

Development and validation of prognostic nomogram for lung cancer patients below the age of 45 years

Lili Dai^{1#}, Wei Wang^{2#}, Qi Liu^{3#}, Tongjia Xia⁴, Qikui Wang⁵, Qingqing Chen⁶, Ning Zhu⁷, Yu Cheng⁸, Ying Yan⁹, Jun Shu^{10*}, Kaixin Qu^{2*}

ABSTRACT

This study aimed to establish a nomogram for the prognostic prediction of patients with early-onset lung cancer (EOLC) in both overall survival (OS) and cancer-specific survival (CSS). EOLC patients diagnosed between 2004 and 2015 were retrieved from the Surveillance, Epidemiology, and End Results (SEER) database and further divided into training and validation sets randomly. The prognostic nomogram for predicting 3-, 5- and 10-years OS and CSS was established based on the relative clinical variables determined by the multivariate Cox analysis results. Furthermore, the predictive performance of nomogram was assessed by concordance index (C-index), calibration curve, receiver operating characteristic (ROC) curve and decision curve analysis (DCA) curve. A total of 1,822 EOLC patients were selected and randomized into a training cohort (1,275, 70%) and a validation cohort (547, 30%). The nomograms were established based on the statistical results of Cox analysis. In training set, the C-indexes for OS and CSS prediction were 0.797 (95% confidence interval [CI]: 0.773-0.818) and 0.794 (95%CI:0.771-0.816). Significant agreement in the calibration curves was noticed in the nomogram models. The results of ROC and DCA indicated nomograms possessed better predict performance compared with TNM-stage and SEER-stage. And the area under the curve (AUC) of the nomogram for OS and CSS prediction in ROC analysis were 0.766 (95%CI:0.745-0.787) and 0.782 (95%CI:0.760-0.804) respectively. The prognostic nomogram provided an accurate prediction of 3-, 5-, and 10-year OS and CSS of EOLC patients which contributed clinicians to optimize individualized treatment plans.

KEYWORDS: Early-onset lung cancer; prognostic nomogram; overall survival; cancer-specific survival; SEER

INTRODUCTION

Cancer is a leading cause of morbidity and mortality worldwide, and among the various types of cancers, lung

cancer (LC) has one of the highest incidences of fatality. LC accounts for nearly 27% of cancer deaths in the United States and 20% in the European Union [1]. However, it is encouraging to note that the 5-year survival rate for patients with LC in the United States has increased from 17.2% in 2009 to 21.7% in 2019 [2]. This progress may be attributable to the combination of personalized treatment, screening of high-risk groups, and early diagnosis. It was found that people aged <40 years had a low incidence of LC, which increased yearly to include people aged 75 to 80 years [1]. In clinical trials today, early-onset LC (EOLC) defines LC in patients aged <45 years. These patients comprise approximately 5% of all patients with LC [3]. Unlike elderly patients, genetic cancer factors are considered as the mainstream cause of EOLC [4], which intensifies the need for the accurate prognosis and individualized treatment.

Presently, the tumor-node-metastasis (TNM) staging system, developed by the Union for International Cancer Control and used by the American Joint Commission on Cancer, is widely accepted as the criterion to predict the prognosis of patients with various cancers involving tumor invasion (T), regional lymph nodes (N), and distant metastasis (M) [5]. Since the popularization of TNM staging in the 1970s, major revisions have been made to *TNM Classification of Malignant Tumours. 8th Edition*, which publishes the latest, internationally agreed-on standards to describe and categorize cancer stage [7]. However, the prognostic assessment based on the TNM staging system

¹Department of Medicine, Funan County People's Hospital, Anhui, China

²Department of Respiratory Medicine, Funan County People's Hospital, Anhui, China

³Department of Endocrinology, Punan Hospital of Pudong District, Shanghai, China

⁴Department of Endocrinology, The First Affiliated Hospital of Anhui Medical University, Anhui, China

⁵Department of Chest Surgery, Anhui Chest Hospital, Anhui, China

⁶Department of Tuberculosis, Anhui Chest Hospital, Anhui, China

⁷Department of Respiratory Medicine, The Second Affiliated Hospital of Xuzhou Medical University, Jiangsu China

⁸Department of Interventional Pulmonary and Endoscopy Center, Anhui Chest Hospital, Anhui, China

⁹Department of Oncology, Anhui Cancer Hospital, Anhui, China

¹⁰Department of Respiratory Medicine, The Fourth Affiliated Hospital of Anhui Medical University, Anhui, China

*Corresponding authors: Kaixin Qu, Department of Respiratory Medicine, Funan County People's Hospital, No. 81 Santa Road, Funan County, Fuyang, 236300, China.
E-mail: qukaixin19770418@163.com

Jun Shu, Department of Respiratory Medicine, The Fourth Affiliated Hospital of Anhui Medical University, No.372 Tunxi Road, Baohe District, Hefei, 230000, China.
E-mail: shujun20170901@163.com

#These authors equally contributed

DOI: <https://dx.doi.org/10.17305/bjbms.2020.5079>

Submitted: 23 August 2020/Accepted: 11 October 2020

Conflict of interest statement: The authors declare no conflict of interests



©The Author(s) (2021). This work is licensed under a Creative Commons Attribution 4.0 International License

has limitations and is deficient in predicting prognosis accurately.

The nomogram has been acceptance in the last decade as a unique, reliable method for predicting tumor prognosis [7]. It has been applied in the prognosis prediction of many cancers including gastric cancer, breast cancer, testicular cancer, and so on [8-11]. As a prognostic model, the nomogram assesses significant related risk factors for the prediction. Specifically, the nomogram can produce accurate predictions for overall survival (OS) and cancer-specific survival (CSS) in patients, due to the multiple clinical variables in the calculation. In this study, we utilized nomograms to predict 3-, 5-, and 10-year OS and CSS in patients with EOLC.

MATERIALS AND METHODS

Data source and patients

Clinicopathological data and individualized prognostic outcomes in patients with EOLC between 2004 and 2015 were obtained from the Surveillance, Epidemiology, and End Results (SEER) database of the National Cancer Institute using SEER*Stat software (version 8.3.5; SEER 18 Regs Custom Data [with additional treatment fields], November 2018 Sub [1975-2016 varying] database). The identification of EOLC patients was based on the exclusion criteria as follows: (I) patients age >45 years old; (II) patients with multiple primaries tumor; (III) the unknown American Joint Committee on Cancer (AJCC) stage; (IV) the unknown TNM stage; (V) patients without surgery. All the eligible EOLC patients included in this study were randomly assigned into the training and validation sets. Local ethics approval or statements were not required because the clinical data used in this study were obtained from the public-access SEER database and thus, the requirement for informed consent was waived.

Study variables

Clinical variables included in this study contained gender, age, race, grade, TNM stage (AJCC, 7th ed.), tumor primary site, SEER stage, chemotherapy and radiotherapy. The age of eligible EOLC patients was divided into three groups (<35, 35-43 and >43; Fig. S1) according to the optimal cut-off value calculated by X-tile software version 3.6.1 (Yale University School of Medicine, US). The tumor primary site contained the following six sites: main bronchus (C34.0), upper lobe (C34.1), middle lobe (C34.2), lower lobe (C34.3), overlapping lesion of lung (C34.8) and not otherwise specified (NOS; C34.9). Moreover, SEER stage comprises three categories: localized, regional, and distant. OS is defined as the time from diagnosis to any cause leading to death or to the date on which data were censored. Moreover, the CSS time analyzed in this study was the survival time from diagnosis

to death associated with cancer, excluding other causes. The cut-off point in this study was December 31, 2016.

