A Novel Prognostic Model Predicts Overall Survival in Patients With Nasopharyngeal Carcinoma Based on Clinical Features and Blood Biomarkers

Changchun Lai
Maoming people's hospital
Chunning Zhang
Maoming people's hospital
Hualiang Lv
Maoming people's hospital
Hanqing Huang
Maoming People's Hospital
Xia Ke
Maoming people's hospital
Chuchan Zhou
Maoming people's hospital
Hao Chen
Sun Yat-sen University Cancer Center
Shulin Chen
Sun Yat-sen University Cancer Center
Lei Zhou (zhouleilab@163.com)
Maoming People's Hospital https://orcid.org/0000-0002-5293-867X

Research article

Keywords: Lasso-Cox regression, model, nomogram, nasopharyngeal carcinoma, prognostic

Posted Date: June 9th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-33442/v1

Version of Record: A version of this preprint was published at Cancer Medicine on May 11th, 2021. See the published version at https://doi.org/10.1002/cam4.3839.

Abstract

Background: To develop and validate a novel prognostic model to estimate overall survival (OS) in nasopharyngeal carcinoma (NPC) patients based on clinical features and blood biomarkers. And assess its incremental value to TNM staging system, clinical treatment, and Epstein-Barr virus DNA (EBV DNA) for individual OS estimation.

Methods: We retrospectively analyzed 519 consecutive NPC. A prognostic model was generated by using the Lasso-Cox regression model in training cohort (n= 346). Then comparison of predictive accuracy between the novel prognostic model, TNM staging, clinical treatment, and EBV DNA using concordance index (C-index), time-dependent ROC (tdROC), and decision curve analysis (DCA). Subsequently, a nomogram for OS incorporating the prognostic model, TNM staging and clinical treatment was built. Finally, we stratified patients into high- risk and low-risk groups according to the model risk score, and the survival time of these two groups was analyzed using Kaplan–Meier survival plots. All the results were validated in the independent validation cohort (n= 173).

Results: Using the Lasso-Cox regression, a prognostic model was established consisting of 13 variables with respect to patient prognosis. The C-index, tdROC and DCA all showed the prognostic model had good predictive accuracy and discriminatory power than TNM staging, clinical treatment and EBV DNA in training cohort. Nomogram consisting of the prognostic model, TNM staging, clinical treatment and EBV DNA shown some superior net benefit. According to the model risk score, we split the patients into two subgroups: low- risk (risk score ≤ -1.423) and high-risk (risk score > -1.423). There had significant differences in OS between the two subgroups of patients. In the validation cohort, similar results were obtained.

Conclusions: The proposed novel prognostic model based on clinical features and serological markers represents a promising signature for estimating OS in NPC patients.

Background

Nasopharyngeal carcinoma (NPC) is a common malignancy of head and neck in Southern China and Southeast Asia(Torre et al. 2015). Distant metastasis is a leading cause of treatment failure in patients with NPC, almost 70% patients are initially diagnosed with locoregionally advanced disease(OuYang et al. 2013). Although the new radiotherapeutic techniques, chemotherapy regimens, and surgical techniques have improved the survivability of NPC patients, the 5-year survival rate remains unsatisfactory(Wang et al. 2010).

Currently, the tumor–node–metastasis (TNM) staging system is usually used to aid in determining the prognosis of cancer patients and in suggesting the treatment strategy. However, the NPC patients with the same TNM stage received similar treatment, there were still large patients showed a poor prognosis(Zhang et al. 2016). Therefore, the TNM staging have some limitations in predicting survival rate of patients with NPC or guiding treatments. This because it is entirely based on the anatomical range

of the existing tumors, but not evaluate the intrinsic biological heterogeneity of tumors(Ng et al. 2014). Consequently, a lot of biomarkers have been researched to improve the prognosis prediction and treatment efficiency of NPC patients, such as patient clinical characteristics(Wang et al. 2014), blood biomarkers(Janvilisri 2015), and radiomics(Zhang et al. 2017). However, most predictive models are integrated with TNM staging system to improve the predictive accuracy for clinical outcome, which makes them not applicable to the patients with uncertain clinical TNM staging. In addition, some models aren't widely used in clinical practice due to time-consuming, high-cost, high risk of radiation exposure and not routine medical examinations in the majority of primary care hospitals.

Recently, more and more blood biomarkers are used to predict clinical outcome in many cancers because they are benefit us by being cost-effective, higher accessibility and non-invasive, as well as by allowing the simple detection. Thus, the current study sought to construct a novel prognostic model predicts overall survival in NPC patients based on clinical features and routine laboratory blood biomarkers. And assess its incremental value to TNM staging system, clinical treatment, and Epstein-Barr virus DNA (EBV DNA) for individual OS estimation. Finally, we validate its effectiveness in patients from the same institution.

Material And Methods

Patient selection and data collection

From January 2009 and December 2011, patients with diagnosed NPC who were treated for the first time at Sun Yat-sen University Cancer Center (Guangzhou, China) were retrospectively enrolled in this study. This study was performed in accordance with the guidelines outlined in the Declaration of Helsinki and was approved by the Clinical Research Ethics Committee of the Sun Yat-sen University Cancer Center, and all patients provided written informed consent at the first visit to our center. The inclusion criteria for the study are as follows: a: pathological evidence of NPC, were not any malignancies besides NPC; b: complete baseline clinical information, blood-biomarkers data, and follow-up data; c: all the blood-biomarkers data were collected one week before anti-tumor therapy.

The following clinicopathologic data were collected for each enrolled patient: gender, age, family history of malignant tumors, smoking index (SI): day × the year of cigarette smoking(Tokarskaya et al. 2002), body mass index (BMI), TNM staging was assigned according to the 8th AJCC TNM classification(OuYang et al. 2017), and clinical treatment. Relevant baseline blood-biomarkers including white blood cell (WBC), neutrophils (N), lymphocyte (L), monocyte (M), platelet (PLT), hemoglobin (HGB), total protein (TP), albumin (ALB), globulin (GLOB), C-reactive protein (CRP), apolipoprotein AI (APOA), apolipoprotein B (APOB), dehydrogenase (LDH), high density lipoprotein (HDL), cystatin C(Cys-C), plasma EBV DNA copy number (EBV DNA), EBV immunoglobulin A/viral capsid antigen (VCA-IgA), EBV immunoglobulin A/early antigen (EA-IgA), neutrophil to lymphocyte ratio (NLR)(Liao et al. 2018), derived neutrophil-lymphocyte ratio (dNLR)(Zhao et al. 2018), lymphocyte to monocyte ratio (LMR), platelet to lymphocyte ratio (PLR), systemic immune-inflammation index (SII): (platelet × neutrophils) /

lymphocyte(Chen et al. 2017), albumin to globulin ratio (AGR), C-reactive protein to albumin ratio (CAR), APOA to APOB ratio (ABR), advanced lung cancer inflammation index (ALI): (BMI × albumin) / NLR(Ozyurek et al. 2018), prognostic nutritional index (PNI): ALB (g/L) + 5 × lymphocyte count × 10^9/L(He et al. 2018), and prognostic index (PI): score 0 for CRP 10 mg/L or less and white cell count 11 × 10^9/L or less, patients with only one of these abnormalities were allocated a score of 1, and if both of them were elevated were allocated a score of 2(Kasymjanova et al. 2010).

Patients Follow Up

The patients' survival data follow-up was done by referring to the clinic attendance records, email, and phone calls, all patients were followed-up after discharge until December 2015. The endpoint of this study was overall survival (OS) was defined as the period from the first time of diagnose to the last follow-up or date of patient death.

