



# Construction of nomograms for nasopharyngeal carcinoma containing primary tumor size and SEER stage

Guangrong Yang<sup>1,2#</sup>, Bangyu Luo<sup>1#</sup>, Qiao Yang<sup>3#</sup>, Mingjing Chen<sup>1</sup>, Xiu Yang<sup>1</sup>, Jianguo Sun<sup>1</sup>

<sup>1</sup>Cancer institute, Xinqiao Hospital, Army Medical University, Chongqing, China; <sup>2</sup>Department of Oncology, Chongqing Qijiang District People's Hospital, Chongqing, China; <sup>3</sup>Department of Ultrasound, The 941<sup>st</sup> Hospital of the PLA Joint Logistic Support Force, Xining, China

**Contributions:** (I) Conception and design: G Yang, J Sun, B Luo, Q Yang; (II) Administrative support: J Sun; (III) Provision of study materials or patients: G Yang, B Luo, Q Yang; (IV) Collection and assembly of data: G Yang, Q Yang; (V) Data analysis and interpretation: G Yang, Q Yang, M Chen. (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

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

**Correspondence to:** Jianguo Sun. Cancer Institute, Xinqiao Hospital, Army Medical University, Chongqing 400037, China. Email: sunjg09@aliyun.com.

**Background:** At present, little attention has been paid to the impact of primary tumor size on the survival of nasopharyngeal carcinoma (NPC). We aimed to construct predictive nomograms for NPC that contain primary tumor size and Surveillance, Epidemiology, and End Results (SEER) stage as factors with the use of a national population-based registry.

**Methods:** A total of 1,574 NPC patients diagnosed between 2004 and 2011 were identified from the SEER database. Univariate and multivariate Cox analysis was used to screen out survival related clinical features. Then, two nomograms of 5-year overall survival (OS) and 5-year cancer-specific survival (CSS) were constructed using the significant clinical features. The concordance index (c-index) and a calibration plot were used to assess the performance of the nomograms. Furthermore, survival analyses were performed by risk score stratification.

**Results:** According to univariate and multivariate analyses, age, race, histology, primary tumor size, SEER stage, radiotherapy and chemotherapy were related to 5-year OS and CSS in NPC. All of these clinical features were included when constructing nomograms. The c-index values were 0.74 and 0.73 for the 5-year OS nomogram in the training cohort and testing cohorts, respectively. For the 5-year CSS nomogram, the c-index values were both 0.73 in the training cohort and testing cohort, respectively. The calibration plots showed that the predicted 5-year OS and CSS outcomes were in good agreement with the observed outcomes. All patients were divided into three groups according to the risk score stratification. Kaplan-Meier survival analyses showed that patients in the low-risk cohort had a greater 5-year survival benefit than patients in the medium-risk and high-risk cohorts ( $P < 0.05$ ).

**Conclusions:** The predictive nomograms, which contain primary tumor size and SEER stage, can predict the rates of 5-year OS and CSS of NPC patients and can be used as an auxiliary prediction tool for future clinical practice.

**Keywords:** Nasopharyngeal carcinoma (NPC); Surveillance, Epidemiology, and End Results (SEER); nomogram; primary tumor size; SEER stage

Submitted May 14, 2020. Accepted for publication Sep 26, 2020.

doi: 10.21037/tcr-20-2080

**View this article at:** <http://dx.doi.org/10.21037/tcr-20-2080>

<sup>^</sup> ORCID: Guangrong Yang, 0000-0001-6223-5938.

## Introduction

Nasopharyngeal carcinoma (NPC) occurs in various countries around the world, but it has the highest incidence in southern China, southeast Asia and northern Africa, with a prevalence of 20 to 30 cases per 100,000 people (1,2). NPC is caused by complex etiological factors, including Epstein-Barr virus (EBV) infection, genetic predisposition and other environmental risk factors, such as smoking, salted-fish consumption and occupational exposures (3). With the application of integrated radiotherapy and chemotherapy and intensity-modulated radiation therapy (IMRT) for NPC, the 5-year survival rate is approximately 80%, and the survival rate of stage I and II NPC can reach more than 90% (4-7). However, after treatment, NPC still has a high probability of local recurrence or distant metastasis, which affects the prognosis of patients. Of course, the efficacy and prognosis of NPC are affected by many factors. Studies have shown that differences in race, age, histology, staging, and choice of treatment options for patients with NPC will significantly affect survival (8-11).

A tool that can make individualized, reliable prognostic predictions for different patients with NPC can fully and effectively evaluate the prognosis and provide a basis for patients to choose appropriate and individualized treatment methods. Nomograms have become one of the most widely used statistical methods in clinical research. Many nomograms have been developed to predict the overall survival (OS) and progression-free survival (PFS) of NPC. The nomogram from Jiang Y includes factors such as age, T stage, N stage, peripheral blood neutrophil-lymphocyte ratio (NLR), and lactate dehydrogenase (LDH) (12). In terms of the OS of NPC, the histogram is better than the TNM staging system in patients receiving local IMRT and chemotherapy. There are also established nomograms using blood biomarkers to predict the OS of NPC (12).

The National Cancer Institute (NCI) of America established the Surveillance, Epidemiology, and End Results (SEER) database in 1973. It is one of the world's most recognized authoritative sources of cancer patient follow-up data, providing reliable data support for clinical research. Some studies have used the SEER database to construct a nomogram of the risk of second primary cancer in non-small-cell lung cancer as well as nomograms of survival prediction for breast cancer and early gastric cancer (13-15). Recently, Wu J screened NPC from the SEER database and constructed 3-, 5-, and 10-year OS prediction nomograms for NPC with age, sex, race, marital status, histology,

TNM staging, radiotherapy, and chemotherapy included as factors (16). However, we found that these studies had some shortcomings. First, some very important clinical features were not included in the nomograms, such as tumor size and SEER stage. Second, the diagnostic period of the screened data was from 2004 to 2015, and the 5- and 10-year survival follow-up time of this part of the data was insufficient. Insufficient follow-up time will obviously affect the accuracy of the model, so we have to question the accuracy of its modeling. Third, the data included in this study contained two versions of the American Joint Committee on Cancer (AJCC) staging system (sixth and seventh editions), although the authors converted the sixth edition to the seventh edition. The author explains the advantages of his model, mainly by comparing the nomograms with the AJCC 7th edition staging. The two versions were converted into stages, and their homogeneity and accuracy are unknown. There are multiple versions of the AJCC TNM staging data in the SEER database because the AJCC TNM staging system is constantly updated, which affects the homogeneity of the data. Therefore, the SEER database also established Derived SS1977 SEER staging, a simple and practical staging method to ensure the homogeneity of staging during data analysis. Studies have shown that the nomogram for early-onset gastric cancer displayed comparable results for TNM stage and SEER stage (15). We believe that when there are multiple versions of AJCC staging in the data, SEER staging can be applied instead. The aim of this study was to construct nomograms that include the primary tumor size and SEER stage, which can better predict the 5-year OS and cancer-specific survival (CSS) of NPC.

We present the following article in accordance with the STROBE reporting checklist (available at <http://dx.doi.org/10.21037/tcr-20-2080>).

## Methods

The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

### *Patients, clinical features and endpoints*

The data of this research came from the American SEER database. The software used was SEER \* Stat (version 8.3.6). NPC patients treated from 2004 to 2011 were selected with the following selection criteria: (I) The diagnosis period ranged from 2004 to 2011 because the clinical

information registered after 2004 is more detailed, and a follow-up time of at least 5 years is guaranteed; (II) the primary tumor site was “nasopharynx”; (III) there was only one primary tumor, and the diagnosis was confirmed by histology; (IV) referring to the International Classification of Diseases-Oncology, Third Edition (ICD-O-3), the pathological diagnosis codes included in this study were as follows: keratinizing squamous cell carcinoma (KSCC): 8070, 8071; differentiated nonkeratinizing squamous cell carcinoma (DNKSCC): 8072, 8073; keratinizing squamous cell carcinoma (UNKSCC): 8020, 8021, 8082, 8083; and other: 8074, 8075, 8010 (8,9). The clinical features included age, sex, race, histology, primary tumor size, SEER stage, radiotherapy and chemotherapy. A total of 3,264 patients with NPC diagnosed from 2004 to 2011 were initially selected, of whom 25 patients with unclear histological diagnostic information were excluded, 169 patients with unknown SEER stages were excluded, and 1,496 patients without baseline length of primary tumor were excluded; thus, 1,574 patients were finally included in the analysis. The clinical features were grouped as follows: age:  $\leq 39$ , 40–69, or  $\geq 70$  years old; sex: male or female; race: white, black, or other races; histology: KSCC, DNKSCC, UNKSCC, or other; primary tumor size:  $\leq 29$ , 30–49, 50–69, or  $\geq 70$  mm; SEER stage: localized (L), regional with direct extension (RE), regional with lymph nodes only (RN), regional with extension and nodes (RE+RN), or distant (D); chemotherapy: no or yes; and radiotherapy: no or yes.

