Development of a prediction model based on LASSO regression to evaluate the risk of non-sentinel lymph node metastasis in Chinese breast cancer patients with 1-2 positive sentinel lymph nodes

Lei Meng
Chongqing Medical University First Affiliated Hospital
Ting Zheng
Chongqing Medical University
Yuanyuan Wang
Chongqing Medical University First Affiliated Hospital
Qi Xiao
Chongqing Medical University First Affiliated Hospital
Junfeng He
Chongqing Medical University First Affiliated Hospital
Zhao Li
Chongqing Medical University First Affiliated Hospital
Jinxiang Tan (✉ tjx1202@163.com)
Chongqing Medical University First Affiliated Hospital  https://orcid.org/0000-0002-2587-8296

Research article

**Keywords:** breast cancer, sentinel lymph node, non-sentinel lymph node, axillary lymph node dissection, prediction model, LASSO

**Posted Date:** August 12th, 2020

**DOI:** https://doi.org/10.21203/rs.3.rs-57669/v1

**License:** This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License]
Abstract

Background: Our study aimed to investigate the correlative factors influencing non-sentinel lymph node (NSLN) metastasis in Chinese breast cancer patients with 1-2 positive sentinel lymph nodes (SLNs) and to develop an intraoperative prediction model based on the least absolute shrinkage and selection operator (LASSO) algorithm to evaluate the risk of NSLN metastasis.

Methods: The factors affecting NSLN status were investigated in a cohort of 714 patients with 1-2 positive SLNs treated at The First Affiliated Hospital of Chongqing Medical University between January 2013 and December 2018. A new mathematical prediction model based on the LASSO algorithm was developed and was validated in a cohort of 131 patients treated between January 2019 and December 2019.

Results: In the training cohort, 266/714 (37.3%) patients had NSLN metastasis. In univariate analysis, the histologic grade (P =0.010), number of positive SLNs (P<0.001), number of negative SLNs (P<0.001), number of SLNs dissected (P<0.001), SLN metastasis ratio (P <0.001), lymphovascular invasion (LVI) status (P <0.001), estrogen receptor (ER) status (P =0.011), human epidermal growth factor receptor 2 (HER2) status (P =0.005), molecular subtype (P =0.001), and risk score (P <0.001) were related with NSLN involvement. In multivariate analysis, the histologic grade (P =0.026), LVI status (P =0.005), number of positive SLNs (P=0.001), number of negative SLNs (P=0.005), SLN metastasis ratio (P =0.005), and molecular subtype (P=0.007) were identified as the independent predictors of NSLN metastasis. A LASSO regression-based mathematical prediction model was developed and had an area under the curve (AUC) of 0.764 (95% CI: 0.729-0.798). In the 131-patient validation cohort, the AUC was 0.777 (95% CI: 0.692-0.862).

Conclusions: We present a new prediction model to assess the risk of NSLN metastasis in Chinese breast cancer patients with 1-2 positive SLNs. The model was further validated in the validation cohort and showed excellent clinical applicability and diagnostic performance. It can be used as an intraoperative clinical tool for clinicians to predict the risk of NSLN metastasis and make the final decision regarding axillary lymph node dissection (ALND).

Background

As the early diagnosis rate and systemic treatment for breast cancer are improving, breast cancer surgery is becoming less traumatic and more individualized. Axillary lymph node (ALN) status is one of the most critical prognostic factors in patients with breast cancer. In the last century, sentinel lymph node biopsy (SLNB) has been indicated to be a reliable method for the axilla stage [1]. Currently, axillary lymph node dissection (ALND) can be safely avoided in breast cancer patients with negative sentinel lymph nodes (SLNs) [2]. However, ALND remains the standard management strategy when breast cancer patients are determined to have positive SLNs [3]. In recent years, data from the American College of Surgeons Oncology Group Z0011 (ACOSOG Z0011) and the International Breast Cancer Study Group 23-01 (IBCSG
23-01) trials showed that further ALND did not result in additional benefit in terms of locoregional recurrence (LRR), disease-free survival (DFS) and overall survival (OS) for patients with limited SLN involvement [4, 5]. However, the majority of patients in these studies underwent breast-conserving surgery with whole-breast irradiation, and the ALND group included fewer patients with good prognostic features of a smaller tumour diameter, hormone receptor positivity, a lower number of positive SLNs, and the absence of lymphovascular invasion (LVI). In addition, neither the specific status of human epidermal growth factor receptor 2 (HER2) and Ki67 nor strict quality control of radiotherapy were shown in these studies. In most developing countries, such as China, ALND is still recommended by most clinicians for patients with positive SLNs for the following reasons: 1. the differences between Eastern and Western populations; 2. the lack of evidence-based medical research in Eastern populations; 3. the uneven distribution of medical resources; and 4. the low rate of breast-conserving surgery (BCR) in China [6]. However, ALND is associated with increased morbidity, such as lymphoedema, paraesthesia and decreased mobility [7]. Moreover, clinical evidence has shown that only 40% of patients with positive SLNs had further axillary involvement, indicating the importance of a prediction model to evaluate the risk of non-sentinel lymph node (NSLN) metastasis in breast cancer patients with 1-2 positive SLNs [8]. Most existing prediction models, such as the Memorial Sloan Kettering Cancer Center (MSKCC) nomogram, Louisville models, MD Anderson Cancer Center score and Tenon score, were developed based on Western populations in developed countries [9-12]. A prediction model designed specifically for the Eastern population is lacking. Therefore, we sought to develop a new intraoperative mathematical prediction model based on a Chinese population to evaluate the risk of NSLN metastasis in Chinese breast cancer patients with 1-2 positive SLNs. Least absolute shrinkage and selection operator (LASSO) regression was used to construct the model.

