

Development and Validation of Nomograms for Predicting Overall and Breast Cancer–Specific Survival in Young Women with Breast Cancer: A Population-Based Study



Yue Gong^{*,†,1}, Peng Ji^{*,†,1}, Wei Sun^{*},
Yi-Zhou Jiang^{*}, Xin Hu^{*} and Zhi-Ming Shao^{*,†,‡}

^{*}Department of Breast Surgery, Key Laboratory of Breast Cancer in Shanghai, Fudan University Shanghai Cancer Center, Fudan University, Shanghai, 200032, China;

[†]Department of Oncology, Shanghai Medical College, Fudan University, Shanghai, 200032, China; [‡]Institutes of Biomedical Science, Fudan University, Shanghai, 200032, China

Abstract

INTRODUCTION: The objective of current study was to develop and validate comprehensive nomograms for predicting the survival of young women with breast cancer. **METHODS:** Women aged <40 years diagnosed with invasive breast cancer between 1990 and 2010 were selected from the Surveillance, Epidemiology, and End Results database and randomly divided into training ($n = 12,465$) and validation ($n = 12,424$) cohorts. A competing-risks model was used to estimate the probability of breast cancer–specific survival (BCSS). We identified and integrated significant prognostic factors for overall survival (OS) and BCSS to construct nomograms. The performance of the nomograms was assessed with respect to calibration, discrimination, and risk group stratification. **RESULTS:** The entire cohort comprised 24,889 patients. The 5- and 10-year probabilities of breast cancer–specific mortality were 11.6% and 20.5%, respectively. Eight independent prognostic factors for both OS and BCSS were identified and integrated for the construction of the nomograms. The calibration curves showed optimal agreement between the predicted and observed probabilities. The *C*-indexes of the nomograms in the training cohort were higher than those of the TNM staging system for predicting OS (0.724 vs 0.694; $P < .001$) and BCSS (0.733 vs 0.702; $P < .001$). Additionally, significant differences in survival were observed in patients stratified into different risk groups within respective TNM categories. **CONCLUSIONS:** We developed and validated novel nomograms that can accurately predict OS and BCSS in young women with breast cancer. These nomograms may help clinicians in making decisions on an individualized basis.

Translational Oncology (2018) 11, 1334–1342

Introduction

Breast cancer is the most frequently diagnosed malignancy and the leading cause of cancer death among women worldwide [1]. However, breast cancer is rare in young women, and approximately 7% of all breast cancers are diagnosed in women under 40 years of age [2]. The incidence of breast cancer in women younger than 40 years has been stable for the past 20 years in most countries [3]. Some studies have demonstrated that breast cancer in younger women is correlated with a more aggressive biology and poorer outcomes than breast cancer in older women [4–7]. Young women have relatively high proportions of estrogen receptor (ER)–negative and progesterone receptor (PR)–negative cancers, human epidermal growth factor 2 (HER2)–positive cancers, and high-grade cancers [8,9]. They are also more likely to be associated with a positive family history and TP53-positive tumors [10,11].

The benefit of chemotherapy in treating breast cancer in women younger than 50 years has been confirmed [12]. However, the answers to many questions remain unknown regarding the selection of therapeutic measures for young women with breast cancer, including whether all young women with breast cancer should receive

Address all correspondence to: Zhi-Ming Shao, MD, PhD or Xin Hu, Department of Breast Surgery, Fudan University Shanghai Cancer Center, 270 Dong-An Road, Shanghai, 200032, P.R. China.

E-mail: zhi_ming_shao@163.com

¹ Both Yue Gong and Peng Ji have contributed equally to this work.

Received 24 July 2018; Revised 18 August 2018; Accepted 20 August 2018

© 2018 The Authors. Published by Elsevier Inc. on behalf of Neoplasia Press, Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1936-5233/18

<https://doi.org/10.1016/j.tranon.2018.08.008>

chemotherapy and whether they are candidates for breast-conserving surgery. Therefore, it is important to divide patients into different risk subgroups that receive certain treatments.

Nomograms have been widely used to estimate a numeric probability of death or recurrence in each patient by combining important prognostic factors [13–15]. To the best of our knowledge, nomograms for predicting overall survival (OS) and breast cancer–specific survival (BCSS) of young women with breast cancer have not been reported. In this study, we aimed to construct comprehensive and practical nomograms for young women with breast cancer based on a large population from the Surveillance, Epidemiology, and End Results (SEER) database. In addition, we compared our nomograms with traditional TNM staging systems to determine their predictive preciseness.

Material and Methods

Study Population

Data for this study were obtained from the current SEER database, which consists of 18 population-based cancer registries. This database represents approximately 28% of the total population in the United States. SEER*Stat Version 8.3.4 (<http://www.seer.cancer.gov/seerstat>) from the National Cancer Institute was used to identify eligible patients [16].

We included female patients, aged 18 to 39 years, who had been diagnosed with breast cancer as a first primary malignancy between 1990 and 2010. Patients diagnosed before 1990 were not included because ER and PR status was not recorded in the SEER database until 1990. Additionally, to ensure adequate follow-up time, patients diagnosed after 2010 were not included. Only histologically confirmed unilateral breast cancer cases were included. Cases diagnosed at autopsy or by death certificate only were excluded. All variables included in the analysis have less than 10% missing values. Other exclusion criteria for this study included patients with unknown race information, unknown specific surgical treatment including mastectomy and breast-conserving surgery, unknown histological grade or grade IV disease, unknown tumor size and number of positive lymph nodes, stage IV breast cancer, unknown ER and PR status, diagnosis with inflammatory breast cancer and Pager's disease, and incomplete survival data. After the exclusion criteria were applied, a total of 24,889 women were eventually eligible for analysis. The flow chart for data selection is shown in Supplementary Figure 1.

Construction of the Nomograms

To establish and validate competing-risks nomograms, the eligible patients were randomly divided into a training ($n = 12,465$) cohort and a validation ($n = 12,424$) cohort.

Race/ethnicity in SEER database was classified into four major groups, including white, black, Asian or Pacific Islander, and American Indian/Alaska Native. Given the small number of American Indian/Alaska Native patients, we incorporate these patients with Asian or Pacific Islander patients into "others" group. Thus, we classified race/ethnicity into three mutually exclusive groups of 1) white, 2) black, and 3) others.

