IVYSPRING INTERNATIONAL PUBLISHER

Journal of Cancer

2019; 10(11): 2443-2449. doi: 10.7150/jca.30386

Research Paper

A Nomogram Predicting Lymph Node Metastasis in T1 Breast Cancer based on the Surveillance, Epidemiology, and End Results Program

Ya-Xin Zhao[1,2*], Yi-Rong Liu[1,2*], Shao Xie[2]✉, Yi-Zhou Jiang[1,2]✉ and Zhi-Ming Shao[1,2,3]✉

1. Department of Breast Surgery, Fudan University Shanghai Cancer Center; Cancer Institute, Fudan University Shanghai Cancer Center, 270 Dong-An Road, Shanghai 200032, People's Republic of China
2. Department of Oncology, Shanghai Medical College, Fudan University, P. R. China
3. Institutes of Biomedical Sciences, Fudan University, Shanghai, P. R. China

*Ya-Xin Zhao and Yi-Rong Liu contributed equally to this work.

✉ Corresponding authors: Shao Xie, PhD, Yi-Zhou Jiang, MD or Zhi-Ming Shao, MD, PhD. Department of Breast Surgery, Fudan University Shanghai Cancer Center, Cancer Institute, Fudan University Shanghai Cancer Center, Department of Oncology, Shanghai Medical College, Fudan University, 270 Dong-An Road, Shanghai, 200032, P. R. China. Tel.: +86-21-64175590; Fax: +86-21-64434556; Email: xshao_2019@163.com, yizhoujiang@fudan.edu.cn or zhimingshao@yahoo.com

Received: 2018.10.03; Accepted: 2019.04.22; Published: 2019.05.27

Abstract

Background: Patients with early stage breast cancer with lymph nodes metastasis were proven to have more aggressive biologically phenotypes. This study aimed to build a nomogram to predict lymph node metastasis in patients with T1 breast cancer.

Methods: We identified female patients with T1 breast cancer diagnosed between 2010 and 2014 in the Surveillance, Epidemiology and End Results database. The patients were randomized into training and validation sets. Univariate and multivariate logistic regressions were carried out to assess the relationships between lymph node metastasis and clinicopathological characteristics. A nomogram was developed and validated by a calibration curve and receptor operating characteristic curve analysis.

Result: Age, race, tumour size, tumour primary site, pathological grade, oestrogen receptor (ER) status, progesterone receptor (PR) status and human epidermal growth factor receptor 2 (HER2) status were independent predictive factors of positive lymph node metastasis in T1 breast cancer. Increasing age, tumour size and pathological grade were positively correlated with the risk of lymph node metastasis. We developed a nomogram to predict lymph node metastasis and further validated it in a validation set, with areas under the receiver operating characteristic curves of 0.733 and 0.741 in the training and validation sets, respectively.

Conclusions: A better understanding of the clinicopathological characteristics of T1 breast cancer patients might important for assessing their lymph node status. The nomogram developed here, if further validated in other large cohorts, might provide additional information regarding lymph node metastasis. Together with sentinel lymph node biopsy, this nomogram can help comprehensively predict lymph node metastasis.

Key words: T1 breast cancer, lymph nodes metastasis, predictive nomogram

Introduction

In developed countries, T1 stage breast cancers have become the most frequently diagnosed invasive breast diseases [1-3]. Patients with lymph node metastasis early in the disease may have biologically aggressive phenotypes, which have been proven to be correlated with a higher risk of distant spread [4]. The known predictors of axillary node metastases include tumour size, lymphovascular invasion, tumour grade and patient age [5, 6].

Sentinel lymph node biopsy (SLNB) was introduced as an alternative to axillary lymph node dissection (ALND) decades ago, with a similar staging capacity [7]. The frequency of axillary lymph node metastases in T1 breast cancer range from 10% to 26% [8], and the risk of missing metastases using SLNB can range from 1% to 4% [9], with a false negative rate of

10% [10, 11]. In addition, SLNB only examines the axillary sentinel nodes, so differences in primary tumour sites lead to significant changes in false-negative rates [10, 11]. Metastasis could also occur in extra-axillary lymph nodes. In studies with internal mammary node biopsies, approximately 7.8% of patients with negative axillary nodes had positive internal mammary biopsies [9].This evidence implies that SLNB might not be sufficient for the diagnosis of lymph node metastasis in T1 breast cancer patients. Combining SLNB and factors predictive of lymph involvement to evaluate the lymph node status could help better assess lymph node metastasis in T1 patients. As a surgical technique, SLNB leads to subjective lymphedema in both patients with both positive and negative nodes [12]. The aim of this study was to screen the clinicopathological characteristics that are associated with lymph node status and combine them into a predictive nomogram, which is a simple graphical representation of a statistical predictive model that generates a numerical probability of a clinical event [13], by using a population-based study cohort collected from the Surveillance, Epidemiology, and End Results (SEER) database. This nomogram was constructed to identify the risk of lymph node metastasis and might provide additional information regarding a patient's lymph node status in addition to that provided by SLNB. Furthermore, patients with lower risk of lymph node involvement might avoid having to undergo SLNB.