Statistical Analyses

Statistical analyses were performed using IBM SPSS Statistical software version 19.0 (IBMCorp., Chicago, IL, USA) and R version 3.6.0 (http://www.R-project.org). Continuous variables were transformed into categorical variables, and the cut-off value of all variables were recognized by the R package "survival" and "survminer"(Diboun et al. 2006). Differences in distributions of clinical characteristics and blood-biomarkers between the training cohort and validation cohort were used to Pearson Chi-square test. We utilized the least absolute shrinkage and selection operator (LASSO) regression to select the most useful prognostic factors in the training cohort. According to the regulation weight λ, LASSO selects variables correlated to the measured outcome by shrinking coefficients weights, down to zero for the ones not correlated to outcome(Tibshirani 1997). The optimal values of the penalty parameter λ were determined through 10-fold cross-validation with the 1 standard error of the minimum criteria (the 1-SE criteria)(Goeman 2010; Tibshirani 1997). Based on the optimal λ value, a list of prognostic variables with associated coefficients was screened out. Then a novel prognostic model was constructed according to calculate the risk score for each patient based on each prognostic variable and its associated coefficient. Comparison of predictive accuracy between the prognostic model, TNM staging, clinical treatment, and EBV DNA for individualized survival was evaluated by concordance index (C-index)(Brentnall et al. 2018), time-dependent ROC (tdROC)(Kamarudin et al. 2017), and decision curve analysis (DCA)(Vickers et al. 2008). Nomograms (by the package of rms in R) for prediction of OS were built based on prognostic model risk score, TNM staging, clinical treatment, and EBV DNA. The calibration plots of nomograms were used to assess the consistency between the predicted survival and the observed survival with bootstrapping (1000 bootstrap resamples)(Shim et al. 2015). Finally, the patients in the training and validation cohort were split into low-risk and high-risk groups according to the optimal cut-off value of prognostic model risk score. Kaplan-Meier method and log-rank tests were used to assess differences in

OS between the predicted high- and low-risk groups. Results with two-sided p values of < 0.05 were considered statistically significant.

Results

Baseline clinical and characteristics

In the present study, 346 eligible patients were analyzed in the training cohort, and 173 patients were included in the validation cohort. The median follow-up was 51.4 months (interquartile range (IQR):42.1–67.0) in the training cohort and 50.4 months (IQR: 41.9–66.0 months) in the validation cohort. In the training cohort, the 1-, 3-, and 5-year OS rates were as follows: 97.4%, 83.8%, and 48.3%. In the validation cohort, the 1-, 3-, and 5-year OS rates were the following: 94.2%, 84.4%, and 42.8%.

The optimal cut-off value for each continuous variable as follows: age (60 years), smoking index (20.0), BMI (26.33 kg/m2), WBC (4.3×10^9/L), Neutrophils (7.0×10^9/L), Lymphocyte (1.41×10^9/L), Monocyte (0.4×10^9/L), Platelet (293.0×10^9/L), HGB (130.0 g/L), NLR (3.91), dNLR (2.46), LMR (3.4), PLR (208.89), SII (1141.96), TP (77.2 g/L), ALB (42.4 g/L), GLOB (33.1 g/L), AGR (1.36), CRP (5.47 mg/L), CAR (0.16), APOA (1.28 g/L), APOB (1.03 g/L), ABR (0.96), LDH (167.5 U/L), HDL (1.16 U/L), Cys-C (0.94 mg/L), ALI (262.33), and PNI (47.35). Patients' clinical characteristics and blood-biomarkers for the patients were listed in Table 1. There was no significant difference in the distribution of clinical characteristics and blood-biomarkers between training cohort and validation cohort.

Table 1
Demographics and clinical characteristics of patients in the training and validation cohort

Characteristic	Training cohort	Validation cohort	χ^2 value	*P* value
	n=(346)	n=(173)		
	No. (%)	No. (%)		
Gender			2.435	0.119
Male	264 (76.3%)	121 (69.9%)		
Female	82 (23.7%)	52 (30.1%)		
Age (years)			0.956	0.328
≤60	310 (89.6%)	150 (86.7%)		
>60	36 (10.4%)	23 (13.3%)		
Family history			0.079	0.778
Yes	90 (26.0%)	47 (27.2%)		
No	256 (74.0%)	126 (72.8%)		
Smoking index[a]			1.661	0.198
≤20.0	226 (65.3%)	103 (59.5%)		
>20.0	120 (34.7%)	70 (40.5%)		
BMI (kg/m^2)			1.250	0.264
≤26.33	298 (86.1%)	155 (89.6%)		
>26.33	48 (13.9%)	18 (10.4%)		
TNM stage[b]			1.965	0.580
I	12 (3.5%)	5 (2.9%)		
II	45 (13.0%)	24 (13.9%)		
III	172 (49.7%)	76 (43.9%)		
IV	117 (33.8%)	68 (39.3%)		
Treatment			0.242	0.623
Rad	58 (16.8%)	32 (18.5%)		
Rad and Che	288 (83.2%)	141 (81.5%)		
WBC ($10^9/L$)			0.007	0.933

≤4.3	57 (16.5%)	29 (16.8%)		
>4.3	289 (83.5%)	144 (83.2%)		
Neutrophils (10^9/L)			0.879	0.348
≤7.0	306 (88.4%)	148 (85.5%)		
>7.0	40 (11.6%)	25 (14.5%)		
Lymphocyte (10^9/L)			0.099	0.753
≤1.41	145 (41.9%)	75 (43.4%)		
>1.41	201 (58.1%)	98 (56.6%)		
Monocyte (10^9/L)			0.466	0.495
≤0.4	175 (50.6%)	82 (47.4%)		
>0.4	171 (49.4%)	91 (52.6%)		
Platelet (10^9/L)				
≤293.0	298 (86.1%)	154 (89.0%)	0.857	0.355
>293.0	48 (13.9%)	19 (11.0%)		
HGB (g/L)			1.130	0.288
≤130.0	106 (30.6%)	61 (35.3%)		
>130.0	240 (69.4%)	112 (64.7%)		
NLR			0.621	0.431
≤3.91	263 (76.0%)	126 (72.8%)		
>3.91	83 (24.0%)	47 (27.2%)		
dNLR			0.692	0.405
≤2.46	254 (73.4%)	121 (69.9%)		
>2.46	92 (26.6%)	52 (30.1%)		
LMR			0.479	0.489
≤3.4	141 (40.8%)	76 (43.9%)		
>3.4	205 (59.2%)	97 (56.1%)		
PLR			0.055	0.815
≤208.89	277 (80.1%)	140 (80.9%)		
>208.89	69 (19.9%)	33 (19.1%)		

SII			0.263	0.608
≤1141.96	294 (85.0%)	144 (83.2%)		
>1141.96	52 (15.0%)	29 (16.8%)		
TP (g/L)			1.585	0.208
≤77.2	273 (78.9%)	128 (74.0%)		
>77.2	73 (1.1%)	45 (26.0%)		
ALB (g/L)			0.148	0.701
≤42.4	132 (38.2%)	63 (36.4%)		
>42.4	214 (61.8%)	110 (63.6%)		
GLOB (g/L)			0.095	0.758
≤33.1	274 (79.2%)	139 (80.3%)		
>33.1	72 (20.8%)	34 (19.7%)		
AGR			1.406	0.236
≤1.36	108 (30.6%)	45 (26.0%)		
>1.36	240 (69.4%)	128 (74.0%)		
CRP (mg/L)			0.087	0.768
≤5.47	268 (77.5%)	132 (76.3%)		
>5.47	78 (22.5%)	41 (23.7%)		
CAR			0.101	0.751
≤0.16	282 (81.56%)	139 (80.3%)		
>0.16	64 (18.5%)	34 (19.7%)		
APOA (g/L)			0.097	0.756
≤1.28	167 (48.3%)	81 (46.8%)		
>1.28	179 (51.7%)	92 (53.2%)		
APOB (g/L)			0.262	0.609
≤1.03	218 (63.0%)	105 (60.7%)		
>1.03	128 (37.0%)	68 (39.3%)		
ABR			0.038	0.845
≤0.96	40 (11.6%)	19 (11.0%)		

