



OPEN

A negative binomial regression model for risk estimation of 0–2 axillary lymph node metastases in breast cancer patients

De Zeng^{1,2,7}, Hao-Yu Lin^{3,7}, Yu-Ling Zhang⁴, Jun-Dong Wu^{2,5}, Kun Lin⁶, Ya Xu⁵ & Chun-Fa Chen⁵✉

Extensive clinical trials indicate that patients with negative sentinel lymph node biopsy do not need axillary lymph node dissection (ALND). However, the ACOSOG Z0011 trial indicates that patients with clinically negative axillary lymph nodes (ALNs) and 1–2 positive sentinel lymph nodes having breast conserving surgery with whole breast radiotherapy do not benefit from ALND. The aim of this study is therefore to identify those patients with 0–2 positive nodes who might avoid ALND. A total of 486 patients were eligible for the study with 212 patients in the modeling group and 274 patients in the validation group, respectively. Clinical lymph node status, histologic grade, estrogen receptor status, and human epidermal growth factor receptor 2 status were found to be significantly associated with ALN metastasis. A negative binomial regression (NBR) model was developed to predict the probability of having 0–2 ALN metastases with the area under the curve of 0.881 (95% confidence interval 0.829–0.921, $P < 0.001$) in the modeling group and 0.758 (95% confidence interval 0.702–0.807, $P < 0.001$) in the validation group. Decision curve analysis demonstrated that the model was clinically useful. The NBR model demonstrated adequate discriminative ability and clinical utility for predicting 0–2 ALN metastases.

The number of axillary lymph node (ALN) metastases is an essential indicator for evaluating the severity and survival of patients with breast cancer¹. A greater number of positive ALNs is associated with higher probability of recurrence and mortality², and more than 6 metastatic ALNs might indicate a high risk of distant metastasis³. The number of involved ALNs is a determinant parameter in treatment decision making, particularly for those with 1–2 positive sentinel lymph nodes (SLNs). For these subgroups of patients and those with negative ALNs, sentinel lymph node biopsy (SLNB) has now become the preferred surgical procedure, rather than axillary lymph node dissection (ALND)^{4–6}.

A number of studies have shown that, in the absence of preoperative ALN status evaluation, including those with T1–4 stage diseases, approximately 51–67% of patients with breast cancer have negative ALNs^{7–10}. About 40% of patients with clinically suspicious ALN metastasis had pathologically negative SLNs^{11–13}. For almost half of the patients with involved SLNs, there was no addition of involved ALNs identified on ALND^{14,15}. Even in the T1–2 stage patients with positive SLNs, approximately 60–73% of patients had no additional non-SLN metastasis after ALND^{4,6}. Extensive clinical trials indicate that patients with negative SLNB do not need ALND, and the ACOSOG Z0011 trial showed that patients with clinical T1–2N0 tumors, without palpable lymph adenopathy, and 1–2 positive SLNs receiving breast conserving surgery and subsequent radiation and adjuvant therapy could be spared from ALND^{6,16}. Therefore, the development of reliable predictors to identify those patients with 0–2 positive nodes who might avoid ALND is needed for treatment quality assurance and improvement.

¹Department of Medical Oncology, Cancer Hospital of Shantou University Medical College, Shantou 515031, China. ²Guangdong Provincial Key Laboratory of Breast Cancer Diagnosis and Treatment, Shantou 515031, China. ³Department of Thyroid and Breast Surgery, The First Affiliated Hospital of Shantou University Medical College, Shantou 515041, China. ⁴Department of Information, Cancer Hospital of Shantou University Medical College, Shantou 515031, China. ⁵The Breast Center, Cancer Hospital of Shantou University Medical College, Shantou 515031, China. ⁶Department of Public Health and Preventive Medicine, Shantou University Medical College, Shantou 515041, China. ⁷These authors contributed equally: De Zeng and Hao-Yu Lin. ✉email: chenchenfa@stu.edu.cn

Several studies have used preoperative ultrasound imaging¹⁷, computed tomography imaging¹⁸, magnetic resonance imaging¹⁷, or ultrasound-guided fine-needle biopsy to evaluate the status of ALNs¹⁹. However, these studies often neglected the distribution of involved nodes and usually investigated the axillary status using a logistic regression model for predicting ALN status, which is not conducive to predicting the number of metastatic nodes. The numbers of metastatic ALNs are count data that often display overdispersion²⁰; hence, negative binomial regression (NBR), which has been well described by Hilbe²¹, is an appropriate choice for modeling this distribution^{20,22,23}. A recent study comprising of 109,618 patients with breast cancer from the Surveillance, Epidemiology, and End Results database indicated that the distribution of overdispersion of metastatic ALNs was similar to a log-concave curve²³. The NBR model provided an excellent fit curve, which could predict the number of involved nodes. In this retrospective study, we aim to first identify factors that affect counts of ALN metastasis and then develop an NBR model to estimate the probability of exact counts of ALN metastasis and calculate the probability of 0–2 ALN metastases, with the ultimate goal of guiding physicians to make appropriate surgical plans preoperatively for breast cancer patients.

Results

Patients characteristic. The demographic information and descriptive statistics are summarized in Table 1. The estrogen receptor (ER) status and Ki67 index between the modeling and the validation groups was significantly different ($P < 0.05$). There were more patients with ER-negative and Ki67 $\geq 14\%$ in the modeling group than in the validation group. A significant difference was also found in the type of axillary surgery between these two groups ($P < 0.001$), with more patients receiving SLNB in the validation group than in the modeling group. Considering the marked difference in the counts of metastatic nodes in patients who underwent SLNB + ALND between the two groups ($P < 0.05$), more metastatic nodes were dissected in the validation group than in the modeling group. The counts of dissected nodes between the two groups were significantly different ($P < 0.05$), and more nodes were dissected in the modeling group than in the validation group. The counts associated with SLNB were also different between the modeling group and the validation group ($P < 0.05$), and more SLNs were dissected in the validation group than in the modeling group. No significant difference was found between the modeling group and the validation group in terms of the other clinicopathologic variables ($P > 0.05$).

The distribution of ALNs in the modeling group and the validation group are shown in Table 2. There were 48.1% of patients with negative ALNs and 20.7% of patients with 1–2 ALNs metastases in the modeling group compared to 54.0% of patients with negative ALNs and 21.4% of patients with 1–2 ALNs metastases in the validation group.

