



# Development of a nomogram for preoperative prediction of lymph node metastasis in non-small cell lung cancer: a SEER-based study

Chufan Zhang<sup>1,2#</sup>, Qian Song<sup>1,2,3#</sup>, Lanlin Zhang<sup>1,2</sup>, Xianghua Wu<sup>1,2^</sup>

<sup>1</sup>Department of Medical Oncology, Fudan University Shanghai Cancer Center, Shanghai, China; <sup>2</sup>Department of Oncology, Shanghai Medical College, Fudan University, Shanghai, China; <sup>3</sup>Institute of Digestive Disease and Department of Medicine and Therapeutics, State Key Laboratory of Digestive Disease, Li Ka Shing Institute of Health Sciences, Shenzhen Research Institute, The Chinese University of Hong Kong, Hong Kong, China

**Contributions:** (I) Conception and design: C Zhang, X Wu; (II) Administrative support: C Zhang; (III) Provision of study materials or patients: C Zhang, Q Song; (IV) Collection and assembly of data: C Zhang, Q Song; (V) Data analysis and interpretation: C Zhang, L Zhang; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

<sup>#</sup>These authors contributed equally to this work.

**Correspondence to:** Xianghua Wu. Department of Medical Oncology, Fudan University Shanghai Cancer Center, 270 Dong-An Road, Shanghai 200032, China. Email: wuxianghua2018@163.com.

**Background:** Lymph node dissection is an important part of lung cancer surgery. Preoperational evaluation of lymph node metastases decides which dissection pattern should be chosen. The present study aimed to develop a nomogram to predict lymph node metastases on the basis of clinicopathological features of non-small cell lung cancer (NSCLC) patients.

**Methods:** A total of 35,138 patients diagnosed with NSCLC from 2010–2015 were selected from the Surveillance, Epidemiology, and End Results (SEER) database. Patients were randomly divided into training cohort and validation cohort. Possible risk factors were included and analyzed by logistic regression models. A nomogram was then constructed and validated.

**Results:** 21.83% of all patients were confirmed with positive lymph node metastasis. Age at diagnosis, sex, stage, T status, tumor size, grade and laterality were identified as predicting factors for lymph node involvement. These variables were included to build the nomogram. The AUC of the model was 0.696 (95% CI, 0.617 to 0.775). The model was further validated in the validation set with AUC 0.693 (95% CI, 0.628 to 0.758). The model presented with good prediction accuracy in both training cohort and validation cohort.

**Conclusions:** We developed a convenient clinical prediction model for regional lymph node metastases in NSCLC patients. The nomogram will help physicians to determine which patients will receive the most benefit from lymph node dissection.

**Keywords:** Non-small cell lung cancer (NSCLC); lymph node metastasis; Surveillance, Epidemiology, and End Results (SEER); nomogram

Submitted Jan 17, 2020. Accepted for publication Jun 02, 2020.

doi: 10.21037/jtd-20-601

**View this article at:** <http://dx.doi.org/10.21037/jtd-20-601>

## Introduction

Lung cancer remains the most frequent malignancy and

one of the leading causes of cancer-specific morbidity and mortality among men and women worldwide (1). Non-small cell lung cancer (NSCLC) is the most common

<sup>^</sup> ORCID: 0000-0001-6914-1598.

histology type and represents 85% of all newly diagnosed lung cancer cases (2). Five-year survival rate of lung cancer is approximately 17% for all stages according to the latest cancer statistics, even though targeted therapy and immunotherapy has started a revolution in management and achieved remarkable progress in recent years (3,4). Considering the relatively low survival rate, a better understanding of cancer biology, earlier and more precise diagnosis, improved staging system and novel therapy is all in urgent need for better outcomes of patients in the future.

Accurate staging is a crucial step after diagnosis to provide optimal treatment and metastases to lymph nodes is one of the key factors in the staging system which is closely related to prognosis (5). PET-CT scan has gradually become the most important examination for lymph node status. Nodes showing greater 18F-FDG uptake at PET without benign calcification or high attenuation >70 household unit (HU) at unenhanced CT were regarded as being positive for malignancy. PET-CT has improved specificity and negative predictive value and help surgeons to determine patients who will benefit the most from the surgery. However high specificity comes with a price of high rate of false positives and possible histological confirmation is recommended according to previous studies (6). Assessment of lymph node metastasis also relies on traditional computed tomography (CT) results and lymph nodes present with short-axis diameters of >1 cm are generally defined as metastatic lesions (7). However, predicting accuracy of CT scan is not satisfactory enough as sensitivity is 43–86% and specificity is 59–83% (8). Other invasive diagnostic methods such as mediastinoscopy and endoscopic ultrasound-guided fine-needle aspiration are valuable but not routinely used in clinical practice (9).

Complete surgery resection is standard care for early stage NSCLC patients confirmed with no distant metastasis (10). Pulmonary and mediastinal lymph nodes dissection is widely performed for patients suspected with lymph node metastases and is believed to be beneficial supported by abundant evidence (11,12). Since evaluation of CT results had its limitation, 15–17% of patients diagnosed as N0 via preoperative CT scan was proved to be N2 according to pathological examination while some other patients classified as N0 came back with positive lymph nodes (13). To decide the risk of lymph node involvement and therefore the extent of lymph node dissection during operation, accurate assessment before surgery become necessary.

