Establishment of a logistic regression model nomogram for clinicopathological characteristics and risk factors with axillary lymph node metastasis in T1 locally advanced breast cancer: a retrospective study

Fang Qian[1]^, Haoyuan Shen[2], Chunyan Deng[3], Chenghao Liu[2], Tingting Su[2], Anli Chen[1], Di Hu[2], Jiacheng Zhu[2]

[1]Postgraduate Training Base of the Xiaogan Central Hospital of Jinzhou Medical University, Xiaogan, China; [2]Department of Thyroid Gland Breast Surgery, Xiaogan Hospital Affiliated to Wuhan University of Science and Technology (Xiaogan Central Hospital), Xiaogan, China; [3]Department of Pediatrics, Xiaogan Hospital Affiliated to Wuhan University of Science and Technology (Xiaogan Central Hospital), Xiaogan, China

Contributions: (I) Conception and design: F Qian, H Shen, C Deng; (II) Administrative support: F Qian, H Shen; (III) Provision of study materials or patients: None; (IV) Collection and assembly of data: F Qian, T Su, A Chen, D Hu, J Zhu; (V) Data analysis and interpretation: F Qian; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Haoyuan Shen, MD, PhD. Department of Thyroid Gland Breast Surgery, Xiaogan Hospital Affiliated to Wuhan University of Science and Technology (Xiaogan Central Hospital), 6 Square Road, Xiaonan District, Xiaogan 432000, China. Email: shhfxgy7679@sina.com.

Background: Although the research reports on locally advanced breast cancer (LABC) are increasing year by year, there are few reports on T1 LABC axillary lymph node metastasis (ALNM). By establishing a prediction model for T1 LABC ALNM, this study provides a reference value for the probability of ALNM of related patients, which helps clinicians to develop a more effective and individualized treatment plan for LABC.

Methods: Cases with pathologically confirmed T1 breast cancer (BC) between 2010 and 2015 in the Surveillance, Epidemiology, and End Results (SEER) database were identified. Logistic regression was used to analyze the correlation between LABC lymph node metastasis and every factor, and the odds ratio (OR) and 95% confidence interval (CI) were used to identify any influencing factors. A nomogram was drawn after incorporating meaningful factors identified in multivariate logistic regression into the model. The receiver operating characteristic (ROC) curve of the model was drawn, and the area under the curve (AUC) and its 95% CI were calculated. Hosmer-Lemeshow goodness-of-fit test and clinical decision curve analysis (DCA) were performed. The results were validated in the validation group.

Results: A total of 200,933 female T1 BC patients were included in this study. Univariate and multivariate logistic regression analysis of T1 BC showed that progesterone receptor (PR)-negative, race, age, lobular carcinoma, micropapillary ductal carcinoma, axillary tail tumor, poor differentiation, and larger tumor diameter increased the probability of ALNM in T1 LABC. A predictive nomogram was established using the above predictors, the AUC of the modeling group was 0.739 (95% CI: 0.732–0.747), and when the AUC cut-off value was 0.026, the specificity and sensitivity of the model were 65.78% and 69.99%, respectively. Validation of the model showed that the AUC of the validation group (n=60,280) was 0.741. When all the risk factors were met, the predicted probability of N2–N3 was 50.40%.

Conclusions: In this study, it was found that PR-negative, Black race, age, lobular carcinoma, micropapillary ductal carcinoma, axillary tail tumor, poor differentiation, and tumor diameter increased the probability of large lymph node metastasis in T1 LABC small tumors.

Keywords: Age; breast cancer (BC); carcinoma; database; diagnosis

---

Introduction

Background

The leading cause of female malignant tumors worldwide has always been breast cancer (BC) as evidenced by data collected over the years. In 2020, the World Health Organization's International Agency for Research on Cancer (IARC) reported 2.26 million new BC cases, making it the type of cancer with the highest rate of incidence (1). Locally advanced breast cancer (LABC) has been introduced both domestically and internationally. The eighth edition of Alternate Joint Communications Center (AJCC) classifies LABC separately, and it defines BC in clinical stage IIIA–C as LABC (2). The incidence of BC is decreasing as the public's awareness of BC deepens and the detection rate increases. However, for patients with LABC at the time of diagnosis, the subsequent treatment becomes quite important due to the high mortality rate.

The development of models related to axillary lymph node metastasis (ALNM) for BC has become particularly important. At present, the types of such models are increasing both internationally and within China. BC ALNM prediction has been greatly improved due to the appearance of models that combine imaging and kinetic curves (3,4). In clinical practice, genetic testing has become a common method for predicting tumor recurrence (5,6). Although the BC ALNM situation can be accurately predicted by the multimodal genetic nomogram, its expensive cost has prevented its widespread use (7,8). Establishing a convenient, economical, and effective model to predict ALNM of LABC is crucial due to the lack of clinically relevant models related to ALNM of T1 LABC.

The aim of this study was to establish a relevant model, which provides a certain reference value for predicting whether ALNM occurs in LABC. At the same time, it is helpful to strengthen the understanding of LABC and enable clinicians to formulate more effective and individualized treatment plans for LABC. We present this article in accordance with the TRIPOD reporting checklist (available at https://gs.amegroups.com/article/view/10.21037/gs-24-34/rc).

Highlight box

Key findings
- In this study, it was found that progesterone receptor (PR)-negative, Black race, age, lobular carcinoma, micropapillary ductal carcinoma, axillary tail tumor, poor differentiation, and tumor diameter increased the probability of large lymph node metastasis in T1 locally advanced breast cancer small tumors.

What is known and what is new?
- Previous studies have shown that tumor diameter, Black race, age, micropapillary ductal carcinoma, poor differentiation, tumor diameter, and histological grade are independent risk factors for the development of axillary lymph node metastasis (ALNM).
- This study found that ALNM is more likely to occur when PR is negative or tumors are in the axillary tail.

What is the implication, and what should change now?
- The PR status and primary tumor location are important for the prognosis of patients. Improving preoperative tumor immunohistochemistry and tumor localization markers can increase the attention of clinicians to a certain extent and improve the prognosis of patients.

