



Nomograms predicting survival for all four subtypes of breast cancer: a SEER-based population study

Jianli Chu^{1,2}, Dehong Yang², Ling Wang¹, Jielai Xia¹

¹Department of Health Statistics, The 4th Military Medical University, Xian 710032, China; ²Center for Clinical Research, Tasly Academy, Tasly Holding Group Co., Ltd., Tianjin 300410, China

Contributions: (I) Conception and design: J Chu, J Xia; (II) Administrative support: J Xia, L Wang; (III) Provision of study materials: J Chu, D Yang; (IV) Collection and assembly of data: J Chu, D Yang; (V) Data analysis and interpretation: J Chu, D Yang, L Wang; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Jielai Xia, Professor. Department of Health Statistics, The 4th Military Medical University, Xian 710032, China.

Email: Xiajielai@fmmu.edu.cn.

Background: The prognosis of female breast cancer (BC) patients is determined by many clinicopathological factors. In this study, we aimed to identify prognostic factors for BC and develop reliable nomograms to predict the 1-, 3-, and 5-year overall survival (OS) and breast cancer-specific survival (BCSS).

Methods: The Surveillance, Epidemiology, and End Results (SEER) database was used to screen 227,989 eligible patients as the study cohort. The whole cohort was randomly divided into a training cohort (n=113,996) and a testing cohort (n=113,993). The log-rank test and Cox proportional hazards analysis were applied to select variables and build nomogram models based on the training cohort. Internal and external validation were performed to evaluate the performance of the models by calculating the C-index and generating calibration plots in the training cohort and testing cohort.

Results: The following factors were included in both the OS and BCSS nomograms: subtypes of BC, metastasis (bone, liver, lung, and brain), age at diagnosis, race, tumor size, grade, number of positive lymph nodes, and marital status. The calibration plots presented excellent consistency between the actual and nomogram-predicted survival probabilities in both the training cohort and testing cohort. The C-index values of the nomograms were 0.796 and 0.793 for OS and 0.856 and 0.853 for BCSS in the training and testing cohorts, respectively.

Conclusions: The established nomograms provide a visualization of the risk of each prognostic factor and can assist clinicians in predicting the 1-, 3-, and 5-year OS and BCSS for all 4 subtypes of BC.

Keywords: Nomograms; breast cancer (BC); all four subtypes; Surveillance, Epidemiology, and End Results (SEER)

Submitted Mar 04, 2020. Accepted for publication Apr 15, 2020.

doi: 10.21037/atm-20-2808

View this article at: <http://dx.doi.org/10.21037/atm-20-2808>

Introduction

Breast cancer (BC) is the most frequently diagnosed cancer in women worldwide, and it is the second leading cause of cancer-related death in women in the United States. In 2020, the American Cancer Society estimates that 276,480 American women will be diagnosed with invasive BC and 42,170 will die of the disease in the United States (1).

Hormone receptors (HRs) [estrogen receptor (ER) and progesterone receptor (PR)] and human epidermal growth factor receptor 2 (HER2) are used as biomarkers for selecting the appropriate therapy and evaluating prognosis in the clinical practice of BC. Patients with ER- and/or PR-positive BC are likely to respond to endocrine treatment and have better survival than those with ER- and/or PR-negative BC (2). The overexpression of HER2

is associated with high-grade tumors, lymph node involvement, a higher relapse rate, and mortality (3). In one study, HER2-targeted therapy led to a significant improvement in the survival of patients with HER2-positive BC (4). Based on these biomarkers, BC can be classified into 4 subtypes: HR+/HER2-, HR+/HER2+, HR-/HER2+ (HER2 overexpression), and HR-/HER2- [triple-negative breast cancer (TNBC)].

In a previous study, the proportions of the BC subtypes were 72.7% for HR+/HER2-, 12.2% for TNBC, 10.3% for HR+/HER2+, and 4.6% for HR-/HER2+ (5). TNBC is more likely to occur among younger women and black women (6). HR-/HER2+ and TNBC tumors are known to be more clinically aggressive and associated with a poorer prognosis than HR+/HER2- tumors (7). Therefore, the subtypes of BC are considered prognostic factors for survival (8). BC can spread to different distant organs, preferentially to the bones, lung, liver, and brain. A previous study showed that bone metastasis accounted for 38.9%, lung metastasis accounted for 17.7%, liver metastasis accounted for 11.9%, multiple metastases accounted for 13.8%, and brain metastasis accounted for 2.5% (9) among 1,038 metastasis cases. Approximately one-third of BC patients will present with distant metastasis, and the 5-year survival rate decreases to 23% when distant metastasis occurs (1).

For nonmetastatic breast cancer, systemic therapy is determined by subtype. TNBC require chemotherapy alone. HR-/HER2+ BC are treated with chemotherapy combined with trastuzumab therapy. HR+/HER2- BC receive endocrine therapy and a minority require chemotherapy. HR+/HER2+ BC require chemotherapy with trastuzumab and endocrine therapy. Systemic therapy for metastatic breast cancer depends on subtype, including standard regimens used in early lines plus agents for later lines. In metastatic HR+/HER2- BC, cyclin-dependent kinase (CDK) 4/6 inhibitor, such as abemaciclib, palbociclib, or ribociclib, is used for first or second line of endocrine therapy. For metastatic TNBC with germline BRCA1/2 mutations, PARP inhibitors, Olaparib and Talazoparib can be therapeutic option. However, more effective therapy for metastatic TNBC is lacking and the clinical trials of antibody-drug conjugates and programmed death ligand 1 (PD-L1) are ongoing. In HER2+ metastatic BC, taxane plus trastuzumab and pertuzumab is used as standard first line therapy, and the antibody-drug conjugate trastuzumab emtansine as second-line therapy (10).

A nomogram is a kind of visual tool based on a prognostic

model that includes the related clinicopathological factors that provides the probabilities of the clinical outcomes of particular individuals. Compared to the traditional tumor-node-metastasis (TNM) staging system, nomograms are able to integrate more important prognostic factors and provide a more precise estimation of prognosis. In our study, subtype and metastatic pattern are included in the nomograms, except for other clinicopathological factors, such as age, race, tumor size, tumor grade, etc. Recently, some previous nomograms had been built for TNBC (11) or brain metastasis BC (12). However, to the best of our knowledge, there was no study to build nomograms for all 4 subtypes of BC based on the updated data of the recent years. In this study, we aimed to build nomograms capable of predicting the survival outcomes of all 4 subtypes of BC patients based on a large population database from the Surveillance, Epidemiology, and End Results (SEER) program.

