



## Research Paper

# A nomogram for cancer-specific survival of lung adenocarcinoma patients: A SEER based analysis

Hong Guo, Doctor of Medicine (M.D.)<sup>a,b,1</sup>, Guole Nie, Doctor of Medicine (M.D.)<sup>c,1</sup>,  
Xin Zhao, Master of Medicine (M.M.)<sup>b,1</sup>, Jialu Liu, Master of Medicine (M.M.)<sup>a</sup>,  
Kaihua Yu, Master of Medicine (M.M.)<sup>a</sup>, Yulan Li, Doctor of Medicine (M.D.)<sup>a,d,\*</sup>

<sup>a</sup> First Clinical Medical College, Lanzhou University, Lanzhou 730000, China

<sup>b</sup> Department of Anesthesiology, Inner Mongolia Hospital of Peking University Cancer Hospital, The Affiliated People's Hospital of Inner Mongolia Medical University, Hohhot 10020, China

<sup>c</sup> Department of Colorectal Hernia Surgery, Binzhou Medical University Hospital, Binzhou, 256600, China

<sup>d</sup> Department of Anesthesiology, First Hospital of Lanzhou University, Lanzhou University, Lanzhou 730000, China

## ARTICLE INFO

## Keywords:

Lung adenocarcinoma (LUAD)  
Cancer-specific survival (CSS)  
Nomogram  
American Joint Committee on Cancer Staging (AJCC)  
SEER

## ABSTRACT

**Background:** Non-small cell lung cancer (NSCLC) accounts for 85 % of lung cancer cases. Among NSCLC subtypes, lung adenocarcinoma (LUAD) stands as the most prevalent. Regrettably, LUAD continues to exhibit a notably unfavorable overall prognosis. This study's primary aim was to develop and validate prognostic tools capable of predicting the likelihood of cancer-specific survival (CSS) in patients with LUAD.

**Methods:** We retrospectively collected 21,099 patients diagnosed with LUAD between 2010 and 2015, and 8290 patients diagnosed between 2004 and 2009 from SEER database. The cohort of 21,099 patients served as the prognostic group for the exploration of LUAD-related prognostic risk factors. The cohort of 8290 patients was designated for external validation. We created a training set and an internal validation set in the prognostic group for the development and internal validation of CSS nomograms. CSS predictors were identified through the least absolute shrinkage and selection operator (Lasso) regression analysis. Prognostic model was constructed via Cox hazard regression analysis, presented in the form of both static and dynamic network-based nomograms.

**Results:** Several independent prognostic factors were incorporated into the construction of nomogram. The nomogram accurately predicted CSS at 1, 3, and 5 years, with respective AUC values of 0.769, 0.761, and 0.748 for the training group, and 0.741, 0.752, and 0.740 for the testing group. The study demonstrated a strong agreement between anticipated and actual CSS values, supported by decision curve analysis (DCA) and time-dependent calibrated curves. High-risk patients based on the nomogram exhibiting significantly lower survival rates compared to their low-risk counterparts according to Kaplan-Meier (K-M) curves. The nomogram demonstrates excellent predictive power in the external validation cohort.

**Conclusions:** A dependable and user-friendly nomogram has been developed, available in both static and online dynamic calculator formats, to facilitate healthcare professionals in accurately estimating the likelihood of CSS for patients diagnosed LUAD.

## Introduction

Lung cancer is a widespread malignancy that exerts a significant toll on public health, particularly in 13 global regions, with a high mortality rate [1]. Broadly classified, lung cancer falls into two primary subtypes: small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC).

Notably, NSCLC constitutes the majority of lung cancer cases, with lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (SCC) being the predominant histological subtypes, accounting for 60 % and 15 % of NSCLC cases, respectively [2]. While prognostic research has been predominantly focused on NSCLC [3,4], there exists a dearth of prognostic models specifically tailored to LUAD.

\* Corresponding author at: First Clinical Medical College, Lanzhou University, Lanzhou 730000, China.

E-mail address: [liyul@lzu.edu.cn](mailto:liyul@lzu.edu.cn) (Y. Li).

<sup>1</sup> Hong Guo and Guole Nie and Xin Zhao contributed equally to this work.

Individual lung cancer prognoses often diverge due to variances in genotype and molecular genetic characteristics, which can significantly impact both prognosis and treatment strategies [5,6]. Presently, the prognosis for NSCLC patients primarily relies on the American Joint Committee on Cancer (AJCC) 8th tumor, node, metastasis (TNM) staging system. However, this system, particularly its N classification that considers lymphatic region involvement, comes with inherent limitations [7]. Consequently, it becomes imperative to establish a dedicated predictive model for LUAD prognosis and optimize therapeutic approaches. Nomograms have emerged as invaluable tools in clinical practice, offering the ability to calculate event probabilities by considering the weighted predictive value of each contributing factor [8]. They hold distinct advantages over traditional TNM staging systems, potentially serving as alternative clinical decision-making aids or even evolving into new standards [9]. Predictive models have gained widespread acceptance as reliable instruments for prognosis assessment and clinical decision-making across various malignancies, encompassing breast cancer [10], gastric cancer [11], and prostate cancer [12].

In our investigation, extracting data from the Surveillance, Epidemiology, and End Results (SEER) database, we delved into prognostic factors capable of foreseeing cancer-specific mortality among LUAD patients. Subsequently, a practical prognostic model was constructed and validated. This model not only enhances clinical management but also facilitates individualized patient counseling, ultimately contributing to the optimization of patient care.

## Methods

### Research approach and data retrieval

A retrospective data analysis was conducted utilizing the SEER\*Stat 8.4.2 program ([www.seer.cancer.gov](http://www.seer.cancer.gov)). Malignant adenocarcinoma cases were identified using the International Classification of Diseases for Oncology, Third Revision (ICD-O-3). Primary tumor locations were specified by CS34.0, CS34.1, CS34.2, CS34.3, and CS34.8 codes, while histological codes encompassed 8140/3, 8141/3, 8144/3, 8250/3, 8211/3, 8255/3, 8260/3, 8310/3, 8323/3, 8480/3, 8481/3, and 8490/3.

Inclusion criteria encompassed histologically confirmed malignant adenocarcinoma originating in the lung or bronchus, a diagnosis period spanning from 2010 to 2015, and available data on both survival duration and cause of death. Exclusion criteria encompassed undifferentiated carcinoma, neuroendocrine tumors, adenosquamous carcinoma, mesenchymal tumors, or adenosarcoma (mixed epithelial and mesenchymal tumor), autopsy or death certificate diagnoses, cases with an unknown reason for mortality, and those with unknown follow-up or survival duration.

Given that the SEER database is publicly accessible and open, ethical review board approval was not mandated. Access to and use of the datasets were facilitated through signed authorizations and licenses provided by the SEER program. Our investigation involved 21,099 patients diagnosed between 2010 and 2015, with a focus on identifying prognostic factors linked to cancer-specific survival (CSS). We divided these patients into training and internal validation groups with a 7:3 ratio. Information extracted from the SEER database included patient age, race, sex, year of diagnosis, histologic subtype, tumor grade, primary site, SEER summary stage (local, regional, or distant), American Joint Committee on Cancer (AJCC) stage, tumor, lymph node, and metastasis (TNM) stage, primary site surgery (Surg.Prim.Site), radiation therapy, chemotherapy, tumor size, number of regional lymph nodes (LNs) examined (Examined.lymph.nodes), number of positive LNs (Positive.lymph.nodes), number of multiple primary tumors (Multi.primary.tumors), survival time, cause of death, and survival status.

### Outcomes and predictors

This study primarily focused on LUAD-related mortality as the primary outcome. Survival time was computed as the interval between the date of diagnosis and either the date of death or the last follow-up visit. Participants with unclear causes of death, unknown survival durations, or zero survival periods were excluded from the analysis. Several factors were considered, including T stage, N stage, M stage, AJCC stage, primary tumor site, primary site surgery (yes or no), radiation therapy (yes or no), chemotherapy (yes or no), tumor size, number of regional lymph nodes examined, number of positive lymph nodes, presence of bone metastasis (Mets.Bone), liver metastasis (Mets.Liver), brain metastasis (Mets.Brain), and the presence of multiple primary tumors. As potential predictors, these parameters were subsequently evaluated. Ultimately, we identified and included 11 variables with non-zero coefficients and a tendency to deviate, applying a 1 standard error threshold in our analysis.

### Conversion of continuous variables

Restricted cubic spline analysis was employed to identify nonlinear relationships between continuous variables and the study outcomes, subsequently transforming them into categorical variables. Age has been classified into two groups based on the optimal cutoff value: <67 and ≥67 years old. Tumor size has been categorized into five groups following a previous study: ≤1, 1.1–3, 3.1–5, 5.1–7, and >7 cm. The number of examined lymph nodes (LNs) was stratified into three categories: <7 and ≥7, based on the optimal cutoff value, while the number of positive LNs was grouped as 0, 1–3, and >3, as per earlier studies. Tumor grades have been categorized as Grades I through IV. SEER summary stages are classified as localized, regional, and distant. For AJCC stages, they have been rebranded as Stages I to IV. The T, N, and M stages of the AJCC were represented as T1, T2, T3, T4; N0, N1, N2, N3; and M0, M1, respectively. Regarding the number of primary tumors, two categories were established: a single primary tumor group and a group with two or more primary tumors.