The goodness of fit for a negative binomial distribution and Poisson distribution. The histograms and fitted curves of the number of ALNs metastases in the present study are shown in Fig. 1A. The mean and median numbers of metastatic ALNs were 2.58 (variance of 22.38) and 1 (range 0 to 35), respectively, in modeling group. The distribution showed a logarithmic concave curve, suggesting that the majority of patients had negative ALNs and then decreased as the number of metastatic ALNs increased. It could be fitted with a Poisson distribution or a Poisson-gamma distribution. However, the goodness of fit with Poisson-gamma distribution was better than that with the Poisson distribution (Fig. 1B,C). In addition, the variance of the number of metastatic ALNs in the present study was nearly 10 times greater than the mean, and the NBR model provided an improved fit for the data and accounted better for overdispersion than did the Poisson regression model, which assumes that the mean and variance are the same. Therefore, we adopted the NBR model to analyze the data in the present study. The lack-of-fit statistics with the NBR model in the modeling group are shown in Table S1.

The result of the negative binomial regression model. As shown in Table 3, several factors were independently correlated with the counts of metastatic ALNs, including clinical lymph node status, histologic grade, ER status, and human epidermal growth factor receptor 2 (HER2) status. The incidence rate ratio (IRR) of positive clinical lymph node status was higher for positive ALN metastasis [IRR = 2.88, 95% confidence interval (CI) 2.29–3.63] than for negative clinical lymph node status. A similar result was also found for histologic grade; the IRR of ALN metastasis in histologic grade 3 was 1.38 (95% CI 1.14–1.69) times higher than that in histologic grade 1–2 patients. Similarly, the IRR of ALN metastasis in HER2-positive patients was 1.33 (95% CI 1.06–1.67) times higher than that in HER2-negative patients. Patients who were ER-positive were more likely to suffer ALN metastasis than ER-negative patients, and the IRR of ALN metastasis in ER-positive patients was 1.48 (95% CI 1.18–1.87) times higher than that in ER-negative patients. Although there was no statistical difference between the size of primary tumor and the number of ALNs involvement in the model, there was a trend for tumor size to be associated with the number of ALNs metastases. The larger the tumor was, the higher the risk of ALNs metastases. For every 1 cm increase in the diameter of the tumor, the risk increased by 0.13 times.

As shown in Fig. 2, the probabilistic distribution of the NBR model fit both the modeling group and the validation group well, and there were no statistically significant differences between the predicted and the observed probability. The prediction between the observed modeling group and the validation group were $P = 1.0$ and $P = 0.736$, respectively.

The discrimination and clinical utility of the NBR model for predicting 0–2 ALN metastases. In the modeling group, the Area under the curve (AUC) for predicting 0–2 ALN metastases was 0.881 (95% CI 0.829–0.921; $P < 0.001$), and the corresponding sensitivity and specificity were 82.2% and 83.3%, respectively (Fig. 3A). In the validation group, the optimal cutoff value of probability was 90.0%, with an AUC for predicting 0–2 ALN metastases was 0.758 (95% CI 0.702–0.807; $P < 0.001$). The sensitivity, specificity and Yondex

Characteristics	Modeling group		Validation group		P
	No	%	No	%	
No. of patients	212	100.0	274	100.0	
Female	212	100.0	274	100.0	
Age (years)					
Median (range)	50.5 (23–91)		50 (24–84)		0.916
Menstrual status					0.373
Pre-menopause	109	51.4	152	55.5	
Post-menopause	103	48.6	122	44.5	
Laterality					0.414
Left	117	55.2	141	51.5	
Right	95	44.8	133	48.5	
Tumor size (cm)					
Median (range)	3.0 (0.7–9.0)		3.0 (1.0–10.0)		0.756
Tumor location					0.143
UOQ	99	46.7	156	56.9	
LOQ	21	9.9	27	9.9	
LIQ	21	9.9	15	5.5	
UIQ	37	17.5	40	14.6	
CR	34	16.0	36	13.1	
Clinical ALN status					0.056
Negative	103	48.6	157	57.3	
Positive	109	51.4	117	42.7	
Stage					0.083
I	36	17.0	59	21.5	
II	116	54.7	160	58.4	
III	60	28.3	55	20.1	
Histologic grade					0.161
1–2	135	63.7	191	69.7	
3	77	36.3	83	30.3	
ER status					0.036
Negative	81	38.2	80	29.2	
Positive	131	61.8	194	70.8	
PR status					0.435
Negative	91	42.9	108	39.4	
Positive	121	57.1	166	60.6	
HER2 status					0.707
Negative	145	68.4	183	66.8	
Positive	67	31.6	91	33.2	
Ki67					0.048
< 14%	21	9.9	44	16.1	
≥ 14%	191	90.1	230	83.9	
Molecular subtype					0.136
HR(+)/HER2(-)	109	51.4	142	51.8	
HR(+)/HER2(+)	29	13.7	56	20.4	
TN	36	17.0	41	15.0	
HER2(+)	38	17.9	35	12.8	
Breast surgery					0.463
Conserving surgery	50	23.6	57	20.8	
Mastectomy	162	76.4	217	79.2	
Axillary surgery					<0.001
SLNB	56	26.4	121	44.2	
SLNB+ALND	43	20.3	66	24.1	
ALND	113	53.3	87	31.7	
ALNs metastasis					0.248
Negative	102	48.1	148	54.0	
1–2 positive ALNs	44	20.8	59	21.5	
Continued					

Characteristics	Modeling group		Validation group		P
	No	%	No	%	
> 2 positive ALNs	66	31.1	67	24.5	
Counts of metastatic nodes (median, rang)	1 (0–35)		0 (0–29)		0.125
SLNB	0 (0–1)		0 (0–2)		0.824
SLNB + ALND	1 (0–10)		2 (0–29)		0.002
ALND	3 (0–35)		1 (0–26)		0.537
Counts of dissected nodes (median, rang)	15 (1–43)		9 (1–32)		0.004
SLNB	2 (1–5)		2 (1–7)		0.036
SLNB + ALND	17.5 (8–32)		17 (6–30)		0.461
ALND	17 (2–43)		18 (7–32)		0.584

Table 1. Comparison of descriptive characteristics between the modeling group and the validation group. Abbreviation: ALN, Axillary lymph node; ALND, axillary lymph node dissection; CR, Central region; ER, estrogen receptor; HER2, human epidermal growth factor receptor 2; HR, hormone receptor; LIQ, Lower inner quadrant; LOQ, Lower outer quadrant; PR, progesterone receptor; SLN, sentinel lymph node; SLNB, sentinel lymph node biopsy; TN, triple negative; UIQ, Upper inner quadrant; UOQ, Upper outer quadrant.

