Am J Cancer Res 2026;16(3):1198-1214
www.ajcr.us /ISSN:2156-6976/ajcr0171804

Original Article

Nomogram outperforms gradient boosting machine for prognostic prediction of laryngeal squamous cell carcinoma: a combined analysis of SEER and single-center data

Difeng Guo[1*], Guanghai Ji[2*], Mengru Jian[1], Kebing Chen[3], Yan Li[1]

[1]*Department of Oncology, The First Affiliated Hospital of Yangtze University, Jingzhou 434000, Hubei, China;*
[2]*Department of Radiology, The First Affiliated Hospital of Yangtze University, Jingzhou 434000, Hubei, China;*
[3]*Department of Pathology, The First Affiliated Hospital of Yangtze University, Jingzhou 434000, Hubei, China.*
[*]*Equal contributors.*

Received January 31, 2026; Accepted March 25, 2026; Epub March 25, 2026; Published March 30, 2026

Abstract: The incidence of laryngeal squamous cell carcinoma (LSCC) remains persistently high, necessitating accurate prognostic prediction for clinical treatment guidance. By comparing the performance of LSCC models based on the Surveillance, Epidemiology, and End Results (SEER) database and those constructed from single-center datasets, this study provides an effective tool for clinical prognosis evaluation. Data from 526 patients with LSCC were extracted from the SEER database. Univariate and multivariate Cox regression analyses were performed to identify independent predictors of overall survival (OS) in LSCC patients. Subsequently, patients were randomly assigned in a 7:3 ratio to the modeling group and the test group. Based on the modeling group data, nomograms and gradient boosting machine (GBM) models were constructed using R software (version 4.4.1) and their performance was evaluated. The testing cohort was utilized to assess the predictive accuracy of the model. In addition, 207 LSCC patients diagnosed at The First Affiliated Hospital of Yangtze University from February 2020 to April 2024 were retrospectively selected as an external validation cohort. Univariate and Multivariate Cox regression analyses determined that age (60-75 years: HR=1.333, P=0.085; >75 years: HR=2.726, P<0.001), tumor size (HR=1.013, P=0.035), radiation (HR=7.555, P<0.001), cause of death (COD, HR=3.996, P<0.001), marital status at diagnosis (HR=1.444, P=0.006), and T stage (HR=1.652, P=0.017) were independent predictive indicators affecting the OS of LSCC patients (*P*<0.05). On this basis, nomogram and GBM models were constructed. The ROC curve showed that the GBM model had an AUC of 0.747, 0.763, and 0.785 at 1-, 2- and 3-years, respectively. For nomogram model, the 1-, 2- and 3-year AUC values reached 0.809, 0.782 and 0.811, respectively. Delong test showed that the AUC values of nomogram were all higher than those of the GBM model (P<0.05). Next, we used nomogram models for verification. AUC values in the verification cohort were 0.783, 0.786, and 0.801, respectively. The AUC values for the external validation cohort were 0.795, 0.760, and 0.783, respectively. The calibration curve shows that the predicted value is basically consistent with the real value. The nomogram model has robust prediction ability and reliable calibration, and its performance is better than that of GBM model.

Keywords: Laryngeal squamous cell carcinoma, SEER, survival, prediction

Introduction

Laryngeal squamous cell carcinoma (LSCC) is the most common malignant tumor of the head and neck, accounting for approximately 95% of all laryngeal malignancies and imposing a heavy burden on global public health due to its high incidence, poor clinical prognosis and potential impact on patients' speech, swallowing and respiratory functions [1]. Epidemiological investigations have confirmed that the incidence of LSCC shows significant variations across different regions and ethnic groups, with risk factors including tobacco and alcohol consumption, human papillomavirus infection, and occupational exposure to harmful substances, among which smoking and heavy drinking remain the most critical modifiable risk factors.

https://doi.org/10.62347/AHUE7844

Despite advances in comprehensive treatment strategies combining surgery, radiotherapy and chemotherapy in recent decades, the 5-year overall survival (OS) rate of LSCC patients remains only 50%-60% [2], and the prognosis of patients with the same clinical stage can vary drastically, highlighting the urgent need for more accurate prognostic assessment tools to guide clinical decision-making.

Accurate prediction of prognosis in LSCC patients is of great significance for developing individualized treatment plans and improving therapeutic efficacy. Currently, in clinical practice, the prognosis of LSCC patients is mostly evaluated by the tumor TNM staging system [3]. However, the TNM staging system has inherent limitations and cannot comprehensively cover individual patient differences and other key prognostic factors. The Surveillance, Epidemiology, and End Results (SEER) data contains a vast amount of cancer patient data, providing valuable support for prognostic prediction studies in LSCC [4]. Single-center clinical data, with detailed patient clinical information and follow-up data, can fully reflect the characteristics of patients in the local area. The complementary advantages of the SEER database repository and single-center data create the possibility of constructing a more universal and accurate prognostic prediction model. The complementary advantages of the SEER database and single-center clinical data create a unique opportunity to construct a prognostic prediction model with both universality (based on large-sample population data) and regional applicability (based on single-center real-world data) for LSCC.

In previous studies, the exploration of prognostic factors for LSCC mostly focused on scattered single - factor analyses or small-scale cohort studies. However, there is a lack of systematic research integrating large-scale databases and multi-center data, especially research on constructing LSCC prognostic prediction models based on the SEER database repository. This research utilizes data from the SEER database repository to thoroughly investigate the crucial factors influencing the prognosis of LSCC patients. Single-center clinical data were used as the testing cohort for external verification of the model. Furthermore, few studies have compared the predictive performance of traditional clinical prognostic models (e.g., Cox nomogram, a visualization tool with high clinical interpretability) with machine learning models (e.g., Gradient Boosting Machine (GBM), a data-driven model with strong fitting ability) in LSCC, which makes it difficult for clinicians to select the most appropriate prognostic assessment tool in clinical practice.

Against this research background, this study first utilizes the large-sample data from the SEER database to thoroughly screen and identify the independent clinical and pathological prognostic factors influencing the OS of LSCC patients through rigorous statistical analysis. We then take the single-center clinical data of LSCC patients from a tertiary hospital in China as the external validation cohort to ensure the regional applicability of the constructed model. On this basis, we construct two prognostic prediction models for LSCC (Cox nomogram and GBM model) and conduct a comprehensive comparison of their predictive efficacy in terms of discrimination, calibration and stability. The innovative value of this study lies in three aspects: first, integrating SEER large-sample data and single-center real-world data to realize the combination of population-based research and local clinical practice, improving the universality and applicability of the model; second, systematically comparing the predictive performance of a traditional interpretable model and a modern machine learning model for LSCC prognosis, filling the research gap in model selection for LSCC prognostic assessment; third, constructing a high-performance prognostic prediction model with clear clinical interpretability, which is more conducive to clinical promotion and application.

Materials and methods

Patient selection

Using the SEER*Stat v8.4.5, the data of patients diagnosed with LSCC from 2016 to 2017 (the follow-up years: 2016 to 2020) were extracted from the SEER database.

URL for the SEER database

Official homepage of the SEER Program: https://seer.cancer.gov/. SEER data access portal (for data application and download): https://seer.cancer.gov/data/access.html.

Sample size calculation was performed based on the minimum event per variable (EPV) principle, which is a standard approach for prognostic model construction in oncology research. To avoid model overfitting and ensure sufficient statistical power, the minimum EPV was set to 10, a commonly accepted threshold in clinical prognostic studies. Prior to patient screening, we reviewed relevant literature on LSCC prognostic factors and estimated that the number of potential independent prognostic variables (e.g., age, TNM stage, treatment methods) to be included in the subsequent Cox regression analysis would be 5-6. Accordingly, the minimum required number of survival events (overall survival events) was 50-60. After applying the predefined inclusion and exclusion criteria to the SEER database, a total of 526 LSCC patients with complete clinical, pathological, and follow-up data were enrolled. This sample size was sufficient to support the subsequent univariate and multivariate Cox regression analyses, as well as the construction and validation of the prognostic models, ensuring the reliability and stability of the study results.

After screening, a total of 526 LSCC patients were included as the research subjects. Information such as the patients' age, gender, race, tumor location, TNM staging, tumor differentiation degree, and treatment methods was collected. The patient enrollment flowchart is shown in **Figure 1**.

