Novel nomograms to predict lymph node metastasis and distant metastasis in resected patients with early-stage non-small cell lung cancer

Yi Tian[1,2], Yu He[2], Xin Li[3], Xiaowen Liu[1]^

[1]Department of Oncology, Xiangya Hospital, Central South University, Changsha, China; [2]Department of Thoracic Surgery, Second Xiangya Hospital, Central South University, Changsha, China; [3]Department of Dermatology, Second Xiangya Hospital, Central South University, Changsha, China

Contributions: (I) Conception and design: X Liu, Y Tian; (II) Administrative support: X Liu; (III) Provision of study materials or patients: Y Tian, Y He; (IV) Collection and assembly of data: Y Tian, Y He, X Li; (V) Data analysis and interpretation: Y Tian; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Xiaowen Liu. Department of Oncology, Xiangya Hospital, Central South University, No. 87 Xiangya Road, Changsha 410008, China. Email: xiaowenlable@gmail.com.

Background: We aimed to formulate and validate nomograms to unravel the significant risk factors associated with lymph node metastasis (LNM) and distant metastasis in early-stage non-small cell lung cancer (NSCLC).

Methods: Totally, 22403 pathologic T1 (pT1) or pT2 NSCLC patients were extracted from the Surveillance, Epidemiology, and End Results (SEER) database and were further randomly assorted into training and testing cohorts. Clinicopathologic risk predictors associated with distant metastasis and LNM were investigated through multivariate logistic regression. The LNM and distant metastasis nomogram models integrating all significant variables were internally validated in the testing cohort. The prediction efficiency of nomograms was assessed via the receiver operating characteristics curve (ROC), calibration plots and decision curve analysis (DCA).

Results: The LNM nomogram displayed a favorable calibration and discrimination in both training and testing cohorts, with an area under the curve (AUC) of 0.721 (95% CI: 0.711–0.734) and 0.713 (95% CI: 0.699–0.727), respectively. And the encouraging prediction efficiency was also revealed in the training cohort (AUC =0.788, 95% CI: 0.761–0.816) and the testing cohort (AUC =0.765, 95% CI: 0.721–0.808) for the distant metastasis nomogram. The calibration plot revealed an optimal consistency between model prediction and practical observation. Both nomograms were endowed with the optimal clinical utility and benefits compared with conventional clinicopathological indicators.

Conclusions: Our nomogram models are a promising tool with robust predictive power to effectively and intuitively predict the occurrence of LNM and distant metastasis in early-stage NSCLC.

Keywords: SEER; early-stage non-small cell lung cancer (NSCLC); lymph node metastasis (LNM); nomogram

---

^ORCID: 0000-0002-2543-4599.

Introduction

Lung cancer is the most common cause of cancer-related death globally (1). About 1.8 million patients are estimated to die from lung cancer in 2018 (2). Non-small cell lung cancer (NSCLC) is the most dominating histologic subtype of all primary lung cancers, occupying about 85% of all cases (3,4). NSCLC is pathologically categorized into squamous cell carcinoma, adenocarcinoma, and large cell carcinoma (5). The 5-year survival rates of lung cancer fluctuate ranging from 4% to 17%, which is primarily ascribed to tumor stage and regional discrepancy (6,7). The unsatisfactory prognosis and frequent recurrence of NSCLC are closely associated with lymph node metastasis (LNM) and distant metastasis (8).

In accordance with the 7th edition of the TNM staging system of lung cancer (9,10), patients can be diagnosed with advanced NSCLC (TNM stage IV) in the presence of distant metastasis, regardless of the pathologic T (pT) classification (10). NSCLC is characterized with formidable metastatic capacity and the bone (25–30%) is the most frequent metastatic site, followed by the lung (20–25%), the brain (15–20%) and the liver (5–10%) (11,12). Approximately 70–80% of patients diagnosed with NSCLC develop LNM in the clinical course of their disease. Furthermore, approximately 40–50% of advanced NSCLC patients (TNM stage III and IV) remain at pT1 or pT2 because of the aggressive characteristics of early-stage NSCLC (13,14). Above mentioned patients merely have 5–year survival rates of 4–15% on account of the progression of distant metastasis and subsequent palliative systemic treatment (6,14). Specifically, NSCLC patients with liver metastasis and abdominal lymph nodes metastasis are correlated with significantly diminished overall survival (OS) (15). For NSCLC patients with a pathological tumor size less than 30 mm, cases with LNM tend to display a much more unsatisfactory median OS than those without LNM (54.4 versus 114.1 months, P<0.001) (16). Thus, distant metastasis and LNM are crucial determinants of the prognosis of early-stage NSCLC (pT1 or pT2). Early detection of metastasis in early-stage NSCLC is essential for improving therapeutic planning and patient's risk stratification in clinical practice.

Currently, there is merely confined information concerning how to estimate the risk factors related to LNM and distant metastasis in early-stage NSCLC individuals. Therefore, we attempted to develop and identify graphical calculating nomograms with an integration of all significant clinicopathological variables, thus accurately quantifying the risk events for tumor progression (17,18). To date, nomograms have not been formulated to predict LNM and distant metastasis in patients with early-stage NSCLC through extracting clinical data from the Surveillance, Epidemiology, and End Results (SEER) database. Herein, our study was designed to establish the first nomograms to confer risk prediction of LNM and distant metastasis in resected early-stage NSCLC patients.

We present the following article in accordance with the TRIPOD reporting checklist (available at http://dx.doi.org/10.21037/apm-20-1756).

Methods

Data source and population selection

The SEER program financed by the National Cancer Institute is composed of 18 population-based cancer registries, covering approximately 28% of the U.S. population (19,20). The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). Considering that all available information from the public SEER database was anonymized, our study was exempted from the ethics committee approval and consent procedure (20,21). Because the SEER database initiated collecting the information for metastatic sites of bone, liver, lung and brain in 2010 and relevant data was renewed until 2015, we queried all pT1 or pT2a NSCLC patients who underwent resection between 2010 and 2015 from the SEER database through SEER*Stat 8.3.6 software. The flowchart of patient selection was shown in *Figure 1*.

A total of 20 clinical parameters were extracted for each patient, including demographic variables (age at diagnosis, sex, ethnicity and survival time), tumor-associated characteristics (primary tumor sites, pathological differentiation, pT stage, N stage, M stage, TNM stage, tumor size, histological type, pleural invasion, the status of distant metastasis, lung metastasis, bone metastasis, brain metastasis, liver metastasis) and cancer-targeted therapy (surgery type). Specifically, histological classifications of NSCLC, including "Adenocarcinoma", "Squamous cell carcinoma", "Large cell lung cancer (LCLC)" and "NSCLC, Not otherwise specified (NSCLC/NOS)" were confirmed based on the International Classification of Diseases for Oncology, 3rd Edition (IDO-O-3). All NSCLC cases were staged based on the 7th edition of the TNM staging system. The exclusion criteria were as follows: patients who
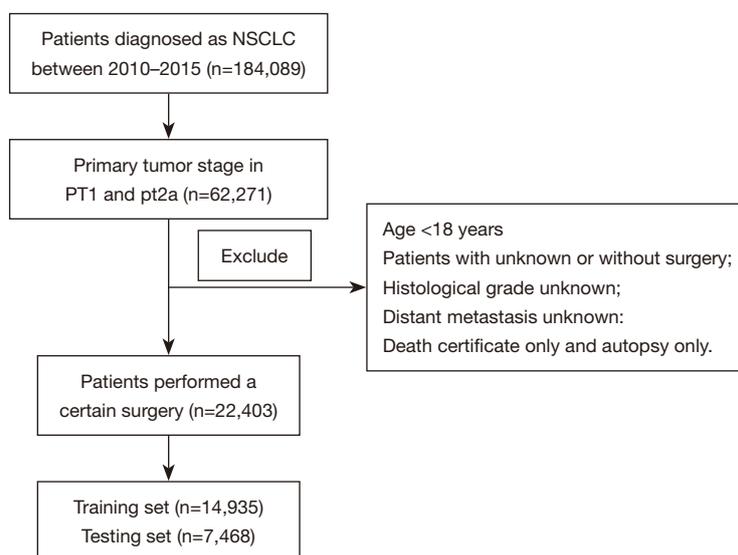
2550

Tian et al. Nomograms for early-stage NSCLC

```
┌─────────────────────────────┐
│  Patients diagnosed as NSCLC │
│  between 2010–2015 (n=184,089)│
└─────────────────────────────┘
              │
              ▼
┌─────────────────────────────┐
│   Primary tumor stage in     │
│   PT1 and pt2a (n=62,271)    │
└─────────────────────────────┘
              │                    ┌──────────────────────────────────────────┐
              │    ┌─────────┐     │ Age <18 years                            │
              ├───▶│ Exclude │────▶│ Patients with unknown or without surgery;│
              │    └─────────┘     │ Histological grade unknown;              │
              │                    │ Distant metastasis unknown:              │
              │                    │ Death certificate only and autopsy only. │
              ▼                    └──────────────────────────────────────────┘
┌─────────────────────────────┐
│    Patients performed a      │
│   certain surgery (n=22,403) │
└─────────────────────────────┘
              │
              ▼
┌─────────────────────────────┐
│   Training set (n=14,935)    │
│   Testing set (n=7,468)      │
└─────────────────────────────┘
```

Figure 1 Flowchart of patient selection. NSCLC, non-small cell lung cancer; pT, pathologic T.

aged <18 years old or were diagnosed merely by autopsy or death certificate, or whose clinical data were insufficient. Eventually, a total of 22,403 enrolled cases extracted from the SEER database were randomly classified into a training cohort (n=14,935) for nomogram establishment and a testing cohort (n=7,468) for internal verification.

