ORIGINAL RESEARCH

A Nomogram To Predict The Overall Survival Of Breast Cancer Patients And Guide ~~The~~ Postoperative Adjuvant Chemotherapy In China

Xin Wang, [iD][1],* Ziwei Feng,[1],*
Yubei Huang,[1] Haixin Li,[1,2]
Ping Cui,[1] Dezheng Wang,[3]
Hongji Dai,[1] Fangfang Song,[1]
Hong Zheng,[1]
Peishan Wang, [iD][1]
Xuchen Cao,[4] Lin Gu,[5]
Jin Zhang,[6] Fengju Song,[1]
Kexin Chen [iD][1]

[1]Department of Epidemiology and Biostatistics, National Clinical Research Center for Cancer, Key Laboratory of Cancer Prevention and Therapy, Tianjin Medical University Cancer Institute and Hospital, Tianjin 300060, People's Republic of China; [2]Department of Cancer Biobank, Tianjin Medical University Cancer Institute and Hospital, Tianjin 300060, People's Republic of China; [3]Center for Non-Communicable Disease Control and Prevention, Tianjin Centers for Disease Control and Prevention, Tianjin 300011, People's Republic of China; [4]The First Department of Breast Cancer, National Clinical Research Center for Cancer, Key Laboratory of Cancer Prevention and Therapy, Tianjin Medical University Cancer Institute and Hospital, Tianjin 300060, People's Republic of China; [5]The Second Department of Breast Cancer, National Clinical Research Center for Cancer, Key Laboratory of Cancer Prevention and Therapy, Tianjin Medical University Cancer Institute and Hospital, Tianjin 300060, People's Republic of China; [6]The Third Department of Breast Cancer, National Clinical Research Center for Cancer, Key Laboratory of Cancer Prevention and Therapy, Tianjin Medical University Cancer Institute and Hospital, Tianjin 300060, People's Republic of China

*These authors contributed equally to this work

Correspondence: Fengju Song;
Kexin Chen
Tianjin Medical University Cancer Institute and Hospital, Huanhu Xi Road, Tiyuan Bei, Hexi District, Tianjin 300060, People's Republic of China
Tel/Fax +86-22-23372231
Email songfengju@163.com;
chenkexin@tjmuch.com

Purpose: We aim to construct a nomogram to predict breast cancer survival and guide postoperative adjuvant chemotherapy in China.

Patients and methods: A total of 5,504 breast cancer patients from the Tianjin Breast Cancer Cases Cohort were included. Multivariable Cox regression was used to investigate the factors associated with overall survival (OS) and a nomogram was constructed based on these prognostic factors. The nomogram was ~~internal and external~~ validated and the performance was evaluated by area under the curve (AUC) and calibration curve. The partial score was also constructed and stratified them into low, moderate and high-risk subgroups for death according to the tripartite grouping method. Multivariate Cox regression analysis and the propensity score matching method were respectively used to test the association between adjuvant chemotherapy and OS in different risk subgroups.

Results: Age, diameter, histological differentiation, lymph node metastasis, estrogen, and progesterone receptor were incorporated into the nomogram and validation results showed this nomogram was well-calibrated to predict the 3-year [AUC =74.1%; 95% confidence interval (CI): 70.1–78.0%] and 5-year overall survival [AUC =72.3%; 95% CI: 69.6–75.1%]. Adjuvant chemotherapy was negatively associated with death in high risk subgroup [Hazard Ratio (HR) = 0.54; 95% CI: 0.37–0.77; *P*<0.001]. However, no significant association were found in groups with low (HR=1.47; 95% CI: 0.52–4.19; *P*=0.47) and moderate risk (HR=0.78; 95% CI: 0.42–1.48; *P*=0.45). The 1:1 PSM generated 822 pairs of well-matched patients and Kaplan-Meier showed the high-risk patients could benefit from chemotherapy, whereas low risk and moderate risk subjects did not appear to benefit from chemotherapy.

Conclusion: Not all ~~of the~~ breast cancer patients benefit equally from chemotherapy. The nomogram could be used to evaluate the overall survival of breast cancer patients and predict the magnitude of benefit and guide adjuvant chemotherapy for breast cancer patients after surgery.

Keywords: breast cancer, prognostic prediction, cohort study, adjuvant chemotherapy

Introduction

Breast cancer is the most common carcinoma in women with high morbidity, and its sharply increasing mortality is affecting women in developing countries, especially in China.[1–3] The development of novel systemic adjuvant therapies has contributed to significant progress in the management of breast cancer.[4,5] Adjuvant chemotherapy, which exhibits great improvement in prognosis for postoperative breast cancer

patients, remains as a frequently used treatment option in the adjuvant settings.[6–8] Survival prediction based on the patients' clinical characteristics such as age, tumor size, and hormone receptor status are very important in prescribing of chemotherapy regimens in clinical practice[9] Moreover, studies also showed not all patients benefit the same degree of cancer recurrence and death reduction from chemotherapy.[10–12] Accordingly, the selection of adjuvant chemotherapy should not only consider the clinical characteristics, but also the potential benefit of patients.[5,12,13] However, how to comprehensively utilize the aforementioned factors to predict breast cancer survival and provide beneficial postoperative adjuvant chemotherapy guidelines for patients remains a complex issue.

A number of models tried to assess breast cancer patient's survival using clinical prognostic variable and gene expression profiling such as PREDICT[14] MammaPrint[15] OncotypeDX[10,16] and the MSKCC nomogram[17] however, few studies have been conducted in China until now. (Table S1) "Adjuvant! Online" is an another well-calibrated and widely used appraisal tool for estimating breast cancer survival and is also recommended by the National Comprehensive Cancer Network (NCCN) to guide postoperative adjuvant chemotherapy.[18–22] However, it was developed based on the Surveillance, Epidemiology, and End-Results (SEER) dataset, whose incidence of breast cancer much higher than that in Chinese women.[2,23] Moreover, the distributions of reproductive factors that may contribute to breast cancer development and prognosis were different between Asian and American women.[24,25] Additionally, validation studies also showed the Adjuvant! Online was overoptimistic in predicting breast cancer survival for patients in China.[26,27] Additionally, de Glas and colleagues also commented that the relatively younger patients included in the study and benefits of therapy were mainly evaluated based on clinical trials that with relatively few proportions of older patients, thus it could not be extrapolated to the general breast cancer patients[28] Therefore, Adjuvant! Online could not be applied for prognosis prediction in Chinese breast cancers and a more suitable prognostic model was warranted to develop and apply in Chinese clinical settings.

In the current study, we aim to construct a survival predictive score for predicting breast cancer OS based on a hospital-based breast cancer survival cohort in Tianjin, P.R. China, so as to assist the clinicians in guiding the implementation of individualized adjuvant chemotherapy regimen for breast cancer patients in China.

