A nomogram prognostic model for large cell lung cancer: analysis from the Surveillance, Epidemiology and End Results Database

Gang Lin, Kang Qi, Bing Liu, Haibo Liu, Jian Li

Department of Thoracic Surgery, Peking University First Affiliated Hospital, Peking University, Beijing, China

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

Correspondence to: Gang Lin, MD. Department of Thoracic Surgery Peking University First Affiliated Hospital Peking University, Dahongluo Street 8, Xicheng District, Beijing 100034, China. Email: lingang0608@163.com.

Background: Currently, there is no reliable method for predicting the prognosis of patients with large cell lung cancer (LCLC). The aim of this study was to develop and validate a nomogram model for accurately predicting the prognosis of patients with LCLC.

Methods: LCLC patients, diagnosed from 2007 to 2009, were identified from the Surveillance, Epidemiology and End Results (SEER) database and used as the training dataset. Significant clinicopathologic variables (P<0.05) in a multivariate Cox regression were selected to build the nomogram. The performance of the nomogram model was evaluated by the concordance index (C-index), the area under the curve (AUC), and internal calibration. LCLC patients diagnosed from 2010 to 2016 in the SEER database were selected as a testing dataset for external validation. The nomogram model was also compared with the currently used American Joint Committee on Cancer (AJCC) tumor-node-metastasis (TNM) staging system (8th edition) by using C-index and a decision curve analysis.

Results: Eight variables—age, sex, race, marital status, T stage, N stage, M stage, and treatment strategy—were statistically significant in the multivariate Cox model and were selected to develop the nomogram model. This model exhibited excellent predictive performance. The C-index and AUC value were 0.761 [95% confidence interval (CI), 0.754 to 0.768] and 0.886 for the training dataset and 0.773 (95% CI, 0.765 to 0.781) and 0.876 for the testing dataset, respectively. This model also predicted three-year and five-year lung cancer-specific survival (LCSS) in both datasets with good fidelity. This nomogram model performs significantly better than the 8th edition AJCC TNM staging system, with a higher C-index (P<0.001) and better net benefits in predicting LCSS in LCLC patients.

Conclusions: We developed and validated a prognostic nomogram model for predicting 3- and 5-year LCSS in LCLC patients with good discrimination and calibration abilities. The nomogram may be useful in assisting clinicians to make individualized decisions for appropriate treatment in LCLC.

Keywords: Large cell lung cancer (LCLC); prognosis; nomogram; prognostic model

Introduction

Large cell lung cancer (LCLC) is a rare pathological type of lung cancer, accounting for 2–3% of non-small cell lung cancers (NSCLC) (1). Compared with other NSCLCs, LCLC grows faster and metastasizes earlier, leading to a higher degree of malignancy (2). Its reported five-year survival rate is about 15–25% (3). Similar to other types of NSCLC, treatment for LCLC consists of surgery, radiation, and chemotherapy, depending on the stage of the tumor and

the general condition of the patient (3,4). Previous studies have reported several clinicopathological factors that affect the prognosis of LCLC, including the LCLC subtype, tumor location, tumor stage, and others (5,6).

Nomogram is a graphical prognostic calculating tool for predicting the prognosis of cancer patients, which may facilitate better prognosis assessment and treatment stratification. Studies have shown that this user-friendly survival prediction tool can improve medical care in patients with gastric cancer, liver cancer, small cell lung cancer, and major salivary gland cancer (7-10). However, to the best of our knowledge, no similar nomogram tool is available for predicting LCLC prognosis until now.

In September 2019, He *et al.* (11) published the only study of nomograms in large cell neuroendocrine carcinoma (LCNEC). In his study, he found that age, gender, tumor stage, N stage, size, and surgery of primary site were the independent prognostic factors of 3- and 5-year overall survival (OS) and developed a nomogram model. He's nomogram has some limitation. He only analyzed the impact of surgery on the prognosis of LCNEC patients but does not analyze the impact of the other two commonly used treatment methods (chemotherapy and radiotherapy) on the prognosis of LCNEC. In addition, LCNEC was not a subtype of LCLC in the 2015 World Health Organization (WHO) Lung Tumors Classification, which had been grouped with other neuroendocrine tumors (12).

The aim of this study was to collect information on LCLC patients from the Surveillance, Epidemiology and End Results (SEER) database, perform statistical analysis to identify prognostic clinicopathological factors and develop and validate a new nomogram model for LCLC patients. A reliable and high-performance nomogram model would be useful for identifying high-risk patients and planning appropriate adjuvant therapeutic strategies for LCLC patients.

Methods

This study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The present study was approved by the Ethics Committee of the Peking University First Hospital (ethics number: 2018-236) and individual consent for this retrospective analysis was waived.

Study population

This retrospective study was conducted by acquiring data from the SEER database using SEER*STAT 8.3.5 software. A flow-chart illustrating the methodology used for identifying cases of LCLC in the SEER database during 2007–2016 (*Figure 1*). Patients with LCLC were identified using International Classification of Diseases for Oncology (ICD-O) topography codes of "8012.2", "8012.3", and "8014.3". All patients enrolled between 2007 and 2016 were selected. Included patients met the following criteria: (I) patients with age ≥18 years; (II) patients with pathologically confirmed LCLC; (III) patients with only one primary tumor; (IV) patients with accurate follow-up information. Exclusion criteria: (I) patients with incomplete data; (II) patients with LCNEC. LCNEC was excluded because its biological, clinical, and prognostic characteristics in advanced stages were similar to those of SCLC. In the 2015 World Health Organization (WHO) Lung Tumors Classification, LCNEC was grouped with other neuroendocrine tumors and was no longer a subtype of LCLC (12). A total of 3,641 LCLC cases were identified from the SEER database, of which 2,971 patients met our inclusion criterion and were included in our study. These patients were divided into two groups based on the year of diagnosis: the training dataset (diagnosed from 2007 to 2009) and the testing dataset (diagnosed from 2010 to 2016). The nomogram model was developed from the 1,669 LCLC patients from the training dataset and validated by the 1,302 cases in the testing dataset.