Establishment and validation of nomogram models

All significant independent risk factors for LNM and distant metastasis were confirmed in early-stage NSCLC cases through univariable and multivariable logistic regression analysis (P<0.05). The SEER internal testing cohort was applied for evaluating the predictive performance of the nomograms associated with LNM and distant metastasis. A bootstrapping strategy with 2,000 samples was further conducted for the internal validation of the nomogram models. The receiver operating characteristics curve (ROC) and calibration plots were used to measure and validate the predictive reliability and accuracy of the models, respectively. Subsequently, decision curve analysis (DCA) curve was established to evaluate the clinical utility and benefits of the nomogram and other conventional clinicopathological parameters.

Statistical analysis

The R language version 3.3.5 was applied for statistical analyses. Means and standard deviation were utilized to represent continuous variables. Variables could be compared through the Chi squared test and Student's t-test. The nomogram models and the corresponding calibration plots as well as ROC curve were calculated by the pROC and rms packages. The cutoff values of the risk scores in the predictive nomogram models were estimated through the maximum Youden index. All cases were further classified into low- and high-risk groups. All tests were 2-tailed and considered statistically significant at P<0.05.

Results

Clinicopathologic features of patients

As shown in Table S1, there were no statistically significant discrepancies in the demographic and clinical characteristics of early-stage NSCLC patients between the training and the validation cohorts (P>0.05). LNM occurred in 1,653 of 14,935 patients (11.06%) in the training group and 697 of 7,468 patients (9.33%) in testing group. Distant metastasis was present in 261 of 14,935 patients (1.75%) in the training cohort and 131 of 7,468 patients (1.75%) in the testing cohort. Specifically, in the training cohort, the proportion of bone metastasis, brain metastasis, liver metastasis was 0.42% (n=63), 0.91% (n=136), 0.15% (n=22), respectively. The corresponding numerical value in the testing cohort was 0.43% (n=32), 0.99% (n=74), 0.09% (n=7), respectively. Table S1 revealed no statistically significant discrepancy in

LNM proportion (P=0.47) or distant metastasis proportion (P=1.00) between both groups.

Correlations between clinicopathologic parameters and LNM and distant metastasis

The existence of LNM was examined stratified by different clinicopathologic variables. Among all clinical parameters, LNM was significantly correlated with age at diagnosis, primary site, pathological differentiation, pT stage, metastasis status, tumor size, pleural invasion, surgery type, brain metastasis, bone metastasis and lung metastasis as well as OS in both the training and testing cohorts (P<0.001) (*Table 1*). Similarly, in the correlation analysis, 11 parameters, namely age, pathological differentiation, lymph node status, histological type, pleural invasion, surgery type, brain metastasis, bone metastasis, liver metastasis and lung metastasis as well as OS were all significantly associated with distant metastasis in two sets (P<0.001) (*Table 2*).

Univariable and multivariable logistic regression analysis in training cohort

Initially, we preliminarily screened the independent risk factors for LNM and distant metastasis through univariate logistic regression analysis (*Table 3*). LNM was significantly correlated with age, gender, ethnicity, primary site, pathological differentiation, pT stage, tumor size, histological type, pleural invasion, surgery type and distant metastasis (P<0.05). Additionally, age, gender, pathological differentiation, pT stage, lymph node status, histological type, pleural invasion and surgery type were significantly related to distant metastasis of early-stage NSCLC (P<0.05).

The abovementioned significant risk predictors in the univariate analysis were all incorporated into the multivariate analysis. The associated risk factors of LNM and distant metastasis and the corresponding OR with 95% CI were presented in *Table 4*. Over 65 years of age (OR =0.77, 95% CI: 0.67–0.81, P<0.001) was negatively associated with the development of LNM. Furthermore, for histological type, patients with NSCLC-subtype squamous cell carcinoma had a significantly diminished risk of LNM compared with those with adenocarcinoma (OR =0.61, 95% CI: 0.55–0.68, P<0.001). Nevertheless, male gender (OR =1.12, 95% CI: 1.02–1.22, P=0.015), T2a stage (OR =1.59, 95% CI: 1.36–1.86, P<0.001), tumor size of 30–50 mm (OR =1.47, 95% CI:1.28–1.69, P<0.001) and distant metastasis (OR =2.94, 95% CI: 2.25–3.83, P<0.001) were positively

associated with LNM. As for primary site of early-stage NSCLC, compared with right upper lobe (RUL), tumor located in the left lower lobe (LLL) displayed the highest risk of LNM (OR =1.37, 95% CI: 1.2–1.57, P<0.001), followed by left upper lobe (LUL) (OR =1.36, 95% CI: 1.22–1.52, P<0.001), right middle lobe (RML) (OR =1.33, 95% CI: 1.1–1.59, P=0.003) and right lower lobe (RLL) (OR =1.31, 95% CI: 1.16–1.49, P<0.001). NSCLC patients with moderately or poorly-differentiated tumors were more prone to develop LNM than those with well-differentiated tumors (OR =2.96, 95% CI: 2.54–3.46, P<0.001 for moderately-differentiated tumors; and OR =4.19, 95% CI: 3.59–4.91, P<0.001 for poorly-differentiated tumors).

Multivariable logistic regression indicated that age >65 years old (OR =0.53, 95% CI: 0.41–0.68, P=0.006) and squamous cell carcinoma (OR =0.44, 95% CI: 0.3–0.63, P<0.001) were negatively correlated with the occurrence of distant metastasis (*Table 4*). Early-stage NSCLC patients treated with lobectomy/bilobectomy (OR =0.35, 95% CI: 0.26–0.46, P<0.001) or pneumonectomy (OR =0.33, 95% CI: 0.13–0.7, P<0.001) showed a significantly diminished risk of distant metastasis compared to those treated with wedge resection. Conversely, male patients (OR =1.43, 95% CI: 1.11–1.85, P<0.001) with pleural invasion (OR =1.66, 95% CI: 1.15–2.42, P<0.001) exhibited an increased risk for developing distant metastasis. Moderately (OR =2.97, 95% CI: 1.74–5.43, P<0.001) or poorly-differentiated (OR =4.96, 95% CI: 2.92–9.05, P<0.001) tumors had a higher risk of the development of distant metastasis than well-differentiated tumors. For lymph node status, patients with more aggressive lymph node involvement had a higher risk of distant metastasis (OR =1.98, 95% CI: 1.33–2.88, P<0.001 for N1; OR =3.11, 95% CI: 2.26–4.23, P<0.001 for N2; OR =15.14, 95% CI: 7.01–31.4, P<0.001 for N3).

Development of nomogram model for LNM and distant metastasis prediction

Significant results of the multivariate regression model were applied for constructing nomogram models of LNM (*Figure 2*) and distant metastasis (*Figure 3*). Moreover, Table S2 revealed the estimated points scales and predictive scores for each parameter in the nomograms. Specifically, *Figure 2* demonstrated that surgery type and degrees of pathological differentiation made the greatest contribution to the presence of LNM, followed by the status of distant metastasis, pT stage, and histological type as well as tumor size. The primary site of tumor, age and gender exerted a

2552

Tian et al. Nomograms for early-stage NSCLC

Table 1 Correlations between clinicopathological variables and LNM in both cohorts

Variables	Training (n=14,935)			Testing (n=7,468)		
	Negative	Positive	P value	Negative	Positive	P value
Age (mean, SD)	67.1±10.3	65.5±10.5	<0.001	66.9±10.3	65.2±10	<0.001
Gender			<0.001			0.03
Female	6,672	1,484		3,294	771	
Male	5,279	1,500		2,689	714	
Race			<0.001			0.16
White	10,010	2,423		4,944	1,197	
Black	1,020	296		535	153	
Unknown	921	265		504	135	
Primary site			<0.001			<0.001
RUL	4,117	862		2,080	458	
RML	746	186		340	85	
RLL	2,121	523		1,060	274	
LUL	2,991	852		1,497	407	
LLL	1,733	466		895	209	
Unknown	243	95		111	52	
Pathological differentiation			<0.001			<0.001
Well; I	2,936	238		1,485	127	
Moderately; II	5,440	1,387		2,704	642	
Poorly; III	3,427	1,304		1,720	691	
Undifferentiated; IV	148	55		74	25	
pT stage			<0.001			<0.001
T1	7,447	1,154		3,760	574	
T2a	4,504	1,830		2,223	991	
Metastasis status			<0.001			<0.001
M0	11,722	2,807		5,878	1,403	
M1a	89	56		32	24	
M1b	140	121		73	58	
Tumor size			<0.001			<0.001
≤30 mm	9,184	1,693		4,640	859	
30–50 mm	2,735	1,282		1,326	618	
Unknown	32	9		17	8	

Table 1 (*continued*)

Table 1 (*continued*)

Variables	Training (n=14,935)			Testing (n=7,468)		
	Negative	Positive	P value	Negative	Positive	P value
Histological type			<0.001			0.003
Adenocarcinoma	7,051	1,921		3,578	921	
Squamous	2,892	677		1,449	358	
Large cell	199	66		96	37	
NSCLC (NOS)	1,809	320		860	169	
Pleural invasion			<0.001			<0.001
No invasion	8,485	1,837		4,253	905	
Invasion	2,010	768		1,015	388	
Unknown	1,456	379		715	192	
Surgery type			<0.001			<0.001
Wedge resection	2,562	329		1,293	154	
Lobectomy/bilobectomy	9,261	2,452		4,613	1,234	
Pneumonectomy	128	203		77	97	
Brain metastasis			<0.001			<0.001
Negative	11,861	2,909		5,931	1,453	
Positive	78	58		47	27	
Unknown	12	17		5	5	
Bone metastasis			<0.001			<0.001
Negative	11,909	2,935		5,967	1,459	
Positive	29	34		10	22	
Unknown	13	15		6	4	
Liver metastasis			<0.001			0.02
Negative	11,928	2,955		5,974	1,477	
Positive	10	12		4	3	
Unknown	13	17		5	5	
Lung metastasis			<0.001			<0.001
Negative	11,902	2,945		5,965	1,470	
Positive	35	19		13	8	
Unknown	14	20		5	7	
Overall survival			<0.001			<0.001
Alive	10,279	1,869		5,157	952	
Death	1,672	1,115		826	533	

LNM, lymph node metastasis; SD, standard deviation; pT, pathologic T; RUL, right upper lobe; RML, right middle lobe; RLL, right lower lobe; LUL, left upper lobe; LLL, left lower lobe; pT, pathologic T; NSCLC (NOS), non-small cell lung cancer, not otherwise specified.

