Validation and update of a lymph node metastasis prediction model for breast cancer

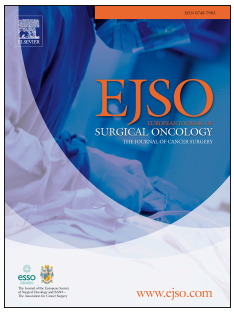
Qiu, Si-Qi; Aarnink, Merel; van Maaren, Marissa C.; Dorrius, Monique D.; Bhattacharya, Arkajyoti; Veltman, Jeroen; Klazen, Caroline A. H.; Korte, Jan H.; Estourgie, Susanne H.; Ott, Pieter

Published in:
EJSO

DOI:
10.1016/j.ejso.2017.12.008

Publication date:
2018

Accepted Manuscript



Validation and update of a lymph node metastasis prediction model for breast cancer

Si-Qi Qiu, Merel Aarnink, Marissa C. van Maaren, Monique D. Dorrius, Arkajyoti Bhattacharya, Jeroen Veltman, Caroline A.H. Klazen, Jan H. Korte, Susanne H. Estourgie, Pieter Ott, Wendy Kelder, Huan-Cheng Zeng, Hendrik Koffijberg, Guo-Jun Zhang, Gooitzen M. van Dam, Sabine Siesling

PII: S0748-7983(18)30037-4

DOI: 10.1016/j.ejso.2017.12.008

Reference: YEJSO 4822

To appear in: *European Journal of Surgical Oncology*

Received Date: 3 November 2017

Revised Date: 30 November 2017

Accepted Date: 21 December 2017

Please cite this article as: Qiu S-Q, Aarnink M, van Maaren MC, Dorrius MD, Bhattacharya A, Veltman J, Klazen CAH, Korte JH, Estourgie SH, Ott P, Kelder W, Zeng H-C, Koffijberg H, Zhang G-J, van Dam GM, Siesling S, Validation and update of a lymph node metastasis prediction model for breast cancer, *European Journal of Surgical Oncology* (2018), doi: 10.1016/j.ejso.2017.12.008.

Validation and update of a lymph node metastasis prediction model for breast cancer

Si-Qi Qiu[a,e]*, Merel Aarnink[f]*, Marissa C. van Maaren[g], Monique D. Dorrius[c], Arkajyoti Bhattacharya[d], Jeroen Veltman[h], Caroline A.H. Klazen[i], Jan H. Korte[j], Susanne H. Estourgie[k], Pieter Ott[l], Wendy Kelder[m], Huan-Cheng Zeng[e], Hendrik Koffijberg[f], Guo-Jun Zhang[n], Gooitzen M. van Dam[a,b], Sabine Siesling[f,g],

[a]Department of Surgery, [b]Department of Nuclear Medicine and Molecular Imaging & Intensive Care, [c]Department of Radiology, [d]Department of Medical Oncology, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands

[e]The Breast Center, Cancer Hospital of Shantou University Medical College, Guangdong, China

[f]Department of Health Technology and Services Research, MIRA Institute for Biomedical Technology and Technical Medicine, University of Twente, Enschede, The Netherlands

[g]Department of Research, Netherlands Comprehensive Cancer Organisation, Utrecht, the Netherlands

[h]Department of Radiology, ZiekenhuisgroepTwente, Almelo, The Netherlands

[i]Department of Radiology, Medisch Spectrum Twente, Enschede, The Netherlands

[j]Department of Radiology, Isala, Zwolle, The Netherlands

[k]Department of Surgery, Medisch Centrum Leeuwarden, Friesland, The Netherlands

[l]Department of Radiology, [m]Department of Surgery, Martini Hospital, Groningen, The Netherlands

[n]Changjiang Scholar's Laboratory of Shantou University Medical College, Guangdong, China

* Both authors contributed equally to this work.

Corresponding author: Sabine Siesling, PhD, Department of Health Technology and Services Research, MIRA Institute for Biomedical Technology and Technical Medicine, University of Twente, Enschede, the Netherlands; Department of Research, Netherlands Comprehensive Cancer Organisation, 3501 DB, Utrecht, the Netherlands (Tel: +31-6-13193806; e-mail address: s.siesling@iknl.nl).

Abstract

Purpose: This study aimed to validate and update a model for predicting the risk of axillary lymph node (ALN) metastasis for assisting clinical decision-making.

Methods: We included breast cancer patients diagnosed at six Dutch hospitals between 2011 and 2015 to validate the original model which includes six variables: clinical tumor size, tumor grade, estrogen receptor status, lymph node longest axis, cortical thickness and hilum status as detected by ultrasonography. Subsequently, we updated the original model using generalized linear model (GLM) tree analysis and by adjusting its intercept and slope. The area under the receiver operator characteristic curve (AUC) and calibration curve were used to assess the original and updated models. Clinical usefulness of the model was evaluated by false-negative rates (FNRs) at different cut-off points for the predictive probability.

Results: Data from 1,416 patients were analyzed. The AUC for the original model was 0.774. Patients were classified into four risk groups by GLM analysis, for which four updated models were created. The AUC for the updated models was 0.812. The calibration curves showed that the updated model predictions were better in agreement with actual observations than the original model predictions. FNRs of the updated models were lower than the preset 10% at all cut-off points when the predictive probability was less than 12.0%.

Conclusions: The original model showed good performance in the Dutch validation population. The updated models resulted in more accurate ALN metastasis prediction and could be useful preoperative tools in selecting low-risk patients for omission of axillary surgery.

Keywords: breast cancer; axillary lymph node metastasis; model; prediction model; axillary surgery omission

Introduction

Axillary lymph node (ALN) status is an important prognostic factor and a major determinant for postoperative treatment decision-making for breast cancer patients[1,2]. ALN staging evolved together with the shift in surgical treatment from the largest tolerable surgery to less invasive surgery. During this process, sentinel lymph node biopsy (SLNB) replaced ALN dissection and has become the standard of care for ALN staging in breast cancer patients with clinically negative ALN for over 10 years. SLNB has significantly reduced the incidence of surgical complications such as upper limb lymphedema and impaired shoulder function, and has improved patients' quality of life without compromising their survival[3–7]. However, the surgical complications from SLNB cannot be ignored. Lymphedema occurs in approximately 5-8% of patients receiving a SLNB and paresthesia in 10-15%[6–12]. In addition, 28-49% of the patients experience shoulder-arm function impairment[11–13]. Notably, 60-70% of the patients receiving a SLNB are shown to have negative SLNs after histopathological analysis and thus do not benefit from the procedure[4,14,15].

Due to early detection through the national screening program, more patients are being diagnosed with early breast cancer and are more often free from ALN metastasis[16]. If patients with a pathologically negative ALN can be preoperatively predicted, omission of axillary surgery could avoid the above-mentioned surgical complications and improve their quality of life, without affecting the postoperative treatment decision-making. Consequently, accurate assessment of the preoperative patients' risk of ALN metastasis is required. However, all currently used imaging modalities have low sensitivity in predicting ALN metastasis, resulting in a false-negative prediction of around 40-70%[17–20]. Therefore, new tools for prediction of preoperative ALN metastasis are urgently needed.