Materials and Methods

Patients

We retrieved the data for 91,364 invasive, stage T1 breast cancer patients registered in the SEER database from January 1, 2010, through December 31, 2014. The patients were selected according to the following criteria: female patients; AJCC stage T1; only one primary tumour; known oestrogen receptor (ER), progesterone receptor (PR) and human epidermal growth factor receptor 2 (HER2) status; known tumour grade; known race; known lymph node status. The T stage was further classified as T1 (0-1 mm), T1a (2-5 mm), T1b (6-10 mm), and T1c (11-20 mm) according to the Breast-Adjusted AJCC Stage (1988+) categories. T and N stage in SEER database are according to Adjusted AJCC(6^{th}). The lymph node status was identified according to the Regional Nodes Positive term in SEER. Patients diagnosed before 2010 were excluded because their HER2 statuses were unknown.

Construction and validation of the nomogram

Patients with known variables were randomly classified into training and validation sets in a 1:1 ratio for development and validation of the nomogram. We screened the clinicopathological characteristics that are associated with lymph node status and found statistically significant variables as follow: age at diagnosis, race, primary tumour site, tumour grade, T stage, and tumour subtype. Variables above were included in the nomogram. We used R (version 3.4.2) to establish the nomogram with R codes attached to the supplementary file. To assess the performance of the nomogram, we used a calibration curve with the bootstrapping method to illustrate the association between the actual probability and the predicted probability of positive lymph nodes in the training set [14, 15]. Receiver operating characteristic (ROC) curves, with the area under curve (AUC) value reported, were applied to evaluate the sensitivity and specificity of the nomogram for predicting lymph node metastasis in both the training and validation sets.

Statistical analyses

To develop a well-calibrated nomogram for the prediction of positive lymph node metastasis, we performed univariate logistic regression analyses to identify correlated variables ($P<0.05$ in univariate logistic regression analysis). The multivariate logistic regression analysis was performed to screen for significant predictors of positive lymph nodes, which were then included in the nomogram construction. Odds ratios (OR) and 95% confidence intervals (CI) were calculated.

We used the chi-square test to evaluate the relationship between the lymph node status and the appropriate variables. According to the nomogram, we calculated the total points for all patients and used Youden's index to identify the best cut-off value. The training and validation sets were stratified into two subgroups according to the cut-off point. Univariate logistic regression analyses were performed to show the correlation between the nomogram and the risk of lymph nodes metastasis.

Fisher's test was performed when necessary, and all reported P-values are two-sided. Only P-values less than 0.05 were deemed statistically significant, unless stated otherwise. SPSS (version 22.0) and R (version 3.4.2) were used to perform the statistical analyses. The R packages rms, pROC, Hmisc and ggplot2, parallel, and Daim (available at URL: http://cran.r-project.org/web/packages/) were used.

Results

Patient Characteristics

Among 91,364 eligible patients, 3819 (4.18%) were positive for lymph node metastases (Table 1). Lymph node status was related to age, race, primary

tumour site, and tumour size, grade, subtype and histological type. Younger patients (age<45) were more likely to have lymph nodes metastases (8.83%), compared to older patients (age=45-64: 4.69%, age>64: 2.76%). Black patients had higher rates of positive lymph nodes than white patients or other patients (6.47% vs. 4.07% and 3.76%). A positive correlation between tumour size and lymph node metastasis was found. Patients with stage T1c cancer had higher rates of positive lymph nodes (6.12%), compared to those with T1mi, T1a and T1b cancer (1.23%, 1.42% and 2.26%, respectively). Of patients with grade III cancer, 8.30% had positive lymph nodes, while 4% of patients with grade II cancer and 1.84% of patients with grade I cancer had positive lymph nodes. Patients whose primary tumour site was the axillary tail of the breast were more likely to have positive lymph nodes (9.26%), while patients whose primary tumour site was the nipple or central portion of the breast ranked second (5.26%). The rate of positive lymph nodes in patients with invasive ductal carcinoma (IDC) was higher than in those with invasive lobular carcinoma (ILC) or other histological types (4.46% vs. 3.85% and 3.33%). Regarding subtype, ER-positive patients had a lower rate of positive lymph nodes than ER-negative patients (3.74% vs. 7.78%). Similarly, PR-positive patients had lower rate of positive lymph nodes than PR-negative patients (3.65% vs. 6.38%). However, 7.67% of HER2-positive patients had positive lymph nodes, while only 3.79% of HER2-negative patients had positive lymph nodes (Table 1).