Materials And Methods

Study Population

Tianjin Breast Cancer Cases Cohort (TBCCC) was established in 2007. All the pathologically confirmed breast cancer patients who were diagnosed or received treatment in the Tianjin Medical University Cancer Institute and Hospital were included. The epidemiological, clinical and therapeutic information were collected by one professional full-time personnel using a structured face-to-face questionnaire. An experienced staff was employed for chasing the survival status once a year by telephone. Moreover, the database was annually updated by checking medical records and by linking to the Tianjin Cancer Death Registry System.

The external validation cohort was abstracted from the Surveillance, Epidemiology, and End Results (SEER) 18 registries cohort[29] SEER program was established in 1973, which is an important data source for cancer epidemiological analyses that comprising approximately 30% of the total US population. SEER*Stat version 8.3.5 was applied to generate the case listing.

Factors And Outcome Definition

Estrogen receptor (ER) and Progesterone receptor (PR) status of breast cancer were assessed by immunohistochemistry and were considered positive if ≥ 1% of positive stained cells.[22,30,31] Human epidermal growth factor receptor-2 (HER-2) status was determined by immunohistochemistry and the subjects with suspicious positive (2+) were additionally tested by fluorescent in situ hybridization. The primary outcome of our study was OS, which was defined as the time from the diagnose date of breast cancer to the date of all-cause death.

Inclusion Criteria

The patients included in the present study fulfilled the following criteria: 1) aged 18–90 years females with unilateral malignant breast cancer; 2) after radical mastectomy; 3) without receiving preoperative chemotherapy or radiotherapy; 4) without distant metastasis. For the SEER cohort, the Chinese female malignant breast cancer patients who were aged 18–90 years, with only one primary site of malignant cancer, TNM I-III stage, histological grade I-IV, diagnosed between 2010 and 2012 (at least 3 years follow-up, as all of the patients, were followed up until 2015), underwent breast surgery, with known ER and PR status and did not have distant metastasis were included.

Ethics Statement

The written informed consent was obtained from each breast cancer patient or each patient's guardian in TBCCC and the current study was approved by the research ethics board of the Tianjin Medical University Cancer Institute and Hospital. The SEER is an open database, and the release of data from the SEER database does not require informed patient consent because cancer is a reportable disease in every state of the United States.

Statistical Analysis

Normally distributed data such as age was summarized as mean± standard deviations (SD). Categorical variables were presented as counts and percentages and the difference between groups were tested by Pearson's chi-square test or Wilcoxon rank-sum test. Multivariable Cox proportional hazards regression was applied to investigate the possible prognostic factors for OS. On the basis of the identified prognostic factors, a nomogram was constructed. The discrimination of the nomogram was evaluated by the receiver operating curve (ROC) and Harrell's Concordance-index (C-index)[32] Area under the curve (AUC) or C-index value of 0.5 indicated no discrimination and a value of 1.0 indicated the perfect separation of patients. Calibration was evaluated by plotting the mean Kaplan-Meier estimate versus the mean nomogram-predicted probability for patients by bootstrapping with 1000 resamples. In order to examine the generalizability of the nomogram, external validation was conducted using the SEER cohort.

The partial score method was used to construct the prognostic partial score (PPS).[33,34] Moreover, the PPS was classified into 3 subgroups as low, moderate, and high-risk group according to the tripartite grouping method. The concordance between the predicted survival and actual outcome of different groups of PPS was plotted by the survival curve.

To investigate the effect of adjuvant chemotherapy on the OS, multivariable Cox regression analysis was conducted to avoid the confounding factors. Moreover, a propensity score matching (PSM) method was used to balance observed covariates between patients received and not received adjuvant chemotherapy[35] Kaplan-Meier method was used and the difference between the curves was tested by the log rank test.

Statistical analyses were carried out using Statistical Package for the Social Sciences (SPSS) version 23.0 software package for Windows (SPSS Inc), R version 3.4.1 (R Foundation for Statistical Computing, Vienna, Austria; www.r-project.org) and Stata statistical software version 12 (StataCorp, College Station, TX). Statistically significant levels were two-tailed and set at $P<0.05$.

Results

Demographic And Clinical Characteristics

A total of 5,504 breast cancer patients were included in the current study (Figure S1A). Of these, the mean age was 51.74±10.26 years (24–89 years), the mean OS was 64.67 ±25.07 months (0–159 months) and the 3-year and 5-year survival rates were 97.2% and 93.8%, respectively. A total of 4,645 (84.4%) patients underwent postoperative chemotherapy, and twenty-seven of 1410 (1.91%) HER-2 positive patients received targeted therapy.

For the SEER cohort, a total of 2,110 records fulfilled the inclusion criteria. The mean survival was 48.83±14.35 months (0–71 months) and the 3-year OS rate was 96.0% (Figure S1B). The demographic and clinical characteristics of the current cohort and the SEER cohort were listed in Table 1.

Construction And Validation Of The Predictive Nomogram

Multivariable Cox regression model showed that age, tumor diameter, lymph node metastasis, and ER and PR expression were independent prognostic factors for OS. Although the histological differentiated grade was borderline significant (P=0.07), it was clinically relevant for predicting OS, thus it was incorporated into the multivariable regression model. (Table 2) On the basis of the above identified prognostic factors, a nomogram was constructed for predicting the 3-year and 5-year OS for breast cancer patients (Figure 1).

The C-index of the nomogram for 3-year and 5-year OS prediction were 0.74 [95% confidence interval (CI): 0.69–0.78] and 0.72 (95% CI: 0.69–0.75), respectively. The calibration curve showed the prediction (solid line) was closely approximates the 45-degree line (Figure 2A and B). Moreover, the ROC curve showed the nomogram exhibited good discrimination to predict the 3-year and 5-year OS and the AUC were 74.1% (95% CI: 70.1–78.0%) and 72.3% (95% CI: 69.6–75.1%), respectively (Figure 2D and E).

Table 1 Distribution Of The Demographics And Clinic Pathologic Characteristics Of Patients In The Current Cohort And SEER Cohort

Factors	The Current Cohort (N=5,504)		The SEER Validation Cohort (N=2,110)	
	No. Of Patients	%	No. Of Patients	%
Age (years)				
18-30	86	1.6	14	0.7
31–35	186	3.4	35	1.7
36–40	466	8.5	113	5.4
41–45	810	14.7	197	9.3
46–50	1058	19.2	334	15.8
51–55	1023	18.6	297	14.1
56–60	813	14.8	274	13.0
61–65	500	9.1	286	13.6
66–70	307	5.6	191	9.1
71–75	193	3.5	150	7.1
76–90	62	1.1	219	10.4
BMI (kg/m^2)			NA	NA
≤25	3056	55.5	NA	NA
>25	2448	44.5	NA	NA
Menopause			NA	NA
No	2522	46.4	NA	NA
Yes	2909	53.6	NA	NA
Smoke			NA	NA
No	4714	88.9	NA	NA
Yes	586	11.1	NA	NA
Drink			NA	NA
No	5157	97.6	NA	NA
Yes	127	2.4	NA	NA
Diameter (cm)			NA	NA
≤2	2425	44.1	1305	61.8
>2	3079	55.9	805	38.2
Histological grade				
Well differentiated	457	8.3	422	20.0
Moderately differentiated	4067	73.9	947	44.9
Poor or undifferentiated	980	17.8	741	35.1
Lymph node metastasis				
No	3023	54.9	1535	72.7
Yes	2481	45.1	575	27.3
ER				
Negative	1702	30.9	395	18.7
Positive	3802	69.1	1715	81.3

(Continued)

Table 1 (Continued).