We previously developed a predictive model for ALN metastasis in a Chinese breast cancer population based on clinicopathological features from the primary tumor and axillary ultrasound[21]. The model was based on six independent predictors for ALN metastasis: clinical tumor size, histological tumor grade, estrogen receptor (ER) status, longest axis, cortical thickness and hilum status of the ALN as detected by ultrasonography. The model was validated on an additional set of 234 Chinese patients, generating an area under the receiver operating characteristic curve (AUC) of 0.864[21], indicating a good performance in ALN metastasis prediction.

In this study we validated the performance of the Chinese model for predicting ALN metastasis in a large Dutch breast cancer population. The model was updated using the Dutch and Chinese patient

data in order to improve its discriminative performance and predictive accuracy and maintain its generalizability in different ethnic groups.

Methods

Patients

We selected all women with primary breast cancer who underwent breast surgery and axillary staging at six participating Dutch hospitals (one university hospital, two teaching and referral hospitals and three general hospitals) between 2011 and 2015 from the Netherlands Cancer Registry (NCR). The NCR records data on all cancer patients in the Netherlands. The inclusion and exclusion criteria were the same with that used in the initial study which developed the model[21]. Patients with one or more ALN(s) detected by a preoperative ultrasound and receiving either a SLNB or ALN dissection were included in this study, irrespective whether the ALNs were palpable or not. Exclusion criteria were use of primary systemic therapy, ductal carcinoma *in situ* and bilateral breast cancer. Patients with lymph node cortical thickness larger than 2.3mm as measured by ultrasound received a final needle aspiration cytology (FNAC) according to Dutch guideline[22]. In addition to the Dutch patients, the Chinese patients (*n*=322) diagnosed at Cancer Hospital of Shantou University Medical College between 2009 and 2014 for developing the model were also used in the present study for updating the original model[21]. This study was approved by all participating hospitals.

Data collection

From the NCR we collected data on age at diagnosis, menopausal status, tumor location, histological grade, ER, progesterone receptor (PgR), human epidermal growth factor receptor 2 (HER2) and pathological ALN status. Data on histological grade, ER, PgR and HER2 status were obtained from surgical resection specimens. Data on clinical tumor size of the primary tumor, and longest axis, cortical thickness and hilum status of the ALN were obtained from measurements by high frequency ultrasound (>10MHz). The ultrasound reports and images were initially checked by one author (M.A.) after receiving training from a radiologist (M.D.D.). In case of uncertainty the image was reviewed by a radiologist of the participating hospital. All researchers involved in the data collection were blinded to the pathological ALN status of the patients.

Staging was coded according to the Tumor, Node and Metastasis (TNM) classification[23]. Tumor grade was scored according to the Nottingham grading system[24]. ER, PgR and HER2 status were

categorized as described previously[21]. Both ER and PgR status were divided into four categories: - (<10%), + (10-25%), ++ (25-75%) and +++ (>75%)[21]. Since the NCR registered the ER and PgR status in 10% steps (0, 10%, 20%, etc.), we replaced the second cut-off point (25%) with 30% and the third (75%) with 80%. Variables related to ALNs were measured on the most suspicious lymph node detected by ultrasonography, which was defined as the lymph node with the thickest cortex and/or absence of a hilum. A lymph node was defined as positive if macrometastases, micrometastases or isolated tumor cells (ITCs) were identified by histopathological analysis[21].

Statistical analysis

Differences of categorical and continuous variables between groups were analyzed using the Chi-square and Mann-Whitney U test, respectively. The clinical value for each predictor in the original model was used to calculate the ALN metastasis probability for each patient in the present study. The AUC was used to evaluate the discriminative performance of the model. For the calibration of the original model, the enrolled Dutch patients were sorted on their predicted probability and grouped into three equally sized groups. For each group, the mean model-predicted probability and the actual percentage (95% confidence interval [CI]) of ALN metastasis were calculated. A calibration plot was drawn showing the mean model-predicted probability against the actual percentage of ALN metastasis, providing information about the predictive accuracy of the model for each group.

To improve its discriminative performance and predictive accuracy, we updated the original model using both the Dutch and the Chinese populations as follows. Generalized linear model (GLM) tree analysis was applied to the Dutch and Chinese patient populations to classify patients into groups with different risk of ALN metastasis based on their clinicopathological characteristics. In each group, the model was updated separately by adjusting the intercept and slope of the original model. Detailed information on GLM tree analysis is provided in the Supplementary Methods. The AUC and calibration plot were used to assess the discriminative performance and predictive accuracy of the updated models. False-negative rates (FNRs) of the updated models at several cut-off points for the predicted probability were calculated to assess their clinical usefulness. FNR was considered acceptable if it was lower than 10%. All statistical tests were two-sided, and a *P*-value <0.05 was considered to be statistically significant. Statistical analyses were performed using the statistical software SPSS, version 19 and R, version 3.3.2.

Results

Patient characteristics

In total, 1,416 out of 2,227 Dutch breast cancer patients were included (Fig.1). Based on Dutch guideline, 468 patients would have received a FNAC before they received any axillary surgery[22]. The comparison of clinicopathological features between the Dutch and Chinese patients is shown in Table 1. All features, except for longest axis of ALN, differed significantly between the two patient populations (P<0.05). The median age of the Dutch patients was 11 years older than the Chinese patients. Dutch patients had a smaller median tumor size and cortical thickness of ALN compared to Chinese patients. A larger proportion of Dutch patients had ER and PgR positive, HER2 negative and low histologically graded tumors. In addition, a smaller proportion of the Dutch patients lacked a hilum compared to the Chinese patients. ALN metastasis was identified in 354 (25.0%) of the Dutch patients. Of these patients, 318 (89.8%) patients had macrometastases and 36 (10.2%) patients had micrometastases or ITCs. Comparison of patients' characteristics among the six Dutch hospitals is shown in Table S1.

Discrimination performance and predictive accuracy of the original Chinese model

The AUC of the original model was 0.774 (95% CI 0.743-0.804) when it was applied to the entire Dutch population (Fig. 2a), whereas the AUCs for each hospital ranged from 0.705 to 0.848 (Fig. S1). The mean model-predicted probability and the actual percentage of patients with ALN metastasis for each group are shown in Table S2. In all groups, except for group 1 (ALN metastasis low-risk group) including patients with a mean model-predicted probability lower than 10.0%, the mean model-predicted probabilities were within the 95% CI of the actual percentage of patients with ALN metastasis. In Fig. 2b, the calibration plot showed less optimal agreement between the model prediction and the actual observation, especially for group 1.