>0.96	306 (88.4%)	154 (89.0%)		
LDH (U/L)			0.004	0.950
≤167.5	193 (55.8%)	96 (55.5%)		
>167.5	153 (44.2%)	77 (44.5%)		
HDL (U/L)			1.114	0.291
≤1.16	179 (51.7%)	81 (46.8%)		
>1.16	167 (48.3%)	92 (53.2%)		
Cys-C (mg/L)			1.640	0.200
≤0.94	222 (64.2%)	101 (58.4%)		
>0.94	124 (35.8%)	72 (41.6%)		
EBV DNA, copy/mL			4.369	0.358
$<10^3$	169 (48.8%)	70 (40.5%)		
10^3-9,999	72 (20.8%)	36 (20.8%)		
10^4-99,999	58 (16.8%)	39 (22.5%)		
10^5-999,999	29 (8.4%)	17 (9.8%)		
$\geq 10^6$	18 (5.2%)	11 (6.4%)		
VCA-IgA			0.081	0.960
<1:80	59 (17.1%)	28 (16.2%)		
1:80–1:320	208 (60.1%)	106 (61.3%)		
≥ 1:640	79 (22.8%)	39 (22.5%)		
EA-IgA			1.338	0.512
<1:10	116 (32.7%)	49 (28.3%)		
1:10–1:20	110 (31.8%)	60 (34.7%)		
≥1:40	123 (35.5%)	64 (37.0%)		
ALI			0.173	0.677
≤262.33	94 (27.2%)	50 (28.9%)		
>262.33	252 (72.8%)	123 (71.1%)		
PNI			0.058	0.810
≤47.35	63 (18.2%)	33 (19.1%)		

>47.35	283 (81.8%)	140 (80.9%)		
PI			0.644	0.725
0	275 (79.5%)	141 (81.5%)		
1	64 (18.5%)	30 (17.3%)		
2	7 (2.0%)	2 (1.2%)		

a: Smoking index: the number of cigarettes smoked each day × the year of cigarette smoking;

b: TNM stage was classified according to the AJCC 8th TNM staging system;

Abbreviations: BMI: body mass index; TNM: Tumor Node Metastasis stage; Rad: radiotherapy; Che: chemotherapy; WBC: white blood cell; HGB: hemoglobin; NLR: neutrophil/lymphocyte ratio; dNLR: neutrophil/WBC-neutrophil ratio; LMR: lymphocyte/monocyte ratio; PLR : platelet/lymphocyte ratio; SII: systemic immune-inflammation index; TP: total protein; ALB: albumin; GLOB: globulin; AGR: ALB/GLOB ratio; CRP: C-reactive protein; CAR: C-reactive protein/albumin ratio; APOA: apolipoprotein AI; APOB: apolipoprotein B; ABR: APOA/APOB ratio; LDH: lactic dehydrogenase; HDL: high density lipoprotein; Cys-C: cystatin C; EBV: Epstein-Barr virus; VCA-IgA: EBV immunoglobulin A/viral capsid antigen; EA-IgA: EBV immunoglobulin A/early antigen; ALI: advanced lung cancer inflammation index; PNI: prognostic nutritional index; PI: prognostic index.

Construction Of The Novel Prognostic Model

In order to find the prognostic variables in the training cohort, we used a LASSO-Cox regression analysis model. Figure 1A showed the change in trajectory of each prognostic variable was analyzed. Moreover, we plotted the partial likelihood deviance versus log (λ) in Fig. 1B, where λ was the tuning parameter. The value of λ was 0.03987 was chosen by 10-fold cross-validation via the 1-SE criteria. So, we obtained 13 variables with nonzero coefficients at the value λ chosen by cross-validation. These prognostic variables included age, BMI, HGB, PLT, LMR, CRP, CAR, GLOB, AGR, LDH, Cys-C, ALI, and PNI. The coefficients of each prognostic variable were presented in Fig. 1C. Then the prognostic model risk score of each patient was computed according to the summation of 13 variables multiplied coefficient from the LASSO regression generated: The prognostic model risk score = -0.680 + (0.569 × age) - (0.280 × BMI + (0.101 × HGB) - (0.554 × PLT) + (0.197 × LMR) - (0.199 × CRP) + (0.186 × CAR) + (1.248 × GLOB) - (0.137 × AGR) - (0.194 × LDH) + (1.248 × Cys-C) - (0.137 × ALI) - (0.194 × PNI). Where each variable was valued as 0 or 1; a value of 0 was assigned when the variable was less than or equal to the corresponding cut-off value, and a value of 1 otherwise.

Comparison of predictive accuracy between the novel prognostic model, TNM Staging, clinical treatment, and EBV DNA

As shown in Table 2, in training cohort, the C-index of the prognostic model was 0.786 (95% confidence interval (CI): 0.728–0.844), which was higher than the C-indices of the TNM staging, clinical treatment, and EBV DNA, with values of 0.740 (95% CI: 0.690–0.790), 0.554 (95% CI: 0.521–0.586), and 0.691 (95% CI: 0.623–0.758), respectively. The C-index for the prognostic model was statistically significantly higher

than the C-index by the clinical treatment (P < 0.001), and EBV DNA (P = 0.013). In the validation cohort, the C-index of prognostic model was both higher than that of TNM staging and clinical treatment, but it was a little lower than that of EBV DNA. Subsequently, we compared the area under ROC curve (AUC) between the novel prognostic model, TNM staging, clinical treatment, and EBV DNA using tdROC. In general, the AUC of novel prognostic model was higher than others both in the training cohort (Fig. 2A) and validation cohort (Fig. 2B). Finally, the DCA displayed the prognostic model had a better overall net benefit than TNM staging, clinical treatment and EBV DNA across a wide range of reasonable threshold probabilities in training cohort (Fig. 3A) and validation cohort (Fig. 3B). These results indicated that the novel prognostic model displayed better accuracy in predicting OS compared with the TNM staging, clinical treatment and EBV DNA.

Table 2
The C-index of the prognostic model, TNM staging, Treatment, and EBV DNA for prediction of OS in the training cohort and validation cohort

Factors		
	C-index (95% CI)	***P***
For training cohort		
Prognostic model	0.786 (0.728 ~ 0.844)	
TNM staging	0.740 (0.690 ~ 0.790)	
Treatment	0.554 (0.521 ~ 0.586)	
EBV DNA	0.691 (0.623 ~ 0.758)	
Prognostic model vs TNM staging		0.067
Prognostic model vs Treatment		< 0.001
Prognostic model vs EBV DNA		0.013
For validation cohort		
Prognostic model	0.697 (0.612 ~ 0.734)	
TNM staging	0.655 (0.575 ~ 0.734)	
Treatment	0.529 (0.470 ~ 0.588)	
EBV DNA	0.734 (0.659 ~ 0.813)	
Prognostic model vs TNM staging		0.310
Prognostic model vs Treatment		< 0.001
Prognostic model vs EBV DNA		0.511
C-index = concordance index; CI = confidence interval; P values are calculated based on normal approximation using function rcorrp.cens in Hmisc package.		

Building And Validating A Predictive Nomogram

The prognostic model risk score, TNM staging, clinical treatment, and EBV DNA were integrated into nomograms to predict the 1-, 3-, and 5-year OS in the training cohort (Fig. 4). Each variable was assigned a corresponding point value based on its contribution to the model. The point values for all the predictor variables are summed to arrive at the "Total Points" axis, and then a line is drawn vertically down from total points to predict the patient's probability of OS at 1-, 3-, and 5-year. Finally, A calibration plot was used to visualize the performance of the nomogram. Nomogram predicted and actual observed outcome at 1-, 3-, and 5-year OS were plotted on the x-axis and y-axis, respectively. The 45° line represented the best prediction, the solid dark red line represented the performance of the nomograms. The calibration curve showed that the 1-, 3-, and 5-year OS predicted by the nomograms were consistent with actual observations (Figs. 5), indicating that the nomograms did well performance. The nomograms and calibration curve in the validation cohort were showed in the supplementary Fig. 1 and supplementary Fig. 2, respectively.