The median follow-up was estimated as the median observed survival time. OS was calculated from the date of diagnosis to the date of death due to any cause, the date of last follow-up, or December 31, 2014. In the training cohort, univariate prognostic factors were determined using the Kaplan-Meier plots and compared using the log-rank tests. Variables that achieved significance at $P < .05$ were

entered into the multivariable analysis via the Cox proportional-hazards model. The independent prognostic factors determined by the multivariate analysis were used to construct a nomogram for OS.

BCSS was measured as the time from the date of diagnosis to the date of death attributed to breast cancer, date of last follow-up, or December 31, 2014. Deaths from other causes were considered competing risks. We used the cumulative incidence function (CIF) to assess the probability of death. Gray's test was conducted to test the difference in CIF among groups [17]. A subdistribution analysis of competing risks was performed to construct a competing-risks model [18]. In the Cox regression model analyzing disease-specific regression, patients who died from other causes were considered as censored at the data of last follow-up. Thus, a nomogram was developed by the integration of associated risk factors to predict 5- and 10-year BCSS of young patients with breast cancer.

Validation and Calibration of the Nomograms

The nomograms were subjected to 1000 bootstrap resamples for internal validation of the training cohort and external validation with the validation cohort. The concordance index (C-index) between the predicted probability and response was used to assess the discrimination performance of the nomograms [19]. The value of the C-index ranges from 0.5 to 1.0, with 0.5 indicating a random chance and 1.0 indicating a perfectly corrected discrimination. Comparison of the C-index of two different models was based on previously described methods [20]. Calibration is the ability of a model to make unbiased estimates of outcome. Marginal estimate versus average predictive probability of the models was used to construct calibration curves. The predictions were expected to fall on a 45° diagonal line in a well-calibrated model.

Risk Group Stratification Based on the Nomogram beyond TNM Staging

In addition to numerically comparing the discrimination ability based on the C-index, we sought to illustrate the independent discrimination ability of the nomogram for OS beyond standard TNM staging. To this end, we determined cutoff values by evenly dividing patients in the training cohort into different risk groups within a certain TNM category according to the total risk scores (from highest to lowest) from the nomogram for OS prediction. These values were then applied to the validation cohort, and the respective Kaplan-Meier survival curves were delineated.

Statistical Analysis

All statistical analyses were performed using R software, version 3.4.0 (<http://www.r-project.org>) and SPSS software, version 22.0 (SPSS, Chicago, IL). The R packages *cmprsk* [21] and *rms* [22] were used for modeling and developing the nomograms. Two-sided P values less than .05 were considered statistically significant.

Results

Patient Characteristics

The entire cohort comprised 24,889 young women with histologically confirmed malignant breast cancer, with 12,465 patients in the training cohort and 12,424 patients in the validation cohort. The demographic and clinical characteristics of the study cohort are shown in Table 1. The majority of tumors were infiltrating

ductal carcinoma (84.1%), and most of the patients were non-Hispanic whites (73.1%). The median survival time was 99 months (interquartile range, 63-149 months). By the end of the last follow-up, 5501 patients (22.1%) had died, including 4897 patients (19.7%) who died from breast cancer and 604 patients (2.4%) who died from other causes.

Overall Survival

The results of the univariate and multivariate analyses are listed in Table 2. All variables except for laterality of breast cancer were significantly correlated with OS ($P < .001$ for all). The significant factors in the univariate analysis were subjected to a multivariate analysis based on a Cox proportional-hazards regression model. Race, histology, tumor grade, tumor size, number of positive lymph nodes, ER status, and surgery type were confirmed to be independently associated with OS ($P < .05$ for all).

Table 1. Demographic and Clinical Characteristics of the Study Cohort

Demographic and Clinical Characteristic	All Patients <i>N</i> = 24,889	Training Cohort <i>N</i> = 12,465	Validation Cohort <i>N</i> = 12,424
Year of diagnosis			
1990-1996	3294 (13.2%)	1676 (13.4%)	1618 (13.0%)
1997-2003	8270 (33.3%)	4197 (33.7%)	4073 (32.8%)
2004-2010	13,325 (53.5%)	6592 (52.9%)	6733 (54.2%)
Race			
White	18,202 (73.1%)	9128 (73.2%)	9074 (73.0%)
Black	3695 (14.9%)	1853 (14.9%)	1842 (14.9%)
Others*	2992 (12.0%)	1484 (11.9%)	1508 (12.1%)
Laterality			
Left	12,378 (49.7%)	6203 (49.8%)	6175 (49.7%)
Right	12,511 (50.3%)	6262 (50.2%)	6249 (50.3%)
Histology			
IDC	20,935 (84.1%)	10,511 (84.3%)	10,424 (83.9%)
ILC	1805 (7.3%)	903 (7.2%)	902 (7.3%)
Others†	2149 (8.6%)	1051 (8.4%)	1098 (8.8%)
Grade			
I	1866 (7.5%)	905 (7.3%)	961 (7.7%)
II	8288 (33.3%)	4178 (33.5%)	4110 (33.1%)
III	14,735 (59.2%)	7382 (59.2%)	7353 (59.2%)
Tumor size (cm)			
≤2	11,833 (47.5%)	5970 (47.9%)	5863 (47.2%)
2-5	10,577 (42.5%)	5275 (42.3%)	5302 (42.7%)
> 5	2479 (10.0%)	1220 (9.8%)	1259 (10.1%)
No. of positive LNs			
0	12,824 (51.5%)	6400 (51.3%)	6424 (51.7%)
1-3	7796 (31.3%)	3922 (31.5%)	3874 (31.2%)
4-9	2875 (11.6%)	1449 (11.6%)	1426 (11.5%)
≥ 10	1394 (5.6%)	694 (5.6%)	700 (5.6%)
ER status			
Positive	15,746 (63.3%)	7858 (63.0%)	7888 (63.5%)
Negative	9143 (36.7%)	4607 (37.0%)	4536 (36.5%)
PR status			
Positive	14,215 (57.1%)	7161 (57.4%)	7054 (56.8%)
Negative	10,674 (42.9%)	5304 (42.6%)	5370 (43.2%)
Surgery			
BCS	11,119 (44.7%)	5500 (44.1%)	5619 (45.2%)
Mastectomy	13,770 (55.3%)	6965 (55.9%)	6805 (54.8%)
Survival months			
Median (IQR)	99 (63-149)	99 (63-150)	99 (63-148)

Abbreviations: BCS, breast-conserving surgery; ER, estrogen receptor; IDC, infiltrating ductal carcinoma; ILC, infiltrating lobular carcinoma; IQR, interquartile range; LN, lymph node; PR, progesterone receptor.

