



# Clinicopathological characteristics and prediction of cancer-specific survival in large cell lung cancer: a population-based study

Yafei Shi<sup>#</sup>, Wei Chen<sup>#</sup>, Chunyu Li, Shuya Qi, Xiaowei Zhou, Yujun Zhang, Ying Li, Guohui Li

Department of Pharmacy, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100021, China

*Contributions:* (I) Conception and design: Y Shi, W Chen; (II) Administrative support: G Li; (III) Provision of study materials or patients: C Li, S Qi, X Zhou; (IV) Collection and assembly of data: Y Zhang, Y Li; (V) Data analysis and interpretation: Y Shi, W Chen; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

<sup>#</sup>These authors contributed equally to this work.

*Correspondence to:* Guohui Li. Department of Pharmacy, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100021, China. Email: lgh0603@cicams.ac.cn.

**Background:** To describe the demographic and clinical characteristics of large cell lung cancer (LCLC) with a population-based database and to find the prognosis factors of cancer-specific survival (CSS) for these patients; also, to develop a nomogram to independently validate and predict the CSS for LCLC based on the identified prognosis factors.

**Methods:** We extracted the LCLC patient's information from the Surveillance, Epidemiology, and End Results (SEER) database [2005–2014] and summarized the characteristics of the extracted factors. We used Cox proportional hazards regression to find the prognosis factors for LCLC patients and to develop the nomogram based on these in a split train cohort from the extracted data. The validation of the developed nomograms was performed in an independent validation cohort from the extracted data, in which the C-index and the average of the time-dependent area under the receiver operating characteristic curve (time-dependent AUC) for CSS in 1-year, 3-year, and 5-year CSS was calculated. The calibration curves were drawn to visualize the performance of the established nomogram.

**Results:** As a result, 4,936 patients with LCLC were identified from the SEER database. Nearly half of LCLC patients were diagnosed with stage IV; only approximately 20% of patients underwent surgery. The prognosis factors that influenced the LCLC patients included age, sex, American Joint Committee on Cancer (AJCC) stage, race, surgery, tumor size, and marital status. The calculated C-index was  $0.701 \pm 0.01$ , and the mean time-dependent AUC for in 1-year, 3-year, and 5-year CSS was 0.88. The calibrated curve showed that the gap between the predicted and observed values for 1-year, 3-year, and 5-year CSS was small.

**Conclusions:** Sex, age, race, marital status, AJCC stage, surgery, and tumor size were shown to all be the independent prognostic factors of CSS in LCLC. The established nomogram can provide more precise evaluation for the survival of LCLC patients and help the clinicians in the individual management of patients.

**Keywords:** Large cell lung cancer (LCLC); characteristics; cancer-specific survival (CSS); nomogram

Submitted Dec 31, 2019. Accepted for publication Mar 20, 2020.

doi: 10.21037/jtd.2020.04.24

View this article at: <http://dx.doi.org/10.21037/jtd.2020.04.24>

## Introduction

Lung cancer is the leading cause of cancer death worldwide (1,2). Large cell lung cancer (LCLC) constitutes a small proportion of lung cancer incidence, accounting for just 9% of all cases of lung cancer (3), and is a sub-type of non-small cell lung carcinoma (NSCLC) without glandular or squamous differentiation. The diagnosis of LCLC is often excluded from adenocarcinoma, squamous, and small cell lung cancer (3). LCLC can be found in any part of the lung, with men being more susceptible (4). The growth and spread of LCLC is quick, which makes it harder to treat. The features and survival outcomes of LCLC are scarcely reported due to its low incidence. As a result, the prognoses of LCLC is still unclear. In general, patients with a solid tumor can be classified by the American Joint Committee on Cancer (AJCC) staging system (5). However, the AJCC classification does not predict the survival of patients with rare or special cancer types (6).

As a kind of decision-making tool for patients with cancer, a nomogram can predict patient survival (7), and nomograms have also been widely used to stratify the treatment and evaluate outcomes (8). To date, however, reports on the use of a nomogram for patients with LCLC are still unavailable.

In this study, we examined the characteristics and prognosis of LCLC by using a United States population-based database. Also, we developed and independently validated a nomogram model based on the selected data to predict the prognosis of LCLC patients visually.

