World Journal of Surgical Oncology

RESEARCH

Open Access

Prognostic value of lymph node metrics in lung squamous cell carcinoma: an analysis of the SEER database

Lei Liu[1†], Qiao Zhang[2†], Shuai Jin[1*†] and Lang Xie[3*†]

Abstract

Introduction Although the Tumor-Node-Metastasis (TNM) staging system is widely used for staging lung squamous cell carcinoma (LSCC), the TNM system primarily emphasizes tumor size and metastasis, without adequately considering lymph node involvement. Consequently, incorporating lymph node metastasis as an additional prognostic factor is essential for predicting outcomes in LSCC patients.

Methods This retrospective study included patients diagnosed with LSCC between 2004 and 2018 and was based on data from the Surveillance, Epidemiology, and End Results (SEER) database of the National Cancer Institute. The primary endpoint of the study was cancer-specific survival (CSS), and demographic characteristics, tumor characteristics, and treatment regimens were incorporated into the predictive model. The study focused on the value of indicators related to pathological lymph node testing, including the lymph node ratio (LNR), regional node positivity (RNP), and lymph node examination count (RNE), in the prediction of cancer-specific survival in LSCC. A prognostic model was established using a multivariate Cox regression model, and the model was evaluated using the C index, Kaplan–Meier, the Akaike information criterion (AIC), decision curve analysis (DCA), continuous net reclassification improvement (NRI), and integrated discrimination improvement (IDI), and the predictive efficacy of different models was compared.

Results A total of 14,200 LSCC patients (2004–2018) were divided into training and validation cohorts. The 10-year CSS rate was approximately 50%, with no significant survival differences between cohorts ($p = 0.8$). The prognostic analysis revealed that models incorporating LNR, RNP, and RNE demonstrated superior performance over the TNM model. The LNR and RNP models demonstrated better model fit, discrimination, and reclassification, with AUC values of 0.695 (training) and 0.665 (validation). The RNP and LNR models showed similar predictive performance, significantly outperforming the TNM and RNE models. Calibration curves and decision curve analysis confirmed the

†Lei Liu, Qiao Zhang, Shuai Jin and Lang Xie has made the same contribution to the manuscript

*Correspondence:
Shuai Jin
jssss@gmc.edu.cn
Lang Xie
xlsmor@163.com

Full list of author information is available at the end of the article

clinical utility and net benefit of the LNR and RNP models in predicting long-term CSS for LSCC patients, highlighting their value in clinical decision-making.

Conclusion This study confirms that RNP status is an independent prognostic factor for CSS in LSCC, with predictive efficacy comparable to LNR, with both models enhancing survival prediction beyond TNM staging.

Keywords Lung squamous cell carcinoma, Regional nodes positive, Lymph node ratio, SEER database, Integrated discrimination improvement, Intelligent retirement

Introduction

Lung cancer is one of the leading causes of cancer-related deaths globally, accounting for 18.0% of all cancer deaths and approximately 180,000 deaths each year. Moreover, it is the leading and second leading cause of death for men and women with cancer, respectively [1]. Lung squamous cell carcinoma (LSCC) is a specific subtype of non-small cell lung cancer (NSCLC), accounting for approximately 35% of all lung cancer cases [2]. Compared to adenocarcinoma, LSCC is characterized by distinct epidemiological features and poorer clinical outcomes, with limited effective targeted therapies [3, 4]. These factors underscore the importance of accurate clinical staging for predicting patient outcomes and guiding appropriate treatment strategies.

Regional node positive (RNP) status is a key factor in pathological evaluation and surgical assessment. Research has demonstrated that RNP status is a significant independent prognostic indicator in various cancers, including thyroid cancer, adrenocortical cancer, and chondrosarcoma [5–7]. In LSCC, prognosis is largely determined by the eighth edition of the American Joint Committee on Cancer Tumor-Node-Metastasis (TNM) staging system, which primarily focuses on tumor metastasis [8]. However, this system does not account for lymphadenectomy or the number of positive lymph nodes, whereas pathological pN staging provides superior prognostic value [9, 10]. Previous studies have shown that factors such as age, lymph node ratio (LNR), and the number of regional nodes examined (RNE) significantly affect survival in NSCLC patients [11, 12]. Furthermore, combining LNR with traditional N staging has been shown to enhance survival prediction compared to N staging alone, and similarly, combining N staging with RNE improves survival prediction [13–15]. Therefore, investigating the prognostic value of RNP status in LSCC patients is essential.

This study used data from the Surveillance, Epidemiology, and End Results (SEER) database to evaluate the role of regional node positivity (RNP) in predicting long-term postoperative survival in patients with LSCC. In addition, we compared multiple models including TNM stage, RNP, and lymph node ratio (LNR) to determine the most effective model for predicting cancer-specific survival (CSS) in patients with LSCC.

Methods

Study design and data source

We used SEER * Stat (version 8.4.0) to download data of patients with LSSC from January 2004 to December 2018 (containing 18 registration states and additional treatment information) [16]. All sample data used in the study were anonymous data in the SEER database; therefore, there was no requirement for patient informed consent and ethical approval by the institution of the investigator. The authors assume responsibility for the content of the study and the results presented in this article, which do not represent the official views of the SEER database or the National Cancer Institute. We report the following findings based on the TRIPOD report checklist [17].

Sample selection

In the SEER database, we selected patients with primary malignant SCC according to the third edition of the International Classification of Disease Oncology (ICD-O-3) and the fifth edition of the WHO Classification of Tumors-Thoracic Tumors criteria. The inclusion criteria were: (1) ICD-O-3 histology was 8070/3, 8071/3, 8072/3, and 8083/3; (2) primary site was C34.0-Main Bronchus, C34.1-Upper lobe, lung, C34.2-Middle lobe, lung, and C34.3-Lower lobe, lung; and (3) pulmonary SCC was the only primary tumor. Exclusion criteria were: (1) tumor size was imprecise; (2) TNM stage was missing; (3) laterality was unclear; (4) age < 18 or > 100 years; (5) race was unknown; (6) surgery was a partial resection; and (7) the lymph node status was unclear or the extent of resection was unclear. The selection process for the study population is shown in Fig. 1.

