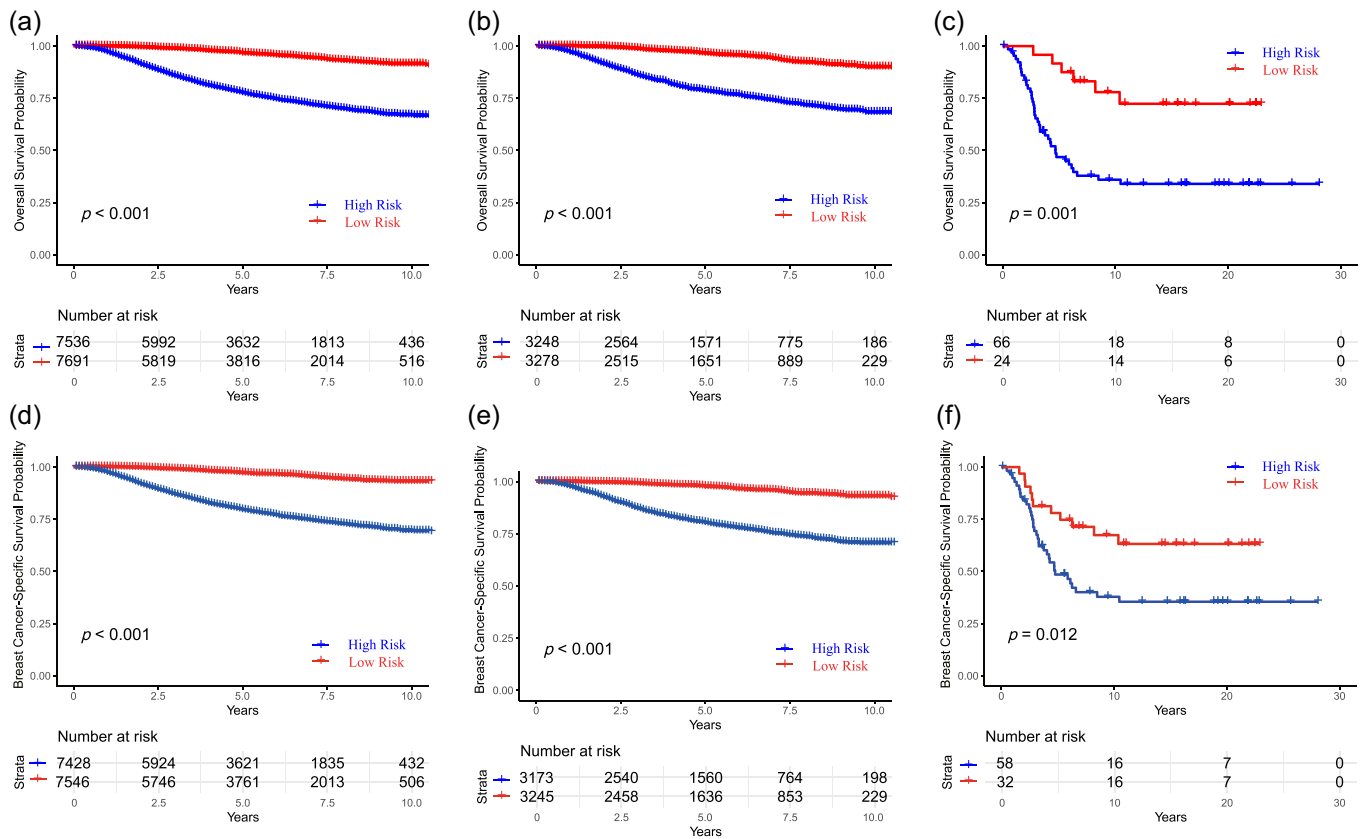
FIGURE 7 Kaplan-Meier curves of OS and BCSS for risk stratification in the training cohort (a, d), the internal validation cohort (b, e), and the external validation cohort (c, f). BCSS, breast cancer-specific survival; OS, overall survival.

combine multiple independent factors to provide a more accurate assessment of YBC patient prognosis.

To the best of our knowledge, two studies have constructed prognostic models for YBC. However, their articles have some shortcomings [12, 13]. In one study, Huang et al. excluded YBC cases with distant metastases and did not predict the prognosis of patients with advanced YBC [12]. Additionally, they only modeled OS in YBC patients. Because such patients often have strong tolerance and are in good basic physical condition, OS may not be the best prognostic indicator for this disease. Second, the short-term survival rate of YBC patients is higher [24], making the prediction of 3- and 5-year OS rates not comprehensive enough. In addition, the authors used their own central data as an external validation cohort. The cohort data from a single center had little heterogeneity and was not widely validated in YBC. Another study by Gong et al. only used the SEER database for modeling and internal verification, with the general performance of the model not being evaluated. Moreover, their study developed and validated the nomogram on a 1:1 scale, which resulted in a relatively small training cohort and failed to adequately account for the weight of influencing factors. Importantly,

both studies only constructed nomograms, not dynamic nomograms. This did not facilitate the assessment of individual patients in clinical practice.

