Heliyon 10 (2024) e30860

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Prognostic factors and novel prediction models for overall survival of patients with submandibular gland cancer: A population-based retrospective cohort study

Shan-shan Yang, MD [a,1], Xiong-gang Yang, PhD [b,1], Xiao-hong Yang, PhD [a,**], Xiao-hua Hu, MD [c,*]

[a] Department of Prosthodontics, Affiliated Stomatological Hospital of Zunyi Medical University, Zunyi Medical University, Zunyi, China
[b] Department of Orthopaedics, The First People's Hospital of Yunnan Province, Affiliated Hospital of Kunming University of Science and Technology, Kunming 650032, China
[c] Department of Oral and Maxillofacial Surgery, Affiliated Stomatological Hospital of Zunyi Medical University, Zunyi Medical University, Zunyi, China

ARTICLE INFO

Keywords:
Submandibular gland cancer
Nomogram
Overall survival
SEER database
Prognostic factors

ABSTRACT

Background: Accurately predicting the survival rate of submandibular gland cancer (SGC) is of significant importance for guiding treatment decision-making and improving patient outcomes. This study was aimed to identify the independent prognostic factors of overall survival (OS) in SGC patients, and develop novel prediction models to aid clinicians in predicting the survival probability.

Materials and methods: Patients diagnosed with primary SGC after the year 2010 were extracted from SEER database and then randomly allocated into training and test samples in a 7:3 ratio. Uni- and multi-variable COX analyses were employed using the training sample to ascertain independent prognostic factors for OS. Subsequently, graphic and online dynamic nomograms were established basing on the independent prognostic factors. We utilized C-index, calibration curve, receiver operating characteristic (ROC) curve, and area under ROC curve (AUC) value to evaluate the discrimination capacity and the consistency between predicted and actual survival.

Results: A total of 527 SGC patients were included (369 assigned to training group and 158 assigned to test group). The multivariable COX analysis showed that age, sex, marital status, tumor histology, summary stage, metastases to bone, and tumor size were independently associated with OS. Novel graphical and online dynamic (URL: https://yangxg1209.shinyapps.io/overall_survival_submandibular_gland_tumor/) nomograms were established. The C-indices (training: 0.77, 95%CI 0.71–0.84; test: 0.77, 95%CI 0.68–0.85) indicate favorable discrimination ability of the model, and the calibration curves demonstrated favorable consistency between the predicted and actual survival rates.

* Corresponding author. Department of Oral and Maxillofacial Surgery, Affiliated Stomatological Hospital of Zunyi Medical University, Zunyi Medical University; No. 143, Dalian Road, Huichuan District, 563000, Zunyi City, Guizhou Province, China.
** Corresponding author. Department of Prosthodontics, Affiliated Stomatological Hospital of Zunyi Medical University, Zunyi Medical University; No. 143, Dalian Road, Huichuan District, 563000, Zunyi City, Guizhou Province, China.
E-mail addresses: 854936065@qq.com (X.-h. Yang), hxh2132000@163.com (X.-h. Hu).
[1] these authors contributed equally to this study.

https://doi.org/10.1016/j.heliyon.2024.e30860
Received 8 November 2023; Received in revised form 12 April 2024; Accepted 7 May 2024
Available online 9 May 2024

Conclusions: Our study identified the independent prognostic factors influencing OS in patients with SGC, and successfully established and validated novel nomograms, which provide accurate prediction of survival rates and allows for personalized risk assessment.

1. Introduction

Salivary gland cancer is a rare malignant tumor of the head and neck, accounting for only 3%–5% of head and neck cancers [1–3]. Among them, approximately 80 % of salivary gland cancers occur in the parotid gland, while only 5–15 % [1,3–7] are located in the submandibular gland. Compared to parotid gland cancer, submandibular gland cancer (SGC) is more aggressive and has a higher degree of malignancy. Therefore, the prognosis of SGC is generally poorer, and treatment may be more challenging. Treatment for SGC typically involves collaboration among a multidisciplinary team, including otolaryngologists, radiation oncologists, and pathologists. The choice of treatment modality depends on various factors such as the pathological type of the tumor, tumor size, extension of tumor, patients' overall performance status, and exist of distant metastasis [8]. Surgical resection is often the preferred treatment method for SGC, with the goal of removing as much tumor tissue as possible. The surgical extent may include the submandibular gland itself, surrounding lymph nodes, and affected muscles and bone structures. Radiation therapy can be used as the primary treatment or in conjunction with surgery. In cases where surgical resection is not feasible, radiation therapy may be the only option. Chemotherapy is commonly administered in advanced stages of SGC or in combination with radiation therapy, either as neoadjuvant (preoperative) or adjuvant (postoperative) treatment. Chemotherapy helps reduce tumor volume, increase surgical resectability, and kill any residual cancer cells following surgery and radiation. The overall goal of treating SGC is to extend survival time and improve quality of life (QoF). Although there is now a well-defined consensus regarding the treatment strategy of SGCs, each patient presents individual differences. Personalized treatment plans must be tailored to the specific disease state of each patient. For example, in choosing surgical strategies, more aggressive procedures may be selected for patients with longer life expectancies, involving extensive resection and combining multiple adjuvant therapies to achieve better local control and longer survival. Conversely, for patients with very short life expectancies, excessively invasive treatment approaches may offer little additional benefit. These patients should prioritize symptom relief and QoF, opting for more limited surgical interventions or palliative/conservative symptomatic treatment to avoid the significant impact on their overall condition caused by invasive surgery and excessive adjuvant treatments.

Thus, accurately predicting the overall survival of SGC is of significant importance for guiding personalized treatment decision-making and improving patient outcomes. The overall survival of SGC is influenced by multiple factors, with common independent risk factors including tumor stage, lymph node metastasis, tumor size and extent of invasion, tumor differentiation, vascular and nerve invasion, age, and others [9]. Denis et al. [10] explored the prognostic factors and treatment outcomes in a cohort of 40 SGC patients, showing that nodal status, high-grade histology, tumor stage and adjuvant radiation-therapy were significant variables for overall survival and treatment outcomes. Bhattacharyya et al. [11] identified the factors affecting survival in 370 patients with SGC, and younger age, decreased tumor grade, and radiation therapy were found to improve the overall survival. Yamada et al. [12] showed that lymph node metastasis ($\geq$N2) was significant prognostic factor for SGC patients. Liu et al. [13] retrospectively reviewed 215 patients with SGC, and multivariate analysis revealed that histological grade, cT classification, cN classification, and age were independent prognostic factors for overall survival.

Although numerous studies have identified a variety of prognostic factors associated with the overall survival of patients with SGC, there is relatively little work that integrates these factors into a predictive model [14,15]. The current clinical prediction models for SGC are subject to several constraints that warrant attention. Diversity in variable selection and assessment methodologies among different studies poses significant challenges for effective comparison and validation. Furthermore, these models were developed based on relatively small-sample cohorts, highlighting the need for enlarged datasets to substantiate their precision and dependability. Lastly, contemporary prediction models predominantly employ regression analyses, which can be intricate and cumbersome for practical application.Therefore, to enhance the accuracy of prognostic assessments, it is necessary to develop a comprehensive risk prediction model that transforms prognostic factors into more quantifiable and clinically referable indicators of survival probability. The nomogram offers a user-friendly graphical tool in statistics and medical research for predicting probabilities and estimating outcomes based on multiple variables. It presents a visual representation of a mathematical model that simplifies intricate calculations into a straightforward graphical format. The key advantage of nomograms is their ability to provide personalized and intuitive predictions or estimations without the need for complex statistical analysis or mathematical equations.

Therefore, the objective of this study was to ascertain the independent prognostic factors for overall survival in patients with SGC using a large sample derived from the Surveillance, Epidemiology, and End Results (SEER) database. Additionally, we aimed to develop novel graphical and online dynamic nomograms to aid clinicians in predicting overall survival and facilitating personalized decision-making regarding treatment strategies.

2. Materials and methods

This study complied with the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) statement [16], and all relevant ethical guidelines and regulations. The study was exempted from requiring informed consent due to the utilization of de-identified data obtained from the SEER database.

2.1. Data source

We obtained our data from the SEER database, a comprehensive cancer registry established by the National Cancer Institute (NCI) in the United States. SEER collects and maintains extensive data on cancer incidence, prevalence, treatment, and survival rates from various regions across the country. Specifically, we extracted information on patients diagnosed with SGC between 2010 and 2020. We mainly collected information on patient demographics, clinical characteristics, treatment modalities, and follow-up survival outcomes.

2.2. Patient selection

We included patients with a confirmed diagnosis of SGC based on the International Classification of Diseases for Oncology, 3rd edition (ICD-*O*-3) codes. The age at diagnosis was restricted in more than 18 years. Patients without complete information on survival outcomes or essential clinicopathological variables were excluded from the study.