Study variables

Data extracted from the SEER database included the following: demographic variables (including age, sex, race, marital status at diagnosis), tumor characteristics (e.g., lateral degree, TNM stage, primary site, histological type, grading, tumor size), treatment status (e.g., primary site surgery, radiotherapy, chemotherapy, lymph node surgery[RNE, RNP]), patient survival months, and CSS. The LNR is obtained by dividing the patient's regional node positivity number by the number of RNE. The LNR was classified into three categories (≤0.15, 0.16–0.5, and ≥0.5), as previously described [10]. Age at

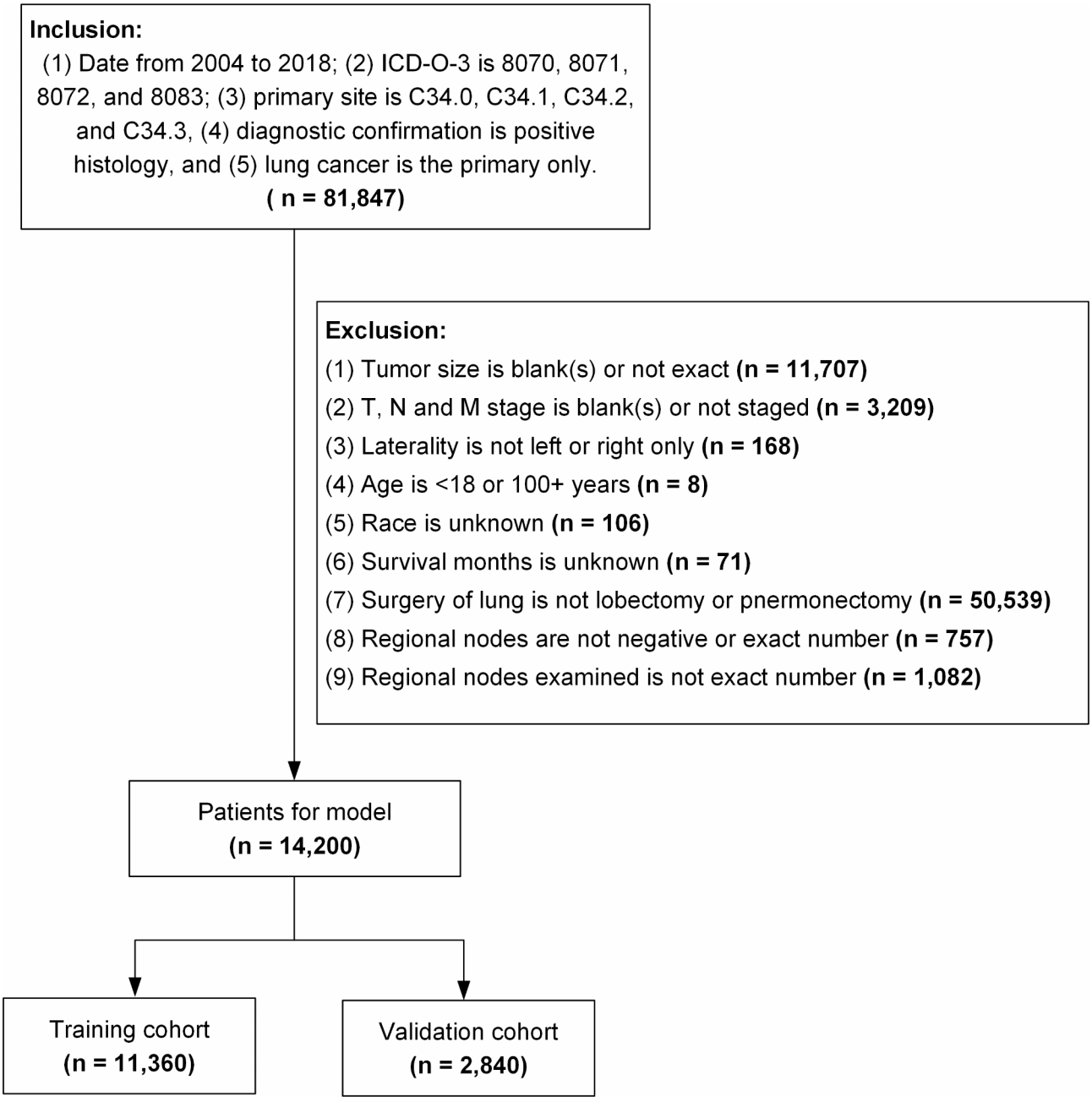

Fig. 1 Study flow diagram. ICD-O-3, third edition of the international classification of disease oncology

diagnosis, tumor size, and the X-tile software (https://medicine.yale.edu/lab/) were used to determine the best cut-off value; the best cut-off values for age were 69 and 76 years, and patients were divided into three age groups (i.e., <69, 69–76, and >76 years). The optimal cut-off values for tumor size were 20 and 38 mm, and patients were divided into three groups (i.e., <20, 20–38, and >38 mm). In this study, negative and positive classifications were made according to whether the number of RNP was 0 or not, and RNE kept the original data in the discriminative model. The endpoint of interest in this study is CSS for LSCC. CSS refers to the time to death of patients due to LSCC itself. The 10-year CSS rate was selected as the terminal event to compare the predictive performance of the models.