Methods

Study cohorts

SEER is a large-scale cancer registration database that covers approximately 34.6% of the U.S. population. The data for this study were selected from 18 registries of the SEER program, which is supported by the National Cancer Institute (NCI). The primary cohort for this study was collected from patients with information on the subtypes of BC in the SEER database from January 1, 2010, to December 31, 2016. The inclusion criteria were defined as follows: (I) female; (II) older than 18 years; (III) diagnosis confirmed by positive histology other than by other methods; (IV) BC as the first and primary cancer by international rules; (V) belonging to 1 of the 4 subtypes, which are HR+/HER2-, HR+/HER2+, HR-/HER2+, and HR-/HER2-; (VI) survival data with complete and available dates and more than 0 days of survival; and (VII) clear clinicopathological information for all the variables of interest including age at diagnosis, race, marital status, breast subtype, tumor size, location, grade, laterality, number of positive lymph nodes, histological subtype, and metastasis site (*Figure 1*).

To develop and validate the nomogram model, the primary cohort was randomly divided into a training cohort (n=113,996) and a validation cohort (n=113,993) by applying 'createDataPartition' function in the package of 'caret' from R, version 3.6.1.

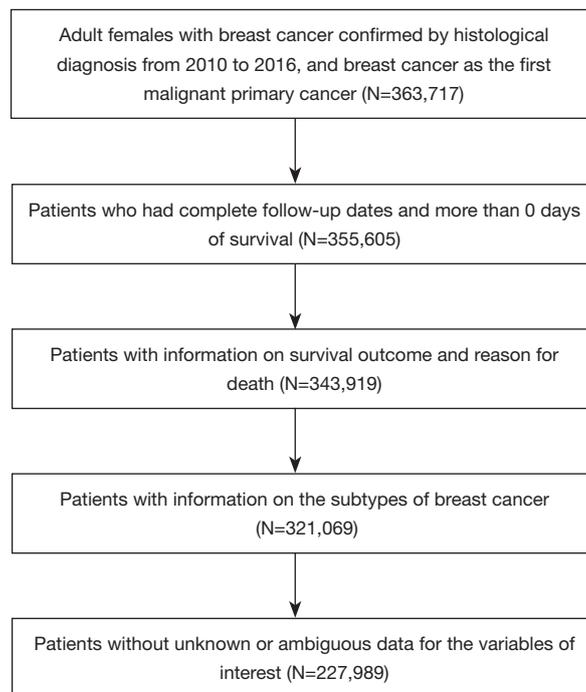


Figure 1 Flowchart of the cohort selection process.

Variables and endpoints

The following variables at diagnosis were selected as the potential prognostic factors: age, race (White, Black, American Indian/Alaska Native, Asian, or Pacific Islander), marital status, laterality (right or left side), tumor subtypes (HR+/HER2-, HR+/HER2+, HR-/HER2+, HR-/HER2-), tumor location (nipple, central portion of the breast, upper-inner quadrant of the breast, lower-inner quadrant of the breast, upper-outer quadrant of the breast, lower-outer quadrant of the breast, axillary tail of the breast, overlapping lesion of the breast), tumor grade (well-differentiated, moderately differentiated, poorly differentiated, undifferentiated or anaplastic), tumor size, number of positive regional nodes, and histological subtype. Marital status was classified as married or unmarried. The latter included single, separated, divorced, widowed, and unmarried/domestic partners. Histological subtype was classified as infiltrating duct carcinoma, lobular carcinoma, and other. The value of age at diagnosis, tumor size, and number of positive regional nodes were transformed into grouped categorical variables according to regular practice.

Both overall survival (OS) and breast cancer-specific survival (BCSS) were used as primary endpoints for this study. OS was defined as the total survival time from the

date of diagnosis to the date of death occurring as a result of all causes of death. BCSS was defined as the time from the date of diagnosis to the date of death caused by BC. The censor time point for this study was December 31, 2016, which was the latest update on the follow-up time.

Statistical analysis

The baseline characteristics of the cohorts were described by summarizing the counts and percentages for each variable of interest.

Based on the training cohort, the risk of each prognostic factor for OS and BCSS was estimated by applying the log-rank test and multivariate analysis. First, univariate analysis was performed by using each of the potential prognostic factors as the only dependent variable. If the P value was significant (<0.05) in the log-rank test of univariate analysis, the factor was included in the multivariate Cox proportional hazards model. Then, the significant factors in multivariate analysis (at least 1 level with P value <0.05) were selected for the final prognostic models to construct the nomograms. The 1-, 3-, and 5-year prognoses of OS and BCSS were predicted by the constructed nomograms.

To confirm the predictive accuracy of the nomogram,

both internal (200 bootstrap resamples based on the training cohort) and external (based on the testing cohort) validations were performed. The performance of the models for predicting the survival outcomes was evaluated by C-index values (13) and calibration plots. The range of the C-index value was from 0.5 to 1.0, which would indicate that the discrimination ability of the models is low to perfect. The calibration plots were generated by comparing the observed survival probabilities with the nomogram-predicted probabilities of OS and BCSS.

Kaplan-Meier curves were utilized to show the impact of each prognostic factor on survival outcomes based on the primary cohort.

Results

Demographics and clinicopathological characteristics

There were 227,989 adult female patients with subtype information who were selected from the SEER database for this study. The patients were randomly allocated into the training cohort and the validation cohort, with 113,996 patients in the training cohort and 113,993 in the testing cohort. Among all the patients, more than half of the patients were between 50–59 (25.78%) and 60–69 (28.35%) years of age. In terms of race, most of the patients were White (79.54%). In terms of tumor subtypes, HR+/HER2– accounted for 74.62% of the total, whereas the proportions for the other 3 subtypes were 10.25% for HR+/HER2+, 3.98% for HR–/HER2+, and 11.14% for TNBC. In general, all the factors had similar proportions between the training cohort and testing cohort, which indicated that patient allocation was performed according to random principles. By the end of the follow-up, 15,027 (6.6%) patients had died in the primary cohort, with 8,586 (3.8%) deaths due to BC and the remaining 6,441 (2.8%) due to other causes. The details of the baseline characteristics are listed in *Table 1*.

Univariate and multivariate Cox proportional hazards analyses

The hazard ratios for OS and BCSS according to all variables in the univariate or multivariate Cox proportional hazards model are listed in *Table 2*. According to the results of univariate and multivariate analyses, we found that the laterality, histology, and location of the tumor were not significant factors for either OS or BCSS. After excluding

the above unqualified variables, we finally had age, race, marital status, subtypes of BC, tumor grade, tumor size, number of positive lymph nodes, bone metastasis, liver metastasis, lung metastasis, and brain metastasis as prognostic factors in the multivariate Cox proportional hazards models for both the OS and BCSS analyses. Among the subgroups of age, the subgroups of 70–79 and ≥ 80 years of age had a significantly higher risk than the younger subgroups. Compared to White patients, Black and American Indian/Alaska Native patients were at higher risk of death, whereas Asian or Pacific Islander patients were at lower risk. The HR+/HER2+ subtype exhibited the lowest risk among the 4 subtypes according to the results of multivariate analysis. The hazard ratios of the other 3 subtypes increased in the following order: HR+/HER2–, HR–/HER2+, and TNBC. The unmarried group also showed a higher risk than the married group. The detailed results for the other factors are presented in *Table 2*. Collectively, each prognostic factor had consistent hazard ratio results between the OS and BCSS analyses.

