Original Article

Establishing a prognostic scoring system and exploring prognostic value of examined lymph node numbers for stage I non-small cell lung cancer: a retrospective study of Surveillance, Epidemiology, and End Results (SEER) database and a Chinese cohort

Siyuan Wang[1], Xin Yin[2], Lingyun Wu[2], Hao Yu[1], Zhongjie Lu[2], Feng Zhao[2], Danfang Yan[2], Senxiang Yan[2]

[1]Zhejiang University School of Medicine, Zhejiang University, Hangzhou, China; [2]Division of Radiotherapy, Department of Radiation Oncology Center, The First Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China

Contributions: (I) Conception and design: S Wang, F Zhao, S Yan; (II) Administrative support: D Yan, S Yan; (III) Provision of study materials or patients: F Zhao, S Yan; (IV) Collection and assembly of data: S Wang, X Yin, L Wu, H Yu; (V) Data analysis and interpretation: S Wang, Z Lu; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Senxiang Yan, MD. Division of Radiotherapy, Department of Radiation Oncology Center, The First Affiliated Hospital, Zhejiang University School of Medicine, No. 79 Qingchun Road, Hangzhou 310003, China. Email: yansenxiang@zju.edu.cn.

Background: There is currently no recognized assessment system to predict disease outcomes for stage I non-small cell lung cancer (NSCLC). This research aimed to develop a prognostic scoring system for predicting 5-year overall survival (OS) of individuals with stage I NSCLC following definitive therapeutic intervention. Additionally, the optimal number of examined lymph nodes (ELNs) count for tumors no larger than 30 mm was determined.

Methods: Patients (n=22,617) diagnosed with stage I NSCLC from 2007 to 2015 who underwent definitive treatment (pulmonary lobectomy, pulmonary sublobectomy, or radiotherapy) were identified from the Surveillance, Epidemiology, and End Results (SEER) database. There were 400 Chinese patients with stage I NSCLC diagnosed in 2017 enrolled for external validation. The nomogram was constructed based on gradient boosting machine. The optimal ELNs in patients with tumors ≤30 mm and node-negative undergoing pulmonary lobectomy or pulmonary sublobectomy were determined using log-rank test and validated by multivariable analysis.

Results: Age at diagnosis, histology, differentiated grade, tumor staging, number of ELNs, and definitive treatment pattern were recognized as important factors for 5-year OS. The prognostic scoring system exhibited superior discrimination accuracy, calibration ability, and net clinical benefit compared to the tumor, node, metastasis (TNM) staging system. For patients with tumors ≤30 mm, more than 10 and 20 ELNs demonstrated the maximum OS difference during lobectomy and sublobectomy, respectively.

Conclusions: This prognostic scoring system will anticipate the prognosis of stage I NSCLC patients after radical treatment, thereby offering individualized treatment recommendations for both clinicians and patients. A minimum of 10 ELNs during lobectomy and 20 ELNs during sublobectomy are necessary for small-sized NSCLC.

Keywords: Lung cancer; prediction model; lymph nodes (LNs)

Submitted Aug 21, 2024. Accepted for publication Nov 26, 2024. Published online Jan 23, 2025.
doi: 10.21037/tcr-24-1474
View this article at: https://dx.doi.org/10.21037/tcr-24-1474

Introduction

Lung cancer ranks as one of the most prevalent forms of malignant neoplasms globally, accounting for more than 2.2 million new cases and approximately 1.8 million fatalities annually (1). Stage I non-small cell lung cancer (NSCLC) is typically viewed as an earlier stage in the progression of the disease, offering patients the potential to be cured and the possibility of long-term survival. Pulmonary lobectomy has been recommended as the first-line treatment for stage I NSCLC, as it has demonstrated an obvious survival benefit to patients (2). However, this surgical procedure typically requires a relatively well physical state of patients and will result in great respiratory function loss (3). Recently, sublobectomy and radiotherapy have emerged as alternative therapeutic options for stage I NSCLC (4), potentially offering minor trauma and improved tolerability compared to lobectomy (5-7). Previous studies focused more on predicting the postoperative survival time of early-stage lung cancer (8,9), yet none of the proposed prognostic scoring systems have included radiotherapy as a radical treatment method along with surgery. The tumor, node, metastasis (TNM) stage is a widely acknowledged prognostic factor for cancer patients. According to the 8th edition of the TNM staging system for lung cancer, the 5-year overall survival (OS) rates for patients with stage IA1, IA2, IA3, and IB NSCLC are 92%, 83%, 77%, and 68%, respectively. However, the impact of other clinicopathological factors on survival outcomes is yet to be fully understood, which may lead to overlooking some patients at risk of early death (10). Therefore, there is a pressing need to establish an accurate prognostic system to guide personalized treatment and evaluate the clinical curative effect in patients with stage I NSCLC.

Accurate lymph node (LN) assessment facilitates the determination of clinical stage classification and eliminates the risk of micrometastasis, thereby improving long-term survival. But in the practical work, patients with clinical stage IA NSCLC often undergo conservative LN sampling instead of radical lymphadenectomy. However, nearly 20 percent of patients will experience pathological T-stage migration due to the existence of high-risk factors (visceral pleural invasion; main bronchus involvement; atelectasis/obstructive pneumonitis extending to hilum) (11). In patients with a higher T stage, more examined lymph nodes (ELNs) should have been examined to confirm an N0 diagnosis (12). Hence, the potential high-risk factors should be foreseen before treatment, and identifying the ideal number of ELNs to be assessed is of significant importance in NSCLC patients with small-sized tumors and clinically LN-negative disease.

The objective of the present study was to identify the clinicopathological factors correlating with the survival outcome of patients diagnosed with stage I NSCLC and to develop an innovative prognostic scoring system for the prediction of the overall 5-year survival rate following radical treatment. The developed prediction model was validated in an external cohort from China. Moreover, the optimal number of ELNs in patients with tumor size ≤30 mm and pathological N0 disease was identified to assess its clinical and prognostic significance. We present this article in accordance with the TRIPOD reporting checklist (available at https://tcr.amegroups.com/article/view/10.21037/tcr-24-1474/rc).

Highlight box

Key findings

- We propose a prognostic scoring system for individuals with stage I non-small cell lung cancer (NSCLC) to predict 5-year overall survival (OS) following definitive treatment (pulmonary lobectomy, pulmonary sublobecomy, or radiotherapy). This model demonstrated excellent prediction performance in both American and Chinese cohorts. Additionally, stage I NSCLC patients with tumor size ≤30 mm should undergo more than 10 and 20 lymph node (LN) dissection during pulmonary lobectomy and sublobectomy, respectively.

What is known and what is new?

- Patients with stage I NSCLC typically have an optimistic prognosis after radical treatment; however, a subset may die within 5 years. Previous studies mainly focused on forecasting postoperative prognosis of early-stage NSCLC, but existing prognostic scoring systems have not included radiotherapy as a curative treatment option.
- Radical clearance of regional LNs is known to help eliminate occult micrometastasis. According to current clinical practice guidelines, systemic LN dissection is recommended as the committed step of lung cancer surgery. However, the relationship between the number of dissected LNs and survival benefit in stage I NSCLC with tumors ≤30 mm remains controversial.

What is the implication, and what should change now?

- As a complement to the tumor, node, metastasis staging system, this prognostic scoring system will enable doctors to make more accurate prognostic anticipation for patients following definitive treatment, thereby contributing to precise medicine for stage I NSCLC. We also propose the minimum number of examined LNs for small-sized stage I NSCLC, providing theoretical evidence for the development of surgical protocols.

Methods

Data collection

Data on patients newly diagnosed with stage I NSCLC between 2007 and 2015 were extracted from 17 population-based cancer registries in the Surveillance, Epidemiology, and End Results (SEER) program, utilizing the SEER*Stat software (version 8.4.3) (13). The SEER program, which is funded by the National Cancer Institute, encompasses roughly 30% of the U.S. population. Additionally, a retrospective collection of data from The First Affiliated Hospital of Zhejiang University was conducted, encompassing the period from January 1, 2017 to December 31, 2017. These data served as an external validation cohort for the study.

Tumor staging was conducted following the guidelines established by the American Joint Committee on Cancer (AJCC; the 8th edition), which categorized the stages as follows: T1aN0M0 for stage IA1, T1bN0M0 for stage IA2, T1cN0M0 for stage IA3, and T2aN0M0 for stage IB. Specifically, the non-sized T2a descriptors included tumors ≤40 mm with main bronchus involvement [Lung Collaborative Stage (CS) Extension codes 200, 210], associated with atelectasis/obstructive pneumonitis that extends to the hilar region (Lung CS Extension codes 400, 550), and tumors with visceral pleural invasion (Lung CS Extension codes 420, 430, 440, 450). The 3rd edition of the International Classification of Diseases for Oncology (ICD-O-3) was used for coding primary tumor location and cancer histology.

Patient selection

Patients diagnosed with stage I primary NSCLC (site codes: C34.0–C34.3) who had received standard definitive treatment were enrolled in the present study. The deadline date for follow-up was December 31, 2021. The histological types of NSCLC include the following (14): squamous cell carcinoma (8051, 8052, 8070–8076, 8078, 8083, 8084, 8090, 8094, 8123), adenocarcinoma (8015, 8050, 8140, 8141, 8143–8145, 8147, 8190, 8201, 8211, 8250–8255, 8260, 8290, 8310, 8320, 8323, 8333, 8401, 8440, 8470, 8471, 8480, 8481, 8490, 8503, 8507, 8550, 8570–8572, 8574, 8576), large cell carcinoma (8012–8014, 8021, 8034, 8082), as well as not otherwise specified (8046, 8003, 8004, 8022, 8030, 8031–8033, 8035, 8120, 8200, 8240, 8241, 8243–8246, 8249, 8430, 8525, 8560, 8562, 8575). Standard treatment procedures included lobectomy (surgery codes: 30, 33, 45–48), sublobectomy (surgery codes: 21, 22), and radiotherapy following NCCN Guidelines® Insights: Non-Small Cell Lung Cancer, Version 3.2024 (4). The exclusion criteria were: (I) age ≤18 years; (II) receiving neoadjuvant systemic therapy; (III) receiving perioperative radiotherapy; (IV) died within 3 months; (V) missing or incomplete information about key variables. Inclusion criteria as well as exclusion criteria were consistent between the development and validation cohort. The detailed inclusion and exclusion process are presented in Figure S1.