Statistical analysis

This study presents continuous variables as mean ± standard deviation, while categorical variables are expressed as counts and percentages. Differences between the training and validation groups were analyzed using Student's t-test for continuous variables and Pearson's χ^2

test for categorical variables. Kaplan–Meier analysis was employed to estimate CSS associated with different variables, and the log-rank test was used to compare the differences between groups to identify potential confounders. Variables with a p-value < 0.2 in the univariate analysis were included in the subsequent multivariate Cox regression analysis. We established five multivariate Cox models to clarify and compare the prognostic effects of lymph node indicators on LSCC CSS. First, a basic model containing only the TNM staging was created and named the TNM model. Second, a limited model was constructed, which included the TNM staging and other clinical parameters except for lymph node-related indicators. Third, the LNR model was developed, which added the LNR indicator to the limited model. Then, the RNP model was established, which combined the RNP indicator with other parameters in the limited model. Finally, the RNE model was constructed, which combined the RNE indicator with all the variables in the limited model. The LNR model, RNP model and RNE model are collectively referred to as the full models. Variable selection was performed using stepwise regression for all models.

Model performance was assessed using measures of model fit (R^2, Akaike Information Criterion [18]), discrimination (Harrell's concordance index [19]), reclassification (continuous net reclassification improvement [NRI] and integrated discrimination improvement [IDI] [20]), calibration plots, and decision curve analysis (DCA) [21].

All *p*-values were two-tailed, and *p*-values < 0.05 denoted statistically significant differences. All statistical analyses were performed using the R programming language and environment (http://www.r-project.org/).

Results

Patient characteristics

Between 2004 and 2018, a total of 14,200 patients were diagnosed with LSCC; these patients were randomly divided into training and validation cohorts in a 7:3 ratio. The results of the training and validation group stratification based on demographic data, tumor manifestations, and treatment-related characteristics are shown in Table 1. In all study samples, 50% of patients were aged < 69 years at diagnosis; >60% of the patients were male. Disease was more frequently observed in the right lung, approximately 70% of patients had T1–2 stage disease, and >70% of patients did not have lymph node metastasis. Only approximately 3% of the patients developed distant metastases. Patients with SCC often have tumors larger than 2 cm in size, and the most common location of lesions was the upper lung. The majority of patients received chemotherapy; more than 90% of patients underwent a lobectomy; and very few patients received radiotherapy. Most patients had an LNR ≤ 0.15. More than 70% of the patients had negative lymph nodes, and the average number of lymph nodes examined in patients was approximately 12. The results of the χ^2 test did not show a difference in performance between the training and validation cohorts, this indicates good comparability between the two groups.

Prognostic analysis

Kaplan–Meier curves demonstrated that the 10-year CSS rate was approximately 50%, with no statistically significant difference in survival between the training and validation groups ($p = 0.8$) (Supplemental Fig. 1). Detailed results such as the TNM model, limited model and full model parameters can be found in Supplementary Table 1. Table 2 present the predicted performance details and

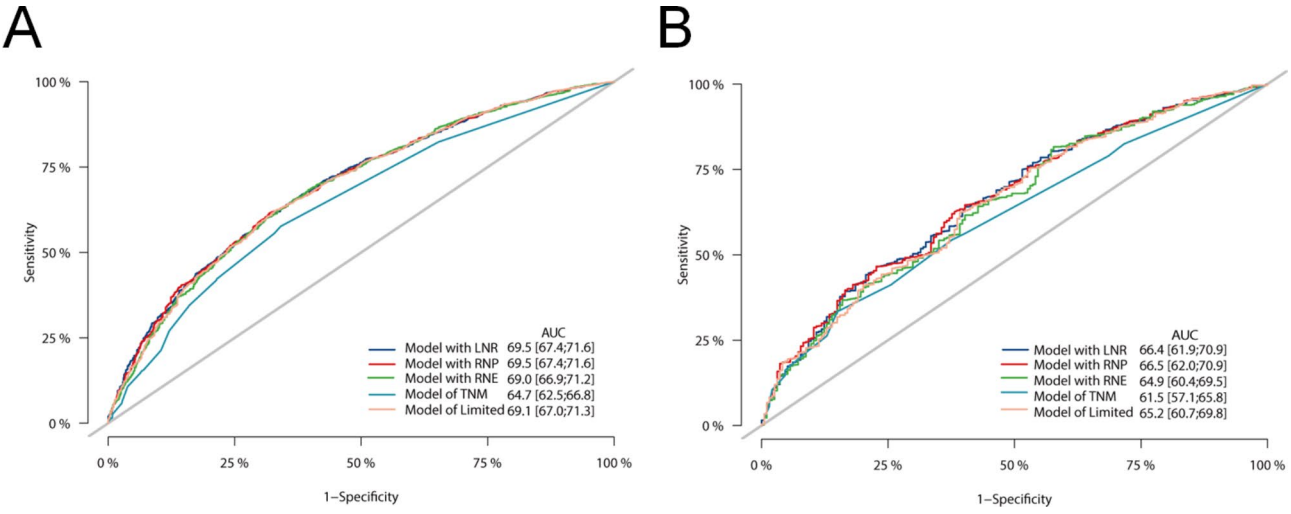

Fig. 2 ROC of Cox models in the (**A**) training and (**B**) validation cohorts predicting 10-year CSS. AUC, area under the receiver operating characteristic curve; CSS, cancer-specific survival; LNR, lymph node ratio; RNE, regional nodes examined; RNP, regional nodes positive