Factors	The Current Cohort (N=5,504)		The SEER Validation Cohort (N=2,110)	
	No. Of Patients	%	No. Of Patients	%
PR				
Negative	1933	35.1	628	29.8
Positive	3571	64.9	1482	70.2
HER-2				
Negative	4094	74.4	1689	80.0
Positive	1410	25.6	421	20.0
Dead				
No	5032	91.4	2037	96.5
Yes	472	8.6	73	3.5
Chemotherapy				
No	859	15.6	1261	59.8
Yes	4645	84.4	849	40.2
Radiotherapy				
No	4687	85.2	1126	53.4
Yes	817	14.8	984	46.6
Endocrine therapy			NA	NA
No	4012	72.9	NA	NA
Yes	1492	27.1	NA	NA

Abbreviations: ER, estrogen receptor; PR, Progesterone receptor; HER-2, human epidermal growth factor receptor-2; NA, not available.

For external validation in the SEER dataset, the nomogram showed an optimal agreement between actual and predicted survival for 3-year OS. (Figure 2C) Additionally, the nomogram showed good discrimination in the validation set with C-index of 0.74 (95% CI: 0.67–0.80) and AUC of 74.0% (95% CI: 68.3–79.8%), which confirmed the exportability of the nomogram (Figure 2F).

Construction And Validation Of The Prognostic Partial Score (PPS)

The PPS was constructed based on the nomogram and three risk groups were modeled. The proportion of the breast cancer patients in the low, moderate, and high-risk subgroups were 34.6% (N =1,907; PPS ≤19.0), 32.1% (N =1,767; PPS=19.5–23.0), and 33.2% (N =1,830; PPS >23.0), respectively. This grouping led to good separation and accurate prediction of the OS probability for 3-year (Figure 2G) and 5-year OS (Figure 2H) in the TBCCC cohort and 3-year OS (Figure 2I) in the SEER validation cohort.

Table 2 Univariate And Multivariate Analysis Of Factors For Predicting The Overall Survival Of The Breast Cancer Patients After Surgery

Factors	Univariate Analysis			Multivariate Analysis		
	Hazard Ratio	95% CI	P	Hazard Ratio	95% CI	P
Age	1.22	1.17 to 1.28	<0.001	1.23	1.18 to 1.29	<0.001
BMI (kg/m^2)						
≤25	1.0	Ref	1.0	1.0	Ref	1.0
>25	1.18	0.98 to 1.41	0.08	NS	NS	NS
Menopause						
No	1.0	Ref	1.0	1.0	Ref	1.0
Yes	1.75	1.44 to 2.12	<0.001	NS	NS	NS
Smoke						
No	1.0	Ref	1.0	1.0	Ref	1.0
Yes	1.46	1.13 to 1.89	0.003	NS	NS	NS
Drink						
No	1.0	Ref	1.0	1.0	Ref	1.0
Yes	0.41	0.15 to 1.09	0.07	NS	NS	NS
Diameter (cm)						
≤2	1.0	Ref	1.0	1.0	Ref	1.0
>2	1.86	1.53 to 2.27	<0.001	1.67	1.37 to 2.03	<0.001
Histological differentiation						
Well differentiated	1.0	Ref	1.0	1.0	Ref	1.0
Moderately differentiated	1.77	1.15 to 2.73	0.009	1.48	0.96 to 2.29	0.08
Poor or undifferentiated	2.45	1.55 to 3.88	<0.001	1.64	1.02 to 2.63	0.04
Lymph node metastasis						
N0	1.0	Ref	1.0	1.0	Ref	1.0
N+	2.42	2.00 to 2.93	<0.001	2.39	1.97 to 2.89	<0.001
ER						
Negative	1.0	Ref	1.0	1.0	Ref	1.0
Positive	0.62	0.52 to 0.75	<0.001	0.77	0.61 to 0.98	0.03
PR						
Negative	1.0	Ref	1.0	1.0	Ref	1.0
Positive	0.59	0.49 to 0.71	<0.001	0.76	0.60 to 0.96	0.02
HER-2						
Negative	1.0	Ref	Ref	1.0	Ref	Ref
Positive	1.23	1.01 to 1.50	0.04	NS	NS	NS

Abbreviations: BMI, body mass index; ER, estrogen receptor; PR, Progesterone receptor; HER-2, human epidermal growth factor receptor-2.

Clinical Utility Of The PPS In Adjuvant Chemotherapy

Multivariable Cox regression analysis showed adjuvant chemotherapy was negatively associated with the risk of death [hazard ratio (HR) =0.65; 95% CI: 0.48–0.87; P=0.004] (Figure 3A). When stratified by the PPS, chemotherapy was significantly associated with better OS for the high-risk subgroup (HR=0.54; 95% CI: 0.37–0.77;

P<0.001). However, no significant association was found in the low-risk subgroup (HR=1.47; 95% CI: 0.52–4.19; P=0.47) or the moderate-risk subgroup (HR=0.78; 95% CI: 0.42–1.48; P=0.45) (Figure 3B–D).

