Original Article

Incidence and risk factors of lymph node metastasis in breast cancer patients without preoperative chemoradiotherapy and neoadjuvant therapy: analysis of SEER data

Mingpeng Luo[1,2,3#], Xixi Lin[1,2#], Dingji Hao[4], Kangle Wang Shen[1,2], Wenxin Wu[1,2], Linbo Wang[1,2], Shanming Ruan[3], Jichun Zhou[1,2]

[1]Department of Surgical Oncology, Sir Run Run Shaw Hospital, Zhejiang University School of Medicine, Hangzhou, China; [2]Biomedical Research Center and Key Laboratory of Biotherapy of Zhejiang Province, Hangzhou, China; [3]The First Affiliated Hospital of Zhejiang Chinese Medical University, Hangzhou, China; [4]Department of Thyroid Breast Hernia Surgery, Tonglu County Hospital of Traditional Chinese Medicine, Hangzhou, China

Contributions: (I) Conception and design: J Zhou, M Luo; (II) Administrative support: L Wang, S Ruan, J Zhou; (III) Provision of study materials or patients: M Luo; (IV) Collection and assembly of data: All authors; (V) Data analysis and interpretation: M Luo, L Wang, S Ruan, J Zhou; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

[#]These authors contributed equally to this work.

Correspondence to: Jichun Zhou, MD, PhD. Department of Surgical Oncology, Sir Run Run Shaw Hospital, Zhejiang University School of Medicine, No. 3 East Qingchun Road, Hangzhou 310016, China; Biomedical Research Center and Key Laboratory of Biotherapy of Zhejiang Province, No. 3 East Qingchun Road, Hangzhou 310016, China. Email: Jichun-zhou@zju.edu.cn; Shanming Ruan, MD. The First Affiliated Hospital of Zhejiang Chinese Medical University, No. 54 Youdian Road, Hangzhou 310014, China. Email: shanmingruan@zcmu.edu.cn; Linbo Wang, MD. Department of Surgical Oncology, Sir Run Run Shaw Hospital, Zhejiang University School of Medicine, No. 3 East Qingchun Road, Hangzhou 310016, China; Biomedical Research Center and Key Laboratory of Biotherapy of Zhejiang Province, No. 3 East Qingchun Road, Hangzhou 310016, China. Email: linbowang@zju.edu.cn.

Background: Breast cancer (BC) is the leading cause of death in the female reproductive system, often linked to lymph node involvement, indicating poor prognosis. This study investigated lymph node metastasis incidence and risk factors in M0 stage BC patients who hadn't received preoperative chemoradiotherapy or neoadjuvant therapy. We explored the influence of various factors on lymph node metastasis.

Methods: We conducted a retrospective analysis using Surveillance, Epidemiology, and End Results data from BC patients diagnosed between 2010 and 2015. Binary logistic regression and propensity score matching (PSM) assessed significant factors in BC patients without preoperative treatment. We developed predictive nomograms and evaluated model performance using the concordance index, calibration curve, area under the curve, and decision curve analysis.

Results: Among 256,504 eligible BC patients, 25.57% had lymph node metastasis. Multivariate logistic regression revealed associations between lymph node metastasis and younger age, African-American ethnicity, central/nipple location, lobular carcinoma, human epidermal growth factor receptor 2 (HER2)-positive status, grade III classification, and T3 stage. PSM confirmed these findings. Interactions were identified between age, race, primary site, histology, breast subtype, grade, and T stage, all influencing lymph node metastasis.

Conclusions: This retrospective study identified lymph node metastasis in female BC patients with distinct clinicopathological characteristics who received no preoperative treatment. We constructed valuable nomograms, revealing that: (I) young age (<35 years), African-American race, central/nipple location, infiltrating duct carcinoma, HER2 positivity, high histological grade (grade III), and larger tumor size are risk factors for regional lymph node metastasis; (II) lymph node metastasis may not solely represent the invasive nature of triple-negative BC; (III) patients with different BC subtypes in T1c–T2 stages may benefit from individualized neoadjuvant treatment strategies.

Keywords: Breast cancer (BC); lymph node metastasis; nomogram; risk factors; propensity score matching (PSM)

Submitted Jun 17, 2023. Accepted for publication Nov 02, 2023. Published online Nov 17, 2023.
doi: 10.21037/gs-23-258
View this article at: https://dx.doi.org/10.21037/gs-23-258

Introduction

Breast cancer (BC) is the leading cause of death in the female reproductive system, accounting for 31% of cancer cases and 15% of cancer-related deaths in women. The incidence of BC in women is increasing by 0.5% annually (1). Based on statistical data, the overall incidence rate of positive sentinel lymph nodes in BC patients is approximately 33%. Patients who test positive for lymph node involvement often experience a higher mortality rate and an elevated risk of disease recurrence (2,3).

The lymphatic vessels within the breast form an open system within the surrounding matrix environment (4). As breast tumors develop, they have a significant chance of invading nearby lymph nodes or lymphatic vessels, using the lymphatic system within the breast for metastasis, resulting in multiple invasive tumor foci. The high heterogeneity of BC contributes to different disease progression and prognosis. Therefore, certain patient and tumor characteristics remain valuable in assessing the status of lymph node metastasis (5). Evaluating the burden of lymph node metastasis in BC is crucial for neoadjuvant therapy (NAT) planning, initial surgical procedures, and guiding axillary treatment (6). Although existing literature has described some risk factors for lymph node metastasis (2,7,8), a comprehensive and in-depth analysis of various clinicopathological factors influencing lymph node metastasis has not been conducted. Moreover, the sample sizes in previous investigations are often small or limited to data from a single medical center, which may lack representativeness. Therefore, based on a large sample of retrospective survey data from the period of 2010 to 2015, this study aims to construct a predictive model to examine the accuracy of previous experiences and provide a detailed analysis and summary of each predictive factor. This will lay the foundation for further refinement of future predictive models.

First, we used chi-square tests to screen for influencing factors with significant differences in these large sample data. Then, we conducted univariate and multivariate logistic regression analyses to analyze the impact of different subgroups within each factor on lymph node positivity. Additionally, the influence of different clinicopathological characteristics on lymph node metastasis in BC may yield inconsistent results due to uncontrolled potential confounding factors. Therefore, it is essential to employ propensity score analysis to assess the association between each independent clinicopathological factor of interest and the outcome (9). To create a balanced cohort, we utilized inverse probability weighting (IPW). Our plan is to evaluate the impact of confounding factors on lymph node metastasis by comparing univariate and multivariate logistic regression analyses before and after propensity score matching (PSM). Subsequently, we will construct a nomogram based on consistent influencing factors identified from the results of univariate logistic regression analysis, multivariate logistic regression analysis, and univariate logistic regression analysis after PSM. We will also investigate potential reasons for any discrepancies between PSM and univariate/multivariate logistic regression in specific clinicopathological characteristics. We present this article in accordance with the TRIPOD reporting checklist (available at https://gs.amegroups.com/article/view/10.21037/gs-23-258/rc).

Highlight box

Key findings

- A nomogram has been developed to predict the probability of lymph node metastasis in a patient who has not received any preoperative treatment.

What is known and what is new?

- Prior to this study, the existing nomograms used for predicting lymph node metastasis in breast cancer (BC) were constructed with limited data and lacked adequate representativeness.
- In this study, a substantial sample size was used to create nomograms that incorporate a range of clinical and pathological factors for the prediction of lymph node metastasis in BC. These nomograms serve as more precise and reliable tools to support clinical medical decision-making.

What is the implication, and what should change now?

- This study, based on the established nomogram, identified certain potential risk factors associated with lymph node metastasis.

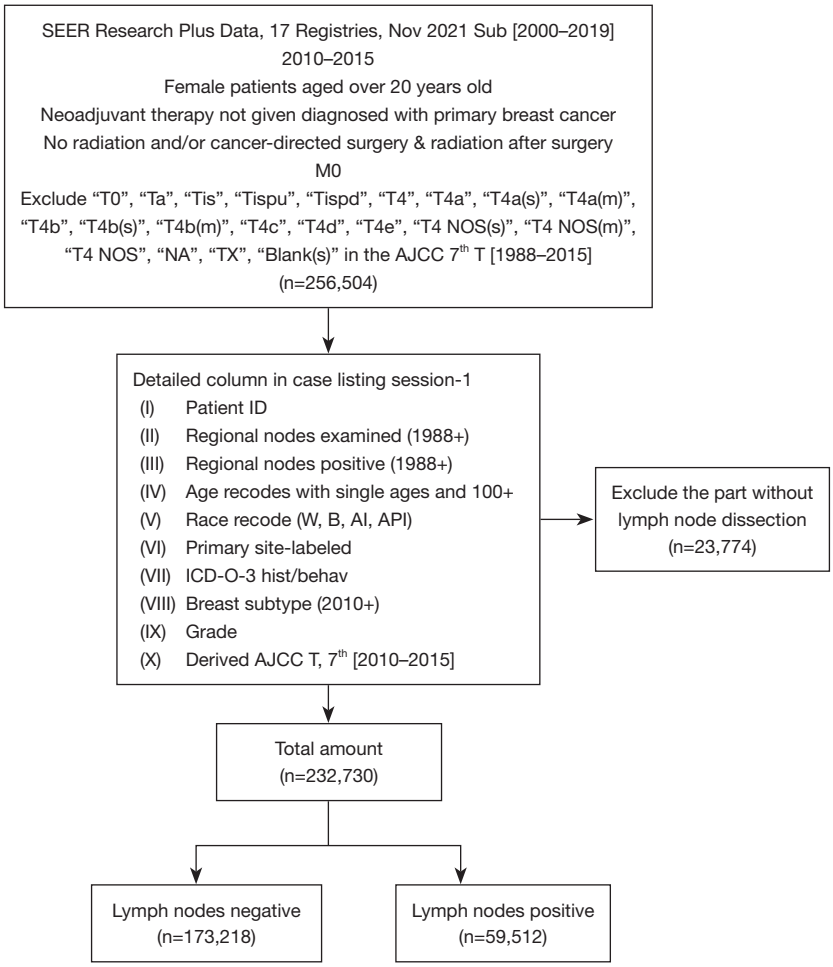

Figure 1 The flow diagram of participant inclusion and exclusion. SEER, Surveillance, Epidemiology, and End Results; NOS, not otherwise specified; AJCC T, the T staging of Cancer Staging Manual of the American Joint Committee on Cancer; W, White; B, Black; AI, American Indian or Alaskan Native; API, Asian or Pacific Islander; ICD-O-3 hist/behav, 3rd edition of the International Classification of Diseases for Oncology histology code and behavior.