Survival analyses of NPC patients according to prognostic model risk score

The optimal cut-off value of prognostic model risk score for predicting survival was determined to be -1.423 by R package "survminer" (Fig. 6A). According to the cutoff value, we classified patients into two different subgroups, of which low-risk group with risk score $\leq$ -1.423, and high-risk group with risk score > -1.423. The distribution of the prognostic model risk score in training and validation cohort were showed in Fig. 6B and Fig. 6C, respectively.

In the training cohort, for the high-risk group, the median OS was 44.4 months (IQR: 24.7–66.1) The 1-, 3- and 5-year probabilities of OS were 95.4%, 63.2%, and 33.3%, respectively. For the low-risk group, the median OS was 61.2 months (IQR: 44.6–67.8). The 1-, 3- and 5-year probabilities of OS were 98.1%, 90.7% and 53.3%, respectively. In the validation cohort, low-risk group also had higher survival probabilities than high-risk group at 1-, 3-, and 5-year, respectively (Table 3). Kaplan–Meier curves were compared to assess the differences in survival between low-risk and high-risk groups. The results showed statistically significant better OS for low-risk group versus high-risk group both in training cohort and validation cohort ($p < 0.05$; Fig. 7).

Table 3
OS and OS rate in high-risk and low-risk groups according to the model risk score in the training and validation cohort

Parameter	Training cohort			Validation cohort		
	High-Risk Group	Low-Risk Group	Total	High-Risk Group	Low -Risk Group	Total
No. of patients	87	259	346	49	124	173
Median (IQR)	44.4 (24.7–66.1)	61.2 (44.6–67.8)	51.4 (42.1–67.0)	45.8 (26.1–64.1)	53.5 (43.0-66.3)	50.4 (41.9–66.0)
No. of OS						
1-Year	83 (95.4%)	254 (98.1%)	337 (97.4%)	44 (89.8%)	119 (96.0%)	163 (94.2%)
3-Year	55 (63.2%)	235 (90.7%)	290 (83.8%)	36 (73.5%)	110 (88.7%)	146 (84.4%)
5-Year	29 (33.3%)	138 (53.3%)	167 (48.3%)	17 (34.7%)	57 (46.0%)	74 (42.8%)
Abbreviations: OS: overall survival; IQR: interquartile range.						

Discussion

In this study, we successfully established a novel prognostic Lasso-Cox regression model based on clinical features and blood-biomarkers for individualized prediction of the OS for NPC patient. The novel prognostic model showed better predictive accuracy and discrimination compared with the traditional AJCC TNM staging system, clinical treatment, and EBV DNA, which successfully split NPC patients into high-risk and low-risk groups, and the two groups of patients exhibited significant differences in the OS.

The present prognostic model consisting of 13 prognostic variables: age, BMI, HGB, PLT, LMR, CRP, CAR, GLOB, AGR, LDH, Cys-C, ALI, and PNI. All the prognosis variables had been reported to be associated with survival in NPC patients except ALI(Du et al. 2015; Du et al. 2014; Hu et al. 2009; Li et al. 2018; Li et al. 2016; Yuan et al. 2016; Zeng et al. 2016), these were credible evidence supporting our analysis results. The ALI was devised to assess the degree of systemic inflammation in patients with advanced non-small cell lung cancer patients(Jafri et al. 2013). Subsequently, ALI had also been demonstrated to be a prognostic factor of survival in some cancers(Hua et al. 2019; Park et al. 2017; Shibutani et al. 2019). The difference between the ALI and other inflammatory markers was that the ALI contained not only indices related to inflammation but also the body mass index (BMI), which was reported to correlate with the sarcopenic status(Shibutani et al. 2019). So, this was the first study to indicate ALI as a prognostic marker in NPC patients.

Subsequently, we compared the predictive accuracy and discrimination of the novel prognostic model with TNM staging, clinical treatment and EBV DNA using C-index, tdROC and DCA. The results all showed the prognostic model had good predictive accuracy and discriminatory power than TNM staging, clinical treatment and EBV DNA in training cohort. Above similar results were observed in the validation cohort except EBV DNA. The C-index of the prognostic model was slightly lower than EBV DNA, but there was no statistical difference. The most likely explanation was that this was a retrospective analysis, and there may have some potential patient selection bias. Then the nomogram consisting of the prognostic model, TNM staging, clinical treatment and EBV DNA shown some superior net benefit. Finally, according to the model risk score, we split the patients into two subgroups: low-risk and high-risk, and there were significant differences in OS between the two subgroups of patients. These results indicated the novel prognostic model had good predictive accuracy and discrimination for estimating OS in NPC patients.

Although previous studies had established some models for predicting NPC survival, this study still had several merits compared to other studies: 1. The prognostic model only included basic clinical and routine laboratory data, which did not include some not routinely available markers, such as EBV DNA(Lertbutsayanukul et al. 2018), and circulating tumor cells (CTC)(Ou et al. 2019; You et al. 2019). and this model was low-cost, non-invasive, no risk of radiation exposure, and convenient. So, this model could widely and safely used in clinical practice, especially in primary hospitals. 2. The prognostic model was constructed by using the newly algorithm LASSO-Cox model, as a statistical method for screening variables to establish the prognostic model, which enabled to adjust for model's over fitting and avoid extreme predictions. So, the predictive accuracy could be improved significantly, and this approach had been applied in many study(Moons et al. 2004; Srivastava et al. 2009; Tibshirani 1997). 3. Many previous models usual integrated with TNM staging and/or clinical treatment to improve the predictive accuracy for clinical outcome(Chang et al. 2013; Li et al. 2018; Liang et al. 2016; Tang et al. 2016; Xia et al. 2013; Yang et al. 2015; Zhang et al. 2013), which made them not applicable to the patients with uncertain clinical TNM staging. This model can be used for patients with TNM staging remained unclear because of it was not include TNM staging.

There also had several drawbacks of this study. First, this was a retrospective analysis and selection bias might exist. Second, the treatment effect heterogeneity for metachronous metastasis patients might bring confounding effects. Third, the endpoint of this study was OS, and the effect of the model for predicting disease-free survival (DFS), distant metastasis free survival (DMFS) and locoregional relapse-free survival (LRFS) in NPC patients were not assessed(Yang et al. 2019). It was better clinical application that the endpoint combined OS with DFS and DMFS. Finally, it was a single-institutional study with a relatively small sample size. Thus, a large-scale and multicenter validation of the model will be needed in the future

Conclusions

In conclusion, we have established a novel prognostic model based on clinical features and blood biomarkers, which showed better predictive accuracy than traditional TNM staging, clinical treatment, and

EBV DNA, and the nomograms comprising the prognostic model, TNM staging, clinical treatment, and EBV DNA can reinforce the prognostic ability of the prognostic model. Therefore, the simple, convenient, low-cost, non-invasive, no risk of radiation exposure, precise and understandable prognostic model was useful for clinicians in decision-making, individual patient counselling and scheduling patients' follow-ups for NPC patients.

Abbreviations

BMI
body mass index; TNM:Tumor Node Metastasis stage; Rad:radiotherapy; Che:chemotherapy; WBC:white blood cell; HGB:hemoglobin; NLR:neutrophil/lymphocyte ratio; dNLR:neutrophil/WBC-neutrophil ratio; LMR:lymphocyte/monocyte ratio; PLR:platelet/lymphocyte ratio; SII:systemic immune-inflammation index; TP:total protein; ALB:albumin; GLOB:globulin; AGR:ALB/GLOB ratio; CRP:C-reactive protein; CAR:C-reactive protein/albumin ratio; APOA:apolipoprotein AI; APOB:apolipoprotein B; ABR:APOA/APOB ratio; LDH:lactic dehydrogenase; HDL:high density lipoprotein; Cys-C:cystatin C; EBV:Epstein-Barr virus; VCA-IgA:EBV immunoglobulin A/viral capsid antigen; EA-IgA:EBV immunoglobulin A/early antigen; ALI:advanced lung cancer inflammation index; PNI:prognostic nutritional index; PI:prognostic index.

Declarations

Acknowledgments

We thank Sun Yat-sen University Cancer Center for providing support on research conditions in this study, and high-level hospital construction research project of Maoming people's hospital.