* Including American Indian/Alaskan native and Asian/Pacific Islander.
† Including other histology of invasive breast cancer except IDC and ILC.

Table 2. Univariate and Multivariate Analysis of OS in the Training Cohort

Variable	Univariate Analysis	Multivariate Analysis		
	<i>P</i> Value	HR	95% CI	<i>P</i> Value
Race	<.001	Reference		<.001
White		Reference		
Black		1.555	1.416-1.707	<.001
Others*		1.040	0.921-1.175	.526
Laterality	.973			
Left				
Right				
Histology	<.001	Reference		<.001
IDC		Reference		
ILC		1.071	0.930-1.233	.342
Others†		0.741	0.634-0.866	<.001
Grade	<.001	Reference		<.001
I		Reference		
II		1.648	1.304-2.083	<.001
III		1.986	1.574-2.507	<.001
Tumor size (cm)	<.001	Reference		<.001
≤2		Reference		
2-5		1.383	1.267-1.510	<.001
>5		1.857	1.646-2.095	<.001
No. of positive LNs	<.001	Reference		<.001
0		Reference		
1-3		1.806	1.641-1.987	<.001
4-9		3.195	2.859-3.571	<.001
≥10		5.361	4.725-6.084	<.001
ER status	<.001	Reference		<.001
Positive		Reference		
Negative		1.221	1.088-1.370	<.001
PR status	<.001	Reference		.425
Positive		Reference		
Negative		1.046	0.936-1.170	.425
Surgery	<.001	Reference		.004
BCS		Reference		
Mastectomy		1.129	1.040-1.226	.004

Abbreviation: HR, hazard ratio.

* Including American Indian/Alaskan native and Asian/Pacific Islander.

† Including other histology of invasive breast cancer except IDC and ILC.

Breast Cancer–Specific Survival

Estimates of probabilities of death resulting from breast cancer and other causes according to clinical characteristics are listed in Table 3. The 5- and 10-year probabilities of death from breast cancer were 11.6% and 20.5%, respectively, while the 5- and 10-year cumulative incidences of death from other causes were 1.1% and 2.6%, respectively. Young black patients exhibited higher cumulative incidence of death than white and “other” patients ($P < .001$ for all outcomes). There was no significant difference between different lateralitys. Tumor grade, tumor size, number of positive lymph nodes, and surgery type were significantly associated with probabilities of death ($P < .05$ for all outcomes). Infiltrating ductal carcinoma and infiltrating lobular carcinoma, negative ER status, and negative PR status were associated with a significantly higher cumulative incidence of death only among patients who died of breast cancer ($P < .001$). All variables significantly correlated with cumulative incidences of death resulting from breast cancer were used to construct the nomogram to predict 5- and 10-year BCSS.

Construction of the Nomograms

Nomograms were constructed based on the Cox regression model to predict 5- and 10-year OS and BCSS (Figure 1). The point assignment of nomograms for OS and BCSS is shown in Supplementary Table 1. Based on the nomograms, tumor grade, tumor size, and number of positive

Table 3. Five- and Ten-year Cumulative Incidences of Death Among Patients in the Training Cohort

Variable	Cumulative Incidence of Death Resulting From Breast Cancer			Cumulative Incidence of Death Resulting From Other Causes		
	5 y	10 y	P Value	5 y	10 y	P Value
All patients	0.116	0.205		0.011	0.026	
Race			<0.001			<0.001
White	0.105	0.192		0.009	0.022	
Black	0.185	0.295		0.020	0.042	
Others*	0.099	0.171		0.012	0.032	
Laterality			0.881			0.753
Left	0.115	0.202		0.011	0.025	
Right	0.118	0.208		0.010	0.026	
Histology			<0.001			0.602
IDC	0.120	0.208		0.011	0.026	
ILC	0.111	0.245		0.011	0.030	
Others†	0.087	0.141		0.013	0.018	
Grade			<0.001			<0.001
I	0.015	0.051		0.009	0.022	
II	0.069	0.170		0.005	0.018	
III	0.156	0.243		0.014	0.031	
Tumor size (cm)			<0.001			<0.001
≤2	0.059	0.123		0.009	0.022	
2-5	0.148	0.256		0.010	0.026	
>5	0.269	0.401		0.027	0.050	
No. of positive LNs			<0.001			<0.001
0	0.057	0.112		0.006	0.017	
1-3	0.117	0.214		0.013	0.028	
4-9	0.245	0.399		0.020	0.039	
≥10	0.397	0.604		0.033	0.091	
ER status			<0.001			0.085
Positive	0.079	0.190		0.008	0.024	
Negative	0.181	0.232		0.015	0.028	
PR status			<0.001			0.106
Positive	0.079	0.186		0.009	0.024	
Negative	0.167	0.231		0.014	0.028	
Surgery			<0.001			0.009
BCS	0.084	0.151		0.009	0.019	
Mastectomy	0.142	0.251		0.012	0.032	

* Including American Indian/Alaskan native and Asian/Pacific Islander.
 † Including other histology of invasive breast cancer except IDC and ILC.

lymph nodes were sharing the largest contribution to prognosis, followed by race and histology. By adding up all points and locating them on the bottom scales, we were easily able to calculate the estimated 5- and 10-year survival probabilities.

Calibration and Validation of the Nomograms

The calibration plots for the OS and BCSS nomograms in the training cohort (Supplementary Figure 2) and validation cohort (Figure 2) demonstrated an acceptable agreement between the nomogram prediction and observed estimates for 5- and 10-year OS and BCSS. As shown in Supplementary Table 2, in the training cohort, the Harrell's C-indexes of the nomograms for the prediction of OS and BCSS were 0.724 [95% confidence interval (CI), 0.714-0.733] and 0.733 (95% CI, 0.723-0.743), respectively, which were significantly higher than those of the TNM staging system for OS (0.694; 95% CI, 0.684-0.704; $P < .001$) and BCSS (0.702; 95% CI, 0.692-0.713; $P < .001$). The C-indexes for the nomogram were similar in the validation cohort: 0.722 (95% CI, 0.712-0.732) for OS and 0.733 (95% CI, 0.723-0.743) for BCSS. Additionally, C-indexes were significantly greater than those of the TNM staging system at 0.699 (95% CI, 0.689-0.709) and 0.710 (95% CI, 0.700-0.720) for OS and BCSS, respectively.