Methods

Data source and patient population

Analysis was performed on data from the National Cancer Institute (NCI)'s Surveillance, Epidemiology, and End Results (SEER) program. SEER provides the definitive information on cancer incidence and survival in the United States (U.S.), currently collects and publishes cancer incidence and survival data from cancer registries covering approximately 48.0% of the U.S. population, and regularly collects data on patient demographics, primary tumor location, tumor morphology and stage of diagnosis, and first course of treatment, and vital status follow-up data are regularly collected. This study analyzed the 17-SEER database that was submitted in November 2021. Patients who had pathologically confirmed clinical stage T1 BC were included; those with unknown clinical stage N were excluded, and patients with unknown information and blank information were excluded. The flow chart for the inclusion of patients is shown in *Figure 1*. Access to the SEER database was obtained for this study. This study
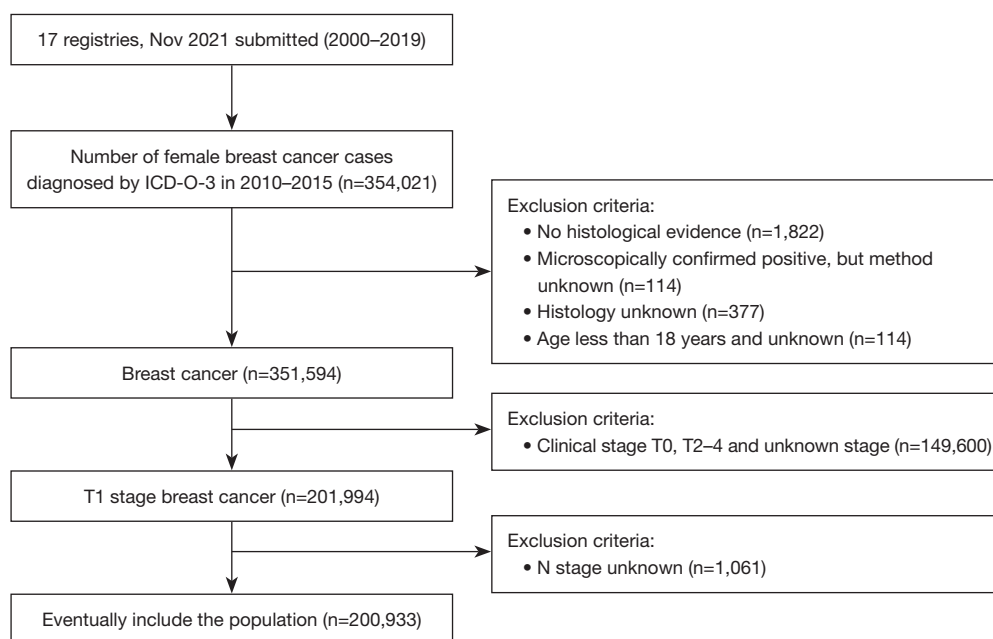
Figure 1 Patient inclusion process. ICD-O-3, International Classification of Disease for Oncology Third Edition.

was conducted in accordance with the principles of the Declaration of Helsinki (as revised in 2013).

Statistical methods

RStudio 4.3.1 software (R Foundation for Statistical Computing, Vienna, Austria) was used to analyze the data. Continuous data were expressed as mean ± standard deviation (M ± SD), and the Kruskal-Wallis rank sum test was used for comparison between the two groups. The percentage of cases was used to express the count data. The basic information of LABC patients and the clinical characteristics of tumors were analyzed by *t*-test and Chi-square test. Bar charts were drawn using the "rio", "ggplot2", "dplyr", "ggrepel", and "RColorBrewer" functions of RStudio 4.3.1, the "echarts4r" package was used for the doughnut drawing, and univariate and multivariate binary logistic regression models were fitted, using the "glm" function. The variables with statistical significance in univariate analysis were included in the multivariate logistic regression analysis to determine the possible influencing factors of LABC ALNM. Logistic regression prediction models were established using the "rms" and "Hmisc" software packages of RStudio. A nomogram was drawn using the "regplot" software package and the

significant factors from the multivariate logistic regression were included in the model. In this study, the sample () function was used to divide the data into the training set (70%) and the test set (30%) by simple random sampling, The receiver operating characteristic (ROC) curve of the nomogram model for the training group and the validation group were drawn respectively. The area under the ROC curve (AUC) and its 95% confidence interval (95% CI), and decision curve analysis (DCA) of the nomogram model for the modeling group (training set) and validation group (test set), were plotted respectively. The calibration curve of the model was drawn using the Bootstrap repeated self-sampling method, and it was tested for goodness of fit using the R language software package "Resource Selection".

The software SPSS 25.0 (IBM Corp., Armonk, NY, USA) was used to clean the data, and P<0.05 was considered statistically significant.

Results

Basic information about the patient and tumor characteristics

A total of 200,933 patients with T1 BC were included in this study. In the modeling group, 137,214 patients

874

Qian et al. Establishment of risk factors model for ALNM in T1 LABC

(97.6%) were in the N0–N1 stage, with an average age of 62.47±12.65 years; there were 3,439 patients (2.4%) in the N2–N3 stage, with an average age of 58.36±12.85 years. In the validation group, 58,827 patients (97.6%) were in the N0–N1 stage, with an average age of 62.53±12.69 years; there were 1,453 patients (2.4%) in N2–N3 stage, with an average age of 58.82±13.02 years. The patients were followed up until 31 December 2020. Differences between groups were analyzed using the *t*-test and Chi-square test (*Table 1*). The results showed that patients with N2–N3 were mainly with large tumor diameter (T1c accounted for 82.3%), outer upper quadrant, invasive ductal carcinoma, and hormone receptor (HR)-positive. T1 BC were mainly located in upper outer quadrant (34.1%) and the overlap

of the breast (23.0%) (*Figure 2A*). In order to better understand the relationship between age and ALNM, a histogram of age and the number of lymph node metastases was drawn with the age difference of 5 years as the spacing, and the results showed that patients in the N2–N3 stage had the highest proportion of 0–35-year-old patients (5.6%). the results showed that for N2–N3 stage had the highest proportion of 0–35-year-old patients (5.6%). Young age was identified as a major risk factor for stage N2–N3 (*Figure 2B*). The doughnut plot of the distribution of pathological types showed that invasive ductal carcinoma (76.8%) and lobular carcinoma (7.9%) were the main tumors in T1 stage, and micropapillary ductal carcinoma accounted for only 0.4% (*Figure 2C*).

Table 1 Clinicopathological characteristics of the modeling group and the validation group of patients with T1 BC

Characteristics	Modeling group				Validation group			
	N0–N1	N2–N3	t/χ^2	P value	N0–N1	N2–N3	t/χ^2	P value
Age (M ± SD)	62.47±12.65	58.36±12.85	18.820	<0.001	62.53±12.69	58.82±13.02	10.997	<0.001
Race, n (%)			88.142	<0.001			56.148	<0.001
White	112,310 (81.9)	2,683 (78.0)			48,116 (81.8)	1,117 (76.9)		
Black	11,792 (8.6)	452 (13.1)			5,095 (8.7)	207 (14.2)		
Other	12,422 (9.1)	291 (8.5)			5,300 (9.0)	119 (8.2)		
Unknown	690 (0.5)	13 (0.4)			316 (0.5)	10 (0.7)		
Pathological type, n (%)			102.090	<0.001			40.987	<0.001
Infiltrating duct carcinoma	105,415 (76.8)	2,632 (76.5)			45,186 (76.8)	1,108 (76.3)		
Lobular carcinoma	10,733 (7.8)	352 (10.2)			4,651 (7.9)	141 (9.7)		
Ductal carcinoma, micropapillary	498 (0.4)	28 (0.8)			197 (0.3)	13 (0.9)		
Metaplastic carcinoma	270 (0.2)	8 (0.2)			120 (0.2)	2 (0.1)		
Medullary carcinoma	226 (0.2)	6 (0.2)			88 (0.1)	1 (0.1)		
Mucinous adenocarcinoma	2,971 (2.2)	9 (0.3)			1,313 (2.2)	6 (0.4)		
Others	17,101 (12.5)	404 (11.7)			7,272 (12.4)	182 (12.5)		
Grade, n (%)			969.380	<0.001			470.380	<0.001
I	42,095 (30.7)	413 (12.0)			17,967 (30.5)	159 (10.9)		
II	60,744 (44.3)	1,507 (43.8)			26,133 (44.4)	635 (43.7)		
III	28,484 (20.8)	1,370 (39.8)			12,184 (20.7)	602 (41.4)		
IV	305 (0.2)	15 (0.4)			126 (0.2)	5 (0.3)		
Unknown	5,586 (4.1)	134 (3.9)			2,417 (4.1)	52 (3.6)		