### Predictor selection

To identify potential risk factors for cancer-specific death (CSD), all variables were analyzed comprehensively for the entire cohort using least absolute shrinkage and selection operator (LASSO) regression. As the penalty term increased, the estimates of components with minimal impact gradually approached zero, effectively performing a filtering role in selecting the most relevant variables. To optimize the model for clinical applicability, a systematic assessment of predictive efficiency was conducted using subsets comprising 5–10 variables. It is noteworthy that the area under the receiver operating characteristic (ROC) curve (AUC) consistently surpassed the robust threshold of 0.740 when integrating 7–10 variables into the model, highlighting its remarkable predictive performance. Subsequently, a rigorous analysis of multicollinearity among the 11 variables was conducted, recognizing its potential to impact both model stability and accuracy. As a result, the final model was constructed employing seven key predictors: the number of positive lymph nodes (LNs), tumor size, age, gender, tumor grade, M stage, and primary surgical site.

### Prognostic nomogram construction and validation

The prognostic nomogram was developed using a training group for construction and a test group for internal validation through Cox proportional hazards regression analysis. This nomogram predicts the probabilities of 1, 3, and 5-year CSS (Cancer-Specific Survival) along with ROC and calibration curves for each time point. An online calculator based on the nomogram was created for generating CSS predictions along with 95 % confidence intervals (CIs). Calibration curves were

**Table 1**  
Baseline characteristics of LUAD patients.

Characteristics	Overall (n = 21,099)	Training group (n = 14,770)	Testing group (n = 6329)	p
Age, years (%)				
<67	9623 (45.6)	6788 (46.0)	2835 (44.8)	0.123
≥67	11,476 (54.4)	7982 (54.0)	3494 (55.2)	
Sex (%)				
Female	12,354 (58.6)	8652 (58.6)	3702 (58.5)	0.920
Male	8745 (41.4)	6118 (41.4)	2627 (41.5)	
Race (%)				
Black	1933 (9.2)	1333 (9.0)	600 (9.5)	0.568
Other	1886 (8.9)	1319 (8.9)	567 (9.0)	
White	17,280 (81.9)	12,118 (82.0)	5162 (81.6)	
Primary Site (%)				
Bronchus/Others	252 (1.2)	169 (1.1)	83 (1.3)	0.733
Lower lobe	6941 (32.9)	4847 (32.8)	2094 (33.1)	
Middle lobe	1192 (5.6)	834 (5.6)	358 (5.7)	
Upper lobe	12,714 (60.3)	8920 (60.4)	3794 (59.9)	
Grade (%)				
I	4852 (23.0)	3411 (23.1)	1441 (22.8)	0.527
II	10,073 (47.7)	7069 (47.9)	3004 (47.5)	
III	6038 (28.6)	4190 (28.4)	1848 (29.2)	0.791
IV	136 (0.6)	100 (0.7)	36 (0.6)	
T (%)				
T1	10,221 (48.4)	7144 (48.4)	3077 (48.6)	0.791
T2	7608 (36.1)	5326 (36.1)	2282 (36.1)	
T3	2528 (12.0)	1788 (12.1)	740 (11.7)	
T4	742 (3.5)	512 (3.5)	230 (3.6)	
N (%)				
N0	16,097 (76.3)	11,290 (76.4)	4807 (76.0)	0.432
N1	2508 (11.9)	1721 (11.7)	787 (12.4)	
N2	2433 (11.5)	1716 (11.6)	717 (11.3)	
N3	61 (0.3)	43 (0.3)	18 (0.3)	
M (%)				
M0	20,443 (96.9)	14,300 (96.8)	6143 (97.1)	0.374
M1	656 (3.1)	470 (3.2)	186 (2.9)	
Surg.Prim.Site (%)				
Sub-lobar	3066 (14.5)	2121 (14.4)	945 (14.9)	0.464
Pneumonectomy	538 (2.5)	371 (2.5)	167 (2.6)	
Lobectomy	17,495 (82.9)	12,278 (83.1)	5217 (82.4)	
Radiation (%)				
None/unknown	18,911 (89.6)	13,201 (89.4)	5710 (90.2)	0.070
Yes	2188 (10.4)	1569 (10.6)	619 (9.8)	
Chemotherapy (%)				
No/unknown	15,388 (72.9)	10,768 (72.9)	4620 (73.0)	0.903
Yes	5711 (27.1)	4002 (27.1)	1709 (27.0)	
Examined.lymph.nodes (%)				
<7	8281 (39.2)	5798 (39.3)	2483 (39.2)	0.987
≥7	12,818 (60.8)	8972 (60.7)	3846 (60.8)	
Positive.lymph.nodes (%)				
0	16,291 (77.2)	11,413 (77.3)	4878 (77.1)	0.893
≤3	3539 (16.8)	2476 (16.8)	1063 (16.8)	
>3	1269 (6.0)	881 (6.0)	388 (6.1)	
Mets.Bone (%)				
No	21,011 (99.6)	14,710 (99.6)	6301 (99.6)	0.797
Yes	88 (0.4)	60 (0.4)	28 (0.4)	
Mets.Brain (%)				
No	20,896 (99.0)	14,627 (99.0)	6269 (99.1)	0.952

**Table 1 (continued)**

Characteristics	Overall (n = 21,099)	Training group (n = 14,770)	Testing group (n = 6329)	p
Mets.Liver (%)				
Yes	203 (1.0)	143 (1.0)	60 (0.9)	0.383
No	21,074 (99.9)	14,750 (99.9)	6324 (99.9)	
Tumor size, cm (%)				
≤1	25 (0.1)	20 (0.1)	5 (0.1)	0.863
1–3	1486 (7.0)	1034 (7.0)	452 (7.1)	
3–5	12,955 (61.4)	9099 (61.6)	3856 (60.9)	
5–7	4482 (21.2)	3121 (21.1)	1361 (21.5)	
>7	1316 (6.2)	910 (6.2)	406 (6.4)	
Multi.primary.tumors (%)				
1	860 (4.1)	606 (4.1)	254 (4.0)	0.173
≥2	12,658 (60.0)	8906 (60.3)	3752 (59.3)	
	8441 (40.0)	5864 (39.7)	2577 (40.7)	

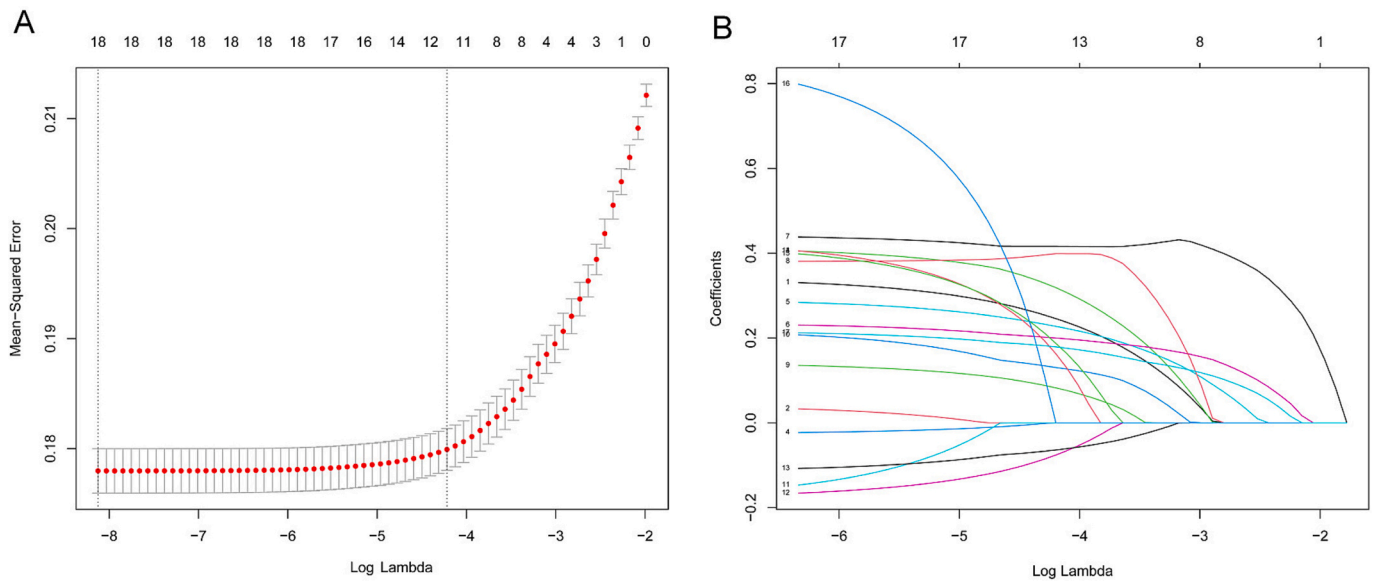
**Table 2**  
Baselines of prognostic cohort and external cohort.