Methods

Patient selection

We searched and downloaded medical records of BC patients from the Surveillance, Epidemiology, and End Results (SEER) database (SEER Research Plus Data, 17 Registries, Nov 2021 Sub, 2000–2019), which covers cancer incidence and survival records of over one-third of the US population. We collected data from 256,504 patients based on the following criteria: (I) January 2010 to December 2015; (II) female patients aged over 20 years old; (III) no NAT given; (IV) diagnosed with primary BC; (V) no radiation and/or cancer-directed surgery & radiation after surgery; (VI) M0 stage; (VII) unilateral and unifocal tumor; (VIII) not classified as "T0", "Ta", "Tis", "Tispu", "Tispd", "T4", "T4a", "T4a(s)", "T4a(m)", "T4b", "T4b(m)", "T4(c)", "T4d", "T4e", "T4 NOS(s)", "T4 NOS(m)", "T4 NOS", "NA", "TX", and "Blank(s)" in the T staging of the 7th Cancer Staging Manual of the American Joint Committee on Cancer (AJCC 7th T 1988–2015); and (IX) exclusion of multiple BC lesions. The flowchart of participant inclusion and exclusion is shown in *Figure 1*. Ethical approval was not required for this study as the clinical data of recruited BC patients were collected from publicly available and anonymized data in the SEER database. The study was conducted in accordance with the

Declaration of Helsinki (as revised in 2013).

Variable description

The demographic and clinical characteristics of the participants are as follows: (I) the independent variables included age at diagnosis (20–44, 45–55, 56+ years), race (White, Black, Asian or Pacific Islander, others), primary site (upper-outer quadrant, upper-inner quadrant, lower-outer quadrant, lower-inner quadrant, central portion & nipple, others), histology (infiltrating duct carcinoma, lobular carcinoma, mucinous adenocarcinoma, others), breast subtype [hormone receptor (HR)$^+$/human epidermal growth factor receptor 2 (HER2)$^-$, HR$^-$/HER2$^-$, HR$^+$/HER2$^+$, HR$^-$/HER2$^+$, unknown; +, positive; –, negative], grade (I = well differentiated, II = moderately differentiated, III = poorly differentiated, IV = undifferentiated, unknown), T stage (derived AJCC, 7th) (T1mic, T1a, T1b, T1c, T2, T3, unknown; mic, microinvasive carcinoma); and (II) the outcome variable is determined based on the information provided in the "Regional nodes positive (1988+)" column, where values greater than 1 are considered to be positive for lymph node involvement.

Statistical analysis

This study employed a retrospective cross-sectional survey for statistical analysis. Patient and tumor characteristics were presented as percentages, and data were analyzed using chi-squared tests or Fisher's exact tests. A P value less than 0.05 was considered statistically significant. All statistical analyses were performed using SPSS version 25.0. In order to visualize the differences more intuitively, relevant analysis results were transformed into stacked bar charts using Microsoft Office Excel 2019. We conducted univariate and multivariate logistic regression analyses using SPSS version 25.0. Furthermore, R software (version 4.2.1) was utilized to perform propensity score reweighting analysis on selected data based on individual independent factors, followed by univariate binary logistic regression analysis. Nomograms based on regression models, calibration curves were generated using various functional packages such as RMS, Foreign, Cmprsk, and other software. A two-tailed P value less than 0.05 was considered statistically significant (*, P<0.05; **, P<0.01). We validated the developed predictive model using internally generated validation data through a 3:1 random allocation. The sensitivity and specificity of the constructed nomogram were reflected using receiver operating characteristic (ROC) curves, while its accuracy was evaluated and validated using calibration plots. Additionally, decision curve analysis (DCA) was utilized to calculate the net benefits for each risk threshold probability.

Results

Baseline characteristics of enrolled patients

Figure 2 presents the overall cohort, which included a total of 232,730 eligible BC patients. Among them, 25.57% (n=59,512) had positive lymph nodes (PLNs), and 74.43% (n=173,218) had negative lymph nodes (NLNs). Additional demographic and clinical characteristics of the recruited BC patients were also provided. Significant differences were observed in age, race, primary site, histology, breast subtype, grade, T stage (derived AJCC, 7th), and other factors. In the age group, the 56+ subgroup had the highest number of patients (n=155,765). Among the three age groups, the 20–44 subgroup had the highest percentage of PLN (35.13%), indicating that PLN is more likely to occur in younger women. In the race group, the largest number of patients were white people (n=187,111), while the highest percentage of PLN was observed in black patients (29.94%). In the analysis and comparison within the group of primary sites, excluding "others", the upper-outer quadrant had the highest number of patients (n=80,331), and the central portion & nipple had the highest percentage of PLN (34.47%). In the group of histology, infiltrating duct carcinoma had the highest number of patients (n=173,008), while lobular carcinoma had the highest percentage of PLN (29.72%). Among the group of breast subtype, the HR$^+$/HER2$^-$ subgroup had the highest number of patients, but the highest percentage of PLN was observed in HR$^-$/HER2$^+$ BC patients (31.16%). In the grade group, grade II had the highest number of patients (n=27,702), while grade III had the largest percentage of PLN (33.83%). In the group of T stage, the largest number of patients belonged to the T1c category (n=88,707), while the highest percentage of PLN was observed in the T3 category. Consistent with our expectations, larger tumors were associated with a higher likelihood of lymph node metastasis. Additionally, the baseline characteristics of the training and validation cohorts are provided in Tables S1-S7.

Logistic regression analysis and propensity score validation of lymph node occurrence in BC patients

Single-factor and multiple-factor logistic regression analyses

Characteristics/Variables	Regional nodes status			negative positive
	Negative n (%)	Positive n (%)	Total n (%)	
	173,218 (74.43)	59,512 (25.57)	232,730 (100.00)	74.43% 25.57%
Age (years)				
20-44	14,201 (64.87)	7,692 (35.13)	21,893 (100.00)	64.87% 35.13%
45-55	39,138 (71.07)	15,934 (28.93)	55,072 (100.00)	71.07% 28.93%
56+	119,879 (76.96)	35,886 (23.04)	155,765 (100.00)	76.96% 23.04%
Race				
White	140,051 (74.85)	47,060 (25.15)	187,111 (100.00)	74.85% 25.15%
Black	15,557 (70.06)	6,647 (29.94)	22,204 (100.00)	70.06% 29.94%
Asian or Pacific Islander	15,861 (75.57)	5,128 (24.43)	20,989 (100.00)	75.57% 24.43%
Others	1,749 (72.09)	677 (27.91)	2,426 (100.00)	72.09% 27.91%
Primary Site				
Upper-outer quadrant	59,123 (73.60)	21,208 (26.40)	80,331 (100.00)	73.60% 26.40%
Upper-inner quadrant	24,992 (83.03)	5,108 (16.97)	30,100 (100.00)	83.03% 16.97%
Lower-outer quadrant	13,334 (73.47)	4,814 (26.53)	18,148 (100.00)	73.47% 26.53%
Lower-inner quadrant	10,969 (80.13)	2,720 (19.87)	13,689 (100.00)	80.13% 19.87%
Central portion & Nipple	7,430 (65.53)	3,909 (34.47)	11,339 (100.00)	65.53% 34.47%
Others	57,370 (72.51)	21,753 (27.49)	79,123 (100.00)	72.51% 27.49%
Histology				
Infiltrating duct carcinoma	128,864 (74.48)	44,144 (25.52)	173,008 (100.00)	74.48% 25.52%
Lobular carcinoma	15,892 (70.28)	6,719 (29.72)	22,611 (100.00)	70.28% 29.72%
Mucinous adenocarcinoma	4,437 (93.16)	326 (6.84)	4,763 (100.00)	93.16% 6.84%
Others	24,025 (74.27)	8,323 (25.73)	32,348 (100.00)	74.27% 25.73%
Breast Subtype				
HR+/HER2-	128,733 (74.82)	43,319 (25.18)	172,052 (100.00)	74.82% 25.18%
HR-/HER2-	15,871 (75.48)	5,155 (24.52)	21,026 (100.00)	75.48% 24.52%
HR+/HER2+	13,536 (69.16)	6,036 (30.84)	19,572 (100.00)	69.16% 30.84%
HR-/HER2+	5,159 (68.84)	2,335 (31.16)	7,494 (100.00)	68.84% 31.16%
Unknown	9,919 (78.81)	2,667 (21.19)	12,586 (100.00)	78.81% 21.19%
Grade				
Grade I	49,161 (84.84)	8,787 (15.16)	57,948 (100.00)	84.84% 15.16%
Grade II	75,561 (73.17)	27,702 (26.83)	103,263 (100.00)	73.17% 26.83%
Grade III	42,205 (66.17)	21,573 (33.83)	63,778 (100.00)	66.17% 33.83%
Grade IV	423 (71.82)	166 (28.18)	589 (100.00)	71.82% 28.18%
Unknown	5,868 (82.05)	1,284 (17.95)	7,152 (100.00)	82.05% 17.95%
Derived AJCC T, 7th				
T1mic	4,730 (95.34)	231 (4.66)	4,961 (100.00)	95.34% 4.66%
T1a	17,914 (94.61)	1,020 (5.39)	18,934 (100.00)	94.61% 5.39%
T1b	41,885 (90.32)	4,491 (9.68)	46,376 (100.00)	90.32% 9.68%
T1c	68,621 (77.36)	20,086 (22.64)	88,707 (100.00)	77.36% 22.64%
T2	36,539 (56.60)	28,012 (43.40)	64,551 (100.00)	56.60% 43.40%
T3	3,208 (36.39)	5,608 (63.61)	8,816 (100.00)	36.39% 63.61%
Unknown	321 (83.38)	64 (16.62)	385 (100.00)	83.38% 16.62%

Figure 2 Clinicopathological features of BC patients diagnosed with and without lymph node metastasis (chi-square test; P<0.001 for each independent factor). HR, hormone receptor; HER2, human epidermal growth factor receptor 2; AJCC T, the T staging of Cancer Staging Manual of the American Joint Committee on Cancer; mic, microinvasive carcinoma; BC, breast cancer.

were conducted to assess the independent risk factors for lymph node occurrence (PLN) in newly diagnosed BC patients. The results showed that the diagnosis age, race, primary site, histology, breast subtype, grade, and T stage (derived AJCC, 7th) was significantly associated with PLN occurrence (*Table 1*).