Independent Predictive Factors in the Training Set

Age, race, and tumour size, grade, primary site, histologic type and subtype were identified as being significantly associated with positive lymph nodes (Supplementary Table 1). All significant factors in the univariate analysis were included in the multivariate logistic regression analysis (Table 2). Histological type was not an independent predictor in our study (P=0.138). All the other variables showed statistically significant predictive capability for positive lymph nodes (P<0.001). The results showed that hormone receptor (HR)-/HER2+(P<0.001, OR=1.90, 95%CI:1.56 -2.31) patients had more positive lymph nodes compared to HR+/HER2+ patients (P<0.001, OR= 1.34, 95%CI:1.16-1.54). There was no significant difference between HR+/HER2- patients and triple negative patients (P=0.275).

Construction and validation of the nomogram

We established a nomogram that incorporated the significant predictive factors from the multivariate analysis (Figure 1). Age, race, primary tumour site, tumour subtype, and T stage, which were shown to be independent predictors in the multivariate logistic regression analysis, were included in the nomogram. By summing the scores of each variable, we can predict the probability of positive lymph nodes in a specific patient. Younger black patients with grade 3, T1c and HR-/HER2+ tumours in the axillary tail of the breast had higher scores. However, older white patients with grade 1, T1mic and HR+/HER2- tumours at inner sites had a lower risk of positive lymph nodes. The risk of positive lymph nodes predicted by our nomogram ranged from 0.01 to 0.5.

To test its performance, the nomogram was subjected to 1000 bootstrap resamples for internal validation with a calibration plot in the training set (Figure 2). The calibration plot showed that the nomogram was well calibrated. We further evaluated the effectiveness of the nomogram at predicting the lymph node status using the ROC curves in both the training (Figure 3A) and validation (Figure 3B) sets. In the training set, the AUC was 0.733 (95% CI: 0.7222-0.7437) and a similar AUC was observed in the validation set (AUC=0.741, 95% CI: 0.7305-0.752). The results confirmed the utility of the nomogram in predicting lymph node metastasis.

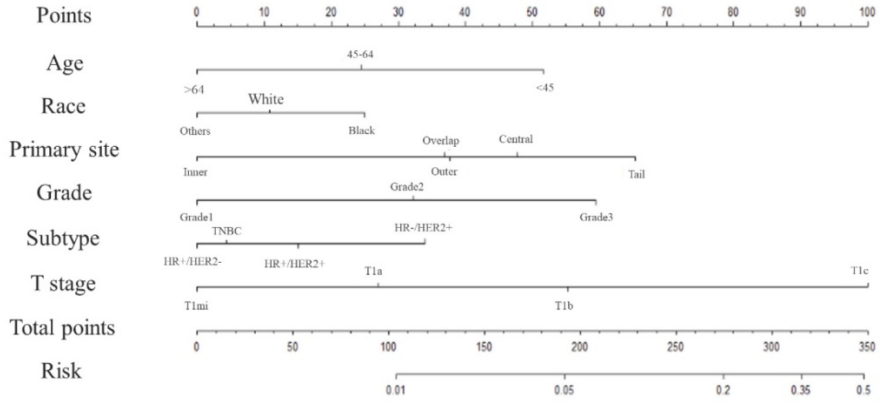

Figure 1. Nomogram predicting the probability of positive lymph nodes.

Table 1. Clinicopathologic characteristics of the cohort by lymph node status.