Characteristics	Prognostic cohort (n = 21,099)	External cohort (n = 8290)	p
Age, years (%)			
<67	9623 (45.6)	3765 (45.4)	0.775
≥67	11,476 (54.4)	4525 (54.6)	
Sex (%)			
Female	12,354 (58.6)	4561 (55.0)	<0.001
Male	8745 (41.4)	3729 (45.0)	
Tumor size, cm (%)			
≤1	1486 (7.0)	469 (5.7)	<0.001
1–3	12,955 (61.4)	4984 (60.1)	
3–5	4482 (21.2)	2011 (24.3)	
5–7	1316 (6.2)	570 (6.9)	
>7	860 (4.1)	256 (3.1)	
M (%)			
M0	20,443 (96.9)	7924 (95.6)	<0.001
M1	656 (3.1)	366 (4.4)	
Grade (%)			
I	4852 (23.0)	1056 (12.7)	<0.001
II	10,073 (47.7)	4098 (49.4)	
III	6038 (28.6)	3067 (37.0)	
IV	136 (0.6)	69 (0.8)	
Surg.Prim.Site (%)			
Lobectomy	17,495 (82.9)	6971 (84.1)	<0.001
Pneumonectomy	538 (2.5)	284 (3.4)	
Sub-lobar	3066 (14.5)	1035 (12.5)	
Positive lymph node (%)			
0	16,291 (77.2)	6023 (72.7)	<0.001
1–3	3539 (16.8)	1719 (20.7)	
>3	1269 (6.0)	548 (6.6)	

generated as part of the internal validation process, involving bootstrapped resampling (1000 iterations) from both the training and testing groups. The nomogram’s time-dependent ROC was compared to AJCC and SEER staging, indicating improved prognostic accuracy with a higher AUC. Kaplan-Meier (K-M) survival curves were generated using the nomogram’s risk scores. External validation included a separate dataset of 8290 patients diagnosed between 2004 and 2009.

**Statistical analysis**

All statistical analyses were performed using R (version 4.3.1). Categorical variables were expressed as numbers and percentages. Potential factors associated with CSS were identified through LASSO regression and multivariate Cox regression analysis. The prognostic nomogram was developed using Cox hazard regression analysis and was presented in both static and dynamic network-based formats. To evaluate the



**Fig. 1.** Selection of predictors using LASSO regression analysis. (A) Plot of mean-squared error versus log lambda. The left vertical dotted line represents the optimal value with the fewest criteria, while the right vertical dotted line corresponds to the single standard error criterion. (B) LASSO coefficient profiles of the 15 variables. LASSO stands for least absolute shrinkage and selection operator.

nomogram’s performance, we conducted calibration, decision curve analysis (DCA), and time-dependent ROC curve analysis. Subsequently, the nomogram was used to assign risk scores to individual patients in both the training and test groups. Patients were then stratified into high and low-risk groups based on the median risk score. Survival rates between these groups were compared using Kaplan-Meier survival curves and log-rank tests. Statistical analysis was carried out using various software packages, including glmnet, caret, mctest, dcurves, pROC, regplot, rms, survival, timeROC, survminer, and DynNom. A p-value of <0.05 was considered statistically significant.

The p-values reported in Tables 1 and 2 were computed using the

Chi-square test implemented through the CreateTableOne function in the R package Tableone. Prognostic models based on independent predictors were developed utilizing the survivor package in R. The nomogram was constructed using the regplot package, while the dynamic nomogram was generated with the DynNom package. ROC curves were plotted using the pROC software, and the AUC was used to evaluate the discriminative ability of the nomogram. Furthermore, time-dependent ROC curves were generated, and AUC values at various time points were compared to the nomogram using the timeROC package. Calibration curves were created using the survival and rms packages, and DCA curves were plotted using the dcurves package. Finally, using the

**Table 3**  
The results of univariate and multivariate Cox analysis.

Characteristics	Univariate analysis			Multivariate analysis		
	HR	p	95%CI	HR	p	95%CI
Age, years						
<67						
≥67	1.357	0.000	1.291–1.426	1.518	0.000	1.443–1.597
Sex						
Female						
Male	1.497	0.000	1.425–1.572	1.389	0.000	1.322–1.459
Grade						
I						
II	1.903	0.000	1.763–2.054	1.547	0.000	1.432–1.672
III	2.921	0.000	2.702–3.159	1.959	0.000	1.805–2.125
IV	2.896	0.000	2.200–3.814	1.782	0.000	1.352–2.350
Positive lymph node						
0						
1–3	2.724	0.000	2.575–2.882	2.211	0.000	2.084–2.345
>3	4.470	0.000	4.151–4.814	3.464	0.000	3.205–3.744
Tumor size, cm						
0						
1–3	1.580	0.000	1.389–1.798	1.421	0.000	1.247–1.618
3–5	2.645	0.000	2.315–3.021	1.934	0.000	1.687–2.217
5–7	3.897	0.000	3.368–4.508	2.483	0.000	2.135–2.887
>7	5.322	0.000	4.577–6.188	3.695	0.000	3.160–4.321
M						
M0						
M1	3.081	0.000	2.786–3.406	2.226	0.000	2.011–2.465
Surg.Prim.Site						
Lobectomy						
Pneumonectomy	2.439	0.000	2.169–2.743	1.160	0.017	1.027–1.310
Sub-lobar	1.107	0.004	1.034–1.186	1.514	0.000	1.410–1.626

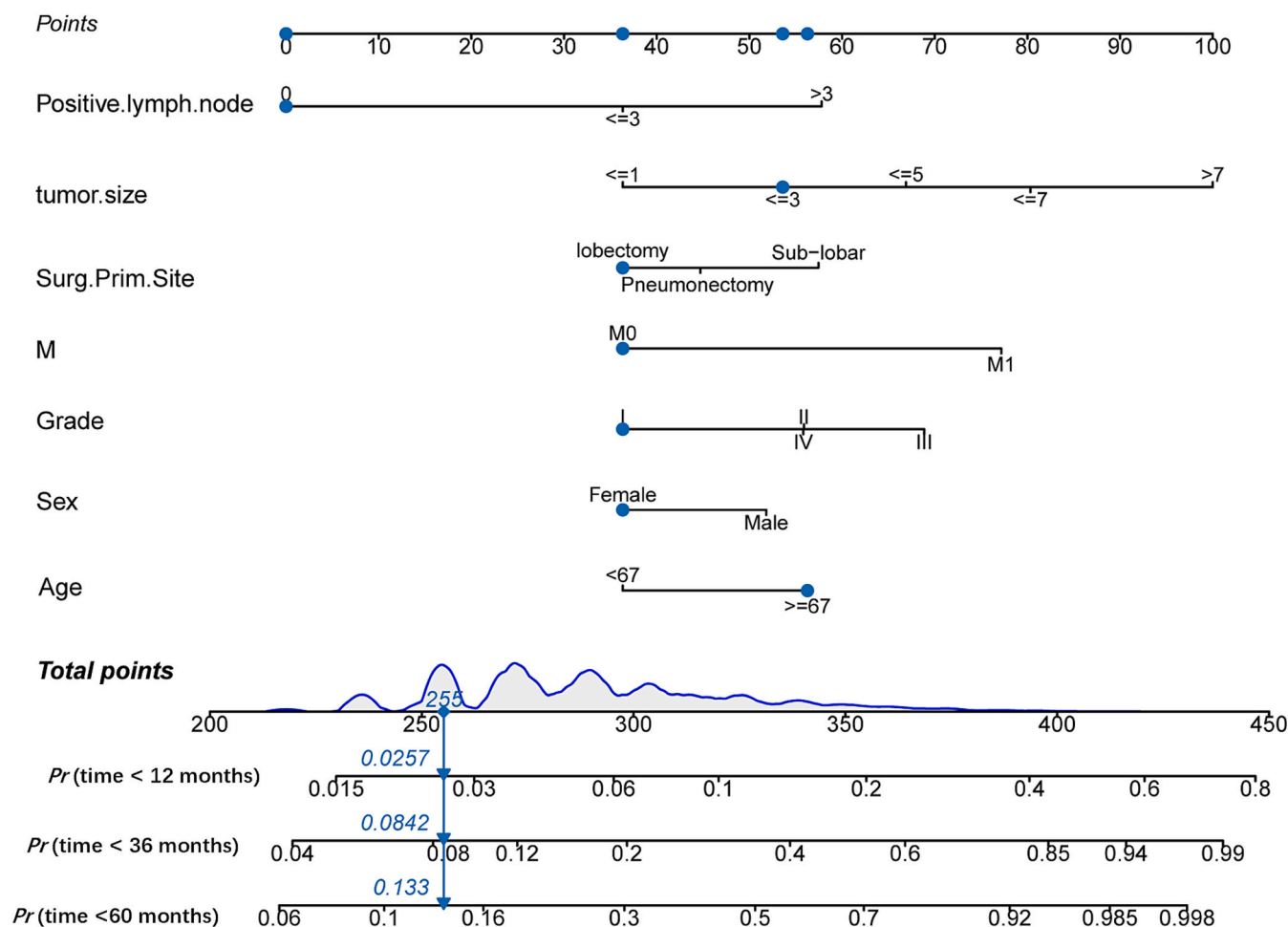


Fig. 2. Nomogram for prognostic assessment in GAD patients.

median risk score predicted by the nomogram, all patients were stratified into high-risk and low-risk groups. The prognostic value of the nomogram was then confirmed by validating survival curves through the logrank test, utilizing the survival package and survminer package to visualize Kaplan-Meier curves.