Performance of the Nomograms in Stratifying Patients According to Risk Scores

We calculated the total points of OS nomogram for every patient in the training cohort and determined the cutoff values by dividing the patients evenly into three subgroups based on total score (0 to 78, 79 to 116, and ≥117). Supplementary Table 3 and Supplementary Figure 3 show that the low-risk subgroup had the best prognosis and the high-risk subgroup had the worst survival. Furthermore, in the validation cohort, patients stratified into different risk subgroups based on cutoff values within each TNM category also exhibited significant differences in survival (Figure 3).

Discussion

Breast cancer in young women has several characteristics that differentiate them from breast cancer in other population [3]. Although several nomograms have been previously reported to predict prognoses in some specific subtypes of breast cancer, no comprehensive nomogram has been developed for young patients with breast cancer [23–25]. In this study, we developed and validated nomograms to predict 5- and 10-year OS and BCSS for breast cancer in young women. Because the SEER database represents approximately 28% of the US population, the nomograms we developed are highly generalizable and provide personalized estimates of OS and BCSS that can be used by patients and clinicians in making personalized treatment decisions and designing clinical studies.

Although most young women with breast cancer experience breast cancer–associated mortality; some of these patients die from other cancers or noncancer causes. Non–breast cancer-related death might preclude the possibility of death resulting from breast cancer, and censoring those events might lead to biased results [26,27]. Therefore, we introduced a competing-risks model in this study. Competing-risks models have been published in recent years for predicting prognoses in thyroid cancer, breast cancer, prostate cancer, and localized renal cell carcinoma [23,25,28–30]. In this study, the 5- and 10-year probabilities of death were 12.7% and 23.1%, respectively. In addition, 5- and 10-year cumulative incidences of death resulting from breast cancer were 11.6% and 20.5%, respectively, indicating a nearly eight-fold higher risk of death from breast cancer than from other causes.

Using log-rank tests, Cox proportional-hazards regression analyses, and competing-risks model, we identified race, histology, tumor grade, tumor size, number of positive lymph nodes, ER status, PR status, and surgery type as independent prognostic factors for both OS and BCSS. These findings were highly concordant with the results of previous studies [31–35]. Previous data have highlighted that young black women have a higher risk of probability of death than young white women [35,36]. Our study confirmed that, after adjustment for other risk factors identified for breast cancer, young white patients have a better OS and BCSS than young black patients. However, there are still some other prognostic markers and molecular profiles that the SEER database did not offer that could be used to predict the survival of breast cancer patients. According to the AJCC 8th edition, HER2 status and multigene panel (such as Oncotype DX) status should also be considered as biology factors that affect the prognosis of breast cancer [37]. Furthermore, a higher number of young patients with breast cancer carry a pathogenic BRCA1 or BRCA2 mutation compared with patients with onset of breast cancer at an older age [38,39]. The cumulative risk of developing breast cancer is relatively high for BRCA1 or BRCA2 carriers [40]. Although

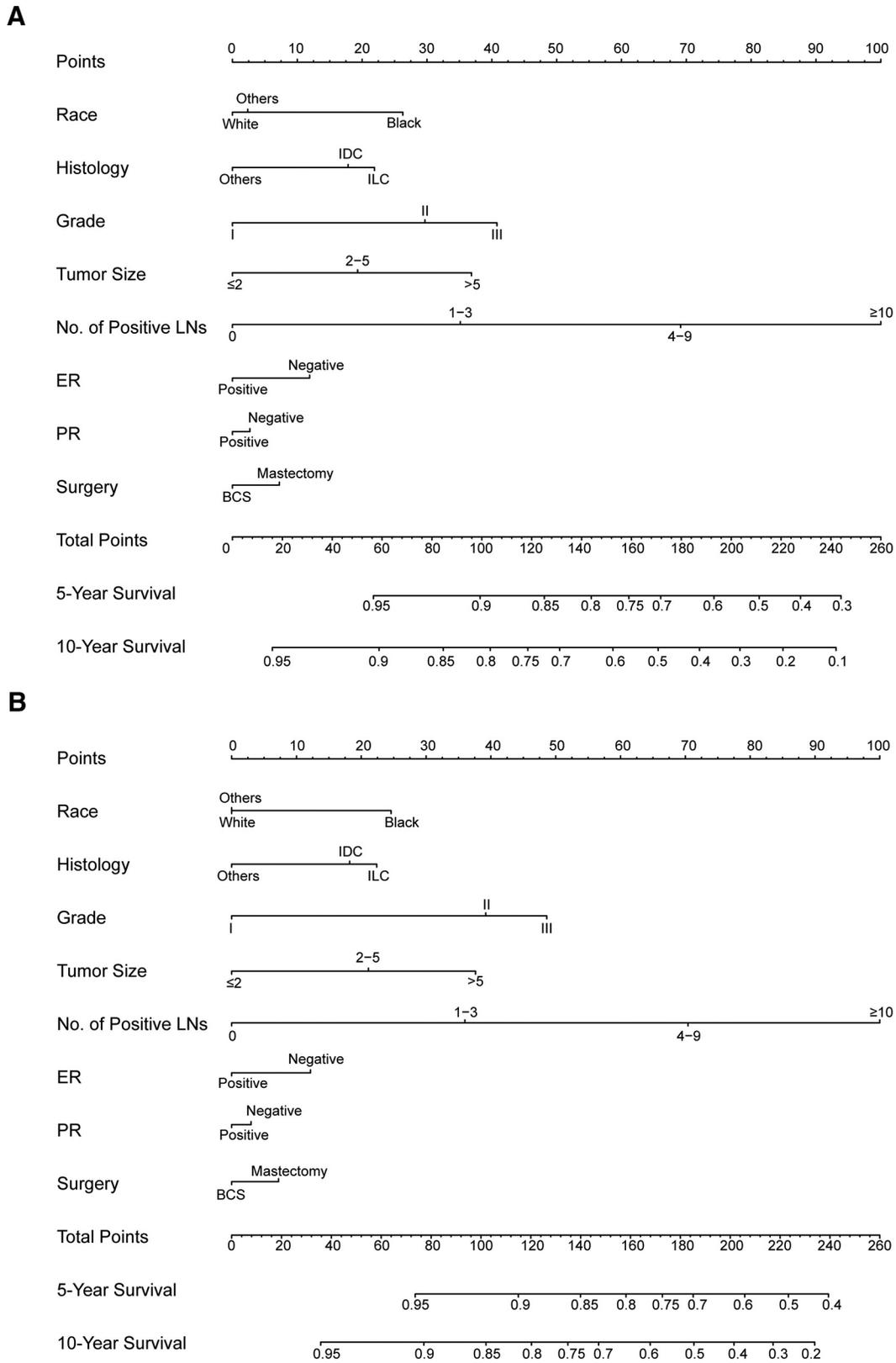


Figure 1. Nomogram for predicting 5- and 10-year probabilities of (A) OS and (B) BCSS of breast cancer in young women. Draw a vertical straight line from the variable value to the axis labeled “Points” to identify points for each variable. Add up all points, and the total points projected on the bottom scales correspond to the 5- and 10-year survival. Abbreviations: *BCS*, breast-conserving surgery; *ER*, estrogen receptor; *IDC*, infiltrating ductal carcinoma; *ILC*, infiltrating lobular carcinoma; *LN*, lymph node.