Count of metastatic axillary lymph nodes	Modeling group	Validation group
	No. (%) (n = 212)	No. (%) (n = 274)
0	102 (48.1)	148 (54.0)
1	27 (12.7)	40 (14.5)
2	17 (8.0)	19 (6.9)
3	15 (7.0)	19 (6.9)
4	12 (5.7)	9 (3.3)
5	5 (2.4)	3 (1.1)
6	8 (3.8)	6 (2.2)
7	4 (1.9)	4 (1.5)
8	7 (3.3)	4 (1.5)
9	1 (0.5)	4 (1.5)
10	3 (1.4)	–
11	2 (0.9)	2 (0.7)
12	3 (1.4)	3 (1.1)
13	2 (0.9)	2 (0.7)
18	–	1 (0.4)
19	–	1 (0.4)
20	–	2 (0.7)
21	–	2 (0.7)
22	–	1 (0.4)
24	1 (0.5)	–
26	1 (0.5)	2 (0.7)
27	–	1 (0.4)
28	1 (0.5)	–
29	–	1 (0.4)
35	1 (0.5)	–

Table 2. Distribution of metastatic axillary lymph nodes.

index were 64.3%, 83.6% and 47.8%, respectively (Fig. 3B). Both of the results indicated a good predictive ability of the NBR model.

The decision curve analysis (DCA) showed that the net benefit with the NBR model for predicting 0–2 ALN metastases in the validation group, with a probability range of 0.58 to 0.90, was superior to either the treat-all or treat-none methods (Fig. 4A). For example, the net benefit was approximately 11% at the 90% probability threshold, which indicated that at this probability threshold was equivalent to indicating 11 patients with 0–2 ALN metastases per 100 patients. At a probability threshold of 90%, the net reduction in interventions was approximately 16 per 100 patients (Fig. 4B). For example, this probability threshold is equivalent to 16 per 100 patients have 0–2 ALN metastases, which may indicate that 16% of patients can avoid unnecessary ALND.

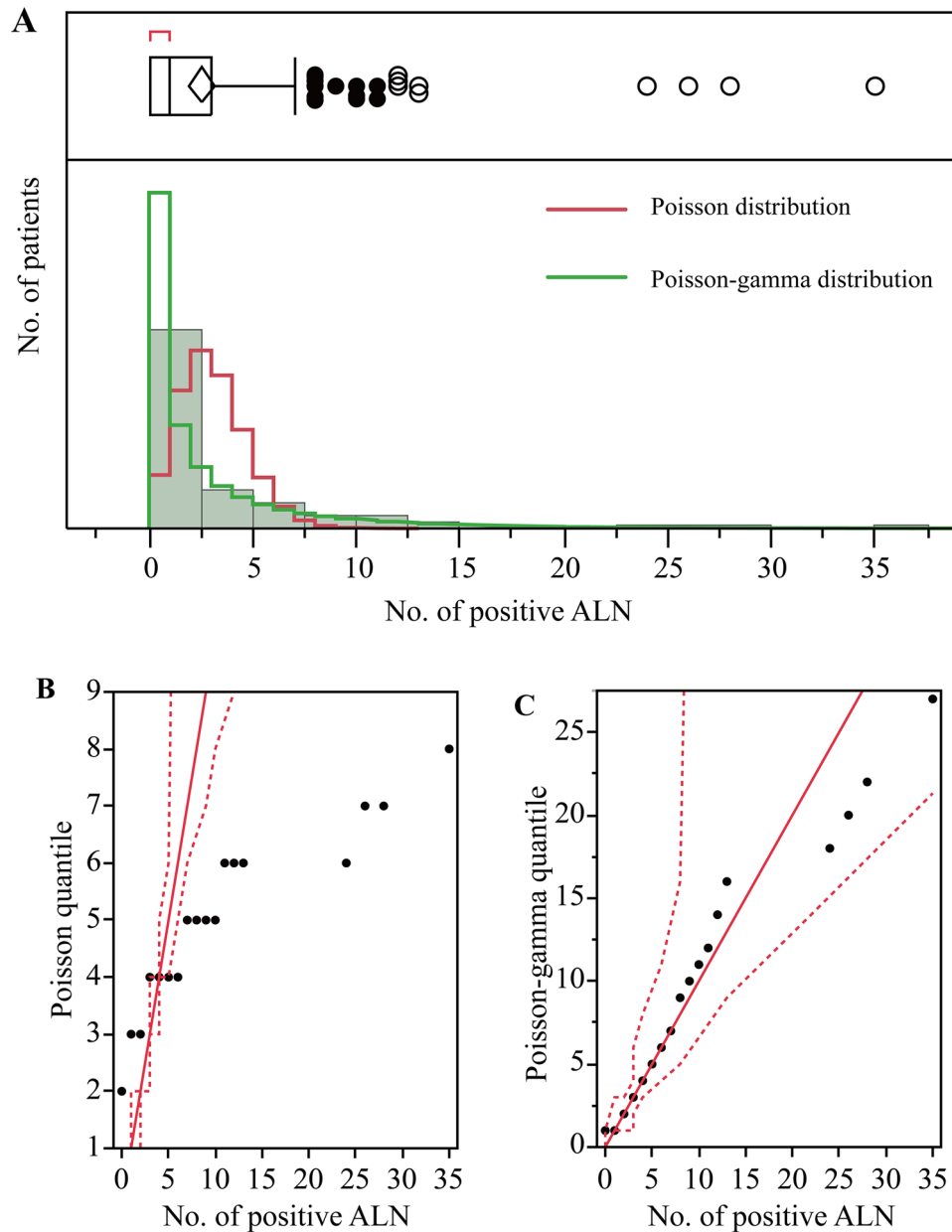


Figure 1. Distribution of metastatic axillary lymph node (ALN) counts. **(A)** Distribution of the metastatic ALNs. The counts of metastatic ALNs approximately follow a Poisson-gamma distribution. In the boxplot above the graph, the vertical line in the bottom of the box indicates the median, the diamond indicates the mean, and the left and right borders of the box mark the 25th and 75th percentiles, respectively. The solid point extending from the right of the box marks the end of 95th percentiles. The hollow points are outliers beyond the 95th percentile. **(B)** Goodness of fit with a Poisson distribution. The red line indicates the fit line and the dotted line shows the 95% confidence interval. **(C)** Goodness of fit with Poisson-gamma distribution. The red line indicates the fit line, and the dotted line shows the 95% confidence interval.