Our study has several advantages. First, we generated nomograms for OS and BCSS with good performance. We also used the calibration curves, ROC curves, and DCA curves to verify different aspects of the nomograms, with all showing good results. Second, because of the better short-term survival of YBC patients, we forecasted the 3-, 5-, and 10- year OS and BCSS rates, which is more comprehensive. Third, we used the METABRIC database for external validation, which still showed a good generalization ability with highly heterogeneous data. Fourth, our prognostic model includes more comprehensive variables and still has good predictive power and clinical utility. Finally, we constructed dynamic nomograms, which are convenient for clinicians to help predict the prognosis of different patients and formulate appropriate treatment strategies.

However, our study also has some limitations. First, we selected the external validation cohort data from Canada and the United Kingdom, making the race default white. Because race is an independent prognostic factor for YBC, the accuracy of the validation may be affected. Second, our

study relied on the SEER and METABRIC databases, which introduces potential selection bias. Cases with incomplete data were excluded, possibly skewing the results. Third, numerous previous studies have shown that many other factors can affect YBC prognosis, such as a higher enrichment of BRCA1/2 and high expression levels of Ki-67 and p53 [34–36]. The absence of the Ki-67 index, BRCA1/2, detailed treatment strategy, and other factors that affect BC in the two databases may lead to a nomogram with a reduced predictive power. In addition, our data were all obtained from online databases, which require further validation in clinical applications. Prospective studies are needed to further test the accuracy of the models. Lastly, we used the TNM staging system, rather than examining T stage, N stage, and M stage individually.

5 | CONCLUSION

The OS and BCSS nomograms generated in this study have good predictive performance and clinical utility for YBC. These nomograms can help guide clinicians when tailoring treatment strategies for individual risk profiles, potentially improving YBC patient outcomes. Future research should explore the integration of these models into routine clinical practice.

AUTHOR CONTRIBUTIONS

Yongxin Li: Conceptualization (equal); data curation (equal); formal analysis (equal); investigation (equal); software (lead). **Xinlong Tao**: Conceptualization (equal); data curation (equal); formal analysis (equal); supervision (equal); validation (equal); writing—original draft (lead). **Yinyin Ye**: Data curation (lead); formal analysis (lead); supervision (equal); visualization (equal). **Yuyao Tang**: Formal analysis (equal); resources (equal); supervision (equal). **Zhengbo Xu**: software (equal); supervision (equal); visualization (equal). **Yaming Tian**: Investigation (equal); methodology (equal); project administration (equal). **Zhen Liu**: Supervision (equal); writing—review and editing (lead). **Jiuda Zhao**: Funding acquisition (equal); supervision (equal); writing—review and editing (equal).

ACKNOWLEDGMENTS
None.

CONFLICT OF INTEREST

Professor Jiuda Zhao is the member of the *Cancer Innovation* Editorial Board. To minimize bias, he was excluded from all editorial decision-making related to the acceptance of this article for publication. The remaining authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request. In addition, the main codes involved in this article are shown in Supplementary Table 1.

ETHICS STATEMENT
Not applicable.

INFORMED CONSENT
Not applicable.