Table 2 Correlations between clinicopathological variables and distant metastasis in both cohorts

Variables	Training (n=14,935)			Testing (n=7,468)		
	Negative	Positive	P value	Negative	Positive	P value
Age (mean, SD)	66.8±10.3	63.3±10.3	<0.001	66.7±10.3	62.5±9.7	<0.001
Gender			<0.001			0.56
Female	8,041	115		3,397	68	
Male	6,633	146		3,340	63	
Race			0.76			0.74
White	12,215	218		6,030	111	
Black	1,291	25		678	10	
Unknown	1,168	18		629	10	
Primary site			0.14			0.41
RUL	4,895	84		2,492	46	
RML	915	17		421	4	
RLL	2,610	34		1,311	23	
LUL	3,760	83		1,864	40	
LLL	2,164	35		1,090	14	
Unknown	330	8		159	4	
Pathological differentiation			<0.001			<0.001
Well; I	3,159	15		1,603	9	
Moderately; II	6,724	103		3,302	44	
Poorly; III	4,600	131		2,338	73	
Undifferentiated; IV	191	12		94	5	
pT stage			<0.001			0.002
T1	8,485	116		4,275	59	
T2a	6,189	145		3,062	72	
Lymph node status*			<0.001			<0.001
N0	11,811	140		5,910	73	
N1	1,502	37		785	19	
N2	1,301	66		613	27	
N3	28	14		16	9	
NX	32	4		13	3	
Tumor size			0.03			0.18
≤30 mm	10,702	175		5,409	90	
30–50 mm	3,933	84		1,904	40	
Unknown	39	2		24	1	

Table 2 (*continued*)

Table 2 (*continued*)

Variables	Training (n=14,935)			Testing (n=7,468)		
	Negative	Positive	P value	Negative	Positive	P value
Histological type			<0.001			<0.001
Adenocarcinoma	8,788	184		4,403	96	
Squamous	3,533	36		1,789	18	
Large cell	253	12		127	6	
NSCLC (NOS)	2,100	29		1,018	11	
Pleural invasion			<0.001			<0.001
No invasion	10,204	118		5,093	65	
Invasion	2,702	76		1,369	34	
Unknown	1,768	67		875	32	
Surgery type			<0.001			<0.001
Wedge resection	2,792	99		139	48	
Lobectomy/bilobectomy	11,558	155		5,766	81	
Pneumonectomy	325	7		172	2	
Brain metastasis			<0.001			<0.001
Negative	14,648	123		7,329	46	
Positive	0	135		0	73	
Unknown	26	3		8	2	
Bone metastasis			<0.001			<0.001
Negative	14,648	196		7,329	97	
Positive	0	63		0	31	
Unknown	26	2		8	3	
Liver metastasis			<0.001			<0.001
Negative	14,648	196		7,329	124	
Positive	0	63		0	5	
Unknown	26	2		8	2	
Lung metastasis			<0.001			<0.001
Negative	14,605	242		7,309	126	
Positive	41	13		18	3	
Unknown	28	6		10	2	
Overall survival			<0.001			<0.001
Alive	12,040	108		6,060	49	
Death	2,634	153		1,277	82	

*, NX indicates unknown stage. RUL, right upper lobe; RML, right middle lobe; RLL, right lower lobe; LUL, left upper lobe; LLL, left lower lobe; pT, pathologic T; NSCLC (NOS), non-small cell lung cancer, not otherwise specified.

Table 3 Univariate logistic regression to analyze risk factors for LNM and distant metastasis

Variables	LNM			Distant metastasis		
	OR	95% CI	P value	OR	95% CI	P value
Age						
≤65	1			1		
>65	0.71	0.65–0.77	<0.001	0.52	0.4–0.66	<0.001
Gender						
Female	1			1		
Male	1.28	1.17–1.38	<0.001	1.53	1.2–1.97	<0.001
Race						
White	1			1		
Black	1.2	1.04–1.37	0.009	1.09	0.7–1.61	0.71
Primary site						
RUL	1			1		
RML	1.19	1–1.42	0.05	1.08	0.62–1.78	0.77
RLL	1.18	1.04–1.33	0.007	0.76	0.5–1.12	0.18
LUL	1.36	1.22–1.51	<0.001	1.29	0.95–1.75	0.11
LLL	1.28	1.13–1.46	<0.001	0.94	0.63–1.39	0.77
Pathological differentiation						
Well; I	1			1		
Moderately; II	3.15	2.73–3.64	<0.001	3.23	1.94–5.78	<0.001
Poorly; III	4.69	4.06–5.45	<0.001	6	3.63–10.7	<0.001
Undifferentiated; IV	4.58	3.25–6.38	<0.001	13.23	5.99–28.6	<0.001
pT stage						
T1	1					
T2a	2.6	2.4–2.8	<0.001	1.7	1.34–2.19	<0.001
Lymph node status						
N0	–	–	–	1		
N1	–	–	–	2.08	1.42–2.96	<0.001
N2	–	–	–	4.28	3.16–5.74	<0.001
N3	–	–	–	42.18	21.2–80.5	<0.001
Tumor size						
≤30 mm	1			1		
30–50 mm	2.54	2.34–2.77	<0.001	1.31	1–1.69	0.05

Table 3 (*continued*)

Table 3 (*continued*)

Variables	LNM			Distant metastasis		
	OR	95% CI	P value	OR	95% CI	P value
Histological type						
Adenocarcinoma	1			1		
Squamous	0.86	0.78–0.95	0.002	0.49	0.33–0.69	<0.001
Large cell	1.22	0.91–1.61	0.17	2.27	1.18–3.94	0.007
NSCLC (NOS)	0.65	0.57–0.74	<0.001	0.66	0.44–0.96	0.04
Pleural invasion						
No invasion	1			1		
Invasion	1.76	1.6–1.94	<0.001	2.43	1.81–3.25	<0.001
Surgery type						
Wedge resection	1			1		
Lobectomy/bilobectomy	2.06	1.83–2.34	<0.001	0.38	0.29–0.49	<0.001
Pneumonectomy	12.35	9.64–15.9	<0.001	0.61	0.25–1.23	0.21
Distant metastasis						
Negative	1			–	–	–
Positive	3.56	2.78–4.56	<0.001	–	–	–

LNM, lymph node metastasis; RUL, right upper lobe; RML, right middle lobe; RLL, right lower lobe; LUL, left upper lobe; LLL, left lower lobe; pT, pathologic T; NSCLC (NOS), non-small cell lung cancer, not otherwise specified.

Table 4 Multivariable logistic regression to analyze risk factors for LNM and distant metastasis

Variables	LNM			Distant metastasis		
	OR	95% CI	P value	OR	95% CI	P value
Age						
≤65	1			1		
>65	0.77	0.67–0.81	<0.001	0.53	0.41–0.68	0.006
Gender						
Female	1			1		
Male	1.12	1.02–1.22	0.015	1.43	1.11–1.85	<0.001
Primary site						
RUL	1			–	–	–
RML	1.33	1.1–1.59	0.003	–	–	–
RLL	1.31	1.16–1.49	<0.001	–	–	–
LUL	1.36	1.22–1.52	<0.001	–	–	–
LLL	1.37	1.2–1.57	<0.001	–	–	–

Table 4 (*continued*)

Table 4 (*continued*)

Variables	LNM			Distant metastasis		
	OR	95% CI	P value	OR	95% CI	P value
Pathological differentiation						
Well; I	1			1		
Moderately; II	2.96	2.54–3.46	<0.001	2.97	1.74–5.43	<0.001
Poorly; III	4.19	3.59–4.91	<0.001	4.96	2.92–9.05	<0.001
Undifferentiated; IV	3.88	2.65–5.61	<0.001	6.88	2.81–16.4	<0.001
pT stage						
T1	1			1		
T2a	1.59	1.36–1.86	<0.001	1.15	0.83–1.59	0.4
Lymph node status						
N0	–	–	–	1		
N1	–	–	–	1.98	1.33–2.88	<0.001
N2	–	–	–	3.11	2.26–4.23	<0.001
N3	–	–	–	15.14	7.01–31.4	<0.001
Tumor size						
≤30 mm	1			–	–	–
30–50 mm	1.47	1.28–1.69	<0.001	–	–	–
Histological type						
Adenocarcinoma	1			1		
Squamous	0.61	0.55–0.68	<0.001	0.44	0.3–0.63	<0.001
Large cell	0.75	0.54–1.02	0.07	1.37	0.65–2.63	0.37
NSCLC (NOS)	0.82	0.71–0.94	0.005	0.9	0.59–1.34	0.63
Pleural invasion						
No invasion	1			1		
Invasion	1.1	0.96–1.26	0.15	1.66	1.15–2.42	<0.001
Surgery type						
Wedge resection	1			1		
Lobectomy/bilobectomy	1.83	1.61–2.09	<0.001	0.35	0.26–0.46	<0.001
Pneumonectomy	8.59	6.56–11.3	<0.001	0.33	0.13–0.7	<0.001
Distant metastasis						
Negative	1			–	–	–
Positive	2.94	2.25–3.83	<0.001	–	–	–