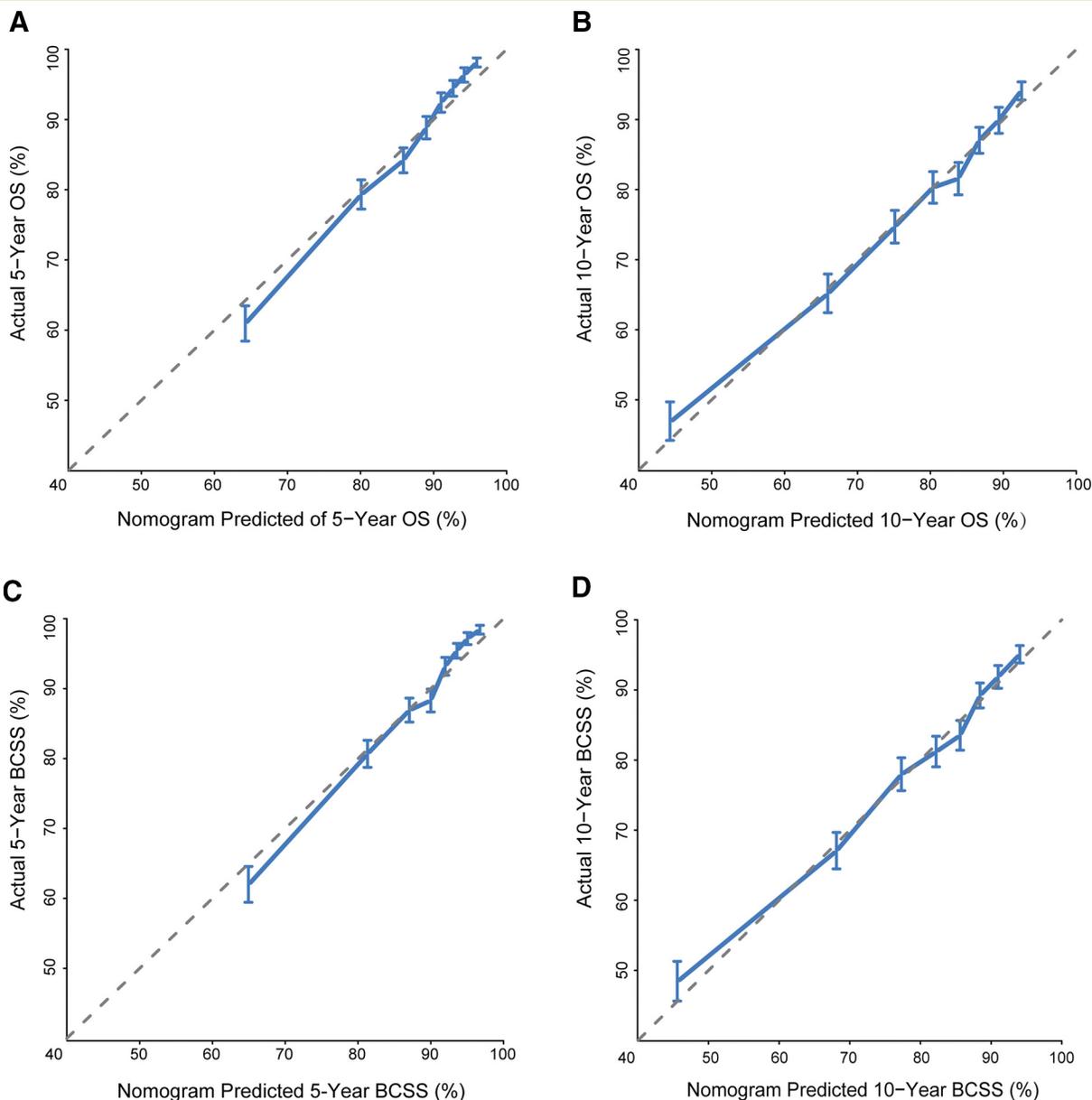


Figure 2. Calibration curves for predicting (A) 5-year and (B) 10-year OS and (C) 5-year and (D) 10-year BCSS in the validation cohort. Nomogram-predicted survival is plotted on the x-axis, and actual survival is plotted on the y-axis. Vertical bars represent 95% CIs measured by Kaplan-Meier analysis. Dashed lines along the 45° line through the origin point represent a perfect calibration model.

whether a germline BRCA1 or BRCA2 mutation has independent prognostic implications after an initial cancer diagnosis is unclear, genetic factor should be considered when applying the nomograms.

In addition, adjuvant therapies including chemotherapy and radiotherapy were not selected as candidate factors due to the lack of complete data for treatment history in the SEER database. Thus, it is difficult to accurately distinguish between the categories “no treatment” and “unknown if patients received treatment.” Another reason for not selecting treatment as candidate factor is that adjuvant therapies are recommended for patients who have a potentially high risk for disease recurrence or death. Thus, if we include adjuvant therapies into the nomograms, it might result in a certain degree of bias.

In our study, calibration plots showed optimal agreement between predicted and actual probabilities of 5- and 10-year OS and BCSS, thereby demonstrating the reliability of the established nomograms.

The C-indexes of our nomograms for OS and BCSS were significantly higher than those for the TNM staging system in both training and validation cohorts, demonstrating good discrimination power. We also separated patients in both cohorts with distinct survival outcomes by stratifying them into three risk groups using total prognostic score. We believe that the identification of subgroups of patients at different risks might have an effect on treatment or care option.