Inclusion Criteria: (1) Patients with a pathological diagnosis of LSCC, and the histological and site codes are: squamous cell neoplasms (8050-8089); (2) The primary tumor site is the larynx; (3) The histopathological type is squamous cell carcinoma (8070); (4) Having complete clinical and pathological information, including age, gender, race, tumor location, TNM staging, differentiation degree, treatment methods (surgical methods, radiation dose, chemotherapy regimen, etc.); (5) Having a clear survival outcome and follow-up time.

Exclusion Criteria: (1) Unclear information about race, histological grading, TNM staging, tumor size, tumor location, surgical/radiation therapy status, etc.; (2) Survival time <1 month; (3) Diagnostic age <18 years old; (4) Only receiving local simple treatment (such as only undergoing neck lymph node dissection without effective treatment for the primary tumor and other similar situations), which makes it impossible to comprehensively evaluate the overall treatment and prognosis of LSCC.

In addition, 207 patients diagnosed with LSCC at The First Affiliated Hospital of Yangtze University from February 2020 to April 2024 were retrospectively included as research subjects. These patients also followed the above inclusion and exclusion criteria. This study was approved by the Ethics Committee of The First Affiliated Hospital of Yangtze University.

Data extraction

The clinical information of 526 LSCC patients sourced from the SEER database was randomly allocated into the modeling cohort and the testing cohort in a 7:3 proportion. The clinical information of LSCC patients diagnosed at a single center from February 2020 to April 2024 was used as an external validation cohort. Drawing on the clinical information from the SEER database, univariate Cox analysis was performed to identify factors linked to overall survival (OS), and multivariate Cox analysis was utilized to determine the independent predictive indicators for LSCC patients. Nomograms and GBM models were constructed, and ROC curves and calibration curves were plotted.

Outcome measures

Primary outcome measure: Overall survival (OS) of LSCC patients, with 1-, 2-, and 3-year as the predefined time points for all outcome indicator measurements.

Secondary outcome measures: Discriminative ability, calibration degree and predictive stability of the prognostic models, all assessed at the above 1-, 2-, and 3-year OS time points. Discrimination Indicator: The model's discrimination was evaluated using the ROC curve and AUC. The AUC value ranges from 0.5 to 1, with values closer to 1 indicating stronger discrimination. The AUC values for the nomogram model were calculated for the modeling cohort, testing cohort, and external validation cohort, respectively. Calibration index: The calibration curve was used to evaluate the calibration degree of the model. If the calibration curve coincides with the 45-degree diagonal, it indicates that the predicted survival probability is consistent with the actual survival probability.

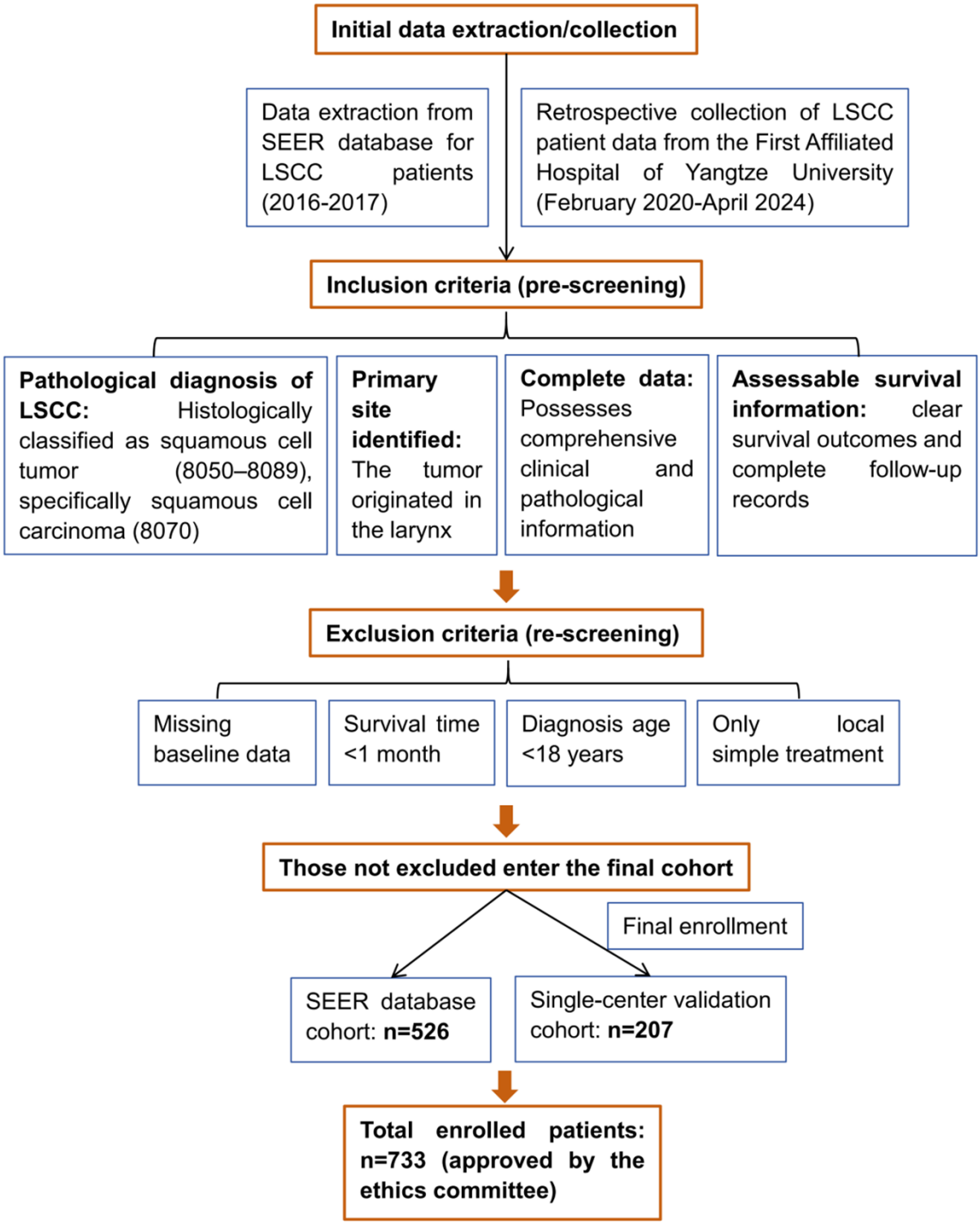

Figure 1. Patient enrollment flowchart.

Model validation: The C-index was calculated through 500 repeated Bootstrap resampling to evaluate the model's discriminative, calibration stability, and accuracy.

Statistical analysis

Measurement data were described as mean ± standard deviation (the rank sum test was used

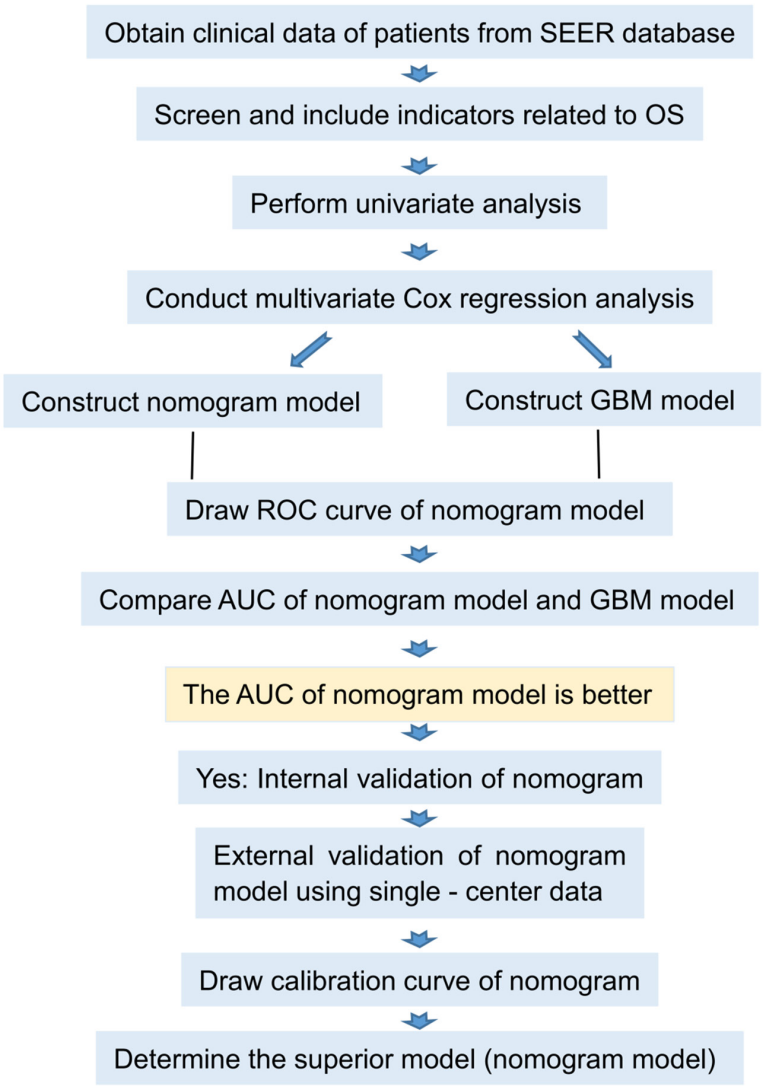

Figure 2. Research flowchart.

for data that do not conform to normal distribution), while count data were presented as case numbers or percentages (Use the chi-square test was used). Statistical analysis was performed using R software (version 4.4.1). A two-tailed *P*-value <0.05 was considered statistically significant.