Table 1 (*continued*)

Table 1 (*continued*)

Project	Modeling group				Validation group			
	N0–N1	N2–N3	t/χ^2	P value	N0–N1	N2–N3	t/χ^2	P value
T, n (%)			1,045.300	<0.001			403.470	<0.001
T1mic	4,515 (3.3)	24 (0.7)			2,010 (3.4)	5 (0.3)		
T1a	16,663 (12.1)	112 (3.3)			7,226 (12.3)	62 (4.3)		
T1b	40,023 (29.2)	453 (13.2)			17,169 (29.2)	205 (14.1)		
T1c	75,611 (55.1)	2,829 (82.3)			32,259 (54.8)	1,178 (81.1)		
T1NOS	402 (0.3)	21 (0.6)			163 (0.3)	3 (0.2)		
ER, n (%)			147.890	<0.001			102.780	<0.001
Positive	118,364 (86.3)	2,747 (79.9)			50,622 (86.1)	1,132 (77.9)		
Negative	16,116 (11.7)	637 (18.5)			6,939 (11.8)	298 (20.5)		
Unknown	2,734 (2.0)	55 (1.6)			1,266 (2.2)	23 (1.6)		
PR, n (%)			175.420	<0.001			98.990	<0.001
Positive	104,647 (76.3)	2,308 (67.1)			44,780 (76.1)	957 (65.9)		
Negative	29,271 (21.3)	1,057 (30.7)			12,526 (21.3)	467 (32.1)		
Unknown	3,296 (2.4)	74 (2.2)			1,521 (2.6)	29 (2.0)		
HER2, n (%)			197.200	<0.001			96.138	<0.001
Positive	15,041 (11.0)	636 (18.5)			6,485 (11.0)	276 (19.0)		
Negative	113,402 (82.6)	2,634 (76.6)			48,447 (82.4)	1,112 (76.5)		
Unknown	8,771 (6.4)	169 (4.9)			3,895 (6.6)	65 (4.5)		
Molecular subtype, n (%)			303.320	<0.001			163.750	<0.001
Luminal A	102,927 (75.0)	2,251 (65.5)			43,968 (74.7)	939 (64.6)		
Luminal B	3,162 (2.3)	149 (4.3)			1,291 (2.2)	51 (3.5)		
Triple negative	10,299 (7.5)	382 (11.1)			4,400 (7.5)	169 (11.6)		
HER2 enriched	4,009 (2.9)	200 (5.8)			1,772 (3.0)	103 (7.1)		
Triple positive	7,828 (5.7)	286 (8.3)			3,404 (5.8)	122 (8.4)		
Unknown	8,989 (6.6)	171 (5.0)			3,992 (6.8)	69 (4.7)		
Primary site, n (%)			267.660	<0.001			155.910	<0.001
Nipple/central portion	6,025 (4.4)	177 (5.1)			2,546 (4.3)	63 (4.3)		
Upper-inner quadrant	18,760 (13.7)	273 (7.9)			7,974 (13.6)	116 (8.0)		
Lower-inner quadrant	8,741 (6.4)	186 (5.4)			3,824 (6.5)	77 (5.3)		
Upper-outer quadrant	46,773 (34.1)	1,204 (35.0)			20,113 (34.2)	472 (32.5)		
Lower-outer quadrant	10,767 (7.8)	337 (9.8)			4,567 (7.8)	147 (10.1)		
Axillary tail	729 (0.5)	47 (1.4)			296 (0.5)	21 (1.4)		
Overlapping lesion	31,642 (23.1)	679 (19.7)			13,660 (23.2)	302 (20.8)		
Breast	13,777 (10.0)	536 (15.6)			5,847 (9.9)	255 (17.5)		

BC, breast cancer; M ± SD, mean ± standard deviation; NOS, not otherwise specified; ER, estrogen receptor; PR, progesterone receptor; HER2, human epidermal growth factor receptor 2.
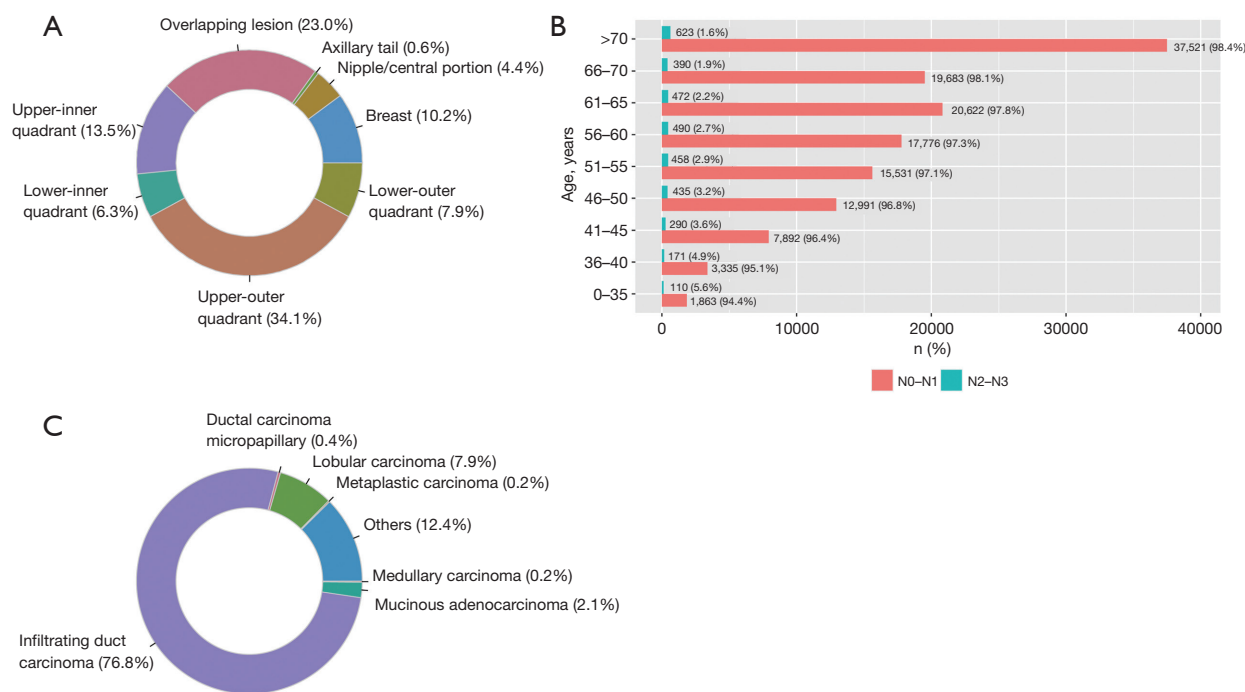
Figure 2 Distribution map of primary tumor location, age, and pathological type in T1 patients. (A) Proportion of primary tumor location; (B) number and proportion of N stages in different age groups; (C) distribution of different pathological types.