**Results**

*Demographic and clinical characteristics*

This study enrolled a cohort of 21,099 patients diagnosed with Lung Adenocarcinoma (LUAD) within the SEER database spanning the years 2010 to 2015. The development of the nomogram utilized data exclusively from the training subset comprising 14,770 patients. Subsequently, internal validation procedures were executed using the testing subset. Table 1 presents a comprehensive overview of the baseline characteristics for the entire patient cohort.

*Predictor selection*

To investigate relevant prognostic factors, we consolidated all patients into a unified cohort. Employing LASSO regression analysis, we adopted a coefficient penalization approach to mitigate variable overfitting and streamline the model, as illustrated in Fig. 1. We similarly analyzed the multicollinearity between the variables and the results are shown in (Fig. S1). We identified seven key variables through a combination of LASSO and multivariate Cox analyses: age, gender, tumor grade, count of involved lymph nodes, tumor size, M stage, and primary

surgical site. For comparison, both univariate and multivariate Cox regression analyses were conducted, and the results are detailed in Table 3. The seven potential variables, each serving as an individual risk factor for CSS, were included in the multivariate Cox regression analysis.

*Prognostic nomogram construction process*

In the training group, a comprehensive set of variables, including the number of positive lymph nodes, tumor size, age, sex, tumor grade, M-stage, and primary surgical site, were used to construct the prognostic nomogram depicted in Fig. 2. Nomogram validation was carried out using a separate testing group. The findings from the multivariate Cox analyses were presented in Fig. 3 as a forest plot illustrating the independent impacts of these predictors on CSS in LUAD patients. This representation included hazard ratios (HRs) and their corresponding 95 % confidence intervals (CIs). Surgery at the primary site emerged as a favorable prognostic factor (Pneumonectomy:HR = 1.160, 95 % CI: 1.027–1.310,  $p < 0.001$ ; Sub-lobar:HR = 1.514, 95 % CI: 1.410–1.626,  $p < 0.001$ ). Conversely, the remaining six prognostic factors were consistently associated with a less favorable prognosis. Grade II (HR = 1.547, 95 % CI: 1.432–1.672,  $p < 0.001$ ), Grade III (HR = 1.959, 95 % CI: 1.805–2.125,  $p < 0.001$ ), or Grade IV (HR = 1.782, 95 % CI: 1.352–2.350,  $p < 0.001$ ) tumors displayed relatively unfavorable prognoses when compared to well-differentiated tumors. Patients aged 67 years or older and male patients were also significantly linked to poorer outcomes (HR = 1.518, 95 % CI: 1.443–1.597,  $p < 0.001$ ; HR = 1.389, 95 % CI: 1.322–1.459,  $p < 0.001$ , respectively). Distant metastasis (HR = 2.226, 95 % CI: 2.011–2.465,  $p < 0.001$ ) was identified as a

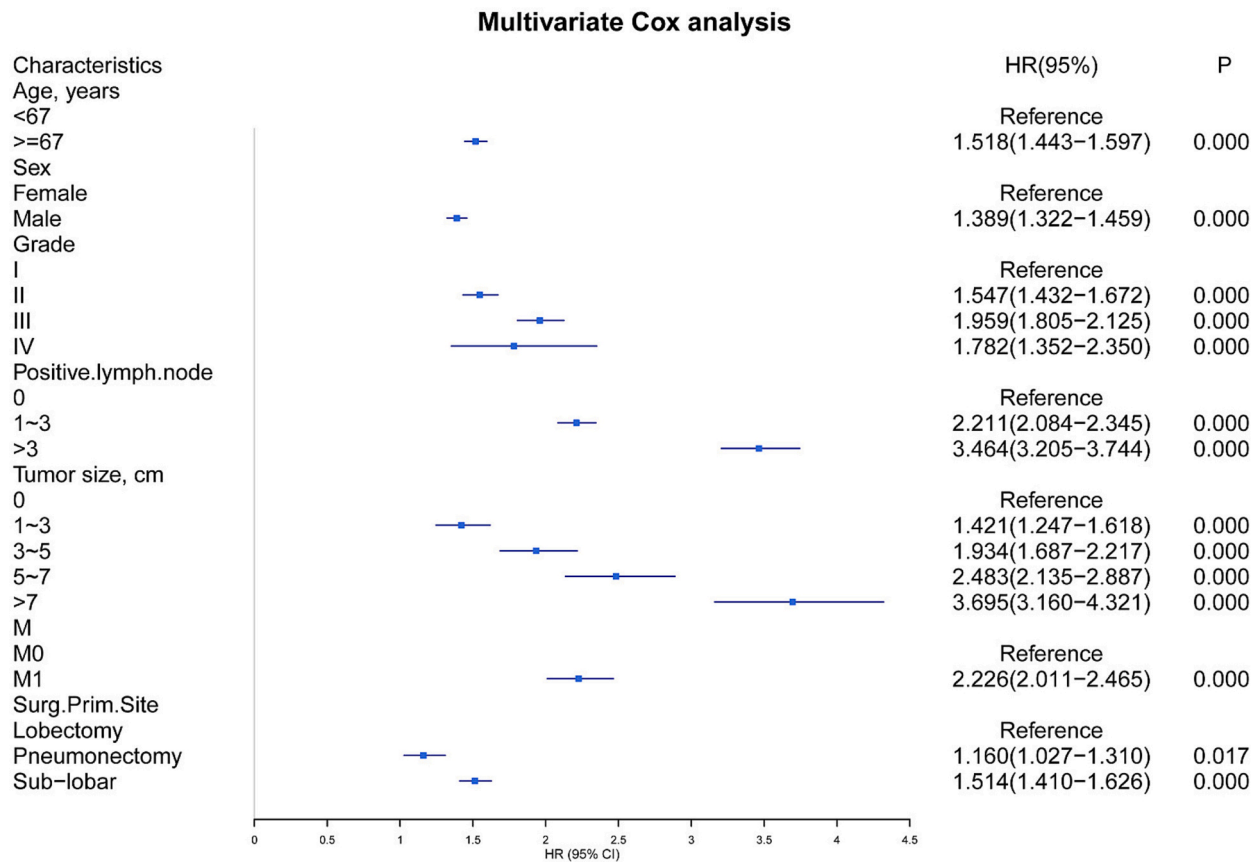


Fig. 3. Multivariate Cox analysis forest plot.

