Nomogram to predict pathological axillary lymph node status after neoadjuvant therapy in triple negative or HER2 positive breast cancer

Shujie Chen
International Peace Maternity and Child Health Hospital

Qinyu Zhang
International Peace Maternity and Child Health Hospital

Min Ji
International Peace Maternity and Child Health Hospital

Li Yang
International Peace Maternity and Child Health Hospital

Jie Wang (✉ jiewang76@hotmail.com)
International Peace Maternity and Child Health Hospital

https://orcid.org/0000-0002-1239-345X

Research Article

Keywords: Breast cancer, Triple negative breast cancer, HER2 positive breast cancer, Lymph nodes, Neoadjuvant therapy, Nomogram

Posted Date: July 11th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3019817/v1

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

**Purpose:** Axillary lymph node (ALN) pathologic complete remission (pCR) rate after neoadjuvant therapy (NAT) is high in triple negative (TN) or human epidermal growth factor receptor 2-positive (HER2+) breast cancer patients. We aimed to identify factors associated with pathological ALN status after NAT in these patients, and establish a nomogram model to avoid unnecessary ALND.

**Methods:** TN or HER2+ breast cancer patients receiving NAT in the Shanghai Jiao Tong University Breast Cancer Database (SJTU-BCDB) were retrospectively included in training cohort and internal validation cohort. Patients at the International Peace Maternity & Child Health Hospital (IPMCH) of China Welfare Institute were retrospectively collected for external validation. Based on univariate and multivariate logistic regression, a nomogram model was constructed to predict the probability of pathologically node-positive disease after NAT (ypN+) in TN or HER2+ patients.

**Results:** 1,686 patients were assigned to the training set, and 723 patients in the validation set. Five independent factors including clinical nodal (cN) stage (P < 0.001), molecular subtype (P < 0.001), Ki67 expression (P = 0.003), tumor grade (P < 0.001), and clinical complete response (P < 0.001) together with clinical tumor (cT) stage were selected to construct the nomogram. The nomogram indicated the areas under ROC curve (AUCs) were 0.782, 0.753 and 0.783 in training cohort, internal validation cohort and external validation cohort, respectively.

**Conclusion:** We developed a nomogram model for predicting the risk of ypN+ in TN or HER2+ breast cancer patients, which may guide the de-escalating ALN surgery after NAT.

1 Introduction

Neoadjuvant therapy (NAT) was used as the standard therapy for locally advanced breast cancer, making initially inoperable breast cancer to be operable [1]. Currently, NAT is further used in the de-escalation treatment of patients with early breast cancer [2], increasing the proportion of breast conserving surgery and decreasing the rate of axillary lymph node dissection (ALND) [3, 4]. With the latest NAT regimen of combining anthracyclines (A) with taxanes (T), breast pathologic complete remission (pCR) rate ranged from 20–46% [5–7]. The addition of molecular targeted drugs to NAT regimen could further increase the breast pCR rate to 40%-70% in patients with human epidermal growth factor receptor 2 positive (HER2+) diseases [7, 8]. For triple negative (TN) breast cancer patients, the adding of platinum or anti-PD-1 antibody would also increase the breast pCR to 50%-70% [9–13].

Axillary lymph node (ALN) pCR is considered to be associated with an excellent prognosis [14, 15], which could reach 30%-49% in breast cancer patients with initial positive axillary lymph nodes [6, 16–21]. However, due to the absence of accuracy methods to predict axillary nodal pCR, ALND is still recommended for the majority of patients with biopsy-confirmed positive ALN [22].
Sentinel lymph node biopsy (SLNB), as a less invasive technique with lower rate of lymphedema, arm pain and other complications, can be used to assess axillary lymph node status in early breast cancer patients. However, in patients treated with NAT, especially those clinical ALN positive patients, SLNB is not well established to replace ALND. The ACOSOG Z1071 study reported that when less than two sentinel lymph nodes (SLN) were detected, the false negative rate (FNR) of SLNB was 12.6%, which did not meet the 10% threshold [23]. The SN FNAC study reported that the FNR of SLNB are 13.3% in patients with positive axillary lymph nodes if SLN metastases of isolated tumor cells has been considered negative [24]. The SENTINA study also reported the FNR of SLNB after NAT was 24.3% and 18.5% in patients removing one ALN and two ALNs, respectively [25], indicating we cannot routinely do SLNB in those clinically node-positive (cN+) patients after NAT.