From the SEER database, a range of demographic and clinical variables were extracted, including: patient ID, year of diagnosis, age at diagnosis, sex, primary site, CS Tumor Size/Extension Evaluation, tumor size, tumor extension, LNs status, distant metastasis, histological subtype, differentiation grade, AJCC T stage, N stage, M stage, ELNs, positive LN, surgical procedure, radiation recode, surgery/radiation sequence, systemic therapy/surgery sequence, follow-up time, year of follow-up recode, survival months, survival status, year of death, cause of death.

For the external validation cohort, the electronic medical record system of The First Affiliated Hospital of Zhejiang University was used to collect the baseline characteristics of patients diagnosed with stage I NSCLC. A senior physician, blinded to other predictor variables, was responsible for determining tumor staging based on imaging and pathological data. The survival time and survival status were obtained through telephone follow-up. The date of the last follow-up was May 3, 2024. The investigators involved in the follow-up were blind to the baseline information and intervention measures of all participants.

Study endpoints

The primary endpoint was OS. It was measured from the date of diagnosis until either the patient's death from any cause or the final follow-up. The secondary endpoint was cause-specific survival (CSS), which was defined as the interval between the diagnosis and the occurrence of death due to lung cancer or the last follow-up. The Cox regression for CSS will serve as a sensitivity analysis to compare with the regression results of the primary endpoint analysis.

Statistical analysis

Continuous variable data were presented as the mean ± standard deviation or median and interquartile range

(IQR) according to the characteristics of data distribution. Categorical variables were presented as frequency distribution and proportion. The *t*-test, analysis of variance, or the Mann-Whitney *U* test was utilized to evaluate disparities in continuous variables. For nominal and ordinal categorical variables, differences were assessed using the chi-square test and the Mann-Whitney *U* test, respectively. Univariate and multivariate Cox proportional hazards regression analyses were conducted to ascertain the independent prognostic factors associated with OS as well as CSS. The median period of follow-up and its IQR for the training and validation cohort were determined using the reverse Kaplan-Meier method. There were no missing values in the training cohort as well as validation cohort.

Subgroup analysis

To determine the optimal number of dissected regional ELNs in clinical stage IA NSCLC, individuals with tumors ≤30 mm and pN0 were divided into the lobectomy and sublobectomy cohorts. The Mann-Whitney *U* test was applied to compare the number of ELNs between lobectomy and sublobectomy cohorts. Log-rank test was performed for each cohort to assess the optimal number of ELNs for maximum OS benefit. Multivariate Cox proportional hazard models, adjusted for age, sex, primary tumor site, histological type, differentiation grade, tumor staging, and additional systemic therapy, were established to account for covariates.

Establishment of prognostic model

A gradient boosting machine (GBM) model based on the decision tree learning method was trained to determine the prognostic value of each predictor to OS. The sample size of individuals incorporated into the training cohort followed the rule of 10 events per variable (15). We employed a sequence of 10,000 decision trees, ensuring that each terminal node contained a minimum of 10 observations. The depth of the decision tree was fine-tuned to 3, and the shrinkage parameter was set to 0.001 to optimize the model. Utilizing the outcomes from the GBM analysis, covariates that had a relative influence exceeding 0.05 were selected and incorporated into the construction of a nomogram. This nomogram was designed to predict the 5-year OS. The process involved summing the scores assigned to each predictive variable, which then allowed for the calculation of an individual's risk score.

The accuracy of the prognostic scoring system was confirmed using several methods: boxplots provided a visual representation of score distribution; the C-index measured the model's discriminatory ability; and receiver operating characteristic (ROC) curves were utilized to evaluate the model's performance in classifying outcomes. Calibration plots were created using the bootstrapping technique with 500 resampling iterations to determine the agreement between the predicted and actual survival probabilities. Decision curve analyses (DCAs) were employed to evaluate the clinical value of the new prognostic scoring system and to measure the net benefit at various threshold probabilities. The restricted cubic spline (RCS) method was applied to fit the relationship between the nomogram score and the risk of all-cause death, and further determine the optimal cut-off points of the nomogram score for risk stratification. The Log-rank test was used to compare the OS of patients across different risk groups. Additionally, the integrated discrimination improvement (IDI) as well as net reclassification improvement (NRI) were computed to compare how well the prediction model performed in comparison to the TNM staging system.

All analyses above were carried out with "survival", "rms", "gbm", "survminer", "ggplot2", "ggpubr", "RcolorBrewer", "survcomp", "timeROC", "ggDCA", "rcssci", "nricens" packages using R software version 4.2.2 (R Foundation for Statistical Computing, Vienna, Austria). Statistical significance for the study was established using a two-sided P value threshold of less than 0.05. It should be noted that all staff and participants, except for statisticians, were kept masked to outcome measurements and analysis results.

This study was conducted following the Declaration of Helsinki (revised 2013) and approved by the Ethics Committee of The First Affiliated Hospital of Zhejiang University (No. IIT20240197B). Individual consent for this retrospective analysis was waived.

Results

Patient characteristics

A total of 22,617 individuals diagnosed with stage I NSCLC who received definitive treatment between the years 2007 and 2015 were identified from the SEER database. Besides, an additional 400 individuals with stage I NSCLC from The First Affiliated Hospital of Zhejiang University, China, were included to serve as an external validation group.

Table 1 presents a comprehensive overview of the baseline demographic and clinicopathological features of the patients included in the analysis.

In the training cohort, there were 9,709 men and 12,908 women who were diagnosed, with the median age being 69 years and an IQR of 62 to 76 years. Almost all NSCLC cases (n=22,518) originated from the pulmonary lobe, whereas those from the primary bronchogenic carcinoma (n=99) were extremely rare. The predominant histology type was adenocarcinoma, representing 13,983 (61.8%) of tumors. More than half of the tumors were moderately to poorly differentiated. A large proportion of patients were diagnosed at stage IA (IA1: 7.3%, IA2: 36.0%, IA3: 26.9%), whereas 29.8% were diagnosed at stage IB. There were 6,151 (27.2%) patients who had no LN examined, while others had at least one ELNs (1–10: 50.3%, 11–15: 11.6%, 16–20: 5.6%, >20: 5.3%). Regarding treatment, 14,426 (63.8%) patients underwent pulmonary lobectomy, 4,021 (17.8%) underwent sublobectomy, 4,170 (18.4%) received radiation as definitive therapy. Additional systemic therapy was administered to 872 (3.9%) patients.

In the training cohort, 12,915 (57.1%) patients died, with a median follow-up of 116 months, of which 6,023 (26.6%) died due to NSCLC. In the validation cohort, 28 (7.0%) patients had died by the end of the study with a median follow-up period of 80 months. Among these patients, 18 of them (4.5%) died specifically due to lung cancer.

Cox regression analysis

Among patients with stage I NSCLC, older age, male sex, squamous cell carcinoma and large cell carcinoma, lower differentiation grade, more advanced stage, fewer number of ELNs, and curative treatment method were identified as independent prognostic factors and were positively associated with worse OS (*Table 2*). Sensitive analysis further confirmed that these factors also influenced CSS independently (*Table 3*). However, the survival advantage of additional systemic therapy was not observed across the entire cohort.

Construction of nomogram

According to the result of GBM analysis (*Table 4*), the curative treatment method was the most important prognostic factor for OS among patients with stage I NSCLC, followed by age at diagnosis. The ranking of the relative importance of these variables is shown in *Figure 1*. Then, those vital clinicopathological factors, including curative treatment method, age at diagnosis, differentiation grade, number of ELNs, histology, and clinical stage, were incorporated into the final prognostic scoring system (*Figure 2*). All selected variables were also independent prognostic factors for both OS and CSS. The score of each parameter and their corresponding 5-year OS rates were presented in *Table 5*.

Calibration and validation of the nomogram

The nomogram score for each patient was calculated and graphically represented as a boxplot (*Figure 3A*), indicating that the developed prognostic scoring system could be a relatively reliable predictor for estimating the likelihood of survival or death in individuals with stage I NSCLC. Within the primary cohort, the area under the curve (AUC) of the developed nomogram as well as TNM staging system was 0.761 and 0.581, respectively (*Figure 3B*). Comparing their performance for predicting 5-year OS, the C-index for the nomogram and TNM staging was 0.712 [95% confidence interval (CI): 0.707–0.716] and 0.590 (95% CI: 0.583–0.597) within the training cohort. The NRI values for 5-year all-cause deaths were 0.503, and the corresponding IDI values were 0.178. The calibration ability of the nomogram and TNM staging system was similar, but the discriminative ability of the nomogram was superior to that of TNM staging system (*Figure 3C*). DCA demonstrated a positive net benefit for the nomogram across a broad range of threshold probabilities (*Figure 3D*).

Within the external validation cohort, the AUC value for forecasting 5-year OS was 0.794 based on the prognostic scoring system, and was 0.716 for TNM staging system. The C-index for the nomogram was 0.825 (95% CI: 0.694–0.956), whereas for the TNM staging system, it was 0.760 (95% CI: 0.625–0.895), in the validation cohort.

Nomogram performance in risk stratification

To deeply analyze the clinical utility of this prognostic scoring system, we introduced a risk-stratification framework based on the total score of each patient in the training cohort. According to the RCS analysis (*Figure 4*), the risk of all-cause death showed an exponential growth trend with respective to the nomogram score. The cut-off score of the nomogram was 210, corresponding to the inflection point of rapidly increasing risk of all-cause death.