Table 1 Description of the training and validation cohorts

Characteristic	Overall *n* = 14,200	Training Cohort *n* = 11,360	Validation Cohort *n* = 2,840	*p*-value
Age, years				0.703
< 69	7,141 (50%)	5,724 (50%)	1,417 (50%)	
69–76	4,614 (32%)	3,695 (33%)	919 (32%)	
> 76	2,445 (17%)	1,941 (17%)	504 (18%)	
Sex				0.897
Female	5,330 (38%)	4,261 (38%)	1,069 (38%)	
Male	8,870 (62%)	7,099 (62%)	1,771 (62%)	
Race				0.180
Black	1,145 (8.1%)	892 (7.9%)	253 (8.9%)	
Other	696 (4.9%)	557 (4.9%)	139 (4.9%)	
White	12,359 (87%)	9,911 (87%)	2,448 (86%)	
Laterality				0.716
Left	6,272 (44%)	5,009 (44%)	1,263 (44%)	
Right	7,928 (56%)	6,351 (56%)	1,577 (56%)	
T stage				0.427
T1	4,638 (33%)	3,735 (33%)	903 (32%)	
T2	5,165 (36%)	4,108 (36%)	1,057 (37%)	
T3	2,884 (20%)	2,321 (20%)	563 (20%)	
T4	1,513 (11%)	1,196 (11%)	317 (11%)	
N stage				0.844
N0	10,138 (71%)	8,106 (71%)	2,032 (72%)	
N1	2,476 (17%)	1,982 (17%)	494 (17%)	
N2	1,533 (11%)	1,232 (11%)	301 (11%)	
N3	53 (0.4%)	40 (0.4%)	13 (0.5%)	
M stage				0.751
M0	13,833 (97%)	11,064 (97%)	2,769 (98%)	
M1	367 (2.6%)	296 (2.6%)	71 (2.5%)	
Marital status				0.672
Married	8,055 (57%)	6,454 (57%)	1,601 (56%)	
Other	6,145 (43%)	4,906 (43%)	1,239 (44%)	
Histologic				0.608
Basaloid SCC	154 (1.1%)	117 (1.0%)	37 (1.3%)	
Kera SCC	1,213 (8.5%)	967 (8.5%)	246 (8.7%)	
Nonkera SCC	667 (4.7%)	530 (4.7%)	137 (4.8%)	
SCC	12,166 (86%)	9,746 (86%)	2,420 (85%)	
Grade				0.501
I–II	6,308 (44%)	5,062 (45%)	1,246 (44%)	
III–IV	6,346 (45%)	5,078 (45%)	1,268 (45%)	
Unknown	1,546 (11%)	1,220 (11%)	326 (11%)	
Tumor size, mm				0.178
< 20	2,372 (17%)	1,928 (17%)	444 (16%)	
20–38	5,668 (40%)	4,504 (40%)	1,164 (41%)	
> 38	6,160 (43%)	4,928 (43%)	1,232 (43%)	
Primary site				0.949
Lower	4,936 (35%)	3,958 (35%)	978 (34%)	
Middle	594 (4.2%)	476 (4.2%)	118 (4.2%)	
Other	465 (3.3%)	368 (3.2%)	97 (3.4%)	
Upper	8,205 (58%)	6,558 (58%)	1,647 (58%)	
Chemotherapy				0.792
No/Unknown	9,856 (69%)	7,879 (69%)	1,977 (70%)	
Yes	4,344 (31%)	3,481 (31%)	863 (30%)	
Surgery				0.975

Table 1 (continued)

Characteristic	Overall *n* = 14,200	Training Cohort *n* = 11,360	Validation Cohort *n* = 2,840	*p*-value
Lobectomy	13,132 (92%)	10,506 (92%)	2,626 (92%)	
Pneumonectomy	1,068 (7.5%)	854 (7.5%)	214 (7.5%)	
Radiotherapy				0.437
No/unknown	12,236 (86%)	9,776 (86%)	2,460 (87%)	
Yes	1,964 (14%)	1,584 (14%)	380 (13%)	
LNR				0.798
≤ 0.15	12,217 (86%)	9,764 (86%)	2,453 (86%)	
0.16–0.5	1,659 (12%)	1,333 (12%)	326 (11%)	
> 0.5	324 (2.3%)	263 (2.3%)	61 (2.1%)	
RNP				0.744
Negative	10,531 (74%)	8,418 (74%)	2,113 (74%)	
Positive	3,669 (26%)	2,942 (26%)	727 (26%)	
RNE	11.72 ± 8.97	11.72 ± 8.96	11.72 ± 9.03	0.995

LNR, lymph node ratio; RNP, regional nodes positive; RNE, regional nodes examined; SCC, squamous cell carcinoma

Table 2 Prediction of performance in the training cohort for the TNM, Limited, and full models

			Full Models		
Variable	**TNM Model**	**Limited Model**	**LNR Model**	**RNP Model**	**RNE Model**
AIC	64,611	64,233	64,135	64,135	64,178
R^2, %	7.5	10.9	11.6	11.4	11.3
Model performance compared with the TNM stage					
NRI		0.201 (0.005, 0.183)			
IDI		0.045 (0.001, 0.041)			
Model performance compared with the limited model					
NRI			0.001 (−0.046, 0.078)	0.158 (0.034, 0.185)	0.005 (−0.028, 0.043)
IDI			0.001 (0.000, 0.002)	0.005 (0.001, 0.008)	0.000 (−0.002, 0.002)
Model performance compared with the LNR model					
NRI				0.037 (−0.060, 0.109)	−0.040 (−0.062, 0.000)
IDI				−0.001 (−0.005, 0.003)	−0.006 (−0.010, −0.001)