An accurate prediction of ALN after NAT is important to help us select a candidate cN+ patients to receive SLNB after NAT. Studies have shown that nodal pCR rates are higher in triple negative (TN) and HER2+ breast cancer patients than those in hormone receptor positive (HR+) HER2-negative (HER2-) patients (4%-18%) [14, 15, 17, 19–21, 26]. The higher rates of nodal pCR in patients with TN and HER2+ diseases make this group more likely to be omitted with ALND. However, there was no accurate model to predict pathologically node-positive disease after NAT (ypN+) in patients with TN or HER2+ diseases [27–32]. In current study, we aimed to analyze factors associated with ypN+ status and established a nomogram model based on these factors to predict the risk of ypN+ after NAT in breast cancer patients with TN or HER2+ diseases, thus to identify the candidate patients to receive the de-escalating SLNB surgery instead of ALND.

2 Materials and methods

2.1 Data resource and patient selection

We collected the data from the Shanghai Jiao Tong University Breast Cancer Database (SJTU-BCDB) between November 2008 and September 2022, which included more than 70,000 breast cancer cases from 40 medical centers in China. The inclusion criteria for retrospective series involved the following: (1) female; (2) HER2+ or TN breast cancer receiving NAT before surgery; (3) invasive breast cancer diagnosed by core needle biopsy (CNB) before NAT with complete clinicopathological information; (4) receiving ≥ 4 cycles of NAT with or without anti-HER2 targeted therapy (trastuzumab ± pertuzumab); (5) undergoing mastectomy or breast conserving surgery (BCS) plus ALND with complete histopathological data after NAT. Exclusion criteria were as follows: male patients; patients receiving neoadjuvant endocrine therapy alone; patients with distant metastases; patients who underwent tumor excision biopsy before NAT; patients diagnosed with occult breast cancer; patients with treatment regimen unavailable; patients who received < 4 cycles of NAT; and patients received SLND and with incomplete pathological information.

For the accuracy and clinical application efficiency of the established model, all patients from SJTU-BCDB were randomly divided into a training cohort and an internal validation cohort at a 7:3 ratio. Figure
1 shows the selection process and the final number of cases included in our present analysis.

The nomogram was validated externally with 108 patients at the International Peace Maternity & Child Health Hospital (IPMCH) of China Welfare Institute, Shanghai Jiaotong University School of Medicine from January 2010 to September 2022. The inclusion criteria and exclusion criteria were as same as the above.

This study was conducted in accordance with the 1964 Helsinki Declaration and its later amendments. Patients’ identity remained anonymous, and the requirement for informed consent was waived due to the retrospective nature of the study.

2.2 Neoadjuvant therapy Regimens

NAT regimens were decided by the physicians according to National Comprehensive Cancer Network breast cancer guidelines, and divided into three categories: regimens containing anthracyclines, such as CEF [cyclophosphamide (C), epirubicin (E), and fluorouracil (F)] and EC; regimens containing taxanes, such as TC [docetaxel (T)], PCb [paclitaxel (P), carboplatin (Cb)], and TCb [docetaxel (T), carboplatin (Cb)]; and regimens combining anthracyclines and taxanes, such as EC-T, TEC, dose-dense EC-weekly P, and ET [epirubicin (E), docetaxel (T)]. HER2+ breast cancer patients were given anti-HER2 targeted therapy (trastuzumab ± pertuzumab) before and after surgery according to the guidelines or physician decision.