Collected variables

Collected data consisted of clinicopathological, treatment, and follow-up information, including the year of diagnosis, sex, age, race, site of primary tumor, laterality, grade, chemotherapy, radiotherapy, marital status, the American Joint Committee on Cancer (AJCC) tumor-node-metastasis (TNM) stage, T category, N category, M category, tumor size, surgical status, survival status, overall survival (OS) time, and lung cancer-specific survival (LCSS) time. Age was a continuous variable, which was converted to a categorical variable in this study. The optimal cutoff values determined by X-tile (version 3.6.1, Yale University) software were 54- and 81-year, respectively. OS was defined as the interval from diagnosis to death or last follow-up, and LCSS was calculated from the date of diagnosis to the date of cancer-specific death. The AJCC TNM stage, T descriptor, N descriptor, and M descriptor were re-classified according to the 8th edition lung cancer staging system defined by the AJCC (13). Treatment methods
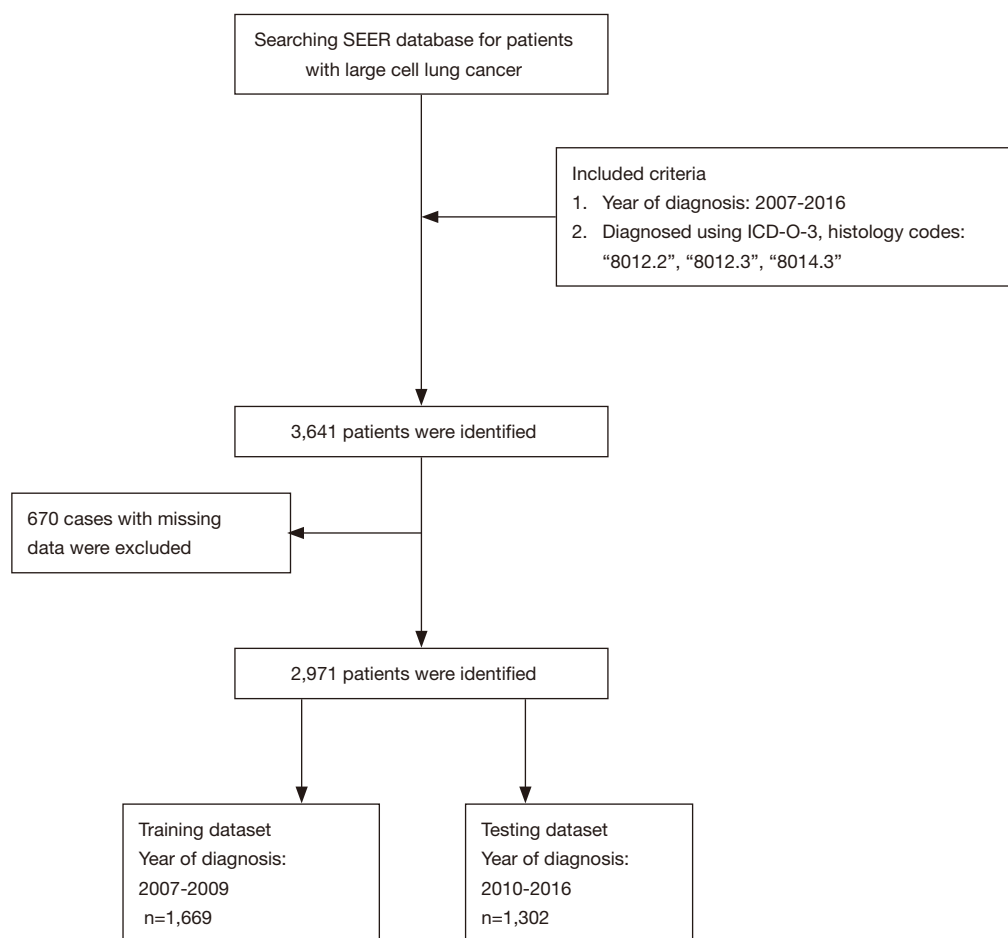
Figure 1 Flow-chart illustrating the methodology used for identifying cases of large cell lung cancer in the SEER database during 2007–2016. SEER, Surveillance, Epidemiology and End Results database.

included surgery, chemotherapy, and radiation therapy, and different treatment methods were combined to form seven treatment strategies, including no treatment, surgery (S), chemotherapy (C), radiotherapy (R), surgery combined with chemotherapy (SC), surgery combined with radiotherapy (SR), chemotherapy combined with radiotherapy (CR), and surgery combined with chemoradiotherapy (SCR).

Establishing nomogram prognostic model

The nomogram was established using the training dataset. Clinicopathological variables, including age, sex, race, marital status, site of the primary tumor, tumor grade, laterality, T descriptor, N descriptor, M descriptor, and treatment strategy were statistically analyzed to select prognostic factors for LCSS. Multivariate Cox regression

analysis was used to estimate the predictive factors and their weights. Akaike information criterion (AIC) was used to develop multivariate models by removing predictors that were less statistically significant starting from a full model containing all predictive variables (14).

Validation of model

The following methods were used to evaluate the predictive performance of this nomogram model. First, a risk score was calculated for each case in the training dataset based on the established Cox regression model. By quartile stratification of the risk score, the patients in the training dataset were divided into four groups, and Kaplan-Meier survival curves were plotted according to the level of the risk score. Second, bootstrapping (1,000 repetitions) was

used to obtain a relatively unbiased estimate of the models' performance (8). The concordance index (C-index) and a calibration plot were used to determine the performance of the model in distinguishing between high-risk and low-risk patients. Third, the area under the curve (AUC) of time-dependent receiver operating characteristics was calculated. Fourth, external validation using cases in the testing dataset was performed, the C-index was calculated, and AUC and the calibration curve were plotted. Finally, the performance of the nomogram model was compared to the AJCC TNM staging system (8th edition) by using the C-index and decision curve analysis (15,16).

Statistical Methods

Continuous variables were converted to categorical variables with the median used as a cutoff value, and categorical variables were denoted as percentages. Differences between groups were evaluated with χ^2 tests. Survival outcome was calculated according to the Kaplan-Meier method, and log-rank tests were used to determine significant differences between survival curves. All statistical analyses were conducted using SPSS 22.0 software (SPSS Inc., Chicago, IL, USA) and R version 3.6.0 (R Foundation for Statistical Computing, Vienna, Austria). R packages "foreign", "survival", "rms", "Hmisc", "rmda", "survivalROC", and "compareC" were used. The source codes of R software in this study were provided in Supplementary (Appendix 1). Results with P value <0.05 were considered statistically significant.

Results

Clinicopathologic characteristics of patients with LCLC

The characteristics of the two datasets are shown in *Table 1*. In comparing the training and testing datasets, T descriptor (P=0.031), M descriptor (P<0.001), and treatment strategy (P=0.047) are significantly different. Other variables, such as age, sex, race, marital status, site of the primary tumor, laterality, N descriptor, and surgical treatment are similar between groups. The median follow-up time is 6.9 months (range, 0 to 119 months) for the training dataset and 6.0 months (range, 0 to 83 months) for the testing dataset. The LCSS and OS in the training dataset are shown in *Figure 2*. The three-year LCSS and OS were 20.2% and 17.4%, respectively. Five-year LCSS and OS were 15.2% and 11.6%, respectively.