AIC, Akaike Information Criterion; IDI, integrated discrimination improvement; LNR, lymph node ratio; NRI, net reclassification improvement; RNP, regional nodes positive; RNE, regional nodes examined; TNM, Tumor–Node–Metastasis

all supporting data for the TNM model, limited model, and full models. Compared with the TNM model, the limited model and full models had a better model fit, with higher R^2 and lower Akaike Information Criterion, better discrimination, and better reclassification (NRI and IDI, 95% CI: >0). Compared with the limited model, the LNR, RNP, and RNE models yielded better results in each aspect. Unlike the NRI (95% CI: −0.046–0.078), the IDI of the LNR model was improved effectively (IDI, 95% CI: >0). The IDI and NRI of the RNP model were both improved (95% CI: >0). The IDI and NRI of the RNE model were consistent with those of the limited model. Compared with the LNR model, the IDI and NRI of the RNP model did not show any advantage, while the performance of the RNE model was worse than that of the LNR model (IDI, 95% CI: <0). The LNR and RNP models showed similar performance, with IDI (-0.001, 95% CI: -0.005-0.0003) and NRI (0.037, 95% CI: -0.060- 0.109). Among the training datasets, the LNR and RNP models performed best in predicting 10-year CSS in LSCC, with an AUC of 0.695 (Fig. 2A). Furthermore, analysis of an observational calibration curve revealed similar outcome prediction and observed risk performance for the LNR and RNP models (Fig. 3A). The DCA analysis also showed that the LNR and RNP models had similar performance in terms of clinical net benefits (Fig. 3C). Based on these results, the performance and effectiveness of the relevant models was further evaluated in the validation cohort.

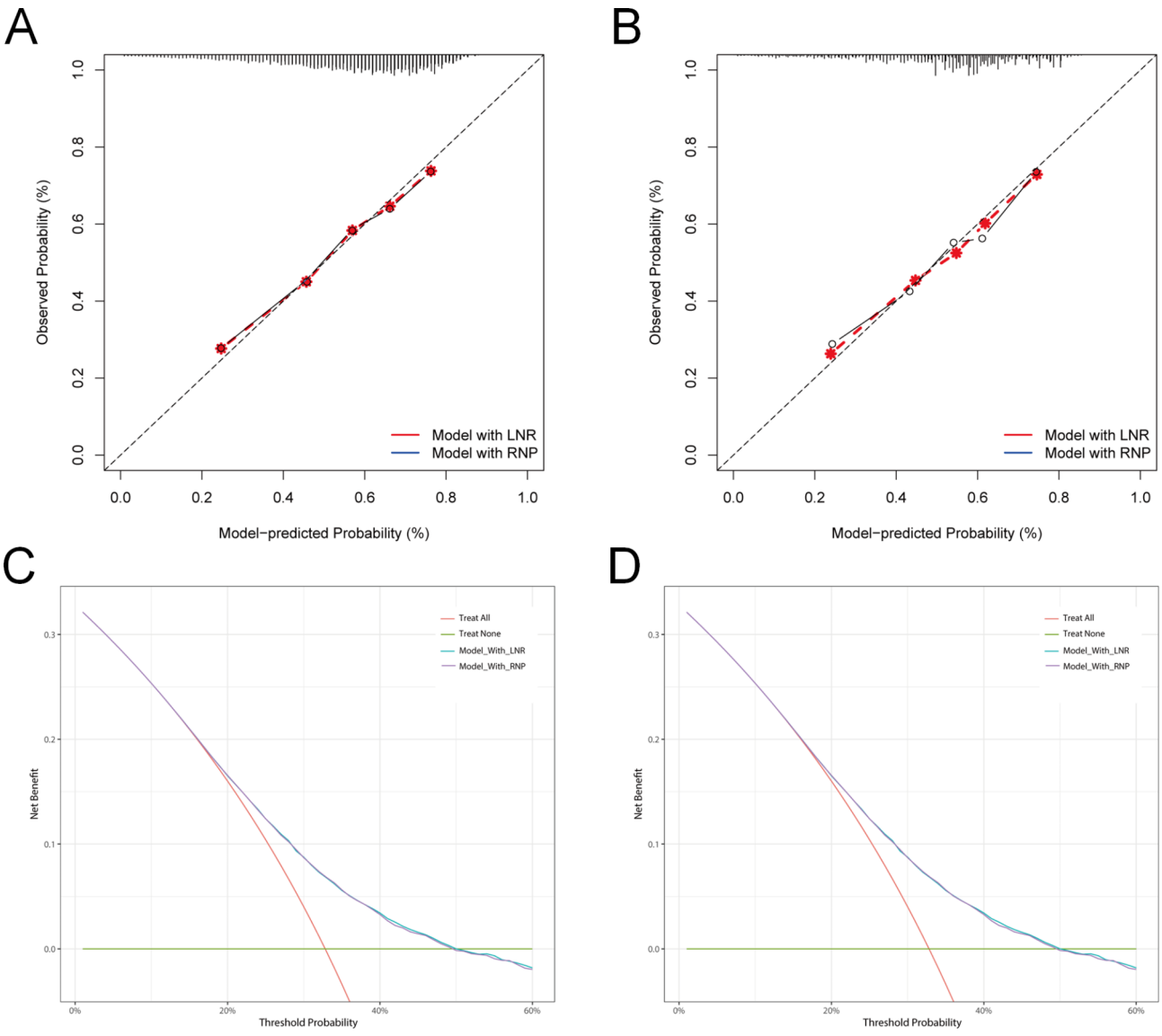

Fig. 3 Calibration curves depicting predicted versus observed 10-year risks of CSS using the LNR and RNP models in the (**A**) training and (**B**) validation cohorts. DCA of CSS using the LNR and RNP models in the (**C**) training and (**D**) validation cohorts. CSS, cancer-specific survival; DCA = decision curve analysis; LNR, lymph node ratio; RNP, regional nodes positive

Validation analysis

Consistent with the predictive efficacy of the training dataset, the AIC and R^2 of the full models were higher than those of the TNM model and limited model in the validation dataset. In validation, the NRI and IDI of the LNR and RNP models were −0.01 (95% CI: -0.115-0.072) and −0.002 (95% CI: -0.015-0.009), respectively. In the validation, the RNP model performed best in predicting 10-year CSS of LSCC, with an AUC of 0.665, while the LNR model was 0.664 (Fig. 2B). Overall, the performance of the RNE model was lower than that of the LNR and RNP models (IDI, 95% CI: <0) (Table 3). The calibration curve (Fig. 3B) and DCA of the validation cohort showed that the LNR and RNP models exhibited similar predictive performance; both models had good predictive ability and clinical net benefits (Fig. 3D). The prediction performance of the LNR model and RNP model is better than that of the TNM model, limited model, and RNE model, both in the training cohort and the validation cohort.