## Methods

### *Ethics statement*

We obtained permission to access the Surveillance, Epidemiology, and End Results (SEER) database with the reference number 10782-Nov2016. The informed consent for patients was not required in this study, as the research data are de-identified and publicly available.

### *Study population*

We obtained data from the November 2013 submission of the SEER Research Data.

In this study, we extracted patients' data from 2005 to 2014 in the SEER database by using SEER\*Stat software version 8.3.5. As an authoritative source of cancer information, the SEER database contains U.S. cancer

incidence and survival data, and there are 18 population-based registries included in the database, representing about 28% of the American population (9,10). In our study, patients diagnosed with large-cell carcinoma of the lung, according to the International Classification of Diseases for Oncology, Third Edition (ICD-O-3), (ICD-O-3:8012/3) were included. Patients with incomplete survival data were excluded. Since the recent version of the World Health Organization (WHO) classification for lung tumors used in the SEER was updated in 2004, we also excluded patients diagnosed before 2005.

### *Covariates*

We collected patient demographic characteristics including age at diagnosis, race, sex, and marital status, and tumor clinicopathological features such as primary site, laterality, grade, AJCC stage (6th). Surgery information was also included. We used the cancer-specific survival (CSS) as the primary outcome, which was defined as a time interval from the diagnosis to death due to this cancer. Data of individuals who died and whose death was attributed to other causes or who were alive on the cutoff date (December, 2017) were censored.

### *Statistical analysis*

Categorical variables for the demographic and clinical characteristics were reported as numbers and percentages. Continuous variables were expressed as the mean  $\pm$  standard deviation (11). The random sampling strategy was used to split the primary cohort into the training cohort and validation cohort. In the training cohort, the optimal cutoff points for continuous variables were determined by using the maximally selected rank statistics in advance of the Cox proportional hazards regression (12,13). The CSS curve was plotted by the Kaplan-Meier method, log-rank test was used for the comparisons between CSS distributions, while proportional hazards regression analysis was used to evaluate the factors influencing CSS and to compute the hazard ratios and the 95% confidence intervals (95% CIs). A P value  $<0.05$  (two-sided) was considered statistically significant. A nomogram was built to predict the probability of 1-year, 3-year, and 5-year CSS based on the final model of Cox proportional hazards regression. In the validation cohort, the total scores of the established nomogram for each patient were calculated. The performance of the scores was assessed by calculating the C-index and the average

of the time-dependent area under the receiver operating characteristic curve (time-dependent AUC) in 1-year, 3-year, and 5-year, CSS. The calibration curves were also drawn to visualize the performance based on the total scores. All the statistical analyses were performed using R version 3.2.5 software (12).

## Results

### Summary of characteristics

A total of 4,963 of LCLC patients from the SEER database were identified; the median CSS time was 6 months, during which time all the patients were included in the study.

The demographic and tumor clinicopathological features for the eligible patients are summarized in *Table 1*, which shows that the mean age of the LCLC patients was  $66.9 \pm 11.2$ . The male/female ratio was 1.37, and most of the recorded patients were White (79.27%). The main primary site labeled was up lobe (52.63%), and laterality was recorded more in the right-origin of primary (56.48%). The mean tumor size was  $49.96 \pm 33.61$ , most of the diagnosis patients were in stage IV (49.77%), and 20.69% of the patients underwent surgery. Based on the random sampling

strategy, all the patients were divided into a training cohort (n=342, 69.05%) and a validation cohort (n=1,536, 30.95%). The training and validation cohort were also listed in *Table 1*. All the characteristics were similar in the training cohort compared with the validation cohort. There were no observed significant CSS differences between the 2 cohorts (P=0.8).