Update of the original model

GLM tree analysis identified two variables (cortical thickness of ALN and clinical tumor size) for partitioning patients according to their ALN status. The identified cut-off points maximized the separation of risk-specific ALN metastasis. The patients were classified into four groups with increasing rates of ALN metastasis. Patients with cortical thickness ≤3.1mm and clinical tumor size ≤0.9cm had the lowest rate (4.1%) of ALN metastasis, while patients with cortical thickness >3.1mm

had the highest rate (70.2%) of ALN metastasis (Fig. 3). The updated models for each group are shown in Table S3.

Discrimination performance and predictive accuracy of the updated models

Based on the GLM tree analysis, the Dutch validation patients were first classified into different groups according to their ALN cortical thickness and clinical tumor size of the primary tumor. Subsequently, their predicted probabilities of ALN metastasis were calculated using the corresponding updated model for each group (Table S3). The AUC of the updated models for the entire Dutch patient population was 0.812 (95% CI 0.784-0.840) (Fig. 2c). The AUCs of the updated models for each hospital ranged from 0.730 to 0.874, which were all higher than the corresponding AUCs of the original model (Fig. S1 and S2). The mean model-predicted probability and the actual percentage of patients with ALN metastasis for each group are shown in Table S4. In all groups, the model-predicted probabilities were within the 95% CI of the actual percentage of patients with ALN metastasis, showing better predictive accuracy compared with the original model, especially for group 1 (Table S2 and S4). The calibration plot for the updated models showed almost perfect agreement between the model prediction and the actual observation (Fig. 2d).

Clinical usefulness of the updated models

FNRs of the updated models at several model-predicted probability cut-off points are shown in Table 2. By using cut-off points, we considered patients with a model-predicted probability less than or equal to the cut-off point to have negative ALNs. Since the generally accepted FNR for SLNB was 10.0%, we considered this percentage acceptable in our study. Using this criterion, the FNR of the updated models (7.9%) was acceptable when the model-predicted probability cut-off point was set at 12.0%. Using the updated models, 415 patients (29.3% of entire study population) could have been selected for axillary surgery omission. When combined with FNAC for those patients with lymph node cortical thickness larger than 2.3mm, the FNR of our model could have been reduced to 6.2%.

Discussion

In this study we validated and updated a previously developed model[21], providing a reliable tool for preoperative assessment of the risk of ALN metastasis. This tool could assist clinicians in determining the optimal treatment strategy of axillary staging for individual patients.

In this study, the Dutch patients presented more favorable prognostic clinicopathological features

compared with the Chinese patients. This disparity could partly be due to the different ethnic backgrounds of the two study populations[25]. In addition, the routinely performed nationwide screening programme in the Netherlands (which is absent in China) may be another contribution to this difference, since this leads to an increasing number of early diagnoses[26]. Despite the differences in clinicopathological features between the two study populations, the original model already showed a good discriminative performance with an AUC of 0.774. This indicates a good generalizability of the model to patients of different ethnic origins. However, the original model underestimated the ALN metastasis probability in the low-risk Dutch patient group with an ALN metastatic rate of 10.0%; a similar underestimation was found in the Chinese study[21]. Moreover, these results may indicate that different models for different ALN metastasis risk groups are required to improve the prediction. In order to refine the predictive accuracy of the model in several risk-based subgroups and maintain generalizability to different ethnic groups, we performed GLM tree analysis using data of both the Dutch and Chinese patients. After creating different risk groups, the discriminative performance and predictive accuracy of the updated models improved in the Dutch patients, especially for low-risk patients with a mean model-predicted probability lower than 10.0%. In a sensitivity analysis, the updated models performed well in the Chinese patients (*n*=322) as well (AUC=0.851 [95%CI, 0.808-0.894], data not shown). These results emphasize the stability of the model in ALN metastasis prediction.

SLNB remains the standard of care for axillary staging in patients with clinically negative nodes. In order to avoid its complications, omission of axillary surgery in ALN metastasis low-risk patients has gained more interest. Several ongoing clinical trials are investigating this topic[27]. Most of these trials enrol patients based on a preoperative ultrasound with or without FNAC. However, the FNR of ALN metastasis prediction by a preoperative ultrasound has been reported to be as high as around 30-55%[17–19,28]. Other imaging modalities e.g. magnetic resonance imaging (MRI) or positron emission tomography-computed tomography (PET-CT) have also been reported to predict the risk of ALN metastasis with a high FNR (MRI 18% vs PET-CT 36%)[29]. Moreover, these imaging modalities are too expensive to be applied for every patient.

Two models predicting the risk of ALN metastasis have been described in literature. The Memorial Sloan Kettering Cancer Center (MSKCC) model, in which 3,786 patients receiving SLNB were included for model development, was published in 2007. This model included eight variables and had

an AUC of 0.754 in ALN metastasis prediction when applied to a validation cohort with 1,545 patients[30]. Recently, another predictive model was reported for patients with negative axillary ultrasound. Comparable to the MSKCC model and our model, this model demonstrated good discrimination with AUC of 0.731 and 0.79 when validated internally and externally, respectively[31]. A disadvantage of these two models is that data for some variables, e.g. lymphovascular invasion, cannot be obtained preoperatively. This will hamper their clinical application as a preoperative predictive model.

To the best of our knowledge, this is the first model that predicts the risk of ALN metastasis in patients with ALNs detected by a preoperative ultrasound. For most of these patients, it is very difficult to judge whether the detected lymph nodes are malignant or not, merely based on their ultrasound morphology. Clinically, a FNAC is performed when suspicious lymph nodes are found. The FNR of ALN metastasis prediction by FNAC is higher than 20%[32]. In contrast, our updated models had a lower FNR (7.9%) in predicting the ALN status when the cut-off point was set at 12.0%. More importantly, this FNR is even lower when combined with a FNAC. These results suggest that axillary surgery omission could be possible for patients with a predicted probability lower than 12.0% by using the updated models, especially when combined with FNAC; this group accounts for 29.3% of the entire patient cohort in this study. These findings underscore the potential clinical value of the updated models in assisting clinical decision making for the selection of ALN metastasis low-risk patients who could be spared axillary surgery. For these low-risk patients, we would still suggest to perform a FNAC according to clinical standard of care management.

The International Breast Cancer Study Group (IBCSG) 23-01 trial was designed to determine whether omission of ALND was warranted in patients with one or more micrometastatic (≤2mm) SLNs and a primary tumor size ≤5cm. Results of this trial support the omission of ALND in patients with micrometastases in SLNs[33]. However, the risk of leaving micrometastases or ITCs in patients without removing any ALNs is unclear. Therefore, in this study, we considered ALNs with micrometastases or ITCs as positive, similar to our previous study[21]. Only 36 patients had ALNs with micrometastases or ITCs. Therefore, we expected it not to have influenced the predictive and discriminative performance of our updated models in predicting ALN macrometastasis (>2mm). This was confirmed by results of a sensitivity analysis where we included patients with ALN macrometastasis only, resulting in an AUC of 0.825 (95%CI 0.797-0.854) (data not shown).