Construction and validation of the nomograms

The nomograms for 1-, 3-, and 5-year OS and BCSS were generated by using the multivariate Cox proportional hazards models as the final prognostic models after the process of factor selection (*Figure 2*). The calibration plots presented excellent consistency between the actual and nomogram-predicted survival probabilities in both the training cohort and the testing cohort (*Figure 3*). The C-index values of the nomograms in the training cohort were 0.794 (95% CI, 0.789–0.800) for OS and 0.855 (95% CI, 0.849–0.861) for BCSS. In the testing cohort, the C-index values were 0.795 (95% CI, 0.790–0.801) for OS and 0.856 (95% CI, 0.850–0.862) for BCSS.

Survival analysis

Kaplan-Meier curves were generated to present the effect of the prognostic factors in the nomograms on OS and BCSS based on the primary cohort. All the prognostic factors in the nomograms were also significant in the primary cohort. This result was consistent with the results of the training cohort, as shown in *Table 2*. From the curves, we found that most of the factors presented the same outcome trends for OS and BCSS except for age. From the curve of the age factor, we found that the subgroup of ≥ 80 years of age had a markedly poorer prognosis for OS than for BCSS.

Table 1 Demographics and baseline characteristics of patients with four subtypes of breast cancer

Characteristics	All patients (n=227,989)		Training cohort (n=113,996)		Testing cohort (113,999)	
	Number of patients	%	Number of patients	%	Number of patients	%
Age						
18–29	1,312	0.58	675	0.59	637	0.56
30–39	10,532	4.62	5,198	4.56	5,334	4.68
40–49	39,364	17.27	19,706	17.29	19,658	17.24
50–59	58,777	25.78	29,349	25.75	29,428	25.82
60–69	64,635	28.35	32,351	28.38	32,284	28.32
70–79	39,275	17.23	19,683	17.27	19,592	17.19
≥80	14,094	6.18	7,034	6.17	7,060	6.19
Race						
White	181,350	79.54	90,727	79.59	90,623	79.50
Black	24,372	10.69	12,116	10.63	12,256	10.75
American Indian/Alaska Native	1,349	0.59	658	0.58	691	0.61
Asian or Pacific Islander	20,918	9.18	10,495	9.21	10,423	9.14
Marital status						
Married	137,213	60.18	68,725	60.29	68,488	60.08
Unmarried	90,776	39.82	45,271	39.71	45,505	39.92
Subtypes						
HR+/HER2–	170,135	74.62	85,005	74.57	85,130	74.68
HR+/HER2+	23,377	10.25	11,684	10.25	11,693	10.26
HR–/HER2+	9,083	3.98	4,545	3.99	4,538	3.98
Triple negative	25,394	11.14	12,762	11.20	12,632	11.08
Histology						
Infiltrating duct carcinoma	174,174	76.40	87,054	76.37	87,120	76.43
Lobular carcinoma	20,107	8.82	9,995	8.77	10,112	8.87
Other	33,708	14.78	16,947	14.87	16,761	14.70
Location						
Nipple	752	0.33	364	0.32	388	0.34
Central portion of breast	11,433	5.01	5,682	4.98	5,751	5.05
Upper-inner quadrant of breast	32,717	14.35	16,389	14.38	16,328	14.32
Lower-inner quadrant of breast	14,266	6.26	7,082	6.21	7,184	6.30
Upper-outer quadrant of breast	89,370	39.20	44,831	39.33	44,539	39.07
Lower-outer quadrant of breast	19,702	8.64	9,862	8.65	9,840	8.63
Axillary tail of breast	1,131	0.50	566	0.50	565	0.50
Overlapping lesion of breast	58,618	25.71	29,220	25.63	29,398	25.79

Table 1 (continued)

Table 1 (continued)

Characteristics	All patients (n=227,989)		Training cohort (n=113,996)		Testing cohort (113,999)	
	Number of patients	%	Number of patients	%	Number of patients	%
Laterality						
Right: origin of primary	112,772	49.46	56,601	49.65	56,171	49.28
Left: origin of primary	115,217	50.54	57,395	50.35	57,822	50.72
Grade						
Well differentiated; Grade I	54,613	23.95	27,335	23.98	27,278	23.93
Moderately differentiated; Grade II	100,854	44.24	50,321	44.14	50,533	44.33
Poorly differentiated; Grade III	72,025	31.59	36,097	31.67	35,928	31.52
Undifferentiated; anaplastic; Grade IV	497	0.22	243	0.21	254	0.22
Tumor size, cm						
≤1	58,349	25.59	29,078	25.51	29,271	25.68
≤2	83,314	36.54	41,809	36.68	41,505	36.41
≤3	45,279	19.86	22,700	19.91	22,579	19.81
≤4	18,282	8.02	9,111	7.99	9,171	8.05
≤5	8,750	3.84	4,359	3.82	4,391	3.85
>5	14,015	6.15	6,939	6.09	7,076	6.21
Positive regional nodes number						
0	162,133	71.11	81,105	71.15	81,028	71.08
1–3	48,687	21.35	24,354	21.36	24,333	21.35
4–9	11,831	5.19	5,831	5.12	6,000	5.26
≥10	5,338	2.34	2,706	2.37	2,632	2.31
Bone metastasis						
No	226,355	99.28	113,193	99.30	113,162	99.27
Yes	1,634	0.72	803	0.70	831	0.73
Brain metastasis						
No	227,892	99.96	113,946	99.96	113,946	99.96
Yes	97	0.04	50	0.04	47	0.04
Liver metastasis						
No	227,393	99.74	113,689	99.73	11,3704	99.75
Yes	596	0.26	307	0.27	289	0.25
Lung metastasis						
No	227,342	99.72	113,662	99.71	113,680	99.73
Yes	647	0.28	334	0.29	313	0.27

Table 2 Univariate and multivariate Cox analysis of overall survival and breast cancer-specific survival