In the age group, the 45–55 subgroup [odds ratio (OR) =0.86; 95% confidence interval (CI): 0.83–0.892, P<0.001] and the 56+ subgroup (OR =0.652; 95% CI: 0.631–0.674, P<0.001) had significantly lower PLN occurrence rates compared to the 20–44 subgroup. In terms of race, Asian or Pacific Islander had a significantly lower risk of PLN occurrence (OR =0.864; 95% CI: 0.834–0.896; P<0.001) compared to black people (OR =1.152; 95% CI: 1.114–1.191; P<0.001). In the group of primary sites, there was no significant difference in PLN occurrence between the upper-outer quadrant and lower-outer quadrant (OR =1; 95% CI: 0.962–1.041; P=0.993). The upper-inner quadrant (OR =0.567; 95% CI: 0.547–0.588; P<0.001) and lower-inner quadrant (OR =0.754; 95% CI: 0.718–0.791; P<0.001) had a lower PLN risk compared to the upper-outer quadrant, while the central portion & nipple had the highest PLN risk (OR =1.387; 95% CI: 1.324–1.452; P<0.001). In terms of histology, mucinous adenocarcinoma had a significantly lower risk of PLN (OR =0.21; 95% CI: 0.187–0.236; P<0.001) compared to infiltrating duct carcinoma. Infiltrating lobular carcinoma showed inconsistent results in single-factor analysis (OR =1.234; 95% CI: 1.197–1.272;

Table 1 Results of univariate and multivariate logistic regression analysis and validation of PSM in BC patients

Characteristics	Total (n=232,730)	Single factor regression analysis		Multi-factor regression analysis		PSM		
		OR (95% CI)	P value	OR (95% CI)	P value	N	OR (95% CI)	P value
Age (years)						113,000		
20–44	21,893	Reference		Reference		12,772	Reference	
45–55	55,072	0.752 (0.727, 0.777)	<0.001	0.86 (0.83, 0.892)	<0.001	28,218	0.87 (0.834, 0.908)	<0.001
56+	155,765	0.553 (0.536, 0.57)	<0.001	0.652 (0.631, 0.674)	<0.001	72,010	0.676 (0.651, 0.702)	<0.001
Race/ethnicity						112,160		
White	187,111	Reference		Reference		88,650	Reference	
Black	22,204	1.272 (1.233, 1.311)	<0.001	1.152 (1.114, 1.191)	<0.001	11,718	1.09 (1.049, 1.133)	<0.001
Asian or Pacific Islander	20,989	0.962 (0.931, 0.995)	0.023	0.864 (0.834, 0.896)	<0.001	10,554	0.879 (0.845, 0.916)	<0.001
Others	2,426	1.152 (1.054, 1.26)	0.002	1.071 (0.972, 1.18)	<0.001	1,238	1.125 (1.006, 1.259)	0.04
Primary site						113,164		
Upper-outer quadrant	80,331	Reference		Reference		39,261	Reference	
Upper-inner quadrant	30,100	0.57 (0.551, 0.589)	<0.001	0.567 (0.547, 0.588)	<0.001	13,097	0.564 (0.541, 0.587)	<0.001
Lower-outer quadrant	18,148	1.006 (0.97, 1.044)	0.729	1 (0.962, 1.041)	0.993	8,993	1.004 (0.959, 1.051)	0.87
Lower-inner quadrant	13,689	0.691 (0.661, 0.723)	<0.001	0.754 (0.718, 0.791)	<0.001	6,048	0.729 (0.691, 0.77)	<0.001
Central portion & nipple	11,339	1.467 (1.407, 1.529)	<0.001	1.387 (1.324, 1.452)	<0.001	6,164	1.359 (1.287, 1.436)	<0.001
Others	79,123	1.057 (1.034, 1.081)	<0.001	1.008 (0.984, 1.032)	0.534	39,601	1.002 (0.974, 1.03)	0.895
Histology						112,268		
Infiltrating duct carcinoma	173,008	Reference		Reference		82,470	Reference	
Lobular carcinoma	22,611	1.234 (1.197, 1.272)	<0.001	0.907 (0.876, 0.939)	<0.001	12,338	0.923 (0.889, 0.959)	<0.001
Mucinous adenocarcinoma	4,763	0.214 (0.192, 0.24)	<0.001	0.21 (0.187, 0.236)	<0.001	1,677	0.218 (0.192, 0.246)	<0.001
Others	32,348	1.011 (0.984, 1.039)	0.418	0.922 (0.895, 0.95)	<0.001	15,783	0.938 (0.907, 0.971)	<0.001
Breast subtype						113,496		
$HR^+/HER2^-$	172,052	Reference		Reference		80,843	Reference	
$HR^-/HER2^-$	21,026	0.965 (0.934, 0.998)	0.037	0.585 (0.563, 0.608)	<0.001	12,220	0.622 (0.598, 0.646)	<0.001
$HR^+/HER2^+$	19,572	1.325 (1.283, 1.369)	<0.001	1.018 (0.982, 1.055)	0.329	10,985	1.035 (0.995, 1.077)	0.089
$HR^-/HER2^+$	7,494	1.345 (1.279, 1.414)	<0.001	1.017 (0.961, 1.076)	0.572	4,169	1.06 (0.996, 1.129)	0.066
Unknown	12,586	0.799 (0.765, 0.835)	<0.001	0.862 (0.821, 0.905)	<0.001	5,279	0.889 (0.841, 0.94)	<0.001
Grade						112,530		
Grade I	57,948	Reference		Reference		19,819	Reference	
Grade II	103,263	2.051 (1.997, 2.106)	<0.001	1.466 (1.425, 1.509)	<0.001	51,184	1.404 (1.359, 1.452)	<0.001
Grade III	63,778	2.86 (2.781, 2.941)	<0.001	1.648 (1.594, 1.704)	<0.001	38,585	1.444 (1.395, 1.495)	<0.001
Grade IV	589	2.196 (1.832, 2.631)	<0.001	1.478 (1.213, 1.801)	<0.001	312	1.329 (1.062, 1.662)	0.013
Unknown	7,152	1.224 (1.148, 1.306)	<0.001	1.191 (1.108, 1.28)	<0.001	2,630	1.117 (1.029, 1.212)	0.008

Table 1 *(continued)*

Table 1 *(continued)*

Characteristics	Total (n=232,730)	Single factor regression analysis		Multi-factor regression analysis		PSM		
		OR (95% CI)	P value	OR (95% CI)	P value	N	OR (95% CI)	P value
Derived AJCC T, 7th						118,970		
T1mic	4,961	Reference		Reference		1,654		
T1a	18,934	1.166 (1.007, 1.35)	0.04	1.233 (1.063, 1.43)	0.006	6,333	1.189 (1.018, 1.387)	0.028
T1b	46,376	2.196 (1.917, 2.514)	<0.001	2.36 (2.055, 2.71)	<0.001	16,942	2.233 (1.935, 2.578)	<0.001
T1c	88,707	5.994 (5.247, 6.846)	<0.001	6.057 (5.288, 6.938)	<0.001	44,050	5.187 (4.507, 5.969)	<0.001
T2	64,551	15.698 (13.743, 17.93)	<0.001	15.122 (13.201, 17.322)	<0.001	42,844	11.673 (10.141, 13.436)	<0.001
T3	8,816	35.795 (31.149, 41.133)	<0.001	34.704 (30.106, 40.004)	<0.001	6,993	25.002 (21.494, 29.082)	<0.001
Unknown	385	4.082 (3.027, 5.506)	<0.001	4.214 (3.117, 5.697)	<0.001	154	4.403 (3.104, 6.244)	<0.001

+, positive; –, negative. PSM, propensity score matching; BC, breast cancer; OR, odds ratio; CI, confidence interval; HR, hormone receptor; HER2, human epidermal growth factor receptor 2; AJCC T, the T staging of Cancer Staging Manual of the American Joint Committee on Cancer; mic, microinvasive carcinoma.

P<0.001) and multiple-factor analysis (OR =0.907; 95% CI: 0.876–0.939; P<0.001). Further evaluation using PSM would be considered. Triple-negative BC (TNBC) (OR =0.585; 95% CI: 0.563–0.608; P<0.001) had a lower risk of PLN compared to $HR^+/HER2^-$ patients. $HR^+/HER2^+$ (OR =1.018; 95% CI: 0.982–1.055; P=0.329) and $HR^-/HER2^+$ patients (OR =1.017; 95% CI: 0.961–1.076; P=0.572) did not show statistically significant differences in PLN risk compared to $HR^+/HER2^-$ patients. In terms of grade, grade II (OR =1.466; 95% CI: 1.425–1.509; P<0.001), grade III (OR =1.648; 95% CI: 1.594–1.704; P<0.001), and grade IV (OR =1.478; 95% CI: 1.213–1.801; P<0.001) were all identified as risk factors for PLN occurrence in BC patients. Furthermore, in the T stage (derived AJCC, 7th), the risk of PLN increased with the size of the tumor, with T3 having the highest risk (OR =34.704; 95% CI: 30.106–40.004; P<0.001). This suggests a correlation between tumor size and the likelihood of lymph node involvement, possibly indicating the progression of the disease.

After PSM, a certain number of patients were successfully matched in each independent factor. The baseline characteristics between the two groups, including diagnostic age (n=113,000), race (n=112,160), primary site (n=113,164), histology (n=112,268), breast subtype (n=113,496), grade (n=112,530), and T stage (derived AJCC, 7th) (n=118,970), achieved good balance [standardized mean difference (SMD) <0.2, see *Tables 1-3* and Tables S1-S7]. In the PSM dataset, the results of the single-factor logistic regression analysis were consistent with the previous results of the multivariable regression analysis. This indicates that the analysis results have a strong level of credibility.

The establishment of the BC lymph node metastasis prediction model

In the original cohort of 232,730 patients, we allocated 174,548 to the training set (75.0%) and 58,182 patients to the validation set (25.0%) (*Tables 2,3*). The flowchart of participant inclusion and exclusion is shown in *Figure 1*. We established a nomogram to visually display the score distribution and the predicted probabilities of risk factors (*Figure 3*). How do we utilize this model? For example, a 46-year-old (score: 7.5) African American (score: 8) female with BC, presenting with a tumor located in the lower-inner quadrant (score: 7.5), histologically diagnosed as invasive ductal carcinoma (score: 46), molecular subtype $HR^-/HER2^+$ (score: 16), unknown grade (score: 5), and T stage classified as T1c (score: 50). The total score is 7.5+8+7.5+46+16+5+50=140, corresponding to an estimated probability of lymph node metastasis of approximately 25%. In clinical practice, it is easy for us to estimate the probability of lymph node metastasis in patients using this nomogram. In order to better utilize

Table 2 Clinicopathological features of the training set and the validation set

Characteristics	Training dataset	Validation dataset	P value
Lymph nodes status, n (%)			
Negative	129,864 (74.4)	43,354 (74.5)	0.587981222
Positive	44,684 (25.6)	14,828 (25.5)	NA
Age (years), n (%)			
20–44	16,464 (9.4)	5,429 (9.3)	0.741465412
45–55	41,269 (23.6)	13,803 (23.7)	NA
56+	116,815 (66.9)	38,950 (66.9)	NA
Race/ethnicity, n (%)			
White	140,277 (80.4)	46,834 (80.5)	0.705380076
Black	16,688 (9.6)	5,516 (9.5)	NA
Asian or Pacific Islander	15,742 (9.0)	5,247 (9.0)	NA
Others	1,841 (1.1)	585 (1.0)	NA
Primary site, n (%)			
Upper-outer quadrant	60,201 (34.5)	20,130 (34.6)	0.69553622
Upper-inner quadrant	22,605 (13.0)	7,495 (12.9)	NA
Lower-outer quadrant	13,582 (7.8)	4,566 (7.8)	NA
Lower-inner quadrant	10,323 (5.9)	3,366 (5.8)	NA
Central portion & nipple	8,551 (4.9)	2,788 (4.8)	NA
Others	59,286 (34.0)	19,837 (34.1)	NA
Histology, n (%)			
Infiltrating duct carcinoma	129,679 (74.3)	43,329 (74.5)	0.324136699
Lobular carcinoma	16,970 (9.7)	5,641 (9.7)	NA
Mucinous adenocarcinoma	3,626 (2.1)	1,137 (2.0)	NA
Others	24,273 (13.9)	8,075 (13.9)	NA
Breast subtype, n (%)			
$HR^+/HER2^-$	129,081 (74.0)	42,971 (73.9)	0.874052188
$HR^-/HER2^-$	15,706 (9.0)	5,320 (9.1)	NA
$HR^+/HER2^+$	14,696 (8.4)	4,876 (8.4)	NA
$HR^-/HER2^+$	5,615 (3.2)	1,879 (3.2)	NA
Unknown	9,450 (5.4)	3,136 (5.4)	NA
Grade, n (%)			
Grade I	43,385 (24.9)	14,563 (25.0)	0.077366972
Grade II	77,711 (44.5)	25,552 (43.9)	NA
Grade III	47,670 (27.3)	16,108 (27.7)	NA
Grade IV	453 (0.3)	136 (0.2)	NA
Unknown	5,329 (3.1)	1,823 (3.1)	NA