Independent prognostic factors for LCSS and OS in the training dataset

As shown in *Table 2*, independent risk factors for LCSS are age, sex, race, T descriptor, N descriptor, M descriptor, marital status, and treatment strategy. The above-mentioned variables are also independent prognostic factors for OS. In addition, although the site of the primary tumor is not a prognostic factor for LCSS (P=0.120), it is an independent risk factor for OS (P=0.048). Other variables, such as tumor grade and laterality, are not prognostic factors for either LCSS or OS.

Three- and five-year LCSS-predicting nomogram model

Based on the results of the Cox multivariate regression analysis, age, sex, race, marital status, T descriptor, N descriptor, M descriptor, and treatment strategies were selected to build the final nomogram model. The three- and five-year LCSS nomogram for the training dataset is shown in *Figure 3*. The treatment strategy shows the largest range of risk scores, followed by the T descriptor, indicating that these two factors have the greatest impact on the prognosis.

Validation of the nomogram model

The predicted patient survival probability curve
Based on the established nomogram model, a risk score was calculated for each case in the training dataset, which ranged from 64 to 317 (median 207). Patients in the training dataset were further divided into four groups based on quartiles of calculated risk scores (group 1: <161, group 2: 161–207, group 3: 208–240 and group 4: >240). Kaplan-Meier survival curves were plotted for each risk level (*Figure 4*). The survival time of different risk levels was statistically different (all P values <0.001), indicating that this nomogram model has a good ability to distinguish between high-risk and low-risk patients.

Calibration curve
Bootstrapping (1,000 repetitions) with resampling of the 1,669 patients was used to obtain a relatively unbiased estimate of the model's performance. Calibration plots for the prediction of three- and five-year LCSS are shown in *Figure 5*, demonstrating a high degree of agreement between the nomogram prediction using the model and actual observations. The C-index for LCSS prediction in the training dataset is 0.761 [95% confidence interval (CI), 0.754 to 0.768] with an AUC of 0.886.

Table 1 Clinicopathologic characteristics of patients with LCLC

Characteristics	Total	Training dataset	Testing dataset	P value
Year of diagnosis	2007–2016	2007–2009	2010–2016	
No. of patients	2,971	1,669 (56.2%)	1,302 (43.8%)	
Age (years)				0.287
<54	427	248 (14.9%)	179 (13.7%)	
54–81	2,248	1,245 (74.6%)	1,003 (77.0%)	
>81	296	176 (10.5%)	120 (9.2%)	
Sex				0.634
Male	1,738	970 (58.1%)	768 (59.0%)	
Female	1,233	699 (41.9%)	534 (41.1%)	
Race				0.320
Caucasian	2,362	1,313 (78.7%)	1,049 (80.6%)	
Black	464	267 (16.0%)	197 (15.1%)	
Other races	145	89 (5.3%)	56 (4.3%)	
Marital status				0.117
Married	1,511	870 (52.1%)	641 (49.2%)	
Unmarried	1,460	799 (47.9%)	661 (50.8%)	
Site of the primary tumor				0.533
Trachea	139	73 (4.4%)	66 (5.1%)	
Upper lobe	1,680	929 (55.7%)	751 (57.7%)	
Middle lobe	123	70 (4.2%)	53 (4.1%)	
Lower lobe	702	416 (24.9%)	286 (22.0%)	
Overlapping lesion	34	19 (1.1%)	15 (1.2%)	
Unknown	293	162 (9.7%)	131 (10.1%)	
Tumor grade				0.338
Well + moderated	32	17 (1.0%)	15 (1.2%)	
Poor + undifferentiated	1,853	1,023 (61.3%)	830 (63.7%)	
Unknown	1,086	629 (37.7%)	457 (35.1%)	
Laterality				0.361
Left-sided	1,189	679 (40.7%)	510 (39.2%)	
Right-sided	1,699	949 (56.9%)	750 (57.6%)	
Unknown	83	41 (2.5%)	42 (3.2%)	
T				0.031
T0	27	9 (0.5%)	18 (1.4%)	
T1	542	305 (18.3%)	237 (18.2%)	
T2	792	441 (26.4%)	351 (27.0%)	
T3	480	292 (17.5%)	188 (14.4%)	
T4	1,130	622 (37.3%)	508 (39.0%)	

Table 1 (*continued*)

Table 1 (*continued*)

Characteristics	Total	Training dataset	Testing dataset	P value
N				0.336
N0	1,046	596 (35.7%)	450 (34.6%)	
N1	265	155 (9.3%)	110 (8.4%)	
N2	1,212	683 (40.9%)	529 (40.6%)	
N3	448	235 (14.1%)	213 (16.4%)	
M				<0.001
M0	1,287	705 (42.2%)	582 (44.7%)	
M1	1,684	964 (57.8%)	720 (55.3%)	
Surgery				0.495
No	2,310	1,290 (77.3%)	1,020 (78.3%)	
Yes	661	379 (22.7%)	282 (21.7%)	
Treatment strategy				0.047
No treatment	717	379 (22.7%)	338 (26.0%)	
C	373	228 (13.7%)	145 (11.1%)	
CR	759	412 (24.7%)	347 (26.7%)	
R	461	271 (16.2%)	190 (14.6%)	
S	373	227 (13.6%)	146 (11.2%)	
SC	130	68 (4.1%)	62 (4.8%)	
SR	48	27 (1.6%)	21(1.6%)	
SRC	110	57 (3.4%)	53 (4.1%)	

C, chemotherapy; CR, chemoradiotherapy; LCLC, large cell lung cancer; M, M descriptor in the 8th TNM staging system of lung cancer; N, N descriptor in the 8th TNM staging system of lung cancer; R, radiotherapy; S, surgery; SC, surgery combined with chemotherapy; SR, surgery combined with radiotherapy; SRC, surgery combined with chemoradiotherapy; T, T descriptor in the 8th TNM staging system of lung cancer.

External validation

An external validation using cases in the testing dataset was performed. The C-index and AUC for LCSS prediction in the testing dataset are 0.773 (95% CI, 0.765 to 0.781) and 0.876, respectively. A calibration curve shows good agreement between prediction and observation in the probability of three- and five-year LCSS (*Figure 6*).

Comparison of predictive ability for LCSS in LCLC patients between the present nomogram model and the AJCC TNM staging system (8th edition)

The performance of the AJCC TNM staging system and the present nomogram model are shown in *Table 3*. The results show that the predictive ability of the current nomogram model is significantly better than the AJCC TNM staging system (P<0.001). The decision curve analyses show that the nomogram model has better net benefits compared with the AJCC 8th edition TNM staging system (*Figure 7*).