Several factors have been revealed to be related with lymph node metastasis by previous studies, including age,

tumor size, pleural invasion etc. (14–16). However, there is no convenient clinical model for risk prediction so far. We intended to build a model with clinicopathological factors to predict the risk of lymph node metastasis for patients with NSCLC. In the present study, we did a retrospective research on data released from SEER database to identify clinicopathologic features correlating with lymph node metastasis in NSCLC patients. Afterwards a nomogram was developed for the first time based on our knowledge of literature to predict the risk of lymph node involvement prior to surgical intervention. We present the following article in accordance with the STROBE reporting checklist (available at <http://dx.doi.org/10.21037/jtd-20-601>).

## Methods

### *Ethical statement*

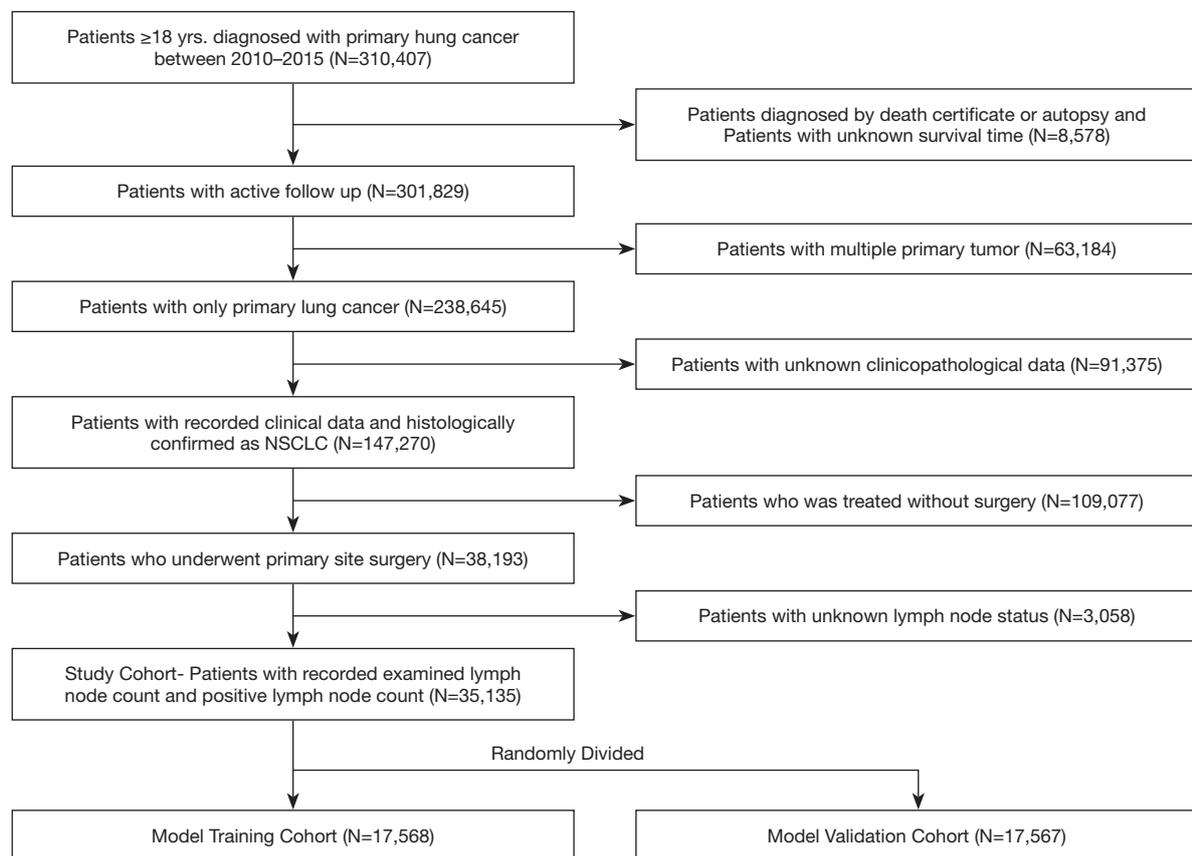
The trial was conducted in accordance with the Declaration of Helsinki and the Harmonized Tripartite Guideline for Good Clinical Practice from the International Conference on Harmonization. This study was approved by Ethical Committee of Shanghai Cancer Center, Fudan University. Patient consent form was not required for data from SEER database as all data was deidentified before release and contained no personally identifying information of patients. No specific funding from public, commercial agencies or individual was received in the present study.

### *Patient population*

Surveillance, Epidemiology, and End Results (SEER) database collects cancer patient information from 18 separate cancer registries all across the United States and covers about one quarter population of the whole country (17). The population entered the database are considered to be able to represent the overall population. SEER\*Stat version 8.3.5 was used to generate a case listing file. A general description of our study design was present in *Figure 1*.