Authors' contributions

All authors contributed to this manuscript, including conception and design (HC, SLC, LZ), acquisition of data (HC), analysis and interpretation of data (CCL, CNZ), material support (SLC), study supervision (HLL, HQH, XK, CCZ), and writing, review and revision of the manuscript (CCL, CNZ, HLL, HQH, XK, CCZ, HC, SLC, LZ).

Funding

This work was supported by the Natural Science Foundation of Guangdong Province (2018A030313622), funded by the Open Project Program of the State Key Laboratory of Proteomics (SKLPO201703), and high-level hospital construction research project of Maoming people's hospital.

Availability of data and materials

The datasets analyzed during the current study are not publicly available due to patient privacy concerns, but are available from the corresponding author on reasonable request.

Ethics approval and consent to participate

This study was approved by the Clinical Research Ethics Committee of the Sun Yat-sen University Cancer Center, and all patients provided written informed consent at the first visit to our center.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. Brentnall AR, Cuzick J. Use of the concordance index for predictors of censored survival data. Stat Methods Med Res. 2018;27:2359–73.
2. Chang H, Gao J, Xu BQ, et al. Haemoglobin, neutrophil to lymphocyte ratio and platelet count improve prognosis prediction of the TNM staging system in nasopharyngeal carcinoma: development and validation in 3,237 patients from a single institution. Clin Oncol (R Coll Radiol). 2013;25:639–46.
3. Chen JH, Zhai ET, Yuan YJ, et al. Systemic immune-inflammation index for predicting prognosis of colorectal cancer. World J Gastroenterol. 2017;23:6261–72.
4. Diboun I, Wernisch L, Orengo CA, et al. Microarray analysis after RNA amplification can detect pronounced differences in gene expression using limma. BMC Genom. 2006;7:252.
5. Du XJ, Tang LL, Mao YP, et al. Value of the prognostic nutritional index and weight loss in predicting metastasis and long-term mortality in nasopharyngeal carcinoma. J Transl Med. 2015;13:364.
6. Du XJ, Tang LL, Mao YP, et al. The pretreatment albumin to globulin ratio has predictive value for long-term mortality in nasopharyngeal carcinoma. PloS one. 2014;9:e94473.
7. Goeman JJ. L1 penalized estimation in the Cox proportional hazards model. Biom J. 2010;52:70–84.
8. He X, Li JP, Liu XH, et al. Prognostic value of C-reactive protein/albumin ratio in predicting overall survival of Chinese cervical cancer patients overall survival: comparison among various inflammation based factors. J Cancer. 2018;9:1877–84.
9. Hu JY, Yi W, Xia YF, et al. Impact of pretherapy body mass index on prognosis of nasopharyngeal carcinoma. Ai Zheng. 2009;28:1043–8.
10. Hua X, Chen J, Wu Y, et al. Prognostic role of the advanced lung cancer inflammation index in cancer patients: a meta-analysis. World J Surg Oncol. 2019;17:177.
11. Jafri SH, Shi R, Mills G. Advance lung cancer inflammation index (ALI) at diagnosis is a prognostic marker in patients with metastatic non-small cell lung cancer (NSCLC): a retrospective review. BMC Cancer. 2013;13:158.

12. Janvilisri T Omics-based identification of biomarkers for nasopharyngeal carcinoma. Dis Markers. 2015; 2015:762128.
13. Kamarudin AN, Cox T, Kolamunnage-Dona R. Time-dependent ROC curve analysis in medical research: current methods and applications. BMC Med Res Methodol. 2017;17:53.
14. Kasymjanova G, Macdonald N, Agulnik JS, et al. The predictive value of pre-treatment inflammatory markers in advanced non-small-cell lung cancer. Current Oncology. 2010;17:52–8.
15. Lertbutsayanukul C, Kannarunimit D, Prayongrat A, et al. Prognostic Value of Plasma EBV DNA for Nasopharyngeal Cancer Patients during Treatment with Intensity-modulated Radiation Therapy and Concurrent Chemotherapy. Radiol Oncol. 2018;52:195–203.
16. 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.
17. Li JP, Chen SL, Liu XM, et al. A Novel Inflammation-Based Stage (I Stage) Predicts Overall Survival of Patients with Nasopharyngeal Carcinoma. Int J Mol Sci. 2016; 17.
18. Liang W, Shen G, Zhang Y, et al. Development and validation of a nomogram for predicting the survival of patients with non-metastatic nasopharyngeal carcinoma after curative treatment. Chin J Cancer. 2016;35:98.
19. Liao LJ, Hsu WL, Wang CT, et al. Prognostic impact of pre-treatment neutrophil-to-lymphocyte ratio (NLR) in nasopharyngeal carcinoma: A retrospective study of 180 Taiwanese patients. Clin Otolaryngol. 2018;43:463–9.
20. Moons KG, Donders AR, Steyerberg EW, et al. Penalized maximum likelihood estimation to directly adjust diagnostic and prognostic prediction models for overoptimism: a clinical example. J Clin Epidemiol. 2004;57:1262–70.
21. Ng WT, Yuen KT, Au KH, et al. Staging of nasopharyngeal carcinoma–the past, the present and the future. Oral Oncol. 2014;50:549–54.
22. Ou G, Xing S, Li J, et al. Circulating tumor cells: a valuable marker of poor prognosis for advanced nasopharyngeal carcinoma. Mol Med. 2019;25:50.
23. OuYang PY, Su Z, Ma XH, et al. Comparison of TNM staging systems for nasopharyngeal carcinoma, and proposal of a new staging system. Br J Cancer. 2013;109:2987–97.
24. OuYang PY, Xiao Y, You KY, et al. Validation and comparison of the 7th and 8th edition of AJCC staging systems for non-metastatic nasopharyngeal carcinoma, and proposed staging systems from Hong Kong, Guangzhou, and Guangxi. Oral Oncol. 2017;72:65–72.
25. Ozyurek BA, Ozdemirel TS, Ozden SB, et al. Does advanced lung inflammation index (ALI) have prognostic significance in metastatic non-small cell lung cancer? Clin Respir J. 2018;12:2013–9.
26. Park YH, Yi HG, Lee MH, et al. Prognostic Value of the Pretreatment Advanced Lung Cancer Inflammation Index (ALI) in Diffuse Large B Cell Lymphoma Patients Treated with R-CHOP Chemotherapy. Acta Haematol. 2017;137:76–85.