Discussion

Currently, the prognosis of patients with LCLC or other types of NSCLC is predicted using the AJCC TNM staging system (17). It has been well established that LCLC has a higher degree of malignancy and a poorer prognosis than other cancers (2). Therefore, a prediction method based on the TNM stage is not accurate enough to meet the needs of LCLC patients. It is, therefore, necessary to establish a new tool for better predicting the prognosis of LCLC based on
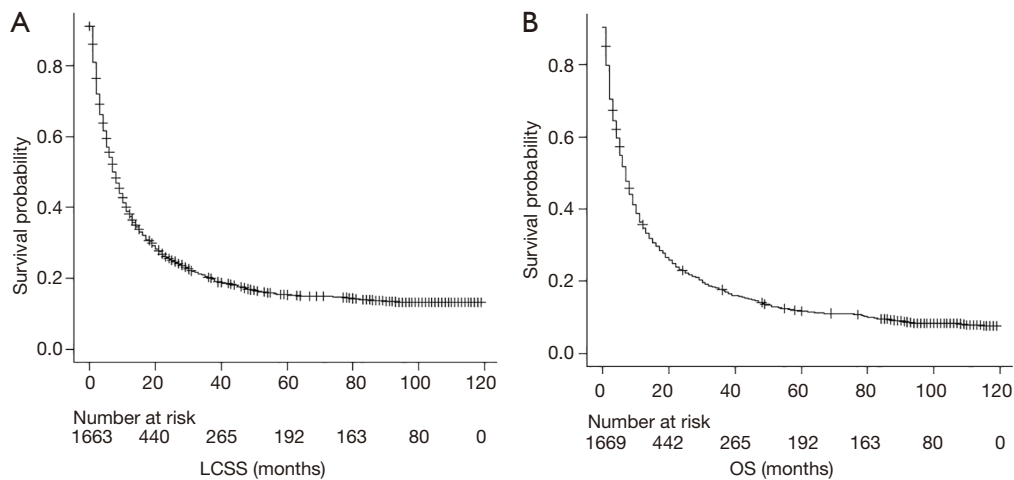
Figure 2 Lung cancer-specific survival (A) and overall survival (B) curves of patients with large cell lung cancer in the training dataset. LCSS, lung cancer-specific survival; OS, overall survival.

Table 2 Multivariate Cox regression analysis for LCSS and OS in patients with LCLC

Characteristics	Subtypes	LCSS			OS		
		HR	95% CI	P value	HR	95% CI	P value
Age (y)	54–81	1			1		
	<54	0.802	0.683–0.941	0.007	0.768	0.658–0.895	0.001
	>81	1.317	1.103–1.572	0.002	1.369	1.159–1.616	<0.001
Sex	Male	1			1		
	Female	0.853	0.761–0.955	0.006	0.840	0.754–0.935	0.001
Race	Caucasian	1			1		
	Black	0.875	0.751–1.019	0.086	0.891	0.772–1.030	0.118
	Other races	0.709	0.553–0.910	0.007	0.678	0.535–0.859	0.001
Marital status	Unmarried	1			1		
	Married	0.890	0.795–0.997	0.044	0.892	0.802–0.993	0.036
Primary site	Upper	1			1		
	Middle	1.131	0.853–1.502	0.393	1.096	0.839–1.430	0.502
	Overlapping	1.135	0.674–1.912	0.633	0.911	0.542–1.531	0.724
	Trachea	1.342	1.038–1.735	0.025	1.415	1.109–1.806	0.005
	Lower	1.126	0.988–1.283	0.075	1.121	0.992–1.268	0.067
	Unknown	1.189	0.971–1.455	0.093	1.172	0.964–1.425	0.112
Grade	W/M	1			1		
	P/U	1.387	0.759–2.533	0.287	1.338	0.784–2.284	0.286
	Unknown	1.344	0.733–2.463	0.339	1.308	0.763–2.241	0.329

Table 2 (*continued*)

Table 2 (*continued*)

Characteristics	Subtypes	LCSS			OS		
		HR	95% CI	P value	HR	95% CI	P value
Laterality	Unknown	1			1		
	Left	0.728	0.495–1.070	0.106	0.698	0.480–1.015	0.060
	Right	0.709	0.484–1.038	0.077	0.680	0.470–0.985	0.042
T							
	T4	1			1		
	T0	0.309	0.122–0.785	0.013	0.398	0.179–0.885	0.024
	T1	0.712	0.603–0.841	<0.001	0.734	0.628–0.858	< 0.001
	T2	0.763	0.662–0.880	< 0.001	0.803	0.702–0.918	0.001
	T3	0.875	0.751–1.019	0.086	0.875	0.755–1.014	0.075
N							
	N3	1			1		
	N0	0.718	0.599–0.860	< 0.001	0.770	0.648–0.915	0.003
	N1	1.131	0.901–1.420	0.290	1.161	0.933–1.444	0.181
	N2	0.964	0.823–1.128	0.646	0.958	0.821–1.117	0.581
M							
	M1	1			1		
	M0	0.487	0.427–0.555	<0.001	0.524	0.463–0.593	<0.001
Treatment							
	S	1			1		
	C	1.869	1.421–2.458	< 0.001	1.504	1.183–1.912	0.001
	CR	2.157	1.508–1.035	< 0.001	1.705	1.372–2.119	<0.001
	R	3.522	2.700–4.594	< 0.001	2.904	2.309–3.654	<0.001
	No	4.923	3.806–6.369	< 0.001	3.951	3.166–4.930	<0.001
	SC	1.216	0.843–1.753	0.295	0.988	0.717–1.363	0.943
	SR	2.087	1.293–3.368	0.003	2.904	2.309–3.654	<0.001
	SRC	1.508	1.035–2.196	0.032	1.143	0.812–1.610	0.444