The primary endpoint was OS, which was defined as the time interval from the diagnosis of NPC to death from any cause or the last follow-up. The secondary endpoint was CSS, which was defined as the time interval from the date of diagnosis of NPC to death caused by NPC.

### **Construction and testing of the nomograms**

First, univariate analyses were performed and survival curves were constructed for all clinical features. Next, all statistically significant clinical features in the univariate analyses ( $P < 0.05$ ) were included in the multivariate analysis to further identify significant predictive clinical features related to OS and CSS. Then, all patients were randomly divided into training and testing cohorts, with a split ratio of 8:2. The chi-square test was used to analyze the differences in baseline clinical characteristics between the training and testing cohorts to ensure that the baseline characteristics of the two cohorts were consistent. Nomograms to predict 5-year OS and CSS were constructed with significant predictive clinical

features using training cohort data. The discrimination of nomograms was determined by the concordance index (c-index), and the calibration was determined by calibration curves. In addition, we validated the predictive ability of nomograms with testing cohort data.

### **Survival analyses by risk score stratification**

The risk score of each patient was calculated based on the nomograms using the Model Predictions function of R software. Receiver operating characteristic (ROC) curves were plotted, and the area under the ROC curves (AUC) was calculated to determine the sensitivity and specificity of the risk score. Then, all patients were divided into three cohorts according to the risk score distribution, and Kaplan–Meier survival analyses were performed by risk score stratification.

### **Statistical analysis**

Cox univariate analyses, multivariate analyses and chi-square tests were conducted with SPSS 23.0 (IBM, Armonk, NY, USA). R package ‘rms’ was used to construct and validate the nomogram. R package ‘survivalROC’ was used to plot ROC and calculate AUC. R packages ‘survminer’ and ‘survival’ were used to performed survival analyses (R software, version 3.5.2, <http://www.r-project.org/>). A two-sided P value  $< 0.05$  was considered statistically significant.

## **Results**

### **Clinical features associated with OS and CSS**

The basic clinical features of all NPC patients are summarized in *Table 1*. Univariate analyses showed that age, race, histology, primary tumor size, SEER stage, radiotherapy and chemotherapy were associated with both OS and CSS (*Tables 1,2*). However, sex did not reach statistical significance. In addition, survival curves of each clinical feature are plotted (*Figures 1,2*). Furthermore, multivariate analyses showed that all significant clinical features in the univariate analyses were still associated with OS and CSS. The multivariate analysis revealed that black, other races, DNKSCC, UNKSCC, other histology, chemotherapy and radiotherapy were protective factors for NPC, while older age, white race, larger size of the primary tumor, regional node infiltration, and distant metastasis were adverse factors for the prognosis. We found

**Table 1** Clinical characteristics, univariate and multivariate analyses of NPC patients with OS as the endpoint

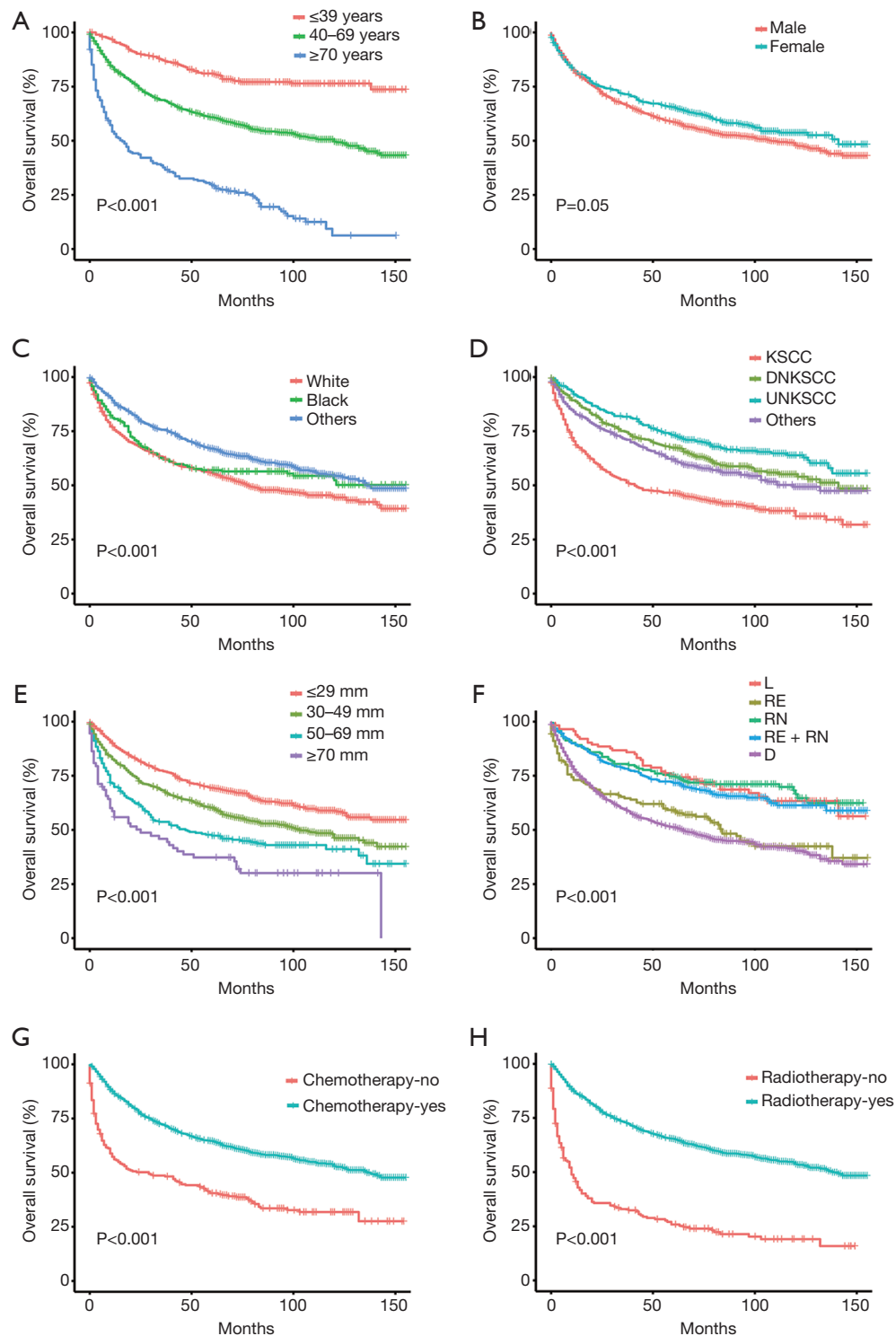
Characteristic	Number (total N=1,574), n (%)	Univariate analysis			Multivariate analysis		
		HR	95% CI	P value	HR	95% CI	P value
<b>Age</b>							
≤39	300 (19.06)	Reference		–	Reference	–	–
40–69	1,072 (68.11)	2.479	1.921–3.199	0.000	2.466	1.901–3.200	0.000
≥70	202 (12.83)	7.073	5.313–9.415	0.000	6.627	4.903–8.957	0.000
<b>Sex</b>							
Male	1,115 (70.84)	Reference	–	–	–	–	–
Female	459 (29.16)	0.849	0.720–1.001	0.051	–	–	–
<b>Race</b>							
White	718 (45.62)	Reference	–	–	Reference	–	–
Black	186 (11.82)	0.826	0.652–1.047	0.114	0.912	0.718–1.160	0.453
Other	670 (42.57)	0.681	0.582–0.795	0.000	0.764	0.648–0.900	0.001
<b>Histology</b>							
KSCC	511 (32.74)	Reference	–	–	Reference	–	–
DNKSCC	365 (23.19)	0.542	0.445–0.659	0.000	0.653	0.534–0.799	0.000
UNKSCC	341 (21.66)	0.418	0.338–0.518	0.000	0.576	0.461–0.720	0.000
Other	357 (22.68)	0.623	0.515–0.755	0.000	0.722	0.591–0.882	0.001
<b>Primary tumor size</b>							
≤29 mm	599 (38.06)	Reference	–	–	Reference	–	–
30–49 mm	653 (41.49)	1.424	1.201–1.689	0.000	1.414	1.188–1.684	0.000
50–69 mm	249 (15.82)	1.956	1.585–2.416	0.000	1.703	1.368–2.121	0.000
≥70 mm	73 (4.64)	2.862	2.106–3.889	0.000	2.434	1.772–3.335	0.000
<b>SEER stage</b>							
L	114 (7.24)	Reference	–	–	Reference	–	–
RE	157 (9.97)	1.995	1.357–2.934	0.000	2.552	1.713–3.801	0.000
RN	184 (11.69)	0.953	0.632–1.437	0.820	1.493	0.978–2.277	0.063
RE+RN	310 (19.70)	1.118	0.772–1.618	0.555	1.879	1.270–2.781	0.002
D	809 (51.40)	2.189	1.571–3.049	0.000	3.507	2.449–5.023	0.000
<b>Chemotherapy</b>							
No	262 (16.65)	Reference	–	–	Reference	–	–
Yes	1,312 (83.35)	0.453	0.381–0.539	0.000	0.573	0.459–0.716	0.000
<b>Radiotherapy</b>							
No	194 (12.33)	Reference	–	–	Reference	–	–
Yes	1,380 (87.67)	0.238	0.236–0.340	0.000	0.374	0.300–0.465	0.000