	Whole cohort				Training set				Validation set			
	NO.	LN negative	LN positive	P	NO.	LN negative	LN positive	P	NO.	LN negative	LN positive	P
Age				<0.001				<0.001				<0.001
<45	8179	7458 (91.18%)	721 (8.82%)		4077	3717 (91.17%)	360 (8.83%)		4102	3741 (91.20%)	361 (8.80%)	
45-64	45088	43007 (95.38%)	2081 (4.62%)		22472	21418 (95.31%)	1054 (4.69%)		22616	21589 (95.46%)	1027 (4.54%)	
>64	37982	36971 (97.34%)	1011 (2.66%)		19079	18553 (97.24%)	526 (2.76%)		18903	18418 (97.43%)	485 (2.57%)	
Race				<0.001				<0.001				<0.001
White	74766	71788 (96.02%)	2978 (3.98%)		37374	35854 (95.93%)	1520 (4.07%)		37392	35934 (96.10%)	1458 (3.90%)	
Black	8194	7679 (93.71%)	515 (6.29%)		4048	3786 (93.53%)	262 (6.47%)		4146	3893 (93.90%)	253 (6.10%)	
Others	8289	7969 (96.14%)	320 (3.86%)		4206	4048 (96.24%)	158 (3.76%)		4083	3921 (96.03%)	162 (3.97%)	
Marital status				0.897				0.563				0.439
Married	55788	53453 (95.81%)	2335 (4.19%)		27874	26701 (95.79%)	1173 (4.21%)		27914	26752 (95.84%)	1162 (4.16%)	
Unmarried	35461	33983 (95.83%)	1478 (4.17%)		17754	16987 (95.63%)	767 (4.32%)		17707	16996 (95.98%)	711 (4.02%)	
Primary site				<0.001				<0.001				<0.001
Central	3577	3391 (94.80%)	186 (5.20%)		1767	1674 (94.74%)	93 (5.26%)		1810	1717 (94.86%)	93 (5.14%)	
Inner	19254	18788 (97.58%)	466 (2.42%)		9658	9422 (97.56%)	236 (2.44%)		9596	9366 (97.60%)	230 (2.40%)	
Outer	39063	37267 (95.40%)	1796 (4.60%)		19419	18511 (95.32%)	908 (4.68%)		19644	18756 (95.48%)	888 (4.52%)	
Tail	427	382 (89.46%)	45 (10.54%)		216	196 (90.74%)	20 (9.26%)		211	186 (88.15%)	25 (11.85%)	
Overlap	28928	27608 (95.44%)	1320 (4.56%)		14568	13885 (95.31%)	683 (4.69%)		14360	13723 (95.56%)	637 (4.44%)	
Grade				<0.001				<0.001				<0.001
I	29859	29335 (98.25%)	524 (1.75%)		14929	14654 (98.16%)	275 (1.84%)		14930	14681 (98.33%)	249 (1.67%)	
II	41097	39482 (96.07%)	1615 (3.93%)		20542	19720 (96.00%)	822(4.00%)		20555	19762 (96.14%)	793 (3.86%)	
III	20293	18619 (91.75%)	1674 (8.25%)		10157	9314 (91.70%)	843 (8.30%)		10136	9305 (91.80%)	831 (8.20%)	
Histology				<0.001				<0.001				<0.001
IDC	71415	68259 (95.58%)	3156 (4.42%)		35654	34063 (95.54%)	1591 (4.46%)		35761	34196 (95.62%)	1565 (4.38%)	
ILC	6607	6364 (96.32%)	243 (3.68%)		3299	3172 (96.15%)	127 (3.85%)		3308	3192 (96.49%)	116 (3.51%)	
Others	13227	12813 (96.87%)	414 (3.13%)		6675	6453 (96.67%)	222 (3.33%)		6552	6360 (97.07%)	192 (2.93%)	
T stage				<0.001				<0.001				<0.001
T1mi	1615	1589 (98.39%)	26 (1.61%)		812	802 (98.77%)	10 (1.23%)		803	787 (98.00%)	16 (2.00%)	
T1a	11060	10909 (98.63%)	151 (1.37%)		5565	5486 (98.58%)	79 (1.42%)		5495	5423 (98.69%)	72 (1.31%)	
T1b	28383	27770 (97.84%)	613 (2.16%)		14309	13985 (97.74%)	324 (2.26%)		14074	13785 (97.95%)	289 (2.05%)	
T1c	50191	47168 (93.98%)	3023 (6.02%)		24942	23415 (93.88%)	1527 (6.12%)		25249	23753 (94.08%)	1496 (5.92%)	
ER status				<0.001				<0.001				<0.001
ER+	79797	76855 (96.31%)	2942 (3.69%)		39884	38391 (96.26%)	1493 (3.74%)		39913	38464 (96.37%)	1449 (3.63%)	
ER-	11452	10581 (92.39%)	871 (7.61%)		5744	5297 (92.22%)	447 (7.78%)		5708	5284 (92.57%)	424 (7.43%)	
PR status				<0.001				<0.001				<0.001
PR+	71263	68723 (96.44%)	2540 (3.56%)		35561	34263 (96.35%)	1298 (3.65%)		35702	34460 (96.52%)	1242 (3.48%)	
PR-	19986	18713 (93.63%)	1273 (6.37%)		10067	9425 (93.62%)	642 (6.38%)		9919	9288 (93.67%)	631 (6.36%)	
HER2 status				<0.001				<0.001				<0.001
HER2+	10813	9969 (92.20%)	844 (7.80%)		5440	5023 (92.33%)	417 (7.67%)		5373	4946 (92.05%)	427 (7.95%)	
HER2-	80436	77467 (96.31%)	2969 (3.69%)		40188	38665 (96.21%)	1523 (3.79%)		40248	38802 (96.41%)	1446 (3.59%)	
Subtype				<0.001				<0.001				<0.001
HR+/HER2-	72579	70145 (96.65%)	2434 (3.35%)		36268	35014 (96.54%)	1254 (3.46%)		36311	35131 (96.75%)	1180 (3.25%)	
HR-/HER2+	2913	2641 (90.67%)	272 (9.33%)		1493	1347 (90.22%)	146 (9.78%)		1420	1294 (91.13%)	126 (8.87%)	
HR+/HER2+	7900	7328 (92.76%)	572 (7.24%)		3947	3676 (93.13%)	271 (6.87%)		3953	3652 (92.39%)	301 (7.61%)	
TNBC	7857	7322 (93.19%)	535 (6.81%)		3920	3651 (93.14%)	269 (6.86%)		3937	3671 (93.24%)	266 (6.76%)	