27. Shibutani M, Maeda K, Nagahara H, et al. The prognostic significance of the advanced lung cancer inflammation index in patients with unresectable metastatic colorectal cancer: a retrospective study. BMC Cancer. 2019;19:241.
28. Shim JH, Jun MJ, Han S, et al. Prognostic nomograms for prediction of recurrence and survival after curative liver resection for hepatocellular carcinoma. Ann Surg. 2015;261:939–46.
29. Srivastava S, Chen L Comparison between the stochastic search variable selection and the least absolute shrinkage and selection operator for genome-wide association studies of rheumatoid arthritis. BMC Proc. 2009; 3 Suppl 7:S21.
30. Tang LQ, Li CF, Li J, et al. Establishment and Validation of Prognostic Nomograms for Endemic Nasopharyngeal Carcinoma. J Natl Cancer Inst. 2016; 108.
31. Tibshirani R. The lasso method for variable selection in the Cox model. Stat Med. 1997;16:385–95.
32. Tokarskaya ZB, Scott BR, Zhuntova GV, et al. Interaction of radiation and smoking in lung cancer induction among workers at the Mayak nuclear enterprise. Health Phys. 2002;83:833–46.
33. Torre LA, Bray F, Siegel RL, et al. Global cancer statistics, 2012. CA Cancer J Clin. 2015;65:87–108.
34. Vickers AJ, Cronin AM, Elkin EB, et al. Extensions to decision curve analysis, a novel method for evaluating diagnostic tests, prediction models and molecular markers. BMC Med Inform Decis Mak. 2008;8:53.
35. Wang S, Ge M, Wang K, et al. The clinical features and prognosis of radiotherapy associated sarcoma (RAS) following radiotherapy for nasopharyngeal carcinoma. Zhonghua Er Bi Yan Hou Tou Jing Wai Ke Za Zhi. 2014;49:955–8.
36. Wang WY, Twu CW, Chen HH, et al. Plasma EBV DNA clearance rate as a novel prognostic marker for metastatic/recurrent nasopharyngeal carcinoma. Clin Cancer Res. 2010;16:1016–24.
37. Xia WX, Zhang HB, Shi JL, et al. A prognostic model predicts the risk of distant metastasis and death for patients with nasopharyngeal carcinoma based on pre-treatment serum C-reactive protein and N-classification. Eur J Cancer. 2013;49:2152–60.
38. Yang L, Hong S, Wang Y, et al. Development and External Validation of Nomograms for Predicting Survival in Nasopharyngeal Carcinoma Patients after Definitive Radiotherapy. Sci Rep. 2015;5:15638.
39. Yang S, Fu X, Huang G, et al. The impact of the interval between the induction of chemotherapy and radiotherapy on the survival of patients with nasopharyngeal carcinoma. Cancer Manag Res. 2019;11:2313–20.
40. You R, Liu YP, Lin M, et al. Relationship of circulating tumor cells and Epstein-Barr virus DNA to progression-free survival and overall survival in metastatic nasopharyngeal carcinoma patients. Int J Cancer. 2019;145:2873–83.
41. Yuan J, Xu M, Li J, et al. Prognostic value of cystatin C in patients with nasopharyngeal carcinoma: a retrospective study of 1063 patients. Clinics. 2016;71:338–43.

42. Zeng Q, Hong MH, Shen LJ, et al. Nomograms for predicting long-term survival in patients with non-metastatic nasopharyngeal carcinoma in an endemic area. Oncotarget. 2016;7:29708–19.
43. Zhang B, Tian J, Dong D, et al. Radiomics Features of Multiparametric MRI as Novel Prognostic Factors in Advanced Nasopharyngeal Carcinoma. Clin Cancer Res. 2017;23:4259–69.
44. Zhang H, Xia W, Lu X, et al. A novel statistical prognostic score model that includes serum CXCL5 levels and clinical classification predicts risk of disease progression and survival of nasopharyngeal carcinoma patients. PloS one. 2013;8:e57830.
45. Zhang L, Huang Y, Hong S, et al. Gemcitabine plus cisplatin versus fluorouracil plus cisplatin in recurrent or metastatic nasopharyngeal carcinoma: a multicentre, randomised, open-label, phase 3 trial. Lancet. 2016;388:1883–92.
46. Zhao L, Chen H, Hu B, et al. Prognostic significance of Ki67 expression and the derived neutrophil-lymphocyte ratio in nasopharyngeal carcinoma. Cancer Manag Res. 2018;10:1919–26.

Figures

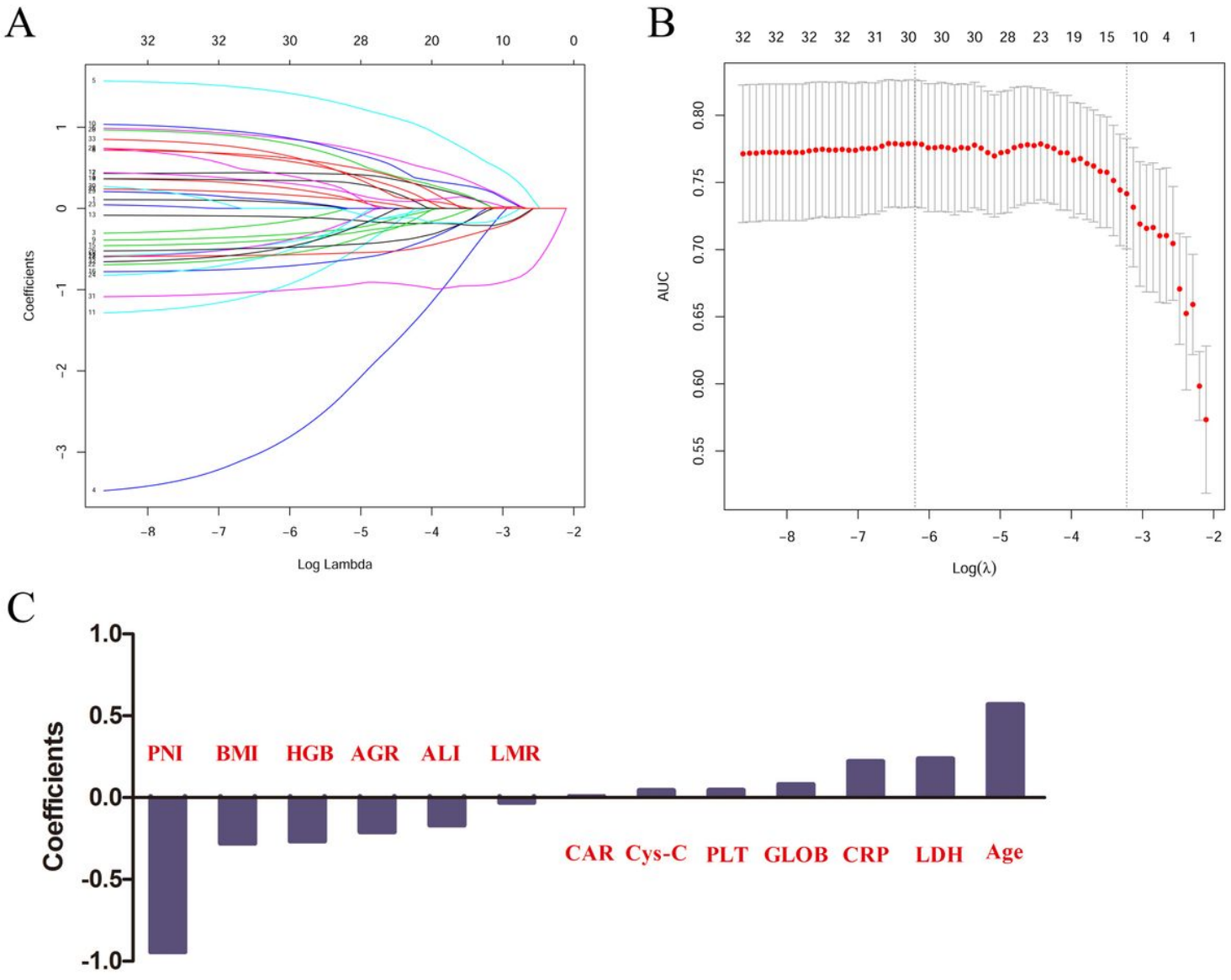

Figure 1

Potential predictors selection using LASSO-Cox regression model.