Discussion

Assessment of axillary status has become a fundamental step in cancer counseling and treatment planning in clinical practice, which might identify low- or high-risk situations and aid in treatment decision making, such as the choice between SLNB or neoadjuvant therapy²⁴. At present, most of the available studies, which addressed the issue of prediction of ALN status, adopted the categorical (dichotomous/polychotomous) outcome variables and applied binary/multinomial logistic regression models to find their predictors^{9,19}. These studies with logistic regression models could merely predict ALN status, such as negative or positive results, while neglecting the actual distribution of the involved nodes. Therefore, we argued that a method of modeling that takes into account the number of nodes, in combination with the nodal status, would allow for a grading of nodal involvement risk and might identify candidate patients for whom SLNB, neoadjuvant chemotherapy, ALND or radiotherapy be

Independent variable	Coefficients	Standard error	Wald χ^2	P	IRR (95% CI)
Intercept	-0.03	0.26	-0.13	0.899	0.97 (0.58–1.60)
Clinical lymph node status					
Negative	-	-	-	-	-
Positive	1.06	0.12	9.00	<0.001	2.88 (2.29–3.63)
Histologic grade					
1–2	-	-	-	-	-
3	0.33	0.10	3.22	0.001	1.38 (1.14–1.69)
ER status					
Negative	-	-	-	-	-
Positive	0.39	0.12	3.36	<0.001	1.48 (1.18–1.87)
HER2 status					
Negative	-	-	-	-	-
Positive	0.28	0.12	2.42	0.016	1.33 (1.06–1.67)
Primary tumor size	0.12	0.07	1.78	0.075	1.13 (0.99–1.29)

Table 3. The fitting result of negative binomial regression model. Abbreviation: CI, confidence interval; ER, estrogen receptor; HER2, human epidermal growth factor receptor 2; IRR, Incidence rate ratio.

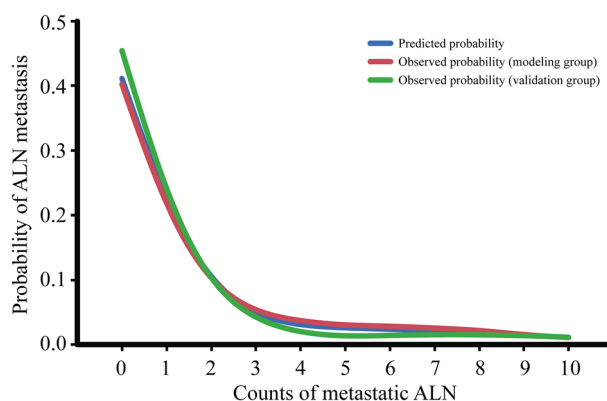


Figure 2. The probabilistic distribution of the negative binomial regression model fitted by the predicted group and observed group. ALN, axillary lymph node.

justified²³. Especially after the ACOSOG Z0011 trial, the treatment strategy for axillary surgery according to the number of metastatic ALNs has become particularly important.

A large database composed of 224,656 breast cancer, 12,404 gastric cancer, 18,015 rectal cancer, 4,117 cervical cancer, 2,443 laryngeal cancer and 9,118 melanoma patients indicated that the number of involved lymph nodes is a set of nonnegative integer values not normally distributed, and the data exhibited considerable heterogeneity within populations²⁰. Consistent with this phenomenon, the frequency distributions for the numbers of involved ALNs in the present study appeared skewed, with a large proportion of individuals with no involved nodes and a smaller proportion with many involved nodes. A statistical description of the frequency distribution of the numbers of involved nodes in an affected population could potentially reveal mechanisms of axillary metastasis, and eventually facilitate predictive models for axillary sampling²². Therefore, we adopted the NBR model for our study. In line with our study, Guern et al. indicated that the NBR model provided a better fit for the number of metastatic ALNs in breast cancer than the Poisson regression model²³.

At present, the exact mechanism of ALN metastasis is unclear. In clinical practice, some patients are more susceptible to ALN metastasis than would be suggested by random events. Our results showed that the variance in the number of metastatic ALNs was nearly 10 times greater than the mean, indicating that the number of involved ALNs in patients with breast cancer is highly variable. This overdispersion of involved ALNs was described by Guern and colleagues and is known as "infectivity", that is, the more involved ALNs a person has, the greater the risk that the same person will subsequently have more involved ALNs²³. Bori and colleagues also described this phenomenon in colorectal cancer²⁵. A cascade process, which we believe is more appropriate than an infection-type process, is easy to understand because lymph nodes are not isolated organs but are interconnected through a complete lymphatic network. In addition, overdispersion means that, as the number of lymph nodes increases, the variance increases. In addition, when the risk increases, individual variability also increases. There are still some clinical paradoxes in that some patients with severe ALN metastasis do not end up with distant metastasis and have a relatively favorable prognosis.

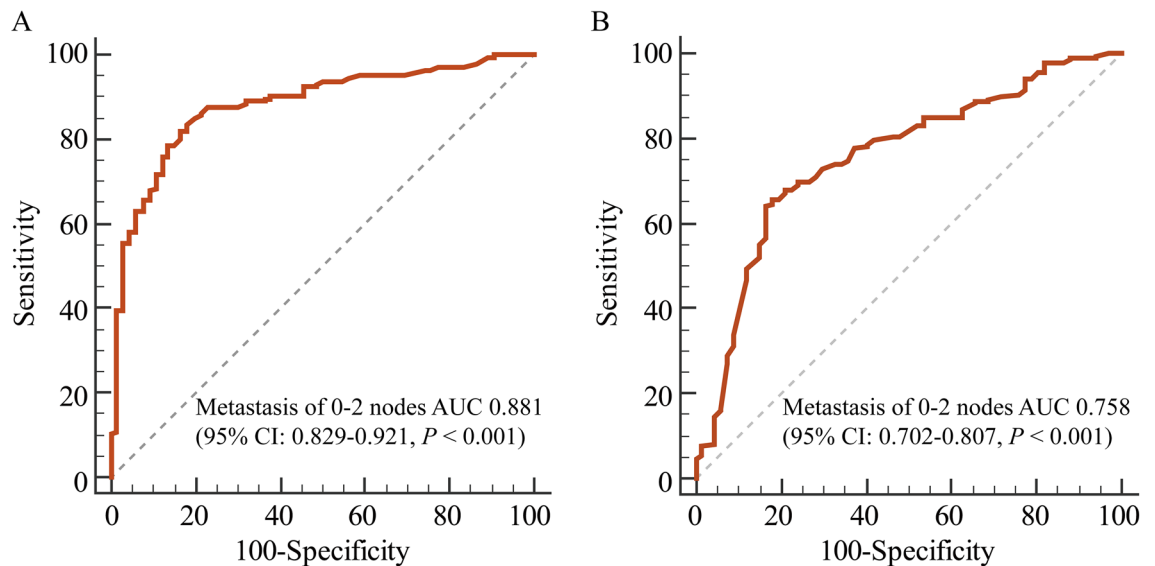


Figure 3. Receiver operating characteristic curve corresponding to the negative binomial regression model. (A) Area under the curve (AUC) for predicting 0–2 metastatic nodes in the modeling group. (B) AUC for predicting 0–2 metastatic nodes in the validation group. CI, confidence interval.