2.3. Collected variables and data processing

We collected various demographic and clinical characteristics, including age, sex (male; female), year of diagnosis, race (white; black; other/unknown), marital status (married; unmarried; widowed; divorced; unknown), tumor history (adenoid cystic carcinoma [ACC]; mucoepidermoid carcinoma [MEC]; squamous cell carcinoma [SCC]; adenocarcinoma; pleomorphic adenoma [PMA]; others), tumor stage (localized; regional; distant; unknown), surgery procedure (performed; recommended but not performed; not recommended; unknown), radiation therapy (yes; no/unknown), chemotherapy (yes; no/unknown), systematic therapy (yes; no; unknown), months from diagnosis to treatment (≤ 1 m; >1 m; unknown), regional lymph node positivity (yes; no; not examined), metastasis to bone (yes; no; unknown), metastasis to brain/liver/lung (yes; no; unknown), metastasis to distant lymph node (yes; no; unknown), median household income (<$50,000; $50,000-$74,999; ≥$75,000), tumor size, and rural/urban status (metropolitan counties; nonmetropolitan counties), from the SEER database. Additionally, overall survival time (time interval between first visit and death or censoring) and vital status were recorded as primary outcome measures.

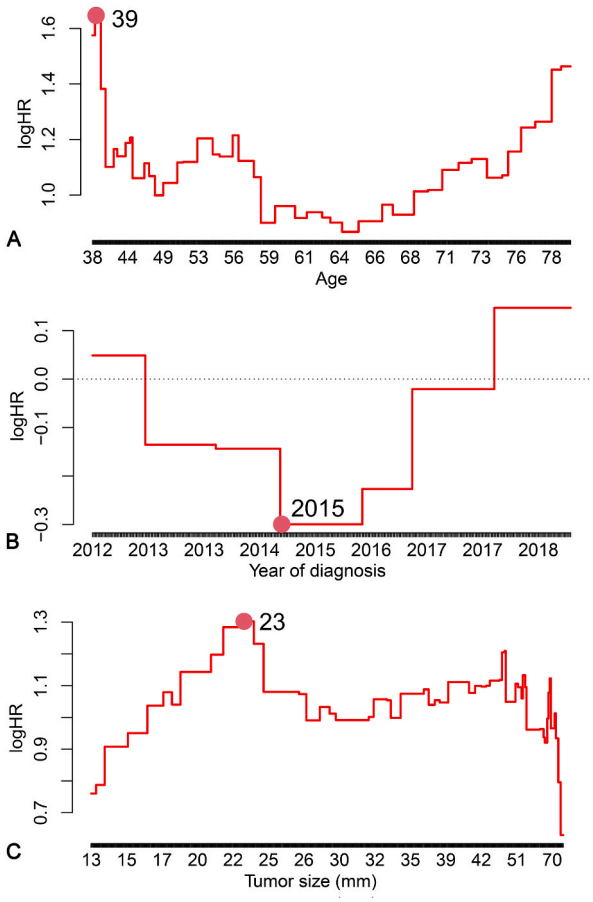

Fig. 1. The results of the running log-rank tests for continuous variables including age (A), year of diagnosis (B), and tumor size (C). The optimal cut-off points were 39 years, 2015, and 23 mm, respectively.

Table 1
Baseline characteristics of the included patients.

Characteristics	Overall (N = 527)	Training (N = 369)	Validation (N = 158)	P-value
Age				
≥39 years	465 (88.2 %)	322 (87.2 %)	143 (90.5 %)	0.290
<39 years	62 (11.8 %)	47 (12.7 %)	15 (9.4 %)	
Race				
Black	48 (9.1 %)	32 (8.7 %)	16 (10.1 %)	0.717
White	373 (70.8 %)	265 (71.8 %)	108 (68.4 %)	
Other	106 (20.2 %)	72 (19.5 %)	34 (21.5 %)	
Sex				
Male	271 (51.4 %)	197 (53.4 %)	74 (46.8 %)	0.199
Female	256 (48.6 %)	172 (46.6 %)	84 (53.2 %)	
Year of diagnosis				
≥2015	299 (56.7 %)	215 (58.2 %)	84 (53.1 %)	0.279
<2015	228 (43.3 %)	154 (41.7 %)	74 (46.8 %)	
Marital status				
Married	293 (55.6 %)	217 (58.8 %)	76 (48.1 %)	0.155
Unmarried	93 (17.7 %)	63 (17.1 %)	30 (19.0 %)	
Widowed	54 (10.2 %)	34 (9.2)	20 (12.7 %)	
Divorced	46 (8.7 %)	27 (7.3 %)	19 (12.0 %)	
Unknown	41 (7.8 %)	28 (7.6 %)	13 (8.2 %)	
Histology types				
Adenoid cystic carcinoma	178 (33.8 %)	124 (33.6 %)	54 (34.2 %)	0.571
Mucoepidermoid carcinoma	93 (17.6 %)	72 (19.5 %)	21 (13.3 %)	
Squamous cell carcinoma	81 (15.4 %)	57 (15.4 %)	24 (15.2 %)	
Adenocarcinoma	37 (7.0 %)	24 (6.5 %)	13 (8.2 %)	
Pleomorphic adenoma	30 (5.7 %)	21 (5.7 %)	9 (5.7 %)	
Others	108 (20.5 %)	71 (19.3 %)	37 (23.5 %)	
Summary stage				
Localized	248 (47.1 %)	176 (47.7 %)	72 (45.6 %)	0.520
Regional	186 (35.3 %)	133 (36.0 %)	53 (33.5 %)	
Distant	72 (13.7 %)	45 (12.2 %)	27 (17.1 %)	
Unknown	21 (4.0 %)	15 (4.1 %)	6 (3.8 %)	
Surgery procedure				
Performed	445 (84.4 %)	313 (84.8 %)	132 (83.5 %)	0.837
Recommended but not performed	12 (2.3 %)	7 (1.9 %)	5 (3.2 %)	
Not recommended	66 (12.5 %)	46 (12.5 %)	20 (12.7 %)	
Unknown	4 (0.8 %)	3 (0.8 %)	1 (0.6 %)	
Radiation therapy				
Yes	311 (59.0 %)	215 (58.3 %)	96 (60.8 %)	0.662
No	216 (41.0 %)	154 (41.7 %)	62 (39.2 %)	
Chemotherapy				
Yes	73 (13.9 %)	51 (13.8 %)	22 (13.9 %)	1.000
No	454 (86.1 %)	318 (86.2 %)	136 (86.1 %)	
Systemic therapy				
Yes	61 (11.6 %)	46 (12.5 %)	15 (9.5 %)	0.407
No	466 (88.4 %)	323 (87.5 %)	143 (90.5 %)	
Months from diagnosis to treatment				
>1 m	79 (15.0 %)	52 (14.1 %)	27 (17.1 %)	0.240
≤1 m	410 (77.8 %)	294 (79.7 %)	116 (73.4 %)	
Unknown	38 (7.2 %)	23 (6.2 %)	15 (9.5 %)	
Positive regional nodes				
Yes	132 (25.0 %)	97 (26.3 %)	35 (22.2 %)	0.265
No	205 (38.9 %)	147 (39.8 %)	58 (36.7 %)	
Not examined	190 (36.1 %)	125 (33.9 %)	65 (41.1 %)	
Metastasis to bone				
Yes	15 (2.8 %)	8 (2.2 %)	7 (4.4 %)	0.230
No	492 (93.4 %)	345 (93.5 %)	147 (93.0 %)	
Unknown	20 (3.8 %)	16 (4.3 %)	4 (2.5 %)	
Metastasis to brain/liver/lung				
Yes	28 (5.3 %)	15 (4.1 %)	13 (8.2 %)	0.137
No	479 (90.9 %)	339 (91.9 %)	140 (88.6 %)	
Unknown	20 (3.8 %)	15 (4.1 %)	5 (3.2 %)	
Metastasis at distant LN				
Yes	5 (0.9 %)	5 (1.4 %)	0 (0.0 %)	0.071
No	232 (44.0 %)	171 (46.3 %)	61 (38.6 %)	
Unknown	290 (55.0 %)	193 (52.3 %)	97 (61.4 %)	
Median household income				
<$50,000	28 (5.3 %)	18 (4.8 %)	10 (6.3 %)	0.733
$50,000 - $74,999	209 (39.7 %)	145 (39.2 %)	64 (40.5 %)	

(*continued on next page*)

Table 1 (*continued*)

Characteristics	Overall (N = 527)	Training (N = 369)	Validation (N = 158)	P-value
≥$75,000	290 (55.0 %)	206 (55.8 %)	84 (53.1 %)	
Tumor size (mm)				
<23	194 (36.8 %)	142 (38.4 %)	52 (32.9 %)	0.440
≥23	282 (53.5 %)	191 (51.7 %)	91 (57.5 %)	
Unknown	51 (9.7 %)	36 (9.7 %)	15 (9.4)	
Rural/Urban				
Metropolitan counties	449 (85.2 %)	314 (85.1 %)	135 (85.4 %)	1.000
Nonmetropolitan counties	78 (14.8 %)	55 (14.9 %)	23 (14.6 %)	

Footnote: LN, lymph node.