### Development of the nomogram

Based on the maximally selected rank statistics, patients were classified into 2 group in terms of age ( $\leq 77$  years,  $> 77$  years) and tumor size ( $\leq 41$  mm,  $> 41$  mm) in the training cohort. In the univariate analysis, age, gender, race, primary site, laterality, AJCC stage, surgery, and marital status were found to be significantly correlated with the CSS in the training cohort (*Table 2*). The potential redundancy was removed according to the AIC-base backward selection procedure in the multivariate Cox proportional hazards regression analysis. The finally recruited independent prognostic factors, including age, sex, race, AJCC stage, surgery, and marital status, were used to construct the nomogram model. The hazard ratios (95% confidence interval, 95% CI) of nomogram parameters are shown in

**Table 1** Demographic and tumor clinicopathological features for the eligible patients

Variable	No. of patients (total =4,963) (%)	Training (n=3,427) (%)	Validation (n=1,536) (%)
Age (mean $\pm$ standard deviation)	66.9 $\pm$ 11.2	66.87 $\pm$ 11.21	66.98 $\pm$ 11.20
Gender			
Male	2,871 (57.85)	2,001 (58.39)	870 (56.64)
Female	2,092 (42.15)	1,426 (41.61)	666 (43.36)
Race			
White	3,934 (79.27)	2,723 (79.46)	1,211 (78.84)
Black	763 (15.37)	517 (15.09)	246 (16.02)
Others	257 (5.18)	182 (5.31)	75 (4.88)
NA	9 (0.18)	5 (0.15)	4 (0.26)
Primary site labeled			
Main bronchus	227 (4.57)	145 (4.23)	82 (5.34)
Upper lobe of the lung	2,612 (52.63)	1,794 (52.35)	818 (53.26)
Middle lobe of the lung	217 (4.37)	151 (4.41)	66 (4.30)
Lower lobe of the lung	1,122 (22.61)	776 (22.64)	346 (22.53)
Overlapping lesion of the lung	58 (1.17)	44 (1.28)	14 (0.91)
Lung NOS	727 (14.65)	517 (15.09)	210 (13.67)

**Table 1** (continued)

Table 1 (continued)

Variable	No. of patients (total =4,963) (%)	Training (n=3,427) (%)	Validation (n=1,536) (%)
Grade			
Grade II	50 (1.01)	38 (1.11)	12 (0.78)
Grade III	1,468 (29.58)	993 (28.98)	475 (30.92)
Grade IV	1,460 (29.42)	1,006 (29.36)	454 (29.56)
Grade I	7 (0.14)	6 (0.18)	1 (0.07)
NA	1,978 (39.85)	1,384 (40.39)	594 (38.67)
Laterality			
Bilateral, single primary	68 (1.37)	53 (1.55)	15 (0.98)
Left-origin of primary	1,900 (38.28)	1,308 (38.17)	592 (38.54)
Only one side-side unspecified	37 (0.75)	26 (0.76)	11 (0.72)
The paired site, but no information concerning laterality	155 (3.12)	105 (3.06)	50 (3.26)
Right-the origin of primary	2,803 (56.48)	1,935 (56.46)	868 (56.51)
AJCC stage			
IA	274 (5.52)	183 (5.34)	91 (5.92)
IB	428 (8.62)	294 (8.58)	134 (8.72)
IIA	35 (0.71)	24 (0.70)	11 (0.72)
IIB	203 (4.09)	139 (4.06)	64 (4.17)
IIIA	468 (9.43)	321 (9.37)	147 (9.57)
IIIB	817 (16.46)	573 (16.72)	244 (15.89)
IV	2,470 (49.77)	1,702 (49.66)	768 (50)
NA	268 (5.40)	191 (5.57)	77 (5.01)
Surgery			
Not performed	3,905 (78.68)	2,689 (78.47)	1,216 (79.17)
Surgery performed	1,027 (20.69)	713 (20.81)	314 (20.44)
NA	31 (0.62)	25 (0.73)	6 (0.39)
Marital status			
Divorced	636 (12.81)	446 (13.01)	190 (12.37)
Married (including common law)	2,555 (51.48)	1,771 (51.68)	784 (51.04)
Separated	53 (1.07)	40 (1.17)	13 (0.85)
Single (never married)	699 (14.08)	473 (13.80)	226 (14.71)
Unmarried or domestic partner	2 (0.04)	1 (0.03)	1 (0.07)
Widowed	873 (17.59)	601 (17.54)	272 (17.71)
NA	145 (2.92)	95 (2.77)	50 (3.26)
Tumour size (mean ± standard deviation)	49.96±33.62	49.92±34.22	50.06±32.26

NA, not applicable; NOS, not otherwise specified; AJCC, American Joint Committee on Cancer.