PSM was conducted, the overall propensity score and all the factors were comparable in all of the groups after matching (Table S2–S5). Results showed patients with adjuvant chemotherapy exhibited a better OS than those
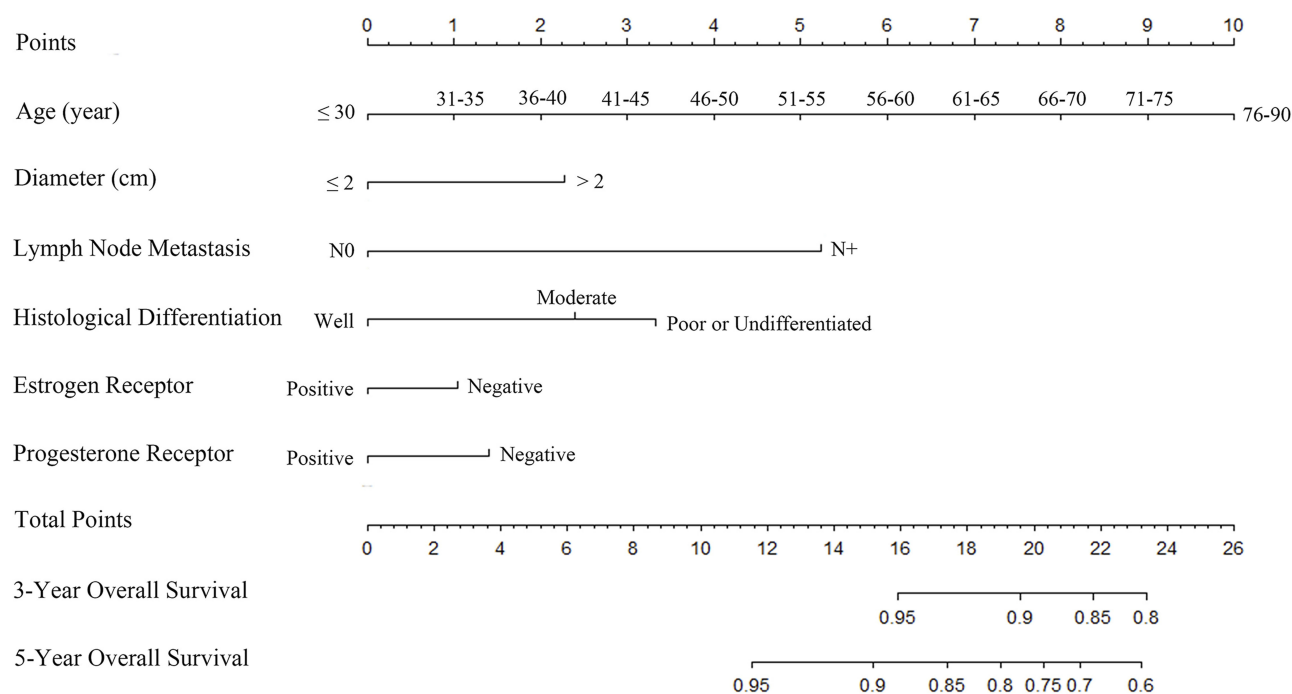
Figure 1 Breast cancer predictive nomogram for predicting 3-year and 5-year overall survival.

who did not receive chemotherapy (96.2% vs 92.7% without chemotherapy; *P*=0.002) (Figure 4A). When further stratified by the PPS group, high-risk patients significantly benefit from receiving chemotherapy (92.4% vs 84.7% without chemotherapy; *P*=0.004). No significant benefit was found in the low-risk group (98.6% vs 98.4% without chemotherapy; *P*=0.96) or the moderate-risk group (95.1% vs 94.4% of without chemotherapy; *P*=0.71) (Figure 4B–D).

Discussion

In the current study, we constructed a well-calibrated prognostic score based on the age, diameter, histological differentiation, lymph node metastasis, estrogen receptor and progesterone receptor to predict the 3-year and 5-year breast cancer survival and then according to the PPS score to stratify the patients into high, moderate and low risk subgroup for death. And then we aim to discuss the benefit of adjuvant chemotherapy in different risk groups. The results showed the subjects with high risk could markedly benefit from the chemotherapy while the ones with moderate and low-risk derived no significant benefit.

As far as we know, this is the first study conducted for predicting the overall survival based on Chinese breast cancer patients. Several demographic and clinical characteristics such as age at diagnosis, tumor size, histological differentiated grade, ER, PR and lymphatic node metastasis were proven to be the independent prognostic factors for breast cancer OS, which were consistent with previous studies.[36–38] It is also reported human epidermal growth factor-2 (HER-2) positivity expression was significantly associated with poor survival and may have implications in the guidance of adjuvant chemotherapy for breast cancer patients.[39–41] However, HER-2 positive expression was not significantly associated with OS in the current study, which may be partly due to the short-term follow up. In the current study, the median follows up period of the patients were about 5 years and the corresponding relative survival rates of breast cancer were reported to be approximate 94% in the current cohort. Accordingly, although the difference may existapparent difference could not be achieved in the short follow-up period. Additionally, this prediction nomogram was constructed to guide the postoperative adjuvant chemotherapy, thus the treatment factors such as radiotherapy and endocrinotherapy after surgery were not included.

Results showed that breast cancer patients at high risk may benefit most from adjuvant chemotherapy, and the ones with moderate and low risk could not earn more survival time after receiving adjuvant chemotherapy, which was also consistent with the 21-gene assay.[10,16] Accordingly, the PPS could guide the allocation of adjuvant chemotherapy for
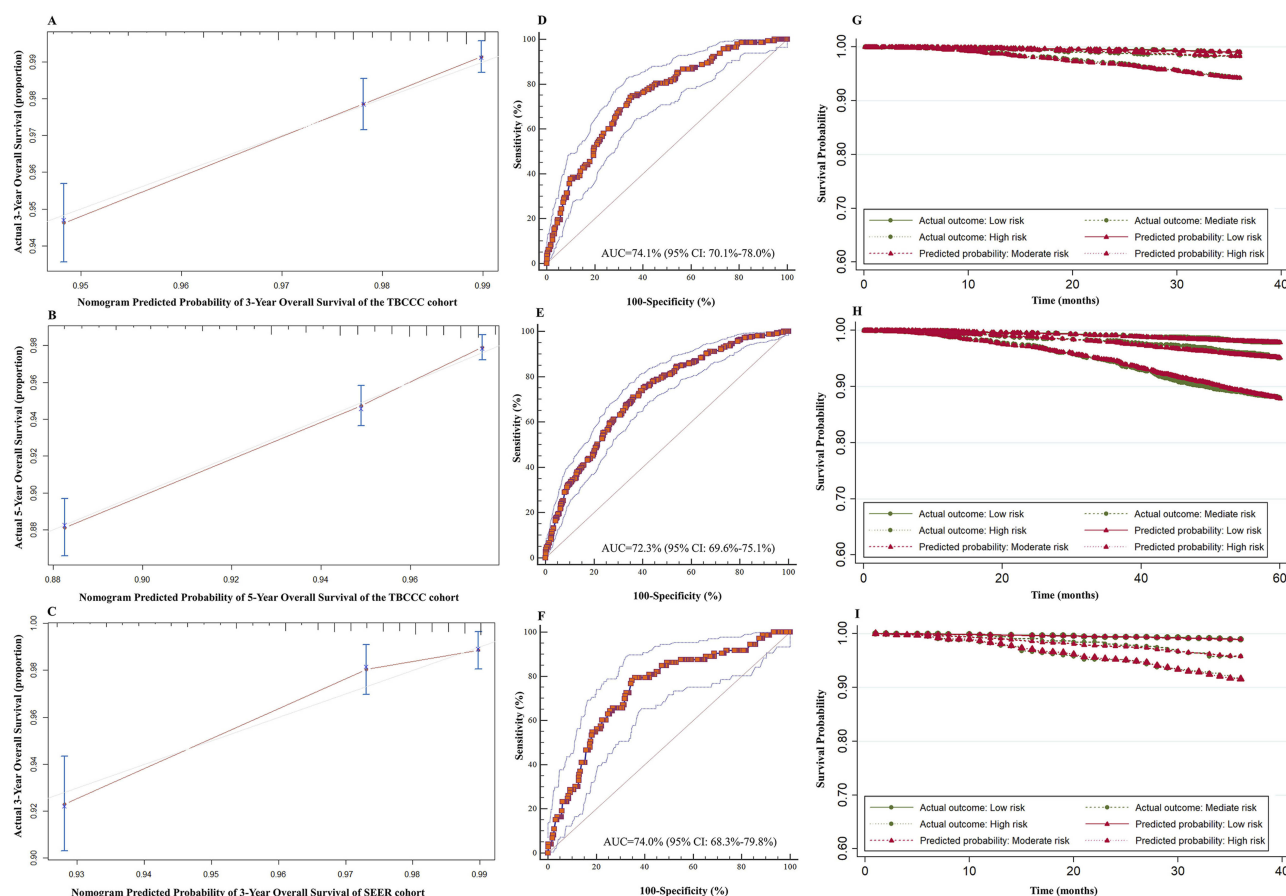
Figure 2 The validation of the predictive nomogram and prognostic partial score. (**A**) Calibration curve for predicting 3-year OS of TBCCC cohort, (**B**) 5-year OS of TBCCC cohort, (**C**) and 3-year OS of SEER cohort, (**D**) ROC curve for predicting 3-year OS of TBCCC cohort, (**E**) 5-year OS of TBCCC cohort, (**F**) and 3-year OS of SEER cohort, (**G**) PPS predicted and observed survival curve for 3-year survival of TBCCC cohort, (**H**) 5-year survival of TBCCC cohort, (**I**) and 3-year survival of SEER cohort.
Abbreviations: ROC, receiver operating curve; TBCCC, Tianjin Breast Cancer Cases Cohort; SEER, Surveillance, Epidemiology, and End Results; OS, overall survival; PPS, prognostic partial score.