Table 2 *(continued)*

Table 2 *(continued)*

Characteristics	Training dataset	Validation dataset	P value
Derived AJCC T, 7th, n (%)			
T1mic	3,734 (2.1)	1,227 (2.1)	0.93774038
T1a	14,211 (8.1)	4,723 (8.1)	NA
T1b	34,758 (19.9)	11,618 (20.0)	NA
T1c	66,591 (38.2)	22,116 (38.0)	NA
T2	48,351 (27.7)	16,200 (27.8)	NA
T3	6,606 (3.8)	2,210 (3.8)	NA
Unknown	297 (0.2)	88 (0.2)	NA

+, positive; –, negative. NA, not available; HR, hormone receptor; HER2, human epidermal growth factor receptor 2; AJCC T, the T staging of Cancer Staging Manual of the American Joint Committee on Cancer; mic, microinvasive carcinoma.

Table 3 Clinicopathological features with and without lymph node metastases in the training and testing set

Characteristics	Training dataset			Validation dataset		
	Regional nodes status		Total	Regional nodes status		Total
	Negative	Positive		Negative	Positive	
Total, n (%)	129,864 (74.40)	44,684 (25.60)	174,548 (100.00)	43,354 (74.51)	14,828 (25.49)	58,182 (100.00)
Age (years), n (%)						
20–44	10,699 (64.98)	5,765 (35.02)	16,464 (100.00)	3,502 (64.51)	1,927 (35.49)	5,429 (100.00)
45–55	29,276 (70.94)	11,993 (29.06)	41,269 (100.00)	9,862 (71.45)	3,941 (28.55)	13,803 (100.00)
56+	89,889 (76.95)	26,926 (23.05)	116,815 (100.00)	29,990 (77.00)	8,960 (23.00)	38,950 (100.00)
Race/ethnicity, n (%)						
White	104,919 (74.79)	35,358 (25.21)	140,277 (100.00)	35,132 (75.01)	11,702 (24.99)	46,834 (100.00)
Black	11,669 (69.92)	5,019 (30.08)	16,688 (100.00)	3,888 (70.49)	1,628 (29.51)	5,516 (100.00)
Asian or Pacific Islander	11,947 (75.89)	3,795 (24.11)	15,742 (100.00)	3,914 (74.60)	1,333 (25.40)	5,247 (100.00)
Others	1,329 (72.19)	512 (27.81)	1,841 (100.00)	420 (71.79)	165 (28.21)	585 (100.00)
Primary site, n (%)						
Upper-outer quadrant	44,305 (73.60)	15,896 (26.40)	60,201 (100.00)	14,818 (73.61)	5,312 (26.39)	20,130 (100.00)
Upper-inner quadrant	18,772 (83.04)	3,833 (16.96)	22,605 (100.00)	6,220 (82.99)	1,275 (17.01)	7,495 (100.00)
Lower-outer quadrant	9,987 (73.53)	3,595 (26.47)	13,582 (100.00)	3,347 (73.30)	1,219 (26.70)	4,566 (100.00)
Lower-inner quadrant	8,277 (80.18)	2,046 (19.82)	10,323 (100.00)	2,692 (79.98)	674 (20.02)	3,366 (100.00)
Central portion & nipple	5,592 (65.40)	2,959 (34.60)	8,551 (100.00)	1,838 (65.93)	950 (34.07)	2,788 (100.00)
Others	42,931 (72.41)	16,355 (27.59)	59,286 (100.00)	14,439 (72.79)	5,398 (27.21)	19,837 (100.00)

Table 3 *(continued)*

Table 3 *(continued)*

Characteristics	Training dataset			Validation dataset		
	Regional nodes status		Total	Regional nodes status		Total
	Negative	Positive		Negative	Positive	
Histology, n (%)						
Infiltrating duct carcinoma	96,524 (74.43)	33,155 (25.57)	129,679 (100.00)	32,340 (74.64)	10,989 (25.36)	43,329 (100.00)
Lobular carcinoma	11,987 (70.64)	4,983 (29.36)	16,970 (100.00)	3,905 (69.23)	1,736 (30.77)	5,641 (100.00)
Mucinous adenocarcinoma	3,382 (93.27)	244 (6.73)	3,626 (100.00)	1,055 (92.79)	82 (7.21)	1,137 (100.00)
Others	17,971 (74.04)	6,302 (25.96)	24,273 (100.00)	6,054 (74.97)	2,021 (25.03)	8,075 (100.00)
Breast subtype, n (%)						
$HR^+/HER2^-$	96,554 (74.80)	32,527 (25.20)	129,081 (100.00)	32,179 (74.89)	10,792 (25.11)	42,971 (100.00)
$HR^-/HER2^-$	11,862 (75.53)	3,844 (24.47)	15,706 (100.00)	4,009 (75.36)	1,311 (24.64)	5,320 (100.00)
$HR^+/HER2^+$	10,144 (69.03)	4,552 (30.97)	14,696 (100.00)	3,392 (69.57)	1,484 (30.43)	4,876 (100.00)
$HR^-/HER2^+$	3,848 (68.53)	1,767 (31.47)	5,615 (100.00)	1,311 (69.77)	568 (30.23)	1,879 (100.00)
Unknown	7,456 (78.90)	1,994 (21.10)	9,450 (100.00)	2,463 (78.54)	673 (21.46)	3,136 (100.00)
Grade, n (%)						
Grade I	36,795 (84.81)	6,590 (15.19)	43,385 (100.00)	12,366 (84.91)	2,197 (15.09)	14,563 (100.00)
Grade II	56,854 (73.16)	20,857 (26.84)	77,711 (100.00)	18,707 (73.21)	6,845 (26.79)	25,552 (100.00)
Grade III	31,518 (66.12)	16,152 (33.88)	47,670 (100.00)	10,687 (66.35)	5,421 (33.65)	16,108 (100.00)
Grade IV	329 (72.63)	124 (27.37)	453 (100.00)	94 (69.12)	42 (30.88)	136 (100.00)
Unknown	4,368 (81.97)	961 (18.03)	5,329 (100.00)	1,500 (82.28)	323 (17.72)	1,823 (100.00)
Derived AJCC T, 7^{th}, n (%)						
T1mic	3,561 (95.37)	173 (4.63)	3,734 (100.00)	1,169 (95.27)	58 (4.73)	1,227 (100.00)
T1a	13,466 (94.76)	745 (5.24)	14,211 (100.00)	4,448 (94.18)	275 (5.82)	4,723 (100.00)
T1b	31,385 (90.30)	3,373 (9.70)	34,758 (100.00)	10,500 (90.38)	1,118 (9.62)	11,618 (100.00)
T1c	51,423 (77.22)	15,168 (22.78)	66,591 (100.00)	17,198 (77.76)	4,918 (22.24)	22,116 (100.00)
T2	27,395 (56.66)	20,956 (43.34)	48,351 (100.00)	9,144 (56.44)	7,056 (43.56)	16,200 (100.00)
T3	2,389 (36.16)	4,217 (63.84)	6,606 (100.00)	819 (37.06)	1,391 (62.94)	2,210 (100.00)
Unknown	245 (82.49)	52 (17.51)	297 (100.00)	76 (86.36)	12 (13.64)	88 (100.00)

+, positive; –, negative. HR, hormone receptor; HER2, human epidermal growth factor receptor 2; AJCC T, the T staging of Cancer Staging Manual of the American Joint Committee on Cancer; mic, microinvasive carcinoma.

this nomogram, we also employed the following objective evaluation methods to analyze the sensitivity, specificity, and accuracy of the model. The calibration curves with similar area under the curve (AUC) values demonstrate the good predictability of our nomogram model (*Figure 4A*,*4B*). DCA suggests that the threshold probability of 0–0.6 is the most favorable predictive factor for lymph node metastasis (*Figure 4C*,*4D*). The calibration curve with a concordance index (C-index) of 0.749 (95% CI: 0.747–0.752) indicates a strong consistency between the observed values and the predicted probabilities (*Figure 4E*,*4F*).

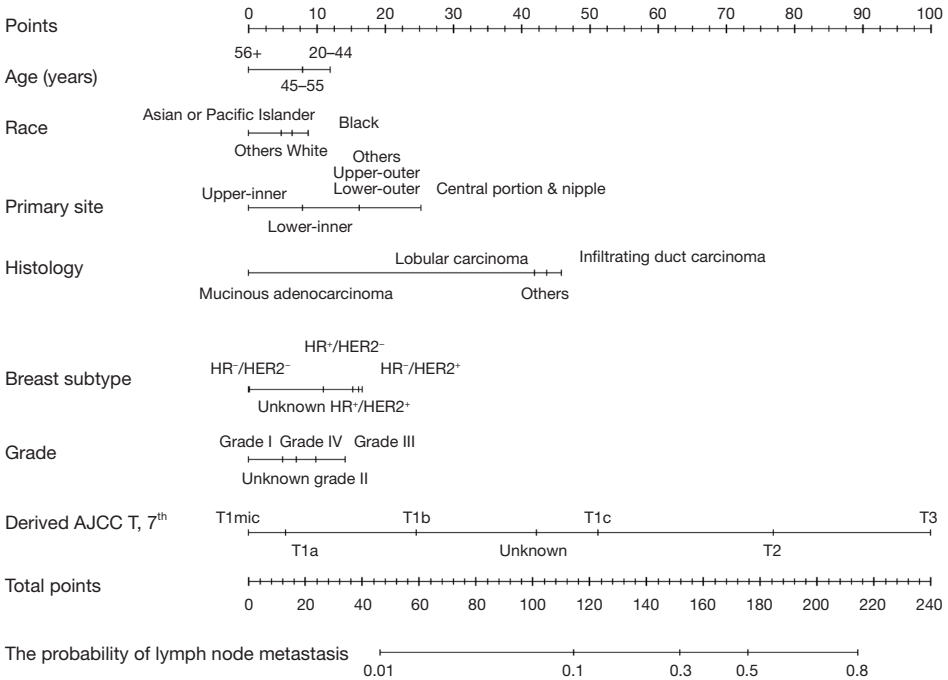

Figure 3 Seven independent factors, including age, race, primary site, histology, breast subtype, grade, derived AJCC T, 7th were included in the nomogram. HR, hormone receptor; HER2, human epidermal growth factor receptor 2; AJCC T, the T staging of Cancer Staging Manual of the American Joint Committee on Cancer; mic, microinvasive carcinoma.