CI, confidence interval; C, chemotherapy; CR, chemoradiotherapy; HR, hazard ratio; LCLC, large cell lung cancer; LCSS, lung cancer-specific survival; M, M descriptor in the 8th TNM staging system of lung cancer; N, N descriptor in the 8th TNM staging system of lung cancer; OS, overall survival; P/U, poorly and undifferentiated; R, radiotherapy; S, surgery; SC, surgery combined with chemotherapy; SR, surgery combined with radiotherapy; SRC, surgery combined with chemoradiotherapy; T, T descriptor in the 8th TNM staging system of lung cancer; W/M, well and moderately differentiated.
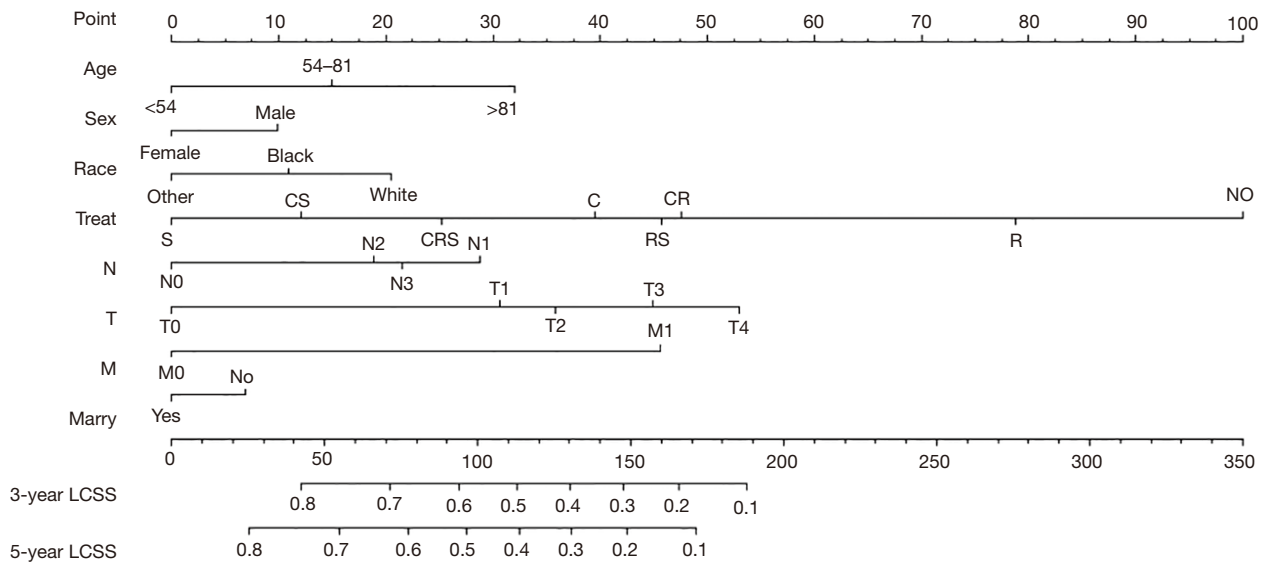
Figure 3 Three- and five-year lung cancer-specific survival nomogram for patients with LCLC. LCLC, large cell lung cancer; LCSS, lung cancer-specific survival; M, M descriptor in the 8th TNM staging system of lung cancer; N, N descriptor in the 8th TNM staging system of lung cancer; Treat-C, chemotherapy; Treat-CR, chemoradiotherapy; Treat-R, radiotherapy; Treat-S, surgery; Treat-SR, surgery combined with radiotherapy; Treat-CRS, surgery combined with chemoradiotherapy; Treat-NO, no treatment; T, T descriptor in the 8th TNM staging system of lung cancer.
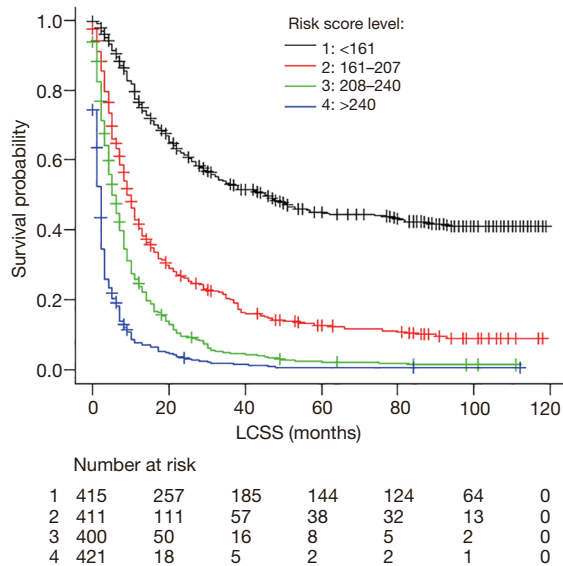
Figure 4 Predicted patient survival probability curve corresponding to risk score levels [1–4]. LCSS, lung cancer-specific survival.

routinely used clinicopathological variables. In this study, we have developed and validated a nomogram prognostic model using a large cohort of LCLC cases from the NIH/NCI SEER database. This nomogram model, consisting of common demographics, staging, and treatment information, predicts the probability of long-term survival for individual LCLC patients very well.

The C-index and AUC of the nomogram model in the training dataset were 0.761 (95% CI, 0.754 to 0.768) and 0.886, respectively. The performance of the model was similar in the testing set, with a C-index of 0.773 (95% CI, 0.765 to 0.781) and an AUC of 0.876, indicating that the nomogram model had a strong predictive ability. In addition, this model had better performance than the conventionally used 8th edition AJCC TNM staging system, with a higher C-index (0.767 vs. 0.676, P<0.001) and an improved prediction of LCSS in LCLC patients. Thus, this comprehensive and personalized risk score calculation method might be used as stratification criteria and applied to clinical practice.