NPC, nasopharyngeal carcinoma; OS, overall survival; HR, hazard ratio; CI, confidence interval; KSCC, keratinizing squamous cell carcinoma; DNKSCC, differentiated nonkeratinizing squamous cell carcinoma; UNKSCC, keratinizing squamous cell carcinoma; SEER, surveillance, epidemiology and end results; L, localized; RE, regional with direct extension; RN, regional with lymph nodes only; RE+RN, regional with extension and nodes; D, distant.

**Table 2** Univariate and multivariate analyses of NPC patients with CSS as the endpoint

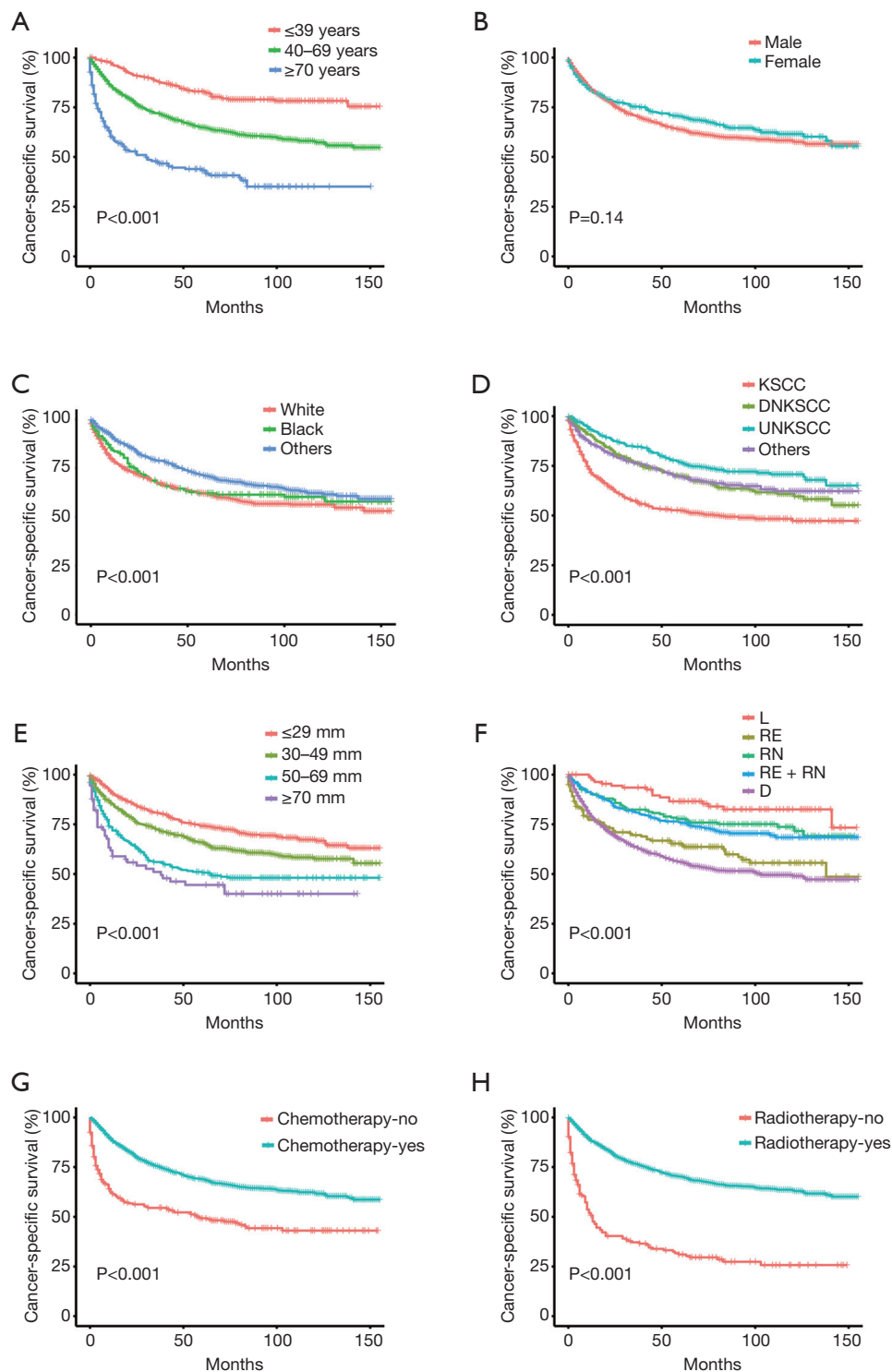
Characteristic	Number (total N=1,574), n (%)	Univariate analysis			Multivariate analysis		
		HR	95% CI	P value	HR	95% CI	P value
<b>Age</b>							
≤39	300 (19.06)	Reference	–	–	Reference	–	–
40–69	1,072 (68.11)	2.204	1.684–2.884	0.000	2.157	1.638–2.841	0.000
≥70	202 (12.83)	5.062	3.701–6.924	0.000	4.796	3.445–6.678	0.000
<b>Sex</b>							
Male	1,115 (70.84)	Reference	–	–	–	–	–
Female	459 (29.16)	0.870	0.725–1.045	0.138	–	–	–
<b>Race</b>							
White	718 (45.62)	Reference	–	–	Reference	–	–
Black	186 (11.82)	0.892	0.687–1.158	0.391	0.956	0.734–1.245	0.737
Other	670 (42.57)	0.718	0.604–0.855	0.000	0.799	0.665–0.960	0.016
<b>Histology</b>							
KSCC	511 (32.74)	Reference	–	–	Reference	–	–
DNKSCC	365 (23.19)	0.584	0.472–0.722	0.000	0.684	0.549–0.852	0.001
UNKSCC	341 (21.66)	0.419	0.330–0.532	0.000	0.555	0.433–0.713	0.000
Other	357 (22.68)	0.563	0.452–0.702	0.000	0.650	0.517–0.819	0.000
<b>Primary tumor size</b>							
≤29 mm	599 (38.06)	Reference	–	–	Reference	–	–
30–49 mm	653 (41.49)	1.398	1.152–1.696	0.001	1.361	1.117–1.658	0.002
50–69 mm	249 (15.82)	2.160	1.716–2.719	0.000	1.764	1.389–2.241	0.000
≥70 mm	73 (4.64)	2.855	2.025–4.025	0.000	2.287	1.606–3.257	0.000
<b>SEER stage</b>							
L	114 (7.24)	Reference	–	–	Reference	–	–
RE	157 (9.97)	3.140	1.856–5.312	0.000	3.971	2.317–6.803	0.000
RN	184 (11.69)	1.666	0.967–2.868	0.066	2.471	1.419–4.300	0.001
RE+RN	310 (19.70)	1.891	1.138–3.140	0.014	3.016	1.778–5.115	0.000
D	809 (51.40)	3.757	2.341–6.031	0.000	5.619	3.414–9.248	0.000
<b>Chemotherapy</b>							
No	262 (16.65)	Reference	–	–	Reference	–	–
Yes	1,312 (83.35)	0.471	0.388–0.572	0.000	0.560	0.436–0.719	0.000
<b>Radiotherapy</b>							
No	194 (12.33)	Reference	–	–	Reference	–	–
Yes	1,380 (87.67)	0.268	0.219–0.327	0.000	0.353	0.277–0.449	0.000