Discussion

The topic of lymph node metastasis in BC has been extensively researched in the past. Previous studies by Van Zee *et al.* constructed a nomogram for predicting the likelihood of additional nodal metastases in BC patients with a positive sentinel node biopsy, but it had limitations as it only focused on preoperative biopsy-positive patients (10). Bevilacqua *et al.* included 3,786 cases of invasive BC between 1996 and 2002, who did not receive NAT, to develop an effective nomogram for predicting the probability of positive sentinel lymph nodes. However, the included clinical features in the nomogram were not sufficiently detailed, for example, the location was categorized only as upper-inner quadrant and other. Additionally, the predictive factors of the nomogram did not include molecular subtypes, which resulted in a lack of precision (2). Li *et al.* constructed a nomogram for lymph node metastasis in T1–2 and non-metastatic (M0) BC patients using the SEER database, but their model did not exclude patients with multiple tumors or those who did not undergo lymph node dissection (11). Gao *et al.* (8) developed nomogram models for stratified prediction of axillary lymph node metastasis in cN0 BC patients using SEER data from 2010 to 2015. However, it's important to note an anomaly in the SEER database during the years 2010 to 2015—it lacks the capability to selectively identify cN0 BC patients. The only columns where cN0 stage patients can be reliably identified are "Derived SEER Combined N [2016–2017]" and "Derived SEER Combined N Src [2016–2017]". In contrast, other columns can solely screen BC patients who have been clinically or pathologically determined as N0 stage. Consequently, the screening method provided in this study lacks the necessary specificity, which may impact the accuracy of the results. In our study, we specifically excluded patients who received NAT or radiation therapy before surgery and excluded patients with M1, T0, Tis, and T4 stages to accurately study the impact of independent factors on lymph node metastasis in female BC. Since traditional level I and II axillary lymph node dissection (ALND) requires at least 10 lymph nodes for pathological evaluation (12), we performed statistical analysis on the number of lymph node biopsies greater than 5, greater than 10, and greater than 12 before establishing the model (*Figure 5*). We found that the lymph node positivity rate in the downloaded

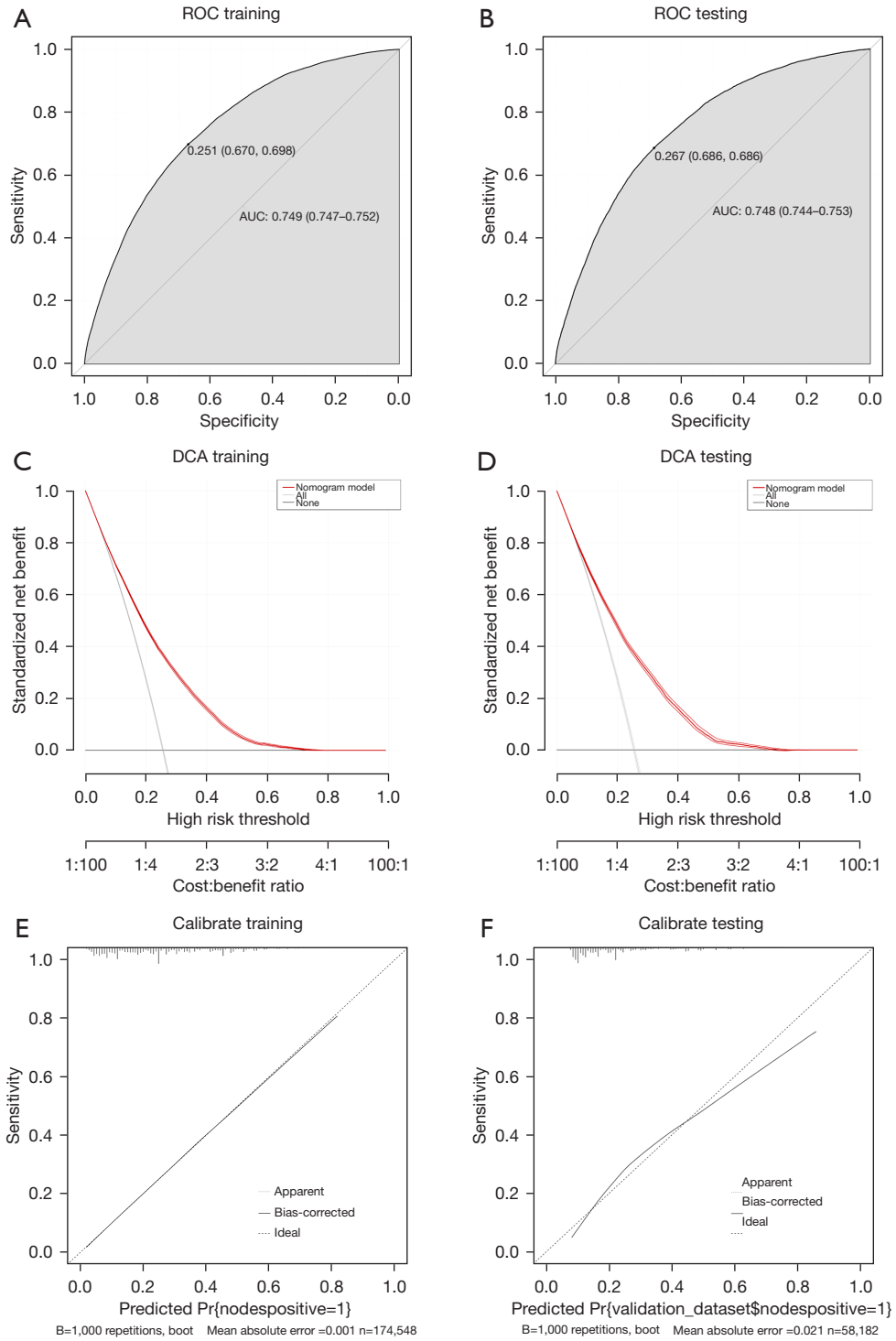

Figure 4 Nomogram of lymph node metastasis risk prediction in BC patients who did not receive any treatment before surgery. (A) The ROC curve of training set. (B) The ROC curve of validation data. (C) The DCA curve of training set. (D) The DCA curve of validation data. (E) The calibration curve of training cohort. (F) The calibration curve of validation data. ROC, receiver operating characteristic; AUC, area under the curve; DCA, decision curve analysis; BC, breast cancer.

Characteristics/Variables	Regional nodes status Negative n (%)	Positive n (%)	Total n (%)
Regional nodes examined>5	23,421 (36.98)	39,913 (63.02)	63,334 (100.00)
Derived AJCC T, 7th			
T1mic	483 (79.44)	125 (20.56)	608 (100.00)
T1a	1,887 (76.55)	578 (23.45)	2,465 (100.00)
T1b	4,628 (66.85)	2,295 (33.15)	6,923 (100.00)
T1c	8,934 (42.74)	11,971 (57.26)	20,905 (100.00)
T2	6,601 (24.54)	20,295 (75.46)	26,896 (100.00)
T3	846 (15.52)	4,605 (84.48)	5,451 (100.00)
Unknown	42 (48.84)	44 (51.16)	86 (100.00)
Regional nodes examined>10	7,348 (20.41)	28,648 (79.59)	35,996 (100.00)
Derived AJCC T, 7th			
T1mic	123 (58.29)	88 (41.71)	211 (100.00)
T1a	508 (56.32)	394 (43.68)	902 (100.00)
T1b	1,175 (43.01)	1,557 (56.99)	2,732 (100.00)
T1c	2,614 (24.24)	8,171 (75.76)	10,785 (100.00)
T2	2,510 (14.46)	14,846 (85.54)	17,356 (100.00)
T3	402 (10.14)	3,562 (89.86)	3,964 (100.00)
Unknown	16 (34.78)	30 (65.22)	46 (100.00)
Regional nodes examined>12	5,053 (17.70)	23,492 (82.30)	28,545 (100.00)
Derived AJCC T, 7th			
T1mic	75 (50.00)	75 (50.00)	150 (100.00)
T1a	352 (53.90)	301 (46.10)	653 (100.00)
T1b	792 (38.80)	1,248 (61.20)	2,040 (100.00)
T1c	1,756 (21.00)	6,623 (79.00)	8,379 (100.00)
T2	1,749 (12.50)	12,209 (87.50)	13,958 (100.00)
T3	317 (9.50)	3,013 (90.50)	3,330 (100.00)
Unknown	12 (34.30)	23 (65.70)	35 (100.00)

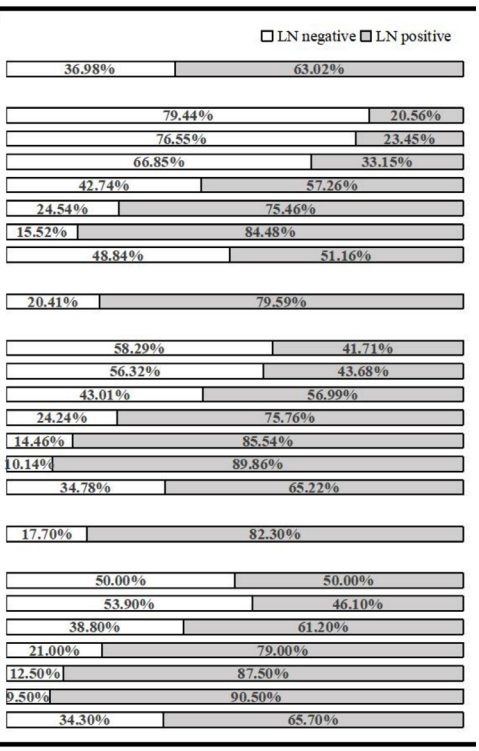

Figure 5 Statistical description of LN status at regional nodes examined >5, >10, and >12. LN, lymph node; AJCC T, the T staging of Cancer Staging Manual of the American Joint Committee on Cancer; mic, microinvasive carcinoma.

data was higher than our expectations, so we reviewed the literature on lymph node positivity rates in different T stages in the past. In previous literature surveys, Iwasaki *et al.* conducted a retrospective study on 823 T1N0M0 invasive BC patients, with a lymph node metastasis rate of 25% (208/823) for T1N0M0 invasive BC (13). Tan *et al.* conducted a retrospective study on 380 early-stage BC patients, with lymph node positivity rates of 4.3% for T1a, 18.9% for T1b, and 27.6% for T1c (14).

In our analysis, we found that the lymph node positivity rates in T1mic BC patients who underwent lymph node dissection were significantly higher than those reported in previous literature. The positivity rate was 20.56% for more than five lymph nodes, 41.71% for more than 10 lymph nodes, and as high as 50.00% for more than 12 lymph nodes. This discrepancy suggests that there may be some bias in the data entry process. One possible explanation is that many patients who undergo lymph node dissection are suspected of having lymph node positivity, prompting physicians to consciously perform the procedure. Based on previous clinical experience data, we excluded patients who did not undergo lymph node dissection and established a more reasonable predictive model.