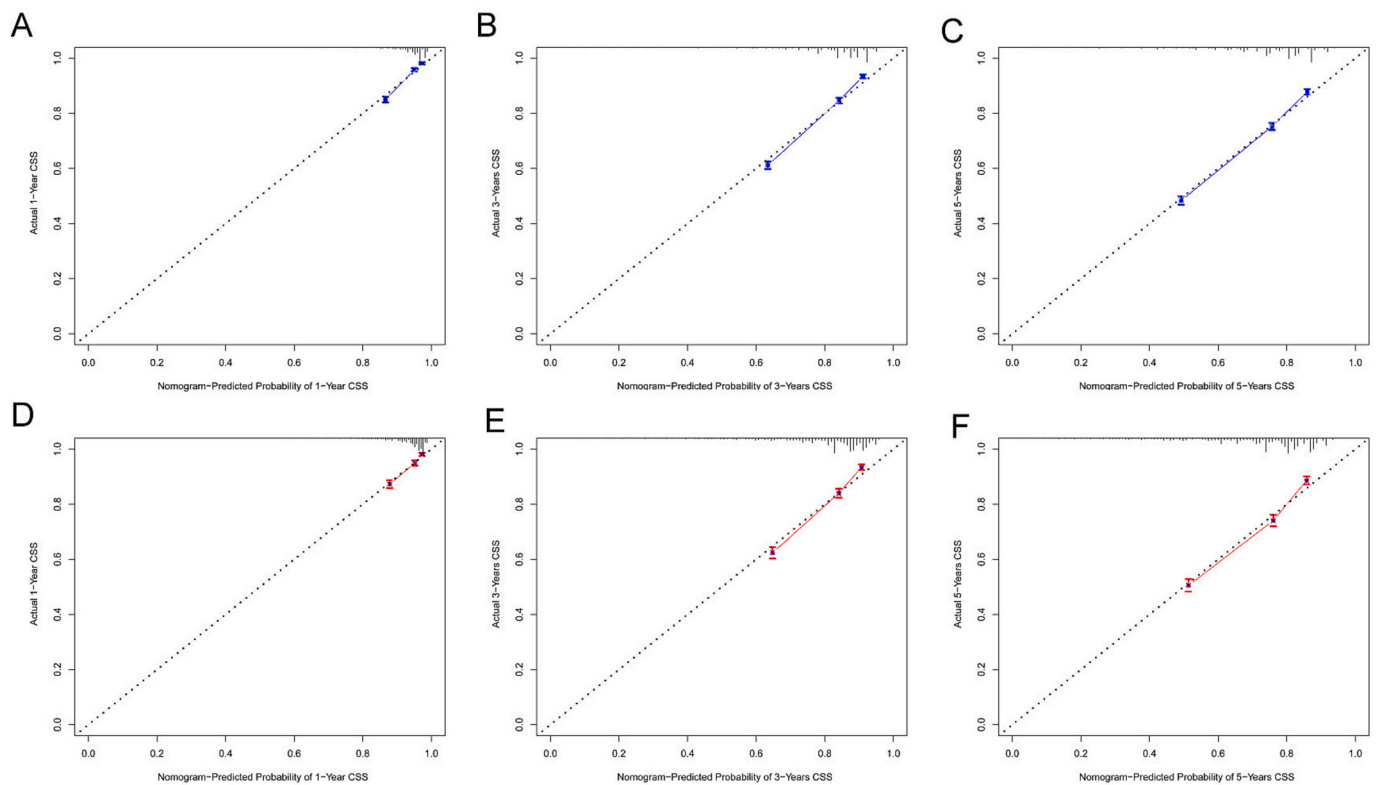
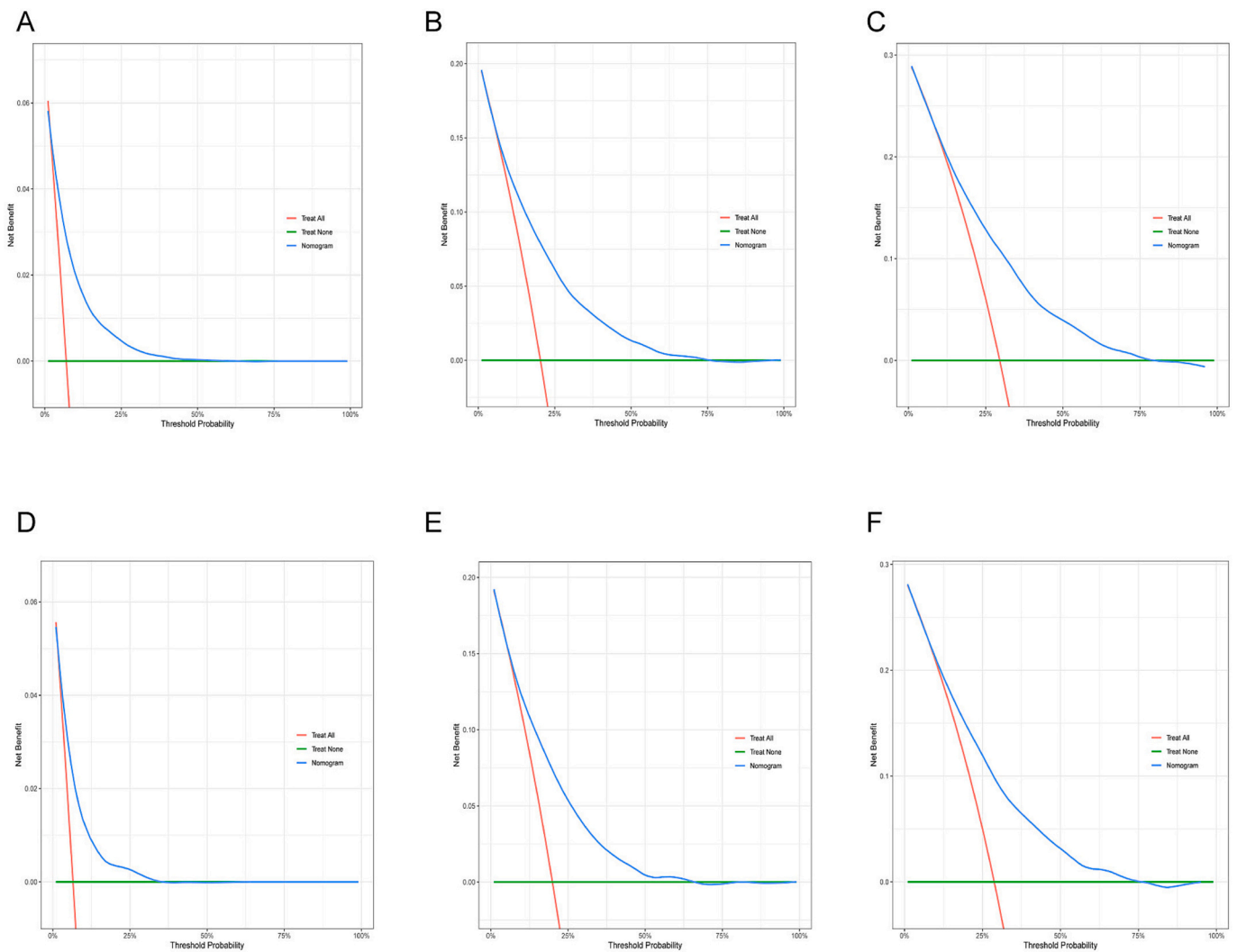


Fig. 4. Nomogram calibration curves. Panels (A), (B), and (C) represent the calibration curves at 1 year, 3 years, and 5 years, respectively, for the training group. Panels (D), (E), and (F) represent the calibration curves at 1 year, 3 years, and 5 years, respectively, for the validation group.



**Fig. 5.** DCA analysis of the nomogram. Panels (A), (B), and (C) represent DCA results at 1 year, 3 years, and 5 years, respectively, for the training group. Panels (D), (E), and (F) represent DCA results at 1 year, 3 years, and 5 years, respectively, for the validation group. DCA, Decision analysis curve.

significant contributor to a poorer prognosis. Additionally, both the lymph node positivity count and tumor size were recognized as risk factors influencing lung cancer prognosis.

#### Internal validation of the nomogram

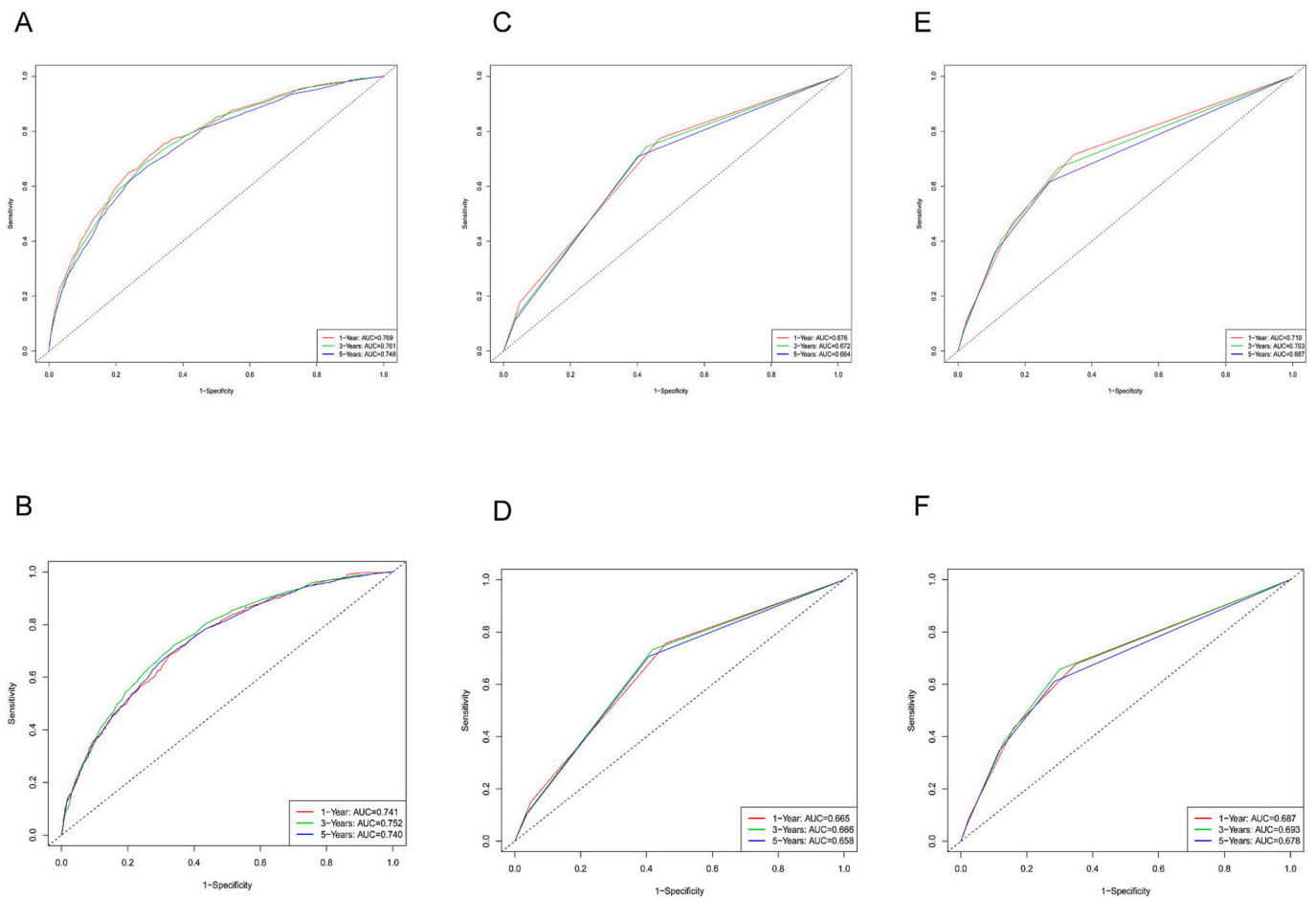
Internal validation was conducted to assess the nomogram's performance, and calibration curves, DCA, and ROC curves were generated over time. The nomogram exhibited robust calibration capabilities, evident from both the training group (Fig. 4A–C) and testing group (Fig. 4D–F) calibration curves. Furthermore, the nomogram demonstrated valuable clinical applicability, as evidenced by DCA plots for both the training (Fig. 5A–C) and testing groups (Fig. 5D–F). Time-dependent ROC curves were concurrently plotted for both groups. For the training group, the nomogram achieved 1-, 3-, and 5-year AUCs of 0.769, 0.761, and 0.748, respectively (Fig. 6A). In the testing group, the AUCs for the nomogram were 0.741, 0.752, and 0.740 for 1-, 3-, and 5-year intervals, respectively (Fig. 6B). Time-dependent ROC curves were also constructed for the AJCC stage (Fig. 6C, D) and SEER stage (Fig. 6E, F) in both the training and testing groups. Notably, the nomogram's AUC outperformed both the AJCC and SEER stages, substantiating its superior discriminatory capability. These findings collectively affirm the reliability of our model.