Results of univariate and multivariate logistic analysis of the modeling group

The significant variables in univariate logistic analysis were included in multivariate logistic regression analysis, and the results showed that patient age, race, pathological type, grade, T stage, progesterone receptor (PR) status, and location were associated with T1 LABC ANLM. The probability of metastasis for T1 BC ANLM was higher with PR-negative tumors located in the upper outer quadrant and axillary tail as the age decreased, grade increased, and T stage increased. The probabilistic risk of developing ANLM was found to be 1.032 times higher in micropapillary ductal carcinoma when compared to invasive ductal carcinoma in this study. Refer to *Table 2* for additional information.

T1 BC ALNM prediction model nomogram and model validation

Meaningful variables from multivariate logistic analysis were included in the prediction model and a nomogram was drawn. The individual score of each influencing factor was obtained by the score scale, and then the score of each factor was added to achieve the total score. The ALNM probability was obtained by comparing the scores of the total

scale below the nomogram. It was found that pathological type and tumor diameter had the greatest influence in T1 BC ALNM. The total score and the probability of ALNM are shown in *Figure 3*; the total score of the first patient was 371 and a probability of ALNM was 5.48% (*Figure 3A*). The minimum score of the model was 200, and the minimum ALNM probability was 0.007% (*Figure 3B*). The maximum score of the model was 447, and the maximum ALNM probability was 50.40% (*Figure 3C*). The ROC curve was plotted by RStudio version 4.3.1, and the AUC of the modeling group was calculated to be 0.739 (95% CI: 0.732–0.747) (*Figure 3D*) and the AUC of the validation group was 0.741 (95% CI: 0.729–0.752) (*Figure 3E*). The Bootstrap repeated self-sampling method was used for internal verification of the nomogram model. The calibration curve was obtained through repeated Bootstrap self-sampling for 1,000 times, and it showed that the absolute error between the simulated curve and the actual curve of the modeling group was 0, and the calibration effect of the model was good (Hosmer-Lemeshow P=0.55) (*Figure 3F*). The absolute error between the simulated curve and actual curve of the validation group was 0.001, and the calibration effect of the model was good (Hosmer-

Table 2 Univariate and multivariate logistic analysis of the influencing factors of ALNM in T1 LABC

Influencing factor	Single factor		Multi-factor	
	OR value (95% CI)	P value	OR value (95% CI)	P value
Age	0.975 (0.972–0.977)	<0.001	0.980 (0.978–0.983)	<0.001
Race				
White	1		1	
Black	1.605 (1.448–1.774)	<0.001	1.304 (1.174–1.445)	<0.001
Other	0.981 (0.866–1.106)	0.75	0.902 (0.795–1.020)	0.10
Unknown	0.789 (0.432–1.308)	0.40	0.702 (0.382–1.171)	0.21
Pathological type				
Infiltrating duct carcinoma	1		1	
Lobular carcinoma	1.314 (1.171–1.468)	<0.001	1.599 (1.418–1.797)	<0.001
Ductal carcinoma, micropapillary	2.252 (1.501–3.236)	<0.001	2.032 (1.347–2.943)	<0.001
Metaplastic carcinoma	1.187 (0.538–2.240)	0.63	0.701 (0.315–1.334)	0.33
Medullary carcinoma	1.063 (0.419–2.186)	0.88	0.506 (0.198–1.047)	0.10
Mucinous adenocarcinoma	0.121 (0.058–0.219)	<0.001	0.195 (0.093–0.352)	<0.001
Others	0.946 (0.850–1.051)	0.31	1.085 (0.972–1.207)	0.14
Grade				
I	1		1	
II	2.529 (2.269–2.825)	<0.001	1.993 (1.785–2.229)	<0.001
III	4.902 (4.392–5.484)	<0.001	3.192 (2.826–3.612)	<0.001
IV	5.013 (2.833–8.196)	<0.001	3.793 (2.126–6.272)	<0.001
Unknown	2.445 (2.001–2.968)	<0.001	2.442 (1.985–2.987)	<0.001
T				
T1mic	1		1	
T1a	1.264 (0.828–2.013)	0.30	1.729 (1.125–2.770)	0.02
T1b	2.129 (1.444–3.302)	<0.001	3.071 (2.062–4.800)	<0.001
T1c	7.039 (4.822–10.832)	<0.001	8.405 (5.698–13.041)	<0.001
T1NOS	9.827 (5.382–17.814)	<0.001	10.361 (5.644–18.884)	<0.001
ER				
Positive	1		1	
Negative	1.703 (1.559–1.858)	<0.001	0.969 (0.721–1.277)	0.83
Unknown	0.867 (0.654–1.123)	0.30	0.588 (0.346–1.032)	0.56
PR				
Positive	1		1	
Negative	1.637 (1.520–1.762)	<0.001	1.244 (1.105–1.398)	<0.001
Unknown	1.018 (0.799–1.276)	0.88	1.854 (1.139–2.856)	0.008

Table 2 (*continued*)

Table 2 (*continued*)