Variables	Overall survival			Overall survival		
	Log-rank test	Univariate analysis	Multivariate analysis	Log-rank test	Univariate analysis	Multivariate analysis
	P value	HR (95% CI)	HR (95% CI)	P value	HR (95% CI)	HR (95% CI)
Age	<0.001			<0.001		
18-29		Reference	Reference		Reference	Reference
30-39		0.87 (0.65-1.15)	0.97 (0.73-1.29)	0.844	0.91 (0.68-1.24)	1.02 (0.75-1.38)
40-49		0.51 (0.39-0.67)	0.77 (0.58-1.01)	0.062	0.48 (0.36-0.64)	0.80 (0.60-1.07)
50-59		0.56 (0.42-0.73)	0.91 (0.69-1.20)	0.501	0.47 (0.35-0.63)	0.89 (0.66-1.19)
60-69		0.63 (0.48-0.83)	1.19 (0.91-1.57)	0.200	0.41 (0.30-0.54)	0.96 (0.72-1.28)
70-79		1.11 (0.85-1.46)	2.22 (1.69-2.91)	<0.001	0.54 (0.40-0.72)	1.42 (1.06-1.90)
≥80		2.97 (2.27-3.90)	4.99 (3.80-6.55)	<0.001	1.17 (0.87-1.57)	2.47 (1.84-3.32)
Race	<0.001			<0.001		
White		Reference	Reference		Reference	Reference
Black		1.62 (1.52-1.72)	1.29 (1.21-1.37)	<0.001	1.98 (1.83-2.13)	1.30 (1.20-1.41)
American Indian/ Alaska Native		1.17 (0.88-1.54)	1.22 (0.92-1.61)	0.169	1.19 (0.82-1.73)	1.09 (0.75-1.58)
Asian or Pacific Islander		0.56 (0.51-0.62)	0.64 (0.57-0.71)	<0.001	0.65 (0.57-0.74)	0.67 (0.59-0.77)
Marital status	<0.001			<0.001		
Married		Reference	Reference		Reference	Reference
Unmarried		1.96 (1.87-2.05)	1.36 (1.30-1.43)	<0.001	1.63 (1.53-1.73)	1.20 (1.12-1.27)
Subtypes	<0.001			<0.001		
HR+/HER2-		Reference	Reference		Reference	Reference
HR+/HER2+		1.09 (1.01-1.19)	0.89 (0.82-0.97)	0.009	1.42 (1.28-1.58)	0.86 (0.77-0.96)
HR-/HER2+		1.67 (1.51-1.85)	1.12 (1.01-1.24)	0.040	2.70 (2.40-3.05)	1.26 (1.11-1.43)
Triple negative		2.76 (2.62-2.91)	2.07 (1.95-2.20)	<0.001	4.64 (4.35-4.96)	2.74 (2.54-2.96)

Table 2 (continued)

Table 2 (continued)

Variables	Overall survival				Overall survival					
	Log-rank test		Univariate analysis		Multivariate analysis		Univariate analysis		Multivariate analysis	
	P value	HR (95% CI)	P value	HR (95% CI)	P value	HR (95% CI)	P value	HR (95% CI)	P value	HR (95% CI)
Histology	0.596									
Infiltrating duct carcinoma							Reference		Excluded	
Lobular carcinoma							0.79 (0.70–0.89)		–	<0.001
Other							0.86 (0.79–0.94)		–	0.001
Location	<0.001									<0.001
Nipple			Reference		Excluded				Reference	Excluded
Central portion			0.91 (0.65–1.26)	0.574	–		0.86 (0.56–1.31)	0.481	–	
Upper-inner quadrant			0.55 (0.40–0.76)	<0.001	–		0.51 (0.33–0.77)	0.002	–	
Lower-inner quadrant			0.63 (0.46–0.88)	0.007	–		0.60 (0.39–0.92)	0.019	–	
Upper-outer quadrant			0.62 (0.45–0.86)	0.004	–		0.60 (0.39–0.90)	0.013	–	
Lower-outer quadrant			0.66 (0.48–0.91)	0.013	–		0.65 (0.43–0.99)	0.046	–	
Axillary tail			0.83 (0.54–1.27)	0.386	–		0.73 (0.42–1.28)	0.277	–	
Overlapping lesion			0.70 (0.50–0.96)	0.026	–		0.67 (0.44–1.01)	0.055	–	
Laterality	0.759									0.989
Right										
Left										
Grade	<0.001									<0.001
Grade I			Reference		Reference		Reference		Reference	
Grade II			1.47 (1.37–1.58)	<0.001	1.10 (1.02–1.19)	0.012	2.73 (2.37–3.13)	<0.001	1.65 (1.43–1.90)	<0.001
Grade III			3.03 (2.82–3.24)	<0.001	1.64 (1.51–1.77)	<0.001	9.30 (8.15–10.61)	<0.001	3.24 (2.81–3.73)	<0.001
Grade IV			3.41 (2.48–4.70)	<0.001	1.92 (1.39–2.65)	<0.001	10.63 (7.23–15.62)	<0.001	3.61 (2.44–5.33)	<0.001
Tumor size, cm	<0.001									<0.001
≤1			Reference		Reference		Reference		Reference	

Table 2 (continued)

Table 2 (continued)

Variables	Overall survival						Overall survival					
	Log-rank test		Univariate analysis		Multivariate analysis		Log-rank test		Univariate analysis		Multivariate analysis	
	P value	HR (95% CI)	P value	HR (95% CI)	P value	HR (95% CI)	P value	HR (95% CI)	P value	HR (95% CI)	P value	HR (95% CI)
≤2		1.61 (1.49–1.75)	<0.001	1.34 (1.24–1.46)	<0.001	2.63 (2.27–3.06)	<0.001	1.81 (1.56–2.11)	<0.001			
≤3		2.79 (2.57–3.02)	<0.001	1.80 (1.65–1.96)	<0.001	6.51 (5.63–7.54)	<0.001	2.86 (2.46–3.32)	<0.001			
≤4		4.42 (4.04–4.83)	<0.001	2.38 (2.16–2.62)	<0.001	11.92 (10.25–13.87)	<0.001	4.05 (3.46–4.74)	<0.001			
≤5		5.20 (4.68–5.77)	<0.001	2.70 (2.41–3.01)	<0.001	14.12 (11.99–16.62)	<0.001	4.48 (3.77–5.31)	<0.001			
>5		7.30 (6.69–7.96)	<0.001	3.29 (2.99–3.62)	<0.001	22.92 (19.80–26.53)	<0.001	6.18 (5.29–7.23)	<0.001			
Positive regional nodes number	<0.001					<0.001						
0		Reference		Reference		Reference		Reference				Reference
1–3		1.94 (1.84–2.05)	<0.001	1.55 (1.47–1.64)	<0.001	3.15 (2.93–3.39)	<0.001	2.13 (1.98–2.30)	<0.001			
4–9		3.99 (3.72–4.27)	<0.001	2.56 (2.38–2.76)	<0.001	7.55 (6.93–8.23)	<0.001	3.72 (3.40–4.08)	<0.001			
≥10		6.68 (6.18–7.23)	<0.001	3.54 (3.25–3.86)	<0.001	13.64 (12.43–14.98)	<0.001	5.61 (5.07–6.21)	<0.001			
Bone metastasis	<0.001					<0.001						
No		Reference		Reference		Reference		Reference				Reference
Yes		7.81 (6.98–8.74)	<0.001	2.29 (2.01–2.61)	<0.001	12.73 (11.30–14.33)	<0.001	2.79 (2.43–3.21)	<0.001			
Brain metastasis	<0.001					<0.001						
No		Reference		Reference		Reference		Reference				Reference
Yes		25.41 (18.14–35.59)	<0.001	4.30 (3.04–6.08)	<0.001	42.06 (29.86–59.26)	<0.001	4.27 (3.00–6.08)	<0.001			
Liver metastasis	<0.001					<0.001						
No		Reference		Reference		Reference		Reference				Reference
Yes		12.81 (10.90–15.04)	<0.001	3.25 (2.69–3.94)	<0.001	20.98 (17.76–24.77)	<0.001	3.14 (2.58–3.83)	<0.001			
Lung metastasis	<0.001					<0.001						
No		Reference		Reference		Reference		Reference				Reference
Yes		12.11 (10.38–14.13)	<0.001	2.17 (1.81–2.60)	<0.001	19.66 (16.74–23.10)	<0.001	2.49 (2.07–3.00)	<0.001			