Our study is the first to utilize the SEER database to construct a nomogram for lymph node metastasis in BC patients who have not received any treatment. The nomogram was well validated. With the risk prediction model we established, we can promptly and scientifically assess and provide personalized treatment for high-risk patients, which has significant implications for their treatment outcomes.

In our present study, we initially conducted comparisons among various subgroups based on clinical and pathological characteristics through univariate and multivariate regression analyses. Following PSM, we noted that there were no substantial alterations in the distinctions between these subgroups when compared to the reference group. This implies that there is no significant interaction among the various clinical-pathological characteristics, and some hints regarding the relationships between individual subgroups can be gleaned from the results of univariate, multivariate, and PSM analyses. Therefore, it is not difficult to conclude that the risk factors for regional lymph node metastasis include young age (<45 years old), black

ethnicity, involvement of the central portion and nipple, infiltrating duct carcinoma, HER2 positivity, grade III, and large tumor size. On the other hand, Asian or Pacific Islander ethnicity, upper-inner and lower-inner quadrants, mucinous adenocarcinoma, triple-negative subtype, low grade, and small tumor size is considered protective factors against lymph node metastasis.

BC accounts for a relatively small proportion in young women (*Figure 2*), approximately 9.4%. However, the positive lymph node ratio (PLNR) is highest among young women and tends to decrease with increasing age. The definition of young women varies among articles (15-17), therefore, this study used the range of perimenopausal women (45–55 years) as the dividing line (18), classifying women into young, perimenopausal, and elderly groups to investigate the impact of age on lymph node metastasis. It has been reported that young women are associated with poor prognosis in BC (19). Some retrospective studies also suggest a higher risk of recurrence after breast-conserving surgery in young women compared to older women (16,20). Patients younger than 40 years old more commonly exhibit lymphocytic stromal reaction, histologic grade 3, and extensive ductal carcinoma in situ (DCIS) (21), which is consistent with the findings of this data analysis, indicating that young women with BC have more aggressive biological behavior compared to older women.

Among the racial categories, Black women are more likely to experience lymph node metastasis. Even after adjusting for independent factors such as age, tumor location, histology, BC subtype, grade, and T stage (*Table 1*), Black women still show a statistically significant difference in lymph node positivity compared to White women, with an OR of 1.09 times higher for Black women (95% CI: 1.049–1.133, P<0.001). This conclusion has been widely validated in previous studies (7,22,23).

Regarding the primary tumor location, we found that the highest risk of lymph node metastasis is in the central portion and nipple, followed by the outer-upper and outer-lower quadrants (*Table 1*). There were no statistically significant differences in lymph node metastasis risk between the outer-upper and outer-lower quadrants in both univariate and multivariate logistic regression analyses and PSM, which aligns with previous research (24).

In terms of molecular subtypes, our analysis through multivariate regression analysis, PSM, and column line chart analysis all showed no significant difference in lymph node metastasis risk among $HR^+/HER2^-$, $HR^+/HER2^+$, and $HR^-/HER2^+$ subtypes (P=0.089; P=0.066). Interestingly, the lymph node metastasis rate of TNBC is lower than that of hormone receptor-positive, $HER2^-$ patients. Although this observation may seem counterintuitive. It has also been reflected in previous studies. In a retrospective study by Si *et al.*, luminal $HER2^+$ was associated with the highest lymph node positivity (49.0%), followed by luminal $HER2^-$ (46.8%), $HER2^+$ (44.4%), luminal A (36.5%), and TNBC (34.7%). The occurrence of LN metastases was lowest in the TNBC subtype (25).

It is possible that the $HR^+/HER2^-$ subtype of BC tumors may have a higher proportion of luminal B subtype. In a retrospective study by Xiong *et al.*, the LNM rates were 72.0% for luminal B $HER2^+$, 51.9% for luminal B, 50.0% for TNBC, and 37.4% for luminal A, with statistically significant differences observed (26). Similarly, in a retrospective study by Cheng *et al.* investigating the relationship between molecular subtypes and clinicopathological features of BC in Chinese women, the basal-like subtype had a significantly lower risk of lymph node metastasis compared to luminal A, luminal B, and $HER2^+$ subtypes (27). Min *et al.* also conducted a retrospective analysis of clinical and pathological features of 16,552 patients who underwent breast surgery at Samsung Medical Center from 2000 to 2015. The results showed a higher incidence of lymph node metastasis in the luminal subtype compared to HER2 and TNBC, with the lowest lymph node metastasis rate observed in the TNBC subtype (28). These clinical studies have reported a lower lymph node metastasis risk in TNBC.

In fact, it is well recognized that TNBC is prone to recurrence and distant metastasis, resulting in a relatively poorer prognosis compared with other BC subtypes (29). It is widely accepted in the field that lymph node metastasis is a precursor event to distant metastasis. Based on this premise, most researchers have traditionally believed that TNBC has a higher likelihood of lymph node involvement. The conclusion drawn from our statistical analysis seems contradictory to previous experience. However, we believe that although TNBC has the lowest risk of lymph node metastasis, this does not exclude its "aggressiveness". Under the same treatment, TNBC has a worse prognosis (29), and it is often more likely to develop visceral metastases, including lung, liver, and brain metastases (30). Indeed, there is a substantial body of research that supports the notion that BC can progress to distant metastasis without necessarily involving axillary lymph node metastasis as an intermediate step (31). This study suggests that the rate of axillary lymph node metastasis in TNBC is relatively lower. However, it is important to note that this result should

not be construed as evidence for a better prognosis among TNBC patients.

In the analysis of T-stage and lymph node positivity, it is easy to observe that the risk of lymph node metastasis increases with tumor size, as also demonstrated in the study by Wu *et al.* (32). However, there are clinical examples of small primary tumors with extensive lymph node metastasis (33) or distant metastasis (34), indicating that a small primary tumor does not always indicate a better prognosis. Although lymph node metastasis is not as immediately life-threatening as lung, liver, or brain metastasis, it still requires our attention. In our analysis, we observed that the incidence of lymph node metastasis in T1c stage BC patients was 22.64%, while it increased to 43.4% in T2 stage patients. Our nomogram further indicates that T1c stage BC receives a higher score, approximately 50 points, compared to the highest scores observed in other subgroups. This suggests that patients in T1c and T2 stages, particularly those facing delayed surgery, may consider NAT as a potential strategy to reduce the risk of lymph node metastasis. However, it's important to note that the decision for adjuvant therapy should be based on additional clinical research findings to ensure its appropriateness.

There are several limitations of our study. The database does not include additional information such as imaging features, Ki-67, vascular embolism, vascular invasion, nerve invasion, etc. (35,36). Including these biomarkers would enhance the accuracy of the developed model. Additionally, The SEER database lacks detailed information on NAT regimens, and this study did not provide a detailed comparison. Further clinical research is needed to validate the efficacy of different NAT regimens for different tumor types.

Conclusions

This retrospective study revealed the lymph node positivity status of female BC patients with distinct clinicopathological characteristics who did not receive any treatment preoperatively. Valuable nomograms were constructed based on these patient populations. The study results demonstrate strong reliability through various calibration and discrimination statistical methods, making them potential tools for guiding clinical diagnosis and individualized treatment.

Risk factors for regional lymph node metastasis include young age (<35 years), Black race, central portion & nipple involvement, infiltrating duct carcinoma, HER2 positivity, grade III, and large tumor size.

Although TNBC has the lowest risk of lymph node metastasis, this does not exclude the "aggressiveness" of TNBC, as it still has a poorer prognosis under the same treatment.

Patients in T1c and T2 stages, particularly those with delayed surgery, may consider NAT to mitigate the potential threat of lymph node metastasis.

Acknowledgments

Funding: The work was supported by the National Natural Science Foundation of China (Nos. 82272855, 81972453, and 81972597), the Natural Science Foundation of Zhejiang Province (Nos. LR22H160011, LY19H160055, LY19H160059, LY18H160005, and LY20H160026), and the Medical and Health Science and Technology of Zhejiang Province (Youth Talent Program) Project (No. 2021RC016). The work was sponsored by the Zheng Shu Medical Elite Scholarship Fund.

Footnote

Reporting Checklist: The authors have completed the TRIPOD reporting checklist. Available at https://gs.amegroups.com/article/view/10.21037/gs-23-258/rc

Peer Review File: Available at https://gs.amegroups.com/article/view/10.21037/gs-23-258/prf

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://gs.amegroups.com/article/view/10.21037/gs-23-258/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. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