### *Patient selection*

A total of 35,138 patients were selected into our study cohort. Inclusion criteria was as followed: (I) patients over 20 years old at diagnosis (II) patients diagnosed with primary lung cancer between 2010 to 2015 with site codes as C34.0–C34.9 (lung and bronchus). (III) patients who received a primary site surgery (IV) positive pathologic confirmation of histologic type



**Figure 1** Flow chart of patient selection from SEER database. Patients included in the study and patients excluded were indicated. NSCLC, non-small cell lung cancer.

as adenocarcinoma (8140–8147, 8255, 8260, 8310, 8323, 8480, 8481, 8490, 8550, 8572), squamous cell carcinoma (8050–8052, 8070–8078), or other (8010, 8012, 8014–8015, 8020–8022, 8030, 8036) (V) active follow up (VI) survival time over 1 month after surgery. Exclusion criteria included (I) diagnoses made by autopsy or death certificate (II) multiple primary tumor (III) no lymph node examined (VI) unknown disease stage at diagnosis. Patients with stage IIIB or stage IV disease were also excluded since surgery was no standard procedure for these patients. Clinicopathologic features extracted were age at diagnosis, sex, race, histology type, clinical stage, T status, tumor size, histologic grade, laterality. We randomly divided all population into two separate sets: a training set with 17,568 patients and a validation set with 17,567 patients.

### Statistical analysis

Statistical analyses were conducted using SPSS for windows (version 24.0, SPSS Inc., Chicago, IL, USA) and RStudio

for windows supported by R programming language and environment version 3.5.1. Pearson's chi-square test or Fisher's exact test when needed were applied to compare all variables between training set and validation set. To identify risk factors for lymph node metastasis, multivariable analysis was performed by binary logistic regression analysis. All features available for preoperative assessment were included into the multivariable analysis.

A logistic regression model-based nomogram was developed using the training set. All variables with a P value <0.05 were included into model construction unless specifically stated otherwise. Nomogram performance was evaluated by both internal validation in the training set and external validation in the validation set. We used bootstrapping method (1,000 repetitions) to generate a calibration curve plotted with observed outcome frequencies and predicted probabilities of lymph node metastasis. Discrimination ability was measured by means of the area under the receiver operating characteristic (ROC) curve

(AUC or C-index). All tests were two-sided and P value <0.05 was considered as significant.

## Results

### Patient characteristics

The clinicopathologic features of all patients were listed in *Table 1*. After selection 35,138 patients were eligible for our study. For all patients, median follow-up time was 16 months (with a range from 1–44 months) 42.2% of all patients were over 70 years old and 57.8% were under. As for race, the majority of the population were white and other races counted for 15.87%. Patients distributed evenly over the diagnosis years. Male and Female patients distributed almost evenly with percentage of 48.32% and 51.68%. 68.03% of patients were histologically classified as adenocarcinoma and 29.16% were squamous cell carcinoma. 41.89% of all lung tumors were located on the left size and 36.76% were small lesions with a diameter under 20 mm. In post operational pathologic examination, 21.83% patients were found with positive lymph node metastasis. There was no significant difference between training and validation set which was consistent with our method of random sampling. We also investigated how different types of surgery will

influence the results of lymph node status. The results were presented in *Table 2*. Most patients went through Resection of one lobe or bilobectomy with mediastinal lymph node dissection, however this procedure held the lowest average positive rate.

### Identification of risk factors correlating with lymph node metastasis

To reveal factors which have influence on lymph node metastasis, we first performed a univariate analysis in both training and validation set as *Table 3* summarized the results. The results showed that factors most strongly related to lymph node metastasis were age at diagnosis, sex, stage, T status, tumor size, grade and laterality. Notably, histology type, which is usually an important indicator for staging and prognosis, had no significant influence on lymph node involvement.

Considering our predictive model will be used before surgery, we entered all features that would be available in preoperational evaluation into a binary logistic regression analysis. These factors included age, sex, race, histology, T status, tumor size, grade and laterality and analysis results were presented in *Table 4*. According to this multivariate

**Table 1** Clinicopathological characteristics of NSCLC patients in the training set and validation set

Characteristics	Training set		Validation set		All patients	
	N=17,568	%	N=17,569	%	N=35,137	%
Age, year						
≤50	661	3.76	699	3.98	1,360	3.87
51–60	3,014	17.16	2,970	16.90	5,984	17.03
61–70	6,447	36.70	6,520	37.11	12,967	36.90
71–80	6,022	34.28	5,937	33.79	11,959	34.04
>80	1,424	8.11	1,443	8.21	2,867	8.16
Race						
White	14,772	84.08	14,788	84.17	29,560	84.13
Black	1,535	8.74	1,595	9.08	3,130	8.91
Other	1,261	7.18	1,186	6.75	2,447	6.96
Sex						
Female	9,009	51.28	9,150	52.08	18,159	51.68
Male	8,559	48.72	8,419	47.92	16,978	48.32

**Table 1** (continued)

Table 1 (continued)