2.3 Data collection and measures

The clinical TNM staging of breast cancer is based on the American Joint Commission on Cancer Staging Manual (8th edition, 2017). Pathological evaluation of tumors was accomplished by at least two independent experienced pathologists. ER, PR, and Ki-67 status was recorded as the percentage of cells that stained positive on immunohistochemistry (IHC). IHC staining positivity for ER and PR was defined as ≥ 1% nuclear staining. ER or PR positive were defined as hormone receptor positive (HR+). The level of Ki-67 expression was classified as high versus low with a cut-off point of 30% [33, 34]. HER2 status with a 3+ score for IHC or HER2 gene amplification confirmed by fluorescent in situ hybridization (FISH) was considered HER2 positivity. Based on HR and HER2 status, the patients in our study were divided into three molecular subtypes: TN (HR-HER2-) [35], HR-HER2+, and HR+HER2+. Other clinicopathologic factors were also extracted, including age, menopausal status, tumor grade, and histologic type.

Clinical tumor size and nodal status at diagnosis, and clinical primary tumor response to NAT were determined mainly using ultrasound by radiologists and assisted with physical examination (PE) by experienced physicians. The clinical primary tumor response was measured according to Response Evaluation Criteria in Solid Tumors (RECIST) [36]. No suspicious lymph node under PE and ultrasound or a negative cytological test result by ultrasound-guided fine-needle aspiration was defined as cN0. Pathological response of the breast and lymph node was evaluated after standard surgery. Patients without any residual invasive cancer in the breast were considered to have achieved a breast pCR [37, 38].
and no breast malignant cells in ALNs were defined as ypN0. In addition to ypN0, others were defined as ypN+.

2.4 Statistical analysis

The statistical analyses were performed using the SPSS version 26.0 software (IBM SPSS Statistics). In descriptive statistics, the continuous variables were described as median and range. The categorical variables were described as frequencies and percentages. The t-test or chi-square test were used for comparison between the training cohort, the internal and external validation cohort, and the Fisher's exact test was used when necessary. Univariate analysis was used to examine factors associated with likelihood of ypN+ in the training cohort. The significant factors were then entered into a multivariate logistic regression stepwise model, which was used to generate a nomogram to predict ypN+ in NAT patients. Results were expressed in odds ratios (OR) and 95% confidence intervals (CI). A two-side P < 0.05 was considered statistically significant.

R Foundation for Statistical Computing (Vienna, Austria) version 4.2.2 was used to establish the nomogram, the calibration curve, the receiver operating characteristic (ROC) curve, and the decision curve analysis (DCA). The training cohort was used to establish the nomogram, and the internal and external validation cohort was used for validation of the model. The predictive accuracy was evaluated via measuring the area under the ROC curve (AUC). The calibration curve was performed to reflect the relationship between the predicted incidence and the actual incidence. The DCA curve was used to evaluate the clinical application value of the nomogram model. The cut-off value of the total score in the nomogram would be determined according to the ROC curve.

3 Results

3.1 Characteristics of patients and NAT response from SJTU-BCDB

A total of 2,409 breast cancer patients from SJTU-BCDB who met the eligibility criteria were selected, of which 1,686 patients were assigned to the training set, and 723 patients in the internal validation set. There were no considerable differences in terms of clinical characteristics, treatment regimens, and NAT response between two cohorts. Table 1 illustrated the clinicopathological of the whole population. The median age was 50 (ranging from 21 to 83) years old. 2,311 (95.9%) patients were diagnosed as invasive ductal carcinoma. 556 (23.1%) patients were classified as cN0 cases. cN1, cN2, and cN3 disease were found in 1,314 (54.5%), 238 (9.9%), and 301 (12.5%) patients, respectively. The proportion of TN, HR-HER2+, and HR+HER2+ molecular subtypes was 25.8%, 20.5%, and 53.7%, respectively. Most patients (77.4%) of the study population were attributed to Ki67≥30%. Approximately one-half of the included patients received NAT regimen containing taxanes (54.7%), and 41.2% of breast cancer patients treated with NAT combining with anthracyclines and taxanes. Targeted therapy was given in 95.9% of the HER2+
breast cancer patients. After NAT, 85.2% (2053/2409) of breast cancer patients received mastectomy, the rest of patients (14.8%, 356/2409) were treated with BCS.