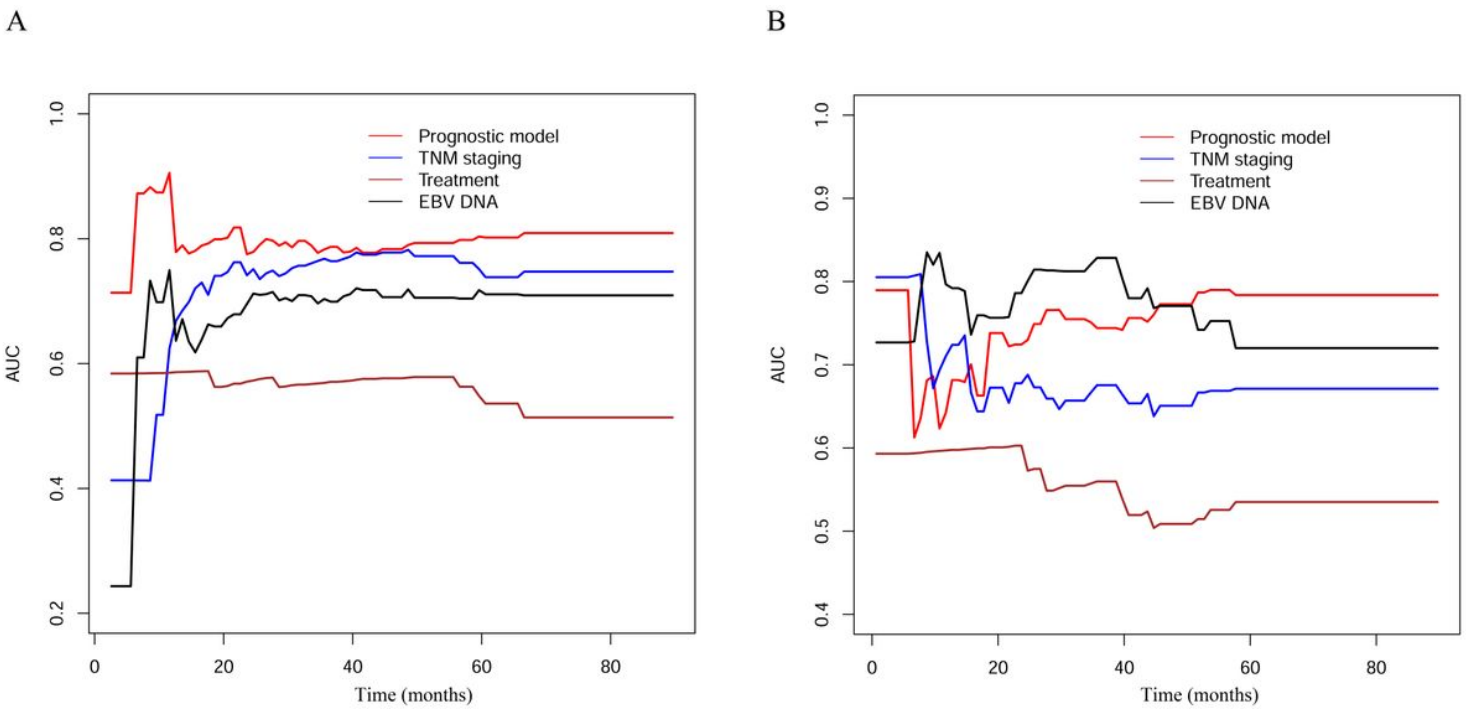

Figure 2

Comparison of predictive accuracy between prognostic model, TNM staging, and clinical treatment using time dependent ROC curves in training cohort and validation cohort.

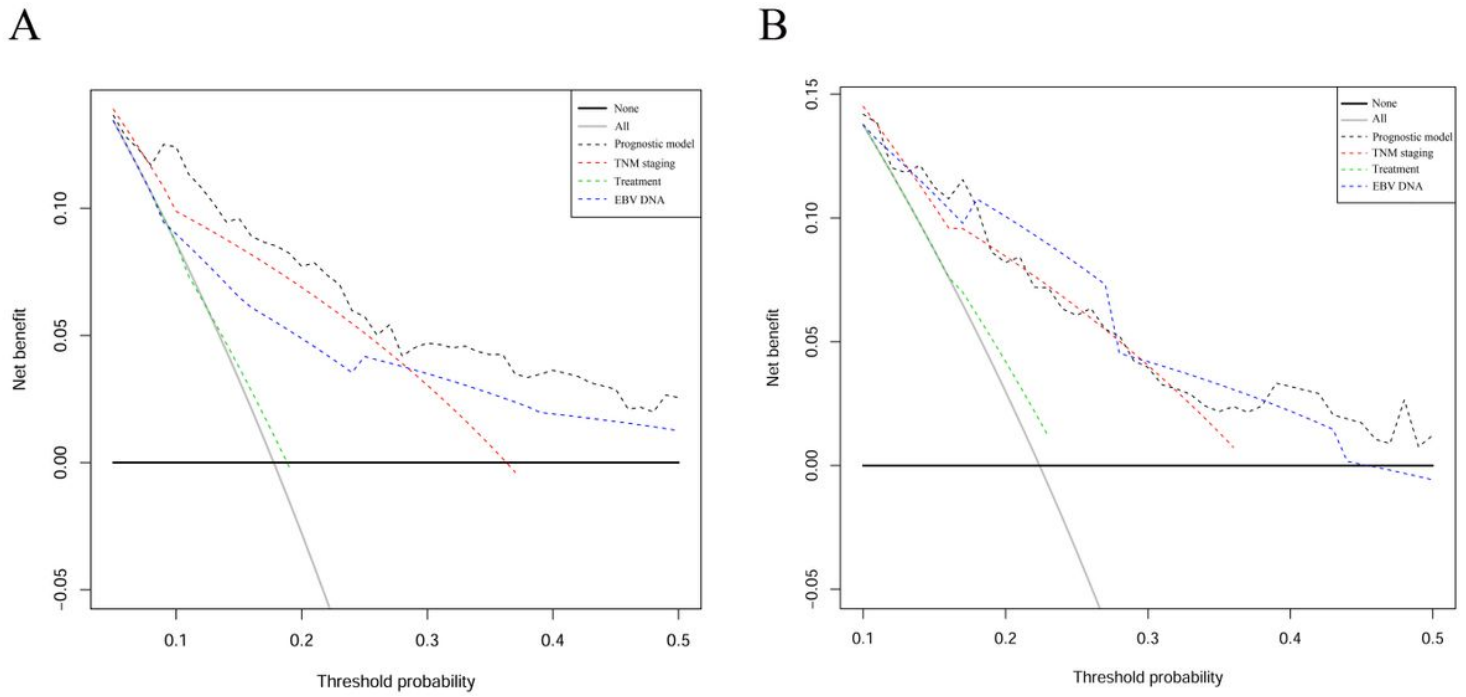

Figure 3

Decision curve analysis for each model in training cohort and validation cohort.