Notes: Central: code C500 and C501; Inner: code C502 and C503; Outer: code C504 and C505; Tail: code C506; Overlap: code C508 and C509. From SEER program coding and staging manual 2015, coding guideline breast C500-C509. Abbreviations: IDC: invasive ductal carcinoma. ILC: invasive ductal carcinoma. ER: estrogen receptor. PR: progesterone receptors. HER2: human epidermal growth factor receptor 2. TNBC: Triple Negative Breast Cancer. LN: lymph nodes.

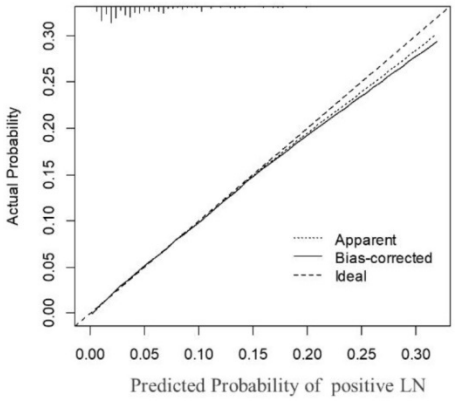

Figure 2. Calibration plot of the nomogram for the probability of positive lymph nodes (bootstrap 1000 repetitions) .

Stratifying patient risk by the nomogram

We determined the cut-off value of total points to predict lymph node metastasis according to Youden's index in the training set. Both the training set and validation set were divided into two groups: the low score group (total points≤182) and the high score group (total points>182). After applying the cut-off value to the training set, we found a significant difference in the probability of lymph node metastasis between the high and low score groups in univariate analysis (OR=4.15, 95% CI:3.77-4.57, P<0.001). The result was consistent in the validation set (OR=4.53, 95% CI 4.10-5.00, P<0.001; Table 3).

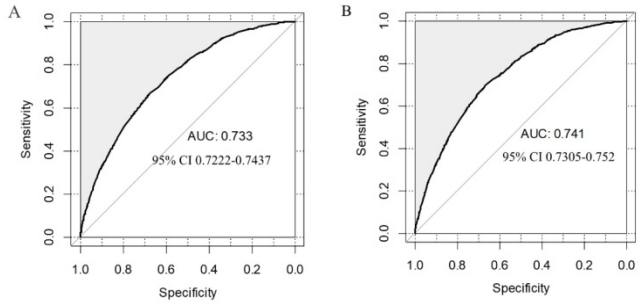

Figure 3. Validation of the nomogram. **A.** Internal validation using receiver operating characteristic (ROC) curve. The area under the ROC curve (AUC) is 0.733, 95% confidence intervals (CI): 0.7222-0.7437. **B.** External validation using ROC. The AUC is 0.741, 95% CI: 0.7305-0.752

Table 2. Multivariate logistic regression analysis of possible variables in predicting positive lymph nodes in training set.

	OR	95% CI	P
Age			P<0.001
<45	1.67	1.47-1.90	
45-64	Reference	Reference	
>64	0.63	0.57-0.70	
Race			P<0.001
White	Reference	Reference	
Black	1.31	1.14-1.50	
Others	0.82	0.69-0.97	
Primary site			P<0.001
Central	1.22	0.97-1.52	
Inner	0.49	0.42-0.57	
Outer	Reference	Reference	
Tail	1.68	1.04-2.73	
Overlap	0.99	0.89-1.10	
Grade			P<0.001
I	0.55	0.48-0.63	
II	Reference	Reference	
III	1.67	1.49-1.87	
Histology			P=0.138
IDC	Reference	Reference	
ILC	0.89	0.77-1.03	
Others	1.11	0.92-1.34	
T stage			P<0.001
T1mi	0.15	0.08-0.29	
T1a	0.25	0.20-0.32	
T1b	0.43	0.38-0.49	
T1c	Reference	Reference	
Subtype			P<0.001
HR+/HER2-	Reference	Reference	
HR-/HER2+	1.90	1.56-2.31	
HR+/HER2+	1.34	1.16-1.54	
TNBC	1.09	0.93-1.27	

Notes: T1mic:0-1mm, T1a:2-5mm, T1b:6-10mm, T1c:11-20mm. Abbreviations: IDC: invasive ductal carcinoma. ILC: invasive ductal carcinoma. ER: estrogen receptor. PR: progesterone receptors. HER2: human epidermal growth factor receptor 2. TNBC: Triple Negative Breast Cancer.