The total breast pCR rate after NAT was 39.5% (951/2409). A nodal pCR was observed in 1533 patients (63.6%) (P < 0.001). Patients with HR-HER2+ disease had a higher breast pCR rate (61.54%) and a higher nodal pCR rate (78.14%) than those with HR+HER2+ and TN disease (P = 0.001) (Fig.2a). In addition, patients who had a better clinical primary tumor response (CR and PR) were more likely to achieve a nodal pCR (Fig.2b). Nodal pCR rates ranged from 82.19% among patients who had a breast complete response (CR) to 24.66% among those who had a breast progression disease (PD) (P < 0.001) (Fig.2b).

3.2 Correlation between ypN+ and clinicopathological characteristics in the training cohort

Using the training cohort, univariate logistic regression analyses were used to identify significant preoperative factors associated with ypN+ (Table 2). Univariate analyses identified clinical tumor (cT) stage (P = 0.006), clinical nodal (cN) stage (P < 0.001), molecular subtype (P < 0.001), Ki67 expression (P < 0.001), tumor grade (P < 0.001), and clinical primary tumor response (P < 0.001) as independent impact factors for pathological nodal response. These factors were then entered into an adjusted multivariable regression to identify predictors of ypN+.

Multivariate analysis (Table 3) indicated that patients who had cN1 (OR: 5.031, 95% CI: 3.579-7.073, P < 0.001), cN2 (OR: 6.486, 95% CI: 4.102-10.256, P < 0.001), and cN3 (OR: 8.679, 95% CI: 5.636-13.364, P < 0.001) disease were less likely to achieve a nodal pCR compared with those who had cN0 disease. Compared with HR-HER2+ disease, TN (OR: 2.618, 95% CI: 1.826-3.752, P < 0.001) and HR+HER2+ (OR: 3.018, 95% CI: 2.175-4.187, P < 0.001) diseases were more likely to have ypN+. Lower Ki67 expression breast cancer patients (OR: 1.506, 95% CI: 1.148-1.976, P = 0.003) were less likely to achieve a nodal pCR. Compared with those whose tumor grade was unknown, patients who had I/II (OR: 2.731, 95% CI: 2.104-3.544, P < 0.001) and III tumor grade (OR: 3.051, 95% CI: 2.233-4.168, P < 0.001) were less likely to achieve a nodal pCR. Patients who had partial response (PR) (OR: 1.849, 95% CI: 1.045-3.271, P = 0.035), stable disease (SD) (OR: 3.212, 95% CI: 1.758-5.870, P < 0.001), and progression disease (PD) (OR: 4.132, 95% CI: 2.046-8.342, P < 0.001) after NAT had a higher rate of ypN+ compared with those who achieved ycT0.

3.3 Construction and internal validation of the nomogram model

A multivariable logistic regression nomogram was developed using variables including cT stage, cN stage, molecular subtype, Ki67 status, tumor grade, and clinical primary tumor response after NAT (Fig.3). The value of each variable was given a score on the “points” line. Then the total sum for each variable is
located on the “total points” line, and a line can be drawn downward to calculate the probability of ypN+. Based on the ROC analysis, the nomogram showed a robust discrimination with an AUC of 0.782 (95% CI: 0.759-0.805) (Fig.4a). The calibration curve of the training cohort showed a high degree of fit between the predicted and actual values, which indicated that the nomogram could well predict ypN+ (Fig.5a).

When the nomogram was applied to the internal validation cohort, the AUC was 0.753 (95% CI: 0.717-0.789), proving that the nomogram had good discriminatory power and provided precise predictions of ypN+ (Fig.4b). The calibration was also good for the validation cohort and indicated that the nomogram was well calibrated (Fig.5b). DCA curve revealed that the nomogram would add more net benefit both in training and internal validation cohort, which indicated that the nomogram had the better clinical predictive power and could serve as an effective diagnostic tool for ypN+ (Fig.6a and Fig.6b).