This study has several strengths. First, it is a multicenter validation study. The six participating hospitals are from different areas of the Netherlands. Therefore, the enrolled patients in our study are representative for the whole breast cancer population in the Netherlands. Our updated models show a very good discriminative performance in all participating hospitals, indicating good generalizability of the models. Second, the necessary data for all of the six variables incorporated in our models can be obtained preoperatively, for example, by a core needle biopsy of the primary tumor and axillary ultrasound examination. This would facilitate the application of our models. Third, our updated models are easy to use. After all the necessary data for the six variables are obtained, patients can be grouped into one of the different risk groups based on their cortical thickness of the lymph node detected by ultrasound and the clinical tumor size, as shown in Fig.3. Subsequently, the probability of ALN metastasis for an individual patient can be calculated using the corresponding formula demonstrated in Table S3. A user-friendly web-based calculator based on our updated models has been developed and will be published online to facilitate the use of our models (https://www.evidencio.com/models/show/999).

There are limitations of this study. First, the ER status and grade of the primary tumor were obtained from surgical specimens and not from core needle biopsies in our study. Since breast cancer is highly heterogeneous, discrepancies between the ER status and tumor grade of core needle biopsies and the whole surgical specimen may arise. Several studies have demonstrated that the discrepancy of ER positive rates analysis on core needle biopsies and subsequent excision specimen from the same patient was only 2-3%[34,35]. We therefore do not expect that this influences the model validity. However, discrepancy of tumor grade between core needle biopsies and excision specimen was reported to be near 30% in a recent meta-analysis[36]. It may be that this discrepancy has influence on the performance of the model. However, it is not clear whether this will also influence the FNR of the models in predicting ALN metastasis low-risk patients. Second, the FNR (7.9%), which we considered might be suitable for selecting patients to avoid axillary surgery, was calculated on the basis of FNR of SLNB from 75.0% of patients received SLNB only. This means that the real FNR of our model is higher than 7.9% and theoretically lower than 17.9%. Given the improvement of systemic treatments for breast cancer, the impact of residual metastatic disease in ALNs on survival of patients has become less important. This has been proved by results from IBCSG 23-01 and Z0011 trials, which demonstrated 13.0% and 27.3% of patients had metastatic non-SLN being left behind in the non-

ALND group without affecting patients' long term recurrence and survival[33,37]. Third, the axillary ultrasound is an operator-dependent technique. Measurement of the three ultrasound variables incorporated in our models may differ when performed by different doctors due to inter-observer variability. However, an experienced radiologist can almost always identify ALNs and access their morphology if they are present[28,38,39]. Therefore, we recommend an experienced radiologist to assess the ALN when using our models. Finally, this is a study based on retrospectively collected data from a cancer registry. Harmonization of data measurement was less optimal. Nevertheless, data was uniformly collected by trained data managers according to strict coding manuals. Together, a validation of our models using core needle biopsies in each center is recommended before using the models in clinical practice to assist decision-making on surgical treatment of the axilla.

Conclusions

We successfully validated and updated our previously published model in a multicenter Dutch breast cancer population. The original model showed good performance in the Dutch population. The updated models resulted in more accurate ALN metastasis predictions and could therefore be useful preoperative tools in selecting low-risk patients for axillary surgery omission.

Acknowledgements

We would like to thank Dr. Rick G. Pleijhuis for kindly developing a user-friendly web-based calculator based on our updated models to facilitate the use of our models. G.J.Z. was funded by Natural Science Foundation Committee (No. 81302331), Major International Collaborative Research Project of NSFC (81320108015), and Guangdong Provincial Key Laboratory on Breast Cancer Diagnosis and Treatment Research.

The study funders did not participate in the design of the study; the collection, analysis, or interpretation of the data; the writing of the manuscript; or the decision to manuscript submission.

Conflict of interest statement

The authors declare no conflict of interest.

References

1. Banerjee M., Gworge J., Song EY., Roy A., Hryniuk W. Tree-based model for breast cancer prognostication. *J Clin Oncol* 2004;**22**(13):2567–75. Doi: 10.1200/JCO.2004.11.141.

2. Soerjomataram I., Louwman MWJ., Ribot JG., Roukema JA., Coebergh JWW. An overview of prognostic factors for long-term survivors of breast cancer. *Breast Cancer Res Treat* 2008;**107**(3):309–30. Doi: 10.1007/s10549-007-9556-1.

3. Veronesi U., Paganelli G., Viale G., et al. Sentinel-lymph-node biopsy as a staging procedure in breast cancer: update of a randomised controlled study. *Lancet Oncol* 2006;**7**(12):983–90. Doi: 10.1016/S1470-2045(06)70947-0.

4. Veronesi U., Paganelli G., Viale G., et al. A randomized comparison of sentinel-node biopsy with routine axillary dissection in breast cancer. *N Engl J Med* 2003;**349**(6):546–53. Doi: 10.1056/NEJMoa012782.

5. Krag DN., Anderson SJ., Julian TB., 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. *Lancet Oncol* 2010;**11**(10):927–33. Doi: 10.1016/S1470-2045(10)70207-2.

6. Ashikaga T., Krag DN., Land SR., et al. Morbidity results from the NSABP B-32 trial comparing sentinel lymph node dissection versus axillary dissection. *J Surg Oncol* 2010;**102**(2):111–8. Doi: 10.1002/jso.21535.

7. Mansel RE., Fallowfield L., Kissin M., et al. Randomized multicenter trial of sentinel node biopsy versus standard axillary treatment in operable breast cancer: The ALMANAC trial. *J Natl Cancer Inst* 2006;**98**(9):599–609. Doi: 10.1093/jnci/djj158.

8. Li CZ., Zhang P., Li RW., Wu CT., Zhang XP., Zhu HC. Axillary lymph node dissection versus sentinel lymph node biopsy alone for early breast cancer with sentinel node metastasis : A meta-analysis. *Eur J Surg Oncol* 2015;**41**(8):958–66. Doi: 10.1016/j.ejso.2015.05.007.

9. Gebruers N., Verbelen H., De Vrieze T., Coeck D., Tjalma W. Incidence and time path of lymphedema in sentinel node negative breast cancer patients: a systematic review. *Arch Phys*

Med Rehabil 2015;**96**(6):1131–9. Doi: 10.1016/j.apmr.2015.01.014.

10. DiSipio T., Rye S., Newman B., Hayes S. Incidence of unilateral arm lymphoedema after breast cancer: a systematic review and meta-analysis. *Lancet Oncol* 2013;**14**(6):500–15. Doi: 10.1016/S1470-2045(13)70076-7.

11. De Groef A., Van Kampen M., Tieto E., et al. Arm lymphoedema and upper limb impairments in sentinel node-negative breast cancer patients: A one year follow up study. *The Breast* 2016;**29**:102–8. Doi: 10.1016/j.breast.2016.07.021.

12. Crane-Okada R., Wascher RA., Elashoff D., Giuliano AE. Long-term morbidity of sentinel node biopsy versus complete axillary dissection for unilateral breast cancer. *Ann Surg Oncol* 2008;**15**(7):1996–2005. Doi: 10.1245/s10434-008-9909-y.