Nevertheless, several limitations should be considered while interpreting our results. First, we excluded a proportion of patients because of missing data for some important variables such as tumor grade, tumor size, and ER and PR status. This might have resulted in some bias in our models. Second, genetic factors and some prognostic parameters including HER2 status, multigene panel status, Ki-67 positivity, body mass index, and smoking status were not recorded in

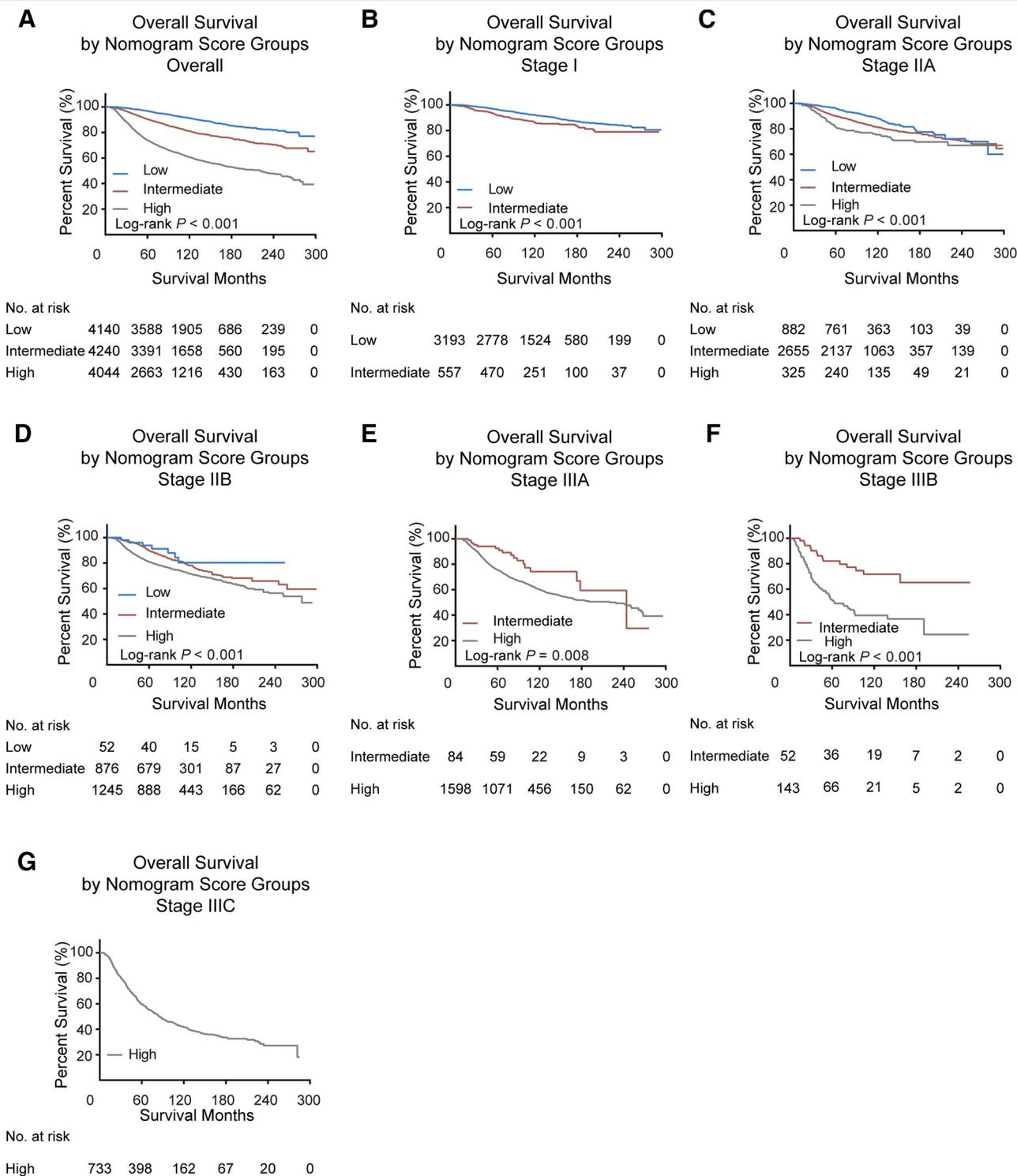


Figure 3. Kaplan-Meier curves for overall survival within each TNM stage (A, all patients; B-G, stage I-III C) according to risk group stratification in the validation cohort. Subgroups with fewer than 20 patients were omitted from the graphs.

the SEER database between 1990 and 2010, but these factors might improve the robustness and effectiveness of the nomograms [41–45]. Third, the long duration of our study period (1990–2010) may affect the results due to the change of therapeutic strategies, including the establishment of breast-conserving surgery and sentinel lymph node biopsy, improvement of chemotherapy, and application of endocrine therapy and targeted therapy. Although information on radiation

therapy and chemotherapy could be accessed from SEER database, they were not recommended for the analysis of survival due to the incompleteness of the variables and biases associated with who receives treatment according to the SEER program. Fourth, young age at diagnosis is a risk factor for recurrence [46,47]. However, the SEER database does not provide information on disease recurrence; thus, we were unable to determine an individualized estimate of the

risk of recurrence. Fifth, our models are limited by the retrospective nature of data collection, and thus, these nomograms must be further validated in a prospective cohort before being applied for clinical use.

Conclusion

Using a larger, population-based cohort, we established and validated novel nomograms for predicting the probability of OS and BCSS in young patients with breast cancer. Our developed nomograms perform excellently in both training and validation cohorts. Thus, these nomograms can assist clinicians to precisely estimate the survival of individuals and to identify patients at high risk of death who need more individualized and specialized treatment strategy.

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.tranon.2018.08.008>.

Declarations

Ethics Approval and Consent to Participate

Our study was approved by Shanghai Cancer Center Ethical Committee. Because cancer is a reportable disease in every state in the United States, informed patient consent is not required for the data released by the SEER database.

Consent for Publication

Not applicable.

Availability of Data and Material

The datasets generated and analyzed during the current study are available from Surveillance, Epidemiology, and End Results (SEER) Program (www.seer.cancer.gov) SEER*Stat Database: Incidence-SEER 18 Regs Research Data + Hurricane Katrina Impacted Louisiana Cases, Nov 2016 Sub (1973-2014 varying), National Cancer Institute, DCCPS, Surveillance Research Program, released April 2017, based on the November 2016 submission.

Competing Interests

The authors declare no competing financial interests.

Funding

This study was supported by a grant from the Ministry of Science and Technology of China (MOST2016YFC0900300, National Key R&D Program of China), grants from the National Natural Science Foundation of China (81672601, 81602311), and grants from the Shanghai Committee of Science and Technology Funds (15410724000, 15411953300). The funders had no role in the study design, collection and analysis of the data, decision to publish, or manuscript preparation.