Methods

Patients

Patients were included in our study according to the following eligibility criteria: 1. diagnosis of primary breast cancer; 2. no signs of ALN involvement discernible by physical examination or imaging; 3. the presence of only one or two positive SLNs; 4. treatment with further ALND; and 5. the availability of complete clinicopathological data. We excluded patients who had undergone neoadjuvant therapy, patients with inflammatory breast cancer or bilateral breast cancer, and patients with a history of breast cancer. The clinicopathological data of breast cancer patients who underwent SLNB and ALND at The First Affiliated Hospital of Chongqing Medical University between January 2013 and December 2018 were collected and analysed retrospectively. A mathematical prediction model was developed. The data of patients treated between January 2019 and December 2019 were used for validation.

Surgical procedures

SLNB was performed by injecting a radiocolloid ($^{99m}$Tc) combined with methylene blue. First, the $^{99m}$Tc-labelled sulfur colloid (Xinke, Beijing, China) was injected into the subareolar area 3-18 hours before
surgery. Then, the surgeon injected methylene blue (Jumpcan, Nanjing, China) into the subareolar area and massaged the breast for 5-10 minutes before the operation. Lymph nodes, any blue-stained nodes and nodes with a high radioactive count (at least 10%) as measured by a gamma detector (neo2000®, Johnson & Johnson, US) were classified as SLNs and removed for preparation of intraoperative frozen sections. All SLNs and ALNs dissected during surgery were routinely submitted for postoperative pathological sections and immunohistochemical (IHC) staining.

**Diagnostic criteria**

The tumours were classified by size as T1, T2 and T3 according to the 8th edition of the American Joint Committee on Cancer (AJCC) staging guidelines [13]. IHC staining was performed to determine the estrogen receptor (ER), progesterone receptor (PR) and Ki67 status. The HER2 status was determined by IHC staining combined with fluorescence in situ hybridization (FISH). Samples that were IHC (-) and IHC (1+) were considered HER2-negative, and samples that were IHC (3+) and FISH (+) were considered HER2-positive. All cases in the study were classified as the luminal A, luminal B, HER2 overexpression or Triple negative subtype by the 2013 St Gallen International Expert Consensus [14].

**Statistical analysis**

The $\chi^2$ test and logistic regression were utilized for univariate and multivariate analysis, respectively. Analyses were performed with SPSS 23.0 software. LASSO regression was performed with the glmnet package in R version 3.6.2 to establish a mathematical prediction model calculating the risk scores (RS) of the patients. The RS of each patient was calculated by the mathematical formula. We used SPSS 23.0 software to generate the receiver operating characteristic (ROC) curve. The performance of the prediction model was assessed by the area under the ROC curve (AUC) in the training cohort and the validation cohort. P<0.05 was considered significant.

**Results**

**General demographic and characteristics**

Ultimately, a total of 845 patients with 1-2 positive SLNs were enrolled in our study; 714 were in the training cohort and 131 were in the validation cohort. All patients were female and ranged in age from 22 to 88 years. The median age was 49 years in both groups. In the training cohort, 266 patients (37.3%) were demonstrated to have NSLN metastasis by analysis of postoperative pathological sections, meaning that the other 448 patients (62.7%) underwent unnecessary ALND. Then, the patients in the training cohort were divided into the low-RS group and the high-RS group, in which 26 patients (10.8%) and 240 patients (50.6%), respectively, were observed to have further NSLN metastasis. More clinicopathological parameters in the training cohort are shown in Table 1. The clinicopathological characteristics of patients in the validation cohort are shown in Table 2.