3.4 External validation of the nomogram model

The external cohort included 108 patients from IPMCH who were enrolled for the external validation of the nomogram. The selection process was shown in Fig.7. A total of 72 (66.7%) patients achieved nodal pCR, and 43 (39.8%) patients achieved breast pCR. The demographic and pathological features of the patients were summarized in Table 4. When the nomogram was applied to the external validation set, the AUC was 0.783 (95% CI: 0.692–0.873) (Fig.8), which showed that the nomogram had good discriminatory power in the external validation data sets. The calibration and DCA curve were also indicated that the nomogram had the better clinical predictive power (Fig.9 and Fig.10).

3.5 Prediction accuracy of different cutoff points

Table 5 showed that the prediction accuracies of ypN+ in our model varied according to the risk cutoff points. The sensitivity and negative predictive values decreased as the cutoff value increased, but the specificity and positive predictive values increased. When predicting the probabilities of patients who were more likely to have ypN+, the patients with false positive rates accounted for 13.2% and 8.8% of those with scores of ≥50% and ≥55%, respectively. Among patients who had a predicted probability of ypN+ ≤20% and ≤25%, the false negative rates accounted for 9.2% and 15.5%, respectively. These results demonstrated that our nomogram model can accurately predict the probability of ypN+ by combining information from routinely available clinicopathological characteristics.

4 Discussion

In our large cohort of breast cancer patients with TN or HER2+ disease from SJTU-BCDB who received NAT, the overall breast pCR rate was 39.5% and the nodal pCR rate was 63.6%. Our results suggested that cN stage, molecular subtype, Ki67 status, tumor grade, and clinical primary tumor response after NAT were independent prognostic factors of ypN+ after NAT. Combined with these independent prognostic factors and their clinical significance, we then established a nomogram model with good application
value to predict the risk of ypN+ in TN or HER2+ patients after NAT, which could help clinicians choose the de-escalating ALN surgery.

Our study found that, first, breast pCR rate and nodal pCR rate both reached high levels in TN or HER2+ breast cancer patients treated with the NAT regimens. Similar to our results, previous researches reported the breast pCR rates of 33.6%-37.5%, 30.9%-50.3% and 7.5%-16.2% in patients with TN, HER2+ and HR+ HER2- diseases [5, 39, 40], respectively, and the nodal pCR rates of 33%-49%, 42%-74% and 4%-18% in patients with TN, HER2+ and HR+ HER2- diseases [14, 15, 17, 19–21, 26], respectively. These results consistently indicated that both breast pCR rates and nodal pCR rates were significantly higher in TN or HER2+ patients than in HR+ HER2- patients. Meanwhile, nodal pCR rates were higher than breast pCR rates in general, indicating SLNB could be potential to replace ALND in certain patients. Secondly, in patients with TN or HER2+ diseases, those with clinical complete response after NAT were more likely to obtain nodal pCR, thus to possibly omit ALND in these patients. Since this group of patients was dominant in number (approximately 80%), large numbers of patients had the opportunity to benefit from an exemption from ALND. In addition, several studies indicated the probability of nodal pCR in patients with primary breast tumor progression. Guo et al. [39] reported a 14.28% rate of nodal pCR in breast PD patients, and Morgan et al. [41] showed a 9.6% nodal pCR rate in patients with breast residual disease. Comparing with these studies, we found a higher nodal pCR rate (24.7%) in breast PD patients. One reason for this was that not all patients had a preoperative lymph node biopsy, which might lead to an overestimation of lymph node stage in those with initial negative lymph nodes. Therefore, patients with breast PD were not excluded from our study, and the proportion of patients who were likely to be exempt from ALND was further expanded.