Influencing factor	Single factor		Multi-factor	
	OR value (95% CI)	P value	OR value (95% CI)	P value
HER2				
Positive	1		1	
Negative	0.549 (0.503–0.600)	<0.001	0.180 (0.007–4.671)	0.23
Unknown	0.456 (0.383–0.540)	<0.001	0.844 (0.175–15.170)	0.87
Molecular subtype				
Luminal A	1		1	
Luminal B	2.155 (1.811–2.543)	<0.001	0.233 (0.009–6.050)	0.31
Triple negative	1.696 (1.517–1.891)	<0.001	0.850 (0.623–1.180)	0.32
HER2 enriched	2.281 (1.962–2.638)	<0.001	0.238 (0.009–6.236)	0.32
Triple positive	1.671 (1.471–1.890)	<0.001	0.209 (0.008–5.415)	0.28
Unknown	0.870 (0.741–1.014)	0.08	0.191 (0.011–0.890)	0.10
Primary site				
Nipple/central portion	1		1	
Upper-inner quadrant	0.495 (0.410–0.601)	<0.001	0.455 (0.376–0.553)	<0.001
Lower-inner quadrant	0.724 (0.588–0.892)	0.002	0.674 (0.546–0.833)	<0.001
Upper-outer quadrant	0.876 (0.745–1.031)	0.11	0.817 (0.696–0.963)	0.14
Lower-outer quadrant	1.065 (0.887–1.287)	0.50	0.937 (0.779–1.132)	0.50
Axillary tail	2.194 (1.561–3.028)	<0.001	1.744 (1.233–2.423)	0.001
Overlapping lesion	0.730 (0.619–0.866)	<0.001	0.684 (0.579–0.813)	0.001
Breast	1.324 (1.117–1.578)	0.001	1.185 (0.997–1.415)	0.06

ALNM, axillary lymph node metastasis; LABC, locally advanced breast cancer; OR, odds ratio; CI, confidence interval; NOS, not otherwise specified; ER, estrogen receptor; PR, progesterone receptor; HER2, human epidermal growth factor receptor 2.

Lemeshow P=0.80) (*Figure 3G*).

Prediction of the clinical validity of the model

DCA was conducted to evaluate the clinical effectiveness of the predictive model. To assess the clinical utility of a specific model, DCA can be used to integrate patient or decision maker preferences (9). From the decision curve of the modeling group, we can see that the maximum net benefit of the model is 30% and the maximum threshold probability is 58.00% (*Figure 3H*). As the threshold probability increased, and the probability of a patient developing N2–N3 decreased. When the model prediction probability was 0.20, the net benefit of the decision curve of the validation group was 16.67%, or the probability of N2–

N3 was 16.67% (*Figure 3I*).

Discussion

The diagnosis and treatment of BC are constantly developing and progressing as a result of the continuous improvement of modern medicine, but the incidence of BC is increasing year by year. The diameter of LABC tumor is large, which is often accompanied by regional lymph node metastasis, and invasion of the chest wall and skin (2), which makes the diagnosis and treatment of LABC somewhat difficult. This study started with T1 BC and provided reference value for predicting T1 LABC ALNM by establishing a logistic regression predictive model.
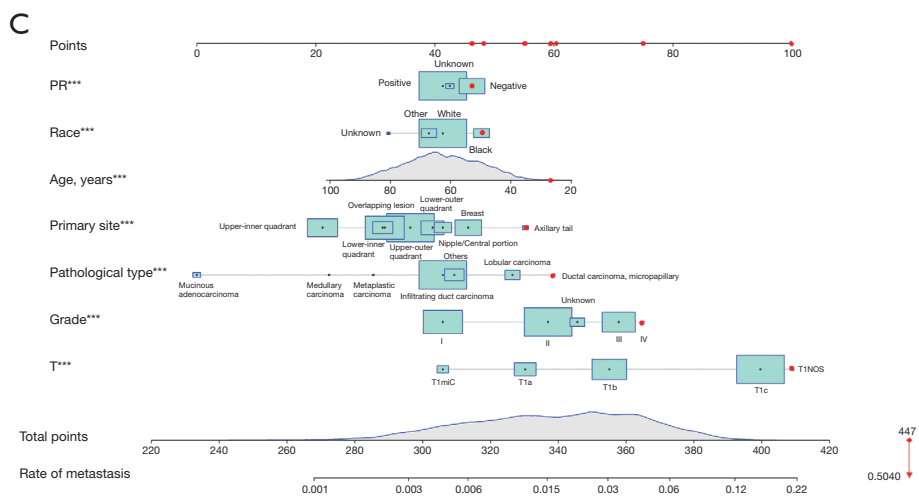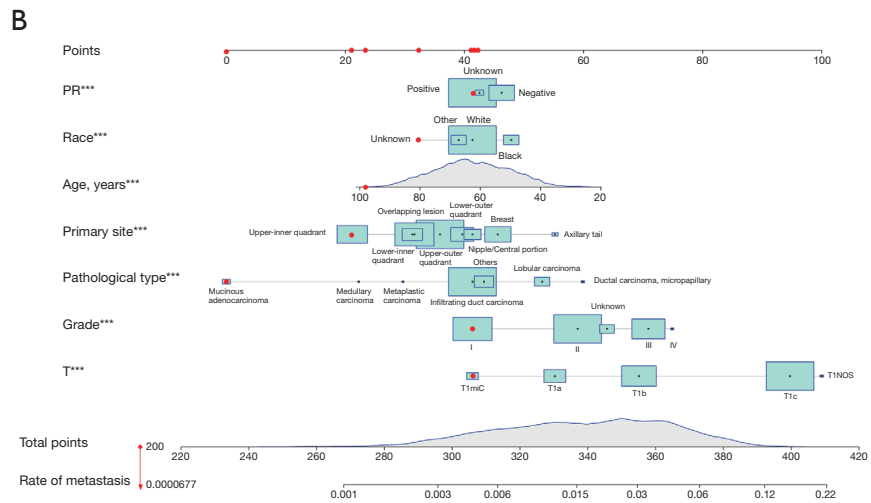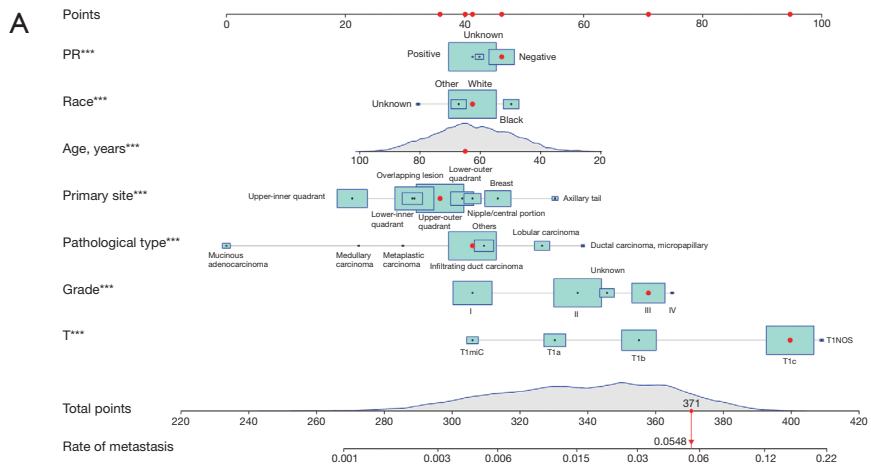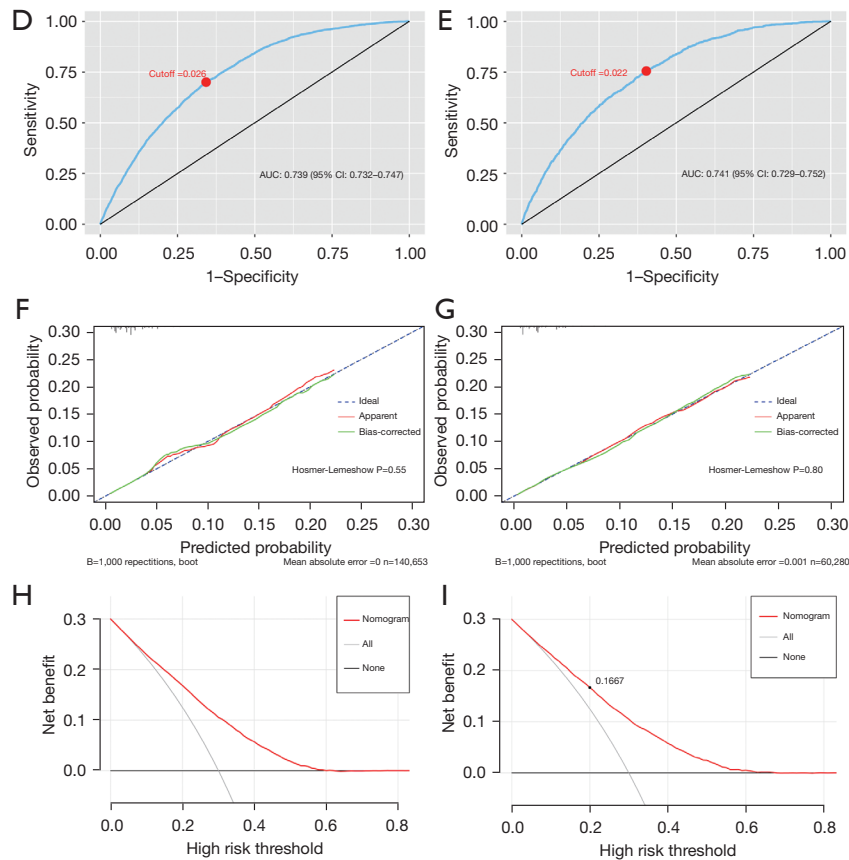
Figure 3 Binary logistic regression prediction model, ROC and DCA. (A-C) Nomogram of LABC logistic regression prediction model in T1 stage. The line segment of each variable is marked with a black dot, which corresponds to the score scale, representing the value range of the variable, and the length of the line segment reflects the contribution of the factor to the outcome event. Points: indicates the individual score of each variable under different values. Total points: represents the total score of the corresponding individual scores after the values of all variables. The blue square size represents the data distribution for each variable in the modeling group. The wavy graph in the total points represents the data distribution of the total scores of the corresponding individual scores after the values of all patients in the modeling group. (A) The red dots in the score and the red dots in the blue squares represent the total score and metastasis probability of all variables corresponding to the 1st patient. (B) The red dots in the score and the red dots in the blue squares represent the minimum total score and metastasis probability for all variables. (C) The red dots in the score and the red dots in the blue squares represent the maximum total score and metastasis probability for all variables. (D) ROC curves for modeling groups. (E) ROC curves for the validation group. (F,G) Calibration curve—apparent: according to the probability of direct prediction by the model, that is, the original prediction probability output by the model; ideal: a perfect prediction situation, where the prediction probability is exactly the same as the observed probability; bias-corrected: the predicted probability corrected by the bootstrap calibration method. (F) modeling group: B=1,000 repetitions, boot and mean absolute error =0 means that the nomogram model is internally verified by the Bootstrap repeated self-sampling method, and the calibration curve is obtained after repeated Bootstrap self-sampling 1,000 times, and the absolute error between the simulated curve and the actual curve is 0, indicating that the trend trajectories of the two curves are basically the same, with strong consistency, and the P value of the Hosmer-Lemeshow goodness-of-fit test is >0.05, so the calibration effect of the model is good. (G) Validation group: mean absolute error =0.001 means that the nomogram model is internally verified by the Bootstrap repeated self-sampling method, and the absolute error between the simulated curve and the actual curve is 0.001. (H,I) DCA: abscissa is the threshold probability and ordinate is the net benefit. The probability of developing N2–N3 is denoted as Pi when the established model reaches a certain value, and when Pi reaches a certain threshold (denoted as Pt), it is defined as positive. The three lines in the figure represent the established model, all positive, all negative, and we can see that with the increase of the threshold probability, the probability of N2–N3 in patients decreases. (H) Modeling group DCA; (I) validation group DCA. ***, P<0.001. PR, progesterone receptor; NOS, not otherwise specified; AUC, area under the curve; CI, confidence interval; ROC, receiver operating characteristic; DCA, decision curve analysis; LABC, locally advanced breast cancer.