breast cancer patients after surgery in China. The chemotherapy for subjects with moderate risk should be considered according to personal needs. Additionally, the choice of receiving chemotherapy as adjuvant therapy is not only determined by the tumor characteristics of the patients, but also the patients' preference.[42,43] However, we were unable to capture the role of patient preference for selection of adjuvant therapy in the current study, which may partly affect the application of the prediction nomogram. Further studies were warranted to adjust the patient preference of adjuvant chemotherapy when investigating the benefit of chemotherapy in different risk subgroups in the future.

Many methods were recommended to avoid the effect of confounding factors such as the multivariable regression analysis and instrumental variable method; however, none of these methods could fully avoid the

bias as multivariable regression analysis could only adjust for the known and currently available confounding factors, and most of the instrumental variables could just explain small part of the variance of the exposure.[44,45] Accordingly, we tried to comprehensively utilize a variety of statistical methods to confirm the results. In the current study, the multivariable Cox regression was conducted to adjust the confounding factors and a propensity score matching method was also conducted to further confirm the benefit of postoperative chemotherapy in the clinics. However, due to the nature of propensity score matching procedure, the patients with extremely probabilities that certainly benefit and none benefit from chemotherapy would be deleted because of the unsuccessful match, which may partly affect the statistical performance and exhibit a relatively conservative result.
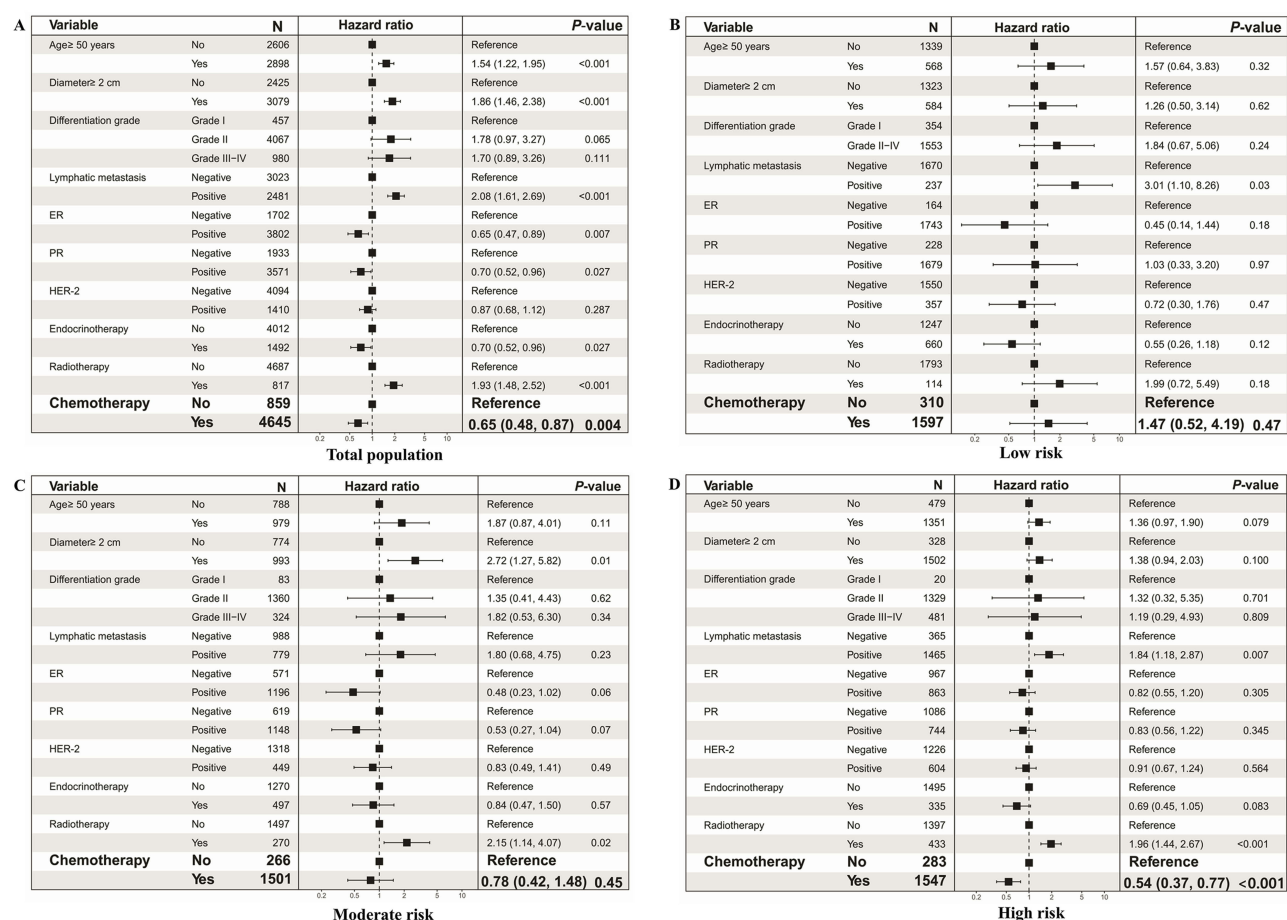
A Total population