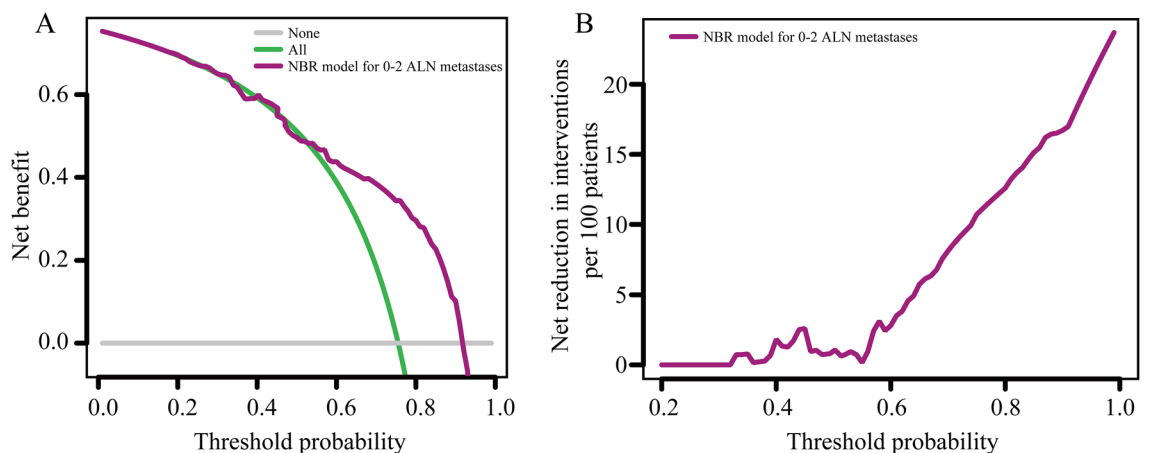


Figure 4. Decision curve analysis of the negative binomial regression (NBR) model in the validation group. (A) Net benefits for predicting 0–2 axillary lymph node (ALN) metastases. (B) Net reduction in interventions per 100 patients at different probability thresholds for predicting 0–2 ALN metastases.

Primary tumor size is the most powerful predictor of ALN status. There is a linear relationship between tumor size and ALN involvement. The chance of ALN involvement significantly increased from pT1mic (9.5%) to pT2 and pT3 tumor specimens (49.9% and 65.6%, respectively)¹⁰. Although tumor size was not significantly associated with the counts of ALN metastases in our study, it could not be excluded in our model. Our results showed a trend that the risk of ALN metastases increases consistently with tumor size increases. This finding agrees with a recent study using Bayesian NBR⁸, and is also supported by several studies showing larger tumor size is correlated with an increased risk of a higher number of positive ALNs^{10,26}. We may conclude that patients with different tumor sizes might undergo surgery to remove individual axillary nodes for refining treatment. A previous study by Coombs, N suggested a similar perspective²⁷. In their results, small tumors less than 20 mm in size tended to potentially benefit from SLNB. In patients with moderate-sized tumors (20–40 mm), the risk of ALNs is relatively high and should be assessed according to tumor characteristics and other work-up results, and patients will probably need a secondary axillary node operation. Patients with a large tumor size (>40 mm) usually have a high probability of ALN metastases, and complete ALND is a preferred surgical procedure.

Presently, we have entered an era of precision medicine that attempts to be less invasive in our surgical procedures and to lower treatment-related long-term morbidity, with an ultimate goal of improving the quality of life for cancer survivors. The advent of SLNB has drastically decreased the complications of ALND. However, the question of whether SLNB alone is good enough in the setting of metastatic ALNs, or whether ALND is still required to improve cancer-related outcomes, remains to be answered. The perspective that ALND does not

improve survival and is therefore not a rational option was derived from the NSABP B-04 trial in which patients received total mastectomy with or without ALND as well as with or without radiation therapy²⁸. After a 25-year follow-up, no significant difference was found in the axillary recurrence rate in both treatment groups²⁹. This result indicated that there might be a subset of patients who do not need ALND or regional radiation therapy. The ACOSOG Z0010³⁰ and NSABP B-32³¹ trials showed that patients with positive ALN by immunohistochemistry staining did not need to proceed to complete ALND without a reduction in overall survival. Since then, evidence has accumulated from studies of the outcomes of patients with involved SLNs who have a low axillary burden do not undergo ALND, and no effects on overall survival and axillary recurrence were found. The IBCSG 23-01 trial demonstrated that omission of ALND did not lead to compromised local recurrence or survival when one or more micro-metastatic SLNs were found in patients with early-stage disease³². The ACOSOG Z0011 trial then further showed that in patients with early-stage breast cancer and 1–2 micro-metastatic or macro-metastatic SLNs, treated with breast conservation therapy and subsequently chemotherapy, omission of ALND did not adversely impact locoregional control and survival³³. The results of the AMAROS trial partly confirmed that mastectomy or conservation surgery in patients with limited positive SLNs after axillary radiotherapy or ALND does not have an effect on axillary recurrence³⁴. All of the above clinical trials supported the idea that ALND is unnecessary in these selected patients. Our study showed that about approximately 70% of patients had 0–2 ALN metastases, which may indicate that SLNB was sufficient for these selected patients. However, we cannot forget that the ultimate goal of management is to cure and control cancer. Therefore, preoperatively evaluating patients with a low ALN burden who would not obtain additional benefit from ALND has become very important.

In the present study, the receiver operating characteristic (ROC) analysis in the modeling group and validation group both indicated a good predictive ability. To further justify clinical utility, DCA was applied to assess whether the NBR model would improve our evaluation of 0–2 ALN metastases in validation group. Combination of ROC and DCA analyses, patient with a probability threshold of >90%, has the greatest chance of avoiding unnecessary ALND.

There are several highlights in the study: (1) To the best of our knowledge, this is the first NBR model for predicting 0–2 ALN metastases in patients with breast cancer. The model is very important for assisting in management axillary surgery in breast cancer patients in the post-Z001 trial era. (2) Our predictive model included seven variables that are routinely available preoperatively, such as primary tumor size, clinical ALN status, histologic grade, ER status, and HER2 status, which can be obtained before surgery. The availability of these data is very important for clinicians to inform patients regarding the possibility of having involved ALNs before undergoing the procedure. (3) Our approach of creating a modeling with data from one hospital and validating it with data from another center is an appropriate measure to reduce bias.

Despite the intriguing findings and strengths in the present study, there are a few limitations that need to be addressed. First, the present study did not include the lympho-vascular invasion variable since it is difficult to identify through a core needle biopsy. Several studies indicated that lympho-vascular invasion was a stronger risk factor associated with ALN metastasis^{3,10,35}. The lack of consideration of lympho-vascular invasion in this study may influence the accuracy of the model. Second, the occurrence of node metastasis partly reflects a time-dependent process, as cancer cells are more likely to develop lymphogenic and/or hematogenic metastases with time. The age of a breast tumor is difficult to calculate. However, the tumor size can indirectly reflect tumor age. In this study, tumor size was also included in the final model, which was related to the incidence of ALN metastasis. Therefore, the lack of define tumor age may not affect the outcome. Third, this study has inherent weaknesses because of its retrospective nature. Fourth, this model can only be applied in patients with a single lesion and is restricted to those with unilateral disease.