**Table 2** Univariate analysis results of the training cohort

Factor	HR (95% CI)	P
Age		
≤77 years	1	
>77 years	1.59 (1.45–1.74)	<0.05
Gender		
Female	1.15 (1.06–1.23)	<0.05
Male	1.15 (1.06–1.23)	
Race		
Black	1	
White	0.96 (0.87–1.06)	0.4066
Others	0.74 (0.61–0.89)	<0.05
NA		
Primary site		
Main bronchus	1	
Upper lobe of lung	0.73 (0.61–0.87)	<0.05
Middle lobe of lung	0.73 (0.57–0.93)	<0.05
Lower lobe of lung	0.84 (0.70–1.01)	0.0682
Overlapping lesion of the lung	0.62 (0.42–0.90)	0.0134
Lung, NOS	1.28 (1.05–1.55)	<0.05
Grade		
Grade I	1	
Grade II	0.74 (0.29–1.90)	0.54
Grade III	0.80 (0.33–1.92)	0.61
Grade IV	0.79 (0.33–1.91)	0.60
NA		
Laterality		
Bilateral, single primary	1	
Left-origin of primary	0.48 (0.36–0.64)	<0.05
Only one side-side unspecified	0.63 (0.38–1.03)	0.067
Paired site, but no information concerning laterality	0.89 (0.6–1.25)	0.502
Right-origin of primary	0.48 (0.36–0.63)	<0.05
AJCC stage		
IA		
IB	1.27 (0.98–1.6)	0.074
IIA	1.35 (0.78–2.34)	0.279
IIB	1.85 (1.38–2.48)	<0.05
IIIA	2.44 (1.91–3.11)	<0.05
IIIB	3.93 (3.13–4.94)	<0.05
IV	6.34 (5.11–7.89)	<0.05
NA		

**Table 2** (continued)**Table 2** (continued)

Factor	HR (95% CI)	P
Surgery		
Not performed	1	
Surgery performed	0.25 (0.22–0.29)	<0.05
NA		
Marital status		
Divorced	0.94 (0.85–1.05)	1
Married (including common law)	0.94 (0.85–1.05)	0.34
Separated	9.25 (0.64–1.33)	0.68
Single (never married)	1.10 (0.96–1.27)	0.17
Unmarried or domestic partner	0.00 (0.00–777.87)	0.97
Widowed	1.19 (1.04–1.36)	<0.05
NA		
Tumor size		
≤41 mm	1	
>41 mm	1.46 (1.34–1.59)	<0.05

HR, hazard ratio; NA, not applicable; NOS, not otherwise specified; AJCC, American Joint Committee on Cancer.

*Table 3.* The detailed scores of each independent prognostic factors are also listed in *Table 3*, and the nomogram is plotted in *Figure 1*. Patients were classified into 7 groups according to the nomogram scores. The CSS curve for these groups is shown in *Figure 2*.

### Validation of the nomogram

The C-index of the nomograms for predicting CSS was  $0.701 \pm 0.01$  in the validation cohort. The average of the time-dependent AUC in 1-year, 3-year, and 5-year CSS was 0.88. The calibration curve (*Figure 3*) for 1-year, 3-year, and 5-year CSS showed little gap between the predictions and actual outcomes in the validation cohort.

### Discussion

In the present study, 4,936 LCLC patients were identified from the SEER database.

Nearly half of LCLC patients were diagnosed with stage IV, and only 20% of patients underwent surgery. This means that most of the LCLC cancer patients were diagnosed with advance stage and the optimal treatment opportunity was often missed. Therefore, an efficient diagnosis and treatment method are urgently needed.