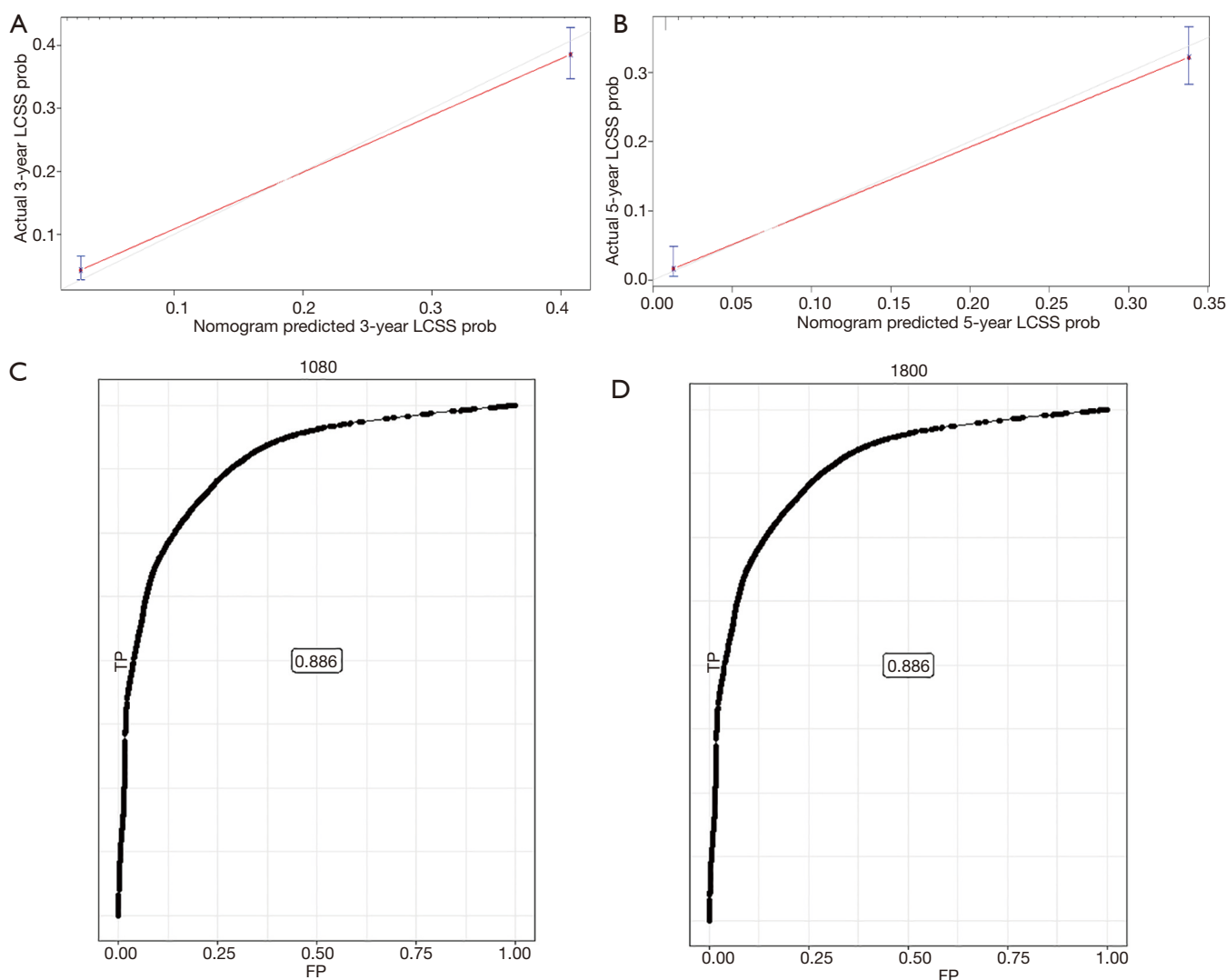In 2019, He et al. (11) developed a nomogram model for LCNEC with a C-index of 0.75 for the training sets

Figure 5 Calibration plots of nomogram predictions of three- (A) and five-year (B) lung cancer-specific survival of patients with large cell lung cancer in the training dataset. Redline: equality of the observed and predicted probability. AUC: 0.861 as calculated from the three- (C) and five-year (D) nomogram prognostic model. AUC, the area under the curve; FP, false positive; LCSS, lung cancer-specific survival; TP, true positive.

and 0.76 for the validation set. The C-index of the current nomogram model was 0.761 and 0.773 in the training and testing datasets, respectively, which was slightly better than the nomogram published by He *et al*. It was worth noting, patients with LCNEC were excluded in our study because its biological, clinical, and prognostic characteristics in advanced stages were similar to those of SCLC (2). In the WHO Classification Lung Tumors Classification [2015], LCNEC had been grouped with other neuroendocrine tumors. This study has some advantages: (I) the number of cases included in the training and testing datasets was large;

(II) this study compared the impact of different treatment strategies on the prognosis of LCLC patients, which was in line with the actual situation of LCLC treatment and was easy for the clinician to use; (III) this nomogram was verified by various methods, and the results show that the nomogram was highly reliable.

Many clinical factors affect the survival of LCLC patients. The nomogram model developed in this study includes the following clinicopathological variables: age, sex, ethnicity, marital status, T descriptor, N descriptor, M descriptor, and treatment strategy. This nomogram model
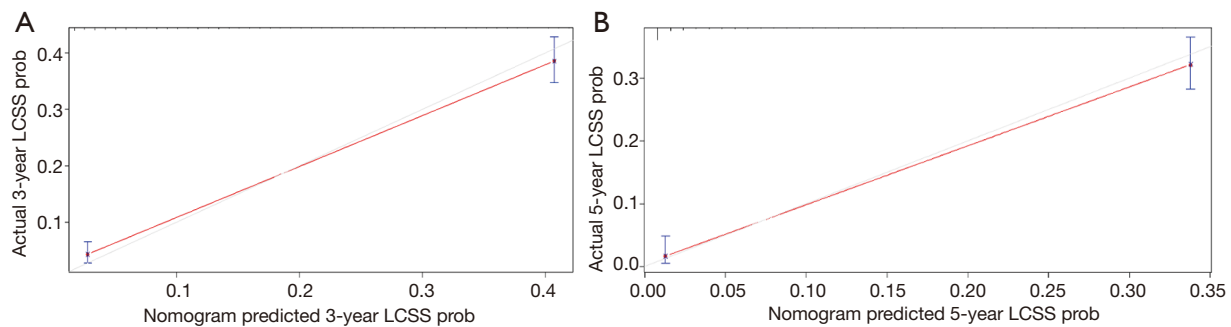
Figure 6 Calibration plots of nomogram predictions of three- (A) and five-year (B) lung cancer-specific survival of patients with large cell lung cancer in the testing dataset. Redline: equality of the observed and predicted probability. LCSS, lung cancer-specific survival.

Table 3 Comparison of predictive ability for lung cancer-specific survival between the nomogram and the AJCC TNM staging system (8th) in the training dataset

Measures	8th AJCC TNM stage	Present nomogram	P value
C-index	0.676	0.767	<0.001
Likelihood ratio test	499.60	933.60	<0.001

AJCC, the American Joint Committee on Cancer; LCLC, large cell lung cancer; TNM, the tumor-node-metastasis stage.

includes more variables than the 8th edition AJCC TNM staging system and thus has a higher predictive ability. Many studies have reported that age is an independent predictor of prognosis. Lara *et al.* (18) studied 114,451 patients with NSCLC and found that patients <50 years old had improved LCSS (HR 0.827, P<0.001). Wu *et al.* (19) built a predictive survival model for LCSS and OS in NSCLC after radical resection and found that age >60 is associated with a worse OS. Gender also has been found to have an impact on the lung cancer patient prognosis over time (20,21). For example, Ferguson *et al.* (20) studied 772 patients with lung cancer and found that females outlived males (P<0.0001). Many previous studies have reported that the prognosis of NSCLC is affected by ethnicity. It was reported that Asian or Hispanic ethnicities have better survival rates compared to those with white or black ancestry (22,23). Our findings from the present study support this conclusion. Marital status also has an impact on survival in various types of cancer, such as breast, prostate, colorectal, gastric, and NSCLC (24-26). Wu *et al.* (26) carried out a population-based study of 70,006 patients with NSCLC and found that married patients have an advantage over the unmarried in both OS and LCSS. The explanation for this phenomenon is that married cancer patients are more likely to have