**Table 1** Clinicopathological characteristics of the 714 patients in the training cohort
| Variable                        | NSLN-negative (%) (N=448) | NSLN-positive (%) (N= 266) | $\chi^2$ | $P$ |
|--------------------------------|---------------------------|-----------------------------|----------|-----|
| Age group (years)              |                           |                             | 0.958    | 0.328 |
| ≤50                            | 271 (64.2)                | 151 (35.8)                  |          |     |
| >50                            | 177 (60.6)                | 115 (39.4)                  |          |     |
| Side                           |                           |                             | 0.273    | 0.602 |
| Left                           | 228 (63.7)                | 130 (36.3)                  |          |     |
| Right                          | 220 (61.8)                | 136 (38.2)                  |          |     |
| Location                       |                           |                             | 2.698    | 0.610 |
| Centre area                    | 16 (53.3)                 | 14 (46.7)                   |          |     |
| Upper inner quadrant           | 84 (67.7)                 | 40 (32.3)                   |          |     |
| Lower inner quadrant           | 49 (60.5)                 | 32 (39.5)                   |          |     |
| Upper outer quadrant           | 180 (62.1)                | 110 (37.9)                  |          |     |
| Under outer quadrant           | 119 (63.0)                | 70 (37.0)                   |          |     |
| Clinical tumour stage          |                           |                             | 3.617    | 0.164 |
| T1                             | 220 (66.3)                | 112 (33.7)                  |          |     |
| T2                             | 224 (59.9)                | 150 (40.1)                  |          |     |
| T3                             | 4 (50.0)                  | 4 (50.0)                    |          |     |
| Histologic type                |                           |                             | 2.494    | 0.287 |
| Invasive ductal carcinoma      | 426 (62.2)                | 259 (37.8)                  |          |     |
| Invasive lobular carcinoma     | 9 (81.8)                  | 2 (18.2)                    |          |     |
| Other                          | 13 (72.2)                 | 266 (37.3)                  |          |     |
| Histologic grade               |                           |                             | 9.204    | 0.010 |
| ₁                               | 17 (70.8)                 | 7 (29.2)                    |          |     |
| ₂                               | 377 (64.8)                | 205 (35.2)                  |          |     |
| ₃                               | 54 (50.0)                 | 54 (50.0)                   |          |     |
| Number of positive SLNs        |                           |                             | 26.285   | <0.001 |
| 1                              | 347 (68.7)                | 158 (31.3)                  |          |     |
| 2                              | 101 (48.3)                | 108 (51.7)                  |          |     |
| Number of negative SLNs        |                           |                             | 97.841   | <0.001 |
| 0                              | 58 (36.5)                 | 101 (63.5)                  |          |     |
| 1                              | 77 (50.0)                 | 77 (50.0)                   |          |     |
| ≥2                             | 313 (78.1)                | 88 (21.9)                   |          |     |
| Number of SLNs dissected       |                           |                             | 68.321   | <0.001 |
| 1-2                            | 110 (42.8)                | 147 (57.2)                  |          |     |
| Clinicopathological Characteristics | 131 Patients in the Validation Cohort |
|------------------------------------|--------------------------------------|
| SLN metastasis ratio               |                                      |
| ≥3                                 | 338 (74.0)                           |
| <0.5                               | 299 (80.6)                           |
| ≥0.5                               | 149 (43.4)                           |
| LVI                                |                                      |
| No                                 | 442 (64.2)                           |
| Yes                                | 6 (24.0)                             |
| ER                                 |                                      |
| Negative                           | 92 (54.4)                            |
| Positive                           | 356 (65.3)                           |
| PR                                 |                                      |
| Negative                           | 137 (58.1)                           |
| Positive                           | 311 (65.1)                           |
| HER2                               |                                      |
| Negative                           | 326 (66.1)                           |
| Positive                           | 122 (55.2)                           |
| Ki67 (%)                           |                                      |
| <14                                | 283 (60.9)                           |
| ≥14                                | 165 (66.3)                           |
| Molecular subtype                  |                                      |
| Luminal A                          | 102 (68.0)                           |
| Luminal B                          | 263 (64.9)                           |
| HER2 overexpression                | 37 (43.5)                            |
| Triple negative                    | 46 (62.2)                            |
| P53                                |                                      |
| Negative                           | 142 (62.3)                           |
| Positive                           | 306 (63.0)                           |
| RS*                                |                                      |
| Low                                | 214 (89.2)                           |
| High                               | 234 (49.4)                           |

NSLN non-sentinel lymph node, SLNs sentinel lymph nodes, LVI lymphovascular invasion, ER estrogen receptor, PR progesterone receptor, HER2 human epidermal growth factor receptor 2, RS risk score. *Patients were divided into Low-RS group and High-RS group when the cutoff value of RS was set as 1.059328005.

Table 2 Clinicopathological characteristics of the 131 patients in the validation cohort
| Variable                              | Validation cohort(%) (N= 238) |
|--------------------------------------|--------------------------------|
| Age group (years)                    |                                |
| \( \leq 50 \)                       | 71 (54.2)                      |
| >50                                  | 60 (45.8)                      |
| Side                                 |                                |
| Left                                 | 58 (44.3)                      |
| Right                                | 73 (55.7)                      |
| Location                             |                                |
| Centre area                          | 11 (8.4)                       |
| Upper inner quadrant                 | 24 (18.3)                      |
| Lower inner quadrant                 | 10 (7.6)                       |
| Upper outer quadrant                 | 67 (51.2)                      |
| Under outer quadrant                 | 19 (14.5)                      |
| Clinical tumour stage                |                                |
| T1                                   | 45 (34.3)                      |
| T2                                   | 83 (63.4)                      |
| T3                                   | 3 (2.3)                        |
| Histologic type                      |                                |
| Invasive ductal carcinoma            | 125 (95.4)                     |
| Invasive lobular carcinoma           | 4 (3.1)                        |
| Other                                | 2 (1.5)                        |
| Histologic grade                     |                                |
| 1                                    | 2 (1.5)                        |
| 2                                    | 103 (78.6)                     |
| 3                                    | 26 (19.9)                      |
| Number of positive SLNs              |                                |
| 1                                    | 81 (61.8)                      |
| 2                                    | 50 (38.2)                      |
| Number of negative SLNs              |                                |
| 0                                    | 26 (19.8)                      |
| 1                                    | 19 (14.5)                      |
| \( \geq 2 \)                        | 86 (65.7)                      |
| Number of SLNs dissected             |                                |
| 1-2                                  | 37 (28.2)                      |
| SLN metastasis ratio | ≥3 | 94 (71.8) |
|----------------------|----|------------|
|                      | ≥0.5 | 76 (58.0) |
|                      | ≥0.5 | 55 (42.0) |
| LVI                  | No | 117 (89.3) |
|                      | Yes | 14 (10.7)  |
| ER                   | Negative | 26 (19.8) |
|                      | Positive | 105 (80.2) |
| PR                   | Negative | 37 (28.2) |
|                      | Positive | 94 (71.8)  |
| HER2                 | Negative | 99 (75.6) |
|                      | Positive | 32 (24.4)  |
| Ki67 (%)             | ≥14 | 34 (26.0)  |
|                      | ≥14 | 97 (74.0)  |
| Molecular subtype    | Luminal A | 31 (23.7) |
|                      | Luminal B | 75 (57.2) |
|                      | HER2-enriched | 8 (6.1) |
|                      | Triple negative | 17 (13.0) |
| P53                  | Negative | 57 (43.5) |
|                      | Positive | 74 (56.5)  |

