Abstract

Objective: Examining the correct number of lymph nodes when diagnosing breast cancer invasion is still a problem. This work aimed to develop a qualification model that estimates the possibility of missing nodes and the number of lymph nodes that need to be examined.

Methods: By analyzing lymph node invasion of 303,760 breast cancer samples with primary tumor stage and the number of examined and positive lymph nodes from the Surveillance, Epidemiology and End Results database using a beta-binomial model, the number of nodes that should be examined was quantified in different stages.

Results: In general, to reduce the possibility of missing positive nodes to less than 10%, 21 lymph nodes should be examined; thus, the current median of dissected nodes (12) is not adequate. The number of nodes needed to be dissected for stages T1, T2, and T3 are 8, 37, and 87, respectively. Currently, the median number of node dissections for these stages were 12, 13, and 14, respectively. The clinical significance of the nodal staging score was validated with survival information.

Conclusion: Currently, the number of lymph nodes dissected in breast cancer are excessive for T1 but insufficient for T2 and T3.

Keywords

Lymph nodes, breast cancer, examination, quantification, invasion, metastasis

Date received: 7 April 2019; accepted: 10 September 2019

Introduction

Breast cancer is the most prevalent cancer among women, with 272,400 new cases and 70,700 deaths in China in 2015.1 As an important adjuvant therapy, radiation is currently recommended for breast cancer subtypes,2 and use of radiation significantly reduces
recurrence, metastasis, and prolongs the survival of breast cancer patients. However, drawbacks of radiation have also been widely reported, including impaired fertility, subsequent malignancies, and lung metastasis. Although some genome-based radiation assessments have been reported and used, currently, use of radiation is largely dependent on detecting lymph node invasion. It has been reported that radiation efficacy is significantly associated with lymph node status.

As lymph node invasion status is the most important clinical indicator for radiation and other adjuvant therapies after surgery, adequate removal of lymph nodes is needed for more accurate diagnosis. However, the retrieving of nodes has significant long-term complications, including lymphedema, sensory morbidity, and restriction of shoulder function. Thus, a balance between diagnostic accuracy and quality of life should be considered. The optimal number of nodes to examine has not been quantitatively defined.

In this study, using a straightforward beta-binominal model and the number of examined lymph nodes and invaded nodes in the Surveillance, Epidemiology and End Results (SEER) database (N=303,760), the relationship between possibly missing invasion-positive lymph nodes and primary tumor stage was estimated. According to the results, the minimum number of nodes that should be examined for stages pT1 to pT2 were 8 and 37, respectively; the effect of nodal dissection in pT3 was unclear and it was not recommended for pT4 stage breast cancer.

**Methods and materials**

**Data collection**

The SEER database includes 17 registries covering 26% of cancer patients in the United States. Breast cancer was selected, and only breast cancer as the primary malignancy was considered to exclude metastases of other cancer types. Patients without records of primary stage, nodal status, regional nodes examined, or positive nodes were excluded. Furthermore, insufficient records or those missing data were also excluded. The stages T1a-c, T2a-c, T3a-c, and T4a-c were converted to T1, T2, T3, and T4, respectively. In total, 303,760 patients were enrolled in this study.

**Hypotheses and calculation processes**

A mathematical model called beta-binomial distribution was used to fit SEER data and to calculate the possibility of missing invasion-positive lymph nodes according to the number dissected and examined. In the following text, true positive means that a lymph node was invaded, and the results showed breast cancer cells had invaded at least one lymph node. True negative refers to when both the true status and result were invasion-negative. False negative is defined as when the true status was invaded, while the results showed no invasion.

To simplify the model, the following hypotheses were employed:

1. All examination results of lymph nodes were correct.
2. The distribution of lymph nodes is equal, i.e., the lymph nodes were exchangeable. Biologically, this assumption is incorrect; however, if a large cohort is employed, the assumption is reasonable, as stated by most previous studies.
3. The sensitivity of true positive is equal to false negative. Sensitivities can only be calculated in node-positive samples.

**Model development and validation steps:**

1. The number of invasion-positive lymph nodes (non-N0 stage) and the number of total nodes dissected and examined was used to estimate the parameters of the beta-binomial (α and β) distribution.
In this procedure, only samples in which at least two lymph nodes were examined were used, because by definition, the node-invasive rate would reach 100% if only one lymph node was examined. Breast cancers were categorized into stages T1 to T4.