Variable		N	Hazard ratio	P-value
Age≥ 50 years	No	2606	Reference	
	Yes	2898	1.54 (1.22, 1.95)	<0.001
Diameter≥ 2 cm	No	2425	Reference	
	Yes	3079	1.86 (1.46, 2.38)	<0.001
Differentiation grade	Grade I	457	Reference	
	Grade II	4067	1.78 (0.97, 3.27)	0.065
	Grade III–IV	980	1.70 (0.89, 3.26)	0.111
Lymphatic metastasis	Negative	3023	Reference	
	Positive	2481	2.08 (1.61, 2.69)	<0.001
ER	Negative	1702	Reference	
	Positive	3802	0.65 (0.47, 0.89)	0.007
PR	Negative	1933	Reference	
	Positive	3571	0.70 (0.52, 0.96)	0.027
HER-2	Negative	4094	Reference	
	Positive	1410	0.87 (0.68, 1.12)	0.287
Endocrinotherapy	No	4012	Reference	
	Yes	1492	0.70 (0.52, 0.96)	0.027
Radiotherapy	No	4687	Reference	
	Yes	817	1.93 (1.48, 2.52)	<0.001
Chemotherapy	**No**	**859**	**Reference**	
	Yes	**4645**	**0.65 (0.48, 0.87)**	**0.004**

B Low risk

Variable		N	Hazard ratio	P-value
Age≥ 50 years	No	1339	Reference	
	Yes	568	1.57 (0.64, 3.83)	0.32
Diameter≥ 2 cm	No	1323	Reference	
	Yes	584	1.26 (0.50, 3.14)	0.62
Differentiation grade	Grade I	354	Reference	
	Grade II–IV	1553	1.84 (0.67, 5.06)	0.24
Lymphatic metastasis	Negative	1670	Reference	
	Positive	237	3.01 (1.10, 8.26)	0.03
ER	Negative	164	Reference	
	Positive	1743	0.45 (0.14, 1.44)	0.18
PR	Negative	228	Reference	
	Positive	1679	1.03 (0.33, 3.20)	0.97
HER-2	Negative	1550	Reference	
	Positive	357	0.72 (0.30, 1.76)	0.47
Endocrinotherapy	No	1247	Reference	
	Yes	660	0.55 (0.26, 1.18)	0.12
Radiotherapy	No	1793	Reference	
	Yes	114	1.99 (0.72, 5.49)	0.18
Chemotherapy	**No**	**310**	**Reference**	
	Yes	**1597**	**1.47 (0.52, 4.19)**	**0.47**

C Moderate risk

Variable		N	Hazard ratio	P-value
Age≥ 50 years	No	788	Reference	
	Yes	979	1.87 (0.87, 4.01)	0.11
Diameter≥ 2 cm	No	774	Reference	
	Yes	993	2.72 (1.27, 5.82)	0.01
Differentiation grade	Grade I	83	Reference	
	Grade II	1360	1.35 (0.41, 4.43)	0.62
	Grade III–IV	324	1.82 (0.53, 6.30)	0.34
Lymphatic metastasis	Negative	988	Reference	
	Positive	779	1.80 (0.68, 4.75)	0.23
ER	Negative	571	Reference	
	Positive	1196	0.48 (0.23, 1.02)	0.06
PR	Negative	619	Reference	
	Positive	1148	0.53 (0.27, 1.04)	0.07
HER-2	Negative	1318	Reference	
	Positive	449	0.83 (0.49, 1.41)	0.49
Endocrinotherapy	No	1270	Reference	
	Yes	497	0.84 (0.47, 1.50)	0.57
Radiotherapy	No	1497	Reference	
	Yes	270	2.15 (1.14, 4.07)	0.02
Chemotherapy	**No**	**266**	**Reference**	
	Yes	**1501**	**0.78 (0.42, 1.48)**	**0.45**

D High risk

Variable		N	Hazard ratio	P-value
Age≥ 50 years	No	479	Reference	
	Yes	1351	1.36 (0.97, 1.90)	0.079
Diameter≥ 2 cm	No	328	Reference	
	Yes	1502	1.38 (0.94, 2.03)	0.100
Differentiation grade	Grade I	20	Reference	
	Grade II	1329	1.32 (0.32, 5.35)	0.701
	Grade III–IV	481	1.19 (0.29, 4.93)	0.809
Lymphatic metastasis	Negative	365	Reference	
	Positive	1465	1.84 (1.18, 2.87)	0.007
ER	Negative	967	Reference	
	Positive	863	0.82 (0.55, 1.20)	0.305
PR	Negative	1086	Reference	
	Positive	744	0.83 (0.56, 1.22)	0.345
HER-2	Negative	1226	Reference	
	Positive	604	0.91 (0.67, 1.24)	0.564
Endocrinotherapy	No	1495	Reference	
	Yes	335	0.69 (0.45, 1.05)	0.083
Radiotherapy	No	1397	Reference	
	Yes	433	1.96 (1.44, 2.67)	<0.001
Chemotherapy	**No**	**283**	**Reference**	
	Yes	**1547**	**0.54 (0.37, 0.77)**	**<0.001**

Figure 3 The association between adjuvant chemotherapy and overall survival in different risk subgroups. (**A**) Forest plot of the association in total population, (**B**) low risk population; (**C**) moderate risk population; (**D**) and high risk population.

There are several limitations to our study. Firstly, the follow-up period was tended to be short, which may affect the performance and the application of the prediction nomogram to some extent, as some important factors such as HER-2 were not included in the model. Secondly, the primary endpoint did not include the disease-free survival, only overall survival, which may partly limit the wide application of the results. Thirdly, due to the data limitation, the predictive model only included the clinical features. The data which reflected the host microenvironment like tumor-infiltrating lymphocytes was not included, which may partly affect the diagnostic accuracy of the predictive nomogram. Fourthly, due to the incomplete record of chemotherapy regimens in the present study, we could just provide a crude dichotomous grouping of chemotherapy (yes vs no) when analysis its benefit in different risk subgroups. In the future, further studies will be conducted to investigate the benefit of the specific chemotherapy agent, cycles and dose in different risk subgroups and help the clinicians to tailor targeted therapy. Moreover, due to the lower proportion of HER-2 positive patients treated by target therapy, we did not conduct further stratified analysis. More studies should be conducted to incorporate these factors into consideration when guiding the adjuvant chemotherapy in the future. Finally, because of the unavailability of data, we were unable to externally validate the clinical utility of the PPS in guiding adjuvant chemotherapy in other centers. Therefore, more studies are needed to further validate the clinical application of PPS in the future.