Combined with univariate and multivariate analyses, we found independent predictors of ypN+ after NAT in breast cancer patients with TN or HER2+ diseases. Most of the previous literatures used single factors to predict whether patients achieved nodal pCR [18, 42–44]. However, there were also many studies [15, 39, 45–47] consistent with our results, which believed that cT stage, cN stage, molecular subtype, Ki67 status, tumor grade, and clinical primary tumor response after NAT had predictive value for ypN+ status evaluation. Moreover, the predictive efficiency of combined multi-factor model was significantly higher than that of single factor. Based on the above impact factors, a multi-factor nomogram was established by weighting the contributions of those factors to ypN+, which could predict the risk of ypN+ accurately and visually.

Our model was established to predict the risk of ypN+ after NAT in patients with TN or HER2+ diseases. Different from models for all molecule types in most previous literatures [27–30], ours was the first model aimed at TN or HER2+ patients for reasons of the higher nodal pCR rate in those patients. In our study, the included variables of nomogram could be obtained during routine examination before NAT, which did not require additional examination and was more consistent with actual clinical operation, showing the feasibility of nomogram as a predictive tool. The AUC values of the training set, internal validation set, and external validation set were all greater than 0.75, indicating a high predictive value of our nomogram. Moreover, all sets were well-validated in calibration curves, which further suggested the high prediction
accuracy of our nomogram in different databases. The DCA was additionally applied to evaluate the model, which filled the gap that the ROC curve and calibration curve could only evaluate the accuracy but not the clinical efficiency.

The current nomogram helped clinicians weigh the risks and benefits of SLNB in TN or HER2+ patients after NAT more appropriately. According to three prospective muti-institutional trails [23–25], the investigators suggest that the FNR of SLNB after NAT less than 10% can be clinically acceptable. In our study, patients with a low predicted probability of ypN+ (≤ 20%) might have a low FNR (9.2%) within the acceptable range of 10%. Consequently, SLNB could be applied instead of ALND to those patients with low risk of ypN+. However, among patients with a predicted probability of ypN+ higher than 55%, the false-positive rate was low enough (8.8%), thus SLNB might be unsuitable.

As far as we known, our nomogram was the first model to provide reference to de-escalation of axillary management in patients with TN or HER2+ diseases. However, there were some limitations in our study. First, as a multi-center retrospective study, there were certain selection biases, such as, including patients without finishing scheduled NAT regimens, and excluding patients with missing data. Although the NAT protocols were used according to the guidelines, but they were still varied. Fortunately, these selection biases could be countered by the strength of a large sample size. Second, not all suspected ALNs were biopsied before NAT (the proportion of fine needle aspiration biopsy was about 66.1%), which could affect the outcome to some extent.

In conclusion, we developed a nomogram model with six predictive factors for predicting the risk of ypN+ after NAT in TN or HER2+ breast cancer patients. Our nomogram with robust discrimination power, calibration, and clinical decision value could guide the de-escalating ALN surgery after NAT.

**Abbreviation**

ALN: axillary lymph node

pCR: pathologic complete remission

NAT: neoadjuvant therapy

TN: triple negative

HER2: human epidermal growth factor receptor 2

HER2+: HER2-positive

ALND: axillary lymph node dissection

SJTU-BCDB: Shanghai Jiao Tong University Breast Cancer Database

IPMCH: the International Peace Maternity & Child Health Hospital
ypN+: pathologically node-positive disease after NAT

ROC: receiver operating characteristic

cN: clinical nodal stage

cT: clinical tumor stage

DCA: decision curve analysis

ER: estrogen receptor

PR: progesterone receptor

AUC: areas under ROC curve

CI: confidence interval

SLNB: sentinel lymph node biopsy

SLN: sentinel lymph node

A: anthracyclines

T: taxanes

FNR: false negative rate

cN+: clinically node-positive disease

HR+: hormone receptor positive

HER2-: HER2-negative

CNB: core needle biopsy

BCS: breast conserving surgery

CEF: cyclophosphamide, epirubicin, and fluorouracil

EC: epirubicin and cyclophosphamide

TC: docetaxel and cyclophosphamide

PCb: paclitaxel and carboplatin

TCb: docetaxel and cyclophosphamide
EC-T: epirubicin and cyclophosphamide followed by docetaxel
TEC: docetaxel, epirubicin and cyclophosphamide
EC-weekly P: epirubicin and cyclophosphamide followed by weekly paclitaxel
ET: epirubicin and docetaxel
IHC: immunohistochemistry
FISH: fluorescent in situ hybridization
HR: hormone receptor
HR-: hormone receptor negative
PE: physical examination