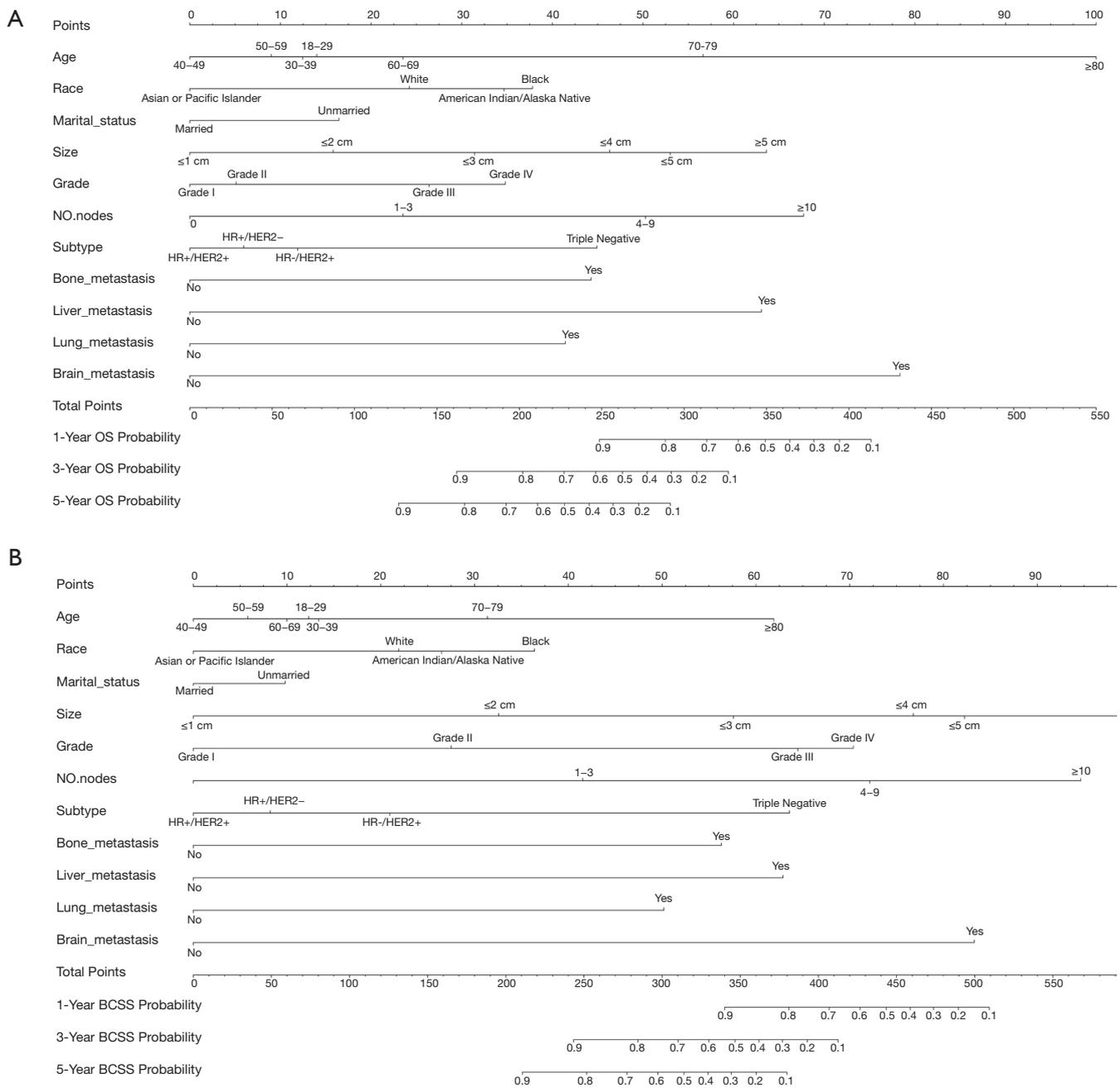


Figure 2 Nomograms for predicting 1-, 3-, and 5-year OS (A) and BCSS (B) for patients with the prognosis factors. The total points are calculated by summing up the points for each factor. The predicted probabilities of OS and BCSS can be obtained by projecting the location of the total points to the bottom scales. NO. nodes: number of positive lymph nodes. OS, overall survival; BCSS, breast cancer-specific survival.

Discussion

The survival prognosis of female BC patients can be affected by multiple factors simultaneously. Therefore, it is necessary to integrate all the possible factors together and

to determine the true factors related to prognosis. Different from the nomograms intended for a specific type of BC, our nomograms were established to predict OS and BCSS for all 4 subtypes of BC based on a large cohort of 227,989