Construction and validation of nomogram

Kaplan-Meier curve and log-rank test were performed to investigated the OS and CSS of EOLC patients. Univariate and multivariate regression analyses were used to evaluate the prognostic factors in patients with EOLC. The Cox proportional hazards model was used as the basis for the construction and verification of nomograms. R software version 3.5.1 (<http://www.R-project.org>) was performed for establishing nomograms. Concordance index (C-index) and calibration curve were performed to evaluate the performance and accuracy of nomograms. The C-index value ranges from 0.50 to 1.00 and shows a positive correlation with the predicted performance of the model. It indicates that and the models accompanied by perfect discrimination ability when the value is 1.00. Moreover, when the calibration curve is applied to a perfectly calibrated model, the prediction will fall on the diagonal 45° in the figure.

In addition, receiver operating characteristic (ROC) curves and decision curve analysis (DCA) were conducted to assess the predicted performance of nomograms, TNM stage, and SEER stage. The statistical software package for social science software (version 20.0; SPSS, Chicago, USA) was applied for all statistical analyses. The results were considered statistically significant as P-value < 0.05 (two-sided).

RESULTS

Demographic and pathologic characteristics

The flow process diagram for retrieving patients is shown in Figure 1. Among all 1,822 patients, there were 1068 males (58.6%), 943 patients (51.8%) aged >43 years, and 1381 white patients (75.8%). In addition, the majority of patients in N stage were in No stage (1087; 59.7%), whereas 1548 (85.0%) were in Mo stage, according to laboratory examinations and postoperative pathological results. Non-small cell LC (NSCLC) was the most prevalent type of pathology in patients with EOLC, accounting for 67.9% (1237) of patients. The most common primary site of tumor in eligible EOLC patients was the upper lobe (925; 50.8%), followed by the lower lobe (578; 31.7%). The treatment protocol for patients included chemotherapy (874; 48.0%) and radiotherapy (508; 27.9%). The demographic and pathologic characteristics of the patients with EOLC are shown in Table 1.

Identification of prognostic factors of OS and CSS

Univariate and multivariate regression analyses were performed to investigate the independent prognostic factors for OS and CSS in patients with EOLC. For OS and CSS, gender, age, race, grade, TNM stage, tumor primary site, SEER stage,

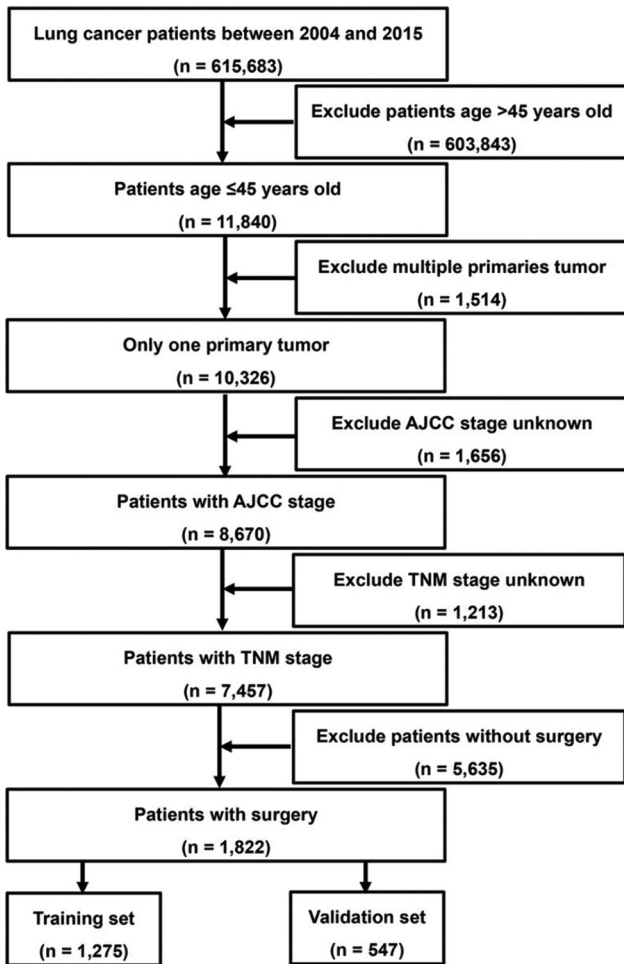


FIGURE 1. Schematic of specific patient screening process.

chemotherapy, and radiotherapy were the prognostic factors according to the univariate analysis. The multivariate analysis was further applied in our study, and it was found that the three variables (gender, chemotherapy, and radiotherapy) were excluded from the prognostic factors (Table 2). Moreover, the results of multivariate analysis also indicated that age, race, grade, TNM stage, tumor primary site, and SEER stage were independent prognostic factors impacting the CSS in patients with EOLC (Table 3). In addition, we further analyzed prognostic factors in patients with EOLC with NSCLC for their maximum percentage of histological type. The results of multivariate analysis indicated that age, race, grade, TNM stage, and chemotherapy were prognostic factors for OS in patients with EOLC with NSCLC, which lost the chemotherapy for CSS (Table S1).

Construction and verification of Nomograms

The clinical variables included in the construction of nomograms were based on the multivariate Cox regression results. The prognostic nomogram for 3-, 5-, and 10-year OS (Figure 2A) comprised age, race, grade, TNM stage, tumor primary site, and SEER stage as independent prognostic factors, and each variable corresponded to a point according to

TABLE 1. Baseline demographic and clinical characteristics with EOLC patients in our study

Variables	All patients n (%)	Training set n (%)	Validation set n (%)
Total	1822	1275 (70.0)	547 (30.0)
Gender			
Male	754 (41.4)	524 (41.1)	230 (42.0)
Female	1068 (58.6)	751 (58.9)	317 (58.0)
Age, years			
<35	292 (16.0)	211 (16.5)	81 (14.8)
35-43	587 (32.2)	414 (32.5)	173 (31.6)
>43	943 (51.8)	650 (51.0)	293 (53.6)
Race			
White	1381 (75.8)	964 (75.6)	417 (76.2)
Black	234 (12.8)	171 (13.4)	63 (11.5)
Others	207 (11.4)	140 (11.0)	67 (12.2)
Grade			
Grade I	351 (19.3)	242 (19.0)	109 (19.9)
Grade II	567 (31.1)	395 (31.0)	172 (31.4)
Grade III	570 (31.3)	396 (31.1)	174 (31.8)
Grade IV	60 (3.3)	45 (3.5)	15 (2.7)
Unknown	274 (15.0)	197 (15.5)	77 (14.1)
T stage			
T1	600 (32.9)	426 (33.4)	174 (31.8)
T2	773 (42.4)	545 (42.7)	228 (41.7)
T3	135 (7.4)	93 (7.3)	42 (7.7)
T4	314 (17.2)	211 (16.5)	103 (18.8)
N stage			
N0	1087 (59.7)	777 (60.9)	310 (56.7)
N1	274 (15.0)	201 (15.8)	73 (13.3)
N2	420 (23.0)	268 (21.0)	152 (27.8)
N3	41 (2.3)	29 (2.3)	12 (2.2)
M stage			
M0	1548 (85.0)	1094 (85.8)	454 (83.0)
M1	274 (15.0)	181 (14.2)	93 (17.0)
Histological type			
Small cell	22 (1.2)	14 (1.1)	8 (1.5)
Non-small cell	1237 (67.9)	860 (67.5)	377 (68.9)
Others	563 (30.9)	401 (31.5)	162 (29.6)
Primary site			
Main bronchus	80 (4.4)	57 (4.5)	23 (4.2)
Upper lobe	925 (50.8)	655 (51.4)	270 (49.4)
Middle lobe	119 (6.5)	87 (6.8)	32 (5.9)
Lower lobe	578 (31.7)	395 (31.0)	183 (33.5)
Overlapping lesion	41 (2.3)	28 (2.2)	13 (2.4)
Lung, NOS	79 (4.3)	53 (4.2)	26 (4.8)
SEER stage			
Localized	689 (37.8)	483 (37.9)	206 (37.7)
Regional	823 (45.2)	578 (45.3)	245 (44.8)
Distant	310 (17.0)	214 (16.8)	96 (17.6)
Chemotherapy			
No/Unknown	948 (52.0)	686 (53.8)	262 (47.9)
Yes	874 (48.0)	589 (46.2)	285 (52.1)
Radiotherapy			
No/Unknown	1314 (72.1)	933 (73.2)	381 (69.7)
Yes	508 (27.9)	342 (26.8)	166 (30.3)