The continuous prognostic factors, including age, year of diagnosis, and tumor size, were transferred into binary variables, using the optimal cut-off points identified by running log-rank test [17]. As a result, the optimal cut-off points for age (<39; ≥ 39), year of diagnosis (<2015; ≥ 2015), and tumor size (<23; ≥ 23; unknown) were 39 years (Fig. 1A), 2015 (Fig. 1B), and 23 mm (Fig. 1C).

2.4. Statistical analysis

A computer algorithm was employed to randomly divide the cohort into training and test samples randomly, with a ratio of 7:3. Appropriate descriptive statistics, including frequencies and percentages, were calculated for the two samples. Pearson's Chi-square test was selected to compare the differences of the prognostic factors between two samples.

Using the training sample, we performed univariate Cox proportional hazards regression analysis to identify potential prognostic factors associated with overall survival. Hazard ratios (HRs) and their corresponding 95 % confidence intervals (CIs) were calculated. Variables with p-values <0.15 were considered statistically significant and selected for further analysis. To determine independent prognostic factors, variables identified in the univariate analysis were included in a multivariate Cox proportional hazards regression model. Adjusted hazard ratios (AHRs) and their corresponding 95 % CIs were estimated to evaluate the association between each variable and overall survival while accounting for potential confounders. Kaplan-Meier survival curves were generated for the independent predictors identified in multivariate analysis, and differences across groups were assessed using the log-rank test.

Based on the results of the multivariate analysis, we developed a graphical nomogram for predicting individual patient's survival probabilities at 12, 24, 60, and 120 months, and median survival time. Calibration curves and concordance indices (C-indices) were used to evaluate the accuracy and discriminative ability of the nomogram. Additionally, we generated receiver operating characteristic (ROC) curve and calculated the areas under ROC curve (AUCs). To continuously illustrate the AUC values, time-dependent AUC curve was generated. The performance assessment of the nomogram were performed both in training and test samples. Finally, we deployed our novel model as an online dynamic widget on the server of shinyapps.ioo.

All statistical analyses were conducted using R version 4.2.1 (R Foundation for Statistical Computing, Vienna, Austria). P-values <0.05 were considered statistically significant.

3. Results

The baseline characteristics of the included patients are listed in Table 1. A total of 527 SGC patients were included in this study,

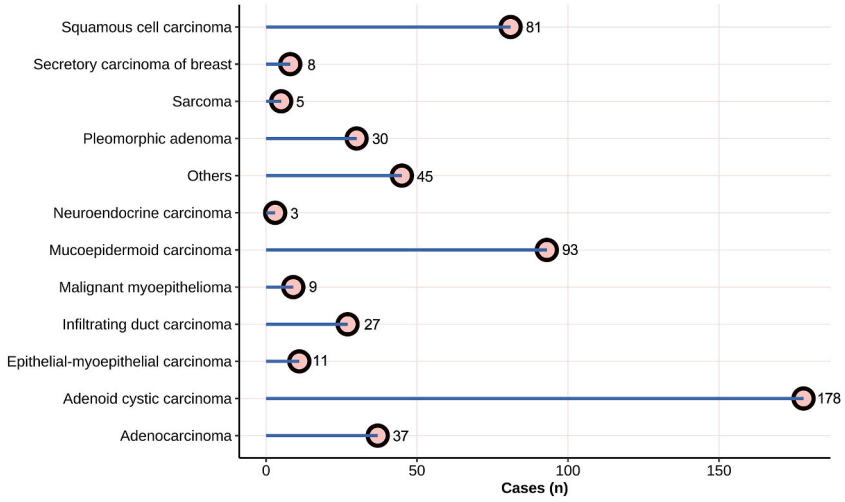

Fig. 2. The detailed distribution of the tumor histology.

Table 2
Results of univariable analysis for prognostic factors of overall survival.

Characteristics	HR	LCI	UCI	P-value
Age				
<39 years	Ref.			
≥39 years	4.464	1.645	12.118	**0.003**[b]
Race				
Black	Ref.			
White	1.392	0.6448	3.005	0.400
Other	0.696	0.2739	1.770	0.447
Sex				
Female	Ref.			
Male	2.497	1.653	3.773	**<0.001**[c]
Year of diagnosis				
<2015	Ref.			
≥2015	0.863	0.571	1.306	0.486
Marital status				
Divorced	Ref.			
Married	0.756	0.387	1.477	0.413
Unmarried	0.435	0.187	1.012	0.053.
Widowed	1.238	0.565	2.716	0.594
Tumor histology				
Adenocarcinoma	Ref.			
Adenoid cystic carcinoma	0.414	0.208	0.828	**0.013**[a]
Mucoepidermoid carcinoma	0.303	0.138	0.669	**0.003**[b]
Pleomorphic adenoma	0.279	0.078	1.003	0.051.
Squamous cell carcinoma	1.042	0.518	2.095	0.908
Others	0.669	0.326	1.373	0.273
Summary stage				
Distant	Ref.			
Localized	0.140	0.083	0.236	**<0.001**[c]
Regional	0.362	0.228	0.576	**<0.001**[c]
Unstaged	0.701	0.291	1.692	0.430
Surgery procedure				
Not recommended	Ref.			
Performed	0.315	0.198	0.501	**<0.001**[c]
Recommended but not performed	0.407	0.122	1.359	0.144.
Radiation therapy				
None/Unknown	Ref.			
Yes	1.004	0.687	1.465	0.986
Chemotherapy				
None/Unknown	Ref.			
Yes	2.671	1.726	4.134	**<0.001**[c]
Systemic therapy				
No	Ref.			
Yes	3.004	1.917	4.708	**<0.001**[c]
Months from diagnosis to treatment				
≤1 m	Ref.			
>1 m	1.448	0.856	2.450	0.168
Positive regional nodes				
No	Ref.			
Yes	2.879	1.814	4.568	**<0.001**[c]
Not examined	1.353	0.833	2.195	0.222
Metastasis to bone				
No	Ref.			
Yes	8.810	3.778	20.546	**<0.001**[c]
Metastasis to brain/liver/lung				
No	Ref.			
Yes	4.747	2.433	9.258	**<0.001**[c]
Metastasis at distant LN				
No	Ref.			
Yes	13.711	4.416	42.572	**<0.001**[c]
Median household income				
$50,000 - $74,999	Ref.			
<$50,000	1.442	0.707	2.939	0.314
≥$75,000	0.854	0.577	1.264	0.431
Tumor size (mm)				
<23	Ref.			
≥23	3.267	2.031	5.256	**<0.001**[c]
Rural/Urban				
Metropolitan counties	Ref.			
Nonmetropolitan counties	1.284	0.783	2.107	0.323

Footnotes: LN, lymph node; HR, hazard ratio; LCI, lower 95 % confidence interval; UCI, upper 95 % confidence interval.p < 0.150.
[a] p < 0.050.
[b] p < 0.010.
[c] p < 0.001.

with 369 assigned to the training group and 158 assigned to test group. The age was less than 39 years in 62 (11.8 %) patients, and ≥39 years in 469 (88.2 %) patients. The male percentage was 51.4 % (271 patients), and 43.3 % (228 patients) of them were diagnosed before 2015. The detailed distribution of the tumor histology is presented in Fig. 2. ACC (n = 178, 33.8), MEC (n = 93, 17.6 %), and SCC (n = 81, 15.4 %) are the top three most common types of SGC. Surgery procedures were performed in 445 (84.4 %) patients, recommended but not performed in 12 (2.3 %) patients, and not recommended in 66 (12.5 %) patients. Radiation, chemical, and