Gland Surg 2024;13(6):871-884 | https://dx.doi.org/10.21037/gs-24-34

Several studies have reported that younger patients are more likely to develop lymph node metastases than older patients (10). This is consistent with the results of this study, which found that age is inversely correlated with ALNM in LABC. The survival rate of younger patients decreased by 5% compared to older patients, according to Miller *et al.*'s findings (11). Single-cell RNA sequencing (scRNA-seq) has revealed that multiple tumor-associated pathways and genes are involved in the regulation of BC lymph node metastasis (12). Galectin-1 is involved in lymph node metastasis after interacting with its membrane receptors (13), and genes such as protein tyrosine phosphatase receptor type C (*PTPRC*) and serglycin (*SRGN*) achieve invasive movement and metastasis of BC through drive-related pathways (12). Previous studies have reported that the high expression of cancer-associated fibroblasts (CAFs) promotes the epithelial-mesenchymal transition (EMT) and remodels the matrix to promote tumor invasion (14). High expression of the *GALNT-1* gene, from the polypeptide N-acetylg alactosaminyltransferase (GALNT) family, is associated with reduced survival in young patients and increases the risk of lymph node metastasis. BC with high expression of GALNT-1 enriches the gene sets for EMT (15), and aging also results decreasing in the relative proportion of fibroblasts (16). Low expression of the glycine-N-acyltransferase-like 1 (*GLYATL1*) gene in young BC patients can promote ALNM (17), and *GATA3* gene deficiency promotes the possibility of tumorigenesis and metastasis (18). Analysis of The Cancer Genome Atlas (TCGA) breast dataset showed that *GATA3* mutations were more common in young patients than other age groups (young *vs.* middle-aged *vs.* elderly =15.2% *vs.* 8.2% *vs.* 9%, P=0.003) (19). Younger patients are more prone to lymph node metastasis, as evidenced by all of these findings.