LNM, lymph node metastasis; RUL, right upper lobe; RML, right middle lobe; RLL, right lower lobe; LUL, left upper lobe; LLL, left lower lobe; pT, pathologic T; NSCLC (NOS), non-small cell lung cancer, not otherwise specified.
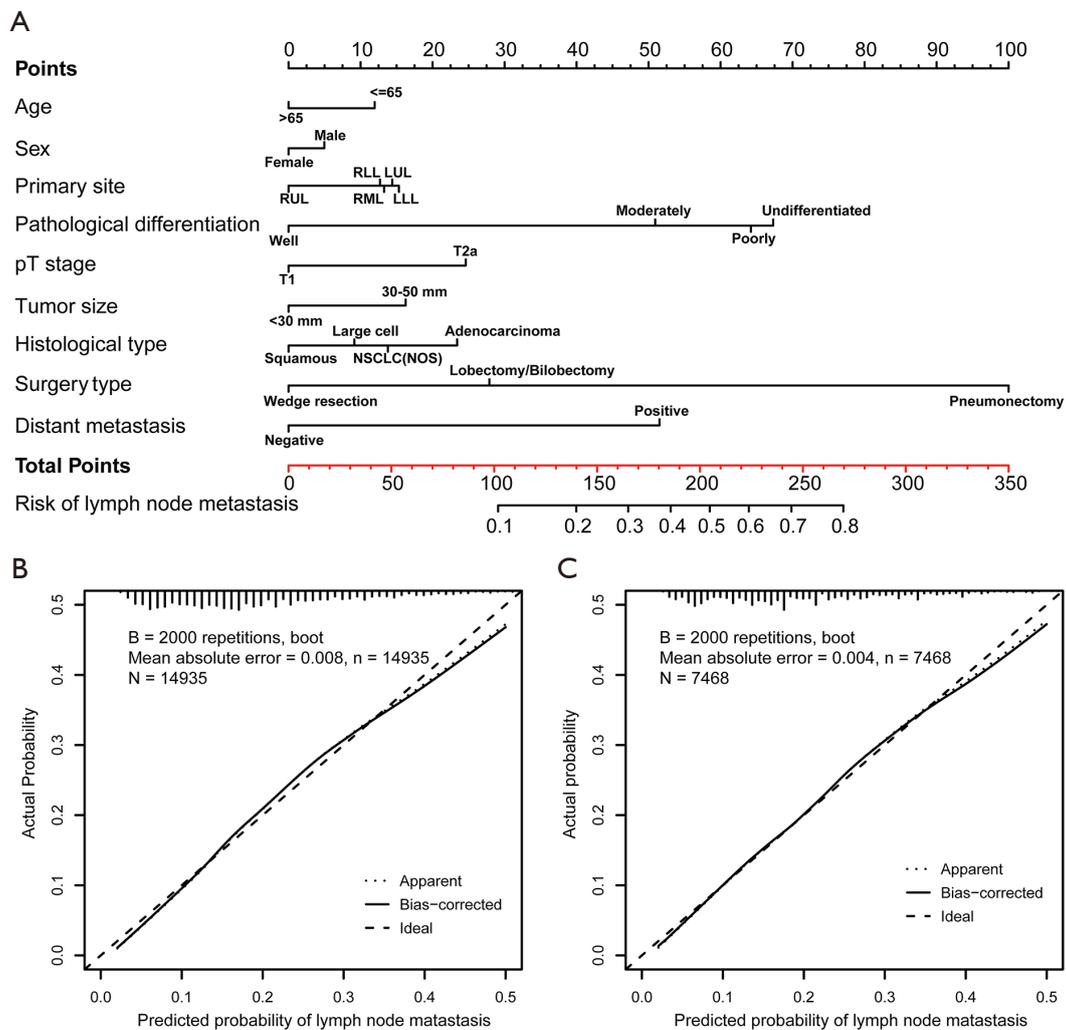

Figure 2 Nomogram models and calibration plots to predict LNM in early-stage NSCLC patients. (A) Nomograms for predicting the risk factors responsible for LNM. Calibration curves to estimate the predictive performance of LNM nomogram in (B) the training cohort and (C) the testing cohort, respectively. LNM, lymph node metastasis; RUL, right upper lobe; RML, right middle lobe; RLL, right lower lobe; LUL, left upper lobe; LLL, left lower lobe; pT, pathologic T; NSCLC (NOS), non-small cell lung cancer, not otherwise specified.

relatively small effect on the risk of LNM. Similarly, the status of lymph node invasion and degrees of pathological differentiation exerted the most noticeable impact on the distant metastasis nomogram, followed by histological type, surgery type, the status of pleural invasion. Age and sex made a few contributions to the risk of distant metastasis (*Figure 3*).

Efficiency and validation of nomogram model for LNM and distant metastasis prediction

The calibration plots to predict LNM (*Figure 2*)

and distant metastasis (*Figure 3*) revealed a favorable concordance between the predicted and observed probabilities in the training and testing set, respectively. We performed a comparison of the predictive accuracy for LNM and distant metastasis between the nomograms and clinicopathological parameters through ROC observation and its corresponding area under the curve (AUC). The predictive value of the LNM nomogram was favorable, achieving an AUC of 0.721 (95% CI: 0.711–0.734) in the training group (*Figure 4*) and 0.713 (95% CI: 0.699–0.727) in the testing cohort (*Figure 4*), respectively, which were more satisfactory than the AUCs of pathological
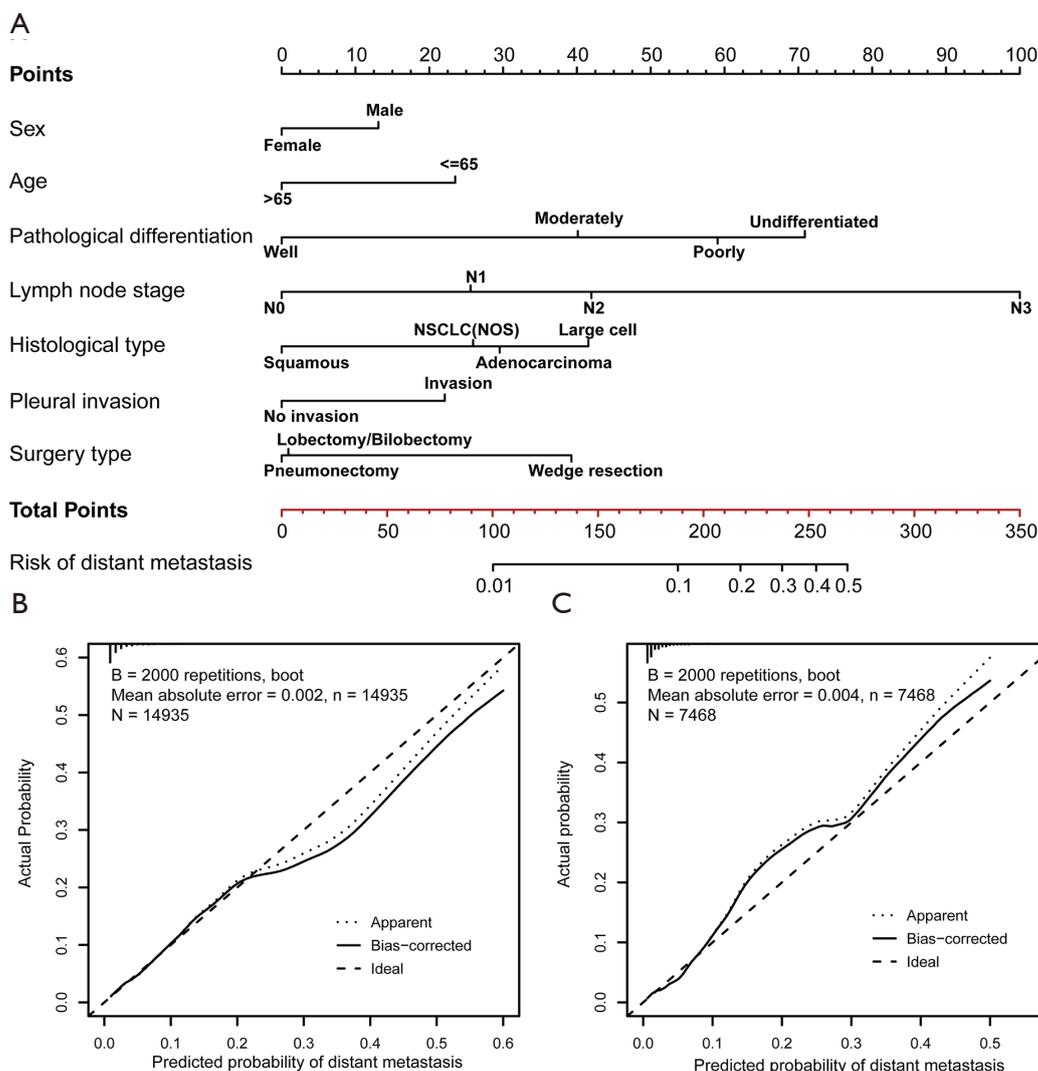
Figure 3 Nomogram models and calibration plots to predict distant metastasis in early-stage NSCLC patients. (A) Nomograms for predicting the risk factors responsible for distant metastasis. Calibration curves to estimate the predictive performance of distant metastasis nomogram in (B) the training cohort and (C) the testing cohort, respectively. NSCLC (NOS), non-small cell lung cancer, not otherwise specified.

differentiation, pT stage and histological type in two cohorts (P<0.001). Analogously, the distant metastasis nomogram model exhibited an AUC value of 0.788 (95% CI: 0.761–0.816) in the training cohort (*Figure 4*) and 0.765 (95% CI: 0.721–0.808) in the testing cohort (*Figure 4*), respectively. These values were higher than those for pathological differentiation, pleural invasion, the status of lymph node invasion and histological type. These results highlighted that the nomogram models conferred a more desirable predictive efficacy than that of the

clinicopathological risk factors alone.