SEER: Surveillance, Epidemiology, and End Results; NOS: Not otherwise specified. EOLC: Early-onset lung cancer

HR. Moreover, the establishment of a prognostic nomogram for CSS (Figure 2B) included age, race, grade, TNM stage, and

TABLE 2. Univariate and multivariate analysis of OS rates

Variables	No. of patients	Univariate analysis	Multivariate analysis ^a	
		<i>P</i> value	HR (95% CI)	<i>P</i> value
Gender		0.001		
Male	524		Reference	
Female	751		-	0.206
Age, years		<0.001		
<35	211		Reference	
35-43	414		1.99 (1.43-2.76)	<0.001
>43	650		1.32 (0.95-1.83)	0.093
Race		0.027		
White	964		Reference	
Black	171		1.16 (0.89-1.50)	0.274
Others	140		0.60 (0.42-0.86)	0.006
Grade		<0.001		
Grade I	242		Reference	
Grade II	395		2.13 (1.40-3.25)	<0.001
Grade III	396		2.94 (1.95-4.46)	<0.001
Grade IV	45		4.48 (2.58-7.78)	<0.001
Unknown	197		2.68 (1.73-4.16)	<0.001
T stage		<0.001		
T1	426		Reference	
T2	545		0.98 (0.75-1.29)	0.876
T3	93		1.50 (1.02-2.21)	0.041
T4	211		2.00 (1.47-2.73)	<0.001
N stage		<0.001		
N0	777		Reference	
N1	201		1.38 (1.03-1.85)	0.029
N2	268		1.89 (1.47-2.43)	<0.001
N3	29		2.26 (1.36-3.75)	0.002
M stage		<0.001		
M0	1094		Reference	
M1	181		3.00 (2.17-4.13)	<0.001
Histological type		0.170		
Small cell	14			
Non-small cell	860			
Others	401			
Primary site		0.006		
Main bronchus	57		Reference	
Upper lobe	655		0.52 (0.33-0.81)	0.004
Middle lobe	87		0.40 (0.22-0.74)	0.004
Lower lobe	395		0.64 (0.40-1.03)	0.064
Overlapping lesion	28		0.64 (0.31-1.34)	0.239
Lung, NOS	53		0.66 (0.37-1.17)	0.158
SEER stage		<0.001		
Localized	483		Reference	
Regional	578		1.51 (1.08-2.12)	0.016
Distant	214		1.30 (0.83-2.03)	0.250
Chemotherapy		<0.001		
No/Unknown	686		Reference	
Yes	589		-	0.061
Radiotherapy		<0.001		
No/Unknown	933		Reference	
Yes	342		-	0.593

OS: Overall survival; SEER: Surveillance, Epidemiology, and End Results. ^aModel was adjusted by gender, age at diagnosis, race, grade, TNM stage, primary site, SEER stage, chemotherapy, and radiotherapy. TNM: Tumor-node-metastasis

TABLE 3. Univariate and multivariate analysis of CSS rates

Variables	No. of patients	Univariate analysis	Multivariate analysis ^a	
		<i>p</i> -value	HR (95% CI)	<i>p</i> -value
Gender		0.001		
Male	524		Reference	
Female	751		-	0.247
Age, years		<0.001		
<35	211		Reference	
35-43	414		2.18 (1.51-3.14)	<0.001
>43	650		1.43 (0.99-2.05)	0.054
Race		0.027		
White	964		Reference	
Black	171		1.00 (0.74-1.34)	0.977
Others	140		0.52 (0.35-0.79)	0.002
Grade		<0.001		
Grade I	242		Reference	
Grade II	395		2.07 (1.30-3.29)	0.002
Grade III	396		3.05 (1.94-4.81)	<0.001
Grade IV	45		4.37 (2.38-8.04)	<0.001
Unknown	197		2.45 (1.50-3.99)	<0.001
T stage		<0.001		
T1	426		Reference	
T2	545		1.16 (0.87-1.56)	0.307
T3	93		2.05 (1.37-3.05)	<0.001
T4	211		2.44 (1.76-3.36)	<0.001
N stage		<0.001		
N0	777		Reference	
N1	201		1.63 (1.22-2.18)	0.001
N2	268		2.47 (1.94-3.15)	<0.001
N3	29		2.85 (1.73-4.70)	<0.001
M stage		<0.001		
M0	1094		Reference	
M1	181		2.96 (2.33-3.77)	<0.001
Histological type		0.057		
Small cell	14			
Non-small cell	860			
Others	401			
Primary site		0.006		
Main bronchus	57		Reference	
Upper lobe	655		0.55 (0.33-0.90)	0.017
Middle lobe	87		0.38 (0.19-0.76)	0.006
Lower lobe	395		0.71 (0.42-1.18)	0.185
Overlapping lesion	28		0.59 (0.25-1.37)	0.218
Lung, NOS	53		0.70 (0.37-1.30)	0.257
SEER stage		<0.001		
Localized	483		Reference	
Regional	578		-	0.029
Distant	214		-	0.334
Chemotherapy		<0.001		
No/Unknown	686		Reference	
Yes	589		-	0.706
Radiotherapy		<0.001		
No/Unknown	933		Reference	
Yes	342		-	0.677

CSS: Cancer-specific survival; SEER: Surveillance, Epidemiology, and End Results. ^aModel was adjusted by gender, age at diagnosis, race, grade, TNM stage, primary site, SEER stage, chemotherapy, and radiotherapy. CSS: Cancer-specific survival; TNM: Tumor-node-metastasis

tumor primary site as the variables. Simultaneously, the prognostic nomograms for OS (Figure S2A) and CSS (Figure S2B)

of patients with EOLC with NSCLC were established according to the Cox regression results.