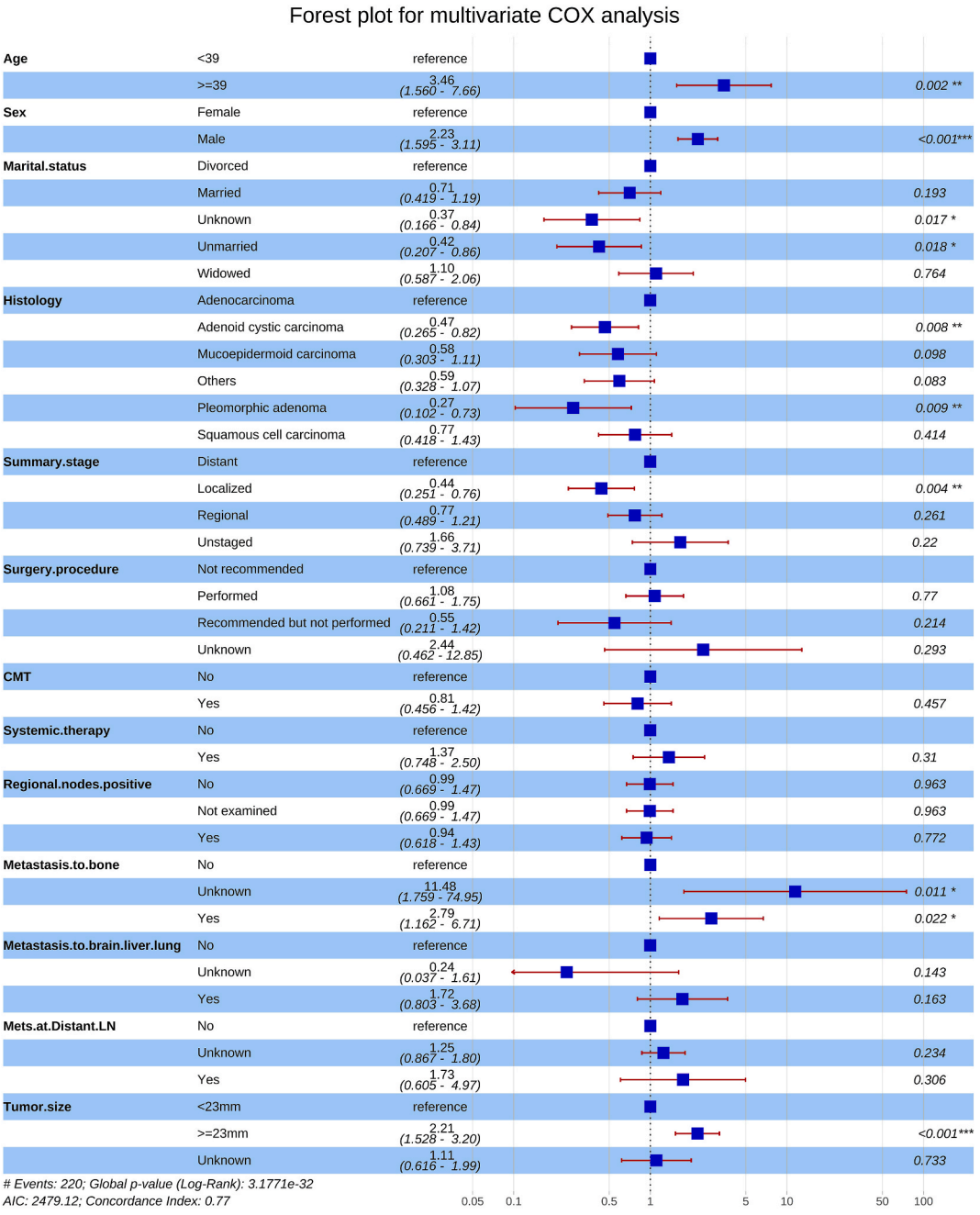

Fig. 3. The forest plot for multivariable COX analysis. The result showed that age, sex, marital status, tumor histology, summary stage, metastasis to bone, and tumor size were identified as independent prognostic factors for overall survival.

systematic therapies were performed in 311 (59.0 %), 73 (13.9 %), and 61 (11.6 %) patients, respectively. Metastasis to bone was confirmed in 15 (2.8 %) patients, and visceral metastasis to brain/liver/lung was confirmed in 28 (5.3 %) patients. The tumor size was larger than 23 mm and less than 23 mm in 282 (53.5 %) and 194 (36.8 %) patients, respectively. The baseline characteristics were similar between training and test samples.

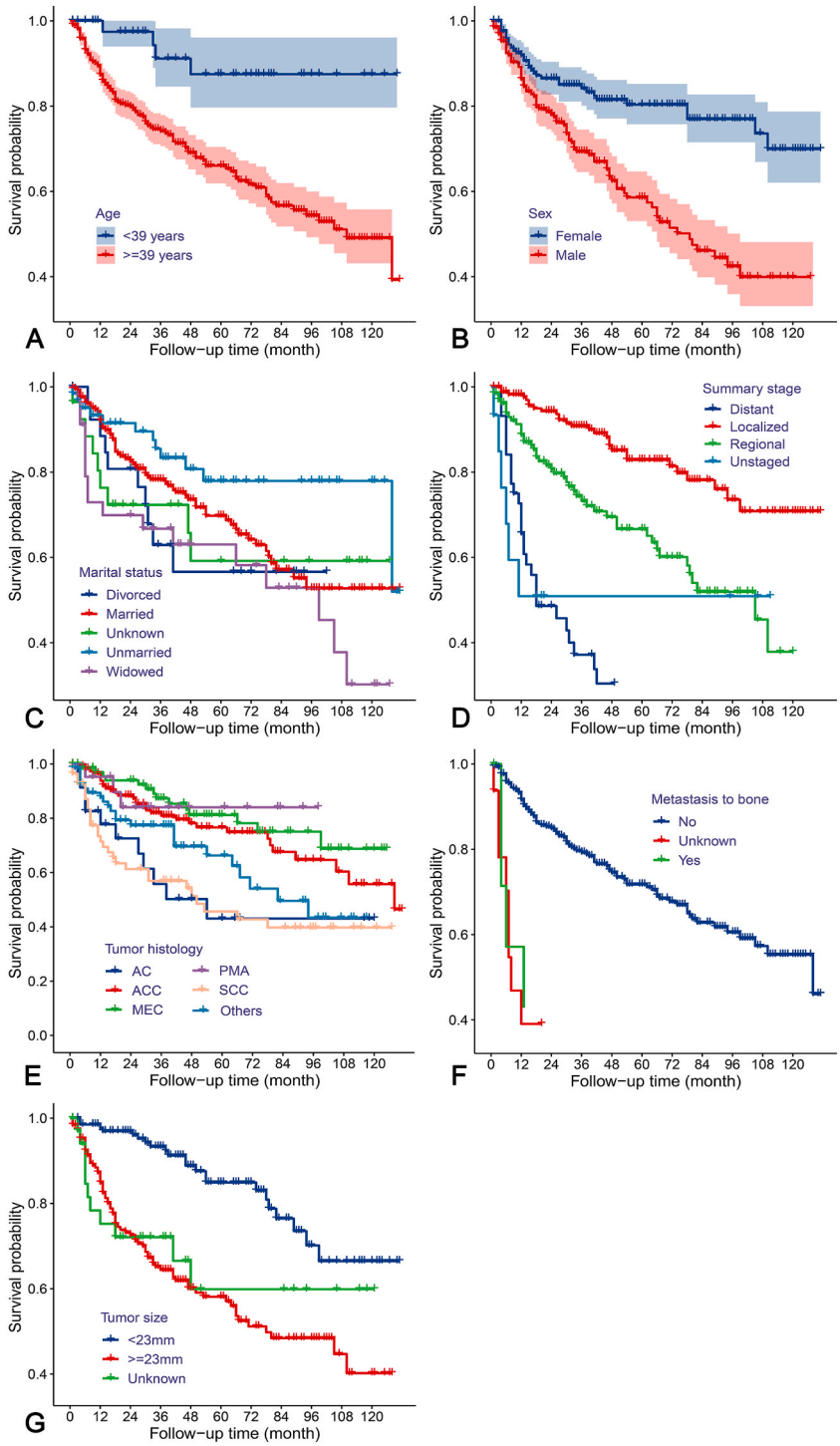

Fig. 4. The Kaplan-Meier survival curves for the independent prognostic factors (A: age; B: sex; C: marital status; D: summary stage; E: tumor histology; F: metastasis to bone; G: tumor size) identified by multivariable analysis. All predictors were proven to be significant by log-rank test ($p < 0.001$***).

3.1. Results of univariable and multivariable COX analyses

The results of the univariable analysis are shown in Table 2. A total of 19 variables were analyzed for the potential predictive value, and the results showed that age ($p = 0.003$**), sex ($p < 0.001$***), tumor histology ($p = 0.003$**), summary stage ($p < 0.001$***), surgery procedures ($p < 0.001$***), chemotherapy ($p < 0.001$***), systematic therapy ($p < 0.001$***), positive regional nodes ($p < 0.001$***), metastases to bone ($p < 0.001$***), brain/liver/lung ($p < 0.001$***), or distant lymph node ($p = 0.003$**), and tumor size ($p < 0.001$***) were significantly associated with the overall survival, while marital status ($p = 0.053$.) was demonstrated to be marginally ($p < 0.150$) associated with overall survival.