Discussion

An accurate assessment of lymph node status is essential in the management of LSCC patients, particularly for treatment decisions and survival prognosis. Through the modeling and validation conducted in this study, we have confirmed that RNP status is an independent prognostic factor for CSS in LSCC patients, with predictive efficacy

Table 3 Prediction of performance in the validation cohort for the TNM, limited, and full models

			Full Models		
Variable	**TNM Model**	**Limited Model**	**LNR Model**	**RNP Model**	**RNE Model**
AIC	64,611	64,222	64,135	64,164	64,178
R^2, %	7.3	10.3	11.4	10.8	10.4
Model performance compared with the TNM model					
NRI		0.160 (0.091, 0.256)			
IDI		0.042 (0.031, 0.071)			
Model performance compared with the limited model					
NRI			0.061 (−0.048, 0.145)	0.115 (−0.003, 0.197)	0.002 (−0.096, 0.070)
IDI			0.010 (0.001, 0.026)	0.008 (0.001, 0.018)	−0.001 (−0.003, 0.002)
Model performance compared with the LNR model					
NRI				−0.010 (−0.115, 0.072)	−0.042 (−0.143, 0.012)
IDI				−0.002 (−0.015, 0.009)	−0.011 (−0.024, −0.002)

AIC, Akaike Information Criterion; IDI, integrated discrimination improvement; LNR, lymph node ratio; NRI, net reclassification improvement; RNP, regional nodes positive; RNE, regional nodes examined; TNM, Tumor–Node–Metastasis

comparable to that of the LNR variable. This finding not only provides a more precise tool for prognostic assessment in LSCC but also offers a solid theoretical foundation for developing individualized treatment strategies based on lymph node status in clinical practice. Specifically, when compared with traditional TNM staging and the RNE model, the RNP and LNR models offer a more accurate reflection of the patient's true prognosis. This aids clinicians in making more informed treatment decisions, ultimately enhancing both patient survival rates and quality of life.

Lymph node status is one of the most important prognostic indicators of NSCLC. Currently, the American Joint Committee on Cancer N staging based on the anatomic location of positive lymph nodes is the most widely used staging system [22]. Compared with the lymph node count, the anatomical definition of lymph node location is more complex and may lead to inconsistencies in staging interpretation and misclassification [23, 24]. However, it has been suggested that use of the anatomic location in this N staging system to distinguish subgroups may overlap in prognostic significance under certain conditions [25, 26]. In this study, model comparisons showed that pathological lymph node status was an independent prognostic factor in patients with LSCC, though the CSS prediction model included N staging. This may be explained by the fact that N staging is defined by pathology or clinical presentation, whereas lymphatic status is defined by pathological examination. In addition, the accuracy of N staging may be affected by an insufficient number of positive lymph nodes, which may result in misclassification of lymph nodes [9]. The staging of LSCC will be more accurate after the RNP and LNR indicators. Through precise staging, clinicians can better identify high-risk patients and develop more individualized treatment plans for these patients.

The LNR index, which combines the number of involved lymph nodes with the total number of detected lymph nodes, could theoretically overcome the limitations of the number-based lymph node classification system. However, an obvious characteristic of the LNR is that it is related to the number of RNE. Therefore, the role of the LNR is questionable when lymph node dissection is low, particularly after thoracoscopic surgery [27]. This study revealed that the effectiveness of RNP status in predicting CSS rate in patients with LSCC did not differ from that of the LNR. These findings suggested that the CSS rate of patients with LSCC was not strongly correlated with the number of RNE. Following the occurrence of lymph node metastasis, the rates of CSS tended to be similar among patients. Thus, RNP status may be a valid variable to replace LNR in predicting LSCC-CSS rates.

Previous studies have confirmed that the NSCLC survival prognosis model containing the LNR indicator has superior predictive ability compared to the TNM staging system [11, 28]. The present study investigated the predictive performance of three lymph node staging variables (i.e., LNR, RNE, and RNP) for determining the long-term CSS rate in patients with LSCC. In the multivariate Cox regression analysis, the LNR and RNP models showed similar predictive ability; however, the ability of the RNE model to predict CSS rates remains to be investigated. The LNR and RNP models showed better prognostic value compared with the TNM model. Models that include RNP and LNR are of great significance in clinical treatment decisions. They can guide doctors

in choosing the appropriate treatment strategy, such as whether more aggressive treatment or close follow-up is required, through more accurate prognostic assessments. The application of RNP and LNR can help improve patient survival rates and quality of life, especially in individualized treatment for lymph node metastasis.

The SEER database, covering 18 states, ensures the generalizability and credibility of the current results in predicting CSS rates in LSCC. Nevertheless, this study had several limitations. Firstly, the SEER database is a large retrospective project, which inherently carries the risk of biases. These biases may stem from the exclusion of patients with incomplete data, leading to a potential underrepresentation of certain subgroups. The reliance on retrospective data also limits the ability to control for unmeasured confounding factors, which could have affected the accuracy of the results. Secondly, the SEER database does not include some key variables that could significantly influence the prognosis of LSCC, such as smoking status, recurrence information, comorbidities, and detailed treatment regimens. These unaccounted factors may introduce residual confounding and limit the ability to fully assess the impact of certain variables on survival outcomes. Finally, the exclusion of patients with missing data for key variables, while necessary for ensuring the completeness of the analysis, may have contributed to selection bias. This exclusion could potentially skew the sample towards patients with more complete medical histories, thus affecting the generalizability of our findings. Additionally, given that the SEER database does not capture certain clinical parameters that are routinely collected in prospective studies, the results of this study should be interpreted with caution.