Characteristics	Training set		Validation set		All patients	
	N=17,568	%	N=17,569	%	N=35,137	%
Year of diagnosis						
2010	2,920	16.62	2,941	16.74	5,861	16.68
2011	2,876	16.37	3,038	17.29	5,914	16.78
2012	2,723	15.50	3,039	17.30	5,762	16.40
2013	2,960	16.85	2,981	16.97	5,941	16.91
2014	2,813	16.01	2,971	16.91	5,784	16.46
2015	2,892	16.46	3,034	17.27	6,016	16.87
Histology						
Adenocarcinoma	11,914	67.82	11,991	68.25	23,905	68.03
Squamous cell carcinoma	5,136	29.23	5,110	29.09	10,246	29.16
Other	518	2.95	468	2.66	986	2.81
Stage						
I	11,023	62.74	11,027	62.76	22,050	62.75
II	3,891	22.15	3,918	22.30	7,809	22.22
III	2,654	15.11	2,624	14.94	5,278	15.02
T status						
T1	7,909	45.02	8,006	45.57	15,915	45.29
T2	6,881	39.17	6,769	38.53	13,650	38.85
T3	2,259	12.86	2,269	12.91	4,528	12.89
T4	521	2.97	525	2.99	1,046	2.98
Size, mm						
≤20	6,446	36.69	6,471	36.83	12,917	36.76
20–30	3,743	21.31	3,829	21.79	7,572	21.55
≥30	7,379	42.00	7,269	41.37	14,648	41.69
Grade						
I	2,679	15.25	2,655	15.11	5,334	15.18
II	8,477	48.25	8,627	49.10	17,104	48.68
III	6,152	35.02	6,043	34.40	12,195	34.71
IV	260	1.48	244	1.39	504	1.43
Laterality						
Left	7,349	41.83	7,371	41.95	14,720	41.89
Right	10,219	58.17	10,198	58.05	20,417	58.11
Regional lymph node						
Positive	3,859	21.97	3,810	21.69	7,669	21.83
Negative	13,709	78.03	13,759	78.31	27,468	78.17

**Table 2** Lymph node status of different surgery procedures

Surgery type	Number	Lymph node examined (average)	Positive lymph node (average)	Positive rate (average)
All	35,144 (100%)	10.609	0.73	8.4%
Excision or resection of less than one lobe- Wedge resection	3,603 (10.25%)	5.72	0.28	8.5%
Resection of at least one lobe or bilobectomy	4,307 (12.25%)	7.86	0.72	11.4%
Resection of [at least one] lobe or bilobectomy WITH mediastinal lymph node dissection	23,171 (65.93%)	11.78	0.758	7.4%
Pneumonectomy	344 (0.97%)	13.38	2.151	20.3%
Pneumonectomy WITH mediastinal lymph node dissection	1,043 (2.97%)	17.358	2.29	16.5%
Others	2,674 (7.6%)	9.20	0.54	8.7%

analysis results, age, sex, T status, tumor size, grade and laterality were independent variables that cast significant influence on lymph node positivity and were furthered taken into nomogram development.

### *Nomogram construction*

We then constructed a nomogram on the basis of a logistic regression model developed by our training set data. Significant risk factors indicated by the logistic regression analysis were fit into the model. The final nomogram was presented in *Figure 2*.

### *Internal and external validation of the nomogram*

Validation was first performed in training set internally. The model had an AUC of 0.696 (95% CI, 0.617 to 0.775), which showed good discrimination. Bootstrapping method (1,000 repetitions) was used and a calibration curve was illustrated in *Figure 3*. There was no obvious deviation between model predicted risk and actual observed risk curve, meaning the model was well calibrated. We further validated the model in the validation set using the same method. Good calibration was observed and the AUC was 0.693 (95% CI, 0.628 to 0.758) demonstrating the nomogram was well fitted.

## **Discussion**

Due to rapid development of novel screening approaches such as PET-CT and widely applied annual health check,

more and more lung cancer patients were detected at an early stage of disease (18). Surgery remains the most effective and possible curable method for these patients and lobectomy with systematic lymph node dissection (LND) is the most common procedure. Traditionally, LND is believed to be capable of providing more accurate stage, detecting occult lesions and improving overall survival (19). However, with the appearance of more early stage lung cancer patients, how to minimize the surgical trauma and shorten the hospital stay should also be taken into consideration. Also, it is not a rare scenario that postoperative pathologic examination reveals no lymph node metastasis. Overly removal of intact lymph nodes will cause longer surgery time, more blood loss and impaired regional immune function which can be harmful for elderly patients and patients with defective pulmonary function (20,21). There is a rising opinion that selective lymph node dissection (SLND) may be a better alternative for systematic lymph node dissection. Han *et al.* proposed that SLND will be a vital component of minimally invasive surgical treatment and may lead to better life quality afterwards (22). Therefore, it's not hard to understand that preoperative lymph node status assessment is of great importance to determine the best lymph node dissection pattern.