Table 1 Comparison of demographic and clinicopathological characteristics between the training and validation cohorts

Characteristics	Training cohort (n=22,617), n (%)	Validation cohort (n=400), n (%)	P value
Age (years)			<0.001
<65	7,399 (32.7)	275 (68.8)	
65–74	8,307 (36.7)	97 (24.3)	
≥75	6,911 (30.6)	28 (6.9)	
Sex			0.47
Female	12,908 (57.1)	221 (55.3)	
Male	9,709 (42.9)	179 (44.7)	
Primary site			0.35
Pulmonary lobe	22,518 (99.6)	400 (100.0)	
Main bronchus	99 (0.4)	0 (0.0)	
Histology			<0.001
Adenocarcinoma	13,983 (61.8)	355 (88.8)	
Squamous cell carcinoma	5,481 (24.2)	33 (8.3)	
Large cell carcinoma	396 (1.8)	2 (0.5)	
Not otherwise specified	2,757 (12.2)	10 (2.4)	
Grade			<0.001
I	4,627 (20.5)	105 (26.3)	
II	8,789 (38.9)	217 (54.3)	
III–IV	6,002 (26.5)	70 (17.4)	
Unknown	3,199 (14.1)	8 (2.0)	
Tumor stage			<0.001
IA1	1,660 (7.3)	122 (30.5)	
IA2	8,146 (36.0)	175 (43.8)	
IA3	6,079 (26.9)	44 (11.0)	
IB	6,732 (29.8)	59 (14.7)	
Examined lymph nodes			<0.001
0	6,151 (27.2)	15 (3.8)	
1–10	11,378 (50.3)	183 (45.8)	
11–15	2,627 (11.6)	94 (23.5)	
16–20	1,262 (5.6)	49 (12.3)	
>20	1,199 (5.3)	59 (14.6)	
Treatment			<0.001
Lobectomy	14,426 (63.8)	233 (58.3)	
Sublobectomy	4,021 (17.8)	164 (41.0)	
Radiotherapy	4,170 (18.4)	3 (0.7)	
Additional systemic therapy			0.25
No	21,745 (96.1)	389 (97.3)	
Yes	872 (3.9)	11 (2.7)	

Grade I: well-differentiated. Grade II: moderately-differentiated. Grade III: poorly-differentiated. Grade IV: undifferentiated.

Table 2 Univariate and multivariate Cox analysis of OS with stage I non-small cell lung cancer patients

Characteristics	Dead patients		Hazard ratio (95% CI)		Adjusted P value
	With all cause (n=12,915)	Without any cause (n=9,702)	Univariate	Multivariate	
Age (years)					
<65	2,784	4,615	Reference	Reference	
65–74	4,732	3,575	1.81 (1.72–1.89)	1.61 (1.54–1.69)	<0.001
≥75	5,399	1,512	3.17 (3.03–3.32)	2.34 (2.23–2.45)	<0.001
Gender					
Female	6,688	6,220	Reference	Reference	
Male	6,227	3,482	1.43 (1.38–1.48)	1.34 (1.29–1.38)	<0.001
Primary site					
Pulmonary lobe	12,854	9,664	Reference	Reference	
Main bronchus	61	38	1.35 (1.05–1.74)	1.13 (0.88–1.46)	0.34
Histology					
Adenocarcinoma	7,300	6,683	Reference	Reference	
Squamous cell carcinoma	4,108	1,373	1.87 (1.80–1.94)	1.21 (1.16–1.26)	<0.001
Large cell carcinoma	282	114	1.65 (1.47–1.86)	1.44 (1.27–1.62)	<0.001
Not otherwise specified	1,225	1,532	0.84 (0.79–0.89)	0.88 (0.82–0.93)	<0.001
Differentiate grade					
I	1,652	2,975	Reference	Reference	
II	5,063	3,726	1.94 (1.84–2.05)	1.70 (1.60–1.80)	<0.001
III–IV	4,083	1,919	2.64 (2.49–2.80)	1.98 (1.87–2.11)	<0.001
Unknown	2,117	1,082	2.73 (2.56–2.91)	1.45 (1.35–1.55)	<0.001
Tumor stage					
IA1	663	997	Reference	Reference	
IA2	4,158	3,988	1.40 (1.29–1.52)	1.20 (1.10–1.30)	<0.001
IA3	3,717	2,362	1.88 (1.73–2.04)	1.43 (1.32–1.56)	<0.001
IB	4,377	2,355	2.09 (1.93–2.27)	1.68 (1.54–1.83)	<0.001
Examined lymph nodes					
0	4,956	1,195	Reference	Reference	
1–10	5,764	5,614	0.38 (0.36–0.39)	0.75 (0.70–0.80)	<0.001
11–15	1,159	1,468	0.32 (0.30–0.34)	0.63 (0.58–0.69)	<0.001
16–20	508	754	0.29 (0.26–0.32)	0.57 (0.51–0.63)	<0.001
>20	528	671	0.32 (0.29–0.35)	0.61 (0.55–0.68)	<0.001
Treatment					
Lobectomy	6,847	7,579	Reference	Reference	
Sublobectomy	2,365	1,656	1.47 (1.40–1.54)	1.32 (1.25–1.40)	<0.001
Radiotherapy	3,703	467	4.15 (3.98–4.32)	2.40 (2.22–2.59)	<0.001
Additional systemic therapy					
No	12,433	9,312	Reference	Reference	
Yes	482	390	0.93 (0.85–1.02)	1.02 (0.93–1.12)	0.69

Grade I: well-differentiated. Grade II: moderately-differentiated. Grade III: poorly-differentiated. Grade IV: undifferentiated. OS, overall survival; CI, confidence interval.

Table 3 Univariate and multivariate Cox analysis of CSS with stage I non-small cell lung cancer patients

Characteristics	Dead patients		Hazard ratio (95% CI)		Adjusted P value
	With lung cancer (n=6,023)	Without lung cancer (n=16,594)	Univariate	Multivariate	
Age (years)					
<65	1,466	5,933	Reference	Reference	
65–74	2,268	6,039	1.57 (1.47–1.68)	1.39 (1.30–1.49)	<0.001
≥75	2,289	4,622	2.30 (2.15–2.45)	1.61 (1.51–1.73)	<0.001
Gender					
Female	3,088	9,820	Reference	Reference	
Male	2,935	6,774	1.43 (1.36–1.50)	1.31 (1.25–1.38)	<0.001
Primary site					
Pulmonary lobe	5,987	16,531	Reference	Reference	
Main bronchus	36	63	1.69 (1.22–2.34)	1.20 (0.86–1.67)	0.27
Histology					
Adenocarcinoma	3,561	10,422	Reference	Reference	
Squamous cell carcinoma	1,745	3,736	1.55 (1.46–1.64)	0.95 (0.89–1.01)	0.09
Large cell carcinoma	151	245	1.81 (1.53–2.12)	1.43 (1.21–1.69)	<0.001
Not otherwise specified	566	2,191	0.80 (0.73–0.87)	0.81 (0.74–0.89)	<0.001
Differentiate grade					
I	611	4,016	Reference	Reference	
II	2,315	6,474	2.32 (2.13–2.54)	2.06 (1.88–2.26)	<0.001
III–IV	2,036	3,966	3.39 (3.10–3.71)	2.54 (2.31–2.79)	<0.001
Unknown	1,061	2,138	3.45 (3.12–3.81)	1.72 (1.54–1.91)	<0.001
Tumor stage					
IA1	248	1,412	Reference	Reference	
IA2	1,699	6,447	1.50 (1.32–1.72)	1.28 (1.12–1.46)	<0.001
IA3	1,698	4,381	2.22 (1.94–2.54)	1.67 (1.46–1.92)	<0.001
IB	2,378	4,354	2.94 (2.58–3.35)	2.32 (2.03–2.66)	<0.001
Examined lymph nodes					
0	2,514	3,637	Reference	Reference	
1–10	2,575	8,803	0.37 (0.35–0.39)	0.76 (0.69–0.84)	<0.001
11–15	508	2,119	0.30 (0.28–0.33)	0.63 (0.56–0.72)	<0.001
16–20	224	1,038	0.28 (0.24–0.32)	0.56 (0.48–0.66)	<0.001
>20	202	997	0.27 (0.23–0.31)	0.54 (0.45–0.64)	<0.001
Treatment					
Lobectomy	2,985	11,441	Reference	Reference	
Sublobectomy	1,069	2,952	1.46 (1.36–1.57)	1.42 (1.30–1.54)	<0.001
Radiotherapy	1,969	2,201	4.28 (4.04–4.54)	2.88 (2.58–3.23)	<0.001
Additional systemic therapy					
No	5,710	16,035	Reference	Reference	
Yes	313	559	1.34 (1.20–1.51)	1.27 (1.13–1.43)	<0.001

Grade I: well-differentiated. Grade II: moderately-differentiated. Grade III: poorly-differentiated. Grade IV: undifferentiated. CSS, cause-specific survival; CI, confidence interval.

Table 4 Variables included in the GBM model and their relative influence for OS of stage I non-small cell lung cancer

Variables	Relative importance
Treatment	33.72
Age	22.94
Grade	14.36
Nodes	10.98
Histology	7.50
Stage	5.55
Sex	4.52
Additional systemic therapy	0.34
Primary site	0.08

GBM, gradient boosting machine; OS, overall survival.

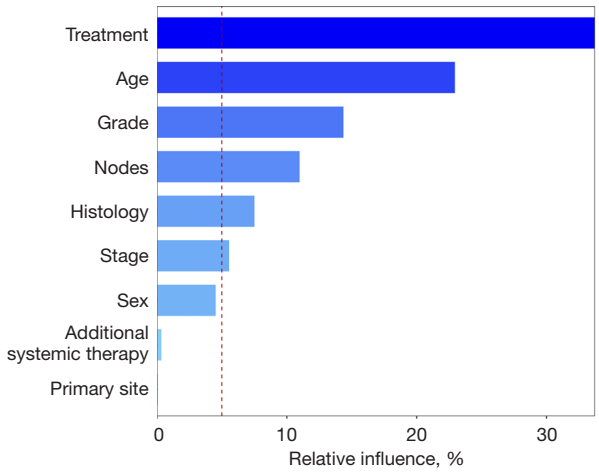

Figure 1 Overview of each variable's relative influence on OS of patients with stage I non-small cell lung cancer in the GBM model. OS, overall survival; GBM, gradient boosting machine.