NPC, nasopharyngeal carcinoma; CSS, cancer-specific survival; HR, hazard ratio; CI, confidence interval; KSCC, keratinizing squamous cell carcinoma; DNKSCC, differentiated nonkeratinizing squamous cell carcinoma; UNKSCC, keratinizing squamous cell carcinoma; SEER, surveillance, epidemiology and end results; L, localized; RE, regional with direct extension; RN, regional with lymph nodes only; RE+RN, regional with extension and nodes; D, distant;



**Figure 1** Effects of different clinical characteristics on the OS of NPC. (A) Age; (B) sex; (C) race; (D) histology; (E) primary tumor size; (F) SEER stage; (G) chemotherapy; (H) radiotherapy. OS, overall survival; NPC, nasopharyngeal carcinoma; KSCC, keratinizing squamous cell carcinoma; DNKSCC, differentiated nonkeratinizing squamous cell carcinoma; UNKSCC, keratinizing squamous cell carcinoma; SEER, surveillance, epidemiology and end results; L, localized; RE, regional with direct extension; RN, regional with lymph nodes only; RE+RN, regional with extension and nodes; D, distant.





**Figure 2** Effects of different clinical characteristics on the CSS of NPC. (A) Age; (B) sex; (C) race; (D) histology; (E) primary tumor size; (F) SEER stage; (G) chemotherapy; (H) radiotherapy. CSS, cancer-specific survival; NPC, nasopharyngeal carcinoma; KSCC, keratinizing squamous cell carcinoma; DNKSCC, differentiated nonkeratinizing squamous cell carcinoma; UNKSCC, keratinizing squamous cell carcinoma; SEER, surveillance, epidemiology and end results; L, localized; RE, regional with direct extension; RN, regional with lymph nodes only; RE+RN, regional with extension and nodes; D, distant.

that compared to a primary tumor size of  $\leq 29$  mm a size of 30–49 mm is a risk factor, and the risks of tumors sized 50–69 and  $\geq 70$  mm progressively increased. We also found that a SEER stage of “RE” was a risk factor compared to “L”, “RN” was a risk factor compared to “L”, “D” had the highest risk for OS and CSS, and “RE” had a higher HR value than “RE+RN”. Hence, age, race, histology, primary tumor size, SEER stage, radiotherapy and chemotherapy were included when constructing the nomograms.

### Construction and testing of the nomograms

As shown in *Table 3*, there was no significant difference in baseline clinical characteristics between the training and testing cohort. A 5-year OS nomogram (*Figure 3A*) and a 5-year CSS nomogram (*Figure 3B*) were constructed with significant clinical features in the multivariate analyses using training cohort data. In the nomograms, each clinical feature was awarded one point according to the scale, and the total number of points could be used to estimate the probability of obtaining 5-year OS or CSS.

To test the predictive ability of the nomograms, the c-index was calculated, and a calibration plot was generated for both the training and testing cohorts. In the training cohort, the c-index was 0.74 (95% CI: 0.73–0.75) for the 5-year OS nomogram, and in the testing cohort, the c-index was 0.73 (95% CI: 0.71–0.75) for the 5-year OS nomogram. Similarly, in the training cohort, the c-index was 0.73 (95% CI: 0.72–0.74) for the 5-year CSS nomogram, and in the verification cohort, the c-index was 0.73 (95% CI: 0.70–0.76) for the 5-year CSS nomogram. The calibration curves showed that the predicted 5-year OS and CSS outcomes were in good agreement with the observed outcomes in both the training and testing cohorts (*Figure 4*).

### Survival analyses by risk score stratification

The risk score of each patient was calculated. For 5-year OS, the risk score ranged from 0.09 to 34.16 in the training cohort (*Figure 5A*), and the risk score ranged from 0.06 to 16.02 in the testing cohort (*Figure 5B*). For 5-year CSS, the risk score ranged from 0.09 to 27.50 in the training cohort (*Figure 5C*), and the risk score ranged from 0.06 to 18.79 in the testing cohort (*Figure 5D*). According to the risk score distribution, all patients were divided into three cohorts: the low-risk cohort, scored from 0 to 1; the medium-risk cohort, scored from 1 to 5; and the high-risk cohort, scored higher than 5. In addition, receiver operating

characteristics (ROCs) curves of risk scores in each cohort were plotted. For 5-year OS, the training cohort had an AUC of 0.864 (*Figure 6A*), and the testing cohort had an AUC of 0.874 (*Figure 6B*). For 5-year CSS, the training cohort had an AUC of 0.871 (*Figure 6C*), and the testing cohort had an AUC of 0.885 (*Figure 6D*). Furthermore, Kaplan–Meier survival analyses of 5-year OS and CSS by risk score stratification were performed. The results showed that patients in the low-risk cohort had more obvious 5-year survival benefits than patients in the medium-risk and high-risk cohorts. These findings indicate that a higher risk score is positively associated with a higher risk of death (*Figure 7*).

### Discussion

Previous studies have constructed nomograms to predict the survival rate of NPC (12,16,17). In particular, Wu J screened NPC from the SEER database and constructed 3-, 5-, and 10-year OS prediction nomograms for NPC. Although our study is also based on NPC in the SEER database, we screened data for at least 5 years of follow-up, constructed a 5-year OS and CSS prediction model, and used the SEER stage instead of the 2 versions of TNM staging. Thus, the homogeneity of staging is guaranteed. Additionally, our model includes the important factor of primary tumor size. Therefore, the model we constructed guarantees that the data have sufficient follow-up times and stage homogeneity to ensure the accuracy of the predictive model. The largest difference in this study is that we also constructed a 5-year CSS nomogram for NPC. We used Cox regression to screen out information related to the OS and CSS of NPC. In particular, no studies have shown whether the baseline length of the primary tumor is related to the survival of NPC. We found that the baseline length of the tumor is closely related to OS and CSS. We included age, sex, race, histology, primary tumor size, SEER stage, radiotherapy and chemotherapy as factors to construct reliable predictive nomograms to help oncologists and patients make more accurate clinical decisions. The OS and CSS of NPC are depend on related to the modality of radiotherapy and chemotherapy. Unfortunately, there are no details of radiotherapy and chemotherapy in SEER database. IMRT was the standard radiotherapy for NPC as early as 2004 (18). And radiotherapy alone has been the standard treatment for early nasopharyngeal carcinoma, induction chemotherapy plus concurrent chemoradiotherapy is a standard treatment regimen for locally advanced nasopharyngeal carcinoma (19,20). Furtherly, since the 1990s, the combination of