Univariate and multivariate Cox analysis were conducted to identify prognostic factors independently linked to the OS of LSCC patients. In the univariate analysis, factors with P<0.05 were included in the multivariate analysis. The multivariate Cox analysis adopted a stepwise forward method to identify independent predictive indicators. Univariate Cox proportional hazards regression analysis was used to screen potential prognostic factors for overall survival (OS) of LSCC patients, with hazard ratios (HRs) and 95% confidence intervals (CIs) calculated. Variables with P<0.05 were included in multivariate Cox regression analysis using a stepwise forward method to identify independent prognostic factors. The proportional hazards assumption was verified by Schoenfeld residuals. Adjusted HRs and 95% CIs from the final multivariate model were used for subsequent model construction.

The nomogram and GBM model were constructed using the independent predictors identified through multivariate Cox analysis. Nomogram is an intuitive visualization tool. By converting the regression coefficient of each independent prognostic factor into the corresponding points, then summing up the points of each factor to obtain the total points, and finally predicting the survival probability of patients according to the total points. The ROC curve was utilized to validate the model's performance. The Delong test was employed to compare the AUC differences between the GBM and nomogram models.

Results

Characteristics of LSCC patients

The research flowchart is shown in **Figure 2**. This study encompassed 526 patients sourced from the SEER database and an additional 207 patients from a single center (**Table 1**).

Age: Among the patients in the SEER database, those aged <60 years accounted for 28.14%, those aged 60-75 years accounted for 54.94%, and those aged >75 years accounted for 16.92%. Among the single-center patients,

Table 1. Characteristics of LSCC patients

Items		SEER patients (526 cases)	Single-center patients (207 cases)	χ^2/Z	P
Vital status	Alive	257 (48.86%)	100 (48.31%)	0.018	0.893
	Dead	269 (51.14%)	107 (51.69%)		
Survival months		40.25±22.08	39.58±22.34	-0.319	0.810
Age (years)	<60	148 (28.14%)	64 (30.92%)	73.096	<0.001
	60-75	289 (54.94%)	110 (53.14%)		
	>75	289 (16.92%)	13 (6.28%)		
Sex	Male	396 (75.29%)	135 (65.22%)	7.542	0.006
	Female	130 (24.71%)	72 (34.78%)		
Race	Black	79 (15.02%)	0 (0%)	651.283	<0.001
	White	429 (81.56%)	0 (0%)		
	Other	18 (3.42%)	207 (100%)		
Combined Summary Stage	Localized	236 (44.87%)	102 (49.28%)	5.218	0.074
	Regional	223 (42.40%)	70 (33.82%)		
	Distant	67 (12.74%)	35 (16.91%)		
Grade Recode	I	84 (15.97%)	50 (24.15%)	45.444	<0.001
	I	304 (57.79%)	78 (37.68%)		
	III	135 (25.67%)	64 (30.92%)		
	IV	3 (0.57%)	15 (7.25%)		
Primary Site	320/321	481 (91.44%)	173 (83.57%)	23.225	<0.001
	322/328	223 (4.37%)	30 (14.49%)		
	329	22 (4.18%)	4 (1.93%)		
Tumor Size (mm)		23.76±12.97	24.53±13.07	-0.712	0.563
RX Summ–Surg Prim Site	Operation	71 (13.50%)	156 (75.36%)	265.927	<0.001
	Unoperated	455 (86.50%)	51 (24.64%)		
RX Summ-Scope Reg LN Sur	Yes	15 (2.85%)	38 (18.36%)	53.243	<0.001
	No	511 (97.15%)	169 (81.64%)		
RX Summ–Surg/Rad Seq	Preoperative or postoperative radiotherapy	35 (6.65%)	27 (13.04%)	7.832	0.005
	Others*	491 (93.35%)	180 (86.96%)		
Radiation	Agree	520 (98.86%)	188 (90.82%)	29.133	<0.001
	Turn down	6 (1.14%)	19 (9.18%)		
RX Summ-Systemic/Sur Seq	Systemic therapy and/or surgical procedures	35 (6.65%)	20 (9.66%)	1.936	0.164
	No systemic therapy and/or surgical procedures	491 (93.35%)	187 (90.34%)		
COD	A	401 (76.24%)	147 (71.01%)	2.146	0.143
	B	125 (23.76%)	60 (28.99%)		
First malignant primary indicator	Yes	441 (83.84%)	52 (25.12%)	232.577	<0.001
	No	85 (16.16%)	155 (74.88%)		
Total number of benign/borderline tumors for patient	0	524 (99.62%)	205 (99.03%)	0.940	0.332
	1	2 (0.38%)	2 (0.97%)		
Marital status at diagnosis	status 1	363 (69.01%)	135 (65.22%)	0.982	0.322
	status 2	163 (30.99%)	72 (34.78%)		
Median household income	<$60000	346 (65.78%)	154 (74.4%)	5.086	0.024
	≥$60000	180 (34.22%)	53 (25.60%)		
T stage	T1-2	281 (53.42%)	108 (52.17%)	0.093	0.761
	T3-4	245 (46.58%)	99 (47.83%)		
N stage	N0	340 (64.64%)	124 (59.90%)	19.969	<0.001
	N1	68 (12.93%)	40 (19.32%)		
	N2	114 (21.67%)	33 (15.94%)		
	N3	4 (0.76%)	10 (4.83%)		
M stage	M0	519 (98.67%)	205 (99.03%)	0.163	0.687
	M1	7 (1.33%)	2 (0.97%)		

Staging of disease	I	103 (19.58%)	42 (20.29%)	1.016	0.797
	II	110 (20.91%)	39 (18.84%)		
	III	156 (29.66%)	68 (32.85%)		
	IV	157 (29.85%)	58 (28.02%)		

Other * includes No radiation and/or no surgery; unknown if surgery and/or radiation given; Radiation before and after surgery; Sequence unknown, but both were given. A is relatively favorable or uncertain for prognosis: Alive, Colon excluding Rectum, Hypertension without Heart Disease, Hypopharynx In situ, benign or unknown behavior neoplasms, Kidney and Renal Pelvis, Larynx, Non-Melanoma Skin, Other Cause of Death, Other Diseases of Arteries, Arterioles, Capillaries, Other Oral Cavity and Pharynx, Prostate, State DC not available or state DC available but no COD, Symptoms, Signs and Ill-Defined Conditions, Tongue, Trachea, Mediastinum and Other Respiratory Organ; B is relatively unfavorable to prognosis: Accidents and Adverse Effects, Cerebrovascular Diseases, Chronic Obstructive Pulmonary Disease and Allied, Diseases of Heart, Diabetes Mellitus, Other Infectious and Parasitic Diseases including HIV, Pneumonia and Influenza, Septicemia Esophagus, Lung and Bronchus, Miscellaneous Malignant Cancer, Pancreas, Stomach Suicide and Self-Inflicted Injury. status 1: Married (including common law), Single (never married); status 2: Divorced, Separated Widowed, Unmarried or Domestic Partner. COD: Cause of Death.

those aged <60 years accounted for 30.92%, those aged 60-75 years accounted for 53.14%, and those aged >75 years accounted for 6.28%.