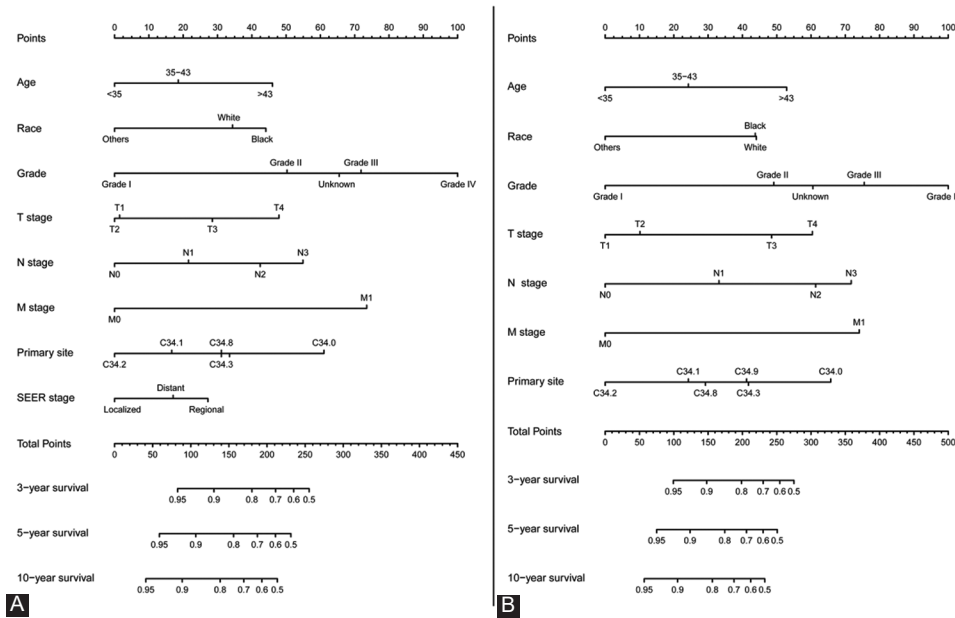


FIGURE 2. The nomogram containing various factors for the 3-, 5-, and 10-year overall survival (OS) and cancer-specific survival (CSS) prediction of early-onset lung cancer patients. (A) Nomogram for OS; (B) Nomogram for CSS.

The time-dependent ROC curves for OS and CSS were plotted to assess the predictive performance of nomograms in different sets. In the training set, the AUC of the nomograms for OS (Figure 3A) and CSS (Figure 3B) was 0.766 (95% CI: 0.745–0.787) and 0.782 (95% CI: 0.760–0.804), respectively (Table 4), which were significantly larger than values for TNM stage and SEER stage. The results in validation set showed the same conclusion; the AUC of nomograms were 0.768 (95% CI: 0.738–0.798) for OS (Figure 3C) and 0.780 (95% CI: 0.748–0.812) for CSS (Figure 3D). Simultaneously, the DCA was applied to verify the clinical utility of nomograms. The results indicated that the nomogram showed comparable clinical applicability for predicting OS and CSS as TNM stage and SEER stage, not only in training set (Figure 4A and B) but also in validation set (Figure 4C and D).

In addition, the concordance index (C-index) was conducted in this study to verify the nomogram. There were significant differences among nomogram, TNM stage, and SEER stage for OS and CSS (Table 5). We therefore used the calibration curve method to compare nomograms with the perfect curves. The results show that the 3-, 5-, and 10-year OS (Figure 5A, C, and E) and CSS (Figure 5B, D, and F) nomograms in the training set possessed excellent consistency with actual observation, which was also found in the validation set (Figure S3). The above results indicated that there was good agreement between the predictions of the nomograms and the actual observations in both the training set and the validation set.

DISCUSSION

At present, the research on patients with EOLC (aged <45 years old) attracted widespread attention due to the rapid

increase in LC morbidity and mortality worldwide. It was strongly suggested that genomic mutation was an important predisposing factor for EOLC [12]. Patients with EOLC usually have poor survival outcomes and a higher proportion of family history with other types of cancers [13,14]. In practice, accurately predicting the prognosis of patients with EOLC and formulating individualized treatments are conducive to improving the survival rate. However, the current pathological staging of tumors based on imaging examinations do not meet the requirements for accurate prognosis prediction of patients with EOLC. There is an urgent need for a reliable system to comprehensively consider multiple prognostic factors in patients with EOLC to accurately predict survival time.

This study focused on the prognosis prediction for patients with EOLC based on the construction of nomograms. First, we established prognostic nomograms for 3-, 5-, and 10-year OS and CSS in patients with EOLC. The clinical variables included in the establishment were determined by the results of Cox regression and comprised age, race, grade, TNM stage, tumor primary site, and SEER stage. In addition, the clinical utility and predictive performance of nomograms were verified by ROC curve, DCA curve, and C-index, indicating that efficacy was better than of TNM stage. Furthermore, the accuracy of predicting 3-, 5-, and 10-year OS and CSS was evaluated by the calibration curve, which showed excellent agreement between the nomogram and the actual observation results. In practice, the AUC in ROC analysis and the C-index were generally higher than 0.760 and 0.790, respectively, for all nomograms, which confirmed the promising predictive ability of nomograms. The results of the DCA curve also supported the good clinical practical value of the nomogram.

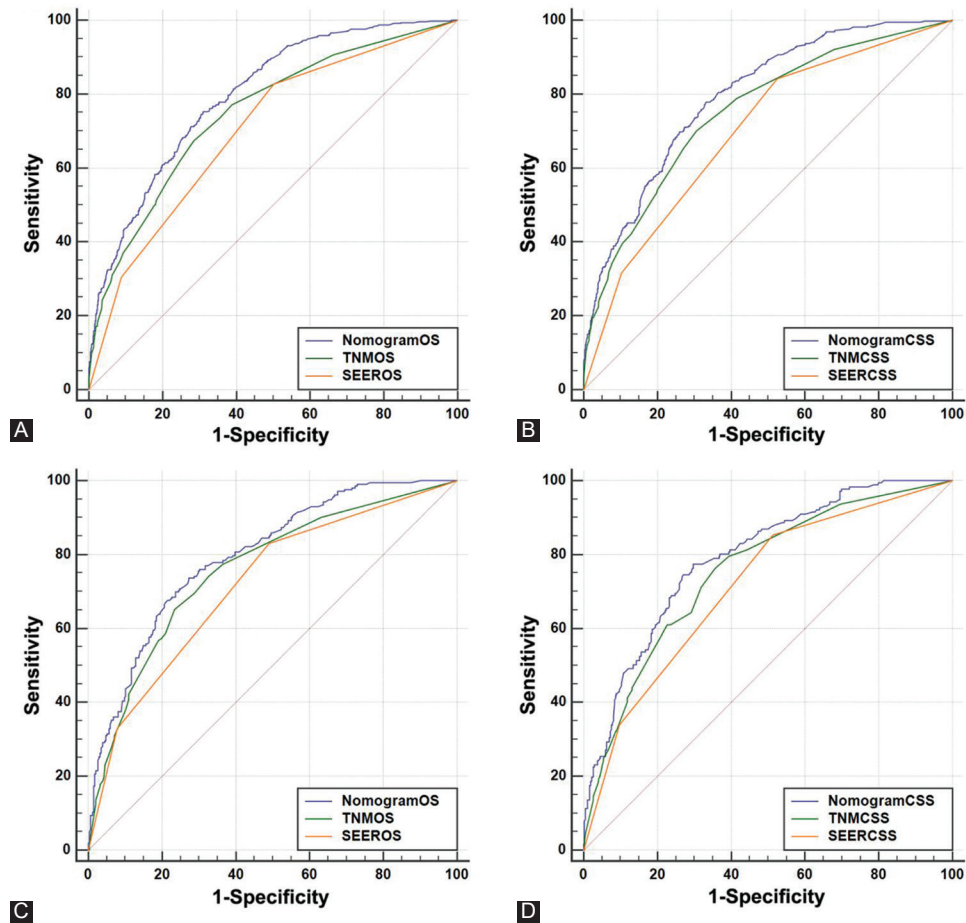


FIGURE 3. Receiver operating characteristic (ROC) verified the predictive value of nomogram, tumor-node-metastasis stage and surveillance, epidemiology, and end results stage in different sets. (A) ROC for OS in training set; (B) ROC for CSS in training set; (C) ROC for OS in validation set; (D) ROC for CSS in validation set.