2. False-negative rates were estimated according to the model and estimated coefficients in the overall datasets and sub-datasets (different primary tumor size), and the observed and corrected prevalence was calculated using the following functions:

\[
FN_{adj,i} = \frac{1 - FN_i \times TP_{adj,i}}{FN_i}
\]

\[
\text{Prev}(pT)_{adj} = \sum \frac{(FN_i + TP_i)}{(FN_i + TP_i + TN_i)}
\]

\[
\text{NSS} = \frac{1 - \text{Pre}(T)}{1 - \text{Prev}(T) + [\text{Prev}(T) \times FN_i]}
\]

where \(FN_{adj,k}\) indicates adjusted false negative values; \(FN\) is observed false negatives; \(TP_{adj,k}\) is the number of true positives; \(T\) indicates primary tumor stage; and \(FN, TP,\) and \(TN\) represent false negatives, true positives, and true negatives, respectively.

3. Overall survival information is not related to lymph nodes dissected or nodal staging score, so it was used for validation. In this step, tumors with no positive lymph nodes were used. Samples in each category (T1, T2, T3, and T4) were divided based on the quartile of the nodal staging score (NSS), and then survival was compared among groups.

**Software**

All analyses in this study were performed using R software (v3.1.0) and R packages. The VGAM (v1.0-3) and bbmle (v1.0.18) packages were used to evaluate \(\alpha\) and \(\beta\) parameters of the beta-binomial model. The survival analysis was performed with R package “survival” using the Kaplan–Meier method.

**Results**

**Data profile**

After removing cases with unavailable data from the SEER database, 303,760 samples were incorporated into this study. Detailed information of primary tumor stage, nodal invasion, and nodes examined are shown in Table 1. The primary tumor stages for most patients were T1 and T2, which accounted for 93.24% of all samples. The mean nodes examined in T1 to T4 ranged from 12.1 to 15.1, and the median nodes examined ranged from 12 to 14.

**False negative diagnosis in overall data**

Using our model, the two parameters for the beta-binominal model, \(\alpha\) and \(\beta\), were estimated to be 0.933 (95% confidence interval: 0.924–0.942) and 1.917 (95% confidence interval: 1.895–1.939), respectively. As shown in Figure 1, the overall probability of missing nodal invasion (1-sensitivity) as a function of the number of examined nodes was calculated. As expected, the probability of missing nodal invasion decreased as the number of examined nodes increased. When only one lymph node was examined, the probability of missing nodal invasion was 67.25%. The minimum number of nodes that were needed to be examined was nine to reach a less than 20% chance of missing invasion, and 21/45 for 10%/5%, respectively. The current median value of examined nodes was 12, and the corresponding probability of missing nodal invasion was 15.52%, suggesting...
Table 1. General information of included samples.

| Stage     | Patients | Nodes examined |
|-----------|----------|----------------|
|           | No.      | %              | Mean | SD  | Median | IQR  |
| T stage   |          |                |      |     |        |      |
| (all patients) |        |                |      |     |        |      |
| 1         | 192453   | 63.35693       | 12.1 | 7.5 | 12     | 6–17 |
| 2         | 90766    | 29.88083       | 13.9 | 7.5 | 13     | 9–18 |
| 3         | 13111    | 4.316237       | 15.1 | 7.9 | 14     | 10–19|
| 4         | 7430     | 2.44601        | 13.7 | 8   | 13     | 8–18 |
| T stage   |          |                |      |     |        |      |
| (N0 patients) |        |                |      |     |        |      |
| 1         | 147272   | 48.48301       | 11.4 | 7.4 | 11     | 5–16 |
| 2         | 44148    | 14.53384       | 12.7 | 7.5 | 12     | 7–17 |
| 3         | 3929     | 1.293455       | 13.3 | 7.5 | 13     | 8–18 |
| 4         | 1403     | 0.461878       | 12.2 | 7.3 | 11     | 7–16 |
| N stage   |          |                |      |     |        |      |
| 0         | 196752   | 64.77219       | 11.8 | 7.5 | 11     | 6–16 |
| 1         | 101939   | 33.55906       | 14.6 | 7.4 | 14     | 10–19|
| 2         | 4720     | 1.553858       | 16.1 | 8.5 | 15     | 10–21|
| 3         | 349      | 0.114893       | 16.3 | 9.7 | 15     | 10–21|

Figure 1. Sensitivity of nodal diagnosis according to the number of examined nodes. The x- and y-axes indicate the number of nodes examined and probability of missing positive nodes, respectively.
that the current number of nodes examined is inadequate.