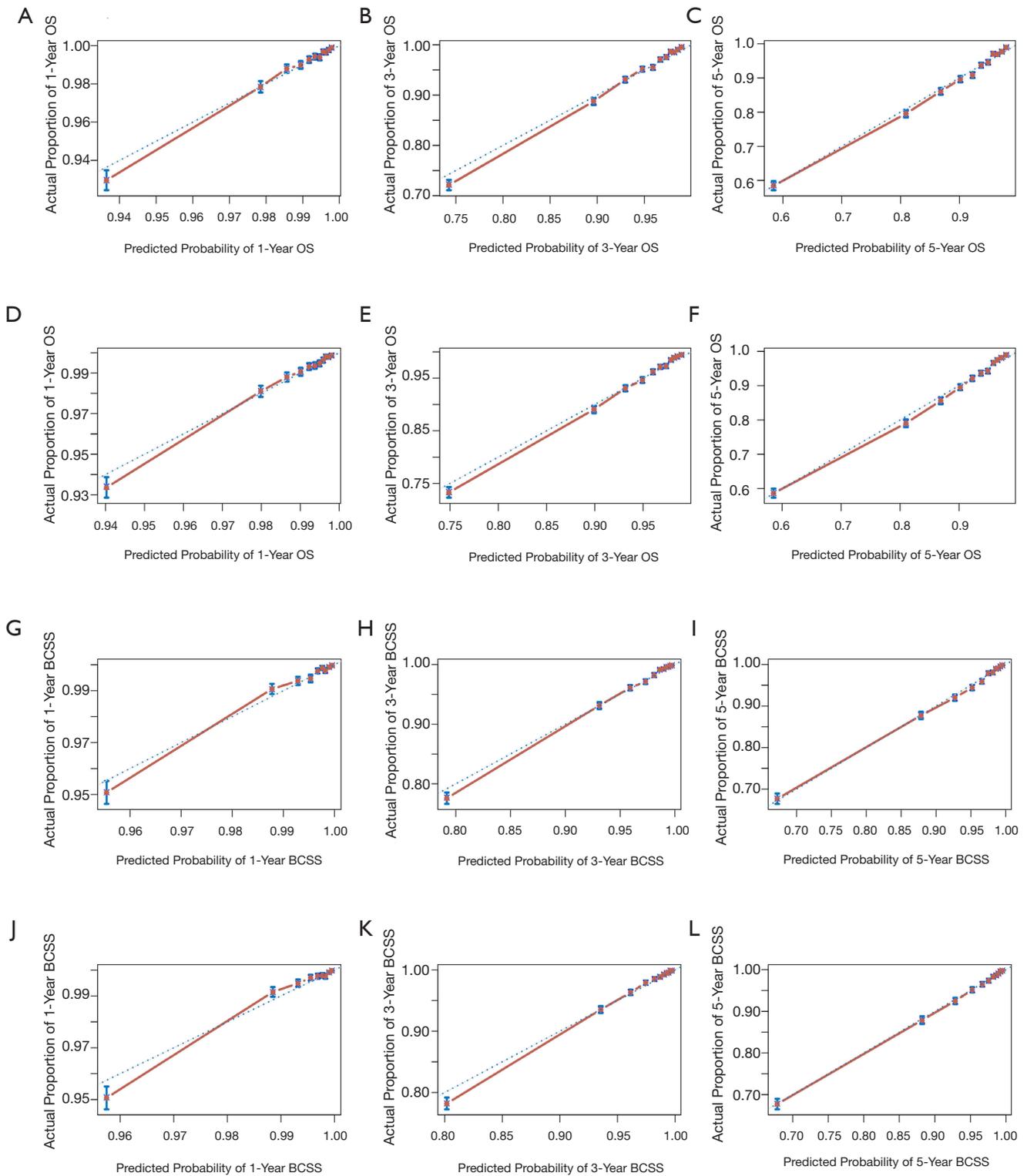


Figure 3 Calibration curves for the 1-, 3-, and 5-year. (A,B,C) Internal calibration curves for OS; (D,E,F) external calibration curves for OS; (G,H,I) internal calibration curves for BCSS; (J,K,L) external calibration curves for BCSS. OS, overall survival; BCSS, breast cancer-specific survival.

patients from the SEER database.

Prognostic factors

In terms of the BC subtypes, a previous study showed that the best survival pattern was observed among women with the HR+/HER2- subtype (survival rate of 92.5% at 4 years), followed by HR+/HER2+ (90.3%), HR-/HER2+ (82.7%), and finally TNBC (77.0%), which had the worst survival. In stage IV BC, there is evidence that the HR+/HER2+ subtype exhibits better survival than the HR+/HER2- subtype (7). In operable invasive BC, the HR-/HER2+ subtype shows better prognosis than TNBC but worse prognosis than the HR+ subtypes regarding both BCSS and OS (8). In our study, however, HR+/HER2+ had a lower survival risk than HR+/HER2- in multivariate analysis. This result might be explained by endocrine therapy and targeted HER-2 therapy for the HR+/HER2+ subtype. The HR-/HER2+ subtype and TNBC subtype had a higher survival risk than the HR+/HER2- subtype, which was consistent with the results of previous studies.

As the critical prognostic factor of BC, the site of metastasis has a strong correlation with survival outcomes. Bone metastasis is the most common metastasis of BC. The median survival time for patients with bone-only metastasis is 7.54 years (14). Lung metastasis is the second most common metastasis in BC patients, with a median survival of 22 months after treatment (15). As the third most common metastasis, liver metastasis leads to a median survival time of only 4–8 months in patients without treatment (16). Compared to the above 3 kinds of metastases, brain metastasis is an infrequent pattern but represents a significant cause of morbidity and mortality in BC. The median survival of brain metastases ranges from 7 months for TNBC to 20 months for the luminal B subtype (17). Likewise, we found that patients with bone metastasis had a longer median survival time than those with liver and lung metastasis. Moreover, brain metastasis led to the shortest median survival time (*Figure 4A,B,C,D,E,F,G,H*).

Age at diagnosis was found to be an important prognostic factor for BC. A previous study showed that the 5-year survival rates of patients aged less than 40, between 40 and 50, and >50 were 54.3%±3.5%, 68.5%±1.9%, and 70.4%±1.3%, respectively (18). In another study, BC-specific mortality at 5 years for age >80 and 70–79 was 25.8% and 17.2%, respectively (19). In this study, we observed a similar trend among the age subgroups. The hazard ratios of OS and BCSS showed a downward trend