**Univariate and multivariate analysis results in the training cohort**

Univariate analysis showed that the histologic grade (P =0.010), number of positive SLNs (P<0.001), number of negative SLNs (P<0.001), number of SLNs dissected (P<0.001), SLN metastasis ratio (P <0.001), LVI status (P <0.001), ER status (P =0.011), HER2 status (P =0.005), molecular subtype (P =0.001), and RS (P <0.001) were associated with NSLN involvement in breast cancer patients with 1-2
positive SLNs (Table 1). In further logistic regression multivariate analysis, the histologic grade ($P = 0.026$), LVI status ($P = 0.005$), number of positive SLNs ($P = 0.001$), number of negative SLNs ($P = 0.005$), SLN metastasis ratio ($P = 0.005$), and molecular subtype ($P = 0.007$) were identified as the independent predictive factors for NSLN metastasis (Table 3). Compared with triple negative breast cancer patients, patients with the HER2 overexpression subtype were more likely to have positive NSLNs, whereas patients with the luminal A and luminal B subtypes showed no significant difference in NSLN metastasis.

**Table 3** Logistic regression multivariate analysis of variables associated with NSLN metastasis

| Variables                  | OR      | 95% CI         | $P$  |
|----------------------------|---------|----------------|------|
| LVI                        | 4.378   | 1.569-12.214   | 0.005|
| Histologic grade           | 1.630   | 1.061-2.505    | 0.026|
| Number of positive SLNs    | 1.898   | 1.299-2.773    | 0.001|
| Number of negative SLNs    | 0.594   | 0.414-0.853    | 0.005|
| SLN metastasis ratio       | 2.414   | 1.307-4.459    | 0.005|
| Molecular subtype          |         |                | 0.007|
| Triple negative            | 1       | Reference      |      |
| Luminal A                  | 0.690   | 0.352-1.351    | 0.279|
| Luminal B                  | 0.873   | 0.487-1.565    | 0.649|
| HER2 overexpression         | 1.990   | 0.972-4.071    | 0.060|

**Development of a mathematical prediction model based on LASSO regression**

In our study, LASSO regression was used to develop a mathematical prediction model. Finally, LASSO regression analysis identified the 13 most powerful factors: namely, the age group, clinical tumour stage, histologic type, number of positive SLNs, number of negative SLNs, number of SLNs dissected, SLN metastasis ratio, ER status, PR status, HER2 status, Ki67 staining percentage, molecular subtype and P53 status (Figure 1). Among these factors, the SLN metastasis ratio was the most influential factor in the RS, with the maximum absolute value of the coefficient. The regression coefficients of each factor are shown in Table 4. The RS of each patient in the study was calculated using the following model equation:

\[
RS = \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n \quad (X: \text{a factor}; \beta: \text{the regression coefficient of that factor})
\]

**Table 4** The regression coefficients of the 13 most powerful factors identified by LASSO regression analysis
### Performance of the prediction model

Furthermore, we generated the ROC curve for the prediction model. The AUC was 0.764 (95% CI: 0.729-0.798), which showed the excellent diagnostic performance of this model (Fig. 2). The RS value of 1.87239924305 was identified as the cutoff value with the highest Youden index. The sensitivity, specificity and overall accuracy of the model were 74.1%, 69.6% and 71.3%, respectively. The false-negative rate (FNR) was as high as 25.9%. However, when the cutoff value of RS was set as 1.059327995, the sensitivity, specificity, total accuracy and FNR were 90.2%, 47.8%, 63.6% and 9.8%, respectively. Our prediction model also showed a satisfying predictive value in the validation cohort of 131 patients, with an AUC of 0.777 (95% CI: 0.692-0.862) (Fig. 3).

### Discussion

In recent years, the results of the ACOSOG Z0011 and IBCSG 23-01 trials showed that neither the DFS nor the OS differed significantly between the SLNB-only and ALND groups among breast cancer patients with limited SLN involvement. Based on the results of these trials, the latest NCCN guidelines also recommended that ALND not be performed in patients with 1-2 involved SLNs who were planning to undergo breast-conserving surgery and subsequent radiotherapy [15]. However, in most developing countries, such as China, the BCR is only approximately 20%, compared to 50%-80% in Western countries [6, 16, 17]. Even in some leading centres in China, the BCR is only 30% [18]. Because of the low BCR in developing countries and the absence of evidence for ALND omission in Eastern populations, most
clinicians in developing countries such as China still hold a conservative view and recommend ALND for patients with positive SLNs [16]. In addition, the ALN status remains one of the most important prognostic factors. In our present study, only 266 (37.3%) of the patients with 1-2 SLN metastases in the training cohort were demonstrated to have NSLN metastasis after ALND, consistent with the results of previous studies [8, 19]. More than 60% of patients thus received unnecessary ALND. Therefore, it is of great importance to accurately predict NSLN metastasis either intraoperatively or preoperatively. Our study retrospectively analysed the clinicopathological data of 714 breast cancer patients with 1-2 positive SLNs in the training cohort to determine the factors associated with axillary involvement. We further developed a new mathematical prediction model based on this Chinese population to evaluate the risk of NSLN metastasis.