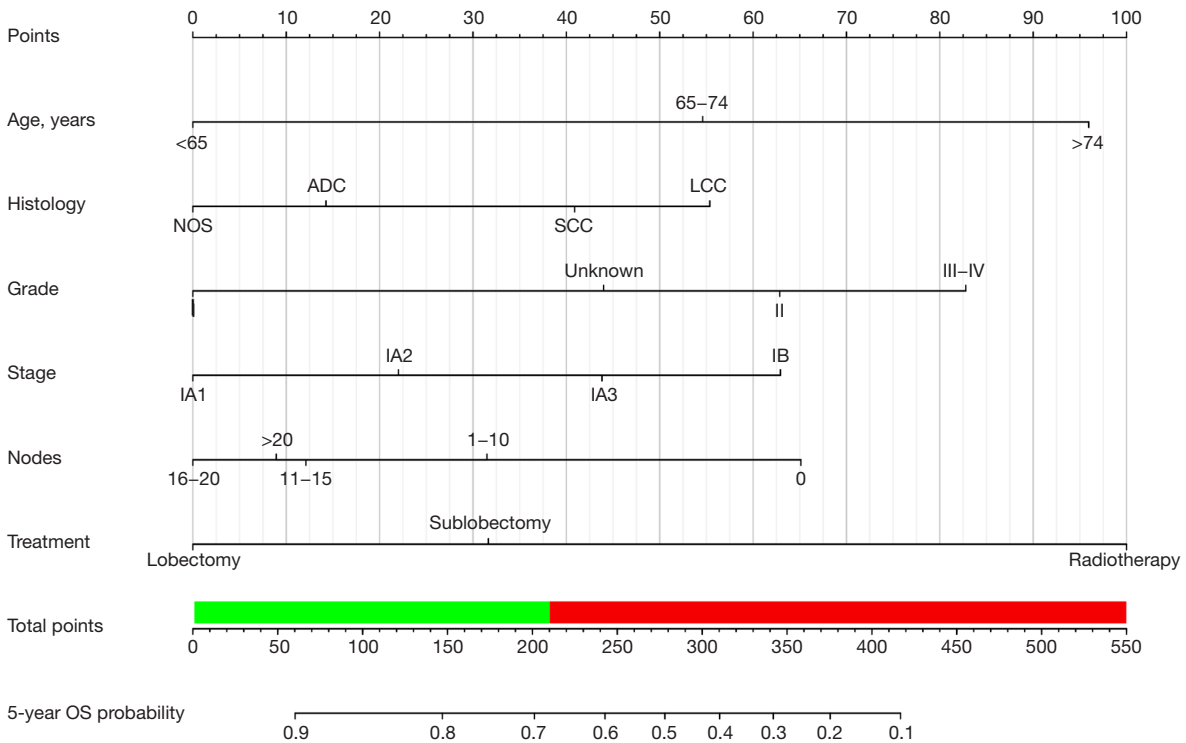

Figure 2 Prognostic score model for predicting the probability of 5-year OS in patients with stage I non-small cell lung cancer following curative treatment. Grade I: well-differentiated. Grade II: moderately-differentiated. Grade III: poorly-differentiated. Grade IV: undifferentiated. NOS, not otherwise specified; ADC, adenocarcinoma; SCC, squamous cell carcinoma; LCC, large cell carcinoma; OS, overall survival.

Table 5 The score of each parameter of the nomogram and the corresponding 5-year OS rate in the development cohort

Characteristics	Score	5-year OS (%)
Age (years)		
<65	0	75.8
65–74	55	63.2
≥75	96	47.1
Histology		
Adenocarcinoma	14	66.9
Squamous cell carcinoma	41	48.2
Large cell carcinoma	55	47.2
Not otherwise specified	0	70.1
Differentiate grade		
I	0	80.9
II	63	63.8
III–IV	83	52.7
Unknown	44	50.1
Tumor stage		
IA1	0	77.5
IA2	22	68.4
IA3	44	58.7
IB	63	54.8
Examined lymph nodes		
0	65	37.1
1–10	32	70.0
11–15	12	75.4
16–20	0	76.5
>20	9	76.6
Treatment		
Lobectomy	0	72.9
Sublobectomy	32	62.2
Radiotherapy	100	26.3

Grade I: well-differentiated. Grade II: moderately-differentiated. Grade III: poorly-differentiated. Grade IV: undifferentiated. OS, overall survival.

Individuals with scores below 210 were designated as the low-risk group, while those with scores of 210 or above were classified as the high-risk group. The median OS of the high-risk group was 51 months (95% CI: 50–53). Conversely, the rate of all-cause mortality was below the threshold of 50% in the low-risk group; thus, the median OS in the low-risk group could not be calculated. Moreover, using the same cut-off value, our prediction model also successfully identified high risk patients from the Chinese cohort. The Kaplan-Meier survival curves for the different risk subgroups are depicted in *Figure 5*.

Subgroup analysis

In the subgroup analysis of stage I NSCLC with small tumor size (≤30 mm), the impact of the number of tumor-free ELNs on OS was assessed across two separate patient populations. The extent of LN examination between the lobectomy cohort (median: 7; IQR: 4–12) and the sublobectomy cohort (median: 1; IQR: 0–4) were of significant difference (P<0.001). Within the cohort of patients who underwent pulmonary lobectomy, the 5-year OS rates varied based on the number of ELNs: 65.6% in patients who had no LNs examined, 73.7% in those with 1–10 ELNs, 78.2% in those with 11–15 ELNs, 78.1% in those with 16–20 ELNs, and 79.6% in those with more than 20 ELNs (*Figure 6A*). The most significant disparity in OS was observed in patients who had 10 or fewer ELNs compared with those who had >10 ELNs (P<0.001; *Figure 6B*). In the pulmonary sublobectomy cohort, the individuals who had no LNs examined had a 5-year OS rate of 57.4%, and those with more than 20 ELNs was 80.0%, which was higher than those with 1–10 ELNs (67.2%), 11–15 ELNs (73.4%) and 16–20 ELNs (69.4%), as shown in *Figure 6C*. The hazard ratio for mortality reached the minimum with more than 20 tumor-free regional LNs (P<0.01; *Figure 6D*).

A multivariate analysis was performed that included age, sex, primary site, histology, differentiation grade, tumor stage, additional systemic therapy, and the number of ELNs. After controlling potential influencing factors, the number of ELNs remained an independent, significant predictor of long-term OS in both the lobectomy (*Table 6*)

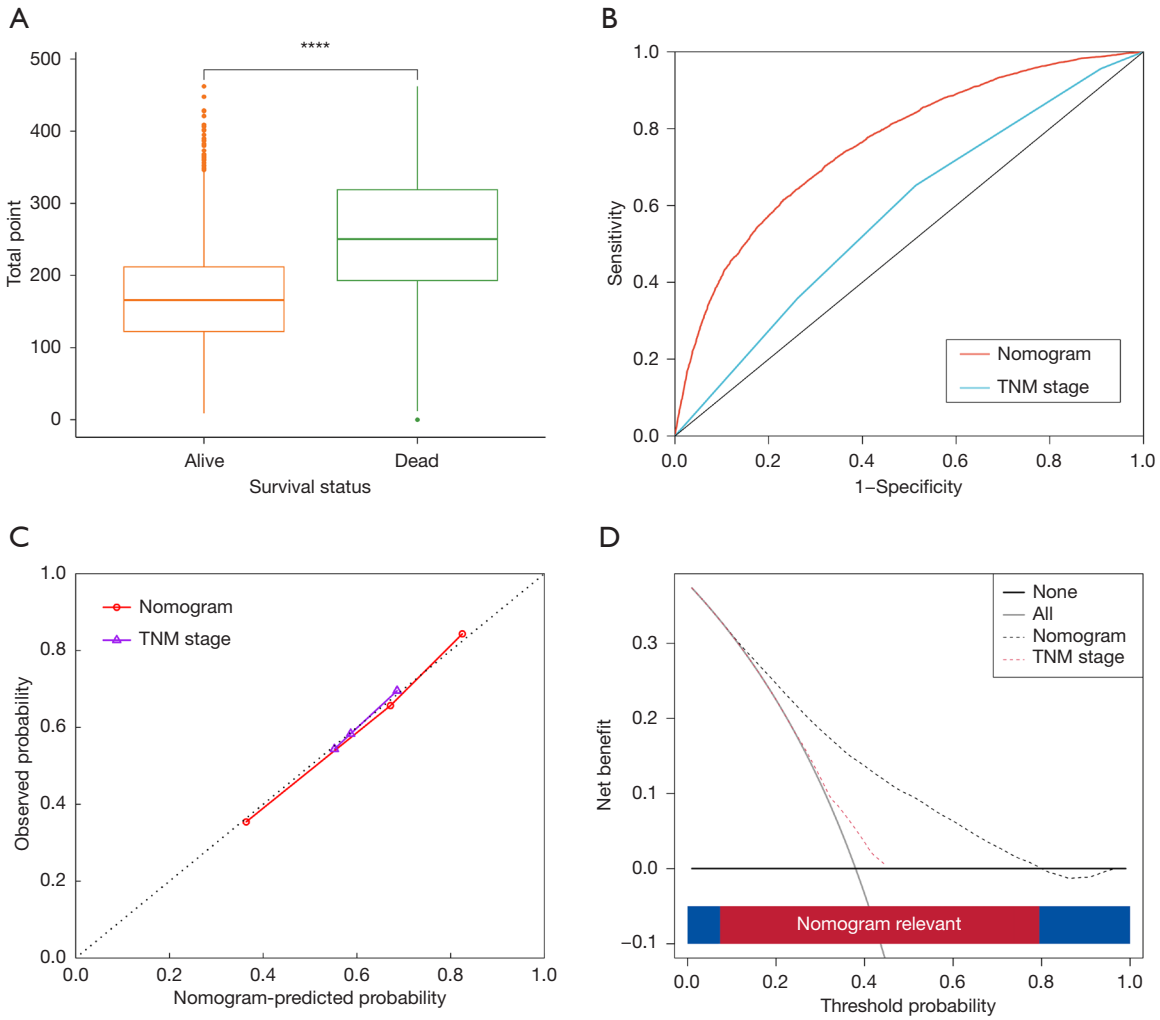

Figure 3 Validation of the performance of the prognostic scoring system for predicting 5-year OS. (A) Boxplot for the nomogram score of the survivors and the dead. (B) Receiver operating characteristic curves of the nomogram and TNM staging system. (C) Calibration curves of the nomogram and the TNM staging system. (D) Decision curve analysis for the nomogram and TNM staging system. ****, P<0.0001. TNM, tumor, node, metastasis; OS, overall survival.

and sublobectomy cohorts (*Table* 7).