Several factors were reported to be relevant to lymph node metastasis in lung cancer. Xia *et al.* revealed that young age increased risk of lymph node involvement and Ding *et al.* reported that tumor size, histologic differentiation and smoking status were correlated to LN positivity (14,23). So far there is no easy to use clinical tool to estimate individual risk of lymph node metastasis. Nomogram is a pictorial

**Table 3** Univariate analyses of risk factors related with regional lymph node metastasis in training set and validation set

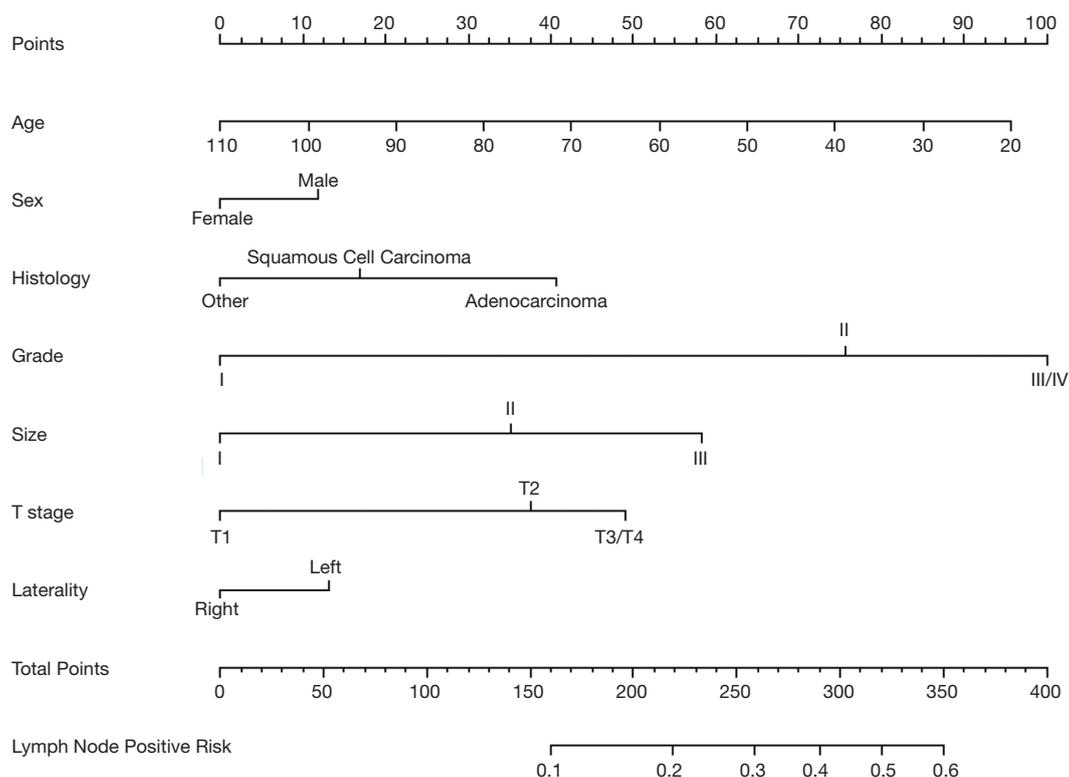
Variables	Training set					Validation set				
	LN metastasis (-)		LN metastasis (+)		P value	LN metastasis (-)		LN metastasis (+)		P value
	No	%	No.	%		No.	%	No.	%	
Age, year					<0.001					<0.001
≤50	471	3.4	190	4.9		483	3.51	216	5.67	
51–60	2,265	16.5	749	19.4		2,196	15.96	774	20.31	
61–70	5,030	36.7	1,417	36.7		5,122	7.233	1,398	36.69	
71–80	4,821	35.2	1,201	31.1		4,766	34.64	1,171	30.73	
>80	1,122	8.2	302	7.8		1,192	8.66	251	6.59	
Race					0.006					0.889
White	11,590	84.54	3,182	82.46		11,590	84.249	3,198	83.94	
Black	1,170	8.53	365	9.46		1,242	9.03	353	9.27	
Other	949	6.92	312	8.08		927	6.74	259	6.80	
Sex					<0.001					<0.001
Female	7,225	52.70	1,784	46.23		7,281	52.92	1,869	49.06	
Male	6,484	47.30	2,075	53.77		6,478	47.08	1,941	50.94	
Histology					0.212					0.522
Adenocarcinoma	9,342	68.15	2,572	66.65		9,418	68.45	2,573	67.53	
Squamous cell carcinoma	3,968	28.94	1,168	30.27		3,980	28.93	1,130	29.66	
Other	399	2.91	119	3.08		361	2.61	107	2.80	
Stage					<0.001					<0.001
I	11,023	80.41	0	0		11,027	80.14	0	0	
II	2,188	15.96	1,703	44.13		2,244	16.40	1,674	43.94	
III	498	3.63	2,156	55.87		488	3.55	2,136	56.06	
T stage					<0.001					<0.001
T1	6,896	50.30	1,011	26.20		6,952	50.53	1,054	27.66	
T2	4,910	35.82	1,971	51.08		4,859	35.32	1,910	50.13	
T3	1,527	11.14	732	18.97		1,569	11.40	700	18.37	
T4	376	2.74	145	3.76		379	2.75	146	3.83	
Size					<0.001					<0.001
≤20 mm (I)	5,675	41.40	771	19.88		5,682	41.30	789	20.70	
20–30 mm (II)	3,001	21.89	742	19.22		3,092	22.47	737	19.34	
≥30 mm (III)	5,033	36.71	2,346	60.79		4,985	36.23	2,284	59.95	
Grade					<0.001					<0.001
I	2,463	17.97	216	5.60		2,448	17.79	207	5.43	
II	6,700	48.87	1,777	46.04		6,853	49.81	1,774	46.56	
III	4,357	31.78	1,795	46.51		4,273	31.06	1,770	46.46	
IV	189	1.38	71	1.84		185	1.34	59	1.55	
Laterality					<0.001					<0.001
Left	5,598	40.83	1,751	45.37		5,600	40.70	1,771	46.48	
Right	8,111	59.17	2,108	54.63		8,159	59.30	2,039	53.52	