Declarations

Funding
This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Competing Interests
The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author Information
Co-first author: Shu-Jie Chen and Qin-Yu Zhang contributed equally to this work.
Contributions: Shu-Jie Chen and Jie Wang contributed equally to the main concept and the study design. Shu-Jie Chen, Qin-Yu Zhang performed data collection, statistical analysis, interpretation and manuscript writing. Li Yang, Min Ji and Jie Wang contributed to the editing and critical revision of the manuscript. All authors contributed to the article and approved the submitted version.
Corresponding authors: Correspondence to Min Ji, Li Yang or Jie Wang.

Acknowledgments
The authors gratefully acknowledge to all the doctors across the country who contributed patients’ data to the SJTU-BCDB and support for this study. The authors also want to express our gratitude to all the patients who participated in the study.
Data Availability Statements

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding author on reasonable request.

References


Figures
Breast cancer patients receiving NAT from November 2008 to September 2022 in the SJTU-BCDB n=3558

Excluded

Male n=7
Neoadjuvant endocrine therapy only n=46
HR+HER2- n=421
Distant metastasis n=61

Operable HER2+ or TN, female breast cancer patients receiving NAT n=3023

Excluded

Incomplete clinical information n=90
Biopsy data unavailable n=83
Previous excision biopsy n=30
Occult breast cancer n=9
Treatment regimen unavailable n=76

Patients treated with mastectomy or breast conserving surgery plus ALND after NAT with complete treatment regimen and histopathological data n=2735

Excluded

Treatment cycles <4 n=102
Patients treated with SLND n=205
Incomplete pathological information n=19

Patients receiving ≥4 cycles of NAT n=2409

7:3

Training set n=1686
Validation set n=723

Figure 1

Selection process of our defined population in SJTU-BCDB

SJTU-BCDB, Shanghai Jiao Tong University Breast Cancer Database; HR, hormone receptor; HER2, human epidermal growth factor receptor 2; TN, triple negative; ALND, axillary lymph nodes dissection; NAT, neoadjuvant therapy; SLND, sentinel lymph nodes dissection.
Figure 2

Pathological complete remission (pCR) rate stratified by molecular subtype and clinical primary tumor response

(a) Percentage of breast pCR and nodal pCR in patients stratified by molecular subtype. (b) Percentage of breast pCR and nodal pCR in patients stratified by clinical primary tumor response.

TN, triple negative; HR, hormone receptor; HER2, human epidermal growth factor receptor-2; CR, complete response; PR, partial response; SD, stable disease; PD, progression disease; pCR, pathological complete remission; ypN, pathologically node status after NAT.
Figure 3

A nomogram for predicting ypN+ in TN or HER2+ patients after NAT. Variables including cT stage, cN stage, molecular subtype, Ki67 status, tumor grade, and clinical primary tumor response were assigned points on the uppermost point scale. A total score could be easily calculated by adding each single score, and then we could estimate the probability of ypN+ by projecting the total score to the bottom scale.

cT, clinical tumor stage; cN, clinical nodal stage; HR, hormone receptor; HER2, human epidermal growth factor receptor-2; TN, triple negative; CR, complete response; PR, partial response; SD, stable disease; PD, progression disease; NAT, neoadjuvant therapy.
Figure 4

ROC plots for predicting ypN+ status for patients with TN or HER2+ breast cancer disease after NAT in (a) the training cohort and (b) the internal validation cohort. The ROC curve in the training cohort indicates an AUC of 0.782 (95% CI: 0.759-0.805), and the ROC indicates an AUC of 0.753 (95% CI: 0.717-0.789) in the internal validation cohort.