The results of the multivariable COX analysis are presented in the forest plot in Fig. 3. As a result, age (AHR = 3.46, 95%CI: 1.56–7.66, $p = 0.002$**), sex (AHR = 2.23, 95%CI: 1.60–3.11, $p < 0.001$***), marital status (divorced vs. unmarried: AHR = 0.42, 95% CI: 0.21–0.86, $p = 0.018$*), tumor histology (adenocarcinoma vs. ACC: AHR = 0.47, 95%CI: 0.27–0.82, $p = 0.008$**; adenocarcinoma vs. PMA: AHR = 0.27, 95%CI: 0.10–0.73, $p = 0.009$**), summary stage (distant vs. localized: AHR = 0.44, 95%CI: 0.25–0.76, $p = 0.004$**), metastasis to bone (no vs. unknown: AHR = 11.48, 95%CI: 1.76–74.95, $p = 0.011$*; no vs. yes: AHR = 2.79, 95%CI: 1.16–6.71, $p = 0.022$*), and tumor size (<23 mm vs. ≥ 23 mm: AHR = 2.21, 95%CI: 1.53–3.20, $p < 0.001$***) were identified as independent prognostic factors for overall survival. The Kaplan-Meier survival curves for these factors identified by multivariable analysis are presented in Fig. 4. All predictors, including age (Fig. 4A), sex (Fig. 4B), marital status (Fig. 4C), summary stage (Fig. 4D), tumor histology (Fig. 4E), metastasis to bone (Fig. 4F), and tumor size (Fig. 4G), were proven to be significant by log-rank test ($p < 0.001$***).

3.2. Establishment and validation of the nomogram

Using the significant factors identified in multivariable analysis, a novel nomogram was generated (see Fig. 5), which could help clinicians predict the 12- (Fig. 5A), 24- (Fig. 5B), 60- (Fig. 5C) and 120-month (Fig. 5D) survival rates, and the median survival time. The C-indexes for the training and test samples were 0.77 (95%CI: 0.71–0.84) and 0.77 (95%CI: 0.68–0.85) respectively, showing a favorable discrimination ability of the novel model. The calibration curves depicting the consistency between predicted and actual survival probabilities for the training and test samples at 12, 24, 60, and 120 months are presented in Fig. 6A–D and Fig. 7A–D respectively, demonstrating excellent prediction accuracy of the nomogram. The ROC curves for training and test samples are available in Fig. 8A–D and Fig. 9A–D. The AUC values for the training sample were 0.847, 0.822, 0.766, and 0.856 at 12, 24, 60, and 120 months, while the AUC values for the test sample were 0.817, 0.848, 0.735, and 0.884 at 12, 24, 60, and 120 months, respectively. The time-

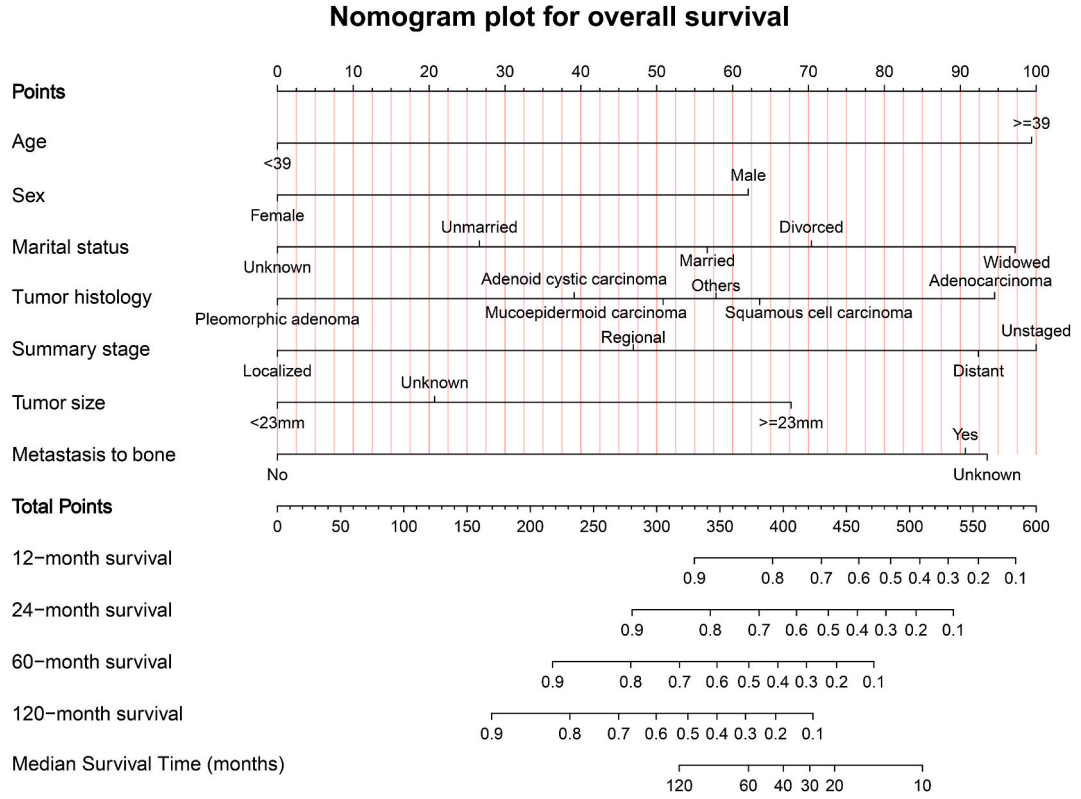

Fig. 5. The graphical nomogram model for guiding the predicting of 12-, 24-, 60- and 120-month survival probabilities, and median survival time.

dependent AUC curves for the two samples are shown in Fig. 10A and B.

Finally, an online dynamic nomogram was deployed, which is available at URL: https://yangxg1209.shinyapps.io/overall_survival_submandibular_gland_tumor/. When users access our dynamic nomogram model, they simply need to select the values corresponding to each predictive factor and click the "Predict" button. They will then obtain a continuous predicted survival curve, and the predicted survival probability at any given time point.

4. Discussion

SGC is a rare malignancy characterized by its diverse clinical manifestations and treatment outcomes. Understanding the prognostic factors associated with overall survival is crucial for clinicians to devise evidence-based treatment plans. This research has determined the independent prognostic factors impacting survival of SGC patients, and has further developed a nomogram model. This model has been rigorously validated, offering physicians a tool to predict survival rates at several time points and the median survival time. Moreover, to facilitate clinical application, an accessible online calculator has been implemented, enabling the swift generation of personalized survival curves and the computation of survival probabilities at any specified time point.

The multivariable COX analysis in our study identified several independent prognostic factors for overall survival. These factors included age, sex, marital status, tumor histology, summary stage, metastasis to bone, and tumor size. The prognostic factors for overall survival in SGC have been widely identified in previous studies [8–10,18–21]. In a similar population-based cohort study, Lee et al. [8] performed multivariable COX analysis and revealed that age, sex, tumor grade, summary stage, and surgery procedure were significantly associated with overall survival. However, the authors did not transfer the results into a prediction model to guide survival prediction. Recently, Westergaard-Nielsen et al. [9] performed a national cohort study of 206 SGC patients, and the multivariable analysis found that age, cervical lymph-node metastases (N+), and vascular invasion had significant impact on the overall survival. In study of Sahin et al. [19], positive nodal stage and positive surgical margin were proven to be significant predictors of overall survival.

The scarcity of published literature on clinical predictive models for SGC can indeed be attributed to its rarity, which often results in limited sample sizes for research and insufficient data for robust modeling. The SEER database, however, offers a unique opportunity to overcome this limitation due to its extensive collection of cancer cases, including those of SGC. By analyzing a substantial dataset