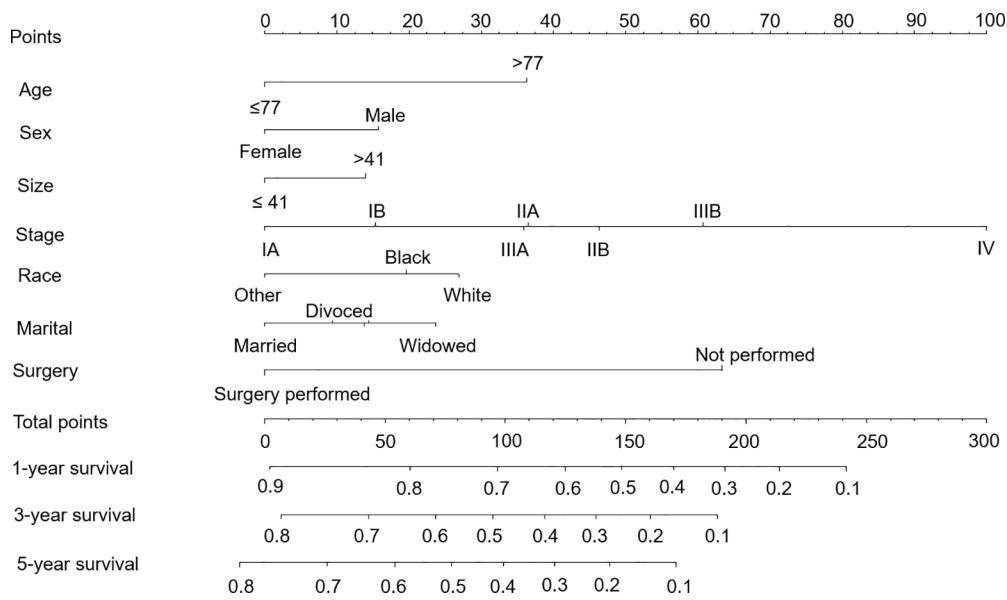
**Table 3** Hazard ratio (95% CI) of nomogram parameters and nomogram scores

Factor	HR (95% CI)	P	Nomogram scores
Age			0
≤77 years	1		0
>77 years	1.60 (1.42–1.80)	<0.05	36
Gender			
Female	1.23 (1.11–1.35)		0
Male	1.23 (1.11–1.35)	<0.05	16
Race			
Black	1		
White	1.10 (0.97–1.25)	0.149	27
Others	0.78 (0.61–0.99)	<0.05	0
AJCC Stage			
IA	1		0
IB	1.22 (0.93–1.61)	0.155	15
IIA	1.60 (0.92–2.79)	0.09	37
IIB	1.82 (1.34–2.48)	<0.05	46
IIIA	1.59 (1.21–2.09)	<0.05	36
IIIB	2.19 (1.68–2.85)	<0.05	61
IV	3.63 (2.83–4.66)	<0.05	100
NA			
Surgery			
Not performed	1		0
Surgery performed	0.44 (0.38–0.51)	<0.05	63
NA			
Marital status			
Divorced	1		9
Married (including common law)	0.89 (0.78–1.01)	0.08	0
Separated	1.07 (0.71–1.61)	0.75	14
Single (never married)	1.06 (0.89–1.26)	0.51	14
Unmarried or domestic partner	NA		–
Widowed	1.20 (1.02–1.41)	0.02	24
NA			
Tumor size			
≤41 mm	1		0
>41 mm	1.20 (1.09–1.31)	<0.05	14

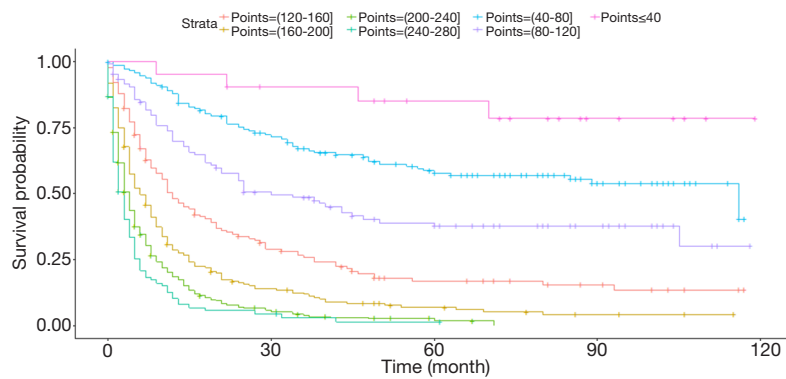
HR, hazard ratio; 95% CI, 95% confidence interval; NA, not applicable; AJCC, American Joint Committee on Cancer.

In this study, the routinely available characteristics of the patients were extracted from the SEER database. Based on these, a nomogram was developed and confirmed (14). The model performance of the developed nomogram was confirmed by the calculated C-index, time-dependent AUC, and calibrated curves. Seven significant factors (age, sex, tumor size, AJCC stage, surgery, race, and marital status) were included in the nomogram. All the factors are routinely available in daily practice, which allows for the nomogram to be easily used in predicting an individual's CSS and making treatment decisions for patients and clinicians.