Combined Summary Stage: In the SEER database, localized lesions accounted for 44.87%, regional lesions accounted for 42.40%, and distant metastases accounted for 12.74%. Among the single-center patients, the corresponding proportions were 49.28%, 33.82%, and 16.91%, respectively.

Tumor size: The mean tumor size of patients in the SEER database was 23.76 ± 12.97 mm, and that of single-center patients was 24.53 ± 13.07 mm.

Radiation: In the SEER database, 98.86% of patients received radiation therapy, and 1.14% refused. Among the single-center patients, the corresponding proportions were 90.82% and 9.18%, respectively.

First malignant primary indicator: In the SEER database, 83.84% of patients had a first malignant primary tumor, and 16.16% did not. Among the single-center patients, the proportions were 25.12% and 74.88%, respectively.

T Stage: Among the patients in the SEER database, 53.42% were in T1-2, and 46.58% were in T3-4. Among the single-center patients, 52.17% were in T1-2, and 47.83% were in T3-4.

N stage: Among the patients in the SEER database, 64.64% were in N0, 12.93% were in N1, 21.67% were in N2, and 0.76% were in N3. Among the single-center patients, 59.90% were in N0, 19.32% were in N1, 15.94% were in N2, and 4.83% were in N3.

Staging of disease: Among the patients in the SEER database, stage I accounted for 19.58%, stage II accounted for 20.91%, stage III accounted for 29.66%, and stage IV accounted for 29.85%. Among the single-center patients, stage I accounted for 20.29%, stage II accounted for 18.84%, stage III accounted for 32.85%, and stage IV accounted for 28.02%.

COD (Cause of Death): In the SEER database, category A accounted for 76.24%, and category B accounted for 23.76%. Among the single-center patients, category A accounted for 71.01%, and category B accounted for 28.99%.

Category A (relatively good prognosis or uncertain): Alive, Colon excluding Rectum, Hypertension without Heart Disease, Hypopharynx In situ, benign or unknown behavior neoplasms, Kidney and Renal Pelvis, Larynx, Non-Melanoma Skin, Other Cause of Death, Other Diseases of Arteries, Arterioles, and Capillaries, Other Oral Cavity and Pharynx, Prostate, State DC not available or state DC available but no COD, Symptoms, Signs and Ill-Defined Conditions, Tongue, Trachea, Mediastinum and Other Respiratory Organ.

Category B (relatively poor prognosis): Accidents and Adverse Effects, Cerebrovascular Diseases, Chronic Obstructive Pulmonary Disease and Allied, Diseases of Heart, Diabetes Mellitus, Other Infectious and Parasitic Diseases including HIV, Pneumonia and Influenza, Septicemia Esophagus, Lung and Bronchus, Miscellaneous Malignant Cancer, Pancreas, Stomach Suicide and Self-Inflicted Injury.

Marital status at diagnosis: In the SEER database, status 1 accounted for 69.01%, and status 2 accounted for 30.99%. Among the single-center patients, the corresponding proportions were 65.22% and 34.78%, respectively.

Status 1: Married (including common law), Single (never married); Status 2: Divorced, Separated Widowed, Unmarried or Domestic Partner.

Table 2. Univariate Cox analysis

Items		HR	95.0% CI	*P*
Age	<60			<0.001
	60-75	1.431	1.052-1.946	0.023
	>75	2.808	1.962-4.018	<0.001
Sex		1.248	0.937-1.664	0.130
Race	Black			0.732
	White	0.898	0.645-1.249	0.522
	Other	0.765	0.343-1.703	0.511
Combined Summary Stage	Localized			<0.001
	Regional	1.634	1.250-1.703	<0.001
	Distant	2.959	2.091-4.187	<0.001
Grade	I			0.123
	II	1.405	0.969-2.036	0.073
	III	1.633	1.091-2.446	0.017
	IV	1.765	0.424-7.35	0.435
Primary Site	320/321			0.114
	322/328	1.706	1.012-2.875	0.045
	329	1.219	0.683-2.177	0.503
Tumor Size		1.03	1.020-1.039	<0.001
RX Summ-Surg Prim Site		1.017	0.719-1.439	0.923
RX Summ-Scope Reg LN Sur		1.132	0.560-2.287	0.730
RX Summ-Surg/Rad Seq		1.053	0.741-1.496	0.772
Radiation		3.829	1.580-9.283	0.003
RX Summ-Systemic/Sur Seq		0.802	0.514-1.253	0.333
COD		4.585	3.582-5.869	<0.001
First malignant primary indicator		1.446	1.072-1.948	0.016
Total number of benign/borderline tumors for patient		0.898	0.126-6.398	0.914
Marital status at diagnosis		1.42	1.105-1.823	0.006
Median household income		0.995	0.773-1.280	0.967
T stage		1.845	1.450-2.349	<0.001
N stage	N0			<0.001
	N1	1.809	1.285-2.547	0.001
	N2	1.808	1.369-2.390	<0.001
	N3	1.432	0.355-5.783	0.614
M stage		1.457	0.543-3.911	0.455
Staging of disease	I			<0.001
	II	1.506	0.974-2.329	0.066
	III	1.93	1.293-2.882	0.001
	IV	2.892	1.962-4.265	<0.001

COD: Cause of Death.

Univariate and multivariate analyses

The univariate Cox analysis results, derived from the SEER database, indicated that variables including Age, Combined Summary Stage, Tumor Size, Radiation, COD, First Malignant Primary Indicator, Marital Status at Diagnosis, T Stage, N Stage, and Edition Stage Group Recode were significantly correlated with the OS of LSCC patients. (P<0.05, **Table 2**). The outcomes of the multivariate Cox analysis revealed that predictors such as Age, Tumor Size, Radiation, COD, Marital Status at Diagnosis, and T Stage were independent predictors influencing the OS of LSCC patients (P<0.05, **Table 3**).

Table 3. Multivariate Cox analysis

Items	Coding/Reference category	HR	95.0% CI	*P*
Age (years)	Reference: <60			<0.001
	60-75	1.333	0.961-1.849	0.085
	>75	2.726	1.836-4.046	<0.001
Combined Summary Stage	Reference: Localized			0.182
	Regional	0.795	0.506-1.249	0.320
	Distant	1.121	0.630-1.995	0.697
Tumor Size (mm)	Reference: continuous variable	1.013	1.001-1.025	0.035
Radiation	Reference: No			<0.001
	Yes	7.555	2.947-19.370	
COD	Reference: A			<0.001
	B	3.996	3.083-5.179	
First malignant primary indicator	Reference: No			0.094
	Yes	1.308	0.955-1.792	
Marital status at diagnosis	Reference: Status 2			0.006
	Status 1	1.444	1.111-1.876	
T stage	Reference: T1-2			0.017
	T3-4	1.652	1.093-2.495	
N stage	Reference: N0			0.084
	N1	1.758	1.122-2.755	0.014
	N2	1.633	0.918-2.907	0.095
	N3	1.650	0.366-7.437	0.514
Staging of disease	Reference: I			0.144
	II	0.295	0.061-1.420	0.128
	III	0.275	0.076-1.003	0.051
	IV	0.242	0.050-1.179	0.079

COD: Cause of Death.

Construction of the nomogram model

Utilizing the independent predictive indicators derived from the multivariate Cox analysis of patients in the SEER database, a Nomogram model (**Figure 3**) and a risk forest plot (**Figure 4**) were constructed using the modeling cohort.

This nomogram integrates variables such as age, tumor size, and radiotherapy, along with their corresponding scores. The total score is the sum of the scores for each variable, and the 1-, 2-, and 3-year OS rates can be derived from the corresponding total score.

In the forest plot, each factor corresponds to the estimated hazard ratio (HR) and 95% confidence interval (95% CI). For example, for the age factor, the model-derived sample size was N=366, with an HR of 1.8, with a 95% CI of (1.38-2.3) and *P*<0.001, indicating that age is a significant factor influencing the risk outcome.

Construction of the GBM model

When building the GBM model, Cox partial deviance was tracked to assess how model performance changed with the number of iterations. As depicted in **Figure 5**, the Cox partial deviance decreased as the number of iterations increased.

The green curve represents the Cox partial deviance during the model training process. It decreased rapidly in the early stage of iterations and leveled off at approximately 500 iterations (referenced by the blue vertical line in the figure), indicating that the model had basically converged at this point and subsequent iterations had limited impact on improving the model performance.