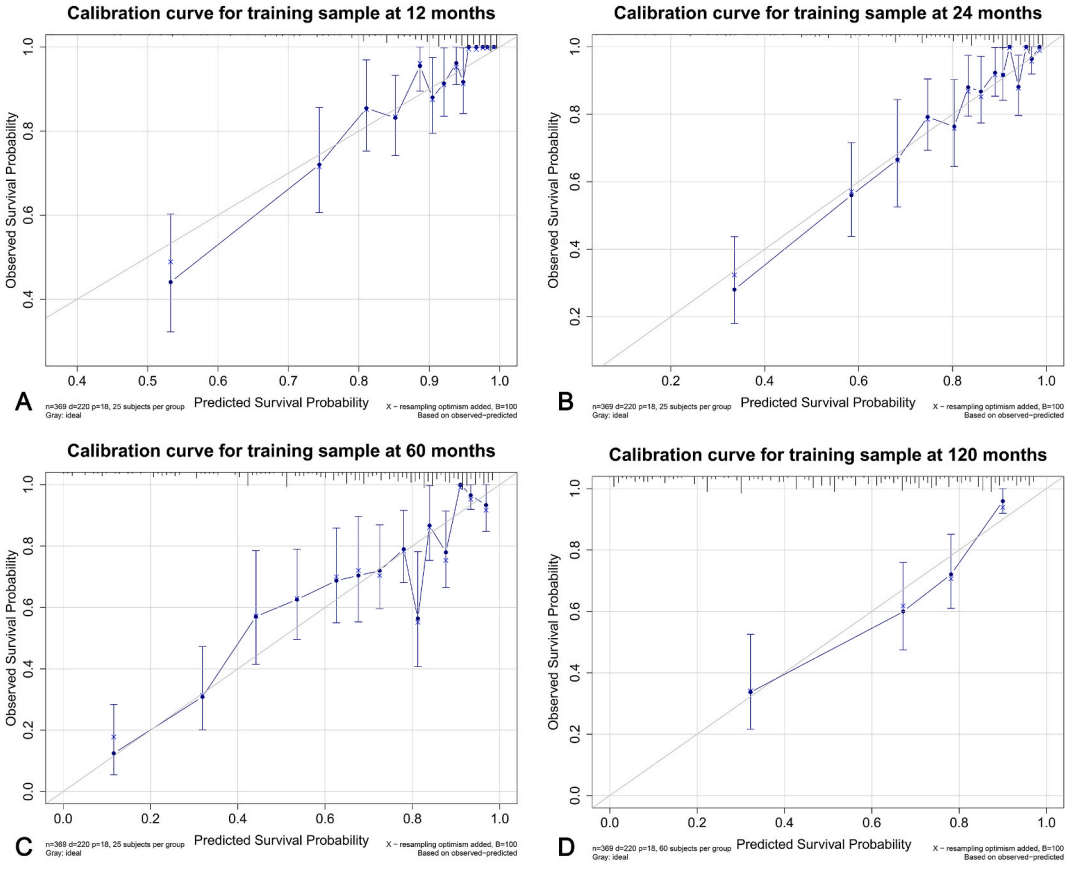

Fig. 6. The calibration curves for training sample at 12 (A), 24 (B), 60 (C), and 120 (D) months. The curves demonstrated favorable consistency between the predicted and actual survival probabilities at each time-point.

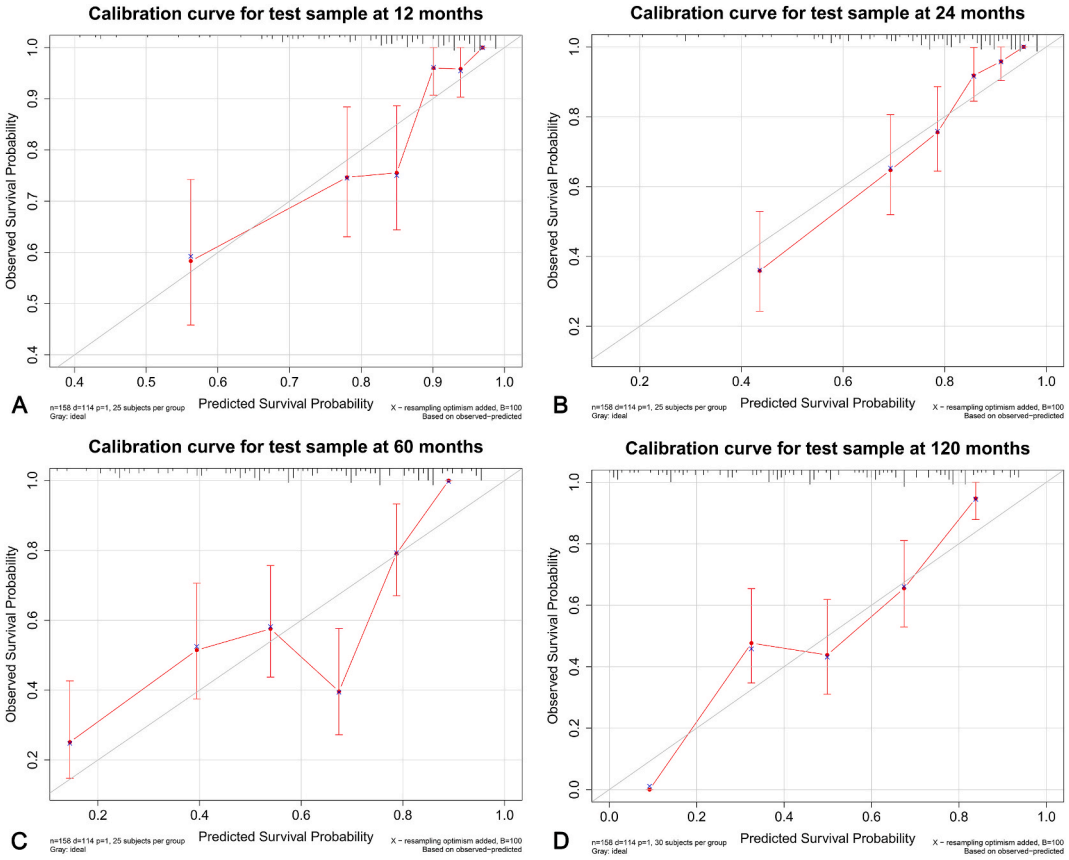

Fig. 7. The calibration curves for test sample at 12 (A), 24 (B), 60 (C), and 120 (D) months. The curves demonstrated favorable consistency between the predicted and actual survival probabilities at each time-point.

comprising 527 cases, this study has been able to surmount the obstacles typically encountered when dealing with less common cancers. The development of a predictive model for SGC in this study represents a significant advancement in the field. It goes beyond mere identification of prognostic factors by integrating multiple clinicopathological variables to create a model capable of providing more precise predictions of outcomes. This approach not only enhances our understanding of the disease's behavior but also has the potential to improve patient management by guiding treatment decisions and facilitating personalized care. The model's reliability is a direct result of the large sample size, which allows for more accurate estimates of the effects of different variables on patient survival. Furthermore, the use of advanced statistical techniques ensures that the model accounts for the complex interplay between various factors, leading to more nuanced predictions.

Nomogram is a graphical tool used in medicine to predict outcomes and make clinical decisions. It consists of a set of calibrated scales that visually represent the relationships between various predictors or risk factors and the likelihood of an event or outcome. A nomogram enables healthcare professionals to estimate individualized probabilities of specific outcomes, such as disease progression, survival rates, or treatment response, by plotting patient-specific values on the respective scales and connecting them to determine the probability. Dynamic nomograms take the concept of traditional nomograms a step further by allowing for real-time adjustments based on new information or updated patient characteristics. They continuously adapt and update the predictions by incorporating the latest data and providing more accurate and personalized prognostic information. This enhances clinical decision-making, facilitates personalized treatment planning, and ultimately improves patient outcomes.

The nomogram has proven to be an indispensable instrument in the realm of oncology, with its application spanning numerous tumor types [22–27]. In our research, we took a pioneering step by developing novel graphical and web-based models dedicated to predicting the survival prospects of patients suffering from SGC. The performance metrics of our nomogram were impressive, showcasing exceptional discriminative prowess.This was substantiated by the calibration plots, which confirmed the model's high degree of accuracy in predicting survival rates at critical junctures—namely, 12, 24, 60, and 120 months—for both training and test datasets. The robustness of our nomogram was further underscored by the Area Under the Curve (AUC) values, which provided compelling evidence of its superior discriminative capabilities. Collectively, these results affirm the trustworthiness of our nomogram as a reliable prognostic tool for SGC survival assessment.

There are several limitations need to be acknowledged here. Firstly, the retrospective nature of the study makes it susceptible to inherent biases and confounding factors that may affect the accuracy and generalizability of the findings. Additionally, the SEER