ROC, receiver operating characteristic; AUC, area under the ROC curve; ypN+, pathologically node-positive disease after NAT; TN, triple negative; HER2, human epidermal growth factor receptor-2; NAT, neoadjuvant therapy.
Figure 5

Calibration curves illustrate the observed and predicted ypN+ status for patients with TN or HER2+ breast cancer disease after NAT in (a) the training cohort and (b) the internal validation cohort. The horizontal axis indicates the predicted probabilities measured by the nomogram, and the vertical axis indicates the actual probabilities.

ypN+, pathologically node-positive disease after NAT; TN, triple negative; HER2, human epidermal growth factor receptor-2; NAT, neoadjuvant therapy.
Figure 6

DCA curves illustrate the net benefit of the nomogram for predicting ypN+ status for patients with TN or HER2+ breast cancer disease after NAT in (a) the training cohort and (b) the internal validation cohort. DCA curve revealed that the nomogram would add more net benefit both in training and internal validation cohort, which indicated that the nomogram had the better clinical predictive power and could serve as an effective diagnostic tool for ypN+

DCA, decision curve analysis; ypN+, pathologically node-positive disease after NAT; TN, triple negative; HER2, human epidermal growth factor receptor-2; NAT, neoadjuvant therapy.
Breast cancer patients receiving NAT from January 2010 to September 2022 in the International Peace Maternity & Child Health Hospital of China Welfare Institute

n=187

Excluded

Neoadjuvant endocrine therapy only n=2
HR+HER2- n=30
Distant metastasis n=4

Operable HER2+ or TN, female breast cancer patients receiving NAT

n=151

Excluded

Incomplete clinical information n=5
Biopsy data unavailable n=3
Previous excision biopsy n=6
Occult breast cancer n=1

Patients treated with mastectomy or breast conserving surgery plus ALND after NAT, with complete treatment regimen and histopathological data

n=136

Excluded

Treatment cycles <4 n=35
Patients treated with SLND n=3

Patients receiving ≥4 cycles of NAT

n=108

Figure 7

Selection process of our defined population in the International Peace Maternity & Child Health Hospital of China Welfare Institute

HR, hormone receptor; HER2, human epidermal growth factor receptor 2; TN, triple negative; ALND, axillary lymph nodes dissection; NAT, neoadjuvant therapy; SLND, sentinel lymph nodes dissection.
Figure 8

ROC plots for predicting ypN+ status for patients with TN or HER2+ breast cancer disease after NAT in the external validation cohort. The ROC curve in the external validation cohort indicates an AUC of 0.783 (95% CI: 0.692-0.873)

ROC, receiver operating characteristic; AUC, area under the ROC curve; ypN+, pathologically node-positive disease after NAT; TN, triple negative; HER2, human epidermal growth factor receptor-2; NAT, neoadjuvant therapy.
Figure 9

Calibration curves illustrate the observed and predicted ypN+ status for patients with TN or HER2+ breast cancer disease after NAT in the external validation cohort. The horizontal axis indicates the predicted probabilities measured by the nomogram, and the vertical axis indicates the actual probabilities ypN+, pathologically node-positive disease after NAT; TN, triple negative; HER2, human epidermal growth factor receptor-2; NAT, neoadjuvant therapy.
Figure 10

DCA curves illustrate the net benefit of the nomogram for predicting ypN+ status for patients with TN or HER2+ breast cancer disease after NAT in the external validation cohort. DCA curve revealed that the nomogram would add more net benefit, which indicated that the nomogram had the better clinical predictive power and could serve as an effective diagnostic tool for ypN+

DCA, decision curve analysis; ypN+, pathologically node-positive disease after NAT; TN, triple negative; HER2, human epidermal growth factor receptor-2; NAT, neoadjuvant therapy.