For further evaluation of the predictive model's effectiveness and

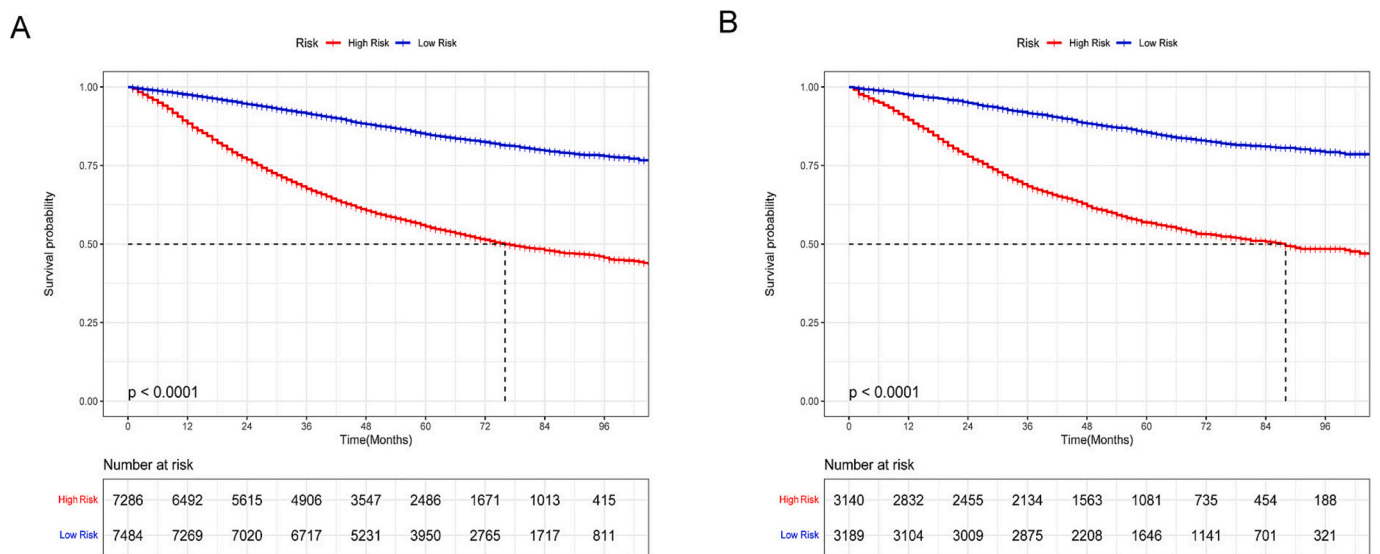
validity, patients were categorized into high-risk and low-risk groups using the median prognostic score obtained from the nomogram. Notably, the K-M survival curves displayed significant divergence ( $p < 0.001$ ), highlighting the model's ability to distinguish individuals at an elevated probability of cancer-related death (Fig. 7). These K-M survival curves underscored the considerably poorer CSS experienced by high-risk individuals in contrast to their low-risk counterparts. Additionally, to enhance the practical utility of our prognostic tool, we have developed an interactive online dynamic nomogram. Fig. 8 visually illustrates how this dynamic nomogram operates, enhancing its clinical usefulness.

#### Nomogram validation through external assessment

Table 2 presents the initial data for the essential variables in both cohorts. External validation, as illustrated by ROC curves (Fig. 9A–C) and DCA plots (Fig. 9D–F), reaffirmed the robust calibration capabilities and clinical applicability of the nomogram. The AUC values for the 1-, 3-, and 5-year time-dependent ROC curves for the nomogram were 0.718, 0.702, and 0.707, respectively (Fig. 10A). Moreover, K-M survival curves displayed significant differences in CSS between the high and low-risk groups (Fig. 10B).



**Fig. 6.** Nomogram, SEER stage, and AJCC stage’s time-dependent ROC curves. Panels (A) and (B) show the 1-, 3-, and 5-year time-dependent ROC curves for the nomogram in the training and validation groups, respectively. Panels (C) and (D) represent the 1-, 3-, and 5-year time-dependent ROC curves for the SEER stage in the training and validation groups, respectively. Panels (E) and (F) depict the 1-, 3-, and 5-year time-dependent ROC curves for the AJCC stage in the training and validation groups, respectively.



**Fig. 7.** Nomogram’s K-M curves. Panels (A) and (B) present the K-M curve for nomogram in the training group and validation group, respectively.

**Discussion**

In this study, age, gender, tumor grade, number of positive lymph

nodes, tumor size, M stage, and primary surgical site were identified as independent prognostic factors for LUAD using SEER data. Subsequently, a nomogram was developed to predict survival outcomes at 1,



Dynamic Nomogram

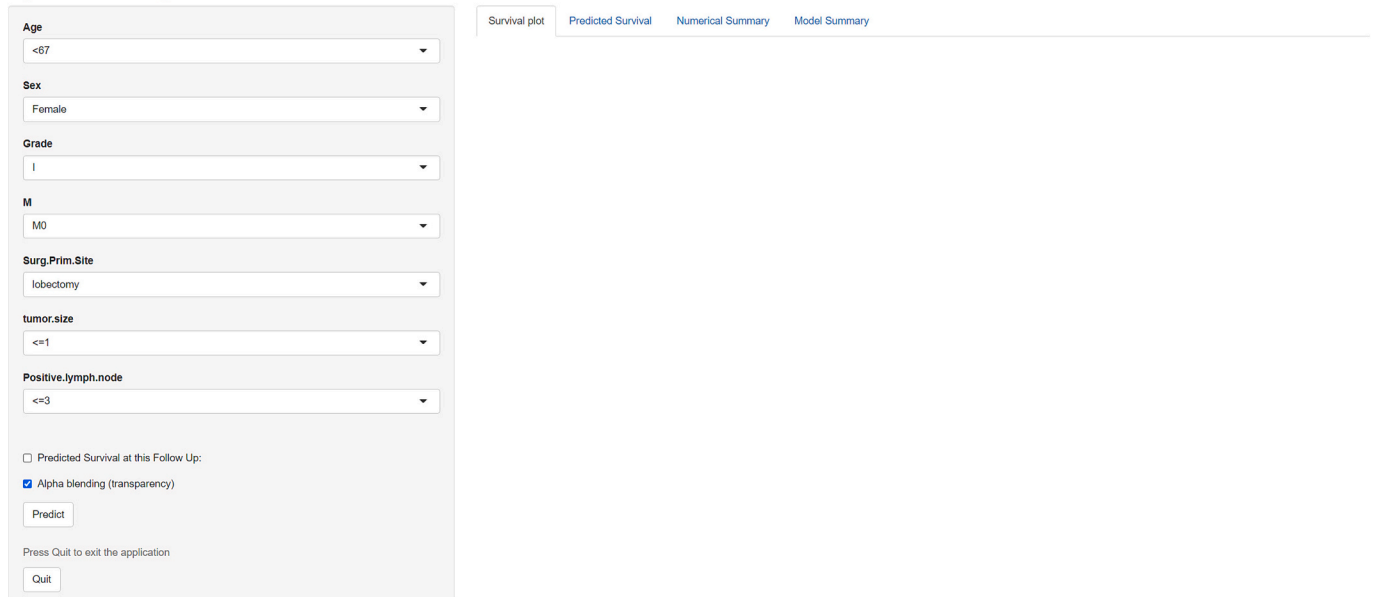


Fig. 8. Dynamic nomogram interface.