Discussion

Our research used GBM to analyze the impact of clinicopathological factors on OS of stage I NSCLC, and selected key variables to construct the new prognostic scoring system to predict 5-year OS of individuals with stage I NSCLC after definitive treatment (including lobectomy, sublobectomy, and radiotherapy). This machine learning-based prognostic scoring system had good performance both in the training and validation cohorts, highlighting its clinical application value of guiding individualized treatment and anticipating prognosis. For NSCLC patients with tumor size ≤30 mm and node-negative disease, more than 10 ELNs was associated with an improved OS for those receiving pulmonary lobectomy, whereas examining more than 20 LNs during pulmonary sublobectomy was beneficial to long-term survival.

In our study, all-cause death and lung cancer-related death probability increased in the elderly and male individuals with stage I NSCLC, which is consistent with previous

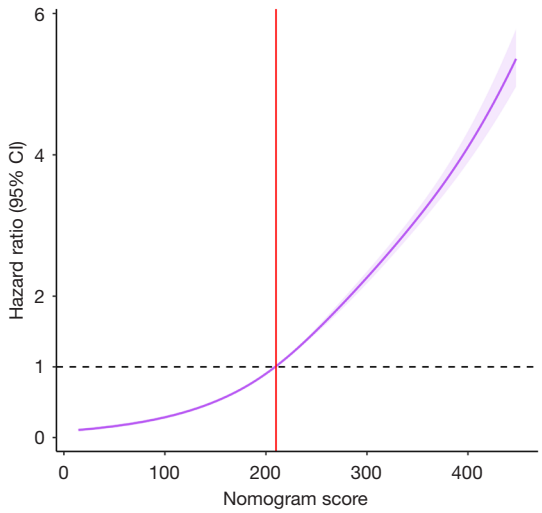

Figure 4 Restricted cubic spline model for the hazard ratio for all-cause mortality according to nomogram score on a continuous scale. CI, confidence interval.

studies (9,16). NSCLC patients with main bronchus tumor generally have a poorer prognosis than those with lobe-specific lung cancer (17). To our knowledge, we report that tumors located in the main bronchus are associated with adverse prognosis; however, the primary tumor site was not an independent prognostic factor in stage I NSCLC. This discrepancy can be explained by the fact that patients in all clinical stages of NSCLC were retrospectively enrolled in the previous study, and main bronchus tumors of advanced stage might have a higher risk of direct invasion to vital adjacent organs than those in early stage (17). Therefore, the primary tumor site may not matter as much in stage I NSCLC; nevertheless, comprehensive examination and aggressive treatment should be considered when there are tumors originating from the main bronchus.

NSCLC is a histologically heterogeneous malignant tumor consisting of a mixture of pathological types, the two most prevalent types of which are adenocarcinoma and squamous carcinoma (18). In this study, stage I lung adenocarcinoma had longer OS than squamous carcinoma, although there was no significant difference in CSS, as previously reported. Nevertheless, a single-center retrospective study involving 356 patients with stage I NSCLC who underwent lung operation concluded that adenocarcinoma was associated with more than twice the risk of tumor recurrence than did non-adenocarcinoma (19). The observed differences can be ascribed to the later onset age for squamous cell carcinoma compared to

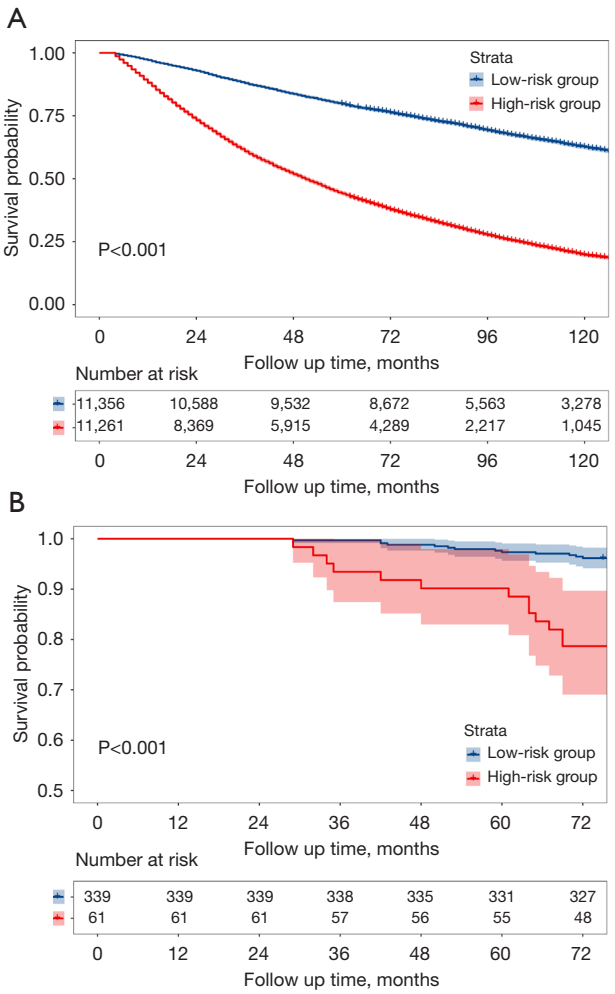

Figure 5 Kaplan-Meier curves for the comparison of OS for patients at low-risk and high-risk in the primary cohort (A) and the external verification cohort (B). OS, overall survival.

adenocarcinoma, along with the reduced life expectancy, typically seen in elderly lung cancer patients when juxtaposed with their younger counterparts (20). Moreover, the frequency of driver gene mutation is much higher in adenocarcinoma than in other histological types, implying that, if tumor recurrence or metastasis occurs, more patients with adenocarcinoma would receive targeted and immune therapies to prolong their survival. Concurrently, in this study, a multivariate Cox regression analysis indicated that the degree of tumor differentiation was a significant factor influencing the survival rates of individuals with stage I NSCLC. In 2011, the terms adenocarcinoma *in situ* and minimally invasive adenocarcinoma were introduced to better characterize lung lesions formerly

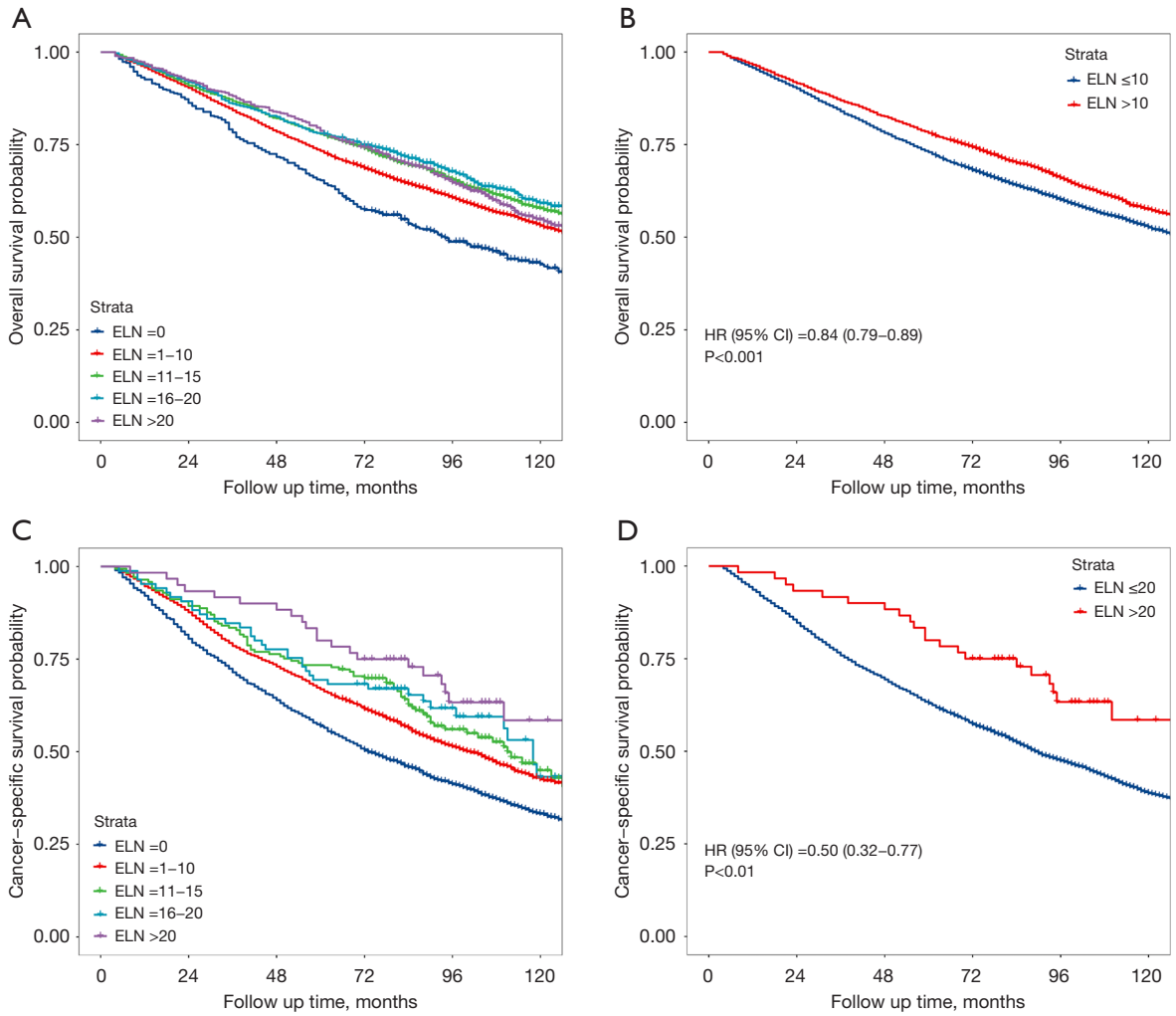

Figure 6 Kaplan-Meier curves of OS for different groups of examined lymph nodes in the lobectomy subgroup (A,B) and sublobectomy group (C,D). ELN, examined lymph node; HR, hazard ratio; CI, confidence interval; OS, overall survival.