Micropapillary ductal carcinoma is a type of pleomorphic nucleolar carcinoma that can be seen with or without papillary discharge and can be observed by microscopic visualization of specialized microstructures floating in the ductal lumen (20), which represents a very small relative proportion. This study found that micropapillary ductal carcinoma comprised only 0.4% of total BC, but exhibited higher grade nuclear figures compared with non-micropapillary carcinoma. During the course of the study, we noted that the risk of lymph node metastasis in micropapillary ductal carcinoma was 1.032 times higher than that in invasive ductal carcinoma. Previous studies have found that micropapillary ductal carcinoma has a higher rate of ipsilateral recurrence compared with non-

papillary ductal carcinoma (21), which may be related to the susceptibility of the micropapillary structure itself. The shed tumor cells wander through the dilated ducts and form extensive metastasis of the tumor (20). Metastatic tumor cells promote the generation of tumor-associated lymphatic vessels by releasing vascular endothelial growth factor (VEGF), thus promoting lymph node metastasis (22). High expression of human epidermal growth factor receptor 2 (HER2), Ki67, and p53 in have been confirmed to be exist in micropapillary ductal carcinoma (20), which to some extent indicates that micropapillary ductal carcinoma is prone to extensive lymph node metastasis. In the course of searching the literature, we found another special type of cancer that was named invasive micropapillary carcinoma (IMPC). Some studies have reported that the probability of lymph node metastasis of IMPC can increase by 17.40% comparing with invasive ductal carcinoma (23). This study found that N2–N3 metastasis occurred 5.57% (41/736) of T1 micropapillary ductal carcinoma, and N2–N3 metastasis occurred 2.42% (3,740/154,341) of invasive ductal carcinoma, 3.11% (493/15,877) of lobular carcinoma, 2.50% (10/400) of metaplastic carcinoma, and 0.35% (15/4,299) of mucinous adenocarcinoma. This further supports that the tendency of micropapillary ductal carcinoma to the lymph node metastasis may be related to the structure of the micropapillary itself, but the specific mechanism has not been reported in the literature, which is still a huge challenge for people.

PR status, estrogen receptor (ER) status, and HER2 status are one of the most important prognostic factors for BC. This study found that PR status was associated with ALNM, and ALNM was more likely to occur when PR is negative, which was consistent with the conclusions of Pathiraja *et al.* (24). Earlier studies have confirmed that patients that ER+/PR− BC are more aggressive and have a worse prognosis than ER+/PR+ patients (25), and high expression of Ki-67 and p53 in all types of BC increases the probability of ALNM. There is a significantly higher expression of Ki-67 and p53 in PR-negative BC compare with PR-positive BC (26). Chemokine receptor 4 (CXCR4) was highly expressed in metastatic BC (MBC), whereas high expression of CXCR4 was significantly associated with PR-negative BC ALNM (27). The relevant contributing factors of ALNM in PR-negative BC have been continuously reported, and the occurrence of the disease is the result of the continuous accumulation and interaction of different factors. Full understanding of its pathological mechanism is helpful for the treatment of the disease, and the exploration

of the mechanism of ALNM in PR-negative BC is a huge challenge for us at present.

Previous studies have reported that tumor diameter and histological grade are independent risk factors for the development of ALNM (28,29). The risk of ALNM of BC with tumor diameter >10 mm was significantly greater than that of BC with a ≤10 mm diameter. This study also confirmed that the tumor diameter and histological grade of T1 BC were proportional to the risk of ALNM occurring. It is an undeniable fact that there is a correlation between primary location and ALNM. Previous studies have reported that tumors located in the central region are more likely to occur ALNM than others (30). However, this study found that compared with other locations, T1 LABC in axillary tail tumors more frequently exhibited ALNM (axillary tail *vs.* nipple/central portion *vs.* upper inner quadrant *vs.* lower inner quadrant *vs.* upper outer quadrant *vs.* outer lower quadrant =6.23% *vs.* 2.73% *vs.* 1.44% *vs.* 2.05% *vs.* 2.44% *vs.* 3.06%). At the same time, we found that the previous studies did not stratify the tumor in the axillary tail alone, which might explain the inconsistency between this study and the previous studies. But regardless of the results, it reminds us that the location of the primary tumor in T1 BC is critical for the prediction of ALNM.

This study aimed to establish a relevant model to predict whether LABC occurs in T1, and to have a guiding role in treatment. For example, this model calculates the probability value of ALNM, and whether the axillary treatment can be changed for LABC patients with high probability to develop N2–N3. The SINODAR-ONE study reported that overall survival (OS) and recurrence-free survival (RFS) were not significantly different in cN0 T1–2 BC patients with 1–2 large metastatic sentinel lymph nodes (SLNs) (31), Another study reported that in women with T1–2 invasive BC with 1–2 SLN metastases (SLNM) OS was similar between patients treated with SLN biopsy (SLNB) and axillary lymph node dissection (ALND) (32). In this study, we collected baseline information to predict the probability of N2–N3. Whether direct ALND can be considered in T1 LABC patients with cN0 but N2–N3 and high economic burden or refusing SLNB for other reasons, and it can avoid the additional burden and false negative risk caused by SLNB. To confirm this conjecture, additional prospective studies are required.

In clinical work, it is common for small tumors to be underestimated, which can result in errors in understanding patients' conditions, leading clinicians to choose more conservative treatment regimens, which can delay treatment. Establishing a clinical prediction model can prevent clinicians from underestimating tumors to some extent and provide a reference value for treatment. Clinical prediction models can be divided into logistic regression models, decision trees, support vector machines, random forest models, and neural networks. The model establishes a logistic regression prediction model according to the outcome variables, the value of which is to predict the risk of extensive lymph node metastasis in advance, so that clinicians can have a preliminary understanding of patients with small tumors, and provide guidance for formulating more effective treatment regimens, The predicted probabilities of this model range from 0.001% to 50.40%, and N2–N3 can be 50.40% or even higher when high risk factors are met, which may draw clinicians' attention to small tumors.

Conclusions

This study found that in terms of tumor characteristics, patients with N2–N3 were mainly moderately and poorly differentiated, with large tumor diameter, upper-outer quadrant, infiltrating duct carcinoma, and HR positive. This study also found that primary tumors were mainly located in the outer upper quadrant, and the incidence of tumors in the axillary tail was the lowest. Age, race, pathological type, grade, T stage, PR status, and primary site were related to ALNM, and the probability of metastasis of T1 LABC ALNM would be increased because for younger age, lobular carcinoma, micropapillary ductal carcinoma, higher grade, higher T stage, PR-negative, and tumors located in the outer upper quadrant and axillary tail. This study found that the probability of developing ANLM in micropapillary ductal carcinoma was increased by 1.032 times compared to infiltrating duct carcinoma.

Acknowledgments

Footnote

Reporting Checklist: The authors have completed the TRIPOD reporting checklist. Available at https://

gs.amegroups.com/article/view/10.21037/gs-24-34/rc

Peer Review File: Available at https://gs.amegroups.com/article/view/10.21037/gs-24-34/prf

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://gs.amegroups.com/article/view/10.21037/gs-24-34/coif). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. This study was conducted in accordance with the principles of the Declaration of Helsinki (as revised in 2013).