False negative diagnosis in different primary stages

The stage of the primary tumor is the most important indicator for nodal examination (Figure 2). The probability of missing nodal invasion was also estimated in different primary tumor TNM stages (T1–T4). To minimize the probability of missing nodal invasion to less than 20%, the least number of nodes to be examined are 3, 15, 36 and 65 for T1, T2, T3, and T4, respectively. While to reach 10%, the number of nodes were 8, 37, and 87 for T1, T2, and T3, respectively. When the number of examined nodes is 12 (the current median value of examined nodes), the probabilities of missing nodal invasion were 7.08%, 21.49%, 38.40%, and 452.00% for T1, T2, T3, and T4, respectively. From the SEER database, the median number of examined nodes in T1, T2, T3, and T4 were 12, 13, 14 and 13 (Table 1), which had corresponding probabilities of missing nodal invasion of 7.08%, 20.42%, 35.47%, and 50.38%, respectively.

Corrected rates were calculated by combining the current rates of positive nodes in different primary tumor stages and our results (Table 2). The corrected node-positive rates were higher than observed rates due to false negatives. These results indicated that the node counts of currently implemented methods are adequate for pT1, but too low for pT2.

Nodal staging score and survival

Follow-up information was not used in our model and was independent from our nodal
staging score. The nodal staging score in samples diagnosed as N0 stage at primary diagnosis was divided by quantiles, and survival differences were compared. Nodal staging score was significantly associated with survival in pT1N0 patients and the pT2N0 group, while not significant in pT3N0 patients and adversely associated with nodal staging score in the pT4N0 group (Figure 3).

### Discussion

Lymph node invasion is highly associated with distant recurrence and relapse in cancers.\textsuperscript{13,14} For breast cancer, it is critical for
therapeutic decisions, including radiation therapy, chemotherapy, and follow up.\textsuperscript{15,16} While the number of lymph nodes examined is controversial for patients: inadequate nodal examination may lead to missing nodal invasion, while excessive nodal examination reduces immune activity and has long term complications.\textsuperscript{11} Thus, balancing diagnostic accuracy with the drawbacks of lymph node removal is critical for improving survival times and the life quality of breast cancer patients.

To investigate whether the current number of nodes removed is reasonable, we fitted the beta-binominal model on the SEER database (303,760 samples). To our knowledge, this is the first study of the number of nodes required to make an accurate diagnosis of nodal invasion status. In this model, a few hypotheses are employed. First, the beta-binominal distribution of nodal invasion was used. According to previous reports, no issues with this hypothesis have been reported.\textsuperscript{12,17–19} Another assumption in this study is that all examined nodes were correct, i.e., no false negatives or false positives were detected. This is reasonable, as all diagnoses were made by experienced pathologist, according to the SEER database.

Our results showed that the probability of missing nodal invasion using the current median examined nodes (12) was 15.52%. To reach an accuracy of 90%, 21 nodes are needed for breast cancer patients when primary tumor stage was not considered. A less than 10% probability of missing nodal invasion demands eight, 37 and 87 lymph nodes be examined in pT1, pT2, and pT3 disease, respectively. Thus, the current median number of examined nodes is sufficient for pT1, but the probability of missing nodal invasion is >20% for pT2 to pT4. This explains why the missing nodal status of pT1 was much less than pT2 and pT3. It is noteworthy that 69.52% and 80.19% of pT3 and pT4 stage breast cancer (corrected rates: 80.06% and 87.46%), respectively, have observed lymph node invasion, indicating that lymph node invasion is common in pT3 and pT4. Additionally, the drawbacks of removing and examining lymph nodes are exaggerated, while the necessity of diagnosing nodal status is rapidly reduced in these patients. This explains why the nodal staging score was not significantly associated with overall survival in pN0T3 patients and inversely correlated with the survival of pN0T4 patients. Thus, nodal examination is recommended for pT1 and pT2 patients, but not for pT3 and pT4 patients, based on there being no benefit of nodal examination in the latter populations.

There were some limitation of this study. Although this was a multicenter study, it was also retrospective in nature. Some important clinical variables were not available for this study due to the lack of enrollment control, and some bias may have existed in patient selection. For example, the distribution of primary tumor stages was unbalanced. Most patients were pT1 and pT2 (N=283,219, 93.24%), while the samples size of the pT4 cohort was 7430 (0.024%).

**Conclusion**

Our study investigated the number of lymph nodes needed to be removed for optimal diagnosis and found that the current number (12 nodes) is sufficient for pT1 patients (eight recommended), but inadequate for pT2 patients (37 recommended). Nodal examination is not recommended for stages pT3 and pT4 based on the prevalence of nodal invasion.