Table 3. Univariate logistic regression analysis of total points in predicting positive lymph nodes in training set and validation set.

	Training set			Validation set		
	OR	95%CI	P	OR	95%CI	P
Group			P<0.001			P<0.001
Low score	Reference	Reference		Reference	Reference	
High score	4.15	3.77-4.57		4.53	4.10-5.00	

Notes: low score:<=182; high score:>182. T1mic:0-1mm, T1a:2-5mm, T1b:6-10mm, T1c:11-20mm.Abbreviations: IDC: invasive ductal carcinoma. ILC: invasive ductal carcinoma. ER: estrogen receptor. PR: progesterone receptors. HER2: human epidermal growth factor receptor 2. TNBC: Triple Negative Breast Cancer.

Discussion

In the present study, we developed a predictive nomogram to evaluate the probability of positive lymph nodes in patients with T1 breast cancer based on the clinical and pathologic characteristics of the primary tumours. To validate the performance of the nomogram, we generated a calibration plot with bootstrap sampling and ROC curves in both the training and validation sets. According to previous studies, biomarkers with AUC between 0.7 and 0.9 have superior accuracy, indicating acceptable discriminations [16, 17]. Thus, we proved that the nomogram has high sensitivity and specificity in predicting the lymph node metastasis.

In the univariate logistic regression analyses, we found that age, race, T stage, and tumour primary site, subtype, grade and histological subtype were related to the lymph node status. These variables were independent predictors of the lymph node status, and their predictive value was confirmed by multivariate logistic regression with the exception of histological subtype, which is in line with the results of previous studies [18, 19]. The results of this study also showed that the risk of lymph node metastasis is positively related to tumour grade and size. Increasing tumour size was significantly associated with an increased risk of lymph node metastasis. Same results have been reported in a previous study [18]. Younger patients had a higher probability of lymph node metastasis. HER2 status was another important predictive factors. Patients with HER2+ status had a higher risk of lymph node metastasis than those who were HER2-; the overexpression of HER2 is now widely recognized as a predictor of poor prognosis in small tumours [20]. Patients with primary tumour sites in the axillary tail of the breast were more likely to have metastatic lymph nodes. These results may remind us that the primary tumour site is important for predicting lymph node metastasis. Regarding histological types, a previous study containing T2 stage tumours reported that histological type was an independent predictor of metastasis [21]; however, in T1 stage tumours we did not find the same result.

As to clinical utility of our nomogram, we could discuss it in different scenarios. It's widely accepted that SLNB is the standard procedure for axillary staging in patients with early breast cancer. However, in some patients, SLNB might not be sufficient for the assessment of metastatic lymph nodes. In the ALMNAC validation study, patients with higher tumour grade had higher false negative rate of SLNB; and the risk of patients with a missed axillary disease after a negative SLNB is 2.2% [22]. The false-negative rate was also higher for tumours in the outer quadrants of the breast [23, 24]. In the present study, we

also found the same result, i.e. tumours with higher grades located in the outer part were independent predictors of lymph node metastasis. In addition, the patients stratified into different risk subgroups according to our nomogram had different prevalence of lymph node metastases. Therefore, patients with different clinicopathologic features might have variable false-negative rate. Combining our nomogram could help understand lymph status better. In the other hand, for patients with metastatic nodes before preoperative/neoadjuvant systemic therapy (NACT), the false-negative rate with SLNB after treatment may range from 10% to 30%. Moreover, performing SLNB leads patients to the risk for long-term complications including permanent lymphedema. Thus for those patients who are predicted to have a lower possibility of lymph node metastasis (<10%) according to our nomogram, but with high false negative SLNB rate (>10%) [11] or high risk for long term SLNB complications, might avoid undergoing SLNB.