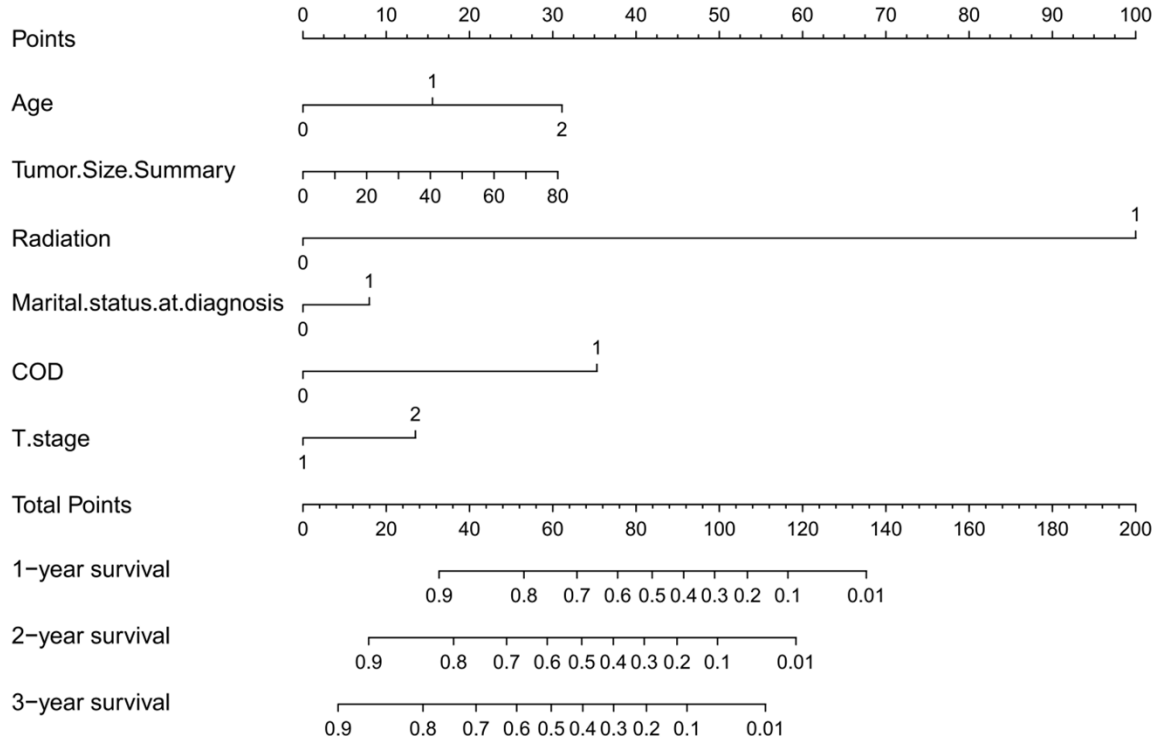

Figure 3. Nomogram based on the modeling cohort. COD: Cause of Death.

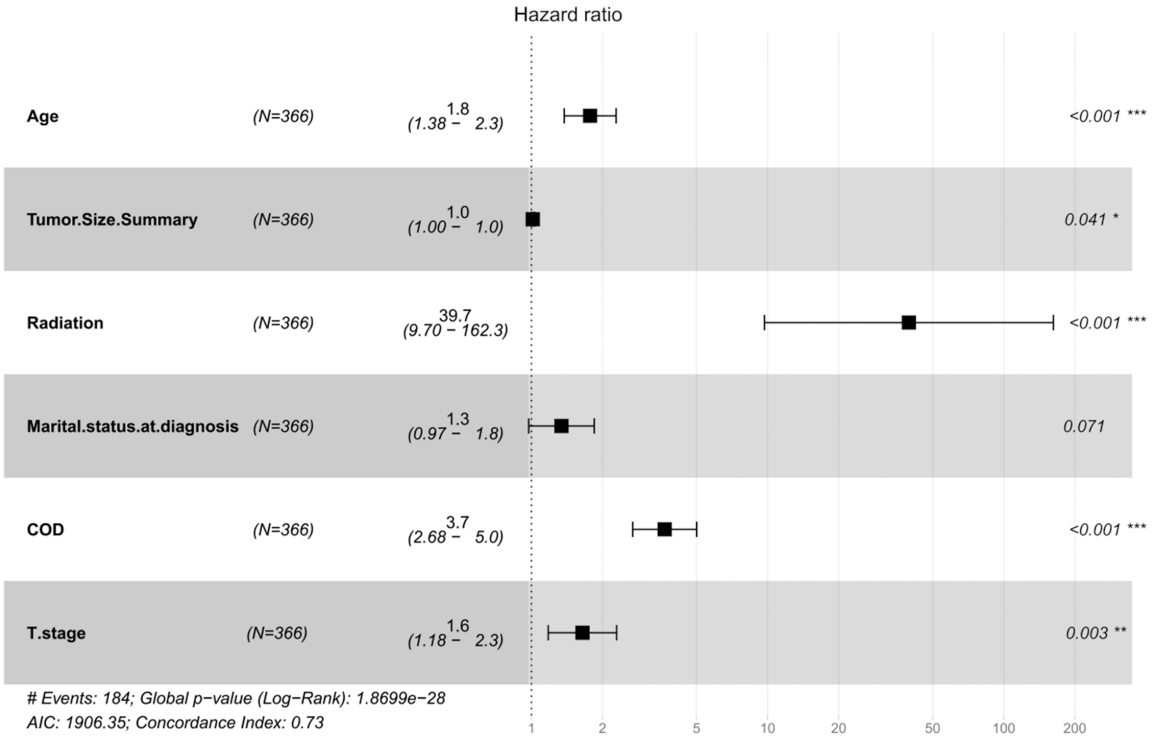

Figure 4. Risk forest plot based on the modeling cohort. COD: Cause of Death.

The black curve represent the Cox partial deviance of the model on the testing cohort. Its downward trend was relatively gentle and gradually stabilized after a certain number of itera-

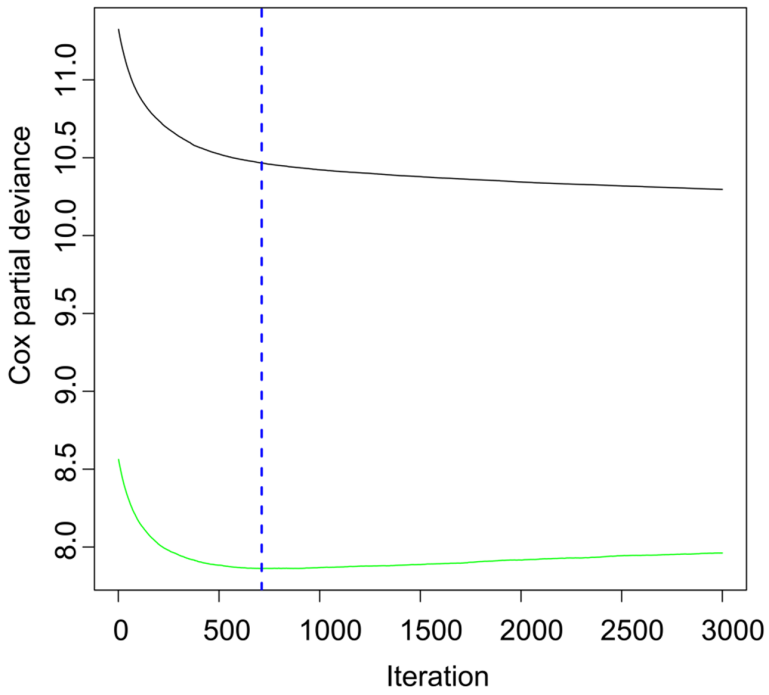

Figure 5. The Gradient Boosting Machine (GBM) model.

tions, reflecting the performance change of the model during the generalization process.

Evaluation of the predictive efficacy of the model

In **Figure 6**, the ROC curves showed that the nomogram model had higher AUC values than the GBM model at 1-, 2-, and 3-year follow-up time points. The nomogram's AUC values were 0.809, 0.782, and 0.811 in the modeling cohort; 0.783, 0.786, and 0.801 in the testing cohort; and 0.795, 0.760, and 0.783 in the external validation cohort. The GBM model's AUC values were 0.747, 0.763, and 0.785. The Delong test confirmed that the nomogram model's AUC values were obviously higher than those of the GBM model ($P<0.05$).