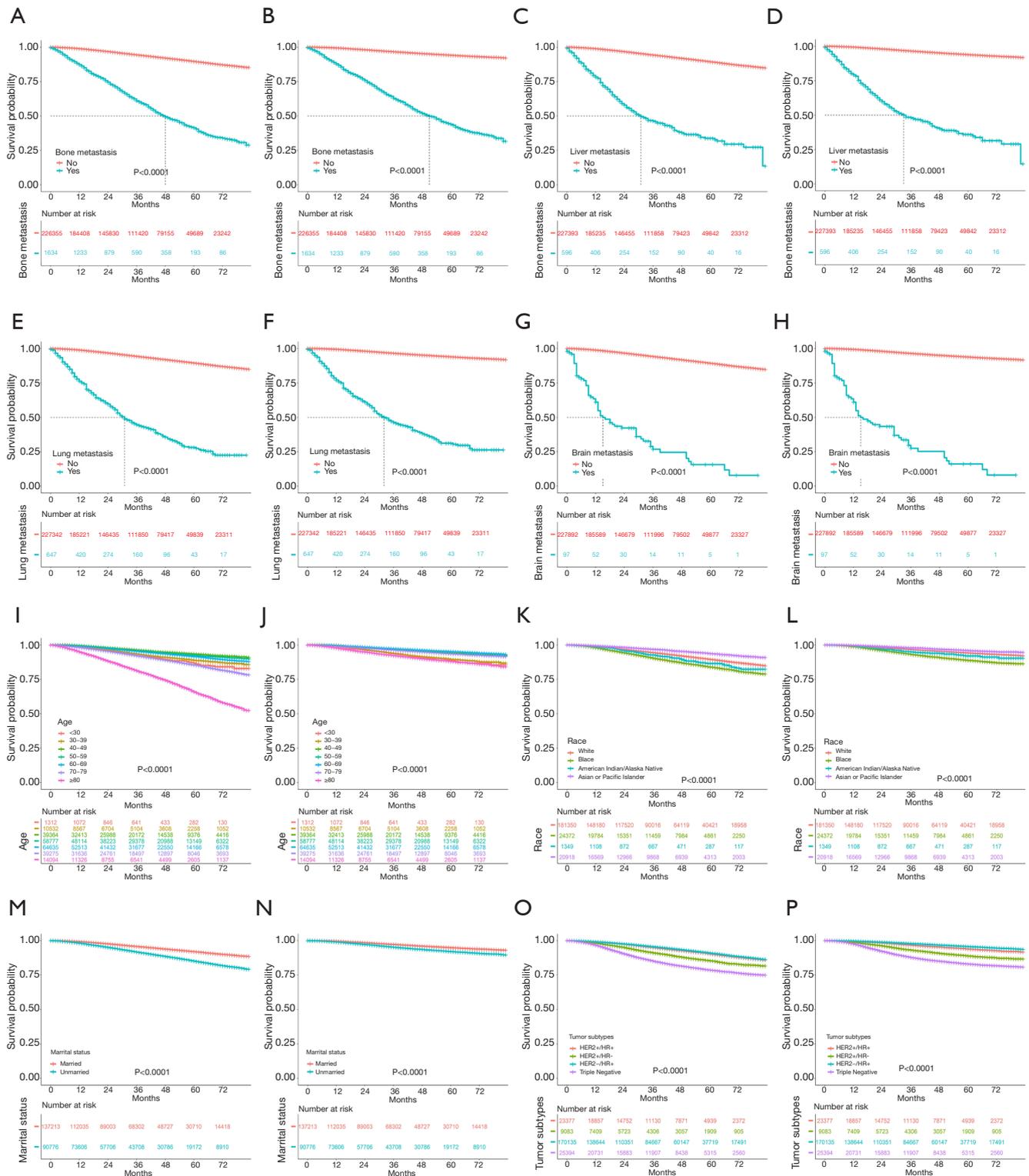
from the subgroup of <40 years of age to the subgroups of 40–49 and 50–59 years of age, and then rose significantly as age increased. In addition to this U-shaped trend of hazard ratios, we also noticed that the BCSS of the ≥80-year-old subgroup was not as bad as the OS (*Figure 4I,J*). This finding indicated that the poor survival prognosis of patients aged ≥80 years old might be due to other reasons unrelated to BC itself.

In a systematic review, it was shown that unmarried patients had a higher risk of metastatic cancer and shorter survival, and unmarried individuals had higher odds of having a later stage of BC at diagnosis. These trends are likely due to the lack of the positive effect of marriage affecting the likelihood of cancer being diagnosed at an early stage (20). This is consistent with our study, which also provides evidence suggesting that unmarried patients are at a higher risk for poor prognosis.

Several studies have shown that race is another prognostic factor for survival outcomes. Compared to White women, all of African American, Hispanic/Latina, Asian American/Pacific Islander, and American Indian/Alaska Native women have lower incidence rates, but they are more likely to be diagnosed at regional/distant stages, which is associated with poorer survival (21,22). Moreover, our data showed that American Indian/Alaska Native and African American BC patients had higher mortality than White patients, whereas Asian/Pacific Islander BC patients had lower mortality than White patients.

Predictive capability of the models

The Cox proportional hazards method was applied to construct the nomogram models. To ensure that the factors indeed contributed to the models, only the qualified factors were selected for the nomogram models. The factors of interest were considered qualified prognostic factors only if they were significant in both the univariate and multivariate analyses. The performance of the models was evaluated by calibration and discrimination through both internal and external validations. Here, discrimination refers to the ability of the models to correctly distinguish patients with events from those without events (23). Calibration is defined as the degree of consistency between the estimated risk generated by the model and the actual observed risk. The calibration plots demonstrated good agreement between the estimated probabilities and the real probabilities for both OS and BCSS. To evaluate the discrimination of the nomograms, the C-index was calculated for both OS and BCSS based on the training and testing cohorts. As shown above, all the values of the C-index were greater than 0.7.