**Table 4** Multivariate analyses of risk factors related with regional lymph node metastasis in training set and validation set

Characteristics	$\beta$	Odds ratio (95% CI)	P value
Age	-0.014	0.986 (0.982–0.990)	<0.001
Race			
White	Reference		
Black	0.031	1.032 (0.906–1.175)	0.636
Other	0.134	1.143 (0.993–1.316)	0.063
Sex			
Male	Reference		
Female	-1.59	0.853 (0.791–0.920)	<0.001
Histology			
Other	Reference		
Adenocarcinoma	0.561	1.752 (1.375–2.232)	<0.001
Squamous cell carcinoma	0.240	1.272 (0.995–1.625)	0.054
T status			
T1	Reference		
T2	0.515	1.674 (1.505–1.864)	<0.001
T3	0.698	2.009 (1.766–2.285)	<0.001
T4	0.558	1.747 (1.407–2.168)	<0.001
Size			
$\leq 20$ mm	Reference		
20–30 mm	0.479	2.215 (1.979–2.480)	<0.001
$\geq 30$ mm	0.795	0.986 (0.982–0.990)	<0.001
Grade			
I	Reference		
II	1.037	2.821 (2.422–3.284)	<0.001
III	1.373	3.948 (3.378–4.614)	<0.001
IV	1.383	3.986 (2.820–5.633)	<0.001
Laterality			<0.001
Left	Reference		
Right	-0.180	0.835 (0.775–0.900)	<0.001

presentation of complex mathematic regression model. It has emerged has a simpler and more advanced method for clinical event prediction and gradually become an important component of decision making (24). Hence, we aimed to develop a nomogram combining clinicopathologic factors to serve as a comprehensive, user friendly, effective clinical prediction model.

In the present study, we utilized data released from large population cancer database SEER and 35,138 patients fit into our selection criteria mentioned above and was included into our study cohort. We first performed univariant and multivariant analysis and age, sex, T status, tumor size, grade and laterality were distinguished as independent variables involving the lymph node metastasis.



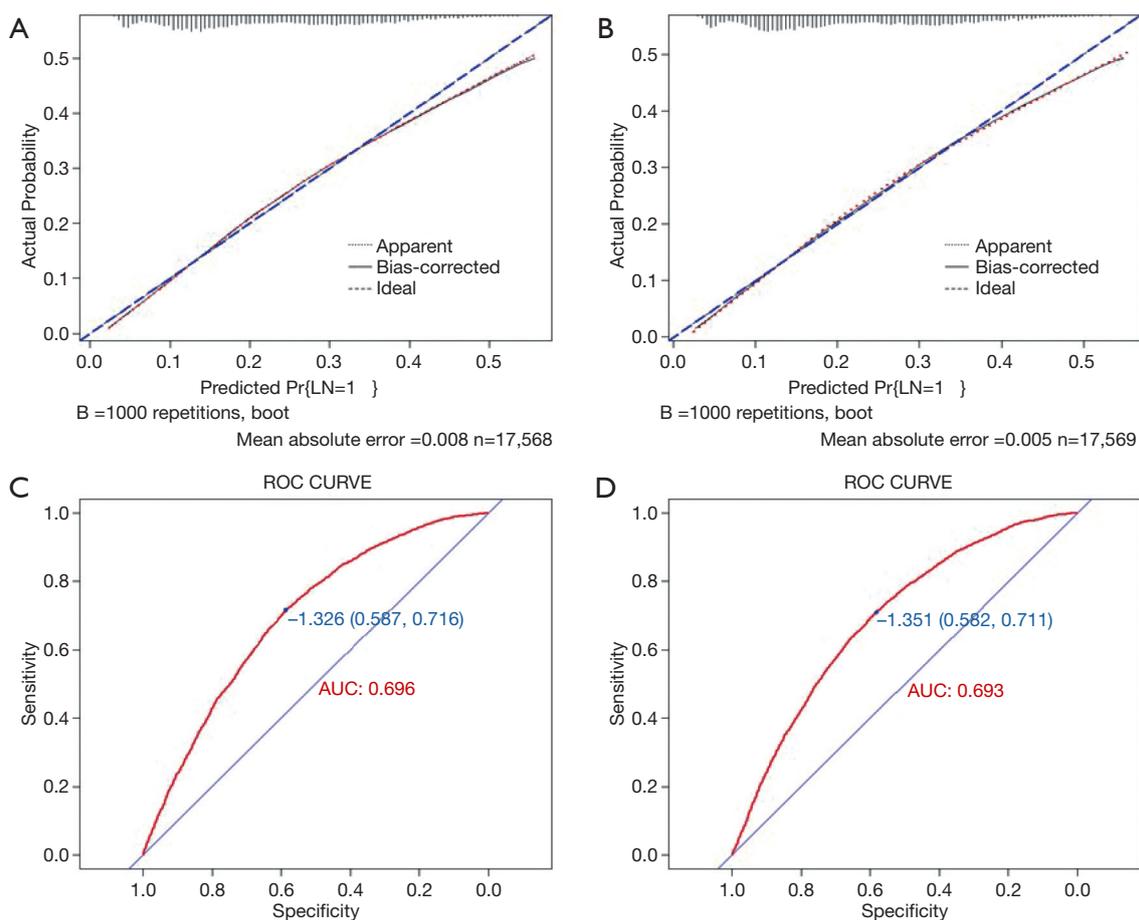
**Figure 2** Nomogram predicting lymph node metastasis in NSCLC patient. First row presented point assignment for each variable. Row 2–8 showed variables included into this model. When using the nomogram for an individual patient, every variable will be assigned with a point basing on clinicopathological features and all points will be summed up. Every score in total point of row 9 will be corresponding with a probability in the last row of risk. NSCLC, non-small cell lung cancer.