References

1. Siegel RL, Miller KD, Fuchs HE, et al. Cancer statistics, 2022. CA Cancer J Clin 2022;72:7-33.
2. 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;25:3670-9.
3. Nottegar A, Veronese N, Senthil M, et al. Extra-nodal extension of sentinel lymph node metastasis is a marker of poor prognosis in breast cancer patients: A systematic review and an exploratory meta-analysis. Eur J Surg Oncol 2016;42:919-25.
4. Asioli S, Eusebi V, Gaetano L, et al. The pre-lymphatic pathway, the rooths of the lymphatic system in breast tissue: a 3D study. Virchows Arch 2008;453:401-6.
5. Voogd AC, Coebergh JW, Repelaer van Driel OJ, et al. The risk of nodal metastases in breast cancer patients with clinically negative lymph nodes: a population-based analysis. Breast Cancer Res Treat 2000;62:63-9.
6. Chang JM, Leung JWT, Moy L, et al. Axillary Nodal Evaluation in Breast Cancer: State of the Art. Radiology 2020;295:500-15.
7. Zhao YX, Liu YR, Xie S, et al. A Nomogram Predicting Lymph Node Metastasis in T1 Breast Cancer based on the Surveillance, Epidemiology, and End Results Program. J Cancer 2019;10:2443-9.
8. 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.
9. Park AR, Chae EY, Cha JH, et al. Preoperative Breast MRI in Women 35 Years of Age and Younger with Breast Cancer: Benefits in Surgical Outcomes by Using Propensity Score Analysis. Radiology 2021;300:39-45.
10. Van Zee KJ, Manasseh DM, Bevilacqua JL, et al. A nomogram for predicting the likelihood of additional nodal metastases in breast cancer patients with a positive sentinel node biopsy. Ann Surg Oncol 2003;10:1140-51.
11. Li H, Tang L, Chen Y, et al. Development and validation of a nomogram for prediction of lymph node metastasis in early-stage breast cancer. Gland Surg 2021;10:901-13.
12. Axelsson CK, Mouridsen HT, Zedeler K. Axillary dissection of level I and II lymph nodes is important in breast cancer classification. The Danish Breast Cancer Cooperative Group (DBCG). Eur J Cancer 1992;28A:1415-8.
13. Iwasaki Y, Fukutomi T, Akashi-Tanaka S, et al. Axillary node metastasis from T1N0M0 breast cancer: possible avoidance of dissection in a subgroup. Jpn J Clin Oncol 1998;28:601-3.
14. 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;46:693-7.
15. Rossi L, Mazzara C, Pagani O. Diagnosis and Treatment of Breast Cancer in Young Women. Curr Treat Options Oncol 2019;20:86.
16. Collins LC, Marotti JD, Gelber S, et al. Pathologic features and molecular phenotype by patient age in a large cohort of young women with breast cancer. Breast Cancer Res Treat 2012;131:1061-6.
17. Gnerlich JL, Deshpande AD, Jeffe DB, et al. Elevated breast cancer mortality in women younger than age 40 years compared with older women is attributed to poorer survival in early-stage disease. J Am Coll Surg 2009;208:341-7.
18. Takahashi TA, Johnson KM. Menopause. Med Clin North Am 2015;99:521-34.
19. Kroman N, Jensen MB, Wohlfahrt J, et al. Factors influencing the effect of age on prognosis in breast cancer: population based study. BMJ 2000;320:474-8.
20. Arvold ND, Taghian AG, Niemierko A, et al. Age, breast cancer subtype approximation, and local recurrence after breast-conserving therapy. J Clin Oncol 2011;29:3885-91.
21. Kurtz JM, Jacquemier J, Amalric R, et al. Why are local recurrences after breast-conserving therapy more frequent in younger patients? J Clin Oncol 1990;8:591-8.
22. Iqbal J, Ginsburg O, Rochon PA, et al. Differences in breast cancer stage at diagnosis and cancer-specific survival by race and ethnicity in the United States. JAMA 2015;313:165-73.
23. McBride R, Hershman D, Tsai WY, et al. Within-stage racial differences in tumor size and number of positive lymph nodes in women with breast cancer. Cancer 2007;110:1201-8.
24. Rinaldi RM, Sapra A, Bellin LS. Breast Lymphatics. In: StatPearls. Treasure Island: StatPearls Publishing; 2023.
25. Si C, Jin Y, Wang H, et al. Association between molecular subtypes and lymph node status in invasive breast cancer. Int J Clin Exp Pathol 2014;7:6800-6.
26. 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.
27. Cheng HT, Huang T, Wang W, et al. Clinicopathological features of breast cancer with different molecular subtypes in Chinese women. J Huazhong Univ Sci Technolog Med Sci 2013;33:117-21.
28. Min SK, Lee SK, Woo J, et al. Relation Between Tumor Size and Lymph Node Metastasis According to Subtypes of Breast Cancer. J Breast Cancer 2021;24:75-84.
29. Bianchini G, De Angelis C, Licata L, Gianni L. Treatment landscape of triple-negative breast cancer - expanded options, evolving needs. Nat Rev Clin Oncol 2022;19:91-113.
30. Rakha EA, Chan S. Metastatic triple-negative breast cancer. Clin Oncol 2011;23:587-600.
31. Venet D, Fimereli D, Rothé F, et al. Phylogenetic reconstruction of breast cancer reveals two routes of metastatic dissemination associated with distinct clinical outcome. EBioMedicine 2020;56:102793.
32. Wu SL, Gai JD, Yu XM, et al. A novel nomogram and risk classification system for predicting lymph node metastasis of breast mucinous carcinoma: A SEER-based study. Cancer Med 2022;11:4767-83.
33. Wo JY, Chen K, 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.
34. Zheng YZ, Wang XM, Fan L, et al. Breast Cancer-Specific Mortality in Small-Sized Tumor with Stage IV Breast Cancer: A Population-Based Study. Oncologist 2021;26:e241-50.
35. Thangarajah F, Malter W, Hamacher S, et al. Predictors of sentinel lymph node metastases in breast cancer-radioactivity and Ki-67. Breast 2016;30:87-91.
36. Chen H, Meng X, Hao X, et al. Correlation Analysis of Pathological Features and Axillary Lymph Node Metastasis in Patients with Invasive Breast Cancer. J Immunol Res 2022;2022:7150304.

Cite this article as: Luo M, Lin X, Hao D, Shen KW, Wu W, Wang L, Ruan S, Zhou J. Incidence and risk factors of lymph node metastasis in breast cancer patients without preoperative chemoradiotherapy and neoadjuvant therapy: analysis of SEER data. Gland Surg 2023;12(11):1508-1524. doi: 10.21037/gs-23-258

Supplementary

Table S1 SMD of age

Characteristics	Level		SMD
	Negative	Positive	
N	56,500	56,500	
Race, n (%)			0.01
White	44,527 (78.81)	44,734 (79.18)	
Black	6,221 (11.01)	6,068 (10.74)	
Asian or Pacific Islander	5,107 (9.04)	5,041 (8.92)	
Others	645 (1.14)	657 (1.16)	
Primary site, n (%)			0.023
Upper-outer quadrant	20,365 (36.04)	20,249 (35.84)	
Upper-inner quadrant	5,101 (9.03)	5,087 (9.00)	
Lower-outer quadrant	4,358 (7.71)	4,661 (8.25)	
Lower-inner quadrant	2,791 (4.94)	2,693 (4.77)	
Central portion & nipple	3,293 (5.83)	3,403 (6.02)	
Others	20,592 (36.45)	20,407 (36.12)	
Histology, n (%)			0.009
Infiltrating duct carcinoma	42,409 (75.06)	42,215 (74.72)	
Lobular carcinoma	5,965 (10.56)	6,093 (10.78)	
Mucinous adenocarcinoma	331 (0.59)	326 (0.58)	
Others	7,795 (13.80)	7,866 (13.92)	
Breast subtype, n (%)			
$HR^+/HER2^-$	40,689 (72.02)	41,128 (72.79)	0.019
$HR^-/HER2^-$	5,126 (9.07)	5,041 (8.92)	
$HR^+/HER2^+$	5,736 (10.15)	5,630 (9.96)	
$HR^-/HER2^+$	2,255 (3.99)	2,141 (3.79)	
Others	2,694 (4.77)	2,560 (4.53)	
Grade, n (%)			0.021
Grade I	8,505 (15.05)	8,722 (15.44)	
Grade II	26,189 (46.35)	26,383 (46.70)	
Grade III	20,277 (35.89)	20,006 (35.41)	
Grade IV	169 (0.30)	153 (0.27)	
Unknown	1,360 (2.41)	1,236 (2.19)	
AJCC T, n (%)			0.009
T1mic	230 (0.41)	231 (0.41)	
T1a	1,027 (1.82)	1,020 (1.81)	
T1b	4,488 (7.94)	4,491 (7.95)	
T1c	20,015 (35.42)	20,086 (35.55)	
T2	27,558 (48.78)	27,453 (48.59)	
T3	3,103 (5.49)	3,155 (5.58)	
Unknown	79 (0.14)	64 (0.11)	

+, positive; –, negative. SMD, standardized mean difference; HR, hormone receptor; HER2, human epidermal growth factor receptor 2; AJCC T, the T staging of Cancer Staging Manual of the American Joint Committee on Cancer; mic, microinvasive carcinoma.

https://dx.doi.org/10.21037/gs-23-258

Table S2 SMD of race

Characteristics	Level		SMD
	Negative	Positive	
N	56,080	56,080	
Age (years), n (%)			0.007
20–44	6,736 (12.01)	6,864 (12.24)	
45–55	14,680 (26.18)	14,659 (26.14)	
56+	34,664 (61.81)	34,557 (61.62)	
Primary site, n (%)			0.012
Upper-outer quadrant	20,275 (36.15)	20,126 (35.89)	
Upper-inner quadrant	5,055 (9.01)	5,075 (9.05)	
Lower-outer quadrant	4,693 (8.37)	4,616 (8.23)	
Lower-inner quadrant	2,743 (4.89)	2,685 (4.79)	
Central portion & nipple	3,362 (6.00)	3,333 (5.94)	
Others	19,952 (35.58)	20,245 (36.10)	
Histology, n (%)			0.008
Infiltrating duct carcinoma	42,044 (74.97)	41,884 (74.69)	
Lobular carcinoma	5,954 (10.62)	6,062 (10.81)	
Mucinous adenocarcinoma	314 (0.56)	326 (0.58)	
Others	7,768 (13.85)	7,808 (13.92)	
Breast subtype, n (%)			0.017
$HR^+/HER2^-$	40,529 (72.27)	40,831 (72.81)	
$HR^-/HER2^-$	5,236 (9.34)	5,020 (8.95)	
$HR^+/HER2^+$	5,530 (9.86)	5,555 (9.91)	
$HR^-/HER2^+$	2,235 (3.99)	2,125 (3.79)	
Others	2,550 (4.55)	2,549 (4.55)	
Grade, n (%)			0.013
Grade I	8,757 (15.62)	8,683 (15.48)	
Grade II	25,818 (46.04)	26,182 (46.69)	
Grade III	20,082 (35.81)	19,824 (35.35)	
Grade IV	152 (0.27)	155 (0.28)	
Unknown	1,271 (2.27)	1,236 (2.20)	
AJCC T, 7^{th}, n (%)			0.01
T1mic	256 (0.46)	231 (0.41)	
T1a	992 (1.77)	1,020 (1.82)	
T1b	4,398 (7.84)	4,491 (8.01)	
T1c	20,161 (35.95)	20,086 (35.82)	
T2	27,121 (48.36)	27,118 (48.36)	
T3	3,087 (5.50)	3,070 (5.47)	
Unknown	65 (0.12)	64 (0.11)	

+, positive; –, negative. SMD, standardized mean difference; HR, hormone receptor; HER2, human epidermal growth factor receptor 2; AJCC T, the T staging of Cancer Staging Manual of the American Joint Committee on Cancer; mic, microinvasive carcinoma.

https://dx.doi.org/10.21037/gs-23-258

Table S3 SMD of primary site

Characteristics	Level		SMD
	Negative	Positive	
N	56,582	56,582	
Age (years), n (%)			0.011
20–44	6,818 (12.05)	7,009 (12.39)	
45–55	15,023 (26.55)	14,852 (26.25)	
56+	34,741 (61.40)	34,721 (61.36)	
Race, n (%)			0.013
White	44,641 (78.90)	44,789 (79.16)	
Black	6,331 (11.19)	6,113 (10.80)	
Asian or Pacific Islander	4,942 (8.73)	5,028 (8.89)	
Others	668 (1.18)	652 (1.15)	
Histology, n (%)			0.011
Infiltrating duct carcinoma	42,387 (74.91)	42,274 (74.71)	
Lobular carcinoma	5,932 (10.48)	6,115 (10.81)	
Mucinous adenocarcinoma	325 (0.57)	326 (0.58)	
Others	7,938 (14.03)	7,867 (13.90)	
Breast subtype, n (%)			0.011
$HR^+/HER2^-$	41,046 (72.54)	41,167 (72.76)	
$HR^-/HER2^-$	5,105 (9.02)	5,068 (8.96)	
$HR^+/HER2^+$	5,610 (9.91)	5,634 (9.96)	
$HR^-/HER2^+$	2,263 (4.00)	2,148 (3.80)	
Others	2,558 (4.52)	2,565 (4.53)	
Grade, n (%)			0.013
Grade I	8,747 (15.46)	8,723 (15.42)	
Grade II	26,046 (46.03)	26,363 (46.59)	
Grade III	20,339 (35.95)	20,099 (35.52)	
Grade IV	155 (0.27)	156 (0.28)	
Unknown	1,295 (2.29)	1,241 (2.19)	
AJCC T, 7^{th}, n (%)			0.01
T1mic	211 (0.37)	231 (0.41)	
T1a	1,038 (1.83)	1,020 (1.80)	
T1b	4,433 (7.83)	4,491 (7.94)	
T1c	20,165 (35.64)	20,086 (35.50)	
T2	27,554 (48.70)	27,486 (48.58)	
T3	3,121 (5.52)	3,204 (5.66)	
Unknown	60 (0.11)	64 (0.11)	