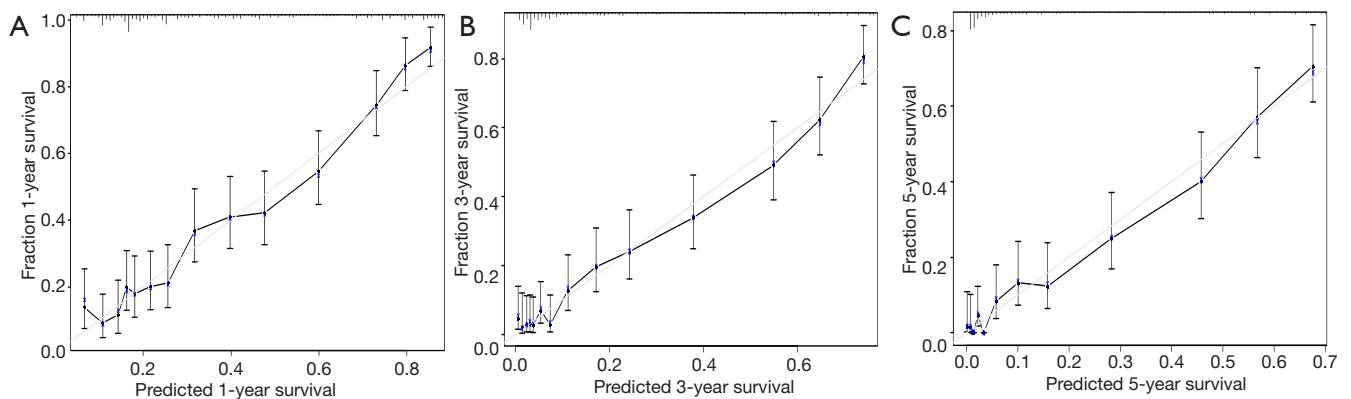
In agreement with other types of NSCLC, the age and sex were all crucial predictors for CSS of LCLC (15). Old age and male patients were associated with a worse prognosis. According to Harrell's guidelines, in this study, patients' age was divided into 2 groups (16), and 77 years old was the best cutoff point. At present, a consistent conclusion has not been reached for the CSS disparities of patients with lung cancer across different races (17). In general, better CSS outcomes would be seen in a race with higher health awareness, and the treatment in these races would be more active. In this study, the significantly different influence of race for LCLC patient's CSS was seen both in univariable and multivariable analysis. The CSS outcome for other races was better than for the White race. Treatment activity might be a possible explanation for this phenomenon, and another reason might be the sample of the SEER database: small proportions of other races were collected in the database, which might have affected the statistical results for this study. As the most commonly used tumor-associated indices, the AJCC stage still contributed most to the established nomogram model, which is in line with other types of NSCLC (18). Tumor size is an essential indicator for the T stage; in this study, it was also found to be an independent risk factor for LCLC. It was confirmed that patients with a tumor size >41 mm show less CSS time than those with a tumor size ≤41 mm using the maximally selected rank statistics. Surgery is the primary treatment for most types of lung cancer. In this study, surgery was also found to be an essential treatment for LCLC in that patients with surgery had a significant decrease in cancer-specific death. Marital status has been confirmed to be associated with the CSS in a series of cancer (19–23). This phenomenon, in our present study, was consistent with the previous study in which married LCLC patients represented by lower nomogram scores had more CSS benefit compared with other types of marital statuses.



**Figure 1** Nomogram for predicting the 1-year, 3-year, and 5-year CSS. CSS, cancer-specific survival.



**Figure 2** The CSS curve for re-grouped patients according to the nomogram scores. CSS, cancer-specific survival.



**Figure 3** Calibration curves for CSS using the validation cohort. CSS, cancer-specific survival.

The nomogram validation is of great importance in the prevention of overfitting for the established model, and also crucial for model generalization (24). In the present study, an independent validation cohort was used for the validation of the nomogram. The calculated C-index ( $0.701 \pm 0.01$ ) confirmed the discriminatory capacity of the established nomogram. Optimal consistency between the predicted CSS and actual observed CSS was seen in the calibration curves for 1-year, 3-year, and 5-year CSS. The time-dependent ROC for the validation cohort was acceptable enough to keep the AUC at a comparatively higher level ( $>0.5$ ) in the prediction of CSS.