Previous studies have shown that LVI is a feature related to poor prognosis and that it promotes local recurrence and distant metastasis of tumours [20]. Several recent studies have recognized LVI as an independent predictor of NSLN metastasis in patients with 1-2 positive SLNs [21, 22]. We arrived at the same conclusion. In our study, 76.0% and 35.8% of patients with LVI and without LVI, respectively, were found to have NSLN involvement, and this difference was significant. It remains controversial whether histologic grade is associated with NSLN metastasis, and the conclusions from Maimaitiali A and Wang XY were inconsistent [23, 24]. Our univariate analysis showed that patients with higher histologic grades were more likely to have at least one positive ALN, and histologic grade remained an independent predictor of NSLN metastasis in the subsequent multivariate analysis (OR = 1.630; 95% CI: 1.061-2.505; P = 0.026).

Moreover, we divided all patients in the training cohort into the luminal A, luminal B, HER2 overexpression and triple negative subtypes according to St Gallen International Expert Consensus (2013 edition) [14]. In these respective molecular subtype groups, 32.0%, 35.1%, 56.5% and 37.8% of the patients exhibited NSLN involvement. Compared to the triple negative type, the HER2 overexpression type was associated with a statistically higher risk of positive NSLNs, but the luminal A and luminal B subtypes were not. Whether NSLN metastasis is associated with the molecular subtype remains controversial. The results from a recent single-centre study of 291 patients demonstrated that patients with luminal B and HER2 overexpression breast cancer had a significantly higher possibility of having at least one positive NSLN than patients with luminal A breast cancer [25]. However, in another retrospective study, investigators failed to identify molecular subtype as an independent predictor of NSLN metastasis. Patients with positive SLNs had the same risk of axillary involvement regardless of their molecular subtypes [21].

The number of positive SLNs, number of negative SLNs, the number of SLNs dissected and the SLN metastasis ratio were important predictors of NSLN metastasis in breast cancer patients with 1-2 positive SLNs. These factors rely heavily on assessment in intraoperative frozen sections. Thus, these values are unclear prior to surgery. Two publications considered the numbers of positive and negative SLNs as the independent risk factors included in their prediction models [26, 27]. The SLN metastasis ratio was incorporated into the model predictions in another clinical study [28]. However, the value of the number of positive SLNs, number of negative SLNs, number of SLNs dissected and SLN metastasis ratio for
predicting NSLN metastasis has not been fully clarified because of the collinearity among these factors. In our study, LASSO regression was used to construct the mathematical model, thus effectively solving the problem of collinearity among these factors.

The MSKCC nomogram, the first model to predict NSLN metastasis in patients with positive SLNs, performed well in the original population, with an AUC of 0.76 [9]. The results from a previous study showed that the AUC of the MSKCC nomogram was less than 0.7, proving that the performance of the MSKCC nomogram was inferior to other models in other populations [29-31]. In addition, the MSKCC nomogram and other previous models were developed based on Western populations in developed countries, and thus hardly apply to the Eastern population. In the present retrospective analysis, we developed a new, LASSO algorithm-based intraoperative mathematical prediction model based on the clinical data of 714 patients in China for evaluating the risk of NSLN metastasis in Chinese breast cancer patients with 1-2 positive SLNs. The LASSO algorithm forces the sum of the absolute value of the regression coefficients to be less than a fixed value by reducing certain coefficients to zero, which helps to effectively construct a simpler model that includes only the most meaningful predictive factors. Ultimately, LASSO regression identified the 13 most powerful predictors: age group, clinical tumour stage, histologic type, number of positive SLNs, number of negative SLNs, number of SLNs dissected, SLN metastasis ratio, ER status, PR status, HER2 status, Ki67 staining percentage, molecular subtype and P53 status. The coefficients of each predictor are shown in Table 4. The higher the absolute value of a regression coefficient, the greater is its influence on the model. In our prediction model, the most powerful predictor was the SLN metastasis ratio, with a coefficient of 0.7672377992. This factor was also included in the Cambridge model and another recent model [32, 33].

Previous evidences showed that the absolute agreement rate of histologic grade and LVI were only 75% and 69%, respectively, between the specimens obtained by core needle biopsy (CNB) and those obtained by surgical excision [34, 35]. The small volume of specimens from CNB and intratumoural heterogeneity are the possible reasons for the low concordance rate of histologic grade and LVI status between the preoperative CNB and postoperative pathology results. Considering that the aim of our study was to develop an intraoperative prediction model, histologic grade and LVI status, which are not entirely available via preoperative or intraoperative evaluation, were excluded from our prediction model, although they were identified as independent risk factors in the retrospective multivariate analysis.

Furthermore, we calculated the RS of each patient in the training cohort according to the model equation. The ROC curve for the prediction model was then generated and shown to have an AUC of 0.764 (95% CI: 0.729-0.798), comparable to that of the MSKCC nomogram [9]. Thus, the predictive power of the model was acceptable. Finally, ROC curve analysis confirmed the cutoff value of the RS to be 1.87239924305. The sensitivity, specificity and total accuracy were 74.1%, 69.6% and 71.3%, respectively. However, the FNR was as high as 25.9%. When the cutoff value of RS was set as 1.059327995, the FNR was only 9.8%, less than the clinically acceptable rate of 10%. Therefore, we believe that ALND may be safely ignored when the RS of a patient, as calculated by the model equation, is less than the cutoff value of 1.059327995. We divided the patients in the training cohort into a low-RS group and a high-RS group.
according to the cutoff value. Significantly more patients had positive NSLNs in the high-RS group than in the low-RS group, which further confirmed the predictive power of our model.