Conclusion

This study demonstrates that RNP status is an independent prognostic factor for CSS in LSCC, comparable to LNR. Incorporating RNP and LNR into prognostic models improves survival prediction over TNM staging, aiding personalized treatment decisions. Further prospective studies are needed to validate these findings and refine LSCC prognosis.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12957-024-03639-4.

Supplementary Material 1

Acknowledgements
The authors would like to thank the SEER team for their assistance in the collection of clinical data. The SEER database is public, and we obtained the data usage protocol for the SEER database.

Author contributions
LX and SJ made an overall plan for the manuscript design and proposal. QZ and LL contributed to data analysis and validation. LL and QZ manuscript writing, SJ and LX revised the manuscript. LL supported manuscript language editing. All authors have reviewed the final version that was submitted.

Funding
No.

Data availability
The data used in this study can be downloaded from the SEER database or obtained from https://github.com/ShinyShine-820/LSCC.

Declarations

Consent for publication
All authors have reviewed the final version of the manuscript and agree with its submission.、.

Competing interests
The authors declare no competing interests.

Conflict of interest
The authors have no conflict of interest to declare.

Ethics approval
The SEER database is freely available for download by researchers, with private patient information removed. Therefore, no additional ethics statement from the author's institution is required.

Author details
[1]School of Biology & Engineering (School of Health Medicine Modern Industry), Guizhou Medical University, Guiyang, 561113, China
[2]Medical Department, The Second People's Hospital of Guiyang(Jinyang Hospital), Guiyang, 550081, China
[3]Department of Hospital Infection Management and Preventive Health Care, Zhejiang Provincial People's Hospital Bijie Hospital, Bijie 551799, China

Received: 8 October 2024 / Accepted: 23 December 2024
Published online: 28 December 2024

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F, 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. https://doi.org/10.3322/caac.21660.
2. Relli V, Trerotola M, Guerra E, Alberti S. Abandoning the notion of Non-small Cell Lung Cancer. Trends Mol Med. 2019;25:585–94. https://doi.org/10.1016/j.molmed.2019.04.012.
3. Langer CJ, Obasaju C, Bunn P, Bonomi P, Gandara D, Hirsch FR, Kim ES, Natale RB, Novello S, Paz-Ares L, Pérol M, Reck M, Ramalingam SS, Reynolds CH, Socinski MA, Spigel DR, Wakelee H, Mayo C, Thatcher N. Incremental Innovation and Progress in Advanced squamous cell Lung Cancer: current status and future impact of treatment. J Thorac Oncol 11 (2016) 2066–81. https://doi.org/10.1016/j.jtho.2016.08.138
4. Hirsch FR, Scagliotti GV, Mulshine JL, Kwon R, Curran WJ, Wu Y-L, Paz-Ares L. Lung cancer: current therapies and new targeted treatments. Lancet. 2017;389:299–311. https://doi.org/10.1016/S0140-6736(16)30958-8.
5. Wan L, Tu C, Li S, Li Z. Regional Lymph Node involvement is Associated with poorer survivorship in patients with Chondrosarcoma: a SEER analysis. Clin Orthop Relat Res. 2019;477:2508–18. https://doi.org/10.1097/CORR.0000000000000846.
6. Min Y, Xiang K, Feng Y, Chen H, Chen J, Wei X, Yin G. Development and validation of a population-based model for predicting the regional lymph node metastasis in adolescent differentiated thyroid carcinoma. Oral Oncol. 2021;121:105507. https://doi.org/10.1016/j.oraloncology.2021.105507.
7. Sada A, Glasgow AE, Lyden ML, Thompson GB, Dy BM, McKenzie TJ, Habermann EB. Positive lymph nodes in Adrenocortical Carcinoma: what does it