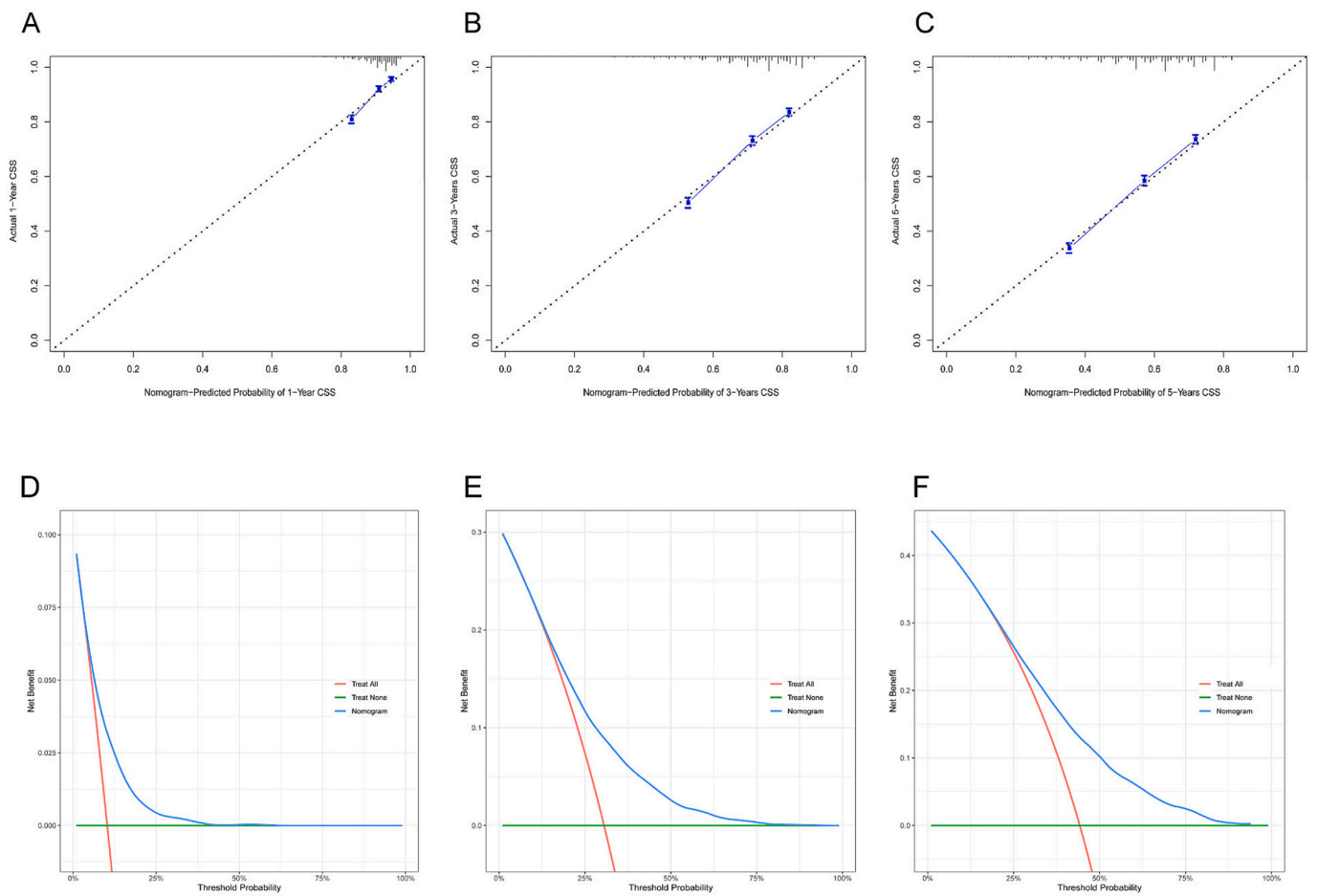
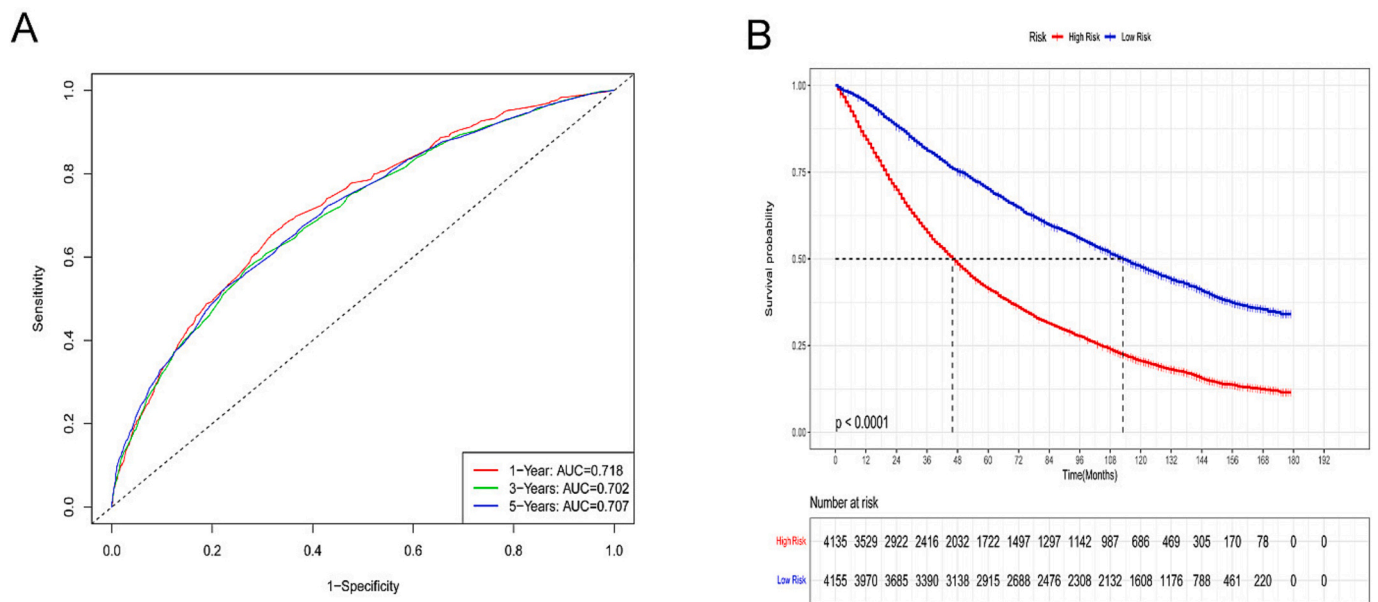


Fig. 9. External validation: calibration curves and DCA. Panels (A), (B), and (C) show the calibration curves for the nomogram at 1 year, 3 years, and 5 years in the external validation group, respectively. Panels (D), (E), and (F) depict the DCA results for the nomogram at 1 year, 3 years, and 5 years in the external validation group, respectively.



**Fig. 10.** External validation: the time-dependent ROC curves and K-M curves. Panel (A) displays the time-dependent ROC curves for the nomogram at 1 year, 3 years, and 5 years in the external validation group. Panel (B) illustrates the K-M curve for the nomogram in the external validation group.

3, and 5 years, demonstrating significant discriminative capabilities. Validation of the model, including calibration curves, DCA, and time-dependent ROC curves, confirmed its outstanding discriminative performance and concordance with actual observations. The prognostic nomogram was applicable to both the training and testing groups, enabling personalized predictions of LUAD-specific survival probabilities at 1-, 3-, and 5-years. To enhance user-friendliness, a static nomogram network was created based on a calculator that takes input values for the seven variables and computes survival rates with 95 % confidence intervals ([https://gflf993823.shinyapps.io/LUAD\\_CSS/](https://gflf993823.shinyapps.io/LUAD_CSS/)). The predictive model assists clinicians in evaluating individual risks and tailoring treatment and follow-up strategies for LUAD patients. High-risk individuals, given their heightened susceptibility to CSD, require comprehensive care and vigilant monitoring. Notably, the prognostic nomogram serves as a valuable complement to the AJCC and SEER staging systems, providing additional pertinent information.

The prognostic nomogram developed in our study demonstrated superior performance compared to other predictive models. In a study by Ma et al., they created a prognostic nomogram for LUAD, which yielded time-dependent AUC values of 0.707, 0.674, and 0.686 at 12, 24, and 36 months in the training cohorts and 0.690, 0.680, and 0.688 in the validation cohorts [13]. Similarly, Song et al. established a prognostic nomogram for NSCLC, reporting AUC values of 0.720, 0.706, and 0.708 at 1, 3, and 5 years in the training cohort, and 0.738, 0.696, and 0.680 in the validation cohorts [14]. The field of medical technology has witnessed significant advancements, leading to the widespread use of molecular diagnostic techniques in the study of malignant tumors. Molecular mechanisms and sequencing technologies have also played pivotal roles in tumor research [15–17]. Factors such as tumor-infiltrating lymphocytes [18], Long non-coding RNAs (LncRNAs) [19], DNA methylation [20], immune-related genes [21], interactions between cuproptosis and ferroptosis [22], and variations in autophagy-related genes [23,24] are all closely linked to lung cancer prognosis. The practical implementation of molecular testing in clinical settings remains challenging, despite the foundation provided by molecular-level studies for the personalized diagnosis and treatment of lung cancer patients.

LUAD, as the most prevalent subtype of NSCLC, presents a significant challenge. Approximately 57 % of patients receive their diagnosis when the disease has already advanced to the metastatic stage, resulting in a

dismal 5-year relative survival rate of only 5 % [25]. While patients undergoing immunotherapy and targeted biologic treatments have shown improved overall survival compared to those receiving traditional cytotoxic chemotherapy for non-small cell lung cancers (NSCLC) [26–28], the collective prognosis for LUAD remains notably unfavorable. Therefore, it is imperative to investigate LUAD patient prognosis by providing tools that offer personalized survival information. The nomogram we have developed serves as one such tool, significantly streamlining the treatment process.