TABLE 4. Comparison of AUC between the nomogram, TNM, and SEER stages in EOLC patients

Characteristics	Training set			Validation set		
	AUC	95% CI	<i>p</i> -value	AUC	95% CI	<i>p</i> -value
OS						
Nomogram	0.766	0.745-0.787		0.768	0.738-0.798	
TNM stage	0.731	0.708-0.754	<0.001	0.736	0.702-0.770	0.003
SEER stage	0.690	0.667-0.713	<0.001	0.709	0.677-0.741	<0.001
CSS						
Nomogram	0.782	0.760-0.804		0.780	0.748-0.812	
TNM stage	0.749	0.725-0.773	<0.001	0.752	0.717-0.787	0.005
SEER stage	0.700	0.676-0.724	<0.001	0.722	0.688-0.756	<0.001

AUC: Area under the curve; CI: confidence interval; SEER: Surveillance, Epidemiology, and End Results. EOLC: Early-onset lung cancer; TNM: Tumor-node-metastasis; AUC: Areas under the curve

Nomograms integrated the biological results into a mathematical model to establish a comprehensive consideration of various clinical characteristics and pathological variables of patients with cancer and then graphically displayed the possibility of clinical results. Nomograms were reported to be more accurate than existing models in predicting patient prognosis [15]. Recently, an increased number of nomograms comprising various clinical variables have been used to predict the prognosis of patients with LC [16-19]. Liang et al [18] analyzed NSCLC patient data in multiple clinical centers and

established the nomogram for postoperative survival prediction. As a multicenter study, it provided patients with resected NSCLC with an accurate individualized prediction of OS and assisted clinicians in decision making. Similarly, Zheng et al [16] developed the nomogram for predicting prognoses in LC with bone metastasis and comprehensively analyzed the independent prognostic factors, which included age, gender, histological types, grade, and others.

In this study, the following clinical variables, including age, race, grade, TNM stage, tumor primary site, and SEER

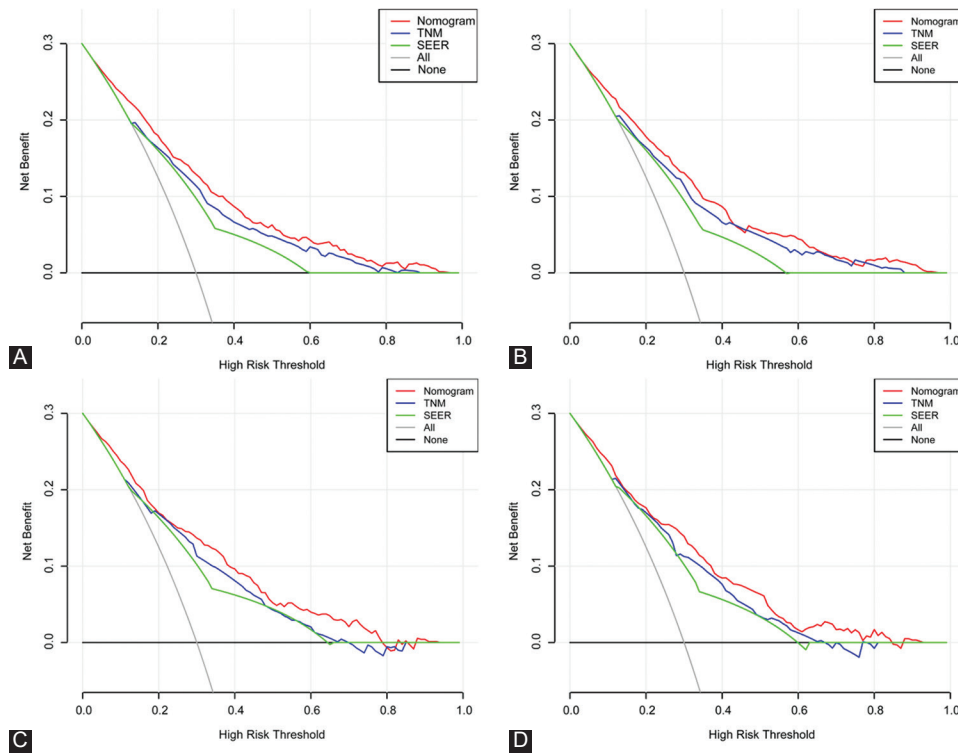


FIGURE 4. Decision curve analysis (DCA) based on nomograms, tumor-node-metastasis-stage and surveillance, epidemiology, and end results stage in different sets. (A) DCA for OS in training set; (B) DCA for CSS in training set; (C) DCA for OS in validation set; (D) DCA for CSS in validation set.

TABLE 5. Comparison of C-indexes between the nomogram, TNM, and SEER stages in EOLC patients

Characteristics	Training set			Validation set		
	HR	95% CI	p-value	HR	95% CI	p-value
OS						
Nomogram	0.797	0.773-0.818		0.797	0.760-0.830	
TNM stage	0.749	0.724-0.773	<0.001	0.759	0.721-0.794	0.005
SEER stage	0.703	0.677-0.728	<0.001	0.718	0.679-0.756	<0.001
CSS						
Nomogram	0.794	0.771-0.816		0.792	0.755-0.825	
TNM stage	0.752	0.728-0.776	<0.001	0.755	0.716-0.790	0.006
SEER stage	0.698	0.672-0.723	<0.001	0.715	0.676-0.753	<0.001

HR: Hazard ratio; CI: Confidence interval; SEER: Surveillance, Epidemiology, and End Results. EOLC: Early-onset lung cancer; TNM: Tumor-node-metastasis

stage, were the independent risk factors that impacted the prognosis of patients with EOLC. Many studies reported age and race as risk factors for the prognosis of various cancers [20,21]. Genetic differences among races as a significant risk factor for tumor prognosis has also been widely recognized [22,23]. Michele et al [24] found that first-degree relatives of patients with EOLC in black races were more susceptible to developing LC, which indicates significant differences among races.

The grade, primary site, and metastasis of the tumor also significantly affects the prognosis of patients [25]. The pathological grade of a tumor was positively correlated with the degree of malignancy and invasion [26]. It has been suggested that cancer cells in high-grade tumors were insensitive to treatment [27], which adversely affects the prognosis

of patients. Tumor site of cancers is as important a factor affecting the prognosis of patients [28,29]. For patients with LC, the primary site of tumor in the right and left lower lobe or in the right middle and left lingual lobe is more susceptible to mediastinal lymph node tumor metastasis [30]. Moreover, lymph node metastasis or distant tumor metastasis represents a poor prognosis and short survival time for patients. In our study, the same results were supported by statistical analysis.