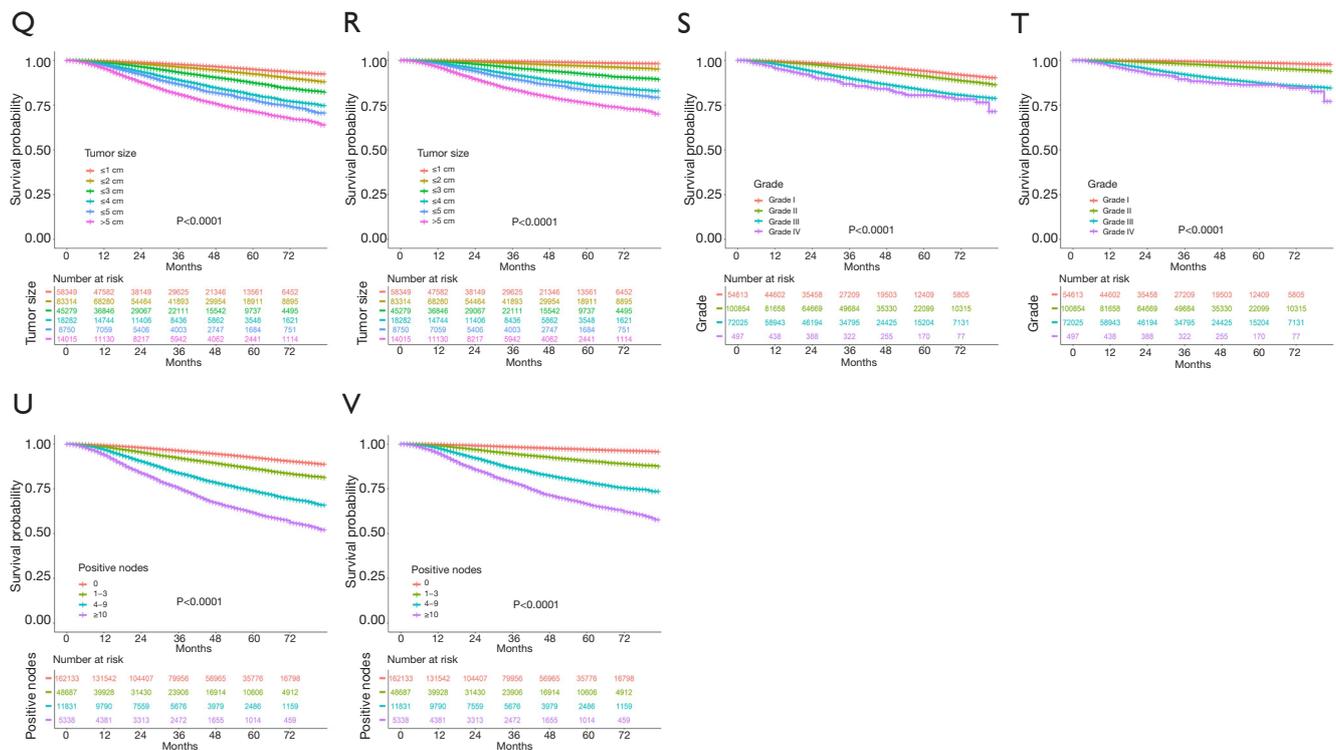


Figure 4 Kaplan-Meier curves of OS and BCSS for each predictor. (A,B) Bone metastasis; (C,D) liver metastasis; (E,F) lung metastasis; (G,H) brain metastasis; (I,J) age; (K,L) race; (M,N) marital status; (O,P) subtype; (O,R) tumor size; (S,T) grade; (U,V) number of positive lymph nodes. OS, overall survival; BCSS, breast cancer-specific survival.

In particular, the C-index values of BCSS were greater than 0.85 in both the training and testing cohorts. These results demonstrated that the nomograms had good discrimination for OS and excellent discrimination for BCSS. In addition, the 95% CI of the C-index values was particularly narrow for both OS and BCSS, which indicated that the established nomograms had a high degree of credibility.

Limitations

Unavoidably, there were also some limitations of this study. First, the treatment variables were not considered prognostic factors because the information on treatment in the SEER database was limited. Second, the study cohort did not include patients with missing or unknown information for any of the involved variables, which may cause selection bias. Third, as a retrospective study, our nomograms need to be confirmed in further prospective studies. Fourth, although internal and external validations could evaluate the performance of the nomograms, it is necessary to validate the nomograms in cohorts outside of the SEER program. Therefore, further prospective studies based on other cohorts

are needed to guarantee the performance of our nomograms.

Conclusions

Based on a large-scale population from the SEER database, we constructed nomograms to predict survival outcomes for all 4 subtypes of BC patients. The established nomograms could provide a visualized estimation of risk for each prognostic factor and assist clinicians in predicting the 1-, 3-, and 5-year OS and BCSS of BC patients.

Acknowledgments

Funding: National Natural Science Foundation of China (Project number: 81973141).

Footnote

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at <http://dx.doi.org/10.21037/atm-20-2808>). 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.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. *CA Cancer J Clin* 2020;70:7-30.
2. 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:R6.
3. Gleeson JP, Keegan NM, Morris PG. Neoadjuvant pertuzumab, T-DM1, weekly paclitaxel and possible anthracyclines in HER2 positive early breast cancer treatment—questions from the KRISTINE study. *Transl Cancer Res* 2018;7:S562-7.
4. Loibl S, Gianni L. HER2-positive breast cancer. *Lancet* 2017;389:2415-29.
5. Howlader N, Altekruse SF, Li CI, et al. US incidence of breast cancer subtypes defined by joint hormone receptor and HER2 status. *J Natl Cancer Inst* 2014. doi: 10.1093/jnci/dju055.
6. Plasilova ML, Hayse B, Killelea BK, et al. Features of triple-negative breast cancer: Analysis of 38,813 cases from the national cancer database. *Medicine (Baltimore)* 2016;95:e4614.
7. Howlader N, Cronin KA, Kurian AW, et al. Differences in Breast Cancer Survival by Molecular Subtypes in the United States. *Cancer Epidemiol Biomarkers Prev* 2018;27:619-26.
8. Brosnan EM, Anders CK. Understanding patterns of brain metastasis in breast cancer and designing rational therapeutic strategies. *Ann Transl Med* 2018;6:163.
9. Largillier R, Ferrero JM, Doyen J, et al. Prognostic factors in 1,038 women with metastatic breast cancer. *Ann Oncol* 2008;19:2012-9.
10. Waks AG, Winer EP. Breast Cancer Treatment: A Review. *JAMA* 2019;321:288-300.
11. Guo LW, Jiang LM, Gong Y, et al. Development and validation of nomograms for predicting overall and breast cancer-specific survival among patients with triple-negative breast cancer. *Cancer Manag Res* 2018;10:5881-94.
12. Xiong Y, Cao H, Zhang Y, et al. Nomogram-Predicted Survival of Breast Cancer Brain Metastasis: a SEER-Based Population Study. *World Neurosurg* 2019;128:e823-34.
13. Harrell FE Jr, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 1996;15:361-87.
14. Parkes A, Warneke CL, Clifton K, et al. Prognostic Factors in Patients with Metastatic Breast Cancer with Bone-Only Metastases. *Oncologist* 2018;23:1282-8.
15. Smid M, Wang Y, Zhang Y, et al. Subtypes of breast cancer show preferential site of relapse. *Cancer Res* 2008;68:3108-14.
16. Adam R, Aloia T, Krissat J, et al. Is Liver Resection Justified for Patients With Hepatic Metastases From Breast Cancer? *Annals of Surgery* 2006;244:897-908.
17. Martin AM, Cagney DN, Catalano PJ, et al. Brain Metastases in Newly Diagnosed Breast Cancer: A Population-Based Study. *JAMA Oncol* 2017;3:1069-77.
18. Liedtke C, Rody A, Gluz O, et al. The prognostic impact of age in different molecular subtypes of breast cancer. *Breast Cancer Res Treat* 2015;152:667-73.
19. Lodi M, Scheer L, Reix N, et al. Breast cancer in elderly women and altered clinico-pathological characteristics: a systematic review. *Breast Cancer Res Treat* 2017;166:657-68.
20. Buja A, Lago L, Lago S, et al. Marital status and stage of cancer at diagnosis: A systematic review. *Eur J Cancer Care (Engl)* 2018. doi: 10.1111/ecc.12755.
21. Warner ET, Tamimi RM, Hughes ME, et al. Racial and Ethnic Differences in Breast Cancer Survival: Mediating Effect of Tumor Characteristics and Sociodemographic and Treatment Factors. *J Clin Oncol* 2015;33:2254-61.
22. Newman LA, Griffith KA, Jatoi I, et al. Meta-analysis of survival in African American and white American patients with breast cancer: ethnicity compared with socioeconomic status. *J Clin Oncol* 2006;24:1342-9.
23. Balachandran VP, Gonen M, Smith JJ, et al. Nomograms in oncology: more than meets the eye. *Lancet Oncol* 2015;16:e173-80.

Cite this article as: Chu J, Yang D, Wang L, Xia J. Nomograms predicting survival for all four subtypes of breast cancer: a SEER-based population study. *Ann Transl Med* 2020;8(8):544. doi: 10.21037/atm-20-2808