termed bronchioloalveolar carcinoma; adenocarcinoma *in situ* has been defined as precursor glandular lesion to differentiate from lung adenocarcinoma (18,21). Since the novel histology classification has not been incorporated into the SEER database, a limited number of individuals diagnosed with lung adenocarcinoma in this study would more accurately be categorized as a separate histology-independent subgroup, which is associated with a favorable prognosis.

Ever since the 1960s, pulmonary lobectomy has continued to be the preferred treatment for individuals with early-stage lung cancer (22). Our research indicated that individuals with stage I NSCLC patients who received a lobectomy experienced a more favorable long-term prognosis compared to those who underwent sublobectomy procedures, such as segmentectomy and wedge resection. This finding aligns with the results of previous studies (8,9). Although sublobectomy could preserve more lung tissue and pulmonary function than lobectomy, in individuals with stage I NSCLC, complete eradication of the malignancy is more likely observed after lobar resection (23). However, our results may have been affected by selection bias; patients undergoing sublobectomy were often older and had poor pulmonary function, which may confounded the efficacy of the operation itself. Recent research has indicated that for patients with NSCLC with tumor size ≤2 cm and non-metastasis, there is no significant difference in the 5-year OS of those undergoing lobectomy and sublobectomy

Table 6 Adjusted hazard ratios of variables that affected OS in stage I non-small cell lung cancer patients with tumors ≤30 mm who underwent pulmonary lobectomy

Variables	Hazard ratio (95% CI)	P value
Age (years)		
<65	Reference	
65–74	1.69 (1.58–1.81)	<0.001
≥75	2.83 (2.63–3.03)	<0.001
Sex		
Female	Reference	
Male	1.36 (1.29–1.44)	<0.001
Primary site		
Pulmonary lobe	Reference	
Main bronchus	0.65 (0.32–1.31)	0.23
Histology		
Adenocarcinoma	Reference	
Squamous cell carcinoma	1.31 (1.22–1.39)	<0.001
Large cell carcinoma	1.36 (1.13–1.63)	<0.001
Not otherwise specified	0.82 (0.73–0.91)	<0.001
Grade		
I	Reference	
II	1.74 (1.60–1.88)	<0.001
III–IV	2.10 (1.92–2.30)	<0.001
Unknown	1.38 (1.20–1.58)	<0.001
Tumor stage		
IA1	Reference	
IA2	1.12 (0.99–1.27)	0.07
IA3	1.31 (1.16–1.48)	<0.001
IB	1.45 (1.27–1.65)	<0.001
Examined lymph nodes		
≤10	Reference	
>10	0.80 (0.76–0.85)	<0.001
Additional systemic therapy		
No	Reference	
Yes	1.12 (0.97–1.30)	0.12

Grade I: well-differentiated. Grade II: moderately-differentiated. Grade III: poorly-differentiated. Grade IV: undifferentiated. OS, overall survival; CI, confidence interval.

Table 7 Adjusted hazard ratios of variables that affected OS in stage I non-small cell lung cancer patients with tumors ≤30 mm who underwent pulmonary sublobectomy

Variables	Hazard ratio (95% CI)	P value
Age (years)		
<65	Reference	
65–74	1.63 (1.45–1.84)	<0.001
≥75	2.53 (2.25–2.85)	<0.001
Sex		
Female	Reference	
Male	1.34 (1.23–1.46)	<0.001
Primary site		
Pulmonary lobe	Reference	
Main bronchus	0.65 (0.09–4.64)	0.67
Histology		
Adenocarcinoma	Reference	
Squamous cell carcinoma	1.28 (1.15–1.41)	<0.001
Large cell carcinoma	1.38 (1.04–1.82)	0.02
Not otherwise specified	0.90 (0.77–1.05)	0.17
Grade		
I	Reference	
II	2.05 (1.80–2.35)	<0.001
III–IV	2.62 (2.27–3.03)	<0.001
Unknown	1.50 (1.23–1.83)	<0.001
Tumor stage		
IA1	Reference	
IA2	1.22 (1.07–1.40)	<0.01
IA3	1.56 (1.34–1.82)	<0.001
IB	1.65 (1.41–1.93)	<0.001
Examined lymph nodes		
≤20	Reference	
>20	0.46 (0.30–0.71)	<0.001
Additional systemic therapy		
No	Reference	
Yes	1.40 (1.10–1.78)	<0.01

Grade I: well-differentiated. Grade II: moderately-differentiated. Grade III: poorly-differentiated. Grade IV: undifferentiated. OS, overall survival; CI, confidence interval.

(5,24-26). The latest clinical study JCOG1211 suggested that the 5-year progression-free survival (PFS)/OS rates were up to 98% in individuals with non-metastatic NSCLC with tumor size ≤30 mm and a consolidation/tumor ratio ≤0.5 after segmentectomy (27). In summary, lobar resection remains the preferred treatment even for early cases of NSCLC. The specific indication for sublobectomy remains unclear and requires further exploration in large-scale clinical trials.

The survival of individuals with stage I/II NSCLC receiving irradiation is better than those receiving the best supportive care (28). Recently, radiotherapy has become a radical treatment for early-stage NSCLC (4). In this study, the clinical outcomes of patients undergoing radiotherapy were associated with a relatively poor prognosis. However, the existence of confounders may have influenced this outcome. First, most patients opting for definitive radiotherapy were not eligible for surgery owing to comorbid conditions (29-31). Second, the cancer staging of patients with NSCLC treated with radiation was often determined by imaging. However, more than 10% of clinical stage I NSCLC patients experienced pathological upstaging after surgery and had a poorer outcome, as previously reported (32-34). Moreover, radiation techniques could affect the positioning accuracy and curative effect of radiotherapy. Several single-arm trials have reported that stereotactic body radiation therapy (SBRT) could offer excellent local cancer control in the early phase of NSCLC (29,31,35-40). A prior meta-analysis of observational studies, enrolling early-stage NSCLC patients treated with SBRT or surgical resection, found that surgery was linked to improved long-term OS; however, the study did not find statistically significant differences in CSS at 1- and 3-year intervals, nor in disease-free survival, local control, and distant control rates between the two treatment groups (41). Recently, a retrospective single-arm study conducted in Japan enrolled 136 individuals with operable stage I NSCLC, and all patients were treated with carbon-ion radiotherapy, and the 5-year OS, CSS, PFS, and local control rates were 81.8%, 91.2%, 65.9%, and 95.8% (95% CI: 92.3–99.5%), respectively (42). To date, there is no high-quality prospective trial comparing the curative effect of early-stage NSCLC in patients treated by surgical operation and radiation; therefore, whether radiotherapy could supersede surgery remains an open question.

The accurate assessment of LN status is important for the diagnosis and staging of lung cancer, especially in patients initially diagnosed through imaging. Prior studies suggested that more extensive dissection of LNs contributes to a sequential reduction in mortality risk for NSCLC patients with pathologically LN-negative (43,44). Consistent with the findings of previous studies, our results suggested the prognosis of stage I NSCLC 30 mm or less to be positively correlated with the number of dissected regional LNs. The possible explanation was that sufficient LN harvest could increase the detection rate of positive LNs, thereby decreasing the risk of under-staging and excising occult micrometastasis (45,46). However, to date, there is no consensus on the exact number of ELNs to consider a full evaluation. In a review involving 442 patients with stage I NSCLC, those with >6 ELNs had lower recurrence and overall mortality rates than those with ≤6 ELNs (47). A population-based study of stage I NSCLC found that 8, 9, 10, and 11 ELNs were optimal thresholds for patients with T1a, T1b, T1c, and T2a tumors to minimize all-cause mortality (48). From a long-term perspective, a minimum of 10 ELNs would significantly enhance the 5-year OS of patients with T1–3N0M0 NSCLC (49). Wu *et al.* reported that more than 15 ELNs may be required to improve 10-year outcomes in patients with stage I NSCLC (50). Recent evidence suggested that the optimal number of ELNs may depend on the extent of lung resection (51). According to the clinical trial ACOSOG Z0030, a minimum of 10 ELNs during pulmonary lobectomy were needed to get reliable pathological confirmation of LN status in early-stage NSCLC (T1–2, N0 or nonhilar N1) (52). Wen *et al.* indicated that patients with T2N0 NSCLC following pulmonary lobectomy or more extensive resection should have at least 12 ELNs to reduce cancer-specific mortality more significantly (53). As in other investigations, we proposed that in NSCLC patients with tumors <30 mm and pathological node-negative disease, more than 10 ELNs resulted in a 20% reduction in adjusted all-cause mortality risk and a 5% increase in 5-year OS within the lobectomy cohort compared with conservative LN examination. In this research, the examination of LNs during sublobectomy was generally more limited than lobectomy. Nevertheless, in the sublobectomy cohort, the most significant OS benefit was observed when the number of ELNs was more than 20. Our result was similar to that of a previous study, showing that the local recurrence rate of early-stage NSCLC (clinical T1N0) after sublobectomy was significantly higher than that after lobectomy (54). This could be attributed to the fact that the resection range of pulmonary lobectomy

includes both the pulmonary lobe of the primary lesion and the corresponding intrapulmonary lymphatic drainage area, whereas that of sublobectomy only includes the primary lesion (55). The SEER database does not contain information on the ELN stations. Recent research proposed that the recurrence-free survival of patients with early-stage NSCLC undergoing a station-based LN examination involving at least three N2 and one N1 station was longer than those undergoing a count-based examination with at least 10 ELNs, whereas the OS rates were found to be comparable between the two approaches (52). However, whether more ELNs while meeting the required number of ELN stations is associated with more survival benefits should be explored further. Overall, complete pathological examination of LN status is of significant prognostic importance for early-stage NSCLC and should be implemented to reduce the incidence of false-negative diagnosis, enabling the provision of proper adjuvant therapy, especially in patients receiving sublobectomy.