Currently, TNM stage, determined by laboratory results and postoperative pathological examination, is the most widely accepted tumor staging system. In practice, clinicians would judge TNM stage based on individual characteristics of the tumor (T), node (N), and metastasis (M) [6]. Chen et al [31] verified the prognostic value of the *TNM Classification of*

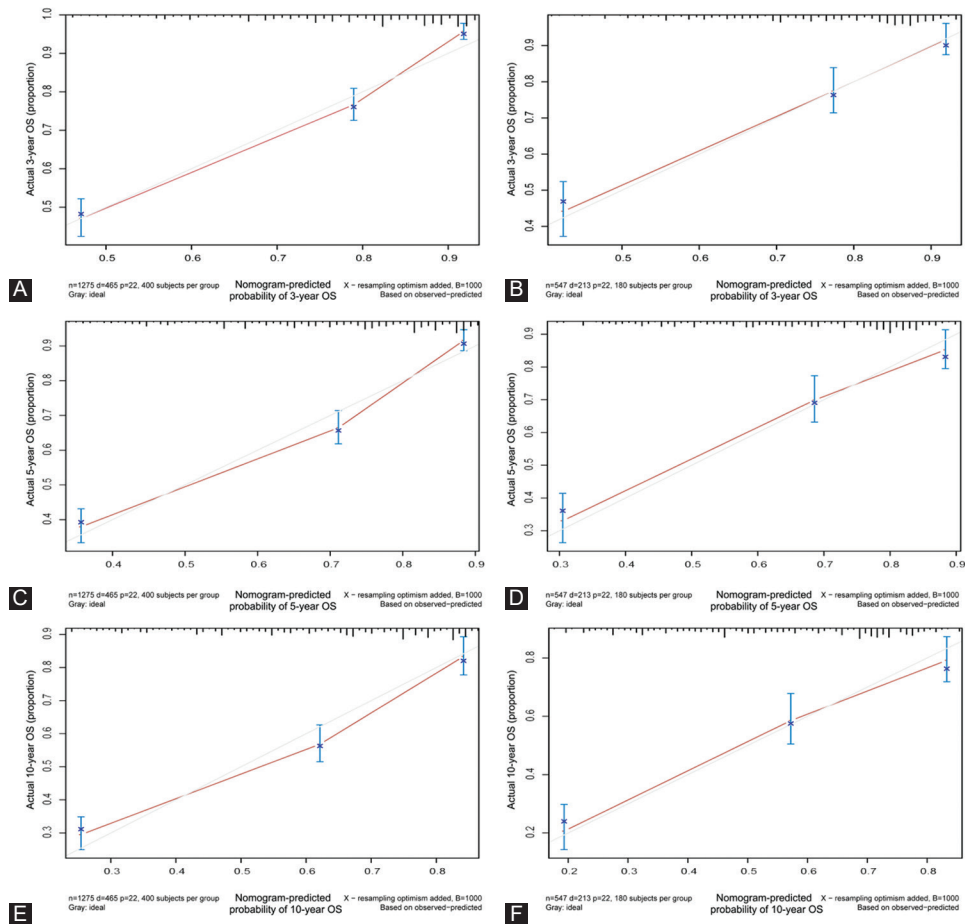


FIGURE 5. Calibration plot of the 3-, 5-, and 10- year OS nomogram in training and validation sets. (A) 3-year OS in training set; (B) 3-year OS in validation set; (C) 5-year OS in training set; (D) 5-year OS in validation set; (E) 10-year OS in training set; (F) 10-year OS in validation set.

Malignant Tumours, 8th Edition TNM staging system for patients with LC and found recurrence-free survival could also be predicted through TNM stage. However, the TNM stage has limitations and could not provide clinicians with individualized prognosis prediction. As shown in this study, patient prognosis was also closely related to a variety of clinical variables except TNM staging, and accurate prediction relied on the comprehensive consideration of all independent risk factors. We successfully established an effective nomogram based on age, race, grade, TNM stage, tumor primary site, and SEER stage, which has been proven a better predictive tool than TNM stage alone. The construction of nomograms would be useful in helping to develop personalized treatment for patients with EOLC.

There are still some limitations to our study. First, the SEER database as a retrospective database includes biases in data collection due to manual recording and other reasons. Second, the clinical data were incomplete; for example, the SEER database failed to record the genetic changes in the patients. Third, the analyzed data did not represent other regions and required external verification. Therefore, it is necessary to conduct multicenter prospective clinical trials to verify the accuracy of nomograms.

CONCLUSIONS

We established prognostic nomograms for 3-, 5-, and 10-year OS and CSS in EOLC patients based on a large amount of clinical data, and this prognostic nomogram has good predictive ability. The models could help clinicians prepare personalized treatment for patients with EOLC.

REFERENCES

1. Malhotra J, Malvezzi M, Negri E, La Vecchia C, Boffetta P. Risk factors for lung cancer worldwide. *European Respiratory Journal*. 2016;48(3):889-902.
2. The L. Lung cancer: some progress, but still a lot more to do. *The Lancet*. 2019;394(10212):1880.
3. Graziano C, Comin CE, Crisci C, Novelli L, Politi L, Messerini L, et al. Functional polymorphisms of the microsomal epoxide hydroxylase gene: A reappraisal on an early-onset lung cancer patients series. *Lung Cancer*. 2009;63(2):187-93.
4. Timofeeva M, Kropp S, Sauter W, Beckmann L, Rosenberger A, Illig T, et al. Genetic polymorphisms of MPO, GSTT1, GSTM1, GSTP1, EPHX1 and NQO1 as risk factors of early-onset lung cancer. *International journal of cancer*. 2010;127(7):1547-61.
5. Amin MB, Greene FL, Edge SB, Compton CC, Gershonwald JE, Brookland RK, et al. *The Eighth Edition AJCC Cancer Staging Manual: Continuing to build a bridge from a population-based to a*