Additionally, the bar charts were formulated to assess the discriminatory efficiency of the LNM and distant metastasis nomogram models following estimating the risk scores. Based on the maximum Youden index, we further established optimal risk score cutoff values of 130 and 102 for the LNM and distant metastasis nomogram, respectively. All enrolled subjects were stratified into low- and high-risk groups. In the training cohort, early-stage NSCLC individuals in the high-risk group had a higher
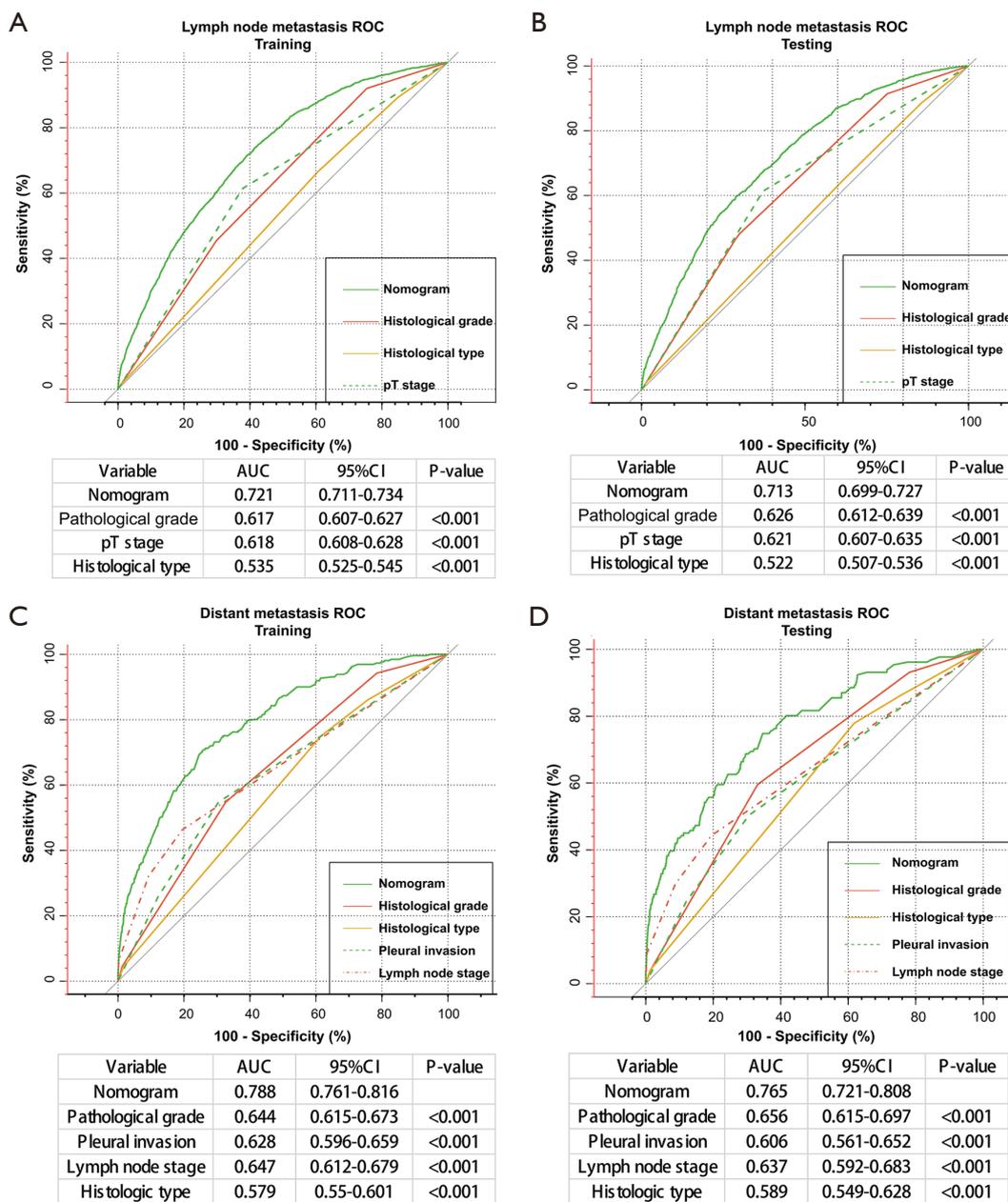
A

Lymph node metastasis ROC
Training

Variable	AUC	95%CI	P-value
Nomogram	0.721	0.711-0.734	
Pathological grade	0.617	0.607-0.627	<0.001
pT stage	0.618	0.608-0.628	<0.001
Histological type	0.535	0.525-0.545	<0.001

B

Lymph node metastasis ROC
Testing

Variable	AUC	95%CI	P-value
Nomogram	0.713	0.699-0.727	
Pathological grade	0.626	0.612-0.639	<0.001
pT stage	0.621	0.607-0.635	<0.001
Histological type	0.522	0.507-0.536	<0.001

C

Distant metastasis ROC
Training

Variable	AUC	95%CI	P-value
Nomogram	0.788	0.761-0.816	
Pathological grade	0.644	0.615-0.673	<0.001
Pleural invasion	0.628	0.596-0.659	<0.001
Lymph node stage	0.647	0.612-0.679	<0.001
Histologic type	0.579	0.55-0.601	<0.001

D

Distant metastasis ROC
Testing

Variable	AUC	95%CI	P-value
Nomogram	0.765	0.721-0.808	
Pathological grade	0.656	0.615-0.697	<0.001
Pleural invasion	0.606	0.561-0.652	<0.001
Lymph node stage	0.637	0.592-0.683	<0.001
Histologic type	0.589	0.549-0.628	<0.001

Figure 4 ROC analysis to evaluate the predictive efficiency of the nomograms for LNM and distant metastasis. ROC curves of LNM and distant metastasis in (A,C) the training cohort and (B,D) the testing cohort, respectively. LNM, lymph node metastasis; ROC, receiver operating characteristic; AUC, area under the curve; 95% CI, 95% confidence interval.

risk of LNM than counterparts in the low-risk group (Figure S1A), which was consistent with the prediction outcome in the testing cohort (Figure S1B). Similarly, patients in the high-risk group suffered from a higher probability for distant metastasis than those in the low-risk group in both cohorts (Figure S1C,D).

Clinical application of nomogram model for LNM and distant metastasis prediction

Furthermore, survival analysis also revealed that early-stage NSCLC subjects who were predicted to develop LNM (Figure S2A,B) or distant metastasis (Figure S2C,D)
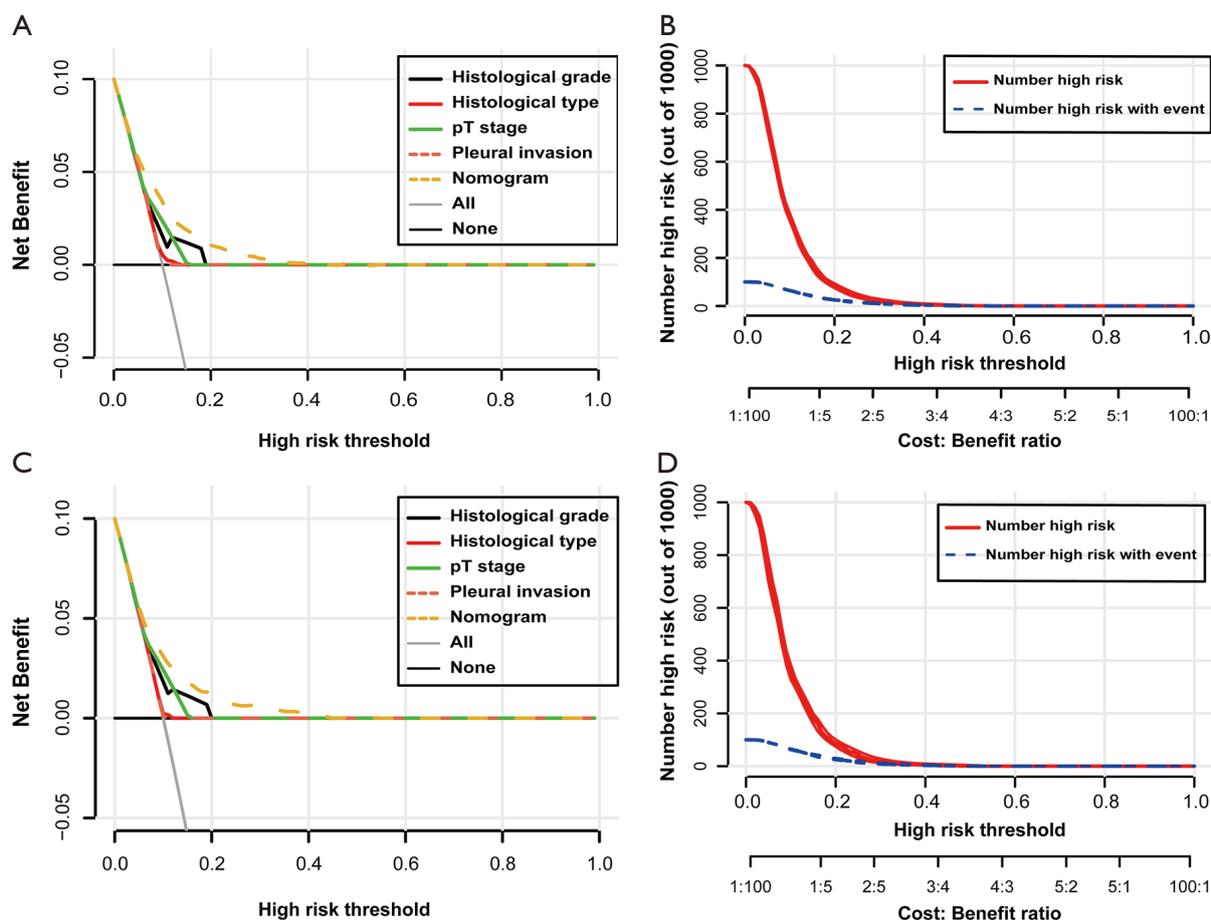
Ann Palliat Med 2021;10(3):2548-2566 | http://dx.doi.org/10.21037/apm-20-1756

Figure 5 DCA curve and clinical impact curve of LNM nomogram in (A,B) the training cohort and (C,D) the testing cohort, respectively. LNM, lymph node metastasis; DCA, decision curve analysis.

were characterized with a significantly decreased OS in both cohorts (P<0.0001). The DCA curve demonstrated that compared with the traditional variables, the LNM nomogram in the training group displayed the optimal clinical usability and benefits, followed by histological grade, pT stage and histological type as well as pleural invasion (*Figure 5*). A roughly consistent trend was also revealed in the LNM nomogram in the testing group (*Figure 5*). Additionally, clinical impact curves illustrated when the risk threshold was within a range of 0 to 0.3, the predicted number of early-stage NSCLC individuals with high risk appeared to be higher than the number of high-risk individuals with an actual LNM occurrence in both training and testing cohorts (*Figure 5*). Similarly, DCA curves showed the highest net benefit of the nomogram for predicting distant metastasis compared to the other four parameters (histological grade, LNM, pleural invasion and

histological type) in both cohorts (*Figure 6*). As showed in *Figure 6*, the cost benefit ratios were admissible within the range of 0 to 0.4.