Nomogram is a kind of practical visualization model, which has been widely used for predicting the incidence rate of clinical events and estimating the prognosis of patients. Zhang *et al.* constructed a postoperative prognostic nomogram based on the SEER database for predicting OS in individuals with stage I NSCLC and validated it in a Chinese cohort (8). They reported good prognostic accuracy and clinical applicability in either the training or validation cohort. Current guidelines have recommended radiotherapy as another definitive treatment for stage I NSCLC. Jacobs developed a prediction model to estimate the risk of mortality from any cause among early-stage NSCLC patients with T1–T2N0 tumors who received radical radiotherapy (56). The developed nomogram revealed that age, sex, comorbidity index, tumor size, and histology contribute to the OS of patients. Nevertheless, the prognostic scoring system for estimating the prognosis of individuals with stage I NSCLC undergoing radical therapy (pulmonary lobectomy, sub-pulmonary lobectomy, and definitive radiotherapy) has not been established. Different from previous researches, we used the GBM to optimize the variable subsets. This decision-tree-based machine learning algorithm has demonstrated the superiority of capturing intricate connections between clinicopathological features and clinic outcomes (57). Moreover, GBM is capable of extracting valuable factors from the high-dimensional dataset, which is an advantage over traditional multiple-stepwise regression (58). Our prognostic scoring system showed well discriminative accuracy and had higher AUC and C-index values for OS prediction than TNM staging. Calibration plots and DCA also confirmed the dependability and superior clinical value of the proposed prediction model. Based on our nomogram, a new risk classification system was proposed to classify patients with stage I NSCLC as being at high or low risk for all-cause death. It is reasonable to believe that binary classification could be simpler and more practical for doctors to distinguish high-risk patients than the current staging system. Patients considered at high risk may require closer medical follow-up than low-risk patients. Moreover, external validation demonstrated that the model may be applied not only to the American population but also to the Asian population. In summary, this novel predictive model will assist clinicians in forecasting the prognosis of individuals with stage I NSCLC after radical therapy, thereby contributing to precise medicine and long-term management of early-stage lung cancer.

There are some limitations in this study. First, all patients included in our study were diagnosed with primary stage I NSCLC. Therefore, this predictive model may not be suitable for secondary primary stage I NSCLC. Second, this prognostic scoring system could not predict mortality within 90 days in stage I NSCLC patients, as early deaths are often influenced by some factors distinct from long-term survival, such as cumulative illness and treatment-related complications *et al.* (59,60). Besides, the forecasting model for predicting 30- and 90-day death after lung cancer surgery has been previously proposed, potentially becoming the standard for early mortality assessment (61). Third, some details are not available within the SEER database, like smoking history, landscape of genetic mutation, surgical margin status, ELN stations, and adjuvant systemic therapy regimen; these factors may potentially influence the prognosis of individuals with stage I NSCLC. Incorporating these factors in the prognostic scoring system would further improve its predictive performance. Furthermore, the assessment of LN status as no metastasis was based on imaging and operative pathology in both the training and validation cohorts. While enhanced computed tomography together with positron emission tomography scans are known for their high negative predictive value, there remains a small subset of patients with an insufficient number of ELNs where undiagnosed LN micrometastasis might still be present. This might affect the results of the analysis to a certain extent. Lastly, this study was an analysis of retrospective data, and the study design might be influenced by unavoidable selection bias. Thus, the

proposed prognosis scoring system and the determined optimal number of ELNs require further validation in prospective clinical trials.

Conclusions

Our study has proposed a new prognostic scoring system to predict 5-year OS of individuals with stage I NSCLC following definitive therapy, so as to instruct individual treatment. For stage I NSCLC with tumor size ≤30 mm, we recommend more than 10 and 20 regional ELNs for patients receiving pulmonary lobectomy and sublobectomy, respectively.

Acknowledgments

We appreciate all the staff members of the National Cancer Institute who were devoted to creation and maintenance of the SEER program. And we also thank The First Affiliated Hospital of Zhejiang University and all enrolled patients for the support of this research.

Footnote

Reporting Checklist: The authors have completed the TRIPOD reporting checklist. Available at https://tcr.amegroups.com/article/view/10.21037/tcr-24-1474/rc

Data Sharing Statement: Available at https://tcr.amegroups.com/article/view/10.21037/tcr-24-1474/dss

Peer Review File: Available at https://tcr.amegroups.com/article/view/10.21037/tcr-24-1474/prf

Funding: This study was supported by grants from the National Natural Science Foundation of China (grant No. 62307031), the Key Research and Development Projects of Zhejiang Provincial Science and Technology Department (No. 2021C03122), and the Zhejiang Provincial Natural Science Foundation (grant No. LQ23H160024).

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://tcr.amegroups.com/article/view/10.21037/tcr-24-1474/coif). 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 study was approved by the Ethics Committee of The First Affiliated Hospital of Zhejiang University (No. IIT20240197B) and individual consent for this retrospective analysis was waived.

References

1. Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin 2021;71:209-49.
2. Churchill ED, Sweet RH, Soutter L, et al. The surgical management of carcinoma of the lung; a study of the cases treated at the Massachusetts General Hospital from 1930 to 1950. J Thorac Surg 1950;20:349-65.
3. Melloul E, Egger B, Krueger T, et al. Mortality, complications and loss of pulmonary function after pneumonectomy vs. sleeve lobectomy in patients younger and older than 70 years. Interact Cardiovasc Thorac Surg 2008;7:986-9.
4. NCCN clinical practice guidelines in oncology, non-small cell lung cancer, version 3.2024. March 12, 2024. Available online: https://www.nccn.org/
5. Altorki N, Wang X, Kozono D, et al. Lobar or Sublobar Resection for Peripheral Stage IA Non-Small-Cell Lung Cancer. N Engl J Med 2023;388:489-98.
6. Scotti V, Bruni A, Francolini G, et al. Stereotactic Ablative Radiotherapy as an Alternative to Lobectomy in Patients With Medically Operable Stage I NSCLC: A Retrospective, Multicenter Analysis. Clin Lung Cancer 2019;20:e53-61.
7. Jiang H, Wu T, Qie P, et al. Comparative analysis of the clinical effects of different thoracoscopic resection in the treatment of Stage I Non-Small Cell Lung Cancer. Pak J Med Sci 2024;40:1644-50.