better financial support, be diagnosed at earlier stages, receive recommended treatments, and make healthier choices due to support from their spouse (27). Treatment strategies significantly affected the prognosis of LCLC patients (28-30). Our study demonstrated that surgery was the most important treatment with the lowest risk score, and cases receiving no treatment had the highest risk score. Interestingly, we found that adjuvant therapy did not improve the prognosis of LCLC patients, compared with surgery alone, which is different from the treatment strategy of LCNEC (4). We analyzed this phenomenon and proposed the following reasons: LCLC is one of the pathological types of NSCLC and having similar biological characteristics as other NSCLCs. LCLC is frequently not sensitive to chemotherapy and surgical resection is the first-line treatment for operable LCLC (31). Regarding the effect of surgical treatment on LCLC, Hanagiri and his colleagues (32) conducted a survey of 975 patients who had undergone resection for NSCLC. He reported that the 5-year survival rate was 61.5% after surgery, and proposing that complete resection is the preferred treatment. However, sometimes, the prognosis for LCLC was dismal even after curative resection. In this case, adjuvant therapy was needed. The indication of adding adjuvant therapy should follow the treatment strategy of other NSCLCs.
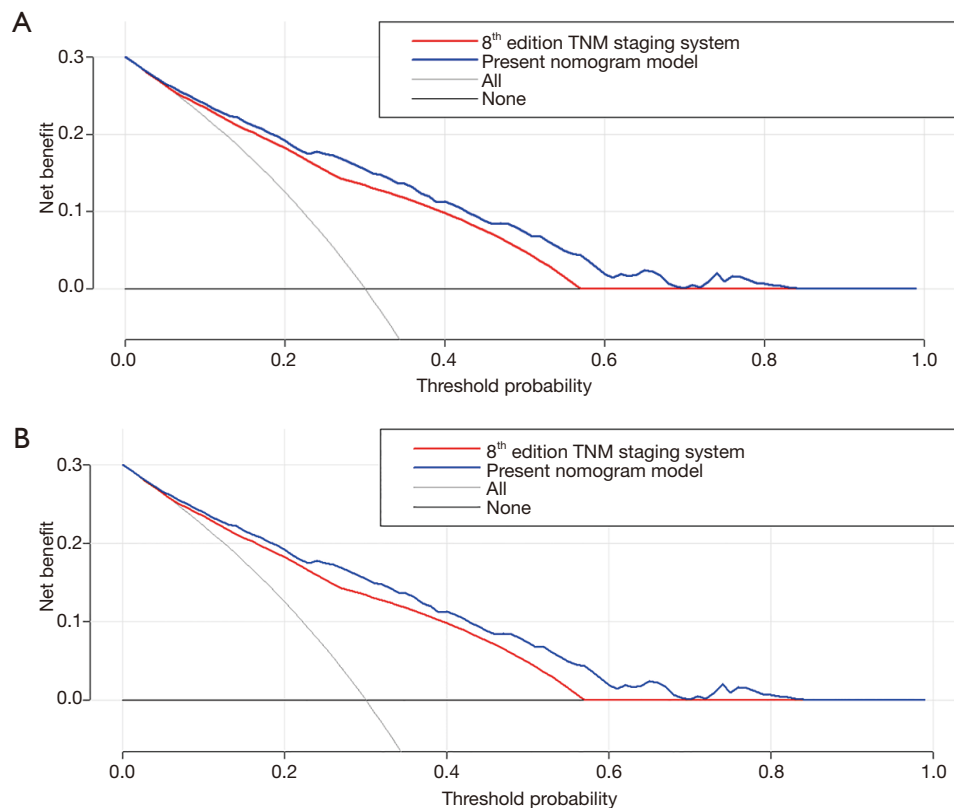
Figure 7 Decision curve analysis of the current nomogram model and the 8th edition AJCC TNM staging system for predicting three- (A) and five-year (B) lung cancer-specific survival. Horizontal line (Threshold probability) referred to the recurrence probability, which was predicted using the nomogram model or the 8th edition TNM staging system. When the recurrence probability exceeded a certain value, the clinician should consider starting treatment for the patient. After starting the treatment, recurrence patients would be benefited from the treatment (benefit), and those who have not relapsed would be harmed by the treatment (disadvantage). The net benefit = benefit – disadvantage. Decision curve analysis was a tool to evaluate the net benefit of starting treatment. The grey lines represent extreme situations. The horizontal line represents that all patients are untreated and the net benefit is zero. The oblique line indicates that all patients have been treated, and the net benefit is a backslash with a negative slope. Both the nomogram model (blue line) and the 8th edition AJCC TNM staging system (red line) have obvious gains in net benefits compared to the extreme curve. The decision curve of the nomogram model is better than that of the AJCC TNM staging system.

It had been reported that platinum-based chemotherapy significantly improved OS (31). Although surgery yields the best prognosis for LCLC, the selection of treatment strategies should be made based on multiple factors, such as the TNM stage of the tumor, the patient's general conditions, and the ability of the hospital to provide a high level of comprehensive treatment for a patient. The optimal treatment plan for a patient should be determined only after the comprehensive evaluation.

This study had several limitations. First, this was a retrospective study, and selection bias could therefore

not be avoided. Second, it should be acknowledged that there are many variables affecting LCLC prognosis, including smoking and chronic obstructive pulmonary disease (COPD), which were not provided by the SEER database. However, our study had certain advantages as well. First, to our knowledge, our study is one of the only studies to predict the prognosis of LCLC patients in a nomogram model. Second, this study used a large-scale subject population to develop a nomogram and performed both internal and external validations, which made the nomogram model highly reliable.

Conclusions

In summary, we developed and validated a new nomogram model for predicting the prognosis of LCLC patients with good calibration. The nomogram performed well with satisfactory discrimination. The ability of this nomogram model to predict the prognosis was not inferior to the 8th edition AJCC TNM staging system. This nomogram may be useful in assisting the clinician in predicting the oncological prognosis of LCLC patients and making decisions regarding appropriate treatment strategies.

Acknowledgments

Footnote

Data Sharing Statement: Available at http://dx.doi.org/10.21037/tlcr-19-517b

Peer Review File: Available at http://dx.doi.org/10.21037/tlcr-19-517b

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at http://dx.doi.org/10.21037/tlcr-19-517b). 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. This study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The ethics committee of Peking University First Hospital approved this study (ethical number: 2018-236) and individual consent for this retrospective analysis was waived.