ORCID
Jiuda Zhao http://orcid.org/0000-0002-1266-8943

REFERENCES

1. Ferlay J, Ervik M, Lam F, Laversanne M, Colombet M, Mery L, et al. Global Cancer Observatory: Cancer Today. Lyon, France: International Agency for Research on Cancer; 2024. Available from: https://gco.iarc.who.int/today
2. Rakha EA, Reis-Filho JS, Baehner F, Dabbs DJ, Decker T, Eusebi V, et al. Breast cancer prognostic classification in the molecular era: the role of histological grade. Breast Cancer Res. 2010;12(4):207. https://doi.org/10.1186/bcr2607
3. Ling YX, Xie YF, Wu HL, Wang XF, Ma JL, Fan L, et al. Prognostic factors and clinical outcomes of breast cancer patients with disease progression during neoadjuvant systemic therapy. Breast. 2023;70:63–9. https://doi.org/10.1016/j.breast.2023.06.004
4. Lyu X, Luo B. Prognostic factors and survival prediction in HER2-positive breast cancer with bone metastases: a retrospective cohort study. Cancer Med. 2021;10(22):8114–26. https://doi.org/10.1002/cam4.4326
5. Pu CC, Yin L, Yan JM. Risk factors and survival prediction of young breast cancer patients with liver metastases: a population-based study. Front Endocrinol. 2023;14:1158759. https://doi.org/10.3389/fendo.2023.1158759
6. Kataoka A, Iwamoto T, Tokunaga E, Tomotaki A, Kumamaru H, Miyata H, et al. Young adult breast cancer patients have a poor prognosis independent of prognostic clinicopathological factors: a study from the Japanese Breast Cancer Registry. Breast Cancer Res Treat. 2016;160(1):163–72. https://doi.org/10.1007/s10549-016-3984-8
7. Paluch-Shimon S, Cardoso F, Partridge AH, Abulkhair O, Azim HA, Bianchi-Micheli G, et al. ESO-ESMO fifth international consensus guidelines for breast cancer in young women (BCY5). Ann Oncol. 2022;33(11):1097–118. https://doi.org/10.1016/j.annonc.2022.07.007
8. Experts Committee on Breast Cancer Chinese Society of Clinical Oncology, the Society of Breast Cancer China Anti-Cancer Association, Chinese Society of Breast Surgery Chinese Medical Association. Expert consensus on the diagnosis and treatment of young breast cancer in China (2022 Edition, in Chinese). Zhonghua Yi Xue Za Zhi. 2023;103(6):387–403. https://doi.org/10.3760/cma.j.cn112137-20220907-01895
9. Bhatia S, Pappo AS, Acquazzino M, Allen-Rhoades WA, Barnett M, Borinstein SC, et al. Adolescent and young adult (AYA) oncology, version 2.2024, NCCN clinical practice

guidelines in oncology. J Natl Compr Cancer Netw. 2023;21(8):851–80. https://doi.org/10.6004/jnccn.2023.0040

10. Tzikas AK, Nemes S, Linderholm BK. A comparison between young and old patients with triple-negative breast cancer: biology, survival and metastatic patterns. Breast Cancer Res Treat. 2020;182(3):643–54. https://doi.org/10.1007/s10549-020-05727-x

11. Murphy BL, Day CN, Hoskin TL, Habermann EB, Boughey JC. Adolescents and young adults with breast cancer have more aggressive disease and treatment than patients in their forties. Ann Surg Oncol. 2019;26(12):3920–30. https://doi.org/10.1245/s10434-019-07653-9

12. Huang X, Luo Z, Liang W, Xie G, Lang X, Gou J, et al. Survival nomogram for young breast cancer patients based on the SEER database and an external validation cohort. Ann Surg Oncol. 2022;29(9):5772–81. https://doi.org/10.1245/s10434-022-11911-8

13. Gong Y, Ji P, Sun W, Jiang YZ, Hu X, Shao ZM. Development and validation of nomograms for predicting overall and breast cancer-specific survival in young women with breast cancer: a population-based study. Transl Oncol. 2018;11(6):1334–42. https://doi.org/10.1016/j.tranon.2018.08.008

14. Wu J, Zhang H, Li L, Hu M, Chen L, Xu B, et al. A nomogram for predicting overall survival in patients with low-grade endometrial stromal sarcoma: a population-based analysis. Cancer Commun. 2020;40(7):301–12. https://doi.org/10.1002/cac2.12067

15. SEER [Internet]. [cited 2024 Apr 20]. Surveillance, Epidemiology, and End Results Program. Available from: https://seer.cancer.gov/index.html

16. Longato E, Vettoretti M, di Camillo B. A practical perspective on the concordance index for the evaluation and selection of prognostic time-to-event models. J Biomed Inf. 2020;108:103496. https://doi.org/10.1016/j.jbi.2020.103496

17. Van Calster B, Wynants L, Verbeek JFM, Verbakel JY, Christodoulou E, Vickers AJ, et al. Reporting and interpreting decision curve analysis: a guide for investigators. Eur Urol. 2018;74(6):796–804. https://doi.org/10.1016/j.eururo.2018.08.038

18. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology. 1982;143(1):29–36. https://doi.org/10.1148/radiology.143.1.7063747

19. Aebi S, Gelber S, Anderson SJ, Láng I, Robidoux A, Martín M, et al. Chemotherapy for isolated locoregional recurrence of breast cancer (CALOR): a randomised trial. Lancet Oncol. 2014;15(2):156–63. https://doi.org/10.1016/s1470-2045(13)70589-8

20. Meattini I, Becherini C, Caini S, Coles CE, Cortes J, Curigliano G, et al. International multidisciplinary consensus on the integration of radiotherapy with new systemic treatments for breast cancer: european Society for Radiotherapy and Oncology (ESTRO)-endorsed recommendations. Lancet Oncol. 2024;25(2):e73–83. https://doi.org/10.1016/S1470-2045(23)00534-X