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

References

1. Ferlay J, Colombet M, Soerjomataram I, et al. Cancer statistics for the year 2020: An overview. Int J Cancer 2021. doi:10.1002/ijc.33588.

2. Giuliano AE, Connolly JL, Edge SB, et al. Breast Cancer-Major changes in the American Joint Committee on Cancer eighth edition cancer staging manual. CA Cancer J Clin 2017;67:290-303.

3. Yu Y, Tan Y, Xie C, et al. Development and Validation of a Preoperative Magnetic Resonance Imaging Radiomics-Based Signature to Predict Axillary Lymph Node Metastasis and Disease-Free Survival in Patients With Early-Stage Breast Cancer. JAMA Netw Open 2020;3:e2028086.

4. Shan YN, Xu W, Wang R, et al. A Nomogram Combined Radiomics and Kinetic Curve Pattern as Imaging Biomarker for Detecting Metastatic Axillary Lymph Node in Invasive Breast Cancer. Front Oncol 2020;10:1463.

5. Cardoso F, van't Veer LJ, Bogaerts J, et al. 70-Gene Signature as an Aid to Treatment Decisions in Early-Stage Breast Cancer. N Engl J Med 2016;375:717-29.

6. Mamounas EP, Russell CA, Lau A, et al. Clinical relevance of the 21-gene Recurrence Score(®) assay in treatment decisions for patients with node-positive breast cancer in the genomic era. NPJ Breast Cancer 2018;4:27.

7. Lai J, Chen Z, Liu J, et al. A radiogenomic multimodal and whole-transcriptome sequencing for preoperative prediction of axillary lymph node metastasis and drug therapeutic response in breast cancer: a retrospective, machine learning and international multicohort study. Int J Surg 2024;110:2162-77.

8. Blok EJ, Bastiaannet E, van den Hout WB, et al. Systematic review of the clinical and economic value of gene expression profiles for invasive early breast cancer available in Europe. Cancer Treat Rev 2018;62:74-90.

9. Fitzgerald M, Saville BR, Lewis RJ. Decision curve analysis. JAMA 2015;313:409-10.

10. Partridge AH, Hughes ME, Warner ET, et al. Subtype-Dependent Relationship Between Young Age at Diagnosis and Breast Cancer Survival. J Clin Oncol 2016;34:3308-14.

11. Miller KD, Fidler-Benaoudia M, Keegan TH, et al. Cancer statistics for adolescents and young adults, 2020. CA Cancer J Clin 2020;70:443-59.

12. Xu K, Wang R, Xie H, et al. Single-cell RNA sequencing reveals cell heterogeneity and transcriptome profile of breast cancer lymph node metastasis. Oncogenesis 2021;10:66.

13. Orozco CA, Martinez-Bosch N, Guerrero PE, et al. Targeting galectin-1 inhibits pancreatic cancer progression by modulating tumor-stroma crosstalk. Proc Natl Acad Sci U S A 2018;115:E3769-78.

14. Pelon F, Bourachot B, Kieffer Y, et al. Cancer-associated fibroblast heterogeneity in axillary lymph nodes drives metastases in breast cancer through complementary mechanisms. Nat Commun 2020;11:404.

15. Oshi M, Ziazadeh D, Wu R, et al. GALNT1 Expression Is Associated with Angiogenesis and Is a Prognostic Biomarker for Breast Cancer in Adolescents and Young Adults (AYA). Cancers (Basel) 2023;15:3489.

16. Li CM, Shapiro H, Tsiobikas C, et al. Aging-Associated Alterations in Mammary Epithelia and Stroma Revealed by Single-Cell RNA Sequencing. Cell Rep 2020;33:108566.

17. Siddig A, Wan Abdul Rahman WF, Mohd Nafi SN, et al. Comparing the Biology of Young versus Old Age Estrogen-Receptor-Positive Breast Cancer through Gene and Protein Expression Analyses. Biomedicines 2023;11:200.

18. Bai F, Zhang LH, Liu X, et al. GATA3 functions

downstream of BRCA1 to suppress EMT in breast cancer. Theranostics 2021;11:8218-33.

19. Luen SJ, Viale G, Nik-Zainal S, et al. Genomic characterisation of hormone receptor-positive breast cancer arising in very young women. Ann Oncol 2023;34:397-409.

20. Castellano I, Marchiò C, Tomatis M, et al. Micropapillary ductal carcinoma in situ of the breast: an inter-institutional study. Mod Pathol 2010;23:260-9.

21. Evers K. Significance of finding micropapillary DCIS on core needle biopsy. Acad Radiol 2011;18:795-6.

22. Karaman S, Detmar M. Mechanisms of lymphatic metastasis. J Clin Invest 2014;124:922-8.

23. Chen AC, Paulino AC, Schwartz MR, et al. Population-based comparison of prognostic factors in invasive micropapillary and invasive ductal carcinoma of the breast. Br J Cancer 2014;111:619-22.

24. Pathiraja TN, Shetty PB, Jelinek J, et al. Progesterone receptor isoform-specific promoter methylation: association of PRA promoter methylation with worse outcome in breast cancer patients. Clin Cancer Res 2011;17:4177-86.

25. Rakha EA, El-Sayed ME, Green AR, et al. Biologic and clinical characteristics of breast cancer with single hormone receptor positive phenotype. J Clin Oncol 2007;25:4772-8.

26. Wintzer HO, Zipfel I, Schulte-Mönting J, et al. Ki-67 immunostaining in human breast tumors and its

relationship to prognosis. Cancer 1991;67:421-8.

27. Woo SU, Bae JW, Kim CH, et al. A significant correlation between nuclear CXCR4 expression and axillary lymph node metastasis in hormonal receptor negative breast cancer. Ann Surg Oncol 2008;15:281-5.

28. Phung MT, Tin Tin S, Elwood JM. Prognostic models for breast cancer: a systematic review. BMC Cancer 2019;19:230.

29. Xiong J, Zuo W, Wu Y, et al. Ultrasonography and clinicopathological features of breast cancer in predicting axillary lymph node metastases. BMC Cancer 2022;22:1155.

30. Gao X, Luo W, He L, et al. Nomogram models for stratified prediction of axillary lymph node metastasis in breast cancer patients (cN0). Front Endocrinol (Lausanne) 2022;13:967062.

31. Tinterri C, Canavese G, Gatzemeier W, et al. Sentinel lymph node biopsy versus axillary lymph node dissection in breast cancer patients undergoing mastectomy with one to two metastatic sentinel lymph nodes: sub-analysis of the SINODAR-ONE multicentre randomized clinical trial and reopening of enrolment. Br J Surg 2023;110:1143-52.

32. Giuliano AE, Ballman KV, McCall L, et al. Effect of Axillary Dissection vs No Axillary Dissection on 10-Year Overall Survival Among Women With Invasive Breast Cancer and Sentinel Node Metastasis: The ACOSOG Z0011 (Alliance) Randomized Clinical Trial. JAMA 2017;318:918-26.