Authors' Contributions

Conception and design: Y. G., P. J., X. H., and Z. M. S. Development of methodology: Y. G., P. J., WS, X. H., and Z. M. S. Acquisition of data: Y. G. and P. J.. Analysis and interpretation of data: Y. G., P. J., X. H., and Z. M. S. Writing, review, and/or revision of manuscript: Y. G., P. J., Y. Z. J., X. H., and Z. M. S. Study supervision: X. H. and Z. M. S. All authors read and approved the final manuscript.

Acknowledgements

We would like to thank SEER for providing open access to the database.

References

- [1] Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, and Jemal A (2015). Global cancer statistics, 2012. *CA Cancer J Clin* **65**(2), 87–108.
- [2] Brinton L, Sherman M, Carreon J, and Anderson W (2008). Recent trends in breast cancer among younger women in the United States. *J Natl Cancer Inst* **100**(22), 1643–1648.
- [3] Narod SA (2012). Breast cancer in young women. *Nat Rev Clin Oncol* **9**(8), 460–470.
- [4] Anders C, Hsu D, Broadwater G, Acharya C, Foekens J, Zhang Y, Wang Y, Marcom P, Marks J, and Febbo P, et al (2008). Young age at diagnosis correlates with worse prognosis and defines a subset of breast cancers with shared patterns of gene expression. *J Clin Oncol* **26**(20), 3324–3330.
- [5] Bleyer A, Barr R, Hayes-Lattin B, Thomas D, Ellis C, and Anderson B (2008). The distinctive biology of cancer in adolescents and young adults. *Nat Rev Cancer* **8**(4), 288–298.
- [6] Bharat A, Aft R, Gao F, and Margenthaler J (2009). Patient and tumor characteristics associated with increased mortality in young women (< or =40 years) with breast cancer. *J Surg Oncol* **100**(3), 248–251.
- [7] Azim H, Michiels S, Bedard P, Singhal S, Criscitiello C, Ignatiadis M, Haibe-Kains B, Piccart M, Sotiriou C, and Loi S (2012). Elucidating prognosis and biology of breast cancer arising in young women using gene expression profiling. *Clin Cancer Res* **18**(5), 1341–1351.
- [8] Collins LC, Marotti JD, Gelber S, Cole K, Ruddy K, Kereakoglow S, Brachtel EF, Schapira L, Come SE, and Winer EP, et al (2012). Pathologic features and molecular phenotype by patient age in a large cohort of young women with breast cancer. *Breast Cancer Res Treat* **131**(3), 1061–1066.
- [9] Lund M, Butler E, Hair B, Ward K, Andrews J, Oprea-Ilie G, Bayakly A, O'Regan R, Vertino P, and Eley J (2010). Age/race differences in HER2 testing and in incidence rates for breast cancer triple subtypes: a population-based study and first report. *Cancer* **116**(11), 2549–2559.
- [10] Althuis M, Brogan D, Daling J, Gammon M, Malone K, Schoenberg J, and Brinton L (2003). Breast cancers among very young premenopausal women (United States). *Cancer Causes Control* **14**(2), 151–160.
- [11] Mouchawar J, Korch C, Byers T, Pitts T, Li E, McCredie M, Giles G, Hopper J, and Southey M (2010). Population-based estimate of the contribution of TP53 mutations to subgroups of early-onset breast cancer: Australian Breast Cancer Family Study. *Cancer Res* **70**(12), 4795–4800.
- [12] Peto R, Davies C, Godwin J, Gray R, Pan H, Clarke M, Cutter D, Darby S, McGale P, and Taylor C, et al (2012). Comparisons between different polychemotherapy regimens for early breast cancer: meta-analyses of long-term outcome among 100,000 women in 123 randomised trials. *Lancet* **379**(9814), 432–444.
- [13] Liang W, Zhang L, Jiang G, Wang Q, Liu L, Liu D, Wang Z, Zhu Z, Deng Q, and Xiong X, et al (2015). Development and validation of a nomogram for predicting survival in patients with resected non-small-cell lung cancer. *J Clin Oncol* **33**(8), 861–869.
- [14] Callegaro D, Miceli R, Bonvalot S, Ferguson P, Strauss DC, Levy A, Griffin A, Hayes AJ, Stacchiotti S, and Pechoux CL, et al (2016). Development and external validation of two nomograms to predict overall survival and occurrence of distant metastases in adults after surgical resection of localised soft-tissue sarcomas of the extremities: a retrospective analysis. *Lancet Oncol* **17**(5), 671–680.
- [15] Fakhry C, Zhang Q, Nguyen-Tân P, Rosenthal D, Weber R, Lambert L, Trotti A, Barrett W, Thorstad W, and Jones C, et al (2017). Development and validation of nomograms predictive of overall and progression-free survival in patients with oropharyngeal cancer. *J Clin Oncol* **35**(36), 4057–4065.
- [16] National Cancer Institute. Surveillance, Epidemiology, and End Results (SEER) Program (www.seer.cancer.gov) SEER*Stat Database, National Cancer Institute, DCCPS, Surveillance Research Program, released April 2017, based on the November 2016 submission.
- [17] Gray RJ (1988). A class of k-sample tests for comparing the cumulative incidence of a competing risk. *Ann Stat* **16**, 1141–1154.
- [18] Fine JP and Gray RJ (1999). A proportional hazards model for the subdistribution of a competing risk. *J Am Stat Assoc* **94**(446), 496–509.
- [19] Wolbers M, Koller MT, Witteman JC, and Steyerberg EW (2009). Prognostic models with competing risks: methods and application to coronary risk prediction. *Epidemiology* **20**(4), 555–561.
- [20] Pencina MJ, D'Agostino RB, D'Agostino RB, and Vasan RS (2008). Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med* **27**(2), 157–172 [discussion 207–112].