Evaluation of the calibration of the Nomogram model

The modeling cohort's C-index was 0.726, suggesting the model can effectively differentiate samples with varying risk levels and has strong discriminative power. The C-index for the test cohort was 0.803, while the external validation cohort recorded a C-index of 0.780.

Figure 7 illustrates the calibration of the predicted 1-year survival rate versus the actual survival rate in the model. The calibration line closely aligns with the 45° diagonal, indicating a high degree of consistency between the model's predictions and the actual survival outcomes.

Modeling plot: The scatter points and line of regression approach the 45° diagonal, indicating that the Nomogram model's prediction of 1-year OS rate for LSCC patients has a certain degree of agreement with the actual data in the training cohort, demonstrating good fit between the model and the modeling data.

Test dataset: The scatter points and line plots are close to the 45° diagonal, indicating that the model maintains good predictive accuracy in the test dataset.

External validation: The scatter plot and the approximate line-diagonal relationship indicate that the model has extrapolation capability in predicting patient survival rates from external single-center data.

Discussion

In recent years, with the advancement of big data and machine learning technologies, the construction of prognostic prediction models based on data warehouses has become a research hotspot [5]. This study established a logistic regression (LR) model for the prognosis of LSCC using the SEER data warehouse and validated it with single-center data, aiming to provide a reliable prognostic assessment tool for clinical practice.

Univariate and multivariate cox regression analyses

In this research, various factors independently linked to the OS of LSCC patients were identified via univariate analysis, including Age, Combined Summary Stage, and Tumor Size. Additional multivariate analysis indicated that Age, Tumor Size, Radiation, COD, Marital Status at Diagnosis, and T Stage serve as independent factors affecting OS. Consistent with previous studies [6, 7], our findings confirm that

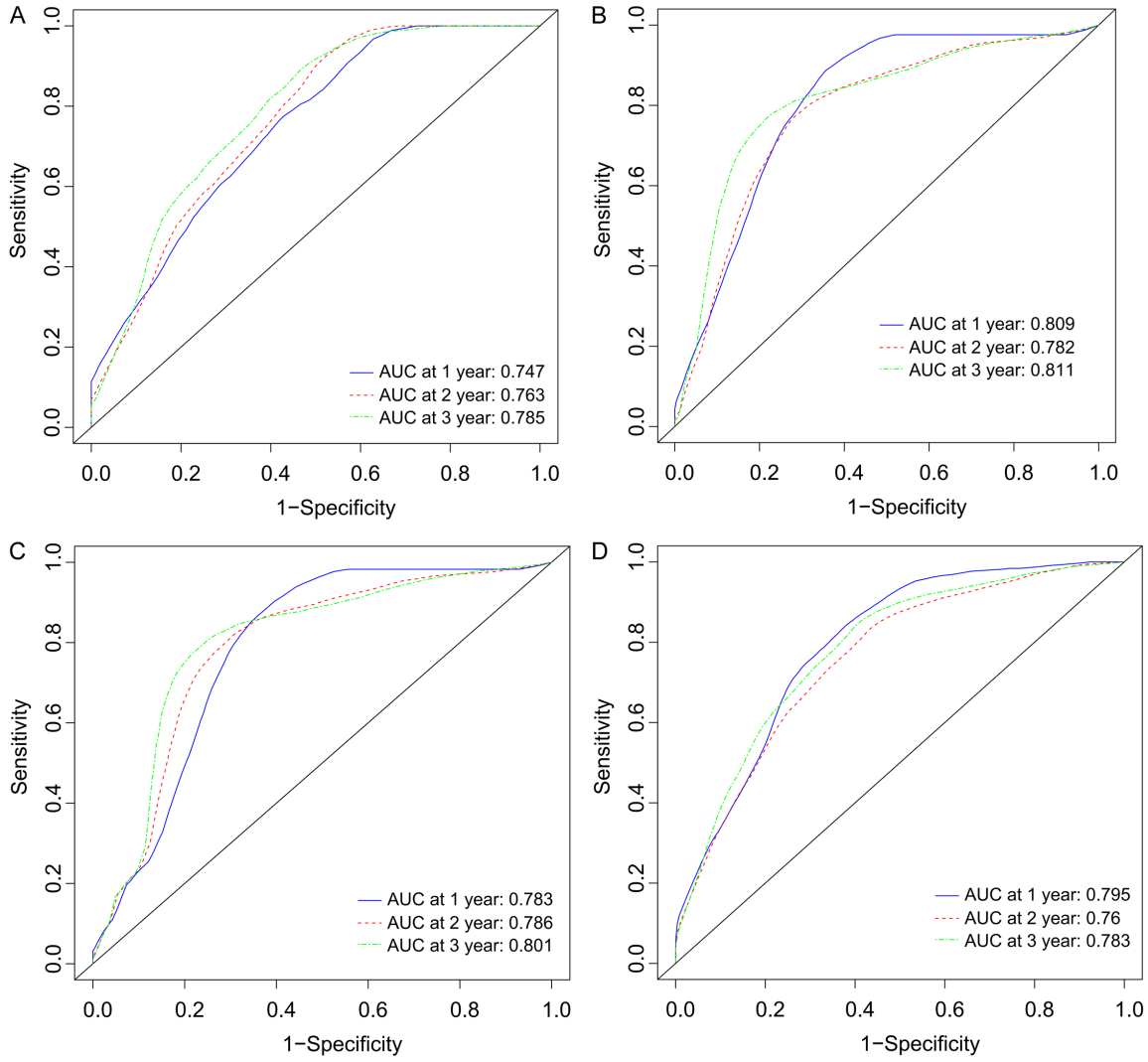

Figure 6. Receiver operating characteristic (ROC) curves for the predictive performance of the models. A. ROC curve of the GBM model; B. ROC curve of the nomogram model; C. ROC curve of the nomogram model in the testing cohort; D. ROC curve of the nomogram model in the external validation cohort.

age and tumor-related indicators (T stage, tumor size) are core prognostic factors for LSCC, which further validates the rationality of our variable selection and the reliability of the study results. As an important prognostic factor, age may be associated with the patients' physical function, treatment tolerance, and tumor biological behavior. With increasing age, the immune function of patients gradually declines, which may render tumors more prone to progression and metastasis [8]. Tumor size directly reflects tumor burden, and larger tumors often have higher invasiveness and metastatic potential [9]. Whether radiation therapy is used and how it is administered notably affect patient outcomes. An appropriate radiation therapy regimen can effectively control tumor growth and improve patients' survival rates [10]. COD is closely related to tumor malignancy and the treatment efficacy. Marital status reflects patients' social support systems to a certain extent. Good social support may help patients better cope with the disease and treatment process, thereby improving the outcomes [11]. The T stage directly indicates the extent of local tumor invasion and is a crucial factor in assessing outcomes.

Construction and evaluation of the Nomogram model

The Nomogram model, developed using the independent predictive indicators from the multivariate analysis, offers clinicians a visual