Discussion

Lung cancer ranks first in terms of incidence and mortality worldwide in 2018. Distant metastasis is the primary reason responsible for unsatisfactory prognosis of early-stage NSCLC patients. LNM is also a considerable prognostic factor associated with tumor recurrence (1,2,22,23). Thus, it is momentous to estimate the risk factors associated with LNM and distant metastasis for early-stage NSCLC patients. In our report, we initially identified the clinicopathologic parameters' difference between early-stage NSCLC patients with LNM and distant metastasis and those without both metastases. We further confirmed
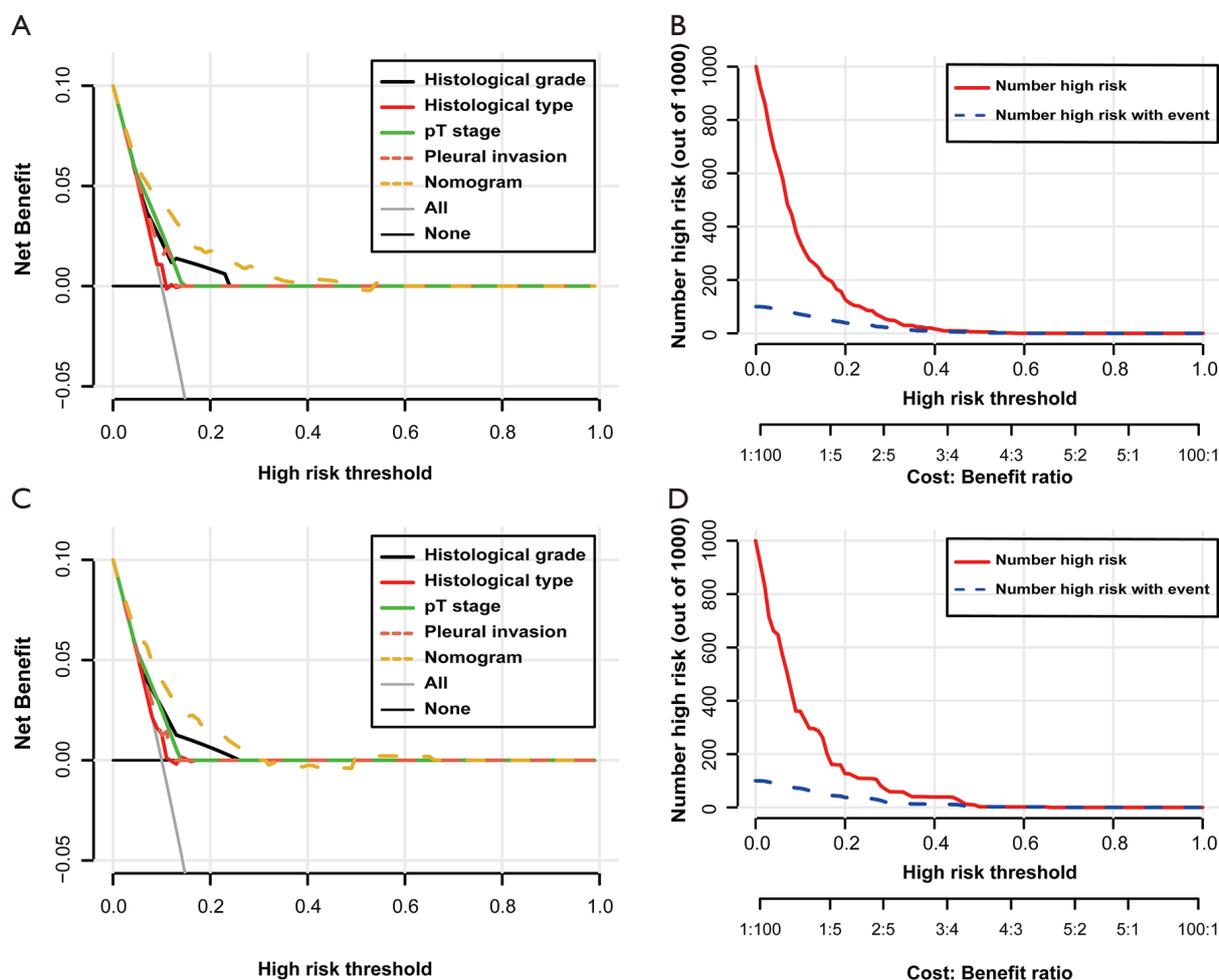
Figure 6 DCA curve and clinical impact curve of distant metastasis nomogram in (A,B) the training cohort and (C,D) the testing cohort, respectively. DCA, decision curve analysis.

certain significant clinical risk factors for LNM and distant metastasis through univariable and multivariable logistic regression analysis. Two visualized nomogram models were formulated and validated to confer prediction for the occurrence of LNM and distant metastasis in early-stage NSCLC patients through including all significant risk predictors filtrated in the multivariable logistic regression analysis. Our two nomograms revealed favorable coherence between model prediction and actual observation in the training and testing cohorts. Moreover, a satisfactory AUC value of the LNM and distant metastasis nomograms was 0.721 (95% CI: 0.711–0.734) and 0.788 (95% CI: 0.761–0.816), respectively, in the training group, highlighting that the predictive performance of nomogram models was significantly superior to that of conventional clinical

parameters, including pathological differentiation, histologic type, pT grade and the status of lymph node invasion.

Specifically, the predictive nomogram for LNM included ten clinical variables, namely surgery type, pathological differentiation, the status of distant metastasis, pT stage, histological type, tumor size, primary site of tumor and age as well as gender, which were basically consistent with the outcome in the previous clinical studies (24-27). Notably, neoplasm located in the LLL was prone to develop LNM, which was not consistent with another report's conclusion that tumor in the RLL had the most obvious tendency for LNM (27). Additionally, we also identified that several traditional and crucial clinical parameters, including pathological differentiation,

2564

Tian et al. Nomograms for early-stage NSCLC

histological type and tumor size, were all independent risk factors associated with LNM. According to our analysis, early-stage NSCLC patients with a tumor size of more than 30 mm and poorly differentiated or undifferentiated adenocarcinoma were more inclined to have a higher grade of lymph node invasion. Therefore, it seems that physicians take these patients more seriously than others and imaging screening is still a reliable and convenient pathway to detect metastasis (21).

The nomogram for distant metastasis incorporated seven clinical factors: the status of lymph node involvement, pathological differentiation, histological type, surgery type, the status of pleural invasion, age and sex. Similarly, Zhang *et al.* retrospectively analyzed 637 resected NSCLC patients and further developed a nomogram to forecast the development of distant metastasis based on three significant clinicopathologic characteristics, including histological type, pathological differentiation, degree of lymph nodes metastasis (22). Wang *et al.* highlighted that locally advanced NSCLC subjects with non-squamous pathological subtypes and more metastatic lymph nodes had a high risk of developing brain metastases (28). Our model also demonstrated that more aggressive LNM, NSCLC subtype adenocarcinoma and poor tumor differentiation were the primary factors responsible for distant metastases. Notably, lymph node involvement played a central role in facilitating distant metastasis in our nomogram. Indeed, a previous report also demonstrated that M1b NSCLC patients displayed higher proportions of N1, N2 and N3 than M1a patients did, and patients with aggressive N stage had a higher probability of developing multiorgan metastases (both P<0.001) (29,30). Intriguingly, our results concurred with those of several previous studies in which younger patients were characterized with an enhanced risk of lymph node invasion and distant metastasis (31,32). In our cohort, male gender was an independent risk factor for developing LNM and distant metastasis in early-stage NSCLC patients. NSCLC, distinct from breast cancer, is generally acknowledged as a hormone-independent cancer. Whether gender potentially exerts a momentous effect on various metastases in hormone-independent tumors needs further investigation (21). As for histological type, our results demonstrated that adenocarcinoma exhibited a higher risk of distant metastasis than other histological subtypes. Indeed, Oliveira *et al.* demonstrated that adenocarcinoma was the most frequent histological subtype (58.3%) to develop bone metastasis while squamous cell carcinoma was significantly related to the lowest risk of bone metastasis

(13.0%) (33). A prior clinical study also found that lung adenocarcinoma occupied over 50% of all the lung cancer cases with bone metastasis (34). Another tumor registry-based research performed in Sweden revealed up to 39% of bone metastasis occurrence in adenocarcinoma (11). Adenocarcinoma subjects also exhibited a 2.86-fold higher risk of brain metastases occurrence compared to non-adenocarcinoma NSCLC patients (22).

Our report had considerable strengths. Firstly, to our knowledge, we established the first predictive nomograms that could estimate the clinicopathological risk factors associated with LNM and distant metastasis in pT1 or pT2a NSCLC patients. Secondly, we extracted the largest number of cases with early-stage NSCLC from the SEER database and randomly classified them into a training cohort and an internal testing cohort, which made our conclusion more convincing. Thirdly, two user-friendly nomogram models performed well and harboured satisfactory accuracy and reliability. All significant risk factors in our nomograms were accessible and prevalent in clinical practice. Thus, our models were convenient for clinicians to perform the individualized risk prediction of each patient. Inevitably, our study has several limitations. Firstly, the representativeness of our sample was partial as all enrolled subjects were from the USA and clinical information concerning cases from additional countries or areas was devoid in the SEER database. Furthermore, the SEER database was devoid of certain crucial clinicopathologic factors of NSCLC, including smoking, body mass index, family history, driven mutations and genetic markers (35,36), which requires further investigations. Ultimately, our predictive nomograms were merely assessed by internal verification owing to the deficiency of data support from another database of our hospital (37). Thus, the applicability of our nomograms to early-stage NSCLC patients remains to be further confirmed through multicenter prospective clinical trials with a large sample size.