13. Kootstra JJ., Dijkstra PU., Rietman H., et al. A longitudinal study of shoulder and arm morbidity in breast cancer survivors 7 years after sentinel lymph node biopsy or axillary lymph node dissection. *Breast Cancer Res Treat* 2013;**139**(1):125–34. Doi: 10.1007/s10549-013-2509-y.

14. Krag DN., Anderson SJ., Julian TB., et al. Technical outcomes of sentinel-lymph-node resection and conventional axillary-lymph-node dissection in patients with clinically node-negative breast cancer: results from the NSABP B-32 randomised phase III trial. *Lancet Oncol* 2007;**8**(10):881–8. Doi: 10.1016/S1470-2045(07)70278-4.

15. Gill G. Sentinel-lymph-node-based management or routine axillary clearance? One-year outcomes of sentinel node biopsy versus axillary clearance (SNAC): a randomized controlled surgical trial. *Ann Surg Oncol* 2009;**16**(2):266–75. Doi: 10.1245/s10434-008-0229-z.

16. Fracheboud J., Otto SJ., van Dijck JAAM., Broeders MJM., Verbeek ALM., de Koning HJ. Decreased rates of advanced breast cancer due to mammography screening in The Netherlands. *Br J Cancer* 2004;**91**(5):861–7. Doi: 10.1038/sj.bjc.6602075.

17. Lee B., Lim AK., Krell J., et al. The efficacy of axillary ultrasound in the detection of nodal metastasis in breast cancer. *Am J Roentgenol* 2013;**200**(3):314–20. Doi: 10.2214/AJR.12.9032.

18. Hyun SJ., Kim EK., Yoon JH., Moon HJ., Kim MJ. Adding MRI to ultrasound and ultrasound-guided fine-needle aspiration reduces the false-negative rate of axillary lymph node metastasis

diagnosis in breast cancer patients. *Clin Radiol* 2015;**70**(7):716–22. Doi: 10.1016/j.crad.2015.03.004.

19. Valente SA., Levine GM., Silverstein MJ., et al. Accuracy of predicting axillary lymph node positivity by physical examination, mammography, ultrasonography, and magnetic resonance imaging. *Ann Surg Oncol* 2012;**19**(6):1825–30. Doi: 10.1245/s10434-011-2200-7.

20. CHo J. Predicting axillary lymph node metastasis in breast cancer with multimodality imaging. *Doi.org* n.d. Doi: 10.1594/ecr2017/b-0926.

21. Qiu S-Q., Zeng H-C., Zhang F., et al. A nomogram to predict the probability of axillary lymph node metastasis in early breast cancer patients with positive axillary ultrasound. *Sci Rep* 2016;**6**(October 2015):21196. Doi: 10.1038/srep21196.

22. NABON. Breast cancer - Dutch Guideline, version 2.0. 2012:57. http://www.oncoline.nl/uploaded/docs/mammacarcinoom/Dutch%20Breast%20Cancer%20Guideline%202012.pdf (accessed 16 November 2017).

23. Sobin L., Gospodarowicz M. *TNM classification of malignant tumors.* 7th ed. NJ: John Wiley&Sons; 2009.

24. Elston CW., Ellis IO. Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. *Histopathology* 1991;**19**(5):403–10. Doi: 10.1111/j.1365-2559.1991.tb00229.x.

25. Yi M., Liu P., Li X., et al. Comparative analysis of clinicopathologic features, treatment, and survival of Asian women with a breast cancer diagnosis residing in the United States. *Cancer* 2012;**118**(17):4117–25. Doi: 10.1002/cncr.27399.

26. Torre LA., Bray F., Siegel RL., Ferlay J., Lortet-Tieulent J., Jemal A. Global cancer statistics , 2012. *CA Cancer J Clin* 2015;**65**:87–108. Doi: 10.3322/caac.21262.

27. Qiu S-Q., Zhang G-J., Jansen L., et al. Evolution in sentinel lymph node biopsy in breast cancer. *Crit Rev Oncol Hematol* 2017. Doi: 10.1016/j.critrevonc.2017.09.010.

28. Stachs A., Thi AT-H., Dieterich M., et al. Assessment of ultrasound features predicting axillary

nodal metastasis in breast cancer: the impact of cortical thickness. *Ultrasound Int Open* 2015;**1**(1):E19-24. Doi: 10.1055/s-0035-1555872.

29. Liang X., Yu J., Wen B., Xie J., Cai Q., Yang Q. MRI and FDG-PET/CT based assessment of axillary lymph node metastasis in early breast cancer: a meta-analysis. *Clin Radiol* 2017;**72**(4):295–301. Doi: 10.1016/j.crad.2016.12.001.

30. Bevilacqua JLB., Kattan MW., Fey J V., Cody HS., Borgen PI., Van Zee KJ. Doctor, what are my chances of having a positive sentinel node? A validated nomogram for risk estimation. *J Clin Oncol* 2007;**25**(24):3670–9. Doi: 10.1200/JCO.2006.08.8013.

31. Meretoja TJ., Heikkilä PS., Mansfield a S., et al. A predictive tool to estimate the risk of axillary metastases in breast cancer patients with negative axillary ultrasound. *Ann Surg Oncol* 2014;**21**(7):2229–36. Doi: 10.1245/s10434-014-3617-6.

32. Houssami N., Ciatto S., Turner RM., Cody HS., Macaskill P. Preoperative ultrasound-guided needle biopsy of axillary nodes in invasive breast cancer. *Ann Surg* 2011;**254**(2):243–51. Doi: 10.1097/SLA.0b013e31821f1564.

33. Galimberti V., Cole BF., Zurrida S., et al. Axillary dissection versus no axillary dissection in patients with sentinel-node micrometastases (IBCSG 23-01): A phase 3 randomised controlled trial. *Lancet Oncol* 2013;**14**(4):297–305. Doi: 10.1016/S1470-2045(13)70035-4.

34. Arnedos M., Nerurkar A., Osin P., A'Hern R., Smith IE., Dowsett M. Discordance between core needle biopsy (CNB) and excisional biopsy (EB) for estrogen receptor (ER), progesterone receptor (PgR) and HER2 status in early breast cancer (EBC). *Ann Oncol* 2009;**20**(12):1948–52. Doi: 10.1093/annonc/mdp234.

35. Chen X., Yuan Y., Gu Z., Shen K. Accuracy of estrogen receptor, progesterone receptor, and HER2 status between core needle and open excision biopsy in breast cancer: A meta-analysis. *Breast Cancer Res Treat* 2012;**134**(3):957–67. Doi: 10.1007/s10549-012-1990-z.

36. Knuttel FM., Menezes GLG., Van Diest PJ., Witkamp AJ., Van Den Bosch MAAJ., Verkooijen HM. Meta-analysis of the concordance of histological grade of breast cancer between core needle biopsy and surgical excision specimen. *Br J Surg* 2016;**103**(6):644–55. Doi:

10.1002/bjs.10128.