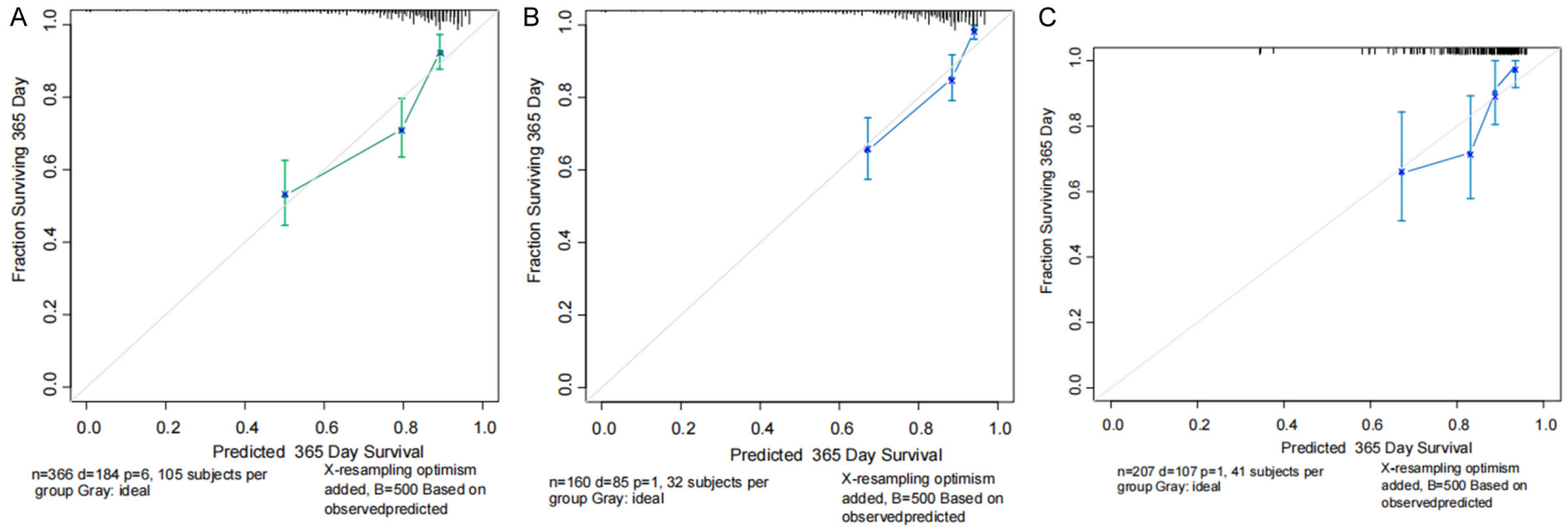

Figure 7. Calibration curves for the predictive performance of the nomogram model. A. Modeling cohort; B. Testing cohort; C. External validation cohort.

and convenient tool for prognostic assessment. Through the Nomogram, clinicians can calculate the total score according to the specific characteristics of patients, and then predict the 1 to 3-year survival rates of patients. The risk forest plot clearly shows the hazard ratios of each factor, enabling clinicians to intuitively understand the influence degree of each factor on the prognosis of patients.

The GBM model and the nomogram model exhibited different performances. The AUC of the GBM model at 1-, 2-, and 3-year were 0.747, 0.763, and 0.785. The AUC values of the nomogram model in the 1, 2, and 3 years in the modeling cohort were 0.809, 0.782, and 0.811, respectively. Based on the Delong tests, the AUC values of the nomogram model at each time point were significantly greater than those of the GBM model, demonstrating that the former had superior predictive performance compared to the latter. Notably, the predictive accuracy of our nomogram (AUC 0.783-0.811) is consistent with or slightly higher than that of Nomogram models reported in previous LSCC studies [12, 13], which confirms the validity of our model. The slight advantage may be attributed to our combined use of SEER database and single-center data, which reduces selection bias and enhances the representativeness of the study population. Results of further testing illustrated the AUC values of the testing cohort of the nomogram model at 1-, 2-, and 3-year were 0.783, 0.786, and 0.801. According to the internal testing data, the nomogram model can stably predict occurrence of events at different time points, suggesting a certain degree of reliability. The AUC values of the nomogram model at 1-, 2-, and 3-year on the external validation cohort were 0.795, 0.760, and 0.783, respectively. There were some differences in values from the testing cohort. These values from the testing cohort, however, still remained at a high level. This indicates that the nomogram model also demonstrates a high level of generalization ability in external samples independent of the modeling data, with a good ability to predict event outcomes over 1-3 years, further confirming the practical effectiveness and robustness of the nomogram model. Combined with the results of the modeling cohort and the comparison with the GBM model (the AUC in the modeling cohort was higher, $P<0.05$), it fully demonstrates the superiority of the nomogram model in terms of predictive performance.

The calibration curves showed that in the modeling cohort, testing cohort, and external validation cohort of the nomogram model, the calibration between the predicted 1-year OS rate by the model and the actual survival rate was good. The scatter points and the broken lines were generally well-aligned with the 45° diagonal, indicating that the model has good calibration and reliability in predicting the 1-year OS rate of patients in different datasets. This means that the survival rate predicted by the model is relatively close to the actually observed survival rate, which can provide clinicians with relatively accurate prognostic information.

Comparison with relevant foreign studies

In recent years, numerous foreign studies have also been dedicated to constructing prognostic prediction models for LSCC. For example, Hu et al. [12] constructed a prognostic model for LSCC based on the data of a large cancer center, using multivariate analysis and machine learning algorithms, and verified the model through internal and external validation cohorts. This study is similar to ours in that both focus on the accuracy and generalization ability of the model. However, there are differences in the included prognostic factors. Li et al. [14] included some gene expression data as prognostic factors, whereas our study constructed primarily the model based on clinicopathological indicators. The difference in included factors is mainly due to the data accessibility: gene expression data are not routinely recorded in the SEER database, and its high detection cost limits its clinical application, which explains why our model prioritizes easily accessible clinical-pathological indicators to ensure practicality, consistent with the clinical needs of primary medical institutions. Although gene expression data can more deeply reflect the biological characteristics of tumors, the high cost of acquisition and detection limits its popularization in routine clinical applications. In contrast, the model of our study is based on easily accessible clinical and pathological information, making it more clinically practical. Wang et al. [15] adopted a different modeling method, using deep learning algorithms to analyze the imaging data of LSCC

patients and constructed a prediction model. This model has achieved good results in predicting the recurrence and metastasis of tumors. However, the interpretation of imaging data are subjective, and it has high requirements for equipment and technology. The Nomogram model of our study does not rely on special imaging equipment and can conduct prognostic assessment through simple clinical data, making it more suitable for popularization and application in primary medical institutions.

Analysis of the mechanisms underlying prognostic factors

In terms of prognostic factors, there are some overlaps but also differences between foreign studies and this study. Most foreign studies regard factors such as age and tumor staging as important prognostic factors, which aligns with the findings of this study [6]. However, some foreign studies [16] also emphasize the influence of the human papillomavirus (HPV) infection status on the prognosis of LSCC. LSCC patients with a positive HPV status often have a better prognosis, and the mechanism may be associated with the changes in the biological behavior of tumor cells caused by HPV infection. In this study, due to the limitations of the data source, the HPV infection status was not analyzed. Future studies can further incorporate factors such as the HPV infection status to improve the prognostic prediction model for LSCC.

The impact of age on the prognosis of LSCC patients is complicated. As people age, their body immune system gradually declines which includes the reduced function of T and B cells and his immune surveillance ability also gets weakened [17, 18]. This allows tumor cells to evade attack by the immune system of the organism and facilitates tumor growth and metastasis. Older adult patients, in addition, often have multiple chronic diseases that coexist. Such diseases may be heart disease, diabetes, etc. The presence of comorbidities may affect patients' tolerance to treatment modalities, such as surgery, radiotherapy and chemotherapy and, consequently, prognosis [19]. The relationship between tumor size and prognosis is primarily because of its invasive and metastatic abilities [20]. Larger tumors might have higher microvascular density [21], which supplies more nutrients and pathways for metastasis for the tumor cells. The proliferation rate of tumor cells is equivalent to the size of the tumor. Cells in larger tumors divide more rapidly compared to smaller tumors. This makes it easier for their cells to break through the basement membrane, penetrate through tissues, and metastasize further to other organs via the lymphatic and hematogenous pathways [22].

The mechanisms by which radiotherapy has a positive impact on the prognosis of LSCC patients mainly include directly killing tumor cells and activating the body's immune response. Radiotherapy destroys the DNA of tumor cells through high-energy rays, leading to the death of tumor cells [23]. Furthermore, radiotherapy can induce tumor cells to release tumor-associated antigens, activate the immune system, and enhance the ability of immune cells to kill tumor cells [24]. Radiotherapy can alter the tumor microenvironment and inhibit tumor angiogenesis, thereby suppressing tumor growth and metastasis [25].

The impact of marital status on the prognosis of LSCC patients may be associated with social support. Married patients typically receive more emotional support, financial support, and daily care. These supports can help patients better cope with the psychological stress and physical discomfort caused by the disease and improve the patients' treatment adherence [13]. At the same time, good social support can also regulate the patients' immune system and enhance the body's resistance to tumors.

Research limitations

While this study has yielded certain outcomes, it still has some limitations. Firstly, this study is mainly based on retrospective data, and it has the inherent limitations of retrospective studies, such as data missing and information bias. Although some measures were taken during the research process to minimize the influence of these biases, they cannot be completely eliminated. Secondly, due to the limitations of data sources, this study did not include certain factors that may influence the prognosis of LSCC, such as HPV infection status and molecular biological characteristics of the tumor. These factors warrant further exploration in future research. Additionally, the sample size of the external validation set in this study was rel-

atively small, which may affect the generalizability and applicability of the model in larger populations. It is necessary to further expand the sample size of the external validation set in future studies to verify the reliability and generalization ability of the model.