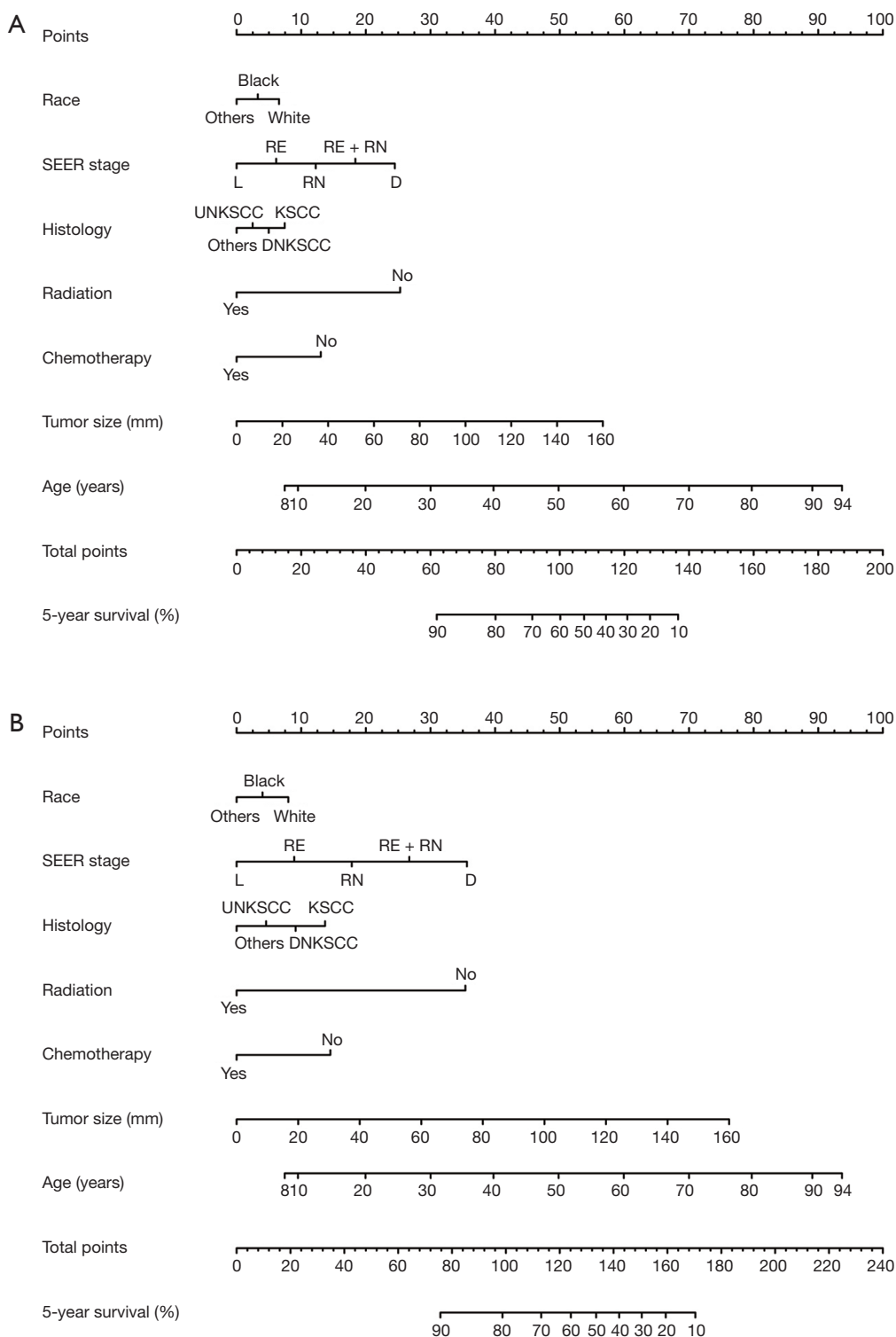
**Table 3** Baseline clinical characteristics of the training cohort and testing cohort

Characteristic	Training (N=1,259)	Testing (N=315)	Total (N=1,574)	P value
Age				0.565
≤39	234	66	300	
40–69	860	212	1,072	
≥70	165	37	202	
Race				0.315
White	579	139	718	
Black	141	45	186	
Other	539	131	670	
Histology				0.364
KSCC	407	104	511	
DNKSCC	282	83	365	
UNKSCC	281	60	341	
Other	289	68	357	
Primary tumor size				0.363
≤29 mm	469	130	599	
30–49 mm	522	131	653	
50–69 mm	207	42	249	
≥70 mm	61	12	73	
SEER stage				0.937
L	90	24	114	
RE	128	29	157	
RN	149	35	184	
RE+RN	246	64	310	
D	646	163	809	
Chemotherapy				0.808
No	211	51	262	
Yes	1,048	264	1,312	
Radiotherapy				0.464
No	159	35	194	
Yes	1,100	280	1,380	

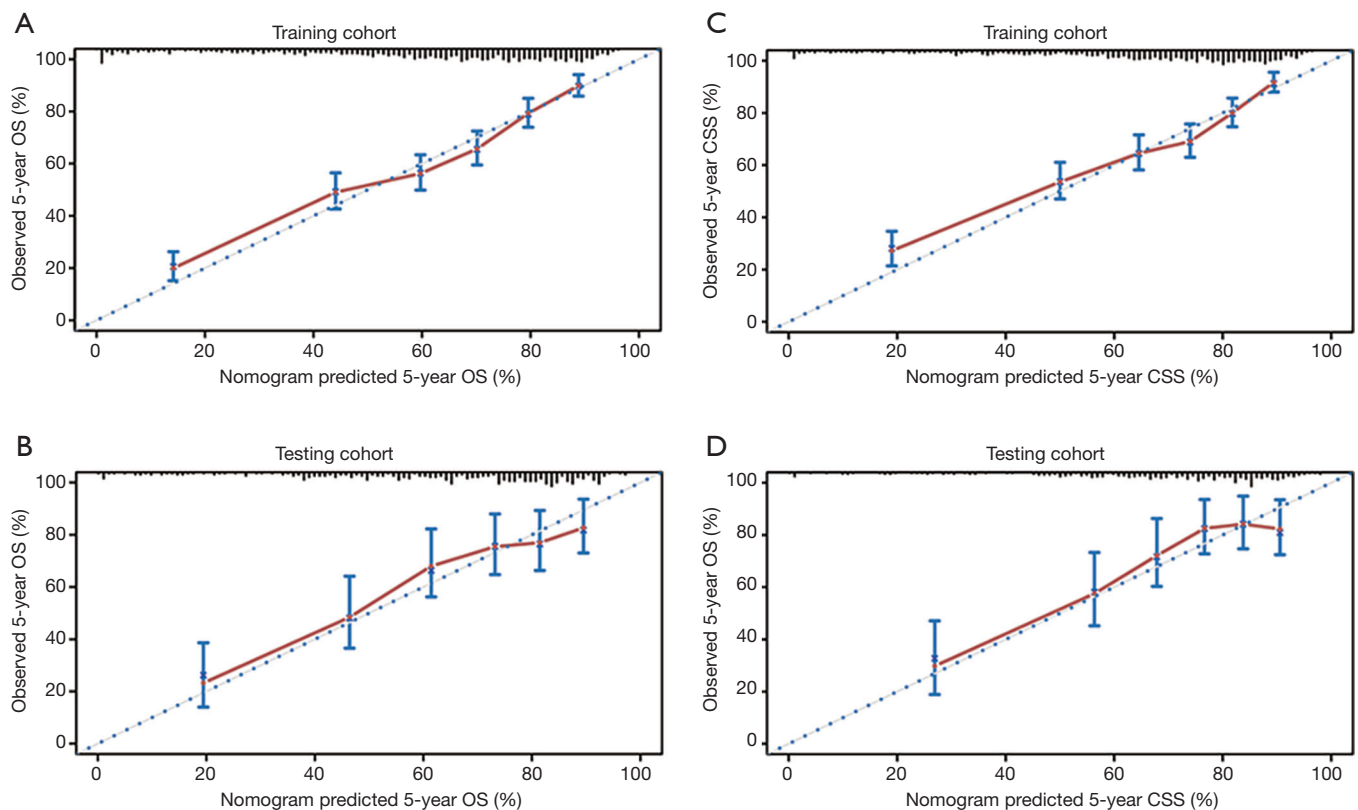
KSCC, keratinizing squamous cell carcinoma; DNKSCC, differentiated nonkeratinizing squamous cell carcinoma; UNKSCC, keratinizing squamous cell carcinoma; SEER, surveillance, epidemiology and end results; L, localized; RE, regional with direct extension; RN, regional with lymph nodes only; RE+RN, regional with extension and nodes; D, distant.

cisplatin and 5-FU has become the standard chemotherapy regimen for NPC (21) Studies have shown that from 2004 to 2011, the treatment of locally advanced nasopharyngeal

carcinoma was mainly based on induction chemotherapy with PF (cisplatin and 5-fluorouracil), TP (docetaxel and cisplatin) and TPF (docetaxel, cisplatin, and 5-fluorouracil)



**Figure 3** Nomograms for NPC patients. (A) 5-year of OS nomogram; (B) 5-year of CSS nomogram. NPC, nasopharyngeal carcinoma; OS, overall survival; CSS, cancer-specific survival; KSCC, keratinizing squamous cell carcinoma; DNKSCC, differentiated nonkeratinizing squamous cell carcinoma; UNKSCC, keratinizing squamous cell carcinoma; SEER, surveillance, epidemiology and end results; L, localized; RE, regional with direct extension; RN, regional with lymph nodes only; RE+RN, regional with extension and nodes; D, distant.