In conclusion, ALN evaluation preoperatively is an essential element in the management of breast cancer. The shift to less invasive surgical procedures and increased adoption of tailored therapy has been paralleled by the shift from ALND to SLNB and even omitting axillary surgery in selected patients. Our study highlights the distribution of ALN metastasis and the factors that may affect the counts of ALN metastases in breast cancer. Our NBR model suggests the relevance of these important predictors and might help physicians to weigh the risks of involved ALNs and determine appropriate surgical procedures.

Materials and methods

Patients. The inclusion criteria of present study were (1) female patients with unilateral breast cancer, (2) patients with pathologically confirmed invasive ductal carcinoma, and (3) patients with complete medical registry and follow-up information. The exclusion criteria were male patients, patients with bilateral or occult breast cancer, patients treated with neoadjuvant chemotherapy, patients with distant metastasis, and patients with pathologically confirmed non-invasive ductal carcinoma. Patients with ductal carcinoma in situ with microinvasion were also excluded due to the difficulty of being diagnosed with core needle biopsy. From Jan 2012 to Dec 2013, 434 consecutive patients with primary breast cancer admitted to The Breast Center in the Cancer Hospital of Shantou University Medical College (SUMC) were reviewed. A total of 212 patients eligible for our study were included in the modeling group (Figure S1A). From Aug 2014 to Dec 2017, 384 consecutive patients with primary breast cancer admitted to the Department of Thyroid and Breast Surgery in the First Affiliated Hospital of SUMC were reviewed. Ultimately, 274 eligible patients were included in the validation group (Figure S1B). The demographic and clinicopathologic characteristics were obtained from the hospital medical records, including age at diagnosis of tumor, primary tumor size based on preoperative physical examination or imaging scans, laterality, tumor location (quadrant), menstrual status, clinical lymph node status, number of ALN metastases, number of ALN dissections, ER status and progesterone receptor (PR) status, Ki67 index, HER2 status, histologic grade and type of surgery.

In the modeling group, ALND was performed when pathological assessment, through either hematoxylin and eosin staining or immunohistochemistry, revealed SLN involvement. According to the Z0011 trial criterion, ALND was performed only in the positive SLNs postoperatively, excluding those with micro-metastasis³⁶.

Pathology. Pathological examinations of tumor samples were analyzed by two experienced pathologists. The number of ALNs with and without metastases at the final pathology review was recorded. Tumor tissues were obtained from surgical resection specimens. The technique used for histopathological analysis is referred to previous publication³⁷. The histological grade scoring for invasive tumors was in line with the Nottingham grading system³⁸. Serial section hematoxylin and eosin staining was performed to examine all nodes postoperatively. Immunohistochemistry staining was further performed to determine the presence or absence of micro-metastases (0.2–2 mm cancer foci) when no cancer cells were detected on hematoxylin and eosin staining. In the present study, micro-metastasis was defined as ALN metastasis. ER³⁹ and PR³⁹ were considered positive if more than 10% of tumor cells demonstrated positive immunostaining. A score of 3+ on immunohistochemistry or amplification on fluorescence in situ hybridization was considered HER2 positivity⁴⁰. According to the Breast Cancer Working Group guidelines, the Ki67 index was determined by the percentage of positive cells among a total number of 1,000 tumor cells, with at least 500 tumor cells being counted⁴¹. Ki67 was considered positive if $\geq 14\%$ tumor cells had nuclear staining⁴².

Statistics. The traditional NBR model, commonly known as negative binomial (NB) 2, is based on a Poisson-gamma mixture distribution allowing the modeling of Poisson heterogeneity using a gamma distribution²¹. In NBR, the mean of u is calculated by incorporating the exposure time t and a set of k regressor variables, and the variance function is $\mu + \alpha\mu^2$. In the present study, the exposure time was set to 1.0. The expression of related quantities is formulated as follows:

$$u_i = \text{Exp}(\beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki}) \quad (1)$$

Often, $x_1 \equiv 1$, in which β_1 is termed the intercept. The regression coefficients $\beta_1, \beta_2, \dots, \beta_k$ are unknown parameters that are estimated from a set of data. Their estimates are symbolized as b_1, b_2, \dots, b_k . Using this notation, the probability distribution function of the NBR model for an observation i is written as follows⁴³:

$$\text{Pr}(y_i | u_i, \alpha) = \frac{\Gamma(y_i + \alpha^{-1})}{\Gamma(y_i + 1)\Gamma(\alpha^{-1})} \left(\frac{1}{1 + \alpha u_i}\right)^{\alpha^{-1}} \left(\frac{\alpha u_i}{1 + \alpha u_i}\right)^{y_i} \quad (2)$$

The NB distribution is a two-parameter model with a mean parameter u and dispersion parameter α . The function $\Gamma(\cdot)$ is the gamma function.

Mann–Whitney U test was used to analyze continuous variables, and the χ^2 test was used for the comparison of categorical variables. The dependent variable was the counts of metastatic ALNs (discrete numerical variable). The independent variables in the model were age at diagnosis of tumor (continuous variable), and clinical tumor size (continuous variable). Categorical variables included menstrual status, laterality, quadrant, clinical lymph node status, histologic grade, ER/PR status, HER2 status and Ki67 index. The maximum likelihood was used to estimate the regression coefficients⁴³.

In the modeling group, we chose the one-way terms options, and set the hierarchical forward to switch with the regression model. We chose a subset size using the maximum likelihood after which the log likelihood was not increased significantly. In the first run, we set the maximum number of variables to find the appropriate number of parameters by looking at changes in the log likelihood. Then, we reset the maximum subset size to this value and made a second run. Next, the variables of primary tumor size, clinical lymph node status, histologic grade, ER status and HER2 status were included in the model. The IRR and 95% CI for each variable were calculated. Then, the model was applied to the validation group by incorporating the eligible variables and calculating each individual patient's probability of having 0–2 ALN metastases. Finally, ROC analysis was used to evaluate the accuracy and discriminative ability of the models in the training cohort and in the validation cohort. DCA was also performed to estimate the clinical usefulness and benefits⁴⁴. Statistical analyses were performed with the JMP statistical software version 13.2.0 (SAS Institute Inc., Cary, NC, USA), the NCSS 11 statistical software (2016) (NCSS, LLC. Kaysville, Utah, USA, ncss.com/software/ncss), and R software with the rms package (version 3.5.2, <http://www.r-project.org>). P values less than 0.05 from a two-sided test were considered statistically significant.