Conclusions

The prognostic nomogram model for LSCC constructed has significant clinical application value and great potential for widespread promotion and application, bringing positive impacts on the precise treatment and prognostic assessment of LSCC patients.

Disclosure of conflict of interest

None.

Address correspondence to: Yan Li, Department of Oncology, The First Affiliated Hospital of Yangtze University, No. 55, Jianghan North Road, Shashi District, Jingzhou 434000, Hubei, China. Tel: +86-0716-8191560; E-mail: oncoliyanjz@163.com

References

[1] Mur T, Jaleel Z, Agarwal P, Edwards H and Levi JR. Paediatric laryngeal squamous cell carcinoma: systematic review and pooled analysis. Clin Otolaryngol 2021; 46: 494-500.

[2] Sun M, Chen S and Fu M. Model establishment of prognostic-related immune genes in laryngeal squamous cell carcinoma. Medicine (Baltimore) 2021; 100: e24263.

[3] Cui J, Wang L, Zhong W, Chen Z, Chen J, Yang H and Liu G. Development and validation of epigenetic signature predict survival for patients with laryngeal squamous cell carcinoma. DNA Cell Biol 2021; 40: 247-264.

[4] Forsyth AM, Camilon PR, Tracy L and Levi JR. Pediatric laryngeal tumors and demographics, management, and survival in laryngeal squamous cell carcinoma. Int J Pediatr Otorhinolaryngol 2021; 140: 110507.

[5] Wang J, Lun L, Jiang X, Wang Y, Li X, Du G and Wang J. APE1 facilitates PD-L1-mediated progression of laryngeal and hypopharyngeal squamous cell carcinoma. Int Immunopharmacol 2021; 97: 107675.

[6] Shaikh N, Morrow V, Stokes C, Chung J, Fancy T, Turner MT and Stokes WA. Factors associated with a prolonged diagnosis-to-treatment interval in laryngeal squamous cell carcinoma. Otolaryngol Head Neck Surg 2022; 166: 1092-1098.

[7] Ghosh S, Kumar S, Chaudhary R and Guha P. High-risk Human papillomavirus infection in squamous cell carcinoma of the larynx: a study from a tertiary care center in North India. Cureus 2023; 15: e34760.

[8] Ikeda H and Togashi Y. Aging, cancer, and antitumor immunity. Int J Clin Oncol 2022; 27: 316-322.

[9] Zhang A, Xu Y, Xu H, Ren J, Meng T, Ni Y, Zhu Q, Zhang WB, Pan YB, Jin J, Bi Y, Wu ZB, Lin S and Lou M. Lactate-induced M2 polarization of tumor-associated macrophages promotes the invasion of pituitary adenoma by secreting CCL17. Theranostics 2021; 11: 3839-3852.

[10] Carvalho GB, Kohler HF, Lira RB, Vartanian JG and Kowalski LP. Survival results of 3786 patients with stage I or II laryngeal squamous cell carcinoma: a study based on a propensity score. Braz J Otorhinolaryngol 2022; 88: 337-344.

[11] Rani B, Sinha AP, Sharma KK, Prasad BV, Srinivasan M and Biswas A. Perception of social support and prevalence of self-reported depressive symptoms among patients with head-and-neck squamous cell carcinoma treated at a tertiary cancer centre in North India. Indian J Palliat Care 2024; 30: 336-341.

[12] Hu WM and Jiang WJ. A prognostic model for laryngeal squamous cell carcinoma based on the mitochondrial metabolism-related genes. Transl Cancer Res 2025; 14: 966-979.

[13] Chen Q, Hu Y, Lin W, Huang Z, Li J, Lu H, Dai R and You L. Studying the impact of marital status on diagnosis and survival prediction in pancreatic ductal carcinoma using machine learning methods. Sci Rep 2024; 14: 5273.

[14] Li Z, Cai H, Zheng J, Chen X, Liu G, Lv Y, Ye H and Cai G. Mitochondrial-related genes markers that predict survival in patients with head and neck squamous cell carcinoma affect immunomodulation through hypoxia, glycolysis, and angiogenesis pathways. Aging (Albany NY) 2023; 15: 10347-10369.

[15] Wang W, Liang H, Zhang Z, Xu C, Wei D, Li W, Qian Y, Zhang L, Liu J and Lei D. Comparing three-dimensional and two-dimensional deep-learning, radiomics, and fusion models for predicting occult lymph node metastasis in laryngeal squamous cell carcinoma based on CT imaging: a multicentre, retrospective, diagnostic study. EClinicalMedicine 2024; 67: 102385.

[16] Yang SM, Wu M, Han FY, Sun YM, Yang JQ and Liu HX. Role of HPV status and PD-L1 expression in prognosis of laryngeal squamous cell carcinoma. Int J Clin Exp Pathol 2021; 14: 107-115.

[17] Song C, Pan W, Brown B, Tang C, Huang Y, Chen H, Peng N, Wang Z, Weber D, Byrne-Steele M, Wu H, Liu H, Deng Y, He N and Li S.

Immune repertoire analysis of normal Chinese donors at different ages. Cell Prolif 2022; 55: e13311.

[18] Lian J, Liu J, Yue Y, Li F, Chen X, Zhang Z, Ping Y, Qin G, Li L, Zhang K, Liu S, Zhang L, Qiao S, Liu N, Zheng Y, Wu J, Zeng Q and Zhang Y. The repertoire features of T cell receptor β-chain of different age and gender groups in healthy Chinese individuals. Immunol Lett 2019; 208: 44-51.

[19] David KA, Sundaram S, Kim SH, Vaca R, Lin Y, Singer S, Malecek MK, Carter J, Zayac A, Kim MS, Reddy N, Ney D, Habib A, Strouse C, Graber J, Bachanova V, Salman S, Vendiola JA, Hossain N, Tsang M, Major A, Bond DA, Agrawal P, Mier-Hicks A, Torka P, Rajakumar P, Venugopal P, Berg S, Glantz M, Goldlust SA, Folstad M, Kumar P, Ollila TA, Cai J, Spurgeon S, Sieg A, Cleveland J, Chang J, Epperla N, Karmali R, Naik S, Martin P, Smith SM, Rubenstein J, Kahl B and Evens AM. Older patients with primary central nervous system lymphoma: survival and prognostication across 20 U.S. cancer centers. Am J Hematol 2023; 98: 900-912.

[20] Zheng L, Liang H, Zhang Q, Shen Z, Sun Y, Zhao X, Gong J, Hou Z, Jiang K, Wang Q, Jin Y and Yin Y. circPTEN1, a circular RNA generated from PTEN, suppresses cancer progression through inhibition of TGF-β/Smad signaling. Mol Cancer 2022; 21: 41.

[21] Izumo W, Kawaida H, Saito R, Nakata Y, Amemiya H, Maruyama S, Takiguchi K, Shoda K, Shiraishi K, Furuya S, Kawaguchi Y, Mochizuki K, Kondo T and Ichikawa D. Evaluation of the details and importance of lymphatic, microvascular, and perineural invasion in patients with non-functioning pancreatic neuroendocrine neoplasms based on tumor size and the 2022 World Health Organization classification: a 23-year retrospective analysis. World J Surg Oncol 2025; 23: 79.

[22] Zhang C, Wang Y, Zhen Z, Li J, Su J and Wu C. mTORC1 Mediates biphasic mechano-response to orchestrate adhesion-dependent cell growth and anoikis resistance. Adv Sci (Weinh) 2024; 11: e2307206.

[23] Sengar D, Pathan NS and Gajbhiye V. D-bait: a siDNA for regulation of DNA-protein kinases against DNA damage and its implications in cancer. Int J Pharm 2025; 673: 125416.

[24] Chen L, Zhang R, Lin Z, Tan Q, Huang Z and Liang B. Radiation therapy in the era of immune treatment for hepatocellular carcinoma. Front Immunol 2023; 14: 1100079.

[25] Guo J, Guo J, Cheng B, Sun X, Zhang H and Ma J. Synergistic effect of stereotactic radiotherapy combined with Karelizumab on patients with advanced NSCLC. J Healthc Eng 2022; 2022: 7875627.