21. Cortes J, Cescon DW, Rugo HS, Nowecki Z, Im SA, Yusof MM, et al. Pembrolizumab plus chemotherapy versus placebo plus chemotherapy for previously untreated locally recurrent inoperable or metastatic triple-negative breast cancer (KEYNOTE-355): a randomised, placebo-controlled, double-blind, phase 3 clinical trial. The Lancet. 2020;396(10265):1817–28. https://doi.org/10.1016/S0140-6736(20)32531-9

22. Schneeweiss A, Chia S, Hickish T, Harvey V, Eniu A, Hegg R, et al. Pertuzumab plus trastuzumab in combination with standard neoadjuvant anthracycline-containing and anthracycline-free chemotherapy regimens in patients with HER2-positive early breast cancer: a randomized phase II cardiac safety study (TRYPHAENA). Ann Oncol. 2013;24(9):2278–84. https://doi.org/10.1093/annonc/mdt182

23. Breast cancer in young women | Nature Reviews Clinical Oncology [Internet]. [cited 2024 Apr 21]. Available from: https://www.nature.com/articles/nrclinonc.2012.102

24. Johnson RH, Anders CK, Litton JK, Ruddy KJ, Bleyer A. Breast cancer in adolescents and young adults. Pediatr Blood Cancer. 2018;65(12):e27397. https://doi.org/10.1002/pbc.27397

25. Shoemaker ML, White MC, Wu M, Weir HK, Romieu I. Differences in breast cancer incidence among young women aged 20–49 years by stage and tumor characteristics, age, race, and ethnicity, 2004–2013. Breast Cancer Res Treat. 2018;169(3):595–606. https://doi.org/10.1007/s10549-018-4699-9

26. Dunnwald LK, Rossing MA, Li CI. Hormone receptor status, tumor characteristics, and prognosis: a prospective cohort of breast cancer patients. Breast Cancer Res. 2007;9(1):R6. https://doi.org/10.1186/bcr1639

27. Cheang MCU, Chia SK, Voduc D, Gao D, Leung S, Snider J, et al. Ki67 index, HER2 status, and prognosis of patients with luminal B breast cancer. JNCI. 2009;101(10):736–50. https://doi.org/10.1093/jnci/djp082

28. US breast cancer mortality trends in young women according to race - Ademuyiwa - 2015 - Cancer - Wiley Online Library [Internet]. [cited 2024 Apr 21]. Available from: https://acsjournals.onlinelibrary.wiley.com/doi/10.1002/cncr.29178

29. Liu G, Kong X, Dai Q, Cheng H, Wang J, Gao J, et al. Clinical features and prognoses of patients with breast cancer who underwent surgery. JAMA Network Open. 2023;6(8):e2331078. https://doi.org/10.1001/jamanetworkopen.2023.31078

30. Radecka B, Litwiniuk M. Breast cancer in young women. Ginekol Pol. 2016;87(9):659–63. https://doi.org/10.5603/GP.2016.0062

31. Yang Y, Wei W, Jin L, He H, Wei M, Shen S, et al. Comparison of the characteristics and prognosis between very young women and older women with breast cancer: a multi-institutional report from China. Front Oncol. 2022;12:783487. https://doi.org/10.3389/fonc.2022.783487

32. Chollet-Hinton L, Anders CK, Tse CK, Bell MB, Yang YC, Carey LA, et al. Breast cancer biologic and etiologic heterogeneity by young age and menopausal status in the *Carolina* breast cancer study: a case-control study. Breast Cancer Res. 2016;18(1):79. https://doi.org/10.1186/s13058-016-0736-y

33. Chen H, Zhou M, Tian W, Meng K, He H. Effect of age on breast cancer patient prognoses: a population-based study using the SEER 18 database. PLoS One. 2016;11(10):e0165409. https://doi.org/10.1371/journal.pone.0165409

34. Haffty BG, Choi DH, Goyal S, Silber A, Ranieri K, Matloff E, et al. Breast cancer in young women (YBC): prevalence of BRCA1/2 mutations and risk of secondary malignancies across diverse racial groups. Ann Oncol. 2009;20(10):1653–9. https://doi.org/10.1093/annonc/mdp051

35. Bao S, He G. Comparing the prognoses of breast-conserving surgeries for differently aged women with early stage breast cancer: use of a propensity score method. Breast J. 2022;2022:1801717. https://doi.org/10.1155/2022/1801717

36. Franco I, Alshalalfa M, Hernandez A, Mahal BA, Nguyen T, Wang L, et al. Genomic characterization of aggressive breast

cancer in younger women. Ann Surg Oncol. 2023;30(12): 7569–78. https://doi.org/10.1245/s10434-023-14080-4

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Li Y, Tao X, Ye Y, Tang Y, Xu Z, Tian Y, et al. Prognostic nomograms for young breast cancer: a retrospective study based on the SEER and METABRIC databases. Cancer Innov. 2024;3:e152. https://doi.org/10.1002/cai2.152