Ethical approval and informed consent. This study was reviewed and approved by the Institutional Review Board of the Cancer Hospital of SUMC and the First Affiliated Hospital of SUMC. It was performed in accordance with the ethical standards laid down in the 1964 declaration of Helsinki and all subsequent revisions. All persons mentioned in the paper gave their informed consent prior to inclusion in the study.

Data availability

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Received: 21 March 2020; Accepted: 2 December 2020

Published online: 14 December 2020

References

- Vinh-Hung, V. *et al.* Ratios of involved nodes in early breast cancer. *Breast Cancer Res.* **6**, R680–688. <https://doi.org/10.1186/bcr934> (2004).
- Colleoni, M. *et al.* Size of breast cancer metastases in axillary lymph nodes: clinical relevance of minimal lymph node involvement. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* **23**, 1379–1389. <https://doi.org/10.1200/JCO.2005.07.094> (2005).
- Tseng, H. S. *et al.* Tumor characteristics of breast cancer in predicting axillary lymph node metastasis. *Med. Sci. Monit. Int. Med. J. Exp. Clin. Res.* **20**, 1155–1161. <https://doi.org/10.12659/MSM.890491> (2014).
- Krag, D. N. *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.* **8**, 881–888. [https://doi.org/10.1016/S1470-2045\(07\)70278-4](https://doi.org/10.1016/S1470-2045(07)70278-4) (2007).
- Carlson, G. W. & Wood, W. C. Management of axillary lymph node metastasis in breast cancer: making progress. *JAMA* **305**, 606–607. <https://doi.org/10.1001/jama.2011.131> (2011).
- Giuliano, A. E. *et al.* Axillary dissection vs no axillary dissection in women with invasive breast cancer and sentinel node metastasis: a randomized clinical trial. *JAMA* **305**, 569–575. <https://doi.org/10.1001/jama.2011.90> (2011).
- Wu, S. G. *et al.* Number of negative lymph nodes is associated with disease-free survival in patients with breast cancer. *BMC Cancer* **15**, 43. <https://doi.org/10.1186/s12885-015-1061-z> (2015).
- Prafulla Kumar Swain, P. K., Gurprit Grover, G., Sangeeta Chakravorty, S., Goel, K. & Singh, V. Estimation of number of involved lymph nodes in breast cancer patients using bayesian regression approach. *J. Stat. Appl. Probab. Lett.* **4**, 17–25. <https://doi.org/10.18576/jsapl/040103> (2017).
- Bevilacqua, J. L. *et al.* Doctor, what are my chances of having a positive sentinel node? A validated nomogram for risk estimation. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* **25**, 3670–3679. <https://doi.org/10.1200/JCO.2006.08.8013> (2007).
- Viale, G. *et al.* Predicting the status of axillary sentinel lymph nodes in 4351 patients with invasive breast carcinoma treated in a single institution. *Cancer* **103**, 492–500. <https://doi.org/10.1002/cncr.20809> (2005).
- Nori, J. *et al.* Role of axillary ultrasound examination in the selection of breast cancer patients for sentinel node biopsy. *Am. J. Surg.* **193**, 16–20. <https://doi.org/10.1016/j.amjsurg.2006.02.021> (2007).
- Lannig, C., Hoffmann, J., Galatius, H. & Engel, U. Assessment of clinical palpation of the axilla as a criterion for performing the sentinel node procedure in breast cancer. *Eur. J. Surg. Oncol. J. Eur. Soc. Surg. Oncol. Br. Assoc. Surg. Oncol.* **33**, 281–284. <https://doi.org/10.1016/j.ejso.2006.09.032> (2007).
- Specht, M. C., Fey, J. V., Borgen, P. I. & Cody, H. S. 3rd. Is the clinically positive axilla in breast cancer really a contraindication to sentinel lymph node biopsy?. *J. Am. Coll. Surg.* **200**, 10–14. <https://doi.org/10.1016/j.jamcollsurg.2004.09.010> (2005).
- Chu, K. U. *et al.* Do all patients with sentinel node metastasis from breast carcinoma need complete axillary node dissection?. *Ann. Surg.* **229**, 536–541. <https://doi.org/10.1097/0000658-199904000-00013> (1999).
- McMasters, K. M. *et al.* Sentinel-lymph-node biopsy for breast cancer—not yet the standard of care. *N. Engl. J. Med.* **339**, 990–995. <https://doi.org/10.1056/NEJM199810013391410> (1998).
- Giuliano, A. E. *et al.* Locoregional recurrence after sentinel lymph node dissection with or without axillary dissection in patients with sentinel lymph node metastases: long-term follow-up from the American College of Surgeons Oncology Group (Alliance) ACOSOG Z0011 randomized trial. *Ann. Surg.* **264**, 413–420. <https://doi.org/10.1097/SLA.0000000000001863> (2016).
- Berg, W. A. *et al.* Diagnostic accuracy of mammography, clinical examination, US, and MR imaging in preoperative assessment of breast cancer. *Radiology* **233**, 830–849. <https://doi.org/10.1148/radiol.2333031484> (2004).
- Chen, C. F. *et al.* Predictive value of preoperative multidetector-row computed tomography for axillary lymph nodes metastasis in patients with breast cancer. *Front. Oncol.* **8**, 666. <https://doi.org/10.3389/fonc.2018.00666> (2018).
- Kim, G. R. *et al.* Preoperative axillary US in early-stage breast cancer: potential to prevent unnecessary axillary lymph node dissection. *Radiology* **288**, 55–63. <https://doi.org/10.1148/radiol.2018171987> (2018).
- Kendal, W. S. The number distribution for involved lymph nodes in cancer. *Math. Biosci.* **205**, 32–43. <https://doi.org/10.1016/j.mbs.2006.01.003> (2007).
- O'Brein, C. M. Negative binomial regression, second edition by Joseph M. Hilbe. *Int. Stat. Rev.* **79**, 483–484 (2011).
- Kendal, W. S. Statistical kinematics of axillary nodal metastases in breast carcinoma. *Clin. Exp. Metas.* **22**, 177–183. <https://doi.org/10.1007/s10585-005-7211-y> (2005).
- Guern, A. S. & Vinh-Hung, V. Statistical distribution of involved axillary lymph nodes in breast cancer. *Bull. Cancer* **95**, 449–455. <https://doi.org/10.1684/bdc.2008.0620> (2008).
- Quan, M. L. & McCready, D. The evolution of lymph node assessment in breast cancer. *J. Surg. Oncol.* **99**, 194–198. <https://doi.org/10.1002/jso.21201> (2009).
- Bori, R., Vinh-Hung, V., Vajda, K., Svebis, M. & Cserni, G. The impact of the largest metastasis size on nodal tumor burden in colorectal carcinomas: implications for the sentinel lymph node theory in cancers of the large intestine. *J. Surg. Oncol.* **95**, 629–634. <https://doi.org/10.1002/jso.20725> (2007).
- Harden, S. P., Neal, A. J., Al-Nasiri, N., Ashley, S. & Querci della Rovere, G. Predicting axillary lymph node metastases in patients with T1 infiltrating ductal carcinoma of the breast. *Breast* **10**, 155–159. <https://doi.org/10.1054/brst.2000.0220> (2001).
- Coombs, N., Chen, W., Taylor, R. & Boyages, J. A decision tool for predicting sentinel node accuracy from breast tumor size and grade. *Breast J.* **13**, 593–598. <https://doi.org/10.1111/j.1524-4741.2007.00507.x> (2007).
- Fisher, B., Wolmark, N., Redmond, C., Deutsch, M. & Fisher, E. R. Findings from NSABP Protocol No. B-04: comparison of radical mastectomy with alternative treatments. II. The clinical and biologic significance of medial-central breast cancers. *Cancer* **48**, 1863–1872. [https://doi.org/10.1002/1097-0142\(19811015\)48:8<3c1863::aid-cnrc2820480825%3e3.0.co;2-u](https://doi.org/10.1002/1097-0142(19811015)48:8<3c1863::aid-cnrc2820480825%3e3.0.co;2-u) (1981).
- Fisher, B. *et al.* Twenty-five-year follow-up of a randomized trial comparing radical mastectomy, total mastectomy, and total mastectomy followed by irradiation. *N. Engl. J. Med.* **347**, 567–575. <https://doi.org/10.1056/NEJMoa020128> (2002).
- Cote, R. *et al.* ACOSOG Z0010: a multicenter prognostic study of sentinel node (SN) and bone marrow (BM) micrometastases in women with clinical T1/T2 N0 M0 breast cancer. *J. Clin. Oncol.* https://doi.org/10.1200/jco.2010.28.18_suppl.cra504 (2010).
- Krag, D. N. *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.* **11**, 927–933. [https://doi.org/10.1016/S1470-2045\(10\)70207-2](https://doi.org/10.1016/S1470-2045(10)70207-2) (2010).
- Galimberti, V. *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. *Lancet Oncol.* **19**, 1385–1393. [https://doi.org/10.1016/S1470-2045\(18\)30380-2](https://doi.org/10.1016/S1470-2045(18)30380-2) (2018).
- Giuliano, A. E. *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* **318**, 918–926. <https://doi.org/10.1001/jama.2017.11470> (2017).
- Donker, M. *et al.* Radiotherapy or surgery of the axilla after a positive sentinel node in breast cancer (EORTC 10981-22023 AMAROS): a randomised, multicentre, open-label, phase 3 non-inferiority trial. *Lancet Oncol.* **15**, 1303–1310. [https://doi.org/10.1016/S1470-2045\(14\)70460-7](https://doi.org/10.1016/S1470-2045(14)70460-7) (2014).
- Chen, K., Liu, J., Li, S. & Jacobs, L. Development of nomograms to predict axillary lymph node status in breast cancer patients. *BMC Cancer* **17**, 561. <https://doi.org/10.1186/s12885-017-3535-7> (2017).