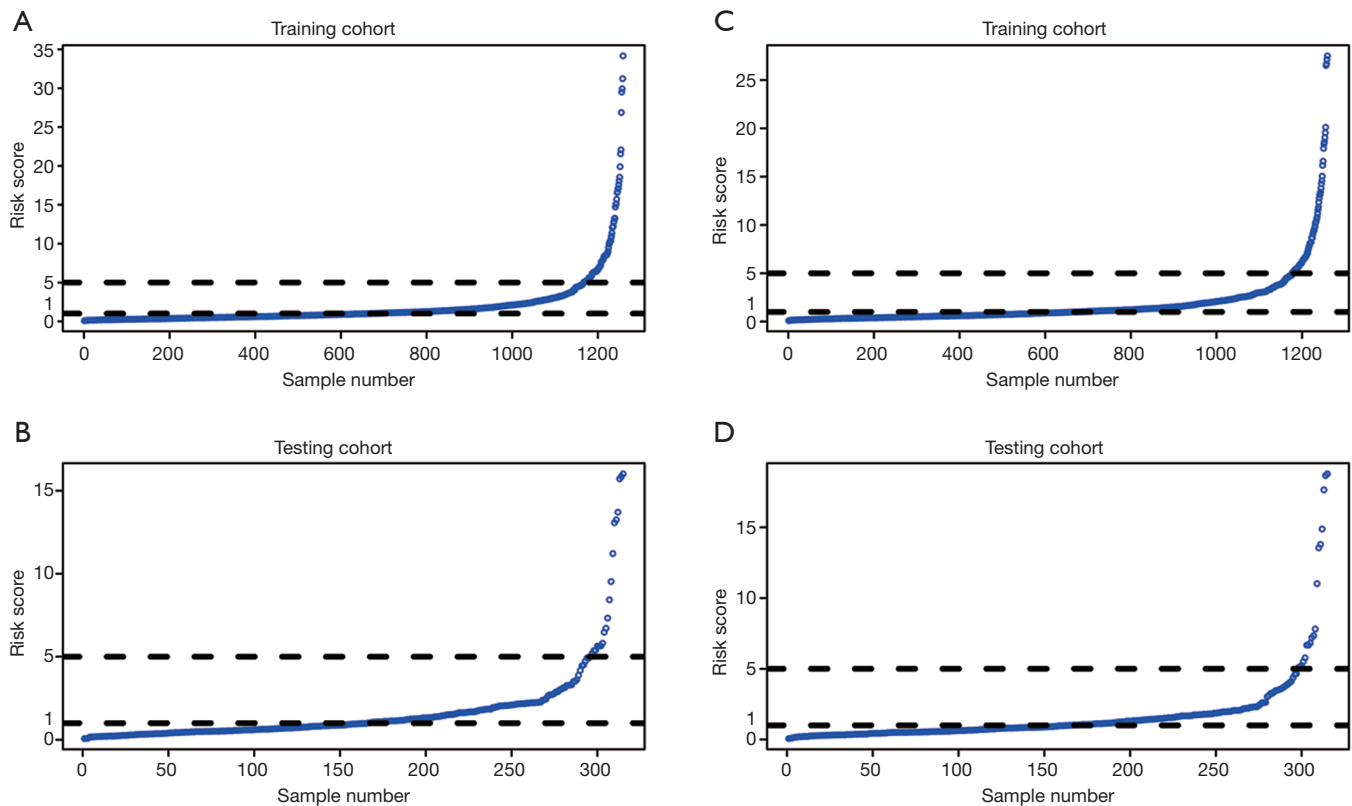


**Figure 4** The calibration curves for predicting OS and CSS. (A) 5-year OS training calibration; (B) 5-year OS testing calibration; (C) 5-year CSS training calibration; (D) 5-year CSS testing calibration (D). OS, overall survival; CSS, cancer-specific survival.

remains followed by cisplatin-based concurrent chemoradiotherapy (22,23). Theoretically, the radiotherapy techniques and chemotherapy regimens among subgroups are similar, and it may not affect the conclusion. To the best of our knowledge, this is the first study to describe the construction and testing of predictive nomograms that combine primary tumor size and SEER staging with other clinical features to predict 5-year OS and CSS for NPC.

The AJCC T stage of NPC is determined based on the anatomical site of primary tumor invasion, not the size of the primary tumor. Some studies have suggested that the primary tumor volume of NPC is an independent risk factor for radiotherapy and affects the prognosis of patients (24). Some studies have also suggested that the tumor volume of NPC is positively correlated with the T stage, and it is suggested that the tumor volume should be included in the TNM stage as a predictive factor. However, this suggestion has not received much attention (25). Prior to this study, no research has focused on the effect of primary tumor size on the survival of NPC. The primary tumor size is recorded in

the SEER database. For example, Yu, C used the primary tumor size of young gastric cancer patients in the SEER database to model prognosis (15), and some studies used the primary tumor size of breast cancer and lung cancer to model prognosis (13,14). However, since the T stage of NPC is not based on the size of the primary tumor, no scholar has ever focused on whether the size of the primary nasopharyngeal tumor in the SEER database is related to the survival prognosis of patients. In this study, given that optimal cutoff categorizing tumor size was identified as a significant independent factor in univariate and multivariate analyses, it is reasonable to presume the potential impact of tumor size in NPC. Kaplan-Meier analysis showed that the size of the primary tumor was closely related to OS and CSS ( $P < 0.05$ ). In fact, based on our results, compared with a primary tumor sized  $\leq 29$  mm, a tumor size of 30–49 mm is a risk factor, and the risks of the 50–69 and  $\geq 70$  mm groups are further increased. It is worth noting that primary tumor size could be of great importance to demonstrate the OS and CSS prognostic risk in this situation. Of note, tumor

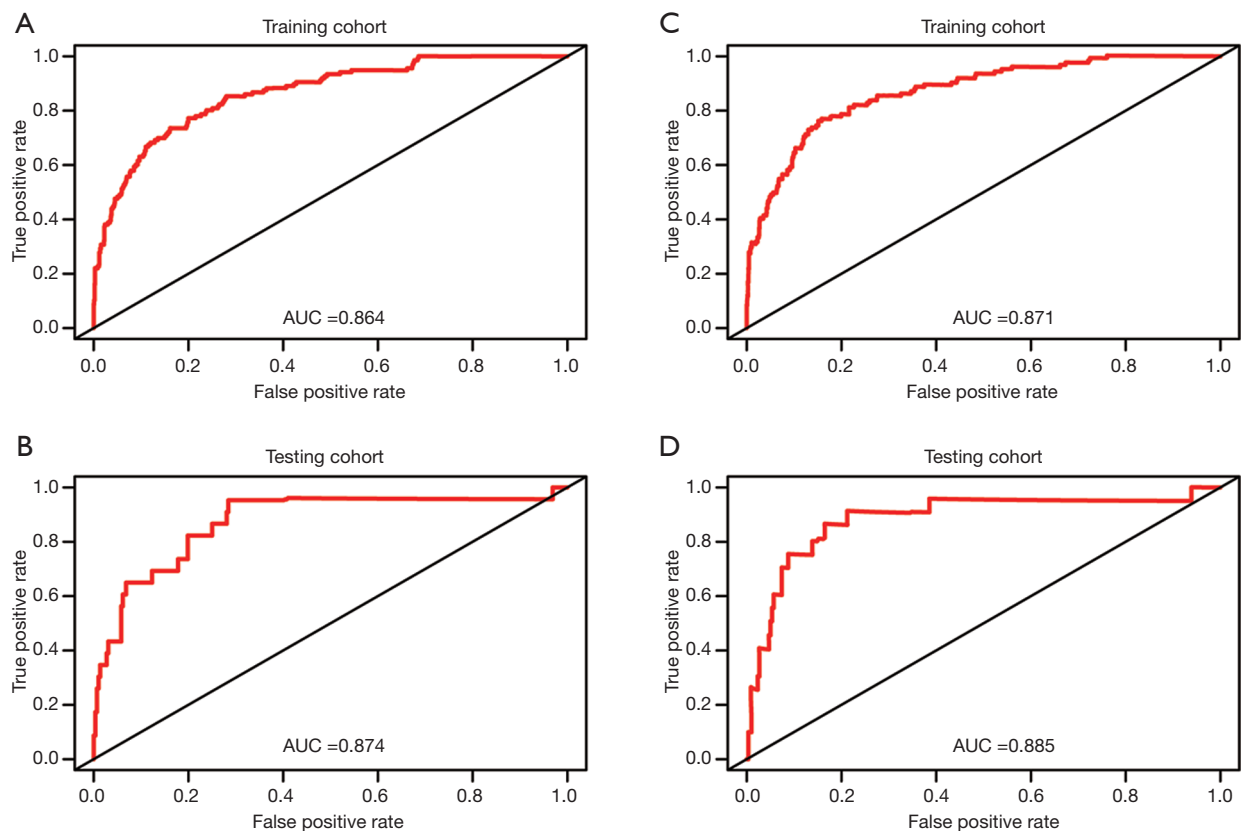


**Figure 5** Risk score distribution. (A) 5-year OS training risk score distribution; (B) 5-year OS testing risk score distribution; (C) 5-year CSS training risk score distribution; (D) 5-year CSS testing risk score distribution. OS, overall survival; CSS, cancer-specific survival.

size  $\geq 70$  mm was a strong predictive indicator for both OS and CSS. It may be potentially correlated with more advanced AJCC stage and occult disseminated tumor cells.