Conclusion

The nomogram was a well-calibrated model for predicting the degree of benefit from adjuvant chemotherapy in Chinese breast cancer patients and it may be utilized in the guidance for the postoperative adjuvant therapy in clinical settings.
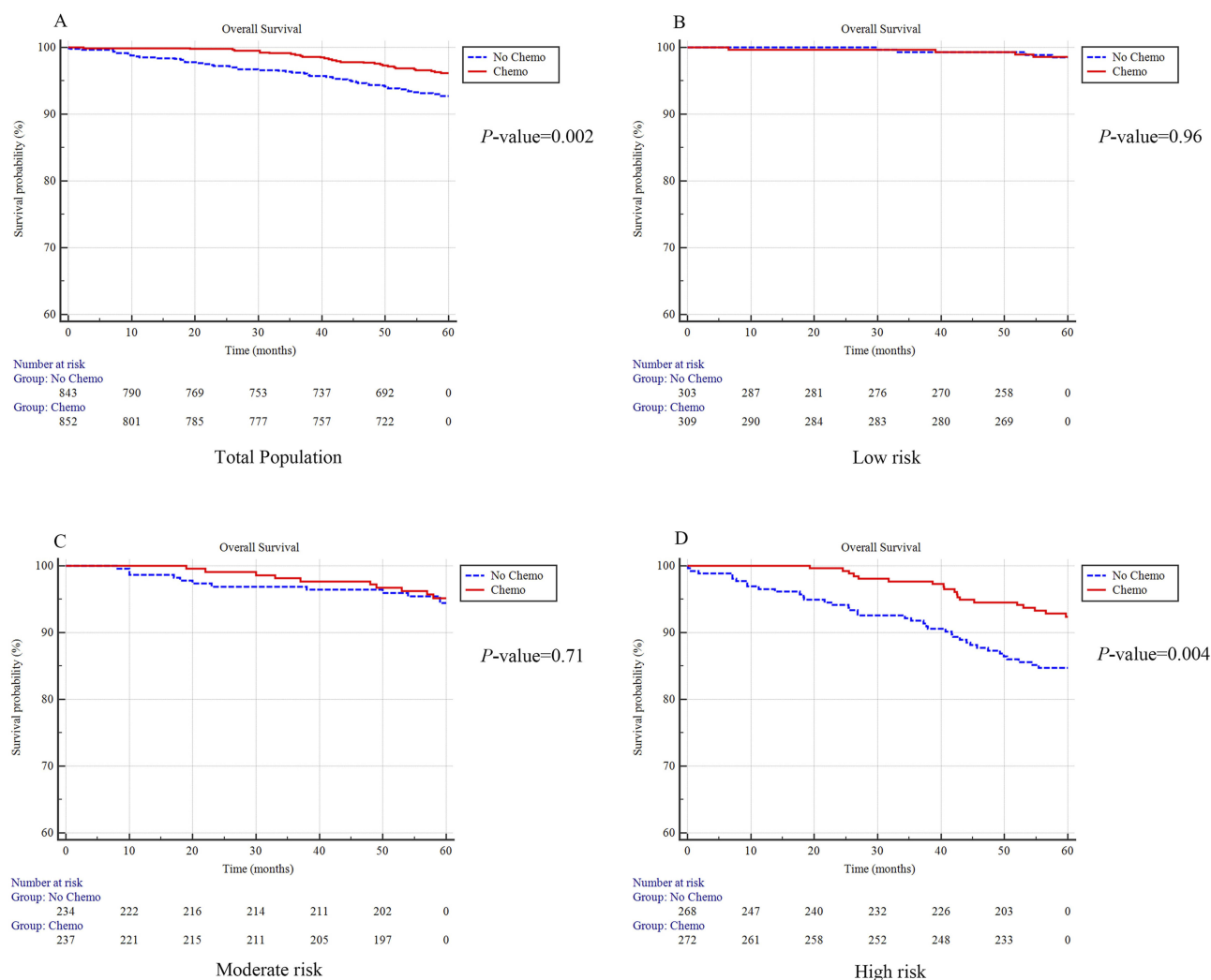
Figure 4 The survival plot for subjects with or without receiving adjuvant chemotherapy after propensity score matching. Survival plot in total population (**A**), low risk population (**B**), moderate risk population (**C**) and high risk population (**D**).

Disclosure

The authors indicated no potential conflicts of interest in this work.

References

1. Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2018;68:394–424. doi:10.3322/caac.v68.6
2. Torre LA, Bray F, Siegel RL, et al. Global cancer statistics, 2012. *CA Cancer J Clin*. 2015;65:87–108. doi:10.3322/caac.21262
3. Chen W, Zheng R, Baade PD, et al. Cancer statistics in China, 2015. *CA Cancer J Clin*. 2016;66:115–132. doi:10.3322/caac.21338

4. Berry DA, Cronin KA, Plevritis SK, et al. Effect of screening and adjuvant therapy on mortality from breast cancer. *N Engl J Med*. 2005;353:1784–1792. doi:10.1056/NEJMoa050518

5. Early Breast Cancer Trialists' Collaborative Group (EBCTCG). Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials. *Lancet*. 2005;365:1687–1717. doi:10.1016/S0140-6736(05)66544-0

6. Rossi L, Stevens D, Pierga JY, et al. Impact of adjuvant chemotherapy on breast cancer survival: a real-world population. *PloS One*. 2015;10:e0132853. doi:10.1371/journal.pone.0132853

7. Muss HB, Berry DA, Cirrincione CT, et al. Adjuvant chemotherapy in older women with early-stage breast cancer. *N Engl J Med*. 2009;360:2055–2065. doi:10.1056/NEJMoa0810266

8. Early Breast Cancer Trialists' Collaborative Group (EBCTCG), Peto R, Davies C, et al. Comparisons between different polychemotherapy regimens for early breast cancer: meta-analyses of long-term outcome among 100,000 women in 123 randomised trials. *Lancet*. 2012;379:432–444.

9. Francescutti V, Farrokhyar F, Tozer R, et al. Primary tumor and patient characteristics in breast cancer as predictors of adjuvant therapy regimen: a regression model. *Cancer Chemother Pharmacol*. 2011;68:661–668. doi:10.1007/s00280-010-1532-5

10. Paik S, Tang G, Shak S, et al. Gene expression and benefit of chemotherapy in women with node-negative, estrogen receptor-positive breast cancer. *J Clin Oncol*. 2006;24:3726–3734. doi:10.1200/JCO.2005.04.7985

11. Fisher B, Jeong JH, Bryant J, et al. Treatment of lymph-node-negative, oestrogen-receptor-positive breast cancer: long-term findings from national surgical adjuvant breast and bowel project randomised clinical trials. *Lancet*. 2004;364:858–868. doi:10.1016/S0140-6736(04)16981-X

12. Bedard PL, Cardoso F. Can some patients avoid adjuvant chemotherapy for early-stage breast cancer? *Nat Rev Clin Oncol*. 2011;8:272–279. doi:10.1038/nrclinonc.2011.19

13. Henderson IC, Patek AJ. The relationship between prognostic and predictive factors in the management of breast cancer. *Breast Cancer Res Treat*. 1998;52:261–288. doi:10.1023/A:1006141703224

14. Wishart GC, Azzato EM, Greenberg DC, et al. PREDICT: a new UK prognostic model that predicts survival following surgery for invasive breast cancer. *Breast Cancer Res*. 2010;12(1):R1. doi:10.1186/bcr2464

15. van 't Veer LJ, Dai H, van de Vijver MJ, et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature*. 2002;415:530–536. doi:10.1038/415530a

16. Paik S, Shak S, Tang G, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med*. 2004;351:2817–2826. doi:10.1056/NEJMoa041588