Numerous investigations have delved into prognostic factors within LUAD. Nguyen et al. unveiled the prognostic potential of the lepidic cell gene signature, offering insights into prognosis and susceptibility to immunotherapy among LUAD patients [29]. Recent research has further delineated micropapillary and solid lymph node metastases as autonomous indicators of unfavorable prognosis in LUAD patients [30]. Our study successfully identified seven independent prognostic determinants, encompassing age, gender, tumor grade, number of positive LNs, tumor dimensions, M stage, and primary surgical locus, for the comprehensive evaluation of LUAD prognosis. By amalgamating these variables, we conceived a novel prognostic nomogram tailored to LUAD patients. Rigorous internal and external validations unequivocally substantiated the superior precision of our model in providing an individualized prognosis assessment vis-à-vis the conventional AJCC staging system. Nevertheless, our study is not without its limitations. Firstly, due to its retrospective nature, the specter of selection bias looms. Secondly, the precise quantification of data pertaining to laboratory tests and imaging data, eluded us, potentially bearing implications of consequence. Lastly, further validation employing external datasets, with particular emphasis on data from healthcare institutions in Asia, is imperative to enhance the model's generalizability.

## Conclusions

This study identified independent prognostic factors for Lung Adenocarcinoma (LUAD), including age, sex, tumor grade, number of positive lymph nodes, tumor size, M stage, and primary site of surgery. A novel nomogram was developed to predict 1-, 3-, and 5-years CSS, demonstrating strong clinical applicability and discriminatory power. The nomogram, coupled with the online prediction tool, provides support to clinicians in appraising mortality risk and formulating tailored

treatment and follow-up approaches. Prospective multicenter investigations are necessary for further validation.

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.sopen.2024.10.003>.

### Ethics statement

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

### Funding statement

This work was supported in part by the Natural Science Foundation of Gansu Province (21JR1RA062).

### CRediT authorship contribution statement

**Hong Guo:** Writing – review & editing. **Guole Nie:** Data curation, Writing – review & editing. **Xin Zhao:** Data curation. **Jialu Liu:** Data curation. **Kaihua Yu:** Data curation. **Yulan Li:** Funding acquisition.

### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### Data availability

Publicly available datasets were analyzed in this study. This data can be found here: <https://seer.cancer.gov/>.

### References

- [1] Ferlay J, Colombet M, Soerjomataram I, Parkin DM, Piñeros M, Znaor A, et al. Cancer statistics for the year 2020: an overview[J]. *Int J Cancer* 2021;149(4):778–89.
- [2] Arbour KC, Riely GJ. Systemic therapy for locally advanced and metastatic non-small cell lung cancer: a review. *JAMA* 2019;322:764–74.
- [3] Wang T, She Y, Yang Y, Liu X, Chen S, Zhong Y, et al. Radiomics for survival risk stratification of clinical and pathologic stage IA pure-solid non-small cell lung cancer. *Radiology* 2022;302:425–34.
- [4] Mukherjee K, Davison N, Malik S, Duszak R, Kokabi N. National utilization, survival, and costs analysis of treatment options for stage I non-small cell lung cancer: a SEER-Medicare database analysis. *Acad Radiol* 2022;29(Suppl. 2):S173–S80.
- [5] Harada G, Yang S-R, Cocco E, Drilon A. Rare molecular subtypes of lung cancer. *Nat Rev Clin Oncol* 2023;20:229–49.
- [6] Zhang Q, Dai Y, Liu H, Sun W, Huang Y, Gong Z, et al. Causes of death and conditional survival estimates of long-term lung cancer survivors. *Front Immunol* 2022;13:1012247.
- [7] Wang S, Yu Y, Xu W, Lv X, Zhang Y, Liu M. Dynamic nomograms combining N classification with ratio-based nodal classifications to predict long-term survival for patients with lung adenocarcinoma after surgery: a SEER population-based study. *BMC Cancer* 2021;21:653.
- [8] Pan Z, You H, Bu Q, Feng X, Zhao F, Li Y, et al. Development and validation of a nomogram for predicting cancer-specific survival in patients with Wilms' tumor. *J Cancer* 2019;10:5299–305.
- [9] Kajiwara Y, Oka S, Tanaka S, Nakamura T, Saito S, Fukunaga Y, et al. Nomogram as a novel predictive tool for lymph node metastasis in T1 colorectal cancer treated with endoscopic resection: a nationwide, multicenter study. *Gastrointest Endosc* 2023;97.
- [10] Veiga LHS, Vo JB, Curtis RE, Mille MM, Lee C, Ramin C, et al. Treatment-related thoracic soft tissue sarcomas in US breast cancer survivors: a retrospective cohort study. *Lancet Oncol* 2022;23:1451–64.
- [11] Nie G, Zhang H, Yan J, Xie D, Zhang H, Li X. Construction and validation of a novel nomogram to predict cancer-specific survival in patients with gastric adenocarcinoma. *Front Oncol* 2023;13:1114847.
- [12] Albertsen PC. Re: trends and practices for managing low-risk prostate cancer: a SEER-Medicare study. *Prostate Cancer Prostatic Dis* 2022;25.
- [13] Ma C, Peng S, Zhu B, Li S, Tan X, Gu Y. The nomogram for the prediction of overall survival in patients with metastatic lung adenocarcinoma undergoing primary site surgery: a retrospective population-based study. *Front Oncol* 2022;12:916498.
- [14] Song X, Xie Y, Zhu Y, Lou Y. Is lobectomy superior to sub-lobectomy in non-small cell lung cancer with pleural invasion? A population-based competing risk analysis. *BMC Cancer* 2022;22:541.
- [15] Yang K, Halima A, Chan TA. Antigen presentation in cancer - mechanisms and clinical implications for immunotherapy. *Nat Rev Clin Oncol* 2023;20:604–23.
- [16] Yang J, Xu J, Wang W, Zhang B, Yu X, Shi S. Epigenetic regulation in the tumor microenvironment: molecular mechanisms and therapeutic targets. *Signal Transduct Target Ther* 2023;8:210.
- [17] Dietlein F, Wang AB, Fagre C, Tang A, Besselink NJM, Cuppen E, et al. Genome-wide analysis of somatic noncoding mutation patterns in cancer. *Science* 2022;376:eabg5601.
- [18] Guo X, Zhang Y, Zheng L, Zheng C, Song J, Zhang Q, et al. Global characterization of T cells in non-small-cell lung cancer by single-cell sequencing. *Nat Med* 2018;24:978–85.
- [19] He Y, Jiang X, Duan L, Xiong Q, Yuan Y, Liu P, et al. LncRNA PKMYT1AR promotes cancer stem cell maintenance in non-small cell lung cancer via activating Wnt signaling pathway. *Mol Cancer* 2021;20:156.
- [20] Kim-Wanner S-Z, Assenov Y, Nair MB, Weichenhan D, Benner A, Becker N, et al. Genome-wide DNA methylation profiling in early stage I lung adenocarcinoma reveals predictive aberrant methylation in the promoter region of the long noncoding RNA PLUT: an exploratory study. *J Thorac Oncol* 2020;15:1338–50.
- [21] Shi X, Li R, Dong X, Chen AM, Liu X, Lu D, et al. IRGS: an immune-related gene classifier for lung adenocarcinoma prognosis. *J Transl Med* 2020;18:55.
- [22] Shen Y, Li D, Liang Q, Yang M, Pan Y, Li H. Cross-talk between cuproptosis and ferroptosis regulators defines the tumor microenvironment for the prediction of prognosis and therapies in lung adenocarcinoma. *Front Immunol* 2022;13:1029092.
- [23] Wen J, Liu H, Wang L, Wang X, Gu N, Liu Z, et al. Potentially functional variants of ATG16L2 predict radiation pneumonitis and outcomes in patients with non-small cell lung cancer after definitive radiotherapy. *J Thorac Oncol* 2018;13:660–75.
- [24] Chen M-M, Guo W, Chen S-M, Guo X-Z, Xu L, Ma X-Y, et al. Xanthine dehydrogenase rewires metabolism and the survival of nutrient deprived lung adenocarcinoma cells by facilitating UPR and autophagic degradation. *Int J Biol Sci* 2023;19:772–88.
- [25] Zhao C, Liu J, Zhou H, Qian X, Sun H, Chen X, et al. NEIL3 may act as a potential prognostic biomarker for lung adenocarcinoma. *Cancer Cell Int* 2021;21:228.
- [26] Halliday PR, Blakely CM, Bivona TG. Emerging targeted therapies for the treatment of non-small cell lung cancer. *Curr Oncol Rep* 2019;21:21.
- [27] Wang M, Herbst RS, Boshoff C. Toward personalized treatment approaches for non-small-cell lung cancer. *Nat Med* 2021;27:1345–56.
- [28] Marinelli D, Mazzotta M, Scalera S, Terrenato I, Sperati F, D'Ambrosio L, et al. KEAP1-driven co-mutations in lung adenocarcinoma unresponsive to immunotherapy despite high tumor mutational burden. *Ann Oncol* 2020;31:1746–54.
- [29] Nguyen TT, Lee H-S, Burt BM, Wu J, Zhang J, Amos CI, et al. A lepidic gene signature predicts patient prognosis and sensitivity to immunotherapy in lung adenocarcinoma. *Genome Med* 2022;14:5.
- [30] Li Y, Byun AJ, Choe JK, Lu S, Restle D, Eguchi T, et al. Micropapillary and solid histologic patterns in N1 and N2 lymph node metastases are independent factors of poor prognosis in patients with stages II to III lung adenocarcinoma. *J Thorac Oncol* 2023;18:608–19.