To evaluate the clinical applicability of the prediction model, a subsequent independent cohort of 131 patients was used for validation. The model still showed impressive performance, with an AUC of 0.777 (95% CI: 0.692-0.862). The present prediction model can thus be considered an intraoperative clinical tool for clinicians to predict the risk of NSLN metastasis in Chinese breast cancer patients with 1-2 positive SLNs and make the decision regarding ALND.

To our knowledge, this study is the first one to apply the LASSO algorithm to develop a prediction model based on an Eastern population. As more than 13 factors were included in our model, it offers a more personalized assessment for breast cancer patients. However, there are a few limitations in our study. First, this was a retrospective study, and a prospective clinical trial is greatly needed. Second, our prediction model should be validated with population data from other centres.

Conclusion

In conclusion, we developed a new intraoperative mathematical prediction model with 13 predictors based on the LASSO algorithm to evaluate the risk of NSLN metastasis in Chinese breast cancer patients with 1-2 positive SLNs. The model performed well in both the training cohort and validation cohort and has good clinical applicability.

Abbreviations

ALN: Axillary lymph node; SLNB: sentinel lymph node biopsy; ALND: axillary lymph node dissection; SLNs: sentinel lymph nodes; ACOSOG Z0011: American College of Surgeons Oncology Group Z0011; IBCSG 23-01: International Breast Cancer Study Group 23-01; LRR: locoregional recurrence; DFS: disease-free survival; OS: overall survival; LVI: lymphovascular invasion; HER2: human epidermal growth factor receptor-2; BCR: rate of breast-conserving surgery; NSLN: non-sentinel lymph node; MSKCC: Memorial Sloan Kettering Cancer Center nomogram; LASSO: Least Absolute Shrinkage and Selection Operator; IHC: immunohistochemical; AJCC: American Joint Committee on Cancer; ER: estrogen receptor; PR: progesterone receptor; FISH: fluorescence in situ hybridization; RS: risk scores; ROC: receiver operating characteristic; AUC: the area under the ROC curve; CNB: core needle biopsy;

Declarations

Acknowledgements

The authors thank all colleagues for participating in this study.

Authors’ contributions
LM developed the study design, performed the data analyses, developed the prediction model and drafted the manuscript. JXT developed the study design, supervised the study, contributed to the critical revision and editing of the final version of the manuscript for publication. TZ contributed to literature review and the writing of the manuscript. YYW performed the data analyses and model validation. QX, JFH and ZL helped to collect and check the study data. All authors read and approved the final manuscript.

**Funding**

This study was supported by the National Natural Science Foundation for Young Scholars of China (81102008); Chongqing Human Resources and Social Security Bureau returned overseas students selected funding project (2013012); Science and health joint medical research project of Chongqing Health Commission.

**Availability of data and materials**

The datasets that support the findings of this study will be available from the corresponding author upon reasonable request.

**Ethics approval and consent to participate**

The study was approved by the Ethics Committee of The First Affiliated Hospital of Chongqing Medical University (2020-309).

**Consent for publication**

Not applicable.

**Competing interests**

The authors have no conflicts of interest.

**Author details**

Lei Meng¹ E-mail: 953172545@qq.com

Ting Zheng² E-mail: 1069797933@qq.com

Yuanyuan Wang¹ E-mail: 2964928845@qq.com

Qi Xiao¹ E-mail: 18716349909@qq.com

Junfeng He¹ E-mail: 1012554324@qq.com

Zhao Li¹ E-mail: 1195272572@qq.com
References

1. Giuliano AE, Kirgan DM, Guenther JM, Morton DL. Lymphatic mapping and sentinel lymphadenectomy for breast cancer. Annals of surgery. 1994;220(3):391-398; discussion 398-401.

2. Krag DN, Anderson SJ, Julian TB, Brown AM, Harlow SP, Costantino JP, Ashikaga T, Weaver DL, Mamounas EP, Jalovec LM et al. Sentinel-lymph-node resection compared with conventional axillary-lymph-node dissection in clinically node-negative patients with breast cancer: overall survival findings from the NSABP B-32 randomised phase 3 trial. The Lancet Oncology. 2010;11(10):927-933.

3. Lyman GH, Giuliano AE, Somerfield MR, Benson AB, 3rd, Bodurka DC, Burstein HJ, Cochran AJ, Cody HS, 3rd, Edge SB, Galper S et al. American Society of Clinical Oncology guideline recommendations for sentinel lymph node biopsy in early-stage breast cancer. Journal of clinical oncology: official journal of the American Society of Clinical Oncology. 2005;23(30):7703-7720.

4. Giuliano AE, Ballman KV, McCall L, Beitsch PD, Brennan MB, Kelemen PR, Ollila DW, Hansen NM, Whitworth PW, Blumencranz PW 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(10):918-926.