**Declaration of conflicting interest**

The authors declare that there is no conflict of interest.
Funding
National Key R&D Program of China (2018YFB1307700) to Liping Sun.

Ethics and consent
All data used in this study are fully publicly accessible, and the authors did not participate in data collection and generation. Thus, written consent is not necessary.

ORCID iD
Liping Sun https://orcid.org/0000-0003-1642-0168

References
1. Chen W, Zheng R, Baade PD, et al. Cancer statistics in China, 2015. CA Cancer J Clin 2016; 66: 115–132.
2. Moran MS, Schnitt SJ, Giuliano AE, et al. Society of Surgical Oncology-American Society for Radiation Oncology consensus guideline on margins for breast-conserving surgery with whole-breast irradiation in stages I and II invasive breast cancer. J Clin Oncol 2014; 32: 1507–1515.
3. Refaat T, Sachdev S, Sathiaseelan V, et al. Hyperthermia and radiation therapy for locally advanced or recurrent breast cancer. Breast 2015; 24: 418–425.
4. Wu SG, Chen Y, Sun JY, et al. Using the lymph nodal ratio to predict the risk of locoregional recurrence in lymph node-positive breast cancer patients treated with mastectomy without radiation therapy. Radiat Oncol 2013; 8: 119.
5. Cleary SF, Liu LM, Graham R, et al. In vitro fertilization of mouse ova by spermatozoa exposed isothermally to radiofrequency radiation. Bioelectromagnetics 1989; 10: 361–369.
6. Jin T, Song T, Deng S, et al. Radiation-induced secondary malignancy in prostate cancer: a systematic review and meta-analysis. Urol Int 2014; 93: 279–288.
7. Koo E, Henderson MA, Dwyer M, et al. Management and prevention of breast cancer after radiation to the chest for childhood, adolescent, and young adulthood malignancy. Ann Surg Oncol 2015; 22: S545–S551.
8. Feyes L, Descamps B, Vanhove C, et al. Radiation-induced lung damage promotes breast cancer lung metastasis through CXCR4 signaling. Oncotarget 2015; 6: 26615–26632.
9. Fisher B, Redmond C, Fisher ER, et al. Ten-year results of a randomized clinical trial comparing radical mastectomy and total mastectomy with or without radiation. N Engl J Med 1985; 312: 674–681.
10. Wu SG, Sun JY, Zhou J, et al. Number of negative lymph nodes can predict survival of breast cancer patients with four or more positive lymph nodes after postmastectomy radiotherapy. Radiat Oncol 2014; 9: 284.
11. Pesce C and Morrow M. The need for lymph node dissection in nonmetastatic breast cancer. Annu Rev Med 2013; 64: 119–129.
12. Gonen M, Schrag D and Weiser MR. Nodal staging score: a tool to assess adequate staging of node-negative colon cancer. J Clin Oncol 2009; 27: 6166–6171.
13. Fujii T, Yajima R, Tatsuki H, et al. Significance of lymphatic invasion combined with size of primary tumor for predicting sentinel lymph node metastasis in patients with breast cancer. Anticancer Res 2015; 35: 3581–3584.
14. Ruiz U, Babeu S, Schwartz MS, et al. Blood vessel invasion and lymph node metastasis: two factors affecting survival in breast cancer. Surgery 1973; 73: 185–190.
15. Hamamoto Y, Ohsumi S, Aogi K, et al. Are there high-risk subgroups for isolated locoregional failure in patients who had T1/2 breast cancer with one to three positive lymph nodes and received mastectomy without radiotherapy? Breast Cancer 2014; 21: 177–182.
16. Citron ML, Berry DA, Cirrincione C, et al. Randomized trial of dose-dense versus conventionally scheduled and sequential versus concurrent combination chemotherapy as postoperative adjuvant treatment of node-positive primary breast cancer: first report of Intergroup Trial C9741/Cancer and Leukemia Group B Trial 9741. J Clin Oncol 2003; 21: 1431–1439.
17. Shariat SF, Ehdaie B, Rink M, et al. Clinical nodal staging scores for bladder cancer: a proposal for preoperative risk assessment. *Eur Urol* 2012; 61: 237–242.

18. Robinson TJ, Thomas S, Dinan MA, et al. How many lymph nodes are enough? Assessing the adequacy of lymph node yield for papillary thyroid cancer. *J Clin Oncol* 2016; 34: 3434–3439.

19. Kluth LA, Abdollah F, Xylinas E, et al. Pathologic nodal staging scores in patients treated with radical prostatectomy: a postoperative decision tool. *Eur Urol* 2014; 66: 439–446.