- more “personalized” approach to cancer staging. *CA Cancer J Clin*. 2017;67(2):93-9.
6. Lim W, Ridge CA, Nicholson AG, Mirsadraee S. The 8(th) lung cancer TNM classification and clinical staging system: review of the changes and clinical implications. *Quantitative imaging in medicine and surgery*. 2018;8(7):709-18.
 7. Balachandran VP, Gonen M, Smith JJ, DeMatteo RP. Nomograms in oncology: more than meets the eye. *Lancet Oncol*. 2015;16(4):e173-80.
 8. Yu C, Zhang Y. Development and validation of prognostic nomogram for young patients with gastric cancer. *Annals of translational medicine*. 2019;7(22):641.
 9. Pan X, Yang W, Chen Y, Tong L, Li C, Li H. Nomogram for predicting the overall survival of patients with inflammatory breast cancer: A SEER-based study. *Breast (Edinburgh, Scotland)*. 2019;47:56-61.
 10. Mao W, Wu J, Kong Q, Li J, Xu B, Chen M. Development and validation of prognostic nomogram for germ cell testicular cancer patients. *Aging (Albany NY)*. 2020;12(21):22095-111.
 11. Eastham JA, Kattan MW, Scardino PTJSiUO. Nomograms as predictive models. 2002;20(2):108-15.
 12. Timofeeva MN, Kropp S, Sauter W, Beckmann L, Rosenberger A, Illig T, et al. CYP450 polymorphisms as risk factors for early-onset lung cancer: gender-specific differences. *Carcinogenesis*. 2009;30(7):1161-9.
 13. Kreuzer M, Kreienbrock L, Müller KM, Gerken M, Wichmann E. Histologic types of lung carcinoma and age at onset. *Cancer*. 1999;85(9):1958-65.
 14. Etzel CJ, Lu M, Merriman K, Liu M, Vaporciyan A, Spitz MR. An epidemiologic study of early onset lung cancer. *Lung Cancer*. 2006;52(2):129-34.
 15. Wang ZX, Qiu MZ, Jiang YM, Zhou ZW, Li GX, Xu RH. Comparison of prognostic nomograms based on different nodal staging systems in patients with resected gastric cancer. *J Cancer*. 2017;8(6):950-8.
 16. Zheng XQ, Huang JF, Lin JL, Chen L, Zhou TT, Chen D, et al. Incidence, prognostic factors, and a nomogram of lung cancer with bone metastasis at initial diagnosis: a population-based study. *Translational lung cancer research*. 2019;8(4):367-79.
 17. Wang S, Yang L, Ci B, Maclean M, Gerber DE, Xiao G, et al. Development and Validation of a Nomogram Prognostic Model for SCLC Patients. *Journal of thoracic oncology : official publication of the International Association for the Study of Lung Cancer*. 2018;13(9):1338-48.
 18. Liang W, Zhang L, Jiang G, Wang Q, Liu L, Liu D, 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(8):861-9.
 19. Wang L, Dong T, Xin B, Xu C, Guo M, Zhang H, et al. Integrative nomogram of CT imaging, clinical, and hematological features for survival prediction of patients with locally advanced non-small cell lung cancer. *European radiology*. 2019;29(6):2958-67.
 20. Zeng C, Wen W, Morgans AK, Pao W, Shu XO, Zheng W. Disparities by Race, Age, and Sex in the Improvement of Survival for Major Cancers: Results From the National Cancer Institute Surveillance, Epidemiology, and End Results (SEER) Program in the United States, 1990 to 2010. *JAMA Oncol*. 2015;1(1):88-96.
 21. Mao W, Wei S, Yang H, Yu Q, Xu M, Guo J, et al. Clinicopathological study of organ metastasis in endometrial cancer. *Future Oncology*. 2020;16(10):525-40.
 22. Özdemir BC, Dotto GP. Racial Differences in Cancer Susceptibility and Survival: More Than the Color of the Skin? *Trends in cancer*. 2017;3(3):181-97.
 23. Freedland S. Prostate cancer: Race and prostate cancer personalized medicine: the future. *Nature reviews Urology*. 2018;15(4):207-8.
 24. Coté ML, Kardia SL, Wenzlaff AS, Ruckdeschel JC, Schwartz AG. Risk of lung cancer among white and black relatives of individuals with early-onset lung cancer. *Jama*. 2005;293(24):3036-42.
 25. Zeng Y, Mayne N, Yang CJ, D'Amico TA, Ng CSH, Liu CC, et al. A Nomogram for Predicting Cancer-Specific Survival of TNM 8th Edition Stage I Non-small-cell Lung Cancer. *Annals of surgical oncology*. 2019;26(7):2053-62.
 26. Pasello G, Zago G, Lunardi F, Urso L, Kern I, Vlacic G, et al. Malignant pleural mesothelioma immune microenvironment and checkpoint expression: correlation with clinical-pathological features and intratumor heterogeneity over time. *Annals of oncology: official journal of the European Society for Medical Oncology*. 2018;29(5):1258-65.
 27. Samadani AA, Norollahi SE, Rashidy-Pour A, Mansour-Ghanaei F, Nemati S, Joukar F, et al. Cancer signaling pathways with a therapeutic approach: An overview in epigenetic regulations of cancer stem cells. *Biomedicine & pharmacotherapy = Biomedecine & pharmacotherapie*. 2018;108:590-9.
 28. Siotos C, McColl M, Psoter K, Gilmore RC, Sebai ME, Broderick KP, et al. Tumor Site and Breast Cancer Prognosis. *Clinical breast cancer*. 2018;18(5):e1045-e52.
 29. Pavlidis N, Pentheroudakis G. Cancer of unknown primary site. *Lancet (London, England)*. 2012;379(9824):1428-35.
 30. Schmulewitz N, Wildi SM, Varadarajulu S, Roberts S, Hawes RH, Hoffman BJ, et al. Accuracy of EUS criteria and primary tumor site for identification of mediastinal lymph node metastasis from non-small-cell lung cancer. *Gastrointestinal Endoscopy*. 2004;59(2):205-12.
 31. Chen K, Chen H, Yang F, Sui X, Li X, Wang J. Validation of the Eighth Edition of the TNM Staging System for Lung Cancer in 2043 Surgically Treated Patients With Non-small-cell Lung Cancer. *Clinical lung cancer*. 2017;18(6):e457-e66.

Related articles published in BJBMS

1. [Development and validation of nomograms for predicting survival of elderly patients with stage I small-cell lung cancer](#)
Yaji Yang et al., BJBMS, 2020
2. [Trends of incidence and prognosis of upper tract urothelial carcinoma](#)
Ming Chen et al., BJBMS, 2020

SUPPLEMENTRY FIGURES

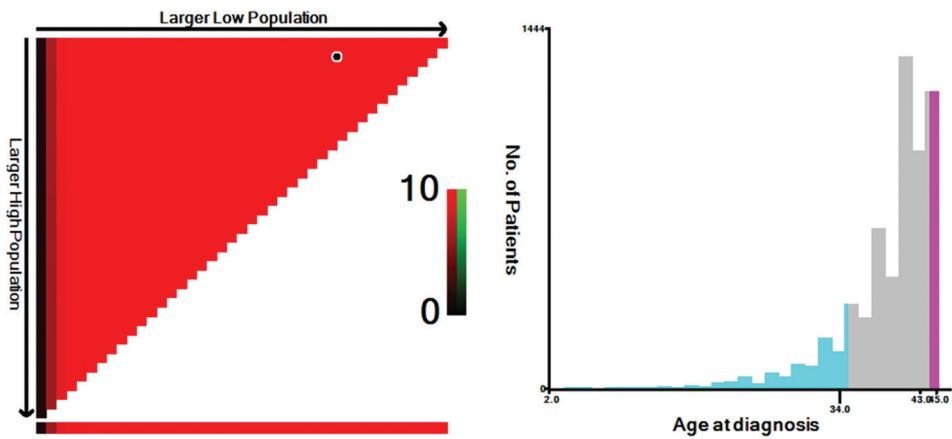


FIGURE S1. Estimation of the cut-off value for the age determined by X-tile software.

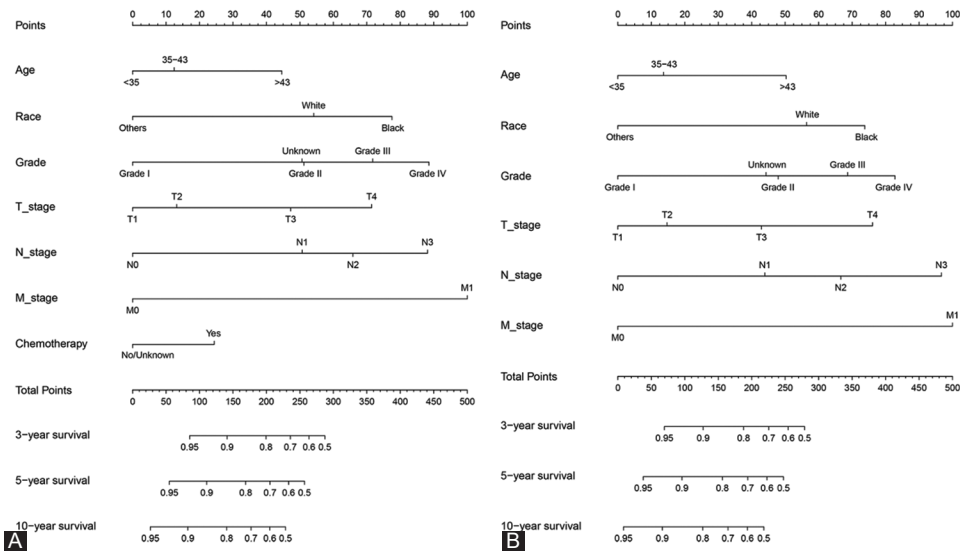


FIGURE S2. The nomogram containing various factors for the 3-, 5-, and 10-year overall survival (OS) and cancer-specific survival (CSS) prediction of early-onset lung cancer (LC) patients with non-small cell LC. (A) Nomogram for OS; (B) Nomogram for CSS.