The SEER database is one of the world's most recognized authoritative sources of tumor patient follow-up data, but there are some limitations. There are multiple versions of the AJCC TNM staging data in the SEER database because the AJCC TNM staging system is constantly updated, which affects the homogeneity of the data. Therefore, the SEER database also established Derived SS1977 SEER staging, a simple and practical staging method to ensure the homogeneity of staging during data analysis. We analyzed the SEER stages by Kaplan-Meier analysis and found that the OS and CSS of NPC in different SEER stages were different ( $P < 0.05$ ). Univariate and multivariate analyses showed that "RE" was a risk factor compared to "L" ( $P < 0.05$ ), "RN" was a risk factor compared to "L" ( $P > 0.05$ ), "D" had the highest risk for OS and CSS, and "RE" had a higher HR value than "RE+RN", which shows that "RE" is associated with a higher risk of death than "RE+RN".

We identified 7 independent predictive factors through univariate and multivariate analyses, including age, race, histology, primary tumor size, SEER stage, radiotherapy and chemotherapy, to construct predictive models. In our study, we found that older patients had a higher risk of death and that individuals of other races had a survival advantage over white individuals, which is consistent with the findings of previous studies (26). In our study, we included NPC patients from nonendemic areas where KSCC is the predominant histological type. For patients from nonendemic areas, the KSCC or DNKSCC subtypes have ERCC1 overexpression compared with the UNKSCC subtype. Patients with overexpression of ERCC1 may have a lower response to radiotherapy (8). A recognized risk factor for nonkeratinizing squamous cell carcinoma (NKSCC) is EBV infection (1), and EBV has been confirmed to have a significant correlation with the progression of NPC (27). For patients with KSCC, the predominant pattern of disease failure is local recurrence, while for NKSCC, distant recurrence is the predominant pattern (8).



**Figure 6** ROCs curves for the nomograms of NPC. (A) 5-year OS training ROC; (B) 5-year OS testing ROC; (C) 5-year CSS training ROC; (D) 5-year CSS testing ROC. ROCs, receiver operating characteristics; OS, overall survival; CSS, cancer-specific survival.

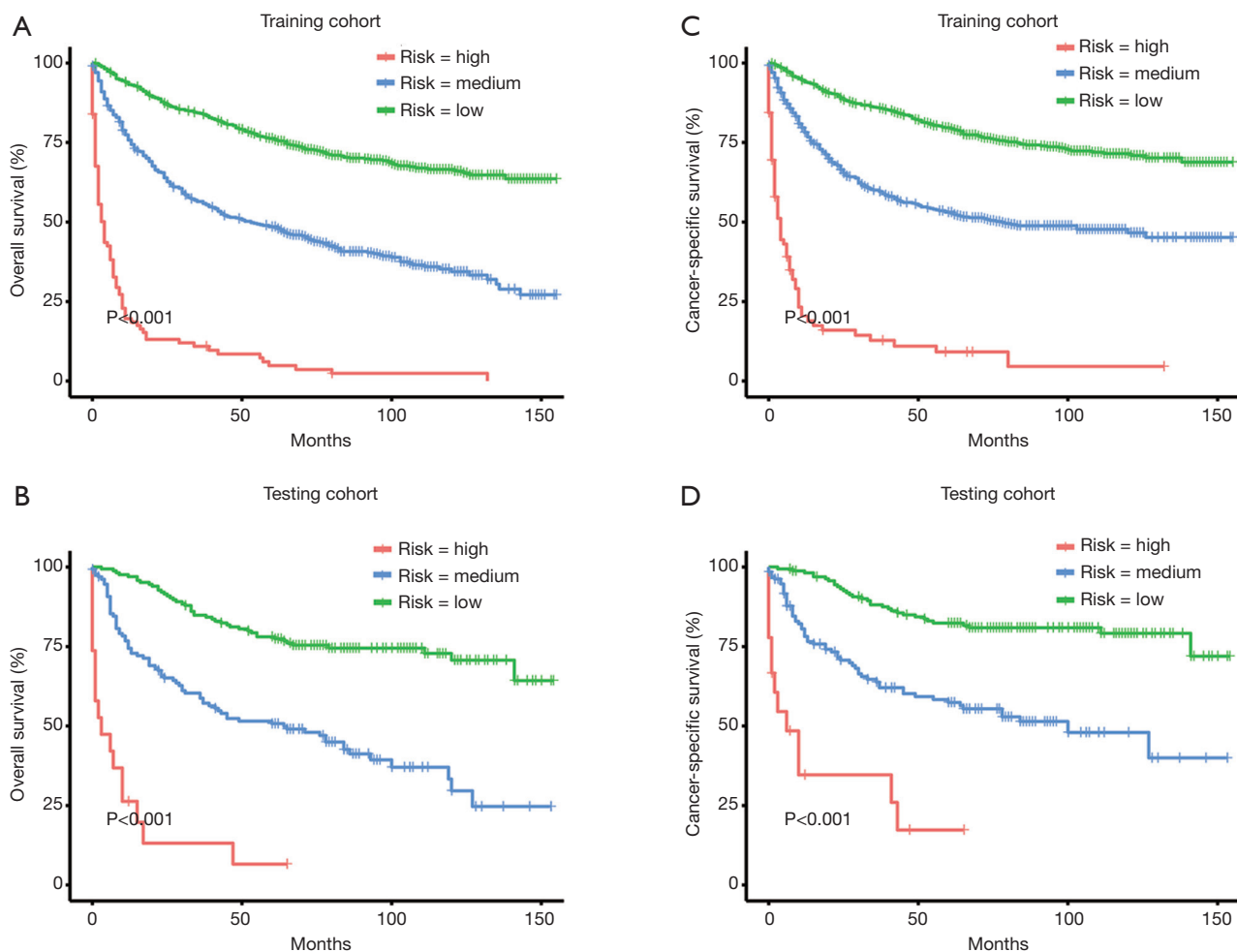
Different pathological types have different biological behaviors and different survival prognoses. Therefore, we included histology as one of the factors in our model. With advances in radiotherapy techniques, the OS of patients with NPC has considerably improved, and concurrent chemoradiotherapy (CCRT) is now established as the standard treatment for locoregionally advanced NPC (28). Therefore, it is necessary to include radiotherapy and chemotherapy as factors in the model.

In our research, to ensure that the data we screened have a follow-up period of at least 5 years, we screened only NPCs with a diagnosis from 2004 to 2011 in the SEER database. The overlap between the calibration curve and the standard reference curve was high, and the predicted outcome of the 5-year survival rate of the predictive model was in good agreement with the observed outcome. Therefore, the predictive model is reliable.

The advantage of this prediction model is that the score is based on some basic clinical indicators, which is different from the traditional TNM staging system. The predictive

model is more personal and facilitates communication between doctors and patients. It may become an auxiliary prediction tool in future clinical practice. Independent predictive factors, such as the primary tumor size, may have a significant impact on individualized survival predictions. Patients with different prognoses may choose different treatment options or nursing and follow-up methods. In addition, oncologists can provide more appropriate adjuvant treatment and monitoring for patients in different risk score cohorts. For example, high-risk patients should receive more intense treatment and follow-up, especially those who are older, those who have larger primary tumors and those who have a wider range of tumor invasion.

This study still has some limitations. First, the clinical features included are limited, and there are many other important factors related to tumor prognosis, such as Performance Status (PS) score, EBV, BMI, specific organ metastasis information, blood biomarkers, and gene mutation status. BMI can reflect the nutritional status of NPC, which is an important prognostic factor of OS of



**Figure 7** Kaplan-Meier curves for all three risk cohorts based on the predictors from the nomograms. (A) 5-year OS training curve by risk score; (B) 5-year OS testing curve by risk score; (C) 5-year CSS training curve by risk score; (D) 5-year CSS testing curve by risk score. OS, overall survival; CSS, cancer-specific survival.

nasopharyngeal carcinoma (29). More clinical features can help us to understand NPC more comprehensively and can be used to construct more accurate predictive nomograms. Second, the specific treatment information was not complete. There was no targeted treatment information available. Finally, the data in the SEER database comes from the United States, but the high incidence areas of NPC are in Asia, and we have only performed an internal validation of the model, not an external validation. It is best to use multiple different data sets and populations from different regions to verify the model to obtain an estimate of the model's performance, thereby determining the applicability of the model. Therefore, the nomograms constructed in this study needs further clinical samples for verification.