Still, some limitations should be considered in the current study. Firstly, since many vital factors identified in previous studies that influence the CSS of lung cancer in SEER database are lacking, many critical factors such as the chemotherapy, radiotherapy, smoking status, and performance status were not obtained in this study, and thus a more detailed understanding of the prognosis factors for LCLC could not be obtained. Additionally, the lack of these factors may affect the accuracy of the nomogram model. Thirdly, as a retrospective study, the selection bias could not be excluded, and thus a validation with a prospective clinical study is still needed.

## Conclusions

This study summarized the clinical features and prognosis of LCLC by using a U.S. population-based cohort from the SEER database, and we found that sex, age, race, marital status, AJCC stage, surgery, and tumor size are all the independent prognostic factors of CSS for LCLC. Furthermore, a visual nomogram was developed to predict the CSS of LCLC. The discrimination of the nomogram was confirmed to be acceptable in the validation cohort. This nomogram can provide a more precise evaluation of CSS for LCLC patients and help clinicians to make individual management decisions.

## Acknowledgments

*Funding:* This work was supported by the CAMS Innovation Fund for Medical Sciences (CIFMS) (Grants Nos. 2016-I2M-1-001, 2017-I2M-1-005, and 2017-I2M-1-003).

## Footnote

*Conflicts of Interest:* All authors have completed the ICMJE

uniform disclosure form (available at <http://dx.doi.org/10.21037/jtd.2020.04.24>). Dr. CL reports grants from CAMS Innovation Fund for Medical Sciences (Grants Nos. 2016-I2M-1-001, 2017-I2M-1-005, and 2017-I2M-1-003), during the conduct of the study; Dr. GL reports grants from CAMS Innovation Fund for Medical Sciences (Grants Nos. 2016-I2M-1-001, 2017-I2M-1-005, and 2017-I2M-1-003), during the conduct of the study; Dr. SQ reports grants from CAMS Innovation Fund for Medical Sciences (Grants Nos. 2016-I2M-1-001, 2017-I2M-1-005, and 2017-I2M-1-003), during the conduct of the study; Dr. WC reports grants from CAMS Innovation Fund for Medical Sciences (Grants Nos. 2016-I2M-1-001, 2017-I2M-1-005, and 2017-I2M-1-003), during the conduct of the study; Dr. YS reports grants from CAMS Innovation Fund for Medical Sciences (Grants Nos. 2016-I2M-1-001, 2017-I2M-1-005, and 2017-I2M-1-003), during the conduct of the study; Dr. YL reports grants from CAMS Innovation Fund for Medical Sciences (Grants Nos. 2016-I2M-1-001, 2017-I2M-1-005, and 2017-I2M-1-003), during the conduct of the study; Dr. YZ reports grants from CAMS Innovation Fund for Medical Sciences (Grants Nos. 2016-I2M-1-001, 2017-I2M-1-005, and 2017-I2M-1-003), during the conduct of the study; Dr. XZ reports grants from CAMS Innovation Fund for Medical Sciences (Grants Nos. 2016-I2M-1-001, 2017-I2M-1-005, and 2017-I2M-1-003), during the conduct of the study.

*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.

*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. Wei J, Yan Y, Chen X, et al. The Roles of Plant-Derived Triptolide on Non-Small Cell Lung Cancer. *Oncol Res* 2019;27:849-58.
2. Yan Y, Xu Z, Qian L, et al. Identification of CAV1