37. Giuliano AE., Ballman K V., McCall L., et al. Effect of axillary dissection vs no axillary dissection on 10-Year overall survival among women with invasive breast cancer and sentinel node metastasis: the ACOSOG Z0011 (Alliance) randomized clinical trial. *JAMA* 2017;**318**(10):918–26. Doi: 10.1001/jama.2017.11470.

38. Pinheiro DJP da C., Elias S., Nazário ACP. Axillary lymph nodes in breast cancer patients: sonographic evaluation. *Radiol Bras* 2014;**47**(4):240–4. Doi: 10.1590/0100-3984.2013.1689.

39. Yang WT., Ahuja A., Tang A. Ultrasonographic demonstration of normal axillary lymph nodes: a learning curve. *J Ultrasound Med* 1995;**14**(11):823–7.

Table 1. Comparison between the Dutch and Chinese patients by clinicopathological characteristics

Characteristics	Dutch patients. No. (%)	Chinese patients. No. (%)	*P*-value
Total patients No.	1,416 (100)	322 (100)	-
Age at diagnosis (year)	.		<0.001
Median (IQR)	61 (52, 69)	50 (43, 57)	
Menopausal status			<0.001
Premenopausal	312 (22.0)	182 (56.5)	
Postmenopausal	1,104 (78.0)	140 (43.5)	
Tumor size (cm)			<0.001
Median (IQR)	1.5 (1.0, 2.2)	3.0 (2.2, 4.0)	
Clinical tumor size by T-class			<0.001
T1a	47 (3.3)	2 (0.6)	
T1b	330 (23.3)	9 (2.8)	
T1c	641 (45.3)	63 (19.6)	
T2 (≤ 3cm)	294 (20.8)	107 (33.2)	
T2 (> 3cm, ≤ 5cm)	101 (7.1)	116 (36.0)	
T3	3 (0.2)	22 (6.8)	
Unknown	0 (0.0)	3 (0.9)	
Tumor location			<0.001
UOQ	573 (40.5)	152 (47.2)	
LOQ	118 (8.3)	42 (13.0)	
UIQ	178 (12.6)	51 (15.8)	
LIQ	127 (9.0)	15 (4.7)	
Central	102 (7.2)	62 (19.3)	
Unknown	318 (22.5)	0 (0.0)	
Histological grade			<0.001
I	375 (26.5)	49 (15.2)	
II	676 (47.7)	104 (32.3)	
III	365 (25.8)	154 (47.8)	
Unknown	0 (0.0)	15 (4.7)	
ER			<0.001
Negative	229 (16.2)	119 (37.0)	
1+	25 (1.8)	22 (6.8)	
2+	81 (5.7)	57 (17.7)	
3+	1,081 (76.3)	124 (38.5)	
PgR			<0.001
Negative	388 (27.4)	132 (41.0)	
1+	105 (7.4)	38 (11.8)	
2+	242 (17.1)	63 (19.6)	
3+	681 (48.1)	89 (27.6)	
HER2			<0.001
Negative	1,242 (87.7)	233 (69.3)	
Positive	162 (11.4)	99 (30.7)	
Unknown	12 (0.8)	0 (0.0)	
Surgical type			<0.001
SLNB only	1062 (75.0)	81(25.2)	
SLNB and ALND/ALND only	335 (23.7)	241(74.8)	
Unknown	19 (1.3)	0 (0.0)	
Longest axis (mm)			0.70
Median (IQR)	12.6 (9.6, 16.9)	13.0 (10.0, 17.0)	
Cortical thickness (mm)			<0.001
Median (IQR)	1.9 (1.3, 2.8)	4.0 (3.0, 6.0)	
Absence of hilum			<0.001
Yes	113 (8.0)	126 (39.1)	
No	1,303 (92.0)	196 (60.9)	
Lymph node metastasis			<0.001
Yes	354 (25.0)	163 (50.6)	
No	1,062 (75.0)	159 (49.4)	

Abbreviations: ALND: axillary lymph node dissection; ER, estrogen receptor; HER2, human epidermal growth factor receptor 2; IQR, interquartile range; LIQ, lower inner quadrant; LOQ, lower outer quadrant; PgR, progesterone receptor; SLNB: sentinel lymph node biopsy; UIQ, upper inner quadrant; UOQ, upper outer quadrant

Table 2. FNRs of the updated models at different model-predicted probability cut-off points

Cut-off point	No. of patients (%)	No. of patients with ALNM	FNR (%)
5%	255 (18.0)	10	2.8
10%	281 (19.7)	13	3.7
11%	323 (22.8)	15	4.2
12%	415 (29.3)	28	7.9
13%	463 (32.7)	37	10.5
14%	540 (38.1)	44	12.4

Abbreviations: ALNM, axillary lymph node metastasis; FNR, false-negative rate

Supplementary Table 1 Comparison of clinicopathological characteristics among six hospitals in the validation group