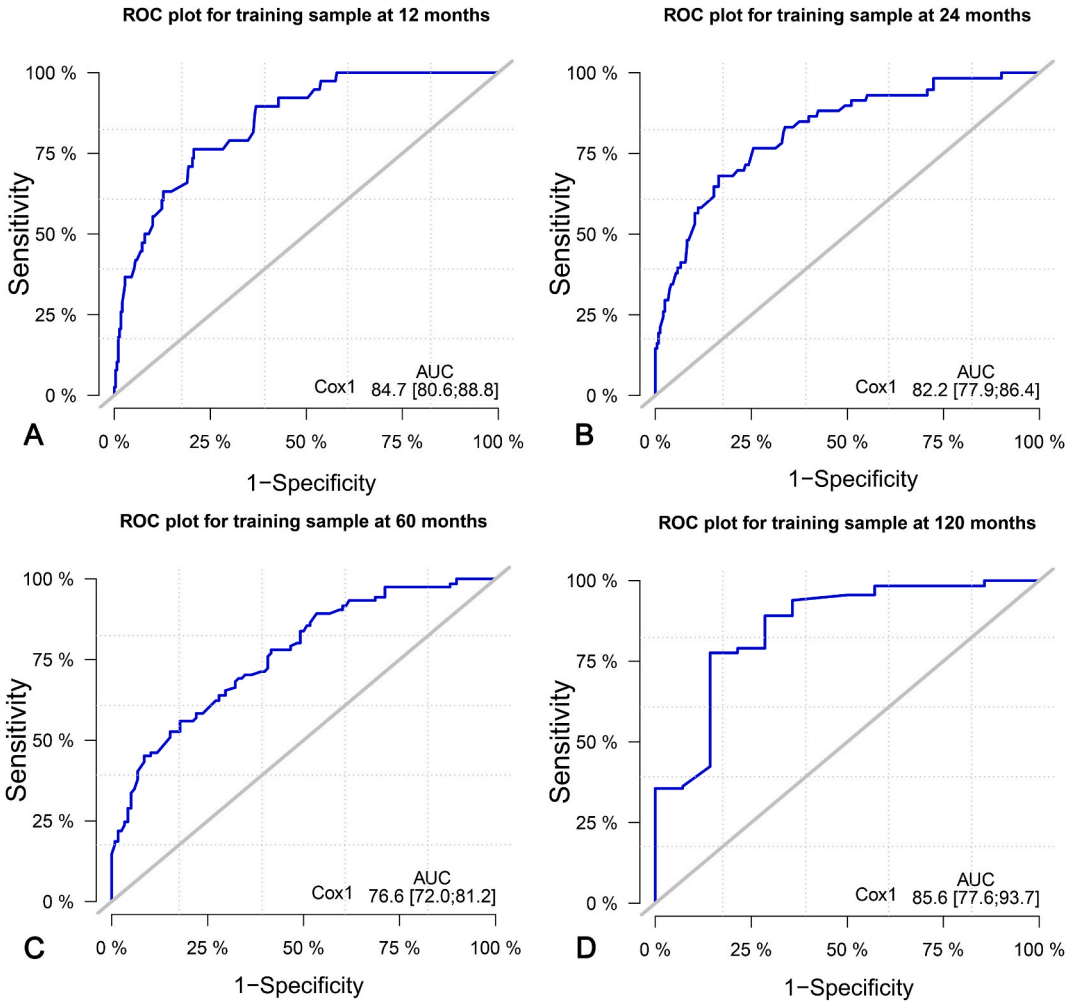

Fig. 8. ROC curves for training sample at 12 (A), 24 (B), 60 (C) and 120 (D) months. The AUC values were 0.847, 0.822, 0.766, and 0.856 at the four time-points. ROC curve, receiver operating characteristic curve; AUC, area under ROC curve.

database, although a comprehensive and widely-used resource, has certain limitations, such as the potential for missing or incomplete data, potential coding errors, and the inability to capture certain characteristics or factors that may influence survival rates. Furthermore, the nomogram model developed and validated in this study should be interpreted with caution as it relies on assumptions and simplifications, and its performance may differ when applied to different populations or settings. Future research should aim to overcome these limitations by conducting prospective studies, incorporating more comprehensive datasets, and validating the nomogram model in diverse patient populations to enhance its reliability and clinical applicability.

5. Conclusions

In conclusion, our study identified the independent prognostic factors for overall survival of patients with SGC, and successfully established and rigorously validated novel graphical and dynamic online nomograms, which provide an accurate prediction of survival rates and allows for personalized risk assessment. The inclusion of multiple prognostic factors in the model enhances its clinical utility and helps clinicians make informed decisions regarding treatment approaches for patients with SGC.

Ethics approval and consent to participate

The study was conducted in accordance with the Declaration of Helsinki, and the use of de-identified data from the SEER database exempted the study from requiring informed consent.

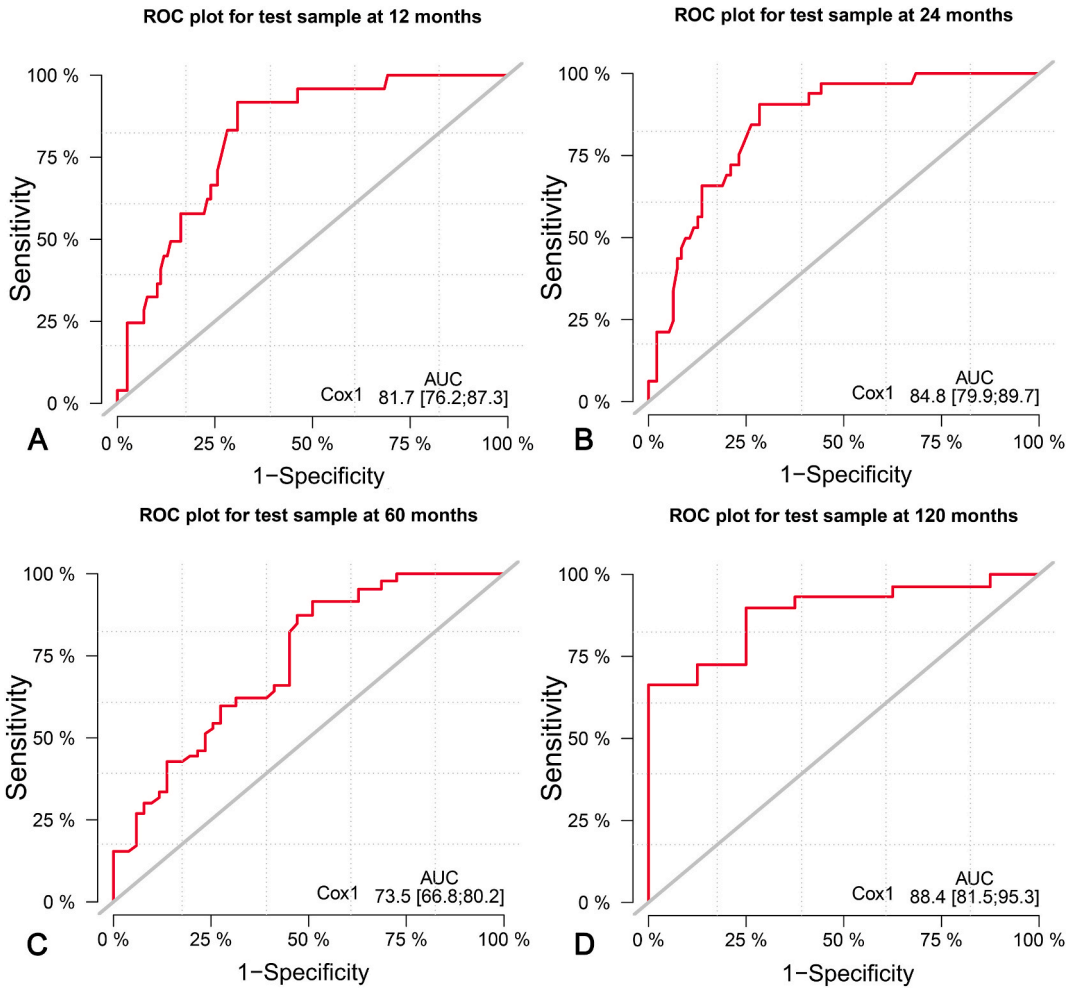

Fig. 9. ROC curves for test sample at 12 (A), 24 (B), 60 (C) and 120 (D) months. The AUC values were 0.817, 0.848, 0.735, and 0.884 at the four time-points. ROC curve, receiver operating characteristic curve; AUC, area under ROC curve.

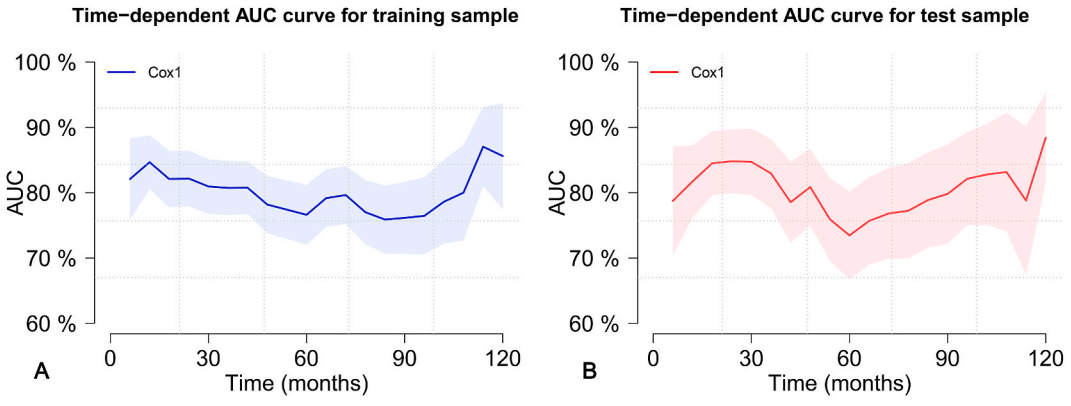

Fig. 10. Time-dependent AUC curves for the training (A) and test (B) samples. AUC, area under ROC curve; ROC curve, receiver operating characteristic curve.

Consent for publication

All authors read and approved the final manuscript.

Data availability statement

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Funding

This work was supported by grant from the Science and Technology Plan Project of Guizhou Province (No. Qiankehe Foundation - ZK[2023] General 537) and Zunyi Medical University Graduate Research Fund (No. ZYK259). The sponsor or funder had no involvement in any part of this study. All authors confirmed the independence of researchers from funding sources.

Institutional review board statement

The study was conducted in accordance with the Declaration of Helsinki, and the use of de-identified data from the SEER database exempted the study from requiring informed consent.