+, positive; –, negative. SMD, standardized mean difference; HR, hormone receptor; HER2, human epidermal growth factor receptor 2; AJCC T, the T staging of Cancer Staging Manual of the American Joint Committee on Cancer; mic, microinvasive carcinoma.

https://dx.doi.org/10.21037/gs-23-258

Table S4 SMD of histology

Characteristics	Level		SMD
	Negative	Positive	
N	56,134	56,134	
Age (years), n (%)			0.018
20–44	6,574 (11.71)	6,871 (12.24)	
45–55	14,936 (26.61)	14,669 (26.13)	
56+	34,624 (61.68)	34,594 (61.63)	
Race, n (%)			0.016
White	44,215 (78.77)	44,509 (79.29)	
Black	6,220 (11.08)	5,966 (10.63)	
Asian or Pacific Islander	5,066 (9.02)	5,006 (8.92)	
Others	633 (1.13)	653 (1.16)	
Primary site, n (%)			0.006
Upper-outer quadrant	20,048 (35.71)	20,100 (35.81)	
Upper-inner quadrant	5,035 (8.97)	5,077 (9.04)	
Lower-outer quadrant	4,689 (8.35)	4,618 (8.23)	
Lower-inner quadrant	2,711 (4.83)	2,680 (4.77)	
Central portion & nipple	3,429 (6.11)	3,400 (6.06)	
Others	20,222 (36.02)	20,259 (36.09)	
Breast subtype, n (%)			0.011
$HR^+/HER2^-$	40,834 (72.74)	40,867 (72.80)	
$HR^-/HER2^-$	5,090 (9.07)	5,042 (8.98)	
$HR^+/HER2^+$	5,596 (9.97)	5,552 (9.89)	
$HR^-/HER2^+$	2,023 (3.60)	2,128 (3.79)	
Unknown	2,591 (4.62)	2,545 (4.53)	
Grade, n (%)			0.012
Grade I	8,786 (15.65)	8,718 (15.53)	
Grade II	25,916 (46.17)	26,221 (46.71)	
Grade III	20,018 (35.66)	19,797 (35.27)	
Grade IV	145 (0.26)	156 (0.28)	
Unknown	1,269 (2.26)	1,242 (2.21)	
AJCC T, 7^{th}, n (%)			0.006
T1mic	220 (0.39)	231 (0.41)	
T1a	1,004 (1.79)	1,020 (1.82)	
T1b	4,492 (8.00)	4,491 (8.00)	
T1c	20,073 (35.76)	20,086 (35.78)	
T2	27,124 (48.32)	27,047 (48.18)	
T3	3,152 (5.62)	3,195 (5.69)	
Unknown	69 (0.12)	64 (0.11)	

+, positive; –, negative. SMD, standardized mean difference; HR, hormone receptor; HER2, human epidermal growth factor receptor 2; AJCC T, the T staging of Cancer Staging Manual of the American Joint Committee on Cancer; mic, microinvasive carcinoma.

https://dx.doi.org/10.21037/gs-23-258

Table S5 SMD of breast subtype

Characteristics	Level		SMD
	Negative	Positive	
N	56,748	56,748	
Age (years), n (%)			0.011
20–44	6,935 (12.22)	7,129 (12.56)	
45–55	15,110 (26.63)	14,950 (26.34)	
56+	34,703 (61.15)	34,669 (61.09)	
Race, n (%)			0.012
White	44,750 (78.86)	44,857 (79.05)	
Black	6,386 (11.25)	6,196 (10.92)	
Asian or Pacific Islander	4,977 (8.77)	5,036 (8.87)	
Others	635 (1.12)	659 (1.16)	
Primary site, n (%)			0.018
Upper-outer quadrant	20,529 (36.18)	20,348 (35.86)	
Upper-inner quadrant	5,152 (9.08)	5,083 (8.96)	
Lower-outer quadrant	4,654 (8.20)	4,679 (8.25)	
Lower-inner quadrant	2,854 (5.03)	2,701 (4.76)	
Central portion & nipple	3,427 (6.04)	3,417 (6.02)	
Others	20,132 (35.48)	20,520 (36.16)	
Histology, n (%)			0.006
Infiltrating duct carcinoma	42,418 (74.75)	42,460 (74.82)	
Lobular carcinoma	6,106 (10.76)	6,093 (10.74)	
Mucinous adenocarcinoma	304 (0.54)	326 (0.57)	
Others	7,920 (13.96)	7,869 (13.87)	
Grade, n (%)			0.011
Grade I	8,750 (15.42)	8,686 (15.31)	
Grade II	26,112 (46.01)	26,400 (46.52)	
Grade III	20,468 (36.07)	20,255 (35.69)	
Grade IV	151 (0.27)	157 (0.28)	
Unknown	1,267 (2.23)	1,250 (2.20)	
AJCC T, 7th, n (%)			0.011
T1mic	225 (0.40)	231 (0.41)	
T1a	1,034 (1.82)	1,020 (1.80)	
T1b	4,342 (7.65)	4,491 (7.91)	
T1c	20,220 (35.63)	20,086 (35.40)	
T2	27,770 (48.94)	27,705 (48.82)	
T3	3,094 (5.45)	3,151 (5.55)	
Unknown	63 (0.11)	64 (0.11)	

SMD, standardized mean difference; AJCC T, the T staging of Cancer Staging Manual of the American Joint Committee on Cancer; mic, microinvasive carcinoma.

https://dx.doi.org/10.21037/gs-23-258

Table S6 SMD of grade

Characteristics	Level		SMD
	Negative	Positive	
N	56,265	56,265	
Age (years), n (%)			0.008
20–44	6,768 (12.03)	6,862 (12.20)	
45–55	14,925 (26.53)	14,745 (26.21)	
56+	34,572 (61.44)	34,658 (61.60)	
Race, n (%)			0.018
White	44,404 (78.92)	44,621 (79.31)	
Black	6,255 (11.12)	5,965 (10.60)	
Asian or Pacific Islander	4,992 (8.87)	5,026 (8.93)	
Others	614 (1.09)	653 (1.16)	
Primary site, n (%)			0.014
Upper-outer quadrant	20,076 (35.68)	20,198 (35.90)	
Upper-inner quadrant	4,916 (8.74)	5,082 (9.03)	
Lower-outer quadrant	4,709 (8.37)	4,618 (8.21)	
Lower-inner quadrant	2,753 (4.89)	2,686 (4.77)	
Central portion & nipple	3,418 (6.07)	3,382 (6.01)	
Others	20,393 (36.24)	20,299 (36.08)	
Histology, n (%)			0.008
Infiltrating duct carcinoma	42,163 (74.94)	41,979 (74.61)	
Lobular carcinoma	5,999 (10.66)	6,089 (10.82)	
Mucinous adenocarcinoma	334 (0.59)	326 (0.58)	
Others	7,769 (13.81)	7,871 (13.99)	
Breast subtype, n (%)			0.023
$HR^+/HER2^-$	40,680 (72.30)	41,003 (72.87)	
$HR^-/HER2^-$	5,335 (9.48)	5,019 (8.92)	
$HR^+/HER2^+$	5,563 (9.89)	5,575 (9.91)	
$HR^-/HER2^+$	2,226 (3.96)	2,116 (3.76)	
Unknown	2,461 (4.37)	2,552 (4.54)	
AJCC T, 7^{th}, n (%)			0.011
T1mic	223 (0.40)	231 (0.41)	
T1a	1,035 (1.84)	1,020 (1.81)	
T1b	4,473 (7.95)	4,491 (7.98)	
T1c	20,059 (35.65)	20,086 (35.70)	
T2	27,320 (48.56)	27,151 (48.26)	
T3	3,093 (5.50)	3,222 (5.73)	
Unknown	62 (0.11)	64 (0.11)	

+, positive. –, negative. SMD, standardized mean difference; HR, hormone receptor; HER2, human epidermal growth factor receptor 2; AJCC T, the T staging of Cancer Staging Manual of the American Joint Committee on Cancer; mic, microinvasive carcinoma.

https://dx.doi.org/10.21037/gs-23-258

Table S7 SMD of derived AJCC T, 7th

Characteristics	Level		SMD
	Negative	Positive	
N	59,485	59,485	
Age (years), n (%)			0.011
20–44	7,584 (12.75)	7,674 (12.90)	
45–55	16,223 (27.27)	15,926 (26.77)	
56+	35,678 (59.98)	35,885 (60.33)	
Race, n (%)			0.013
White	46,908 (78.86)	47,051 (79.10)	
Black	6,839 (11.50)	6,631 (11.15)	
Asian or Pacific Islander	5,103 (8.58)	5,126 (8.62)	
Others	635 (1.07)	677 (1.14)	
Primary site, n (%)			0.006
Upper-outer quadrant	21,330 (35.86)	21,205 (35.65)	
Upper-inner quadrant	5,064 (8.51)	5,108 (8.59)	
Lower-outer quadrant	4,837 (8.13)	4,813 (8.09)	
Lower-inner quadrant	2,732 (4.59)	2,720 (4.57)	
Central portion & nipple	3,828 (6.44)	3,891 (6.54)	
Others	21,694 (36.47)	21,748 (36.56)	
Histology, n (%)			0.011
Infiltrating duct carcinoma	43,985 (73.94)	44,137 (74.20)	
Lobular carcinoma	6,661 (11.20)	6,707 (11.28)	
Mucinous adenocarcinoma	313 (0.53)	326 (0.55)	
Others	8,526 (14.33)	8,315 (13.98)	
Breast subtype, n (%)			0.015
$HR^+/HER2^-$	43,040 (72.35)	43,298 (72.79)	
$HR^-/HER2^-$	5,344 (8.98)	5,155 (8.67)	
$HR^+/HER2^+$	5,954 (10.01)	6,031 (10.14)	
$HR^-/HER2^+$	2,428 (4.08)	2,334 (3.92)	
Unknown	2,719 (4.57)	2,667 (4.48)	
Grade, n (%)			0.006
Grade I	8,761 (14.73)	8,787 (14.77)	
Grade II	27,632 (46.45)	27,700 (46.57)	
Grade III	21,622 (36.35)	21,548 (36.22)	
Grade IV	153 (0.26)	166 (0.28)	
Unknown	1,317 (2.21)	1,284 (2.16)	

+, positive; –, negative. SMD, standardized mean difference; AJCC T, the T staging of Cancer Staging Manual of the American Joint Committee on Cancer; HR, hormone receptor; HER2, human epidermal growth factor receptor 2.

https://dx.doi.org/10.21037/gs-23-258