Characteristics	Hospital 1 No. (%)	Hospital 2 No. (%)	Hospital 3 No. (%)	Hospital 4 No. (%)	Hospital 5 No. (%)	Hospital 6 No. (%)
No. of patients	204 (100)	355 (100)	108 (100)	287 (100)	300 (100)	162 (100)
Age at diagnosis (year)						
Median (IQR)	63 (53, 71)	61 (52, 69)	62 (54, 70)	61 (52, 69)	61 (51, 69)	61 (50, 68)
Menopausal status						
Premenopausal	38 (18.6)	78 (22.0)	17 (15.7)	68 (23.7)	73 (24.3)	38 (23.5)
Postmenopausal	166 (81.4)	277 (78.0)	91 (84.3)	219 (76.3)	227 (75.7)	124 (76.5)
Tumor size (cm) [&]						
Median (IQR)	1.3 (1.0, 2.0)	1.4 (1.0, 2.0)	1.5 (1.2, 2.0)	1.6 (1.1, 2.3)	1.5 (1.0, 2.2)	1.8 (1.1, 2.4)
Clinical tumor size by T-class [&]						
T1	160 (78.4)	268 (75.5)	82 (75.9)	199 (69.3)	208 (69.3)	101 (62.3)
T2-T3	44 (21.6)	87 (24.5)	26 (24.1)	88 (30.7)	92 (30.7)	61 (37.7)
Tumor location [&]						
UOQ	76 (37.3)	143 (40.3)	37 (34.3)	125 (43.6)	118 (39.3)	74 (45.7)
LOQ	19 (9.3)	28 (7.9)	9 (8.3)	23 (8.0)	23 (7.7)	16 (9.9)
UIQ	26 (12.7)	33 (9.3)	12 (11.1)	34 (11.8)	52 (17.3)	21 (13.0)
LIQ	20 (9.8)	33 (9.3)	7 (6.5)	17 (5.9)	31 (10.3)	19 (11.7)
Central	20 (9.8)	18 (5.1)	8 (7.4)	31 (10.8)	17 (5.7)	8 (4.9)
Unknown	43 (21.1)	100 (28.2)	35 (32.4)	57 (19.9)	59 (19.7)	24 (14.8)
Histological grade [&]						
I	53 (26.0)	115 (32.4)	23 (21.3)	93 (32.4)	56 (18.7)	35 (21.6)
II	110 (53.9)	158 (44.5)	42 (38.9)	121 (42.2)	171 (57.0)	74 (45.7)
III	41 (20.1)	82 (23.1)	43 (39.8)	73 (25.4)	73 (24.3)	53 (32.7)
ER [&]						
Negative	22 (10.8)	67 (18.9)	18 (16.7)	51 (17.8)	40 (13.3)	31 (19.1)
1+	4 (2.0)	4 (1.1)	3 (2.8)	4 (1.4)	3 (1.0)	7 (4.3)
2+	5 (2.5)	19 (5.4)	3 (2.8)	25 (8.7)	23 (7.7)	6 (3.7)
3+	173 (84.8)	265 (74.6)	84 (77.8)	207 (72.1)	234 (78.0)	118 (72.8)
PR						
Negative	46 (22.5)	101 (28.5)	30 (29.6)	82 (28.6)	78 (26.0)	51 (31.5)
1+	10 (4.9)	27 (7.6)	2 (1.9)	25 (8.7)	31 (10.3)	10 (6.2)
2+	43 (21.1)	53 (14.9)	16 (14.8)	45 (15.7)	52 (17.3)	33 (20.4)
3+	105 (51.5)	174 (49.0)	60 (55.6)	135 (47.0)	139 (46.3)	68 (42.0)
Her-2						
Negative	187 (92.2)	312 (87.9)	93 (86.1)	249 (86.8)	265 (88.3)	136 (84.0)
Positive	16 (7.8)	38 (10.7)	12 (11.1)	35 (12.2)	35 (11.7)	26 (16.0)
Unknown	1 (0.5)	5 (1.4)	3 (2.8)	3 (1.0)	0 (0.0)	0 (0.0)
Surgical type [&]						
SLNB only	156 (76.5)	275 (77.5)	82 (75.9)	207 (72.1)	244 (81.3)	98 (60.5)
ALND *	44 (21.6)	79 (22.3)	24 (22.2)	73 (25.5)	53 (17.7)	62 (38.3)
Unknown	4 (1.9)	1 (0.2)	2 (1.9)	7 (2.4)	3 (1.0)	2 (1.2)
Longest axis (mm) [&]						
Median (IQR)	12.4 (10.3, 16.7)	11.1 (8.2, 14.9)	14.0 (10.5, 18.3)	13.4 (10.0, 16.8)	14.1 (10.8, 18.9)	12.2 (9.0, 16.6)
Cortical thickness (mm) [&]						
Median (IQR)	2.0 (1.4, 2.9)	1.7 (1.3, 2.4)	1.6 (1.0, 2.7)	1.7 (1.2, 2.5)	1.9 (1.4, 2.9)	2.3 (1.5, 4.1)
Absence of hilum [&]						
Yes	10 (4.9)	24 (6.8)	8 (7.4)	17 (5.9)	24 (8.0)	30 (18.5)
No	194 (95.1)	331 (93.2)	100 (92.6)	270 (94.1)	276 (92.0)	132 (81.5)
Lymph node metastases [&]						
Yes	47 (23.0)	86 (24.2)	25 (23.1)	78 (27.2)	55 (18.3)	63 (38.9)
No	157 (77.0)	269 (75.8)	83 (76.9)	209 (72.8)	245 (81.7)	99 (61.1)

ER: estrogen receptor; HER-2: human epidermal growth factor receptor 2; IQR: interquartile range, which is the 25th percentile and the 75th percentile; LIQ: lower inner quadrant; LOQ: lower outer quadrant; PR: progesterone receptor; UIQ: upper inner quadrant; UOQ: upper outer quadrant; [&] Significantly different from at least one other patient set; * SLNB and ALND or ALND only

Supplementary Table 2. Actual versus predicted rate of axillary lymph node metastasis for the original model

Group	Patient number	Proportion of ALN metastasis	
		Actual % (95% CI)	Mean predicted %
1	472	10.0 (7.3-12.7)	6.8
2	472	17.6 (14.1-21.0)	18.9
3	472	47.5 (42.9-52.0)	51.3

Abbreviations: ALN, axillary lymph node; CI, confidence interval

Supplementary Table 3. Correction factor for slope, new intercept and updated model for each group classified by GLM tree analysis

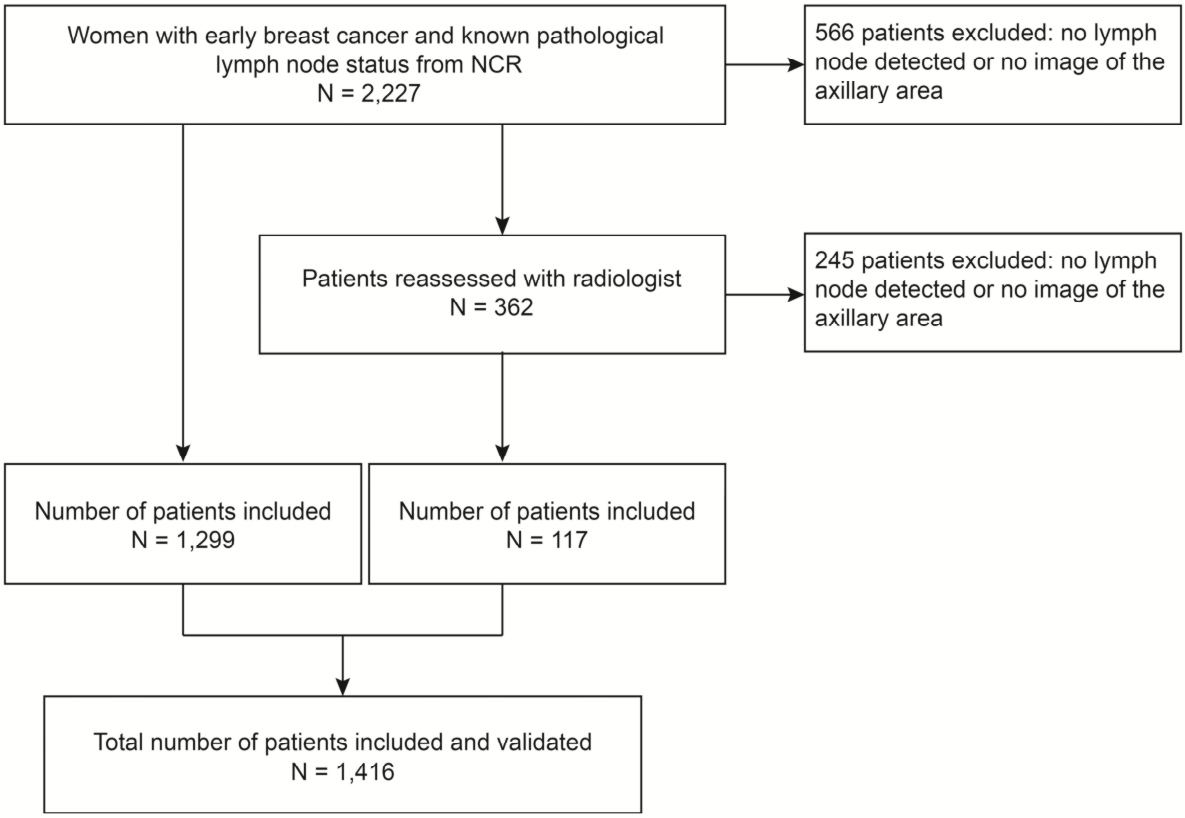
Group	Correction factor for slope	New intercept	Updated model
1	0.1406	-3.6344	ln p/1-p = 0.0089**a* + 0.0389**b* + 0.1997**c* + 0.2112 (*d1*) + 0.2939 (*d2*) + 0.0429**e* + 0.0533**f* - 3.6344
2	0.2451	-2.7284	ln p/1-p = 0.0154**a* + 0.0679**b* + 0.3480**c* + 0.3681 (*d1*) + 0.5123 (*d2*) + 0.0748**e* + 0.0929**f* - 2.7284
3	0.6454	-4.0356	ln p/1-p = 0.0407**a* + 0.1788**b* + 0.9165**c* + 0.9694 (*d1*) + 1.3489 (*d2*) + 0.1968**e* + 0.2446**f* – 4.0356
4	0.8074	-4.0805	ln p/1-p = 0.0509**a* + 0.2236**b* + 1.1465**c* + 1.2127 (*d1*)+ 1.6875 (*d2*) + 0.2463**e* + 0.3060**f* – 4.0805