5. Galimberti V, Cole BF, Viale G, Veronesi P, Vicini E, Intra M, Mazzarol G, Massarut S, Zgajnar J, Taffurelli M et al. Axillary dissection versus no axillary dissection in patients with breast cancer and sentinel-node micrometastases (IBCSG 23-01): 10-year follow-up of a randomised, controlled phase 3 trial. The Lancet Oncology. 2018;19(10):1385-1393.

6. Huang NS, Liu MY, Chen JJ, Yang BL, Xue JY, Quan CL, Mo M, Liu GY, Shen ZZ, Shao ZM et al. Surgical management of breast cancer in China: A 15-year single-center retrospective study of 18,502 patients. Medicine. 2016;95(45): e4201.

7. Lucci A, McCall LM, Beitsch PD, Whitworth PW, Reintgen DS, Blumencranz PW, Leitch AM, Saha S, Hunt KK, Giuliano AE. Surgical complications associated with sentinel lymph node dissection (SLND) plus axillary lymph node dissection compared with SLND alone in the American College of Surgeons Oncology Group Trial Z0011. Journal of clinical oncology: official journal of the American Society of Clinical Oncology. 2007;25(24):3657-3663.

8. Straver ME, Meijnen P, van Tienhoven G, van de Velde CJ, Mansel RE, Bogaerts J, Duez N, Cataliotti L, Klinkenbijl JH, Westenberg HA et al. Sentinel node identification rate and nodal involvement in the EORTC 10981-22023 AMAROS trial. Annals of surgical oncology. 2010;17(7):1854-1861.
9. Van Zee KJ, Manasseh DM, Bevilacqua JL, Boolbol SK, Fey JV, Tan LK, Borgen PI, Cody HS, 3rd, Kattan MW. A nomogram for predicting the likelihood of additional nodal metastases in breast cancer patients with a positive sentinel node biopsy. Annals of surgical oncology. 2003;10(10):1140-1151.

10. Chagpar AB, Scoggins CR, Martin RC, 2nd, Carlson DJ, Laidley AL, El-Eid SE, McGlothlin TQ, McMasters KM. Prediction of sentinel lymph node-only disease in women with invasive breast cancer. American journal of surgery. 2006;192(6):882-887.

11. Meretoja TJ, Leidenius MH, Heikkilä PS, Boross G, Sejben I, Regitnig P, Luschin-Ebengreuth G, Žgajnar J, Perhavec A, Gazic B et al. International multicenter tool to predict the risk of nonsentinel node metastases in breast cancer. Journal of the National Cancer Institute. 2012;104(24):1888-1896.

12. Barranger E, Coutant C, Flahault A, Delpech Y, Darai E, Uzan S. An axilla scoring system to predict non-sentinel lymph node status in breast cancer patients with sentinel lymph node involvement. Breast cancer research and treatment. 2005;91(2):113-119.

13. Amin MB, Greene FL, Edge SB, Compton CC, Gershenwald JE, Brookland RK, Meyer L, Gress DM, Byrd DR, Winchester DP. The Eighth Edition AJCC Cancer Staging Manual: Continuing to build a bridge from a population-based to a more “personalized” approach to cancer staging. CA: a cancer journal for clinicians. 2017;67(2):93-99.

14. Goldhirsch A, Winer EP, Coates AS, Gelber RD, Piccart-Gebhart M, Thurlimann B, Senn HJ. Personalizing the treatment of women with early breast cancer: highlights of the St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2013. Annals of oncology: official journal of the European Society for Medical Oncology. 2013;24(9):2206-2223.

15. Riba MB, Donovan KA, Andersen B, Braun I, Breitbart WS, Brewer BW, Buchmann LO, Clark MM, Collins M, Corbett C et al. Distress Management, Version 3.2019, NCCN Clinical Practice Guidelines in Oncology. Journal of the National Comprehensive Cancer Network: JNCCN. 2019;17(10):1229-1249.

16. Yang B, Ren G, Song E, Pan D, Zhang J, Wang Y, Liao N, Tang J, Wang X, Cui S et al. Current Status and Factors Influencing Surgical Options for Breast Cancer in China: A Nationwide Cross-Sectional Survey of 110 Hospitals. Oncologist. 2020;

17. Chiu AS, Thomas P; Killelea BK, Horowitz N, Chagpar AB, Lannin DR. Regional variation in breast cancer surgery: Results from the National Cancer Database (NCDB). American journal of surgery. 2017;214(5):907-913.

18. Bao X, Sun K, Tian X, Yin Q, Jin M, Yu N, Jiang H, Zhang J, Hu Y. Present and changing trends in surgical modalities and neoadjuvant chemotherapy administration for female breast cancer in Beijing, China: A 10-year (2006-2015) retrospective hospitalization summary report-based study. Thoracic cancer. 2018;9(6):707-717.

19. Maimaiti Aili, Wu D, Liu Z, Liu H, Muyiduli X, Fan Z. Analysis of factors related to non-sentinel lymph node metastasis in 296 sentinel lymph node-positive Chinese breast cancer patients. Cancer biology & medicine. 2018;15(3):282-289.
20. Shen SD, Zhong SZ, Wang CZ, Huang WH. Correlation of lymphovascular invasion with clinicopathological factors in invasive breast cancer: a meta-analysis. International journal of clinical and experimental medicine. 2015;8(10):17789-17795.

21. Nowikiewicz T, Wnuk P, Malkowski B, Kurylcio A, Kowalewski J, Zegarski W. Application of artificial neural networks for predicting presence of non-sentinel lymph node metastases in breast cancer patients with positive sentinel lymph node biopsies. Archives of medical science: AMS. 2017;13(6):1399-1407.