Comparing with former nomograms, our nomogram has some advantages. First, our nomogram is the first one to predict any lymph nodes involved including internal mammary nodes and infraclavicular lymph nodes, while most of former nomograms focus on only sentinel lymph nodes or non-sentinel lymph nodes. Reyal F reported a predictive nomogram only for sentinel nodes with an AUC of 0.73 [25].In Charles C's study, they compared several predictive nomograms focusing on non-sentinel lymph nodes including MSKCC nomogram, Mayo nomogram and Cambridge nomogram. Second, our nomogram can be widely used with less restrictions. Tenon score dependent on sentinel lymph nodes status may less convenient and accurate for clinicians [26]. The efficiency of Ngo C's model depends on the number of positive axillary lymph nodes. The AUC is 0.74 for patients without lymph nodes involvement and 0.70 for patients with one or two involved nodes [27]. The AUC values of the Chen's nomogram for positive axillary lymph nodes is 0.788 and for pN2-3 disease is 0.680 which may not accurately enough [28]. These model have high quality only on some specific conditions. Third, our model is the first one to be specific on T1 breast cancer.

One limitation of our study was that we could not obtain additional information from the SEER database, including the presence of lymphatic or vascular invasion, which have been demonstrated to be significant factors for axillary lymph node metastasis in some single centre investigations [18,29]. Including these predictive factors may improve the sensitivity and specificity of our nomogram. Another limitation is that we have not validated the nomogram in patients with SLNB. The information about whether patients underwent SLNB was unavailable in SEER database, therefore we tried to validate the nomogram in our own database which has the axillary surgery methods info. However, the nomogram was build based on white majority population, and race is an important factor. Thus we could not validate the nomogram in our own database with almost Asian patients. We also tried to manually deleted the race factor from nomogram, which turns out the modified nomogram won't work at all. This further validated the race's role in predicting. In addition, our study was a retrospective study, and therefore it was challenging to avoid selection bias or information bias. Most of the population in this study was American; whether similar biological characteristics of patients with small tumours and lymph node metastasis exist in an Asian population needs further investigation.

Conclusion

In conclusion, we developed a predictive nomogram for lymph node metastasis in patients with T1 breast cancer. The nomogram can help identify patients at low risk of lymph node metastasis who might be exempted from undergoing SLNB. In addition, combining the nomogram with SLNB might provide comprehensive information lymph node status after further validation is performed.

Abbreviations

ER: oestrogen receptor; PR: progesterone receptor; HER2: human epidermal growth factor receptor 2; SLNB: sentinel lymph node biopsy; ALND: axillary lymph node dissection; ROC: receiver operating characteristic curves; AUC: area under curve; IDC: invasive ductal carcinoma; ILC: invasive ductal carcinoma; neoadjuvant systemic therapy: NACT.

Supplementary Material

Supplementary table 1.
http://www.jcancer.org/v10p2443s1.pdf

Acknowledgements

The authors are grateful to Jiong Wu, Guang-Yu Liu and Zhen-Zhou Shen for their advice of construction of the nomogram.

Funding

This work was supported by grants from the National Natural Science Foundation of China (815 02278, 81572583, 81372848, 81370075), the Training Plan of Excellent Talents in Shanghai Municipality Health System (2017YQ038), the "Chen Guang" project supported by Shanghai Municipal Education

Commission and Shanghai Education Development Foundation (17CG01), the Training Plan of Excellent Talents of Fudan University Shanghai Cancer Center (YJYQ201602), the Municipal Project for Developing Emerging and Frontier Technology in Shanghai Hospitals (SHDC12010116), the Cooperation Project of Conquering Major Diseases in Shanghai Municipality Health System (2013ZYJB0302), the Innovation Team of Ministry of Education (IRT1223), and the Shanghai Key Laboratory of Breast Cancer (12DZ2260100).

The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript authors.

Ethics approval and consent to participate

All the procedures followed were in accordance with the Helsinki Declaration (1964, amended in 1975, 1983, 1989, 1996 and 2000) of the World Medical Association. The data released by the SEER database was publicly available and does not require informed patient consent because cancer is a reportable disease in every state in the United States.

Availability of data and materials

Surveillance, Epidemiology, and End Results (SEER) database.

Competing Interests

The authors have declared that no competing interest exists.