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

References

1. Howlader N, Noone AM, Krapcho M. SEER Cancer Statistics Review, 1975-2013. Bethesda (MD): National Cancer Institute; 2015.
2. Asamura H, Kameya T, Matsuno Y, et al. Neuroendocrine neoplasms of the lung: a prognostic spectrum. J Clin Oncol 2006;24:70-6.
3. Gollard R, Jhatakia S, Elliott M, et al. Large cell/neuroendocrine carcinoma. Lung Cancer 2010;69:13-8.
4. Lo Russo G, Pusceddu S, Proto C, et al. Treatment of lung large cell neuroendocrine carcinoma. Tumour Biol 2016;37:7047-57.
5. Sun YH, Lin SW, Hsieh CC, et al. Treatment outcomes of patients with different subtypes of large cell carcinoma of the lung. Ann Thorac Surg 2014;98:1013-9.
6. Liang R, Chen TX, Wang ZQ, et al. A retrospective analysis of the clinicopathological characteristics of large cell carcinoma of the lung. Exp Ther Med 2015;9:197-202.
7. Kim SY, Yoon MJ, Park YI, et al. Nomograms predicting survival of patients with unresectable or metastatic gastric cancer who receive combination cytotoxic chemotherapy as first-line treatment. Gastric Cancer 2018;21:453-63.
8. Wang Y, Li J, Xia Y, et al. Prognostic nomogram for intrahepatic cholangiocarcinoma after partial hepatectomy. J Clin Oncol 2013;31:1188-95.
9. Wang S, Yang L, Ci B, et al. Development and Validation of a Nomogram Prognostic Model for SCLC Patients. J Thorac Oncol 2018;13:1338-48.
10. Lu CH, Liu CT, Chang PH, et al. Develop and validation a nomogram to predict the recurrent probability in patients with major salivary gland cancer. J Cancer 2017;8:2247-55.
11. He Y, Liu H, Wang S, et al. Prognostic nomogram predicts overall survival in pulmonary large cell neuroendocrine carcinoma. PLoS One 2019;14:e0223275.
12. Travis WD, Brambilla E, Nicholson AG, et al. The 2015 World Health Organization Classification of Lung Tumors: Impact of Genetic, Clinical and Radiologic Advances Since the 2004 Classification. J Thorac Oncol 2015;10:1243-60.
13. Detterbeck FC, Boffa DJ, Kim AW, et al. The Eighth Edition Lung Cancer Stage Classification. Chest 2017;151:193-203.
14. Harrell FE Jr, 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.

15. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. Med Decis Making 2006;26:565-74.

16. Ma K, Dong B, Wang L, et al. Nomograms for predicting overall survival and cancer-specific survival in patients with surgically resected intrahepatic cholangiocarcinoma. Cancer Manag Res 2019;11:6907-29.

17. Wang J, Wu N, Zheng Q, et al. Evaluation of the 7th edition of the TNM classification for lung cancer at a single institution. J Cancer Res Clin Oncol 2014;140:1189-95.

18. Lara MS, Brunson A, Wun T, et al. Predictors of survival for younger patients less than 50 years of age with non-small cell lung cancer (NSCLC): a California Cancer Registry analysis. Lung Cancer 2014;85:264-9.

19. Wu CY, Fu JY, Wu CF, et al. Survival Prediction Model Using Clinico-Pathologic Characteristics for Nonsmall Cell Lung Cancer Patients After Curative Resection. Medicine (Baltimore) 2015;94:e2013.

20. Ferguson MK, Skosey C, Hoffman PC, et al. Sex-associated differences in presentation and survival in patients with lung cancer. J Clin Oncol 1990;8:1402-7.

21. Torre LA, Siegel RL, Jemal A. Lung Cancer Statistics. Adv Exp Med Biol 2016;893:1-19.

22. Soneji S, Tanner NT, Silvestri GA, et al. Racial and Ethnic Disparities in Early-Stage Lung Cancer Survival. Chest 2017;152:587-97.

23. Patel MI, Wang A, Kapphahn K, et al. Racial and Ethnic Variations in Lung Cancer Incidence and Mortality: Results From the Women's Health Initiative. J Clin Oncol 2016;34:360-8.

24. Aizer AA, Chen MH, McCarthy EP, et al. Marital status and survival in patients with cancer. J Clin Oncol 2013;31:3869-76.

25. Li Q, Gan L, Liang L, et al. The influence of marital status on stage at diagnosis and survival of patients with colorectal cancer. Oncotarget 2015;6:7339-47.

26. Wu Y, Ai Z, Xu G. Marital status and survival in patients with non-small cell lung cancer: an analysis of 70006 patients in the SEER database. Oncotarget 2017;8:103518-34.

27. Chang SM, Barker FG 2nd. Marital status, treatment, and survival in patients with glioblastoma multiforme: a population based study. Cancer 2005;104:1975-84.

28. Veronesi G, Morandi U, Alloisio M, et al. Large cell neuroendocrine carcinoma of the lung: a retrospective analysis of 144 surgical cases. Lung Cancer 2006;53:111-5.

29. Saji H, Tsuboi M, Matsubayashi J, et al. Clinical response of large cell neuroendocrine carcinoma of the lung to perioperative adjuvant chemotherapy. Anticancer Drugs 2010;21:89-93.

30. Iyoda A, Hiroshima K, Moriya Y, et al. Prospective study of adjuvant chemotherapy for pulmonary large cell neuroendocrine carcinoma. Ann Thorac Surg 2006;82:1802-7.

31. Tiseo M, Bartolotti M, Gelsomino F, et al. First-line treatment in advanced non-small-cell lung cancer: the emerging role of the histologic subtype. Expert Rev Anticancer Ther 2009;9:425-35.

32. Hanagiri T, Oka S, Takenaka S, et al. Results of surgical resection for patients with large cell carcinoma of the lung. Int J Surg 2010;8:391-4.

R program used in this study

```
# for Cox regression
cox1<-cph (Surv (survival months, status) ~var list,
x=T, y=T, data=mydata, method=("breslow"), surv=T, time.inc=36 or 60)
cox1.step <- step (cox1, direction=("both"))
cox1.step=step(cox1)

# for plotting the nomogram
plot (nomogram (cox1.step, fun=list (surv2, surv3), lp=F,
funlabel=c ('3-Year LCSS','5-Year LCSS'),
maxscale=100, fun.at=c (0.95,0.85,0.80,0.70,0.6,0.5,0.4,0.3,0.2,0.1)),
xfrac=.3)

# for plotting calibration curve
calnv<- calibrate (cox1.step, cmethod='KM', method='boot', u=30 or 60, m=500, B=1000)
plot(calnv)

# for decision curve
DC1 <- decision_curve (status~var list,
data = mydata,
study.design = "case-control",
policy = "opt-in",
bootstraps = 500,
confidence.intervals =0.95,
population.prevalence= 0.3))
plot_decision_curve(DC1, curve.names = "nomogram model")
```