22. Kuru B, Sullu Y, Yuruker S, Koray Bayrak I, Ozen N. Factors predicting non-sentinel lymph node metastasis in T1-2 invasive breast cancer with 1-2 axillary sentinel lymph node metastases: Presentation of Ondokuz Mayis scoring system. Journal of BUON: official journal of the Balkan Union of Oncology. 2016;21(5):1129-1136.

23. Maimaitiali A, Wu D, Liu Z, Liu H, Muyiduli X, Fan Z. Erratum to Analysis of factors related to non-sentinel lymph node metastasis in 296 sentinel lymph node-positive Chinese breast cancer patients. Cancer biology & medicine. 2018;15(4):478.

24. Wang XY, Wang JT, Guo T, Kong XY, Chen L, Zhai J, Gao YQ, Fang Y, Wang J. Risk factors and a predictive nomogram for non-sentinel lymph node metastases in Chinese breast cancer patients with one or two sentinel lymph node macrometastases and mastectomy. Current oncology (Toronto, Ont). 2019;26(2): e210-e215.

25. Wang NN, Yang ZJ, Wang X, Chen LX, Zhao HM, Cao WF, Zhang B. A mathematical prediction model incorporating molecular subtype for risk of non-sentinel lymph node metastasis in sentinel lymph node-positive breast cancer patients: a retrospective analysis and nomogram development. Breast cancer (Tokyo, Japan). 2018;25(6):629-638.

26. Chen JY, Chen JJ, Xue JY, Chen Y, Liu GY, Han QX, Yang JT, Shen ZZ, Shao ZM, Wu J. Predicting Non-sentinel Lymph Node Metastasis in a Chinese Breast Cancer Population with 1-2 Positive Sentinel Nodes: Development and Assessment of a New Predictive Nomogram. World journal of surgery. 2015;39(12):2919-2927.

27. Bi X, Wang Y, Li M, Chen P, Zhou Z, Liu Y, Zhao T, Zhang Z, Wang C, Sun X et al. Validation of the Memorial Sloan Kettering Cancer Center nomogram for predicting non-sentinel lymph node metastasis in sentinel lymph node-positive breast-cancer patients. OncoTargets and therapy. 2015;8:487-493.

28. Houvenaeghel G, Nos C, Giard S, Mognotte H, Esterni B, Jacquemier J, Buttarelli M, Classe JM, Cohen M, Rouanet P et al. A nomogram predictive of non-sentinel lymph node involvement in breast cancer patients with a sentinel lymph node micrometastasis. European journal of surgical oncology: the journal of the European Society of Surgical Oncology and the British Association of Surgical Oncology. 2009;35(7):690-695.

29. Fougo JL, Senra FS, Araujo C, Dias T, Afonso M, Leal C, Dinis-Ribeiro M. Validating the MSKCC nomogram and a clinical decision rule in the prediction of non-sentinel node metastases in a Portuguese population of breast cancer patients. Breast (Edinburgh, Scotland). 2011;20(2):134-140.
30. Moghaddam Y, Falzon M, Fulford L, Williams NR, Keshtgar MR. Comparison of three mathematical models for predicting the risk of additional axillary nodal metastases after positive sentinel lymph node biopsy in early breast cancer. The British journal of surgery. 2010;97(11):1646-1652.

31. Coufal O, Pavlik T, Fabian P, Bori R, Boross G, Sejben I, Maraz R, Koca J, Krejci E, Horakova I et al. Predicting non-sentinel lymph node status after positive sentinel biopsy in breast cancer: what model performs the best in a Czech population? Pathology oncology research: POR. 2009;15(4):733-740.

32. Pal A, Provenzano E, Duffy SW, Pinder SE, Purushotham AD. A model for predicting non-sentinel lymph node metastatic disease when the sentinel lymph node is positive. The British journal of surgery. 2008;95(3):302-309.

33. Gruber I, Henzel M, Schonfisch B, Stabler A, Taran FA, Hahn M, Rohm C, Helms G, Oberlechner E, Wiesinger B et al. Prediction of Non-sentinel Lymph Node Metastases After Positive Sentinel Lymph Nodes Using Nomograms. Anticancer research. 2018;38(7):4047-4056.

34. Usami S, Moriya T, Amari M, Suzuki A, Ishida T, Sasano H, Ohuchi N. Reliability of prognostic factors in breast carcinoma determined by core needle biopsy. Japanese journal of clinical oncology. 2007;37(4):250-255.

35. Harris GC, Denley HE, Pinder SE, Lee AH, Ellis IO, Elston CW, Evans A. Correlation of histologic prognostic factors in core biopsies and therapeutic excisions of invasive breast carcinoma. The American journal of surgical pathology. 2003;27(1):11-15.

Figures
Identification of the influencing factors by LASSO regression. LASSO regression identified the 13 most powerful predictors: age group, clinical tumour stage, histologic type, number of positive SLNs, number of negative SLNs, number of SLNs dissected, SLN metastasis ratio, ER status, PR status, HER2 status, Ki67 staining percentage, molecular subtype and P53 status. Plot B shows the coefficients of each predictor when these 13 predictors were included in LASSO regression model.
ROC curve for the training cohort. The AUC was 0.764 (95% CI: 0.729-0.798).

Figure 3

ROC curve for the validation cohort. The AUC was 0.777 (95% CI: 0.692-0.862).