## Conclusions

This study was based on SEER data, and multivariate analysis showed that clinical characteristics such as race, histology, SEER stage, radiotherapy, chemotherapy, primary tumor size, and age were related to NPC survival prognosis. The predictive nomograms, which contain primary tumor size and SEER stage, can predict the 5-year OS and CSS rates of NPC patients and can be used as an auxiliary prediction tool for future clinical practice.

## Acknowledgments

The authors are grateful to all the staff in the National



Cancer Institute (USA) for their contribution to the SEER program.

*Funding:* This study was supported by the grant from the National Key Research & Development Program (No. 2016YFC0106400).

### Footnote

*Reporting Checklist:* The authors have completed the STROBE reporting checklist. Available at <http://dx.doi.org/10.21037/tcr-20-2080>

*Conflicts of Interest:* All authors have completed the ICMJE uniform disclosure form (available at <http://dx.doi.org/10.21037/tcr-20-2080>). The authors have no conflicts of interest to declare.

*Ethical Statement:* The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The author GR Yang has gotten the access to the SEER database (accession number: 17254-Nov2018).

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

### References

1. Carioli G, Negri E. Global trends in nasopharyngeal cancer mortality since 1970 and predictions for 2020: Focus on low-risk areas. *Int J Cancer* 2017;140:2256-64.
2. Chua ML, Chan AT. Gemcitabine: a game changer in nasopharyngeal carcinoma. *Lancet* 2016;388:1853-4.
3. Tang LL, Chen WQ, Xue WQ, et al. Global trends in incidence and mortality of nasopharyngeal carcinoma. *Cancer Lett* 2016;374:22-30.
4. Chen C, Yi W, Gao J, et al. Alternative endpoints to the 5-year overall survival and locoregional control for nasopharyngeal carcinoma: A retrospective analysis of 2,450 patients. *Mol Clin Oncol* 2014;2:385-92.
5. Peng G, Wang T, Yang KY, et al. A prospective, randomized study comparing outcomes and toxicities of intensity-modulated radiotherapy vs. conventional two-dimensional radiotherapy for the treatment of nasopharyngeal carcinoma. *Radiother Oncol* 2012;104:286-93.
6. Mao YP, Xie FY, Liu LZ, et al. Re-evaluation of 6th edition of AJCC staging system for nasopharyngeal carcinoma and proposed improvement based on magnetic resonance imaging. *Int J Radiat Oncol Biol Phys* 2009;73:1326-34.
7. Li Y, Ou X, Hu C. Validation and suggestion of eighth T classifications of the UICC/AJCC staging system for nasopharyngeal carcinoma patients: a retrospective analysis. *Jpn J Clin Oncol* 2018;48:927-33.
8. Wu SG, Lian CL, Wang J, et al. The effect of histological subtypes on survival outcome in nasopharyngeal carcinoma after extensive follow up. *Ann Transl Med* 2019;7:768.
9. Sun J, Huang Z, Hu Z, et al. Benefits of local tumor excision and pharyngectomy on the survival of nasopharyngeal carcinoma patients: a retrospective observational study based on SEER database. *J Transl Med* 2017;15:116.
10. Xu Y, Huang T, Fan L, et al. Patterns and Prognostic Value of Lymph Node Metastasis on Distant Metastasis and Survival in Nasopharyngeal Carcinoma: A Surveillance, Epidemiology, and End Results Study, 2006-2015. 2019;2019:4094395.
11. Guo R, Wu H, Wang J, et al. Lymph Node Status and Outcomes for Nasopharyngeal Carcinoma According to Histological Subtypes: A SEER Population-Based Retrospective Analysis. 2019;36:3123-33.
12. Li J, Chen S, Peng S, et al. Prognostic nomogram for patients with Nasopharyngeal Carcinoma incorporating hematological biomarkers and clinical characteristics. *Int J Biol Sci* 2018;14:549-56.
13. Zhou H, Shen J, Zhang Y, et al. Risk of second primary malignancy after non-small cell lung cancer: a competing risk nomogram based on the SEER database. *Ann Transl Med* 2019;7:439.
14. Wang Y, Yang Y, Chen Z, et al. Development and validation of a novel nomogram for predicting distant metastasis-free survival among breast cancer patients. *Ann Transl Med* 2019;7:537.
15. Yu C, Zhang Y. Development and validation of prognostic nomogram for young patients with gastric cancer. *Ann Transl Med* 2019;7:641.
16. Wu J, Zhou Q, Pan Z, et al. Development and validation

- of a nomogram for predicting long-term overall survival in nasopharyngeal carcinoma: A population-based study. *Medicine (Baltimore)* 2020;99:e18974.
17. Jiang Y, Qu S, Pan X, et al. Prognostic Nomogram For Locoregionally Advanced Nasopharyngeal Carcinoma. *Sci Rep* 2020;10:861.
  18. Qu S, Liang ZG, Zhu XD. Advances and challenges in intensity-modulated radiotherapy for nasopharyngeal carcinoma. *Asian Pac J Cancer Prev* 2015;16:1687-92.
  19. Chang JT, Ko JY, Hong RL. Recent advances in the treatment of nasopharyngeal carcinoma. *J Formos Med Assoc* 2004;103:496-510.
  20. Wang Q, Xu G, Xia Y, et al. Comparison of induction chemotherapy plus concurrent chemoradiotherapy and induction chemotherapy plus radiotherapy in locally advanced nasopharyngeal carcinoma. *Oral Oncol* 2020;111:104925.
  21. Vokes EE, Athanasiadis I. Chemotherapy of squamous cell carcinoma of head and neck: the future is now. *Ann Oncol* 1996;7:15-29.
  22. Vermorken JB, Remenar E, van Herpen C, et al. Cisplatin, fluorouracil, and docetaxel in unresectable head and neck cancer. *N Engl J Med* 2007;357:1695-704.
  23. He Y, Guo T, Wang J, et al. Which induction chemotherapy regimen followed by cisplatin-based concurrent chemoradiotherapy is the best choice among PF, TP and TPF for locoregionally advanced nasopharyngeal carcinoma? *Ann Transl Med* 2019;7:104.
  24. Chen C, Fei Z, Pan J, et al. Significance of primary tumor volume and T-stage on prognosis in nasopharyngeal carcinoma treated with intensity-modulated radiation therapy. *Jpn J Clin Oncol* 2011;41:537-42.
  25. Zhou JY, Chong VF, Khoo JB, et al. The relationship between nasopharyngeal carcinoma tumor volume and TNM T-classification: a quantitative analysis. *Eur Arch Otorhinolaryngol* 2007;264:169-74.
  26. Lv JW, Huang XD, Chen YP, et al. A National Study of Survival Trends and Conditional Survival in Nasopharyngeal Carcinoma: Analysis of the National Population-Based Surveillance Epidemiology and End Results Registry. *Cancer Res Treat* 2018;50:324-34.
  27. Liao C, Zhou Q, Zhang Z, et al. Epstein-Barr virus-encoded latent membrane protein 1 promotes extracellular vesicle secretion through syndecan-2 and synaptotagmin-like-4 in nasopharyngeal carcinoma cells. *Cancer Sci* 2020;111:857-68.
  28. Sun XS, Xiao BB, Lin C, et al. Establishment and validation of two nomograms to predict the benefit of concurrent chemotherapy in stage II-IVa nasopharyngeal carcinoma patients with different risk factors: Analysis based on a large cohort. *Cancer Med* 2020;9:1661-70.
  29. Hong JS, Hua YJ, Su L, et al. Modified-Nutrition Index is a Significant Prognostic Factor for the Overall Survival of the Nasopharyngeal Carcinoma Patients who Undergo Intensity-modulated Radiotherapy. *Nutr Cancer* 2017;69:1011-8.

**Cite this article as:** Yang G, Luo B, Yang Q, Chen M, Yang X, Sun J. Construction of nomograms for nasopharyngeal carcinoma containing primary tumor size and SEER stage. *Transl Cancer Res* 2020;9(11):6939-6954. doi: 10.21037/tcr-20-2080