References

[1] Bland KI Menck HR, Scott-Conner CE, et al. The National Cancer Data Base 10-year survey of breast carcinoma treatment at hospitals in the United States. Cancer. 1998;83:1262-73.
[2] Fracheboud J, Otto SJ, Van DJ, et al. Decreased rates of advanced breast cancer due to mammography screening in The Netherlands. Br J Cancer. 2004;91:78.
[3] Houvenaeghel G, Goncalves A, Classe JM, et al. Characteristics and clinical outcome of T1 breast cancer: a multicenter retrospective cohort study. Ann Oncol. 2014;25:623-28.
[4] Jennifer YW, Kun C, Neville BA, et al. Effect of Very Small Tumor Size on Cancer-Specific Mortality in Node-Positive Breast Cancer. J Clin Oncol. 2011;29:2619-27.
[5] Donegan W L. Tumor-related prognostic factors for breast cancer. CA Cancer J Clin. 1997;47:28-51.
[6] Lyman GH, Giuliano AE, Somerfield MR, et al. American Society of Clinical Oncology guideline recommendations for sentinel lymph node biopsy in early-stage breast cancer. J Clin Oncol. 2005;23:7703-20.
[7] Shoup M, Malinzak L, Weisenberger J, et al. Predictors of axillary lymph node metastasis in T1 breast carcinoma. Am Surg. 1999;65:752-53.
[8] 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;25:3670-79.
[9] Hindié E, Groheux D, Brenot-Rossi I, et al. The sentinel node procedure in breast cancer: nuclear medicine as the starting point. J Nucl Med. 2011;52:405-14.
[10] Krag DN, Anderson SJ, Julian TB, et al. Technical outcomes of sentinel-lymph-node resection and conventional axillary-lymph-node dissection in patients with clinically node-negative breast cancer: results from the NSABP B-32 randomised phase III trial. Lancet Oncol.2007;8:881-8.
[11] GH L, S T, SB E, et al. Sentinel lymph node biopsy for patients with early-stage breast cancer: American Society of Clinical Oncology clinical practice guideline update. J Clin Oncol. 2014;32:1365-83.
[12] Purushotham AD, Upponi S, Klevesath MB, et al. Morbidity after sentinel lymph node biopsy in primary breast cancer: results from a randomized controlled trial. J Clin Oncol. 2005;23:4312-21.
[13] Iasonos A, Schrag D, Raj GV, et al. How to build and interpret a nomogram for cancer prognosis. J Clin Oncol. 2008;26:1364-70.
[14] Jr HF, 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.
[15] Steyerberg EW, Jr HF, Borsboom G, et al. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. J Clin Epidemiol. 2001;54:774-81.
[16] Qiu PF, Liu JJ, Wang YS, et al. Risk factors for sentinel lymph node metastasis and validation study of the MSKCC nomogram in breast cancer patients. Jpn J Clin Oncol. 2012;42:1002-7.
[17] Koca B, Kuru B, Ozen N, et al. A breast cancer nomogram for prediction of non-sentinel node metastasis - validation of fourteen existing models. Asian Pac J Cancer Prev. 2014;15:1481-8.
[18] Rivadeneira DE, Simmons RM, Christos PJ, et al. Predictive factors associated with axillary lymph node metastases in T1a and T1b breast carcinomas: analysis in more than 900 patients. J Am Coll Surg. 2000;191:1-6.
[19] Gajdos C, Tartter PI, Bleiweiss IJ. Lymphatic invasion, tumor size, and age are independent predictors of axillary lymph node metastases in women with T1 breast cancers. Ann Surg. 1999;230:692-6.
[20] Gonzalez Angulo AM, Litton JK, Broglio KR, et al. High risk of recurrence for patients with breast cancer who have human epidermal growth factor receptor 2-positive, node-negative tumors 1 cm or smaller. J Clin Oncol. 2009;27:5700-6.
[21] Tan LG, Tan YY, Heng D, et al. Predictors of axillary lymph node metastases in women with early breast cancer in Singapore. Singapore Med J. 2005;46:693-97.
[22] Goyal A, Newcombe RG, Chhabra A, et al. Factors affecting failed localisation and false-negative rates of sentinel node biopsy in breast cancer--results of the ALMANAC validation phase. Breast Cancer Res Treat. 2006;99:203-8.
[23] Krag D, Weaver D, Ashikaga T, et al. The Sentinel Node in Breast Cancer − A Multicenter Validation Study. N Engl J Med. 1998;339:941-6.
[24] Hiram S, Cody III. Clinicopathologic factors associated with false-negative sentinel lymph-node biopsy in breast cancer. Ann Surg. 2006;244:324.
[25] Reyal F, Rouzier R, Depont-Hazelzet B, et al. The Molecular Subtype Classification Is a Determinant of Sentinel Node Positivity in Early Breast Carcinoma. PloS One. 2011;6:1211-22.
[26] Coutant C, Olivier C, Lambaudie E, et al. Comparison of Models to Predict Nonsentinel Lymph Node Status in Breast Cancer Patients With Metastatic Sentinel Lymph Nodes: A Prospective Multicenter Study. J Clin Oncol. 2009;27:2800-8.
[27] Ngô C, Mouttet D, De YR, et al. Validation over time of a nomogram including HER2 status to predict the sentinel node positivity in early breast carcinoma. Eur J Surg Oncol. 2012;38:1211-17.
[28] Chen K, Liu J, Li S, et al. Development of nomograms to predict axillary lymph node status in breast cancer patients. BMC Cancer. 2017;17:561.
[29] Lauria R, Perrone F, Carlomagno C, et al. The prognostic value of lymphatic and blood vessel invasion in operable breast cancer. Cancer. 1995;76:1772-8.