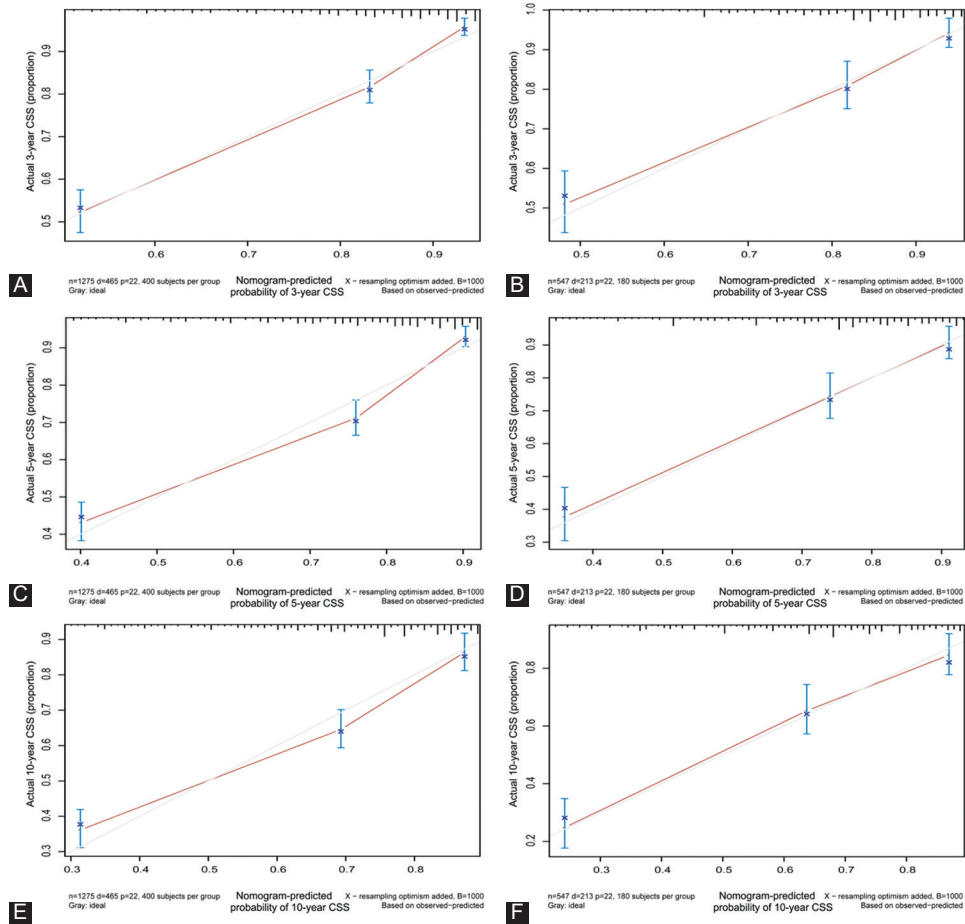


FIGURE S3. Calibration plot of the 3-, 5-, and 10- year CSS nomogram in training and validation sets. (A) 3-year CSS in training set; (B) 3-year CSS in validation set; (C) 5-year CSS in training set; (D) 5-year CSS in validation set; (E) 10-year CSS in training set; (F) 10-year CSS in validation set.

SUPPLEMENTRY TABLE**TABLE S1.** Multivariate analysis of OS and CSS rates of patients with non-small cell lung cancer in the training set

Variables	Overall survival		Cancer-specific survival	
	Multivariate analysis ^a		Multivariate analysis ^b	
	HR (95% CI)	<i>p</i> -value	HR (95% CI)	<i>p</i> -value
Gender				
Male	Reference		Reference	
Female	-	0.168	-	0.121
Age, years				
<35	Reference		Reference	
35-43	1.77 (1.20-2.62)	0.004	1.87 (1.21-2.87)	0.005
>43	1.17 (0.80-1.73)	0.416	1.19 (0.78-1.82)	0.433
Race				
White	Reference		Reference	
Black	1.35 (0.99-1.83)	0.055	1.24 (0.88-1.75)	0.225
Others	0.50 (0.32-0.79)	0.003	0.50 (0.31-0.81)	0.005
Grade				
Grade I	Reference		Reference	
Grade II	1.93 (1.20-3.11)	0.007	1.81 (1.07-3.07)	0.026
Grade III	2.51 (1.56-4.02)	<0.001	2.34 (1.40-3.92)	0.001
Grade IV	3.11 (1.62-5.99)	0.001	2.79 (1.37-5.71)	0.005
Unknown	1.92 (1.15-3.20)	0.013	1.74 (0.99-3.05)	0.055
T stage				
T1	Reference		Reference	
T2	1.18 (0.88-1.60)	0.270	1.20 (0.86-1.67)	0.276
T3	1.83 (1.18-2.83)	0.007	1.70 (1.05-2.75)	0.031
T4	2.49 (1.77-3.50)	<0.001	2.56 (1.77-3.69)	<0.001
N stage				
N0	Reference		Reference	
N1	1.91 (1.39-2.62)	<0.001	1.72 (1.22-2.42)	0.002
N2	2.32 (1.76-3.06)	<0.001	2.28 (1.72-3.02)	<0.001
N3	3.03 (1.69-5.45)	<0.001	3.24 (1.81-5.82)	<0.001
M stage				
M0	Reference		Reference	
M1	3.57 (2.72-4.67)	<0.001	3.42 (2.57-4.54)	<0.001
Primary site				
Main bronchus	Reference		Reference	
Upper lobe	-	0.024	-	0.013
Middle lobe	-	0.696	-	0.374
Lower lobe	-	0.101	-	0.033
Overlapping lesion	-	0.556	-	0.713
Lung, NOS	-	0.618	-	0.611
SEER stage				
Localized	Reference		Reference	
Regional	-	0.080	-	0.123
Distant	-	0.965	-	0.572
Chemotherapy				
No/Unknown	Reference		Reference	
Yes	1.35 (1.03-1.77)	0.029	-	0.512
Radiotherapy				
No/Unknown	Reference		Reference	
Yes	-	0.389	-	0.354

OS: Overall survival; CSS: cancer-specific survival; HR: Hazard ratio; CI: Confidence intervals; SEER: Surveillance, Epidemiology, and End Results. ^{a,b}Model was adjusted by gender, age, race, grade, AJCC stage, TNM stage, histology, SEER stage, chemotherapy, and radiotherapy. TNM: Tumor-node-metastasis; AJCC: American Joint Commission on Cancer