36. Giuliano, A. E. *et al.* Locoregional recurrence after sentinel lymph node dissection with or without axillary dissection in patients with sentinel lymph node metastases: the American College of Surgeons Oncology Group Z0011 randomized trial. *Ann. Surg.* **252**, 426–432; discussion 432–423. <https://doi.org/10.1097/SLA.0b013e3181f08f32> (2010).
37. Chen, C. *et al.* Decrease in the Ki67 index during neoadjuvant chemotherapy predicts favorable relapse-free survival in patients with locally advanced breast cancer. *Cancer Biol. Med.* **16**, 575–586. <https://doi.org/10.20892/j.issn.2095-3941.2018.0423> (2019).
38. Elston, C. W. & Ellis, I. O. 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* **19**, 403–410. <https://doi.org/10.1111/j.1365-2559.1991.tb00229.x> (1991).
39. Ogawa, Y. *et al.* Immunohistochemical assessment for estrogen receptor and progesterone receptor status in breast cancer: analysis for a cut-off point as the predictor for endocrine therapy. *Breast Cancer* **11**, 267–275 (2004).
40. Wolff, A. C. *et al.* Recommendations for human epidermal growth factor receptor 2 testing in breast cancer: American Society of Clinical Oncology/College of American Pathologists clinical practice guideline update. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* **31**, 3997–4013. <https://doi.org/10.1200/JCO.2013.50.9984> (2013).
41. Dowsett, M. *et al.* Assessment of Ki67 in breast cancer: recommendations from the International Ki67 in Breast Cancer working group. *J. Natl. Cancer Inst.* **103**, 1656–1664. <https://doi.org/10.1093/jnci/djr393> (2011).
42. Cheang, M. C. *et al.* Ki67 index, HER2 status, and prognosis of patients with luminal B breast cancer. *J. Natl. Cancer Inst.* **101**, 736–750. <https://doi.org/10.1093/jnci/djp082> (2009).
43. Cameron, A. C. & Trivedi, P. K. *Regression Analysis of Count Data* 2nd edn. (Cambridge University Press, Cambridge, 2013).
44. Vickers, A. J. & Elkin, E. B. Decision curve analysis: a novel method for evaluating prediction models. *Med. Decis. Mak. Int. J. Soc. Med. Decis. Mak.* **26**, 565–574. <https://doi.org/10.1177/0272989X06295361> (2006).

Acknowledgements

This work was supported by the Natural Science Foundation of Guangdong Province, China (Grant Nos. 2018A030313562, 2019A1515010239), the Science and Technology Project of Shantou, China (Grant Nos. 190816095262235, 190819145262877 and 170828211930352), the Youth Scientific Research Fund of Cancer Hospital affiliated to Shantou University Medical College (No. 2020A003), the 2018 Special Fund for Science and Technology Innovation Strategy of Guangdong Province (No. Shantou Government 2018-49), and the 2019 Science and Technology Special Fund of Guangdong Province (No. Shantou Government 2019-132).

Author contributions

C.C. and D.Z. conceived and designed the project. Y.X. and Y.Z. collected the patients' characteristic data. J.W., Y.X. and H.L. prepared the figures and tables. C.C., Y.Z. and K.L. analyzed and interpreted the data. C.C., H.L. and D.Z. wrote the manuscript. C.C. and D.Z. approved the final version to be submitted. All authors wrote informed consent and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-020-79016-4>.

Correspondence and requests for materials should be addressed to C.-F.C.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020