- and DCN as potential predictive biomarkers for lung adenocarcinoma. *Am J Physiol Lung Cell Mol Physiol* 2019;316:L630-43.
3. Sholl LM. Large-cell carcinoma of the lung: a diagnostic category redefined by immunohistochemistry and genomics. *Curr Opin Pulm Med* 2014;20:324-31.
  4. Monica V, Ceppi P, Righi L, et al. Desmocollin-3: a new marker of squamous differentiation in undifferentiated large-cell carcinoma of the lung. *Mod Pathol* 2009;22:709-17.
  5. Moitra D, Mandal RK. Automated AJCC (7th edition) staging of non-small cell lung cancer (NSCLC) using deep convolutional neural network (CNN) and recurrent neural network (RNN). *Health Inf Sci Syst* 2019;7:14.
  6. Liang W, Zhang L, Jiang G, et al. Development and validation of a nomogram for predicting survival in patients with resected non-small-cell lung cancer. *J Clin Oncol* 2015;33:861-9.
  7. Zhou H, Shen J, Zhang Y, et al. Risk of second primary malignancy after non-small cell lung cancer: a competing risk nomogram based on the SEER database. *Ann Transl Med* 2019;7:439.
  8. Asare EA, Liu L, Hess KR, et al. Development of a model to predict breast cancer survival using data from the National Cancer Data Base. *Surgery* 2016;159:495-502.
  9. Kuo TM, Mobley LR. How generalizable are the SEER registries to the cancer populations of the USA? *Cancer Causes Control* 2016;27:1117-26.
  10. Xie JC, Yang S, Liu XY, et al. Effect of marital status on survival in glioblastoma multiforme by demographics, education, economic factors, and insurance status. *Cancer Med* 2018;7:3722-42.
  11. Wang X, Xu Z, Chen X, et al. A tropomyosin receptor kinase family protein, NTRK2 is a potential predictive biomarker for lung adenocarcinoma. *PeerJ* 2019;7:e7125.
  12. Liu Y, Zhai X, Li J, et al. Is there an optimal time to initiate adjuvant chemotherapy to predict benefit of survival in non-small cell lung cancer? *Chin J Cancer Res* 2017;29:263-71.
  13. Wei S, Tian J, Song X, et al. Causes of death and competing risk analysis of the associated factors for non-small cell lung cancer using the Surveillance, Epidemiology, and End Results database. *J Cancer Res Clin Oncol* 2018;144:145-55.
  14. Wang Y, Yang Y, Chen Z, et al. Development and validation of a novel nomogram for predicting distant metastasis-free survival among breast cancer patients. *Ann Transl Med* 2019;7:537.
  15. Mao Q, Xia W, Dong G, et al. A nomogram to predict the survival of stage IIIA-N2 non-small cell lung cancer after surgery. *J Thorac Cardiovasc Surg* 2018;155:1784-92.e3.
  16. Harrell FE Jr, Califf RM, Pryor DB, et al. Evaluating the yield of medical tests. *Jama* 1982;247:2543-6.
  17. Varlotto JM, McKie K, Volland RP, et al. The Role of Race and Economic Characteristics in the Presentation and Survival of Patients With Surgically Resected Non-Small Cell Lung Cancer. *Front Oncol* 2018;8:146.
  18. Xiao HF, Zhang BH, Liao XZ, et al. Development and validation of two prognostic nomograms for predicting survival in patients with non-small cell and small cell lung cancer. *Oncotarget* 2017;8:64303-16.
  19. 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.
  20. Feng Y, Dai W, Li Y, et al. The effect of marital status by age on patients with colorectal cancer over the past decades: a SEER-based analysis. *Int J Colorectal Dis* 2018;33:1001-10.
  21. Li Z, Wang K, Zhang X, et al. Marital status and survival in patients with rectal cancer: A population-based STROBE cohort study. *Medicine (Baltimore)* 2018;97:e0637.
  22. Huang TB, Zhou GC, Dong CP, et al. Marital status independently predicts prostate cancer survival in men who underwent radical prostatectomy: An analysis of 95,846 individuals. *Oncol Lett* 2018;15:4737-44.
  23. Wang H, Wang L, Kabirov I, et al. Impact of marital status on renal cancer patient survival. *Oncotarget* 2017;8:70204-13.
  24. Iasonos A, Schrag D, Raj GV, et al. How to build and interpret a nomogram for cancer prognosis. *J Clin Oncol* 2008;26:1364-70.

**Cite this article as:** Shi Y, Chen W, Li C, Qi S, Zhou X, Zhang Y, Li Y, Li G. Clinicopathological characteristics and prediction of cancer-specific survival in large cell lung cancer: a population-based study. *J Thorac Dis* 2020;12(5):2261-2269. doi: 10.21037/jtd.2020.04.24