- [21] Gray B (). cmprsk: Subdistribution Analysis of Competing Risks. R package version 2.2–7. <http://CRAN.R-project.org/package=cmprsk>.
- [22] Frank E and Harrell J (). rms: Regression Modeling Strategies. R Package version 5.1–1. <http://CRAN.R-project.org/package=rms>.
- [23] Sun W, Jiang YZ, Liu YR, Ma D, and Shao ZM (2016). Nomograms to estimate long-term overall survival and breast cancer-specific survival of patients with luminal breast cancer. *Oncotarget* **7**(14), 20496–20506.
- [24] Mazouni C, Spyrtos F, Romain S, Fina F, Bonnier P, Ouafik LH, and Martin PM (2012). A nomogram to predict individual prognosis in node-negative breast carcinoma. *Eur J Cancer* **48**(16), 2954–2961.
- [25] Hanrahan EO, Gonzalez-Angulo AM, Giordano SH, Rouzier R, Broglio KR, Hortobagyi GN, and Valero V (2007). *J Clin Oncol* **25**(31), 4952–4960.
- [26] Gooley T, Leisenring W, Crowley J, and Storer B (1999). Estimation of failure probabilities in the presence of competing risks: new representations of old estimators. *Stat Med* **18**(6), 695–706.
- [27] Southern D, Faris P, Brant R, Galbraith P, Norris C, Knudtson M, and Ghali W (2006). Kaplan-Meier methods yielded misleading results in competing risk scenarios. *J Clin Epidemiol* **59**(10), 1110–1114.
- [28] Yang L, Shen W, and Sakamoto N (2013). Population-based study evaluating and predicting the probability of death resulting from thyroid cancer and other causes among patients with thyroid cancer. *J Clin Oncol* **31**(4), 468–474.
- [29] Stephenson A, Kattan M, Eastham J, Bianco F, Yossepowitch O, Vickers A, Klein E, Wood D, and Scardino P (2009). Prostate cancer-specific mortality after radical prostatectomy for patients treated in the prostate-specific antigen era. *J Clin Oncol* **27**(26), 4300–4305.
- [30] Kutikov A, Egleston B, Wong Y, and Uzzo R (2010). Evaluating overall survival and competing risks of death in patients with localized renal cell carcinoma using a comprehensive nomogram. *J Clin Oncol* **28**(2), 311–317.
- [31] Saadatmand S, Bretveld R, Siesling S, and Tilanus-Linthorst M (2015). Influence of tumour stage at breast cancer detection on survival in modern times: population based study in 173,797 patients. *BMJ* **351**, h4901.
- [32] Colzani E, Liljegren A, Johansson A, Adolfsson J, Hellborg H, Hall P, and Czene K (2011). Prognosis of patients with breast cancer: causes of death and effects of time since diagnosis, age, and tumor characteristics. *J Clin Oncol* **29**(30), 4014–4021.
- [33] Pestalozzi BC, Zahrieh D, Mallon E, Gusterson BA, Price KN, Gelber RD, Holmberg SB, Lindtner J, Snyder R, and Thürlimann B, et al (2008). Distinct clinical and prognostic features of infiltrating lobular carcinoma of the breast: combined results of 15 international breast cancer study group clinical trials. *J Clin Oncol* **26**(18), 3006–3014.
- [34] Dunnwald L, Rossing M, and Li C (2007). Hormone receptor status, tumor characteristics, and prognosis: a prospective cohort of breast cancer patients. *Breast Cancer Res* **9**(1), R6.
- [35] Ademuyiwa FO, Gao F, Hao L, Morgensztern D, Aft RL, Ma CX, and Ellis MJ (2015). US breast cancer mortality trends in young women according to race. *Cancer* **121**(9), 1469–1476.
- [36] Liu P, Li X, Mittendorf EA, Li J, Du XL, He J, Ren Y, Yang J, Hunt KK, and Yi M (2013). Comparison of clinicopathologic features and survival in young American women aged 18–39 years in different ethnic groups with breast cancer. *Br J Cancer* **109**(5), 1302–1309.
- [37] Giuliano AE, Connolly JL, Edge SB, Mittendorf EA, Rugo HS, Solin LJ, Weaver DL, Winchester DJ, and Hortobagyi GN (2017). Breast cancer—major changes in the American Joint Committee on Cancer eighth edition cancer staging manual. *CA Cancer J Clin* **67**(4), 290–303.
- [38] Malone K, Daling J, Doody D, Hsu L, Bernstein L, Coates R, Marchbanks P, Simon M, McDonald J, and Norman S, et al (2006). Prevalence and predictors of BRCA1 and BRCA2 mutations in a population-based study of breast cancer in white and black American women ages 35 to 64 years. *Cancer Res* **66**(16), 8297–8308.
- [39] Anglian Breast Cancer Study Group (2000). Prevalence and penetrance of BRCA1 and BRCA2 mutations in a population-based series of breast cancer cases. *Br J Cancer* **83**(10), 1301–1308.
- [40] Kuchenbaecker K, Hopper J, Barnes D, Phillips K, Mooij T, Roos-Blom M, Jervis S, Leeuwen F, Milne R, and Andrieu N, et al (2017). Risks of breast, ovarian, and contralateral breast cancer for BRCA1 and BRCA2 mutation carriers. *JAMA* **317**(23), 2402–2416.
- [41] Chan DS, Vieira AR, Aune D, Bandera EV, Greenwood DC, McTiernan A, Navarro Rosenblatt D, Thune I, Vieira R, and Norat T (2014). Body mass index and survival in women with breast cancer-systematic literature review and meta-analysis of 82 follow-up studies. *Ann Oncol* **25**(10), 1901–1914.
- [42] Warren GW, Kasza KA, Reid ME, Cummings KM, and Marshall JR (2013). Smoking at diagnosis and survival in cancer patients. *Int J Cancer* **132**(2), 401–410.
- [43] Yerushalmi R, Woods R, Ravdin P, Hayes M, and Gelmon K (2010). Ki67 in breast cancer: prognostic and predictive potential. *Lancet Oncol* **11**(2), 174–183.
- [44] Slamon D, Clark G, Wong S, Levin W, Ullrich A, and McGuire W (1987). Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. *Science* **235**(4785), 177–182.
- [45] Copson E, Maishman T, Tapper W, Cutress R, Greville-Heygate S, Altman D, Eccles B, Gerty S, Durcan L, and Jones L, et al (2018). Germline BRCA mutation and outcome in young-onset breast cancer (POSH): a prospective cohort study. *Lancet Oncol* **19**(2), 169–180.
- [46] Bock G, Hage J, Putter H, Bonnema J, Bartelink H, and Velde C (2006). Isolated loco-regional recurrence of breast cancer is more common in young patients and following breast conserving therapy: long-term results of European Organisation for Research and Treatment of Cancer studies. *Eur J Cancer* **42**(3), 351–356.
- [47] Wapnir I, Anderson S, Mamounas E, Geyer C, Jeong J, Tan-Chiu E, Fisher B, and Wolmark N (2006). Prognosis after ipsilateral breast tumor recurrence and locoregional recurrences in five National Surgical Adjuvant Breast and Bowel Project node-positive adjuvant breast cancer trials. *J Clin Oncol* **24**(13), 2028–2037.