Patients with younger age, poorly differentiation, higher T stage and larger size tumor are facing higher risk. We took these preoperative assessment available factors into our nomogram construction and the model showed good discrimination and calibration. When using the nomogram for a specific patient, first collect all the clinicopathologic feature listed and put them into the scale and get a score for each variable. Next add all scores up to get a total score and match with a probability on the risk scale. This model enables physicians to calculate a specific lymph node metastasis risk for each patient and help them decide which patients will possibly benefit most from more extensive lymph node dissection and make better surgery plan.

Recent research has put much effort into understanding the molecular mechanism underlying lymph node metastasis in lung cancer. Moriya *et al.* reported a gene profiling method and selected sets of gene predictor for lymph node metastasis (21). Gomez *et al.* reported higher expression of E-cadherin,  $\gamma$ -catenin, p27, and p53 in patients with positive

lymph node while p16 and Rb were expressed in negative cases (25). With more and more clear interpretation of lymph node metastasis molecular biology in the future, our model would serve as a basic platform to identify high risk patients for further gene or protein level profiling examinations to achieve more accurate staging, more individualized management and hopefully better outcomes.

There are certain limitations of our study that we need to point out. Firstly, as this is a retrospective study, selection bias should be considered. White people constituted vast majority of the study cohort which can make prediction less accurate for people of other races and influence the model application in other countries. Secondly, There is some missing detailed information in SEER database like smoking history, genetic mutations, accurate recording of positive lymph node location and any preoperative chemotherapy or regional radiotherapy which will affect lymph node status. EGFR mutation is a very important factor in lung cancer management today and may have certain effect on



**Figure 3** Calibration curve and discrimination curves of the nomogram in training cohort and validation cohort. (A,B) Calibration curves of training cohort and validation curve. The x-axis showed predicted probability of the model and y-axis shows actual probabilities. (C,D) Receiver operating characteristic (ROC) curve for discrimination in the training and validation cohorts. Area under the curve (AUC) was 0.696 (95% CI, 0.617 to 0.775) and 0.693 (95% CI, 0.628 to 0.758) which showed that the model presented with good performance.

lymph node status. Also, it would be better if PET scan results are also taken into consideration. These variables will be a major part of our future research to improve this model. Thirdly, further external validation of more multi-institution cohorts from around the world are required to confirm our results.

### Acknowledgments

*Funding:* None.

### Footnote

*Reporting Checklist:* The authors have completed the STROBE reporting checklist. Available at <http://dx.doi.org/10.21037/jtd-20-601>

<http://dx.doi.org/10.21037/jtd-20-601>

*Conflicts of Interest:* All authors have completed the ICMJE uniform disclosure form (available at <http://dx.doi.org/10.21037/jtd-20-601>). The authors have no conflicts of interest to declare.

*Ethical Statement:* The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The trial was conducted in accordance with the Declaration of Helsinki (as revised in 2013) and the Harmonized Tripartite Guideline for Good Clinical Practice from the International Conference on Harmonization. This study was approved

by Ethical Committee of Shanghai Cancer Center, Fudan University. Patient consent form was not required for data from SEER database as all data was deidentified before release and contained no personally identifying information of patients.