17. Bevilacqua JL, Kattan MW, Fey JV, et al. Doctor, what are my chances of having a positive sentinel node? A validated nomogram for risk estimation. *J Clin Oncol*. 2007;25(24):3670–3679. doi:10.1200/JCO.2006.08.8013

18. Adjuvant! Online Web Site. Available from: http://www.adjuvanton line.com. Accessed December 1, 2018.

19. Ravdin PM, Siminoff LA, Davis GJ, et al. Computer program to assist in making decisions about adjuvant therapy for women with early breast cancer. *J Clin Oncol*. 2001;19:980–991. doi:10.1200/JCO.2001.19.4.980

20. Olivotto IA, Bajdik CD, Ravdin PM, et al. Population-based validation of the prognostic model ADJUVANT! for early breast cancer. *J Clin Oncol*. 2005;23:2716–2725. doi:10.1200/JCO.2005.06.178

21. Mook S, Schmidt MK, Rutgers EJ, et al. Calibration and discriminatory accuracy of prognosis calculation for breast cancer with the online adjuvant! program: a hospital-based retrospective cohort study. *Lancet Oncol*. 2009;10:1070–1076. doi:10.1016/S1470-2045(09)70254-2

22. National Comprehensive Cancer Network. Breast Cancer (Version 3. 2017). Available from: https://www.nccn.org/professionals/physi cian_gls/PDF/breast.pdf. Accessed December 1, 2018.

23. Bray F, Ferlay J, Laversanne M, et al. Cancer incidence in five continents: inclusion criteria, highlights from volume X and the global status of cancer registration. *Int J Cancer*. 2015;137:2060–2071. doi:10.1002/ijc.29670

24. Morabia A, Costanza MC. International variability in ages at menarche, first livebirth, and menopause. World Health Organization Collaborative Study of Neoplasia and Steroid Contraceptives. *Am J Epidemiol*. 1998;148(12):1195–1205. doi:10.1093/oxfordjournals.aje. a009609

25. Gail MH, Brinton LA, Byar DP, et al. Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. *J Natl Cancer Inst*. 1989;81(24):1879–1886. doi:10.1093/jnci/81.24.1879

26. Bhoo-Pathy N, Yip CH, Hartman M, et al. Adjuvant! Online is overoptimistic in predicting survival of Asian breast cancer patients. *Eur J Cancer*. 2012;48:982–989. doi:10.1016/j.ejca.2012.01.034

27. Yao-Lung K, Dar-Ren C, Tsai-Wang C. Accuracy validation of adjuvant! online in Taiwanese breast cancer patients-a 10-year analysis. *BMC Med Inform Decis Mak*. 2012;12:108. doi:10.1186/1472-6947-12-108

28. de Glas NA, van de Water W, Engelhardt EG, et al. Validity of adjuvant! Online program in older patients with breast cancer: a population-based study. *Lancet Oncol*. 2014;15(7):722–729. doi:10.1016/S1470-2045(14)70200-1

29. Surveillance E, Results E. (SEER) Program Research Data (1973–2014), National Cancer Institute, DCCPS, Surveillance Research Program, Surveillance Systems Branch. Available from: http://www.seer.cancer.gov. Accessed December 1, 2018.

30. Hammond ME, Hayes DF, Dowsett M, et al. American Society of Clinical Oncology/College Of American Pathologists guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer. *J Clin Oncol*. 2010;28:2784–2795. doi:10.1200/JCO.2009.25.6529

31. Allred DC, Carlson RW, Berry DA, et al. NCCN task force report: estrogen receptor and progesterone receptor testing in breast cancer by immunohistochemistry. *J Natl Compr Canc Netw*. 2009;7:S1–S21. quiz S22-S23. doi:10.6004/jnccn.2009.0079

32. Pencina MJ, D'Agostino RB Sr. Evaluating discrimination of risk prediction models: the C statistic. *JAMA*. 2015;314:1063–1064. doi:10.1001/jama.2015.11082

33. Chow E, Abdolell M, Panzarella T, et al. Predictive model for survival in patients with advanced cancer. *J Clin Oncol*. 2008;26:5863–5869. doi:10.1200/JCO.2008.17.1363

34. Chow E, Ding K, Parulekar WR, et al. Predictive model for survival in patients having repeat radiation treatment for painful bone metastases. *Radiother Oncol*. 2016;118:547–551. doi:10.1016/j.radonc.2015.10.018

35. Austin PC, McDonald KM. The relative ability of different propensity score methods to balance measured covariates between treated and untreated subjects in observational studies. *Med Decis Making*. 2009;29:661–677. doi:10.1177/0272989X09341755

36. Korhonen T, Kuukasjarvi T, Huhtala H, et al. The impact of lobular and ductal breast cancer histology on the metastatic behavior and long term survival of breast cancer patients. *Breast*. 2013;22:1119–1124. doi:10.1016/j.breast.2013.06.001

37. Carter CL, Allen C, Henson DE. Relation of tumor size, lymph node status, and survival in 24,740 breast cancer cases. *Cancer*. 1989;63:181–187. doi:10.1002/(ISSN)1097-0142

38. Song Q, Huang R, Li J, et al. The diverse distribution of risk factors between breast cancer subtypes of ER, PR and HER2: a 10-year retrospective multi-center study in China. *PloS One*. 2013;8:e72175. doi:10.1371/journal.pone.0072175

39. Agrup M, Stal O, Olsen K, et al. C-erbB-2 overexpression and survival in early onset breast cancer. *Breast Cancer Res Treat*. 2000;63:23–29. doi:10.1023/A:1006498721508

40. Jelovac D, Wolff AC. The adjuvant treatment of HER2-positive breast cancer. *Curr Treat Options Oncol*. 2012;13:230–239. doi:10.1007/s11864-012-0186-4

41. Stal O, Sullivan S, Wingren S, et al. C-erbB-2 expression and benefit from adjuvant chemotherapy and radiotherapy of breast cancer. *Eur J Cancer*. 1995;31A:2185–2190. doi:10.1016/0959-8049(95)00344-4

42. Griggs JJ, Hawley ST, Graff JJ, et al. Factors associated with receipt of breast cancer adjuvant chemotherapy in a diverse population-based sample. *J Clin Oncol*. 2012;30:3058–3064. doi:10.1200/JCO.2012.41.9564

43. Zhang Y, Gao H, Bu Y, et al. Factors associated with receipt of adjuvant chemotherapy among married women with breast cancer. *World J Surg Oncol*. 2013;11:286. doi:10.1186/1477-7819-11-286

44. Brookhart MA, Schneeweiss S. Preference-based instrumental variable methods for the estimation of treatment effects: assessing validity and interpreting results. *Int J Biostat*. 2007;3:14. doi:10.2202/1557-4679.1072

45. Bosco JL, Silliman RA, Thwin SS, et al. A most stubborn bias: no adjustment method fully resolves confounding by indication in observational studies. *J Clin Epidemiol*. 2010;63:64–74. doi:10.1016/j.jclinepi.2009.03.001