Availability of data and material

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

CRediT authorship contribution statement

Shan-shan Yang: Writing – original draft, Software, Methodology, Investigation, Data curation. **Xiong-gang Yang:** Writing – original draft, Validation, Software, Methodology. **Xiao-hong Yang:** Validation, Project administration. **Xiao-hua Hu:** Validation, Project administration.

Declaration of competing interest

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

Acknowledgements

Not applicable.

List of abbreviations

SGC	submandibular gland cancer
OS	overall survival
ROC	receiver operating characteristic
AUC	area under ROC curve
SEER	Surveillance, Epidemiology, and End Results
TRIPOD	Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis
NCI	National Cancer Institute
ICD-*O*-3	International Classification of Diseases for Oncology, 3rd edition
HR	hazard ratio
CI	confidence interval
AHR	adjusted hazard ratio
ACC	adenoid cystic carcinoma
MEC	mucoepidermoid carcinoma
SCC	squamous cell carcinoma
PMA	pleomorphic adenoma

References

[1] R.H. Spiro, Salivary neoplasms: overview of a 35-year experience with 2,807 patients, Head Neck Surg. 8 (3) (1986) 177–184, https://doi.org/10.1002/hed.2890080309.

[2] J.A. Pinkston, P. Cole, Incidence rates of salivary gland tumors: results from a population-based study, Otolaryngol. Head Neck Surg. 120 (6) (1999) 834–840, https://doi.org/10.1016/S0194-5998(99)70323-2.

[3] J.W. Eveson, R.A. Cawson, Salivary gland tumours. A review of 2410 cases with particular reference to histological types, site, age and sex distribution, J. Pathol. 146 (1) (1985) 51–58, https://doi.org/10.1002/path.1711460106.

[4] M. Westergaard-Nielsen, C. Godballe, J.G. Eriksen, et al., Salivary gland carcinoma in Denmark: a national update and follow-up on incidence, histology, and outcome, Eur. Arch. Oto-Rhino-Laryngol. 278 (4) (2021) 1179–1188, https://doi.org/10.1007/s00405-020-06205-2.

[5] P. Wahlberg, H. Anderson, A. Biörklund, T. Möller, R. Perfekt, Carcinoma of the parotid and submandibular glands–a study of survival in 2465 patients, Oral Oncol. 38 (7) (2002) 706–713, https://doi.org/10.1016/s1368-8375(02)00007-6.

[6] L. Barnes, J.W. Eveson, P. Reichart, D. Sidransky, Pathology and genetics of head and neck tumours, in: WHO Classification of Tumours, third ed., Volume 9, 2005. Available at: https://publications.iarc.fr/Book-And-Report-Series/Who-Classification-Of-Tumours/Pathology-And-Genetics-Of-Head-And-Neck-Tumours-2005.

[7] R.B. Bell, E.J. Dierks, L. Homer, B.E. Potter, Management and outcome of patients with malignant salivary gland tumors, J. Oral Maxillofac. Surg. 63 (7) (2005) 917–928, https://doi.org/10.1016/j.joms.2005.03.006.

[8] R.J. Lee, A.P. Tan, E.L. Tong, N. Satyadev, R.E. Christensen, Epidemiology, prognostic factors, and treatment of malignant submandibular gland tumors: a population-based cohort analysis, JAMA Otolaryngol Head Neck Surg 141 (2015) 905–912, https://doi.org/10.1001/jamaoto.2015.1745.

[9] M. Westergaard-Nielsen, C. Godballe, J.G. Eriksen, et al., Epidemiology, outcomes, and prognostic factors in submandibular gland carcinomas: a national DAHANCA study, Eur. Arch. Oto-Rhino-Laryngol. 280 (7) (2023) 3405–3413, https://doi.org/10.1007/s00405-023-07940-y.

[10] B. Denis, K. Aleksandar, I. Mijatov, I. Miroslav, Pathological nodal status as a main predictive factor of survival and treatment outcomes of submandibular salivary gland cancers: a 25-year single center experience, J Stomatol Oral Maxillofac Surg (2023) 101462, https://doi.org/10.1016/j.jormas.2023.101462.

[11] N. Bhattacharyya, Survival and prognosis for cancer of the submandibular gland, J. Oral Maxillofac. Surg. 62 (4) (2004) 427–430, https://doi.org/10.1016/j.joms.2003.06.012.

[12] K. Yamada, K. Honda, H. Tamaki, et al., Survival in patients with submandibular gland carcinoma - results of a multi-institutional retrospective study, Auris Nasus Larynx 45 (5) (2018) 1066–1072, https://doi.org/10.1016/j.anl.2018.01.011.

[13] S. Liu, W. Xu, L. Liu, et al., Prognostic factors and treatment considerations of submandibular gland carcinomas: a population-based study, Oral Dis. (2022), https://doi.org/10.1111/odi.14313.

[14] C.D. Jacobs, I. Barak, S.H. Jung, et al., Prediction model to estimate overall survival benefit of postoperative radiotherapy for resected major salivary gland cancers, Oral Oncol. 132 (2022) 105955, https://doi.org/10.1016/j.oraloncology.2022.105955.

[15] S. Mallik, J. Agarwal, T. Gupta, et al., Prognostic factors and outcome analysis of submandibular gland cancer: a clinical audit, J. Oral Maxillofac. Surg. 68 (9) (2010) 2104–2110, https://doi.org/10.1016/j.joms.2009.09.030.

[16] G.S. Collins, J.B. Reitsma, D.G. Altman, K.G. Moons, Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement, Ann. Intern. Med. 162 (1) (2015) 55–63, https://doi.org/10.7326/m14-0697.

[17] R. Rami-Porta, V. Bolejack, J. Crowley, et al., The IASLC lung cancer staging project: proposals for the revisions of the T descriptors in the forthcoming eighth edition of the TNM classification for lung cancer, J. Thorac. Oncol. 10 (7) (2015) 990–1003, https://doi.org/10.1097/jto.0000000000000559.

[18] H.M. Rayess, A. Dezube, I. Bawab, et al., Tumor differentiation as a prognostic factor for major salivary gland malignancies, Otolaryngol. Head Neck Surg. 157 (3) (2017) 454–461, https://doi.org/10.1177/0194599817700593.

[19] M.M. Şahin, V. Demircan, M. Göcek, et al., Submandibular gland cancers: prognostic factors and survival analysis, Eur. Arch. Oto-Rhino-Laryngol. 277 (8) (2020) 2307–2313, https://doi.org/10.1007/s00405-020-05924-w.

[20] D. Lombardi, R. Accorona, A. Lambert, et al., Long-term outcomes and prognosis in submandibular gland malignant tumors: a multicenter study, Laryngoscope 128 (12) (2018) 2745–2750, https://doi.org/10.1002/lary.27236.

[21] Y. Liu, L. Qin, R. Zhuang, X. Huang, M. Su, Z. Han, Nodal stage: is it a prognostic factor for submandibular gland cancer? J. Oral Maxillofac. Surg. 76 (8) (2018) 1794–1799, https://doi.org/10.1016/j.joms.2017.11.012.

[22] W.S. Peng, X. Xing, Y.J. Li, et al., Prognostic nomograms for nasopharyngeal carcinoma with nodal features and potential indication for N staging system: validation and comparison of seven N stage schemes, Oral Oncol. 144 (2023) 106438, https://doi.org/10.1016/j.oraloncology.2023.106438.

[23] J.X. Xu, S.L. Qin, H.W. Wei, Y.Y. Chen, Y.C. Peng, L.N. Qi, Prognostic factors and an innovative nomogram model for patients with hepatocellular carcinoma treated with postoperative adjuvant transarterial chemoembolization, Ann. Med. 55 (1) (2023) 2199219, https://doi.org/10.1080/07853890.2023.2199219.

[24] L. Liu, F.Z. Sun, P.Y. Zhang, Y. Xiao, H.X. Ni, Development and validation a model for predicting overall survival of bladder cancer with lung metastasis: a population-based study, Eur. J. Med. Res. 28 (1) (2023) 279, https://doi.org/10.1186/s40001-023-01261-w.

[25] S. Wang, Y. Chen, J. Sun, R. Mo, Q. Tan, Development and validation of two online dynamic nomograms for patients with non-distant metastatic cutaneous melanoma based on surgical approaches, Cancer Med. (2023), https://doi.org/10.1002/cam4.6448.

[26] R. Yang, X. Yu, P. Zeng, Construction and validation of a SEER-based prognostic nomogram for young and middle-aged males patients with hepatocellular carcinoma, J. Cancer Res. Clin. Oncol. 149 (12) (2023) 10099–10108, https://doi.org/10.1007/s00432-023-04901-0.

[27] X.G. Yang, J.T. Feng, F. Wang, et al., Development and validation of a prognostic nomogram for the overall survival of patients living with spinal metastases, J. Neuro Oncol. 145 (1) (2019) 167–176, https://doi.org/10.1007/s11060-019-03284-y.