Conclusions

Based on the identified clinicopathologic risk predictors in a large retrospective population-based cohort, our study presented the first novel nomogram models with favorable predictive accuracy to make an individualized distinction and prediction for LNM and distant metastasis in resected patients with pT1 and pT2a NSCLC, which is instrumental in prognostic evaluation, personalized therapy and follow-up management in clinical practice.

Footnote

Reporting Checklist: The authors have completed the TRIPOD reporting checklist. Available at http://dx.doi.org/10.21037/apm-20-1756

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at http://dx.doi.org/10.21037/apm-20-1756). 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 study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The SEER database is publicly available and our study was performed based on the guideline of this database. All patient information was anonymized and de-identified in the SEER database. Thus, our study was exempted from the ethics committee approval and patients' informed consent.

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

1. Zhang Z, Tang H, Chen P, et al. Demystifying the manipulation of host immunity, metabolism, and extraintestinal tumors by the gut microbiome. Signal Transduct Target Ther 2019;4:41.
2. Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018;68:394-424.
3. Rotow J, Bivona TG. Understanding and targeting resistance mechanisms in NSCLC. Nat Rev Cancer 2017;17:637-58.
4. Thomas A, Liu SV, Subramaniam DS, et al. Refining the treatment of NSCLC according to histological and molecular subtypes. Nat Rev Clin Oncol 2015;12:511-26.
5. Liu WJ, Du Y, Wen R, et al. Drug resistance to targeted therapeutic strategies in non-small cell lung cancer. Pharmacol Ther 2020;206:107438.
6. Hirsch FR, Scagliotti GV, Mulshine JL, et al. Lung cancer: current therapies and new targeted treatments. Lancet 2017;389:299-311.
7. Zhang Z, Chen P, Xie H, et al. Using circulating tumor DNA as a novel biomarker to screen and diagnose hepatocellular carcinoma: A systematic review and meta-analysis. Cancer Med 2020;9:1349-64.
8. Tan WL, Chua KLM, Lin CC, et al. Asian Thoracic Oncology Research Group expert consensus statement on optimal management of stage III NSCLC. J Thorac Oncol 2020;15:324-43.
9. Diederich S. Lung cancer staging update: the revised TNM classification. Cancer Imaging 2010;10 Spec no A:S134-5.
10. Goldstraw P. The 7th Edition of TNM in Lung Cancer: what now? J Thorac Oncol 2009;4:671-3.
11. Riihimaki M, Hemminki A, Fallah M, et al. Metastatic sites and survival in lung cancer. Lung Cancer 2014;86:78-84.
12. Li J, Zhu H, Sun L, et al. Prognostic value of site-specific metastases in lung cancer: A population based study. J Cancer 2019;10:3079-86.
13. Ruffini E, Asioli S, Filosso PL, et al. Significance of the presence of microscopic vascular invasion after complete resection of Stage I-II pT1-T2N0 non-small cell lung cancer and its relation with T-Size categories: did the 2009 7th edition of the TNM staging system miss something? J Thorac Oncol 2011;6:319-26.
14. Goldstraw P, Chansky K, Crowley J, et al. The IASLC Lung Cancer Staging Project: Proposals for Revision of the TNM Stage Groupings in the Forthcoming (Eighth) Edition of the TNM Classification for Lung Cancer. J Thorac Oncol 2016;11:39-51.
15. Ashour Badawy A, Khedr G, Omar A, et al. Site of Metastases as Prognostic Factors in Unselected Population of Stage IV Non-Small Cell Lung Cancer. Asian Pac J

Cancer Prev 2018;19:1907-10.

16. Zhu WY, Fang KX, He JY, et al. A Prediction Rule for Overall Survival in Non-Small-Cell Lung Cancer Patients with a Pathological Tumor Size Less Than 30 mm. Dis Markers 2019;2019:8435893.

17. Kluth LA, Black PC, Bochner BH, et al. Prognostic and Prediction Tools in Bladder Cancer: A Comprehensive Review of the Literature. Eur Urol 2015;68:238-53.

18. Balachandran VP, Gonen M, Smith JJ, et al. Nomograms in oncology: more than meets the eye. Lancet Oncol 2015;16:e173-80.

19. Furqan M, Tien YY, Schroeder MC, et al. Lobar versus sub-lobar surgery for pulmonary typical carcinoid, a population-based analysis. J Thorac Dis 2018;10:5850-9.

20. Zhang Z, Xie H, Chen P, et al. Development and Identification of a Nomogram Prognostic Model for Patients with Primary Clear Cell Carcinoma of the Liver. Med Sci Monit 2020;26:e919789.

21. Mok TSK, Wu YL, Kudaba I, et al. Pembrolizumab versus chemotherapy for previously untreated, PD-L1-expressing, locally advanced or metastatic non-small-cell lung cancer (KEYNOTE-042): a randomised, open-label, controlled, phase 3 trial. Lancet 2019;393:1819-30.

22. Zhang F, Zheng W, Ying L, et al. A Nomogram to Predict Brain Metastases of Resected Non-Small Cell Lung Cancer Patients. Ann Surg Oncol 2016;23:3033-9.

23. Yang X, Pan X, Liu H, et al. A new approach to predict lymph node metastasis in solid lung adenocarcinoma: a radiomics nomogram. J Thorac Dis 2018;10:S807-S819.

24. Cerfolio RJ, Bryant AS, Eloubeidi MA. Routine mediastinoscopy and esophageal ultrasound fine-needle aspiration in patients with non-small cell lung cancer who are clinically N2 negative: a prospective study. Chest 2006;130:1791-5.

25. Gomez-Caro A, Garcia S, Reguart N, et al. Incidence of occult mediastinal node involvement in cN0 non-small-cell lung cancer patients after negative uptake of positron emission tomography/computer tomography scan. Eur J Cardiothorac Surg 2010;37:1168-74.

26. Shafazand S, Gould MK. A clinical prediction rule to estimate the probability of mediastinal metastasis in patients with non-small cell lung cancer. J Thorac Oncol 2006;1:953-9.

27. Jiang L, Jiang S, Lin Y, et al. Nomogram to Predict Occult N2 Lymph Nodes Metastases in Patients With Squamous Nonsmall Cell Lung Cancer. Medicine (Baltimore) 2015;94:e2054.

28. Wang SY, Ye X, Ou W, et al. Risk of cerebral metastases for postoperative locally advanced non-small-cell lung cancer. Lung Cancer 2009;64:238-43.

29. Yang J, Peng A, Wang B, et al. The prognostic impact of lymph node metastasis in patients with non-small cell lung cancer and distant organ metastasis. Clin Exp Metastasis 2019;36:457-66.

30. Dai C, Ren Y, Xie D, et al. Does Lymph Node Metastasis Have a Negative Prognostic Impact in Patients with NSCLC and M1a Disease? J Thorac Oncol 2016;11:1745-54.

31. Chen T, Zhou F, Jiang W, et al. Age at diagnosis is a heterogeneous factor for non-small cell lung cancer patients. J Thorac Dis 2019;11:2251-66.

32. 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.

33. Oliveira MB, Mello FC, Paschoal ME. The relationship between lung cancer histology and the clinicopathological characteristics of bone metastases. Lung Cancer 2016;96:19-24.

34. Sugiura H, Yamada K, Sugiura T, et al. Predictors of survival in patients with bone metastasis of lung cancer. Clin Orthop Relat Res 2008;466:729-36.

35. Fillon M. Tobacco control initiatives cut the number of lung cancer deaths in California by 28. CA Cancer J Clin 2019;69:83-5.

36. Karuppasamy R, Veerappapillai S, Maiti S, et al. Current progress and future perspectives of polypharmacology: From the view of non-small cell lung cancer. Semin Cancer Biol 2021;68:84-91.

37. Zheng XQ, Huang JF, Lin JL, et al. Incidence, prognostic factors, and a nomogram of lung cancer with bone metastasis at initial diagnosis: a population-based study. Transl Lung Cancer Res 2019;8:367-79.