*Open Access Statement:* This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

## References

- Gridelli C, Rossi A, Carbone DP, et al. Non-small-cell lung cancer. *Nat Rev Dis Primers* 2015;1:15009.
- Jamal-Hanjani M, Wilson GA, McGranahan N, et al. Tracking the Evolution of Non-Small-Cell Lung Cancer. *N Engl J Med* 2017;376:2109-21.
- Herzberg B, Campo MJ, Gainor JF. Immune checkpoint inhibitors in non-small cell lung cancer. *Oncologist* 2017;22:81-8.
- Masri M, McManus M, Mudad R. Treatment of Advanced Non-Small Cell Lung Cancer in the Era of Targeted Therapy. *Curr Pulmonol Rep* 2018;7:79-91.
- Reck M, Rabe KF. Precision diagnosis and treatment for advanced non-small-cell lung cancer. *N Engl J Med* 2017;377:849-61.
- de Castro ABG, Domínguez JF, Bolton RD, et al. PET-CT in presurgical lymph node staging in non-small cell lung cancer: The importance of false-negative and false-positive findings. *Radiologia* 2017;59:147-58.
- El-Sherief AH, Lau CT, Obuchowski NA, et al. Cross-Disciplinary Analysis of Lymph Node Classification in Lung Cancer on CT Scanning. *Chest* 2017;151:776-85.
- Peng Z, Liu Q, Li M, et al. Comparison of (11)C-choline PET/CT and enhanced CT in the evaluation of patients with pulmonary abnormalities and locoregional lymph node involvement in lung cancer. *Clin Lung Cancer* 2012;13:312-20.
- Eloubeidi MA, Cerfolio RJ, Chen VK, et al. Endoscopic ultrasound-guided fine needle aspiration of mediastinal lymph node in patients with suspected lung cancer after positron emission tomography and computed tomography scans. *Ann Thorac Surg* 2005;79:263-8.
- Gagliasso M, Migliaretti G, Ardisson F. Assessing the prognostic impact of the International Association for the Study of Lung Cancer proposed definitions of complete, uncertain, and incomplete resection in non-small cell lung cancer surgery. *Lung Cancer* 2017;111:124-30.
- Keller SM, Adak S, Wagner H, et al. Mediastinal lymph node dissection improves survival in patients with stages II and IIIa non-small cell lung cancer. *Ann Thorac Surg* 2000;70:358-65.
- Lardinois D, Suter H, Hakki H, et al. Morbidity, survival, and site of recurrence after mediastinal lymph-node dissection versus systematic sampling after complete resection for non-small cell lung cancer. *Ann Thorac Surg* 2005;80:268-74; discussion 274-5.
- Park HK, Jeon K, Koh WJ, et al. Occult nodal metastasis in patients with non-small cell lung cancer at clinical stage IA by PET/CT. *Respirology* 2010;15:1179-84.
- Ding N, Mao Y, Gao S, et al. Predictors of lymph node metastasis and possible selective lymph node dissection in clinical stage IA non-small cell lung cancer. *J Thorac Dis* 2018;10:4061-8.
- Xia W, Wang A, Jin M, et al. Young age increases risk for lymph node positivity but decreases risk for non-small cell lung cancer death. *Cancer Manag Res* 2018;10:41-8.
- Yu X, Li Y, Shi C, et al. Risk factors of lymph node metastasis in patients with non-small cell lung cancer  $\leq$  2 cm in size: A monocentric population-based analysis. *Thorac Cancer* 2018;9:3-9.
- Soneji S, Beltran-Sanchez H, Sox HC. Assessing progress in reducing the burden of cancer mortality, 1985-2005. *J Clin Oncol* 2014;32:444-8.
- Senan S, Paul MA, Lagerwaard FJ. Treatment of early-stage lung cancer detected by screening: surgery or stereotactic ablative radiotherapy? *Lancet Oncol* 2013;14:e270-4.
- Doddoli C, Aragon A, Barlesi F, et al. Does the extent of lymph node dissection influence outcome in patients with stage I non-small-cell lung cancer? *Eur J Cardiothorac Surg* 2005;27:680-5.
- Ishiguro F, Matsuo K, Fukui T, et al. Effect of selective lymph node dissection based on patterns of lobe-specific lymph node metastases on patient outcome in patients with resectable non-small cell lung cancer: A large-scale retrospective cohort study applying a propensity score. *J Thorac Cardiovasc Surg* 2010;139:1001-6.
- Moriya Y, Iyoda A, Kasai Y, et al. Prediction of lymph node metastasis by gene expression profiling in patients with

- primary resected lung cancer. *Lung Cancer* 2009;64:86-91.
22. Han H, Chen H. Selective lymph node dissection in early-stage non-small cell lung cancer. *J Thorac Dis* 2017;9:2102-7.
  23. Zhao F, Zhou Y, Ge PF, et al. A prediction model for lymph node metastases using pathologic features in patients intraoperatively diagnosed as stage I non-small cell lung cancer. *BMC Cancer* 2017;17:267.
  24. Balachandran VP, Gonen M, Smith JJ, et al. Nomograms in oncology: more than meets the eye. *Lancet Oncol* 2015;16:e173-80.
  25. Gomez AM, Jarabo Sarceda JR, Garcia-Asenjo JA, et al. Relationship of immunohistochemical biomarker expression and lymph node involvement in patients undergoing surgical treatment of NSCLC with long-term follow-up. *Tumour Biol* 2014;35:4551-9.

**Cite this article as:** Zhang C, Song Q, Zhang L, Wu X. Development of a nomogram for preoperative prediction of lymph node metastasis in non-small cell lung cancer: a SEER-based study. *J Thorac Dis* 2020;12(7):3651-3662. doi: 10.21037/jtd-20-601