Mean? World J Surg. 2021;45:188–94. https://doi.org/10.1007/s00268-020-05801-x.
8. Liu X, Sun W, Wu J, Feng Y, Mao L, Chen M, Yang X, Wang H, Chi K, Yang Y, Lin D. Major pathologic response assessment and clinical significance of metastatic lymph nodes after neoadjuvant therapy for non-small cell lung cancer. Mod Pathol. 2021;34:1990–8. https://doi.org/10.1038/s41379-021-00871-1.
9. Liang W, He J, Shen Y, Shen J, He Q, Zhang J, Jiang G, Wang Q, Liu L, Gao S, Liu D, Wang Z, Zhu Z, Ng CSH, Liu C, Horsleben Petersen R, Rocco G, D'Amico T, Brunelli A, Chen H, Zhi X, Liu B, Yang Y, Chen W, Zhou Q, He J. Impact of examined Lymph Node count on precise staging and long-term survival of Resected non–small-cell lung Cancer: a Population Study of the US SEER database and a Chinese multi-institutional Registry. JCO. 2017;35:1162–70. https://doi.org/10.1200/JCO.2016.67.5140.
10. Jonnalagadda S, Arcinega J, Smith C, Wisnivesky JP. Validation of the lymph node ratio as a prognostic factor in patients with N1 nonsmall cell lung cancer: LNR as a prognostic factor in N1 NSCLC. Cancer. 2011;117:4724–31. https://doi.org/10.1002/cncr.26093.
11. Deng W, Xu T, Wang Y, Xu Y, Yang P, Gomez D, Liao Z. Log odds of positive lymph nodes may predict survival benefit in patients with node-positive non-small cell lung cancer. Lung Cancer. 2018;122:60–6. https://doi.org/10.1016/j.lungcan.2018.05.016.
12. Li Q, Zhan P, Yuan D, Lv T, Krupnick AS, Passaro A, Brunelli A, Smeltzer MP, Osarogiagbon RU, Song Y. Prognostic value of lymph node ratio in patients with pathological N1 non-small cell lung cancer: a systematic review with meta-analysis. Transl Lung Cancer Res. 2016;5:258–64. https://doi.org/10.21037/tlcr.2016.06.13.
13. Wen Y-S, Xi K-X, Xi K-X, Zhang R-S, Wang G-M, Huang Z-R, Zhang L-J. The number of resected lymph nodes is associated with the long-term survival outcome in patients with T2 N0 non-small cell lung cancer. CMAR Volume. 2018;10:6869–77. https://doi.org/10.2147/CMAR.S186047.
14. Ding X, Hui Z, Dai H, Fan C, Men Y, Ji W, Liang J, Lv J, Zhou Z, Feng Q, Xiao Z, Chen D, Zhang H, Yin W, Lu N, He J, Wang L. A proposal for combination of lymph node ratio and anatomic location of involved lymph nodes for nodal classification in non–small cell Lung Cancer. J Thorac Oncol. 2016;11:1565–73. https://doi.org/10.1016/j.jtho.2016.05.004.
15. Wisnivesky JP, Arciniega J, Mhango G, Mandeli J, Halm EA. Lymph node ratio as a prognostic factor in elderly patients with pathological N1 non-small cell lung cancer. Thorax. 2011;66:287–93. https://doi.org/10.1136/thx.2010.148601.
16. Surveillance E, Results E. (SEER) Program (www.seer.cancer.gov) SEER*Stat Database: Incidence - SEER Research Plus Data, 18 Registries, Nov 2020 Sub (2000–2018) - Linked To County Attributes - Total U.S., 1969–2019 Counties, National Cancer Institute, DCCPS, Surveillance Research Program, released April 2021, based on the November 2020 submission., (n.d.).
17. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. BMJ. 2015;350:g7594. https://doi.org/10.1136/bmj.g7594.
18. Van Houwelingen JC, Le Cessie S. Predictive value of statistical models. Stat Med. 1990;9:1303–25. https://doi.org/10.1002/sim.4780091109.
19. Harrell FE, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med. 1996;15:361–87. https://doi.org/10.1002/(SICI)1097-0258(19960229)15:4%3C361::AID-SIM168%3E3.0.CO;2-4.
20. Chambless LE, Cummiskey CP, Cui G. Several methods to assess improvement in risk prediction models: extension to survival analysis. Stat Med. 2011;30:22–38. https://doi.org/10.1002/sim.4026.
21. Vickers AJ, Holland F. Decision curve analysis to evaluate the clinical benefit of prediction models. Spine J. 2021;21:1643–8. https://doi.org/10.1016/j.spinee.2021.02.024.
22. Asamura H, Chansky K, Crowley J, Goldstraw P, Rusch VW, Vansteenkiste JF, Watanabe H, Wu Y-L, Zielinski M, Ball D, Rami-Porta R, International Association for the Study of Lung Cancer Staging and Prognostic Factors Committee. Advisory Board Members, and participating institutions, the International Association for the study of Lung Cancer Lung Cancer Staging Project: proposals for the revision of the N descriptors in the Forthcoming 8th Edition of the TNM classification for Lung Cancer. J Thorac Oncol. 2015;10:1675–84. https://doi.org/10.1097/JTO.0000000000000678.
23. Ramirez RA, Wang CG, Miller LE, Adair CA, Berry A, Yu X, O'Brien TF, Osarogiagbon RU. Incomplete intrapulmonary lymph node retrieval after routine pathologic examination of resected lung cancer. J Clin Oncol. 2012;30:2823–8. https://doi.org/10.1200/JCO.2011.39.2589.
24. Park BJ, Kim TH, Shin S, Kim HK, Choi YS, Kim J, Zo JI, Shim YM, Cho JH. Recommended change in the N descriptor proposed by the International Association for the study of Lung Cancer: a validation study. J Thorac Oncol. 2019;14:1962–9. https://doi.org/10.1016/j.jtho.2019.07.034.
25. Yun JK, Lee GD, Choi S, Kim HR, Kim Y-H, Kim DK, Park S-I. Comparison between lymph node station- and zone-based classification for the future revision of node descriptors proposed by the International Association for the study of Lung Cancer in surgically resected patients with non-small-cell lung cancer. Eur J Cardiothorac Surg. 2019;56:849–57. https://doi.org/10.1093/ejcts/ezz147.
26. Yang M, Zhang H, Ma Z, Gong L, Chen C, Ren P, Shang X, Tang P, Jiang H, Yu Z. Log odds of positive lymph nodes is a novel prognostic indicator for advanced ESCC after surgical resection. J Thorac Dis. 2017;9:1182–9. https://doi.org/10.21037/jtd.2017.03.187.
27. Merritt RE, Hoang CD, Shrager JB. Lymph node evaluation achieved by open lobectomy compared with thoracoscopic lobectomy for N0 lung cancer. Ann Thorac Surg. 2013;96:1171–7. https://doi.org/10.1016/j.athoracsur.2013.05.044.
28. He C, Ni M, Liu J, Teng X, Ke L, Matsuura Y, Okuda K, Sakairi Y, Cheng J, Yu L, Lv W, Hu J. A survival nomogram model for patients with resectable non-small cell lung cancer and lymph node metastasis (N1 or N2) based on the Surveillance, Epidemiology, and end results database and single-center data. Transl Lung Cancer Res. 2024;13:573–86. https://doi.org/10.21037/tlcr-24-119.

Publisher's note