Table S1 Clinical features of the included early-stage NSCLC patients

Variables	Total (n=22403)	Training (n=14935)	Test (n=7468)	P value
Age (mean, SD)	66.7±10.3	66.7±10.3	66.6±10.3	0.36
Gender				0.81
Female	12,221	8,156	4,065	
Male	10,182	6,779	3,403	
Race				0.15
White	18,574	12,433	6,141	
Black	2,004	1,316	688	
Unknown	1,825	1,186	639	
Primary site				0.63
RUL	7,517	4,979	2,538	
RML	1,357	932	425	
RLL	3,978	2,644	1,334	
LUL	5,747	3,843	1,904	
LLL	3,303	2,199	1,104	
Unknown	501	338	163	
Pathological differentiation				0.62
Well; I	4,786	3,174	1612	
Moderately; II	10,173	6,827	3346	
Poorly; III	7,142	4,731	2411	
Undifferentiated; IV	302	203	99	
pT stage				0.53
T1	12,935	8,601	4,334	
T2a	9,468	6,334	3,134	
Lymph node status*			5,983	0.47
N0	17,934	11,915	804	
N1	2,343	1,539	640	
N2	2,007	42	25	
N3	67	36	16	
NX	52	36	16	
Metastasis status				0.25
M0	21,810	14,529	7,281	
M1a	201	145	56	
M1b	392	261	131	
TNM stage				0.47
I	17,600	11,722	5,878	
II	2,262	1,485	777	
III	1,905	1,291	614	
IV	593	406	187	
Unknown	43	31	12	

Table S1 (*continued*)

Table S1 (*continued*)

Variables	Total (n=22403)	Training (n=14935)	Test (n=7468)	P value
Tumor size				0.29
≤30mm	16,576	10,877	5,499	
30-50 mm	5,961	4,017	1,944	
Unknown	66	41	25	
Histological type				0.79
Adenocarcinoma	13,471	8,972	4,499	
Squamous	5,376	3,569	1,807	
Large cell	398	265	133	
NSCLC (NOS)	3,158	2,129	1,029	
Pleural invasion				0.91
No invasion	15,480	10,322	5158	
Invasion	4,181	2,778	1403	
Unknown	2,742	1,835	907	
Surgery type				0.86
Wedge resection	4,338	2,891	1,447	
Lobectomy/Bilobectomy	17,560	11,713	5,847	
Pneumonectomy	505	331	174	
Distant metastasis				1
Negative	22,011	14,674	7,337	
Positive	392	261	131	
Bone metastasis				0.65
Negative	22,270	14,844	7,426	
Positive	95	63	32	
Unknown	38	28	10	
Brain metastasis				0.5
Negative	22,154	14,770	7,384	
Positive	210	136	74	
Unknown	39	29	10	
Liver metastasis				0.31
Negative	22,334	14,883	7,451	
Positive	29	22	7	
Unknown	40	30	10	
Lung metastasis				0.35
Negative	22,282	14,847	7,435	
Positive	75	54	21	
Unknown	46	34	12	
Overall survival				0.41
Alive	18,257	12,148	6,109	
Death	4,146	2,787	1,359	

*, N_x means that there is lymph node metastasis, but unknown for which stage. SD, standard deviation; RUL, right upper lobe; RML, right middle lobe; RLL, right lower lobe; LUL, left upper lobe; LLL, left lower lobe; pT, pathologic T; NSCLC (NOS), Non-small cell lung cancer, Not otherwise specified.

Table S2 Point distributions and predictive scores for every parameter in both nomograms

Variables	Nomogram score	
	LNM	Distant metastasis
Age		
≤65	12	23
>65	0	0
Gender		
Female	0	0
Male	5	13
Primary site		
RUL	0	–
RML	13	–
RLL	13	–
LUL	14	–
LLL	15	–
Pathological differentiation		
Well; I	0	0
Moderately; II	51	40
Poorly; III	64	59
Undifferentiated; IV	67	70
pT stage		
T1	0	–
T2a	25	–
Lymph node status		
N0	–	0
N1	–	25
N2	–	42
N3	–	100

Table S2 (*continued*)

Table S2 (*continued*)

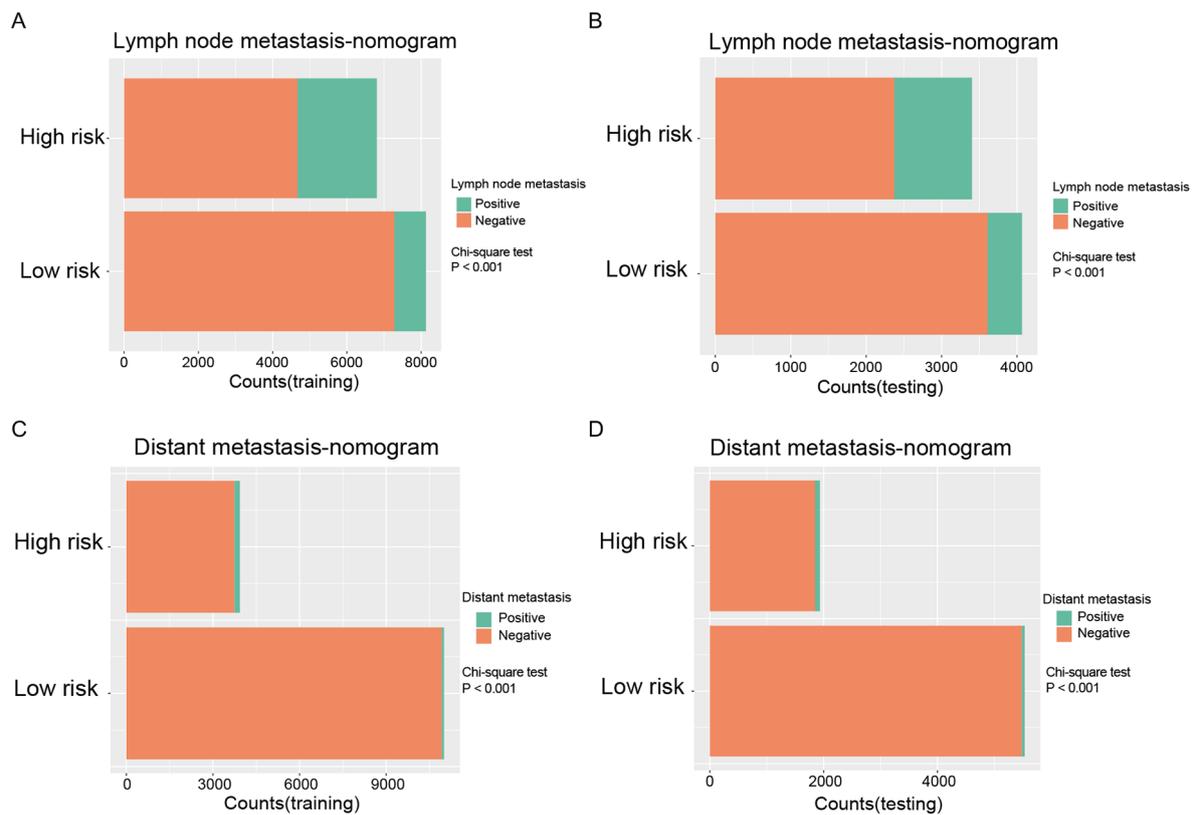| Variables	Nomogram score	
	LNM	Distant metastasis
Tumor size		
≤30mm	0	–
30-50 mm	16	–
Histological type		
Adenocarcinoma	23	30
Squamous	0	0
Large cell	9	42
NSCLC (NOS)	14	26
Pleural invasion		
No invasion	–	0
Invasion	–	22
Surgery type		
Wedge resection	0	39
Lobectomy/bilobectomy	28	1
Pneumonectomy	100	0
Distant metastasis		
Negative	0	–
Positive	52	–

Figure S1 Discriminatory efficiency of the predictive nomogram models for LNM and distant metastasis. Risk stratification for the nomogram models was performed through the maximum Youden index, and their performance in differentiating LNM and distant metastasis in (A,C) the training cohort and (B,D) the testing cohort were plotted.
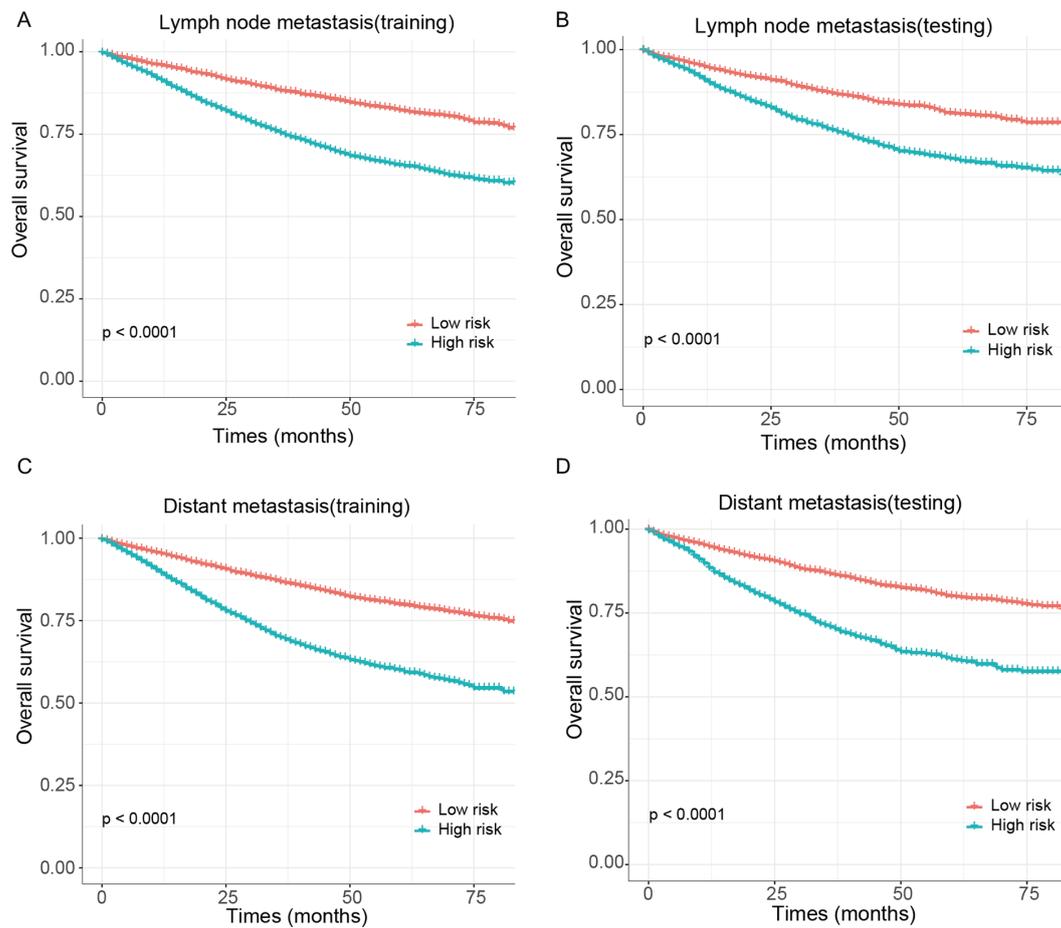
Figure S2 Survival analysis of patients with LNM and distant metastasis in (A,C) the training cohort and (B,D) the testing cohort, respectively. OS, overall survival.