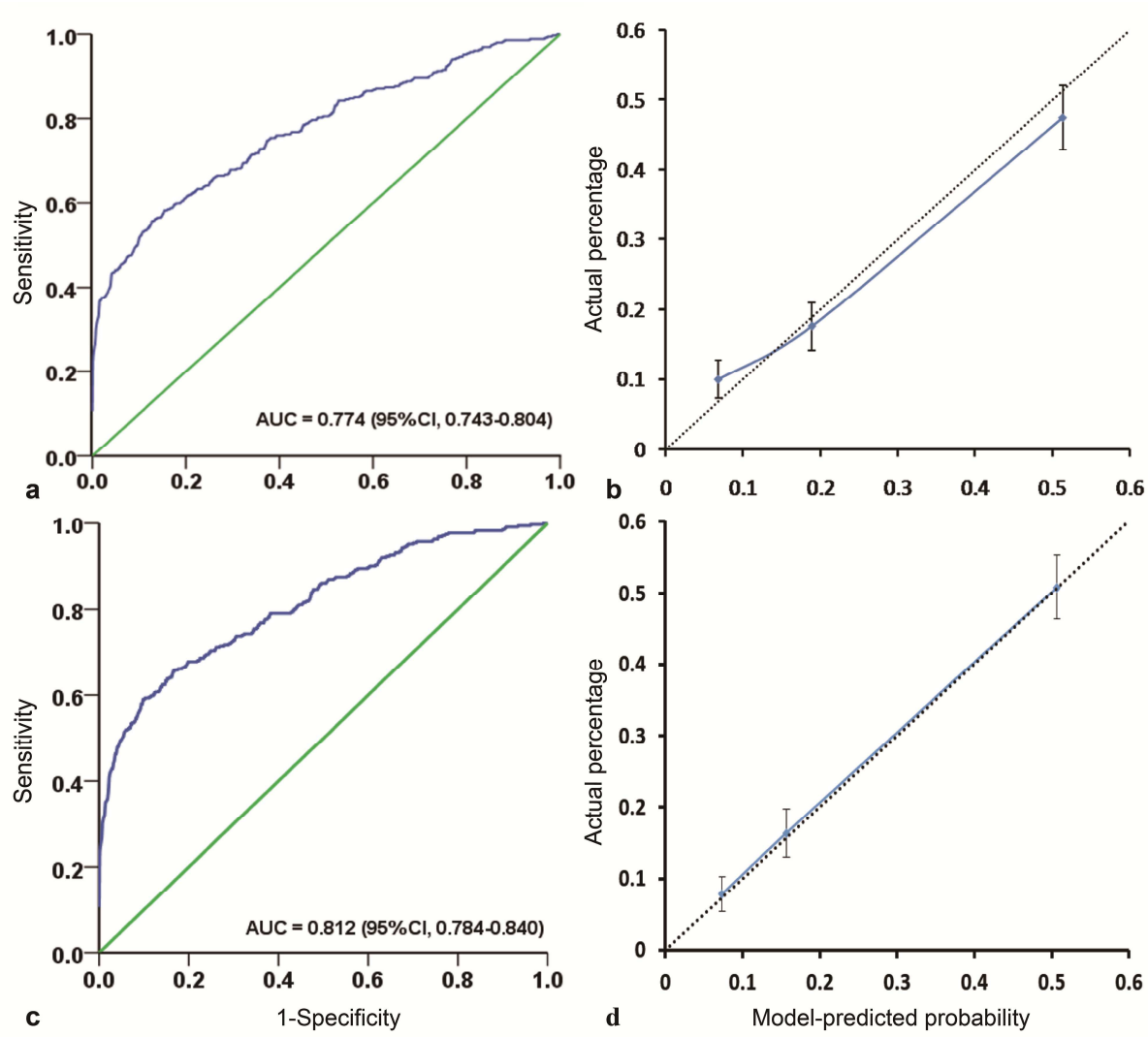
Abbreviations: GLM: Generalized linear model, a: longest axis of lymph nodes, continuous variable, b: cortical thickness of lymph nodes, continuous variable, c: hilum of lymph nodes, c=0 when presence and c=1 when absence, d1 and d2: d1=1 and d2=0 when histological tumor grade 2, d1=0 and d2=1 when histological tumor grade 3, e: clinical tumor size, continuous variable, f: tumor estrogen receptor status, f=0 when ER is negative, f=1 when ER+, f=2 when ER2+ and f=3 when ER3+, p in the formulas: the predictive probability of axillary lymph node metastasis.

Supplementary Table 4. Actual versus predicted rate of axillary lymph node metastasis for the updated models

Group	Patient number	Proportion of ALN metastasis	
		Actual % (95% CI)	Mean predicted %
1	472	7.8 (5.4-10.3)	7.4
2	472	16.3 (13.0-19.7)	15.7
3	472	50.8 (46.3-55.4)	50.7

Abbreviations: ALN, axillary lymph node; CI, confidence interval




Sensitivity
AUC = 0.774 (95%CI, 0.743-0.804)
a
Actual percentage
b
Sensitivity
AUC = 0.812 (95%CI, 0.784-0.840)
1-Specificity
c
Actual percentage
Model-predicted probability
d