8. Zhang H, Zeng J, Li X, et al. The nomogram for the prediction of overall survival after surgery in patients in early-stage NSCLC based on SEER database and external validation cohort. Cancer Med 2024;13:e6751.
9. Wang ZH, Deng L. Establishment and Validation of a Predictive Nomogram for Postoperative Survival of Stage I Non-Small Cell Lung Cancer. Int J Gen Med 2022;15:7287-98.
10. 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.
11. Ou SH, Zell JA, Ziogas A, et al. Prognostic significance of the non-size-based AJCC T2 descriptors: visceral pleura invasion, hilar atelectasis, or obstructive pneumonitis in stage IB non-small cell lung cancer is dependent on tumor size. Chest 2008;133:662-9.
12. Tan KS, Hsu M, Adusumilli PS. Pathologic node-negative lung cancer: Adequacy of lymph node yield and a tool to assess the risk of occult nodal disease. Lung Cancer 2022;174:60-6.
13. Surveillance, Epidemiology, and End Results (SEER) program. SEER*Stat Database: Incidence - seer research data, 17 registries, nov 2023 sub (2000-2021) - linked to county attributes - time dependent (1990-2022) income/rurality, 1969-2022 counties, national cancer institute, dccps, surveillance research program. Released April 2024, based on the November 2023 submission. Available online: https://seer.cancer.gov
14. Ganti AK, Klein AB, Cotarla I, et al. Update of Incidence, Prevalence, Survival, and Initial Treatment in Patients With Non-Small Cell Lung Cancer in the US. JAMA Oncol 2021;7:1824-32.
15. Riley RD, Ensor J, Snell KIE, et al. Calculating the sample size required for developing a clinical prediction model. BMJ 2020;368:m441.
16. Zhou H, Zhang Y, Qiu Z, et al. Nomogram to Predict Cause-Specific Mortality in Patients With Surgically Resected Stage I Non-Small-Cell Lung Cancer: A Competing Risk Analysis. Clin Lung Cancer 2018;19:e195-203.
17. Li C, Liu J, Lin J, et al. Poor survival of non-small-cell lung cancer patients with main bronchus tumor: a large population-based study. Future Oncol 2019;15:2819-27.
18. Travis WD, Brambilla E, Noguchi M, et al. International association for the study of lung cancer/american thoracic society/european respiratory society international multidisciplinary classification of lung adenocarcinoma. J Thorac Oncol 2011;6:244-85.
19. Wu CF, Fu JY, Yeh CJ, et al. Recurrence Risk Factors Analysis for Stage I Non-small Cell Lung Cancer. Medicine (Baltimore) 2015;94:e1337.
20. Kreuzer M, Kreienbrock L, Müller KM, et al. Histologic types of lung carcinoma and age at onset. Cancer 1999;85:1958-65.
21. Nicholson AG, Tsao MS, Beasley MB, et al. The 2021 WHO Classification of Lung Tumors: Impact of Advances Since 2015. J Thorac Oncol 2022;17:362-87.
22. Cahan WG. Radical lobectomy. J Thorac Cardiovasc Surg 1960;39:555-72.
23. Wang C, Wu Y, Shao J, et al. Clinicopathological variables influencing overall survival, recurrence and post-recurrence survival in resected stage I non-small-cell lung cancer. BMC Cancer 2020;20:150.
24. Koike T, Yamato Y, Yoshiya K, et al. Intentional limited pulmonary resection for peripheral T1 N0 M0 small-sized lung cancer. J Thorac Cardiovasc Surg 2003;125:924-8.
25. Okada M, Koike T, Higashiyama M, et al. Radical sublobar resection for small-sized non-small cell lung cancer: a multicenter study. J Thorac Cardiovasc Surg 2006;132:769-75.
26. Fan J, Wang L, Jiang GN, et al. Sublobectomy versus lobectomy for stage I non-small-cell lung cancer, a meta-analysis of published studies. Ann Surg Oncol 2012;19:661-8.
27. Aokage K, Suzuki K, Saji H, et al. Segmentectomy for ground-glass-dominant lung cancer with a tumour diameter of 3 cm or less including ground-glass opacity (JCOG1211): a multicentre, single-arm, confirmatory, phase 3 trial. Lancet Respir Med 2023;11:540-9.
28. Rowell NP, Williams CJ. Radical radiotherapy for stage I/II non-small cell lung cancer in patients not sufficiently fit for or declining surgery (medically inoperable). Cochrane Database Syst Rev 2001;(1):CD002935.
29. Baumann P, Nyman J, Hoyer M, et al. Outcome in a prospective phase II trial of medically inoperable stage I non-small-cell lung cancer patients treated with stereotactic body radiotherapy. J Clin Oncol 2009;27:3290-6.
30. Onishi H, Shirato H, Nagata Y, et al. Hypofractionated stereotactic radiotherapy (HypoFXSRT) for stage I non-small cell lung cancer: updated results of 257 patients in a Japanese multi-institutional study. J Thorac Oncol 2007;2:S94-100.
31. Timmerman R, Paulus R, Galvin J, et al. Stereotactic body

radiation therapy for inoperable early stage lung cancer. JAMA 2010;303:1070-6.
32. Boffa DJ, Kosinski AS, Paul S, et al. Lymph node evaluation by open or video-assisted approaches in 11,500 anatomic lung cancer resections. Ann Thorac Surg 2012;94:347-53; discussion 353.
33. Licht PB, Jørgensen OD, Ladegaard L, et al. A national study of nodal upstaging after thoracoscopic versus open lobectomy for clinical stage I lung cancer. Ann Thorac Surg 2013;96:943-9; discussion 949-50.
34. Stiles BM, Servais EL, Lee PC, et al. Point: Clinical stage IA non-small cell lung cancer determined by computed tomography and positron emission tomography is frequently not pathologic IA non-small cell lung cancer: the problem of understaging. J Thorac Cardiovasc Surg 2009;137:13-9.
35. Kocak Uzel E, Bagci Kilic M, Morcali H, et al. Stereotactic body radiation therapy for stage I medically operable non-small cell lung cancer. Sci Rep 2023;13:10384.
36. Sun B, Brooks ED, Komaki RU, et al. 7-year follow-up after stereotactic ablative radiotherapy for patients with stage I non-small cell lung cancer: Results of a phase 2 clinical trial. Cancer 2017;123:3031-9.
37. Bral S, Gevaert T, Linthout N, et al. Prospective, risk-adapted strategy of stereotactic body radiotherapy for early-stage non-small-cell lung cancer: results of a Phase II trial. Int J Radiat Oncol Biol Phys 2011;80:1343-9.
38. Ricardi U, Filippi AR, Guarneri A, et al. Stereotactic body radiation therapy for early stage non-small cell lung cancer: results of a prospective trial. Lung Cancer 2010;68:72-7.
39. Fakiris AJ, McGarry RC, Yiannoutsos CT, et al. Stereotactic body radiation therapy for early-stage non-small-cell lung carcinoma: four-year results of a prospective phase II study. Int J Radiat Oncol Biol Phys 2009;75:677-82.
40. Hoyer M, Roed H, Hansen AT, et al. Prospective study on stereotactic radiotherapy of limited-stage non–small-cell lung cancer. International Journal of Radiation Oncology*Biology*Physics 2006;66:S128-35.
41. Zhang B, Zhu F, Ma X, et al. Matched-pair comparisons of stereotactic body radiotherapy (SBRT) versus surgery for the treatment of early stage non-small cell lung cancer: a systematic review and meta-analysis. Radiother Oncol 2014;112:250-5.
42. Kubo N, Suefuji H, Nakajima M, et al. Five-Year Survival Outcomes After Carbon-Ion Radiotherapy for Operable Stage I NSCLC: A Japanese National Registry Study (J-CROS-LUNG). J Thorac Oncol 2024;19:491-9.
43. Becker DJ, Levy BP, Gold HT, et al. Influence of Extent of Lymph Node Evaluation on Survival for Pathologically Lymph Node Negative Non-Small Cell Lung Cancer. Am J Clin Oncol 2018;41:820-5.
44. Osarogiagbon RU, Ogbata O, Yu X. Number of lymph nodes associated with maximal reduction of long-term mortality risk in pathologic node-negative non-small cell lung cancer. Ann Thorac Surg 2014;97:385-93.
45. Bollen EC, van Duin CJ, Theunissen PH, et al. Mediastinal lymph node dissection in resected lung cancer: morbidity and accuracy of staging. Ann Thorac Surg 1993;55:961-6.
46. Gallina FT, Marinelli D, Tajè R, et al. Analysis of predictive factors of unforeseen nodal metastases in resected clinical stage I NSCLC. Front Oncol 2023;13:1229939.
47. Gajra A, Newman N, Gamble GP, et al. Effect of number of lymph nodes sampled on outcome in patients with stage I non-small-cell lung cancer. J Clin Oncol 2003;21:1029-34.
48. Dai J, Liu M, Yang Y, et al. Optimal Lymph Node Examination and Adjuvant Chemotherapy for Stage I Lung Cancer. J Thorac Oncol 2019;14:1277-85.
49. Zhu Z, Song Z, Jiao W, et al. A large real-world cohort study of examined lymph node standards for adequate nodal staging in early non-small cell lung cancer. Transl Lung Cancer Res 2021;10:815-25.
50. Wu YC, Lin CF, Hsu WH, et al. Long-term results of pathological stage I non-small cell lung cancer: validation of using the number of totally removed lymph nodes as a staging control. Eur J Cardiothorac Surg 2003;24:994-1001.
51. Dezube AR, Mazzola E, Bravo-Iñiguez CE, et al. Analysis of Lymph Node Sampling Minimums in Early Stage Non-Small-Cell Lung Cancer. Semin Thorac Cardiovasc Surg 2021;33:834-45.
52. Darling GE, Allen MS, Decker PA, et al. Randomized trial of mediastinal lymph node sampling versus complete lymphadenectomy during pulmonary resection in the patient with N0 or N1 (less than hilar) non-small cell carcinoma: results of the American College of Surgery Oncology Group Z0030 Trial. J Thorac Cardiovasc Surg 2011;141:662-70.
53. Wen YS, Xi KX, Xi KX, et al. The number of resected lymph nodes is associated with the long-term survival outcome in patients with T2 N0 non-small cell lung cancer. Cancer Manag Res 2018;10:6869-77.

54. Ginsberg RJ, Rubinstein LV. Randomized trial of lobectomy versus limited resection for T1 N0 non-small cell lung cancer. Lung Cancer Study Group. Ann Thorac Surg 1995;60:615-22; discussion 622-3.
55. Rami-Porta R, Tsuboi M. Sublobar resection for lung cancer. Eur Respir J 2009;33:426-35.
56. Jacobs CD, Mehta K, Gao J, et al. Nomogram Predicting Overall Survival Benefit of Stereotactic Ablative Radiotherapy for Early-Stage Non-Small Cell Lung Cancer. Clin Lung Cancer 2022;23:177-84.
57. Natekin A, Knoll A. Gradient boosting machines, a tutorial. Front Neurorobot 2013;7:21.
58. Cygu S, Seow H, Dushoff J, et al. Comparing machine learning approaches to incorporate time-varying covariates in predicting cancer survival time. Sci Rep 2023;13:1370.
59. van Rossum PSN, Wolfhagen N, van Bockel LW, et al. Real-World Acute Toxicity and 90-Day Mortality in Patients With Stage I NSCLC Treated With Stereotactic Body Radiotherapy. J Thorac Oncol 2024;19:1550-63.
60. Quero-Valenzuela F, Piedra-Fernández I, Hernández-Escobar F, et al. Half the deaths after surgery for lung cancer occur after discharge. Surg Oncol 2018;27:630-4.
61. Powell HA, Tata LJ, Baldwin DR, et al. Early mortality after surgical resection for lung cancer: an analysis of the English National Lung cancer audit. Thorax 2013;68:826-34.

Cite this article as: Wang S, Yin X, Wu L, Yu H, Lu Z, Zhao F, Yan D, Yan S. Establishing a prognostic scoring system and exploring prognostic value of examined lymph node numbers for stage I non-small cell lung cancer: a retrospective study of Surveillance, Epidemiology, and End Results (SEER) database and a Chinese cohort. Transl Cancer Res 2025;14(1):404-423. doi: 10.21037/tcr-24-1474