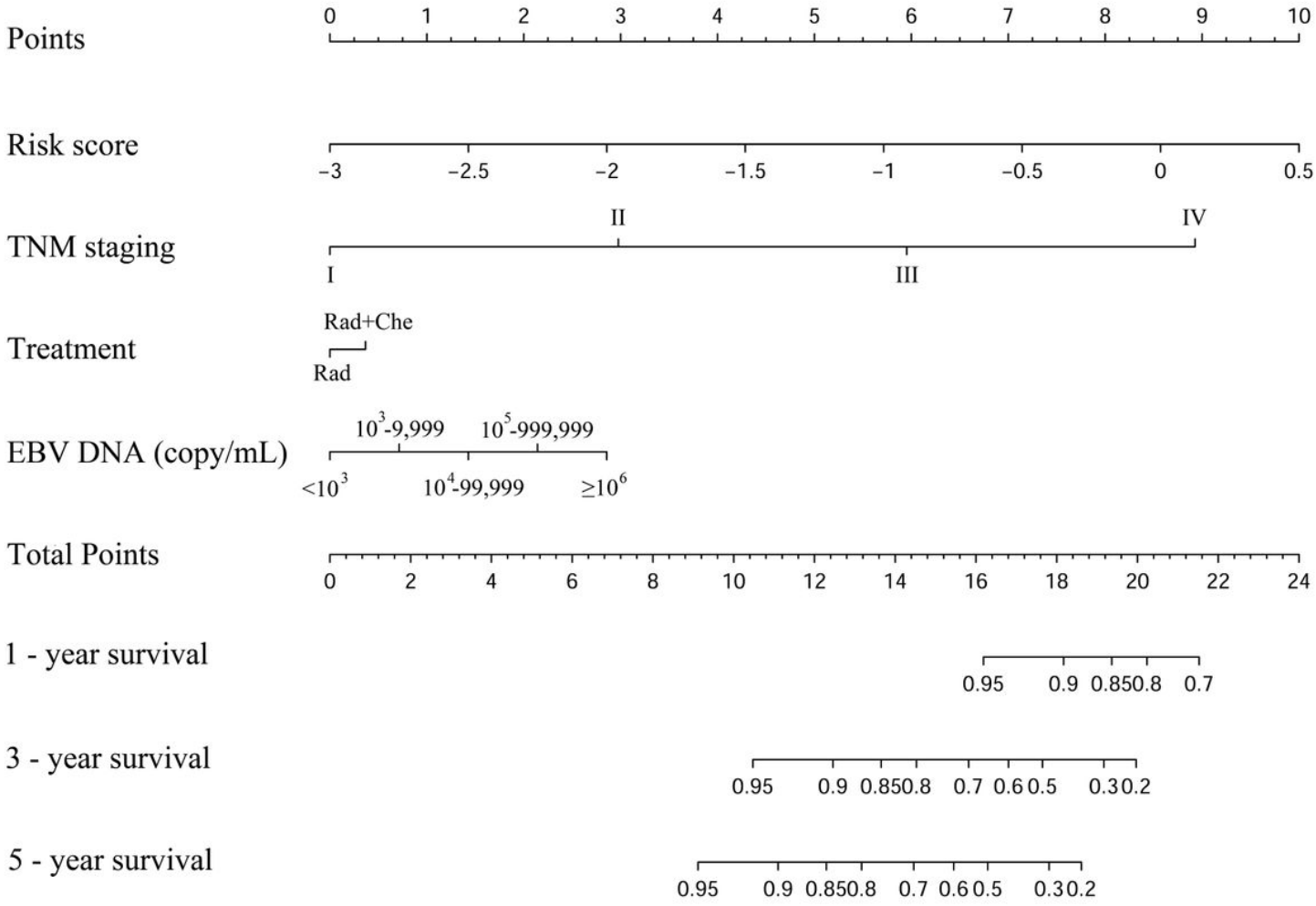

Figure 4

The nomogram was used to estimate OS for NPC patients in training cohort.

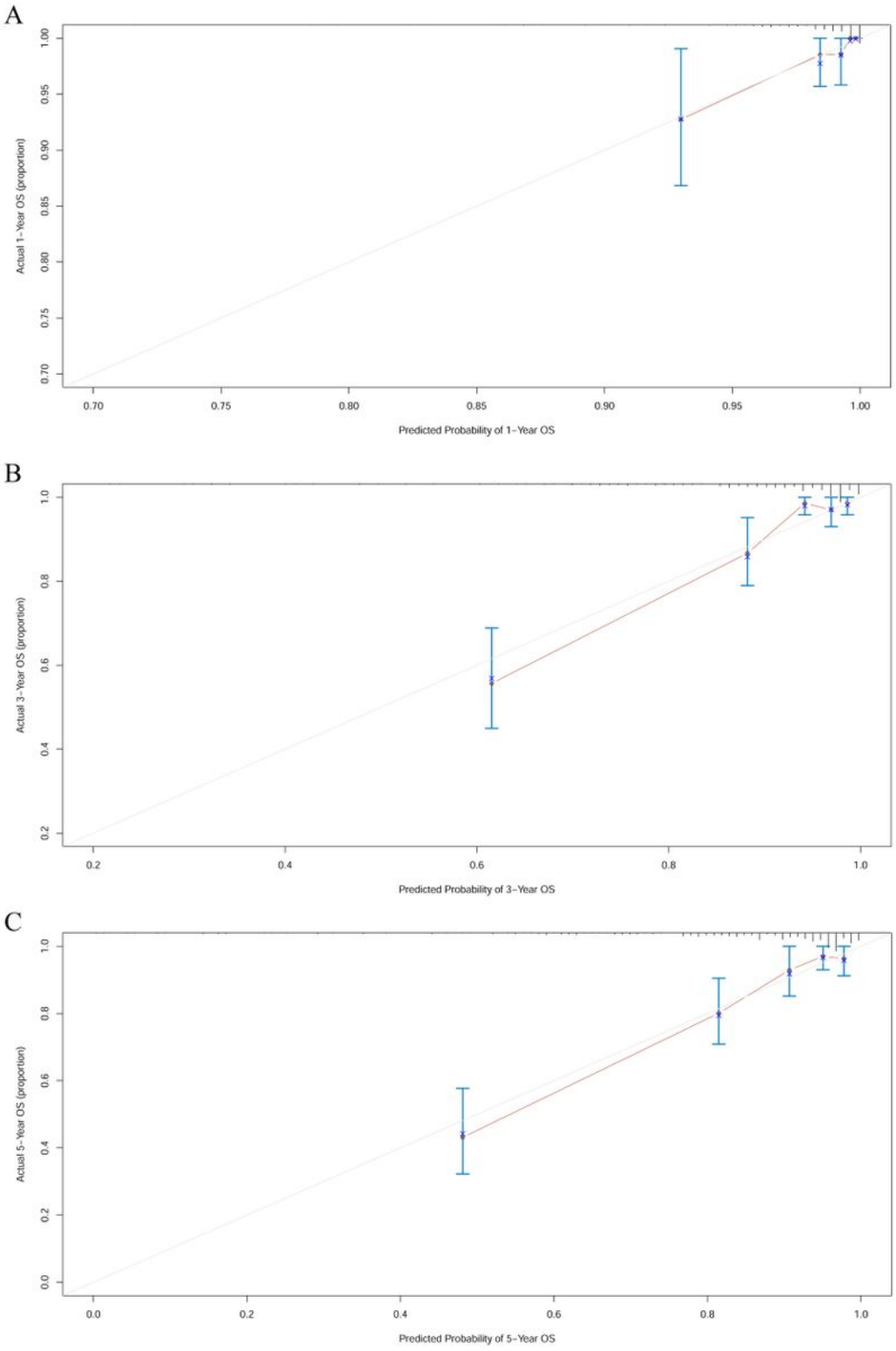

Figure 5

The calibration plot for the nomograms at 1-, 3-, 5- year in training cohort.

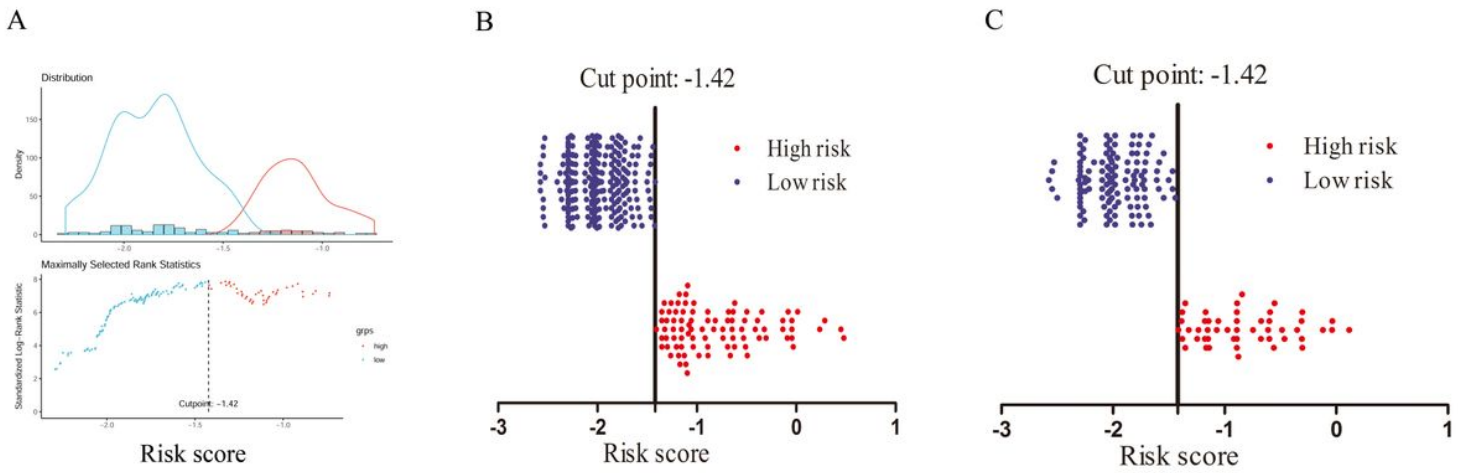

Figure 6

The optimal cut-off value of prognostic model risk score using R package "survival", and the distribution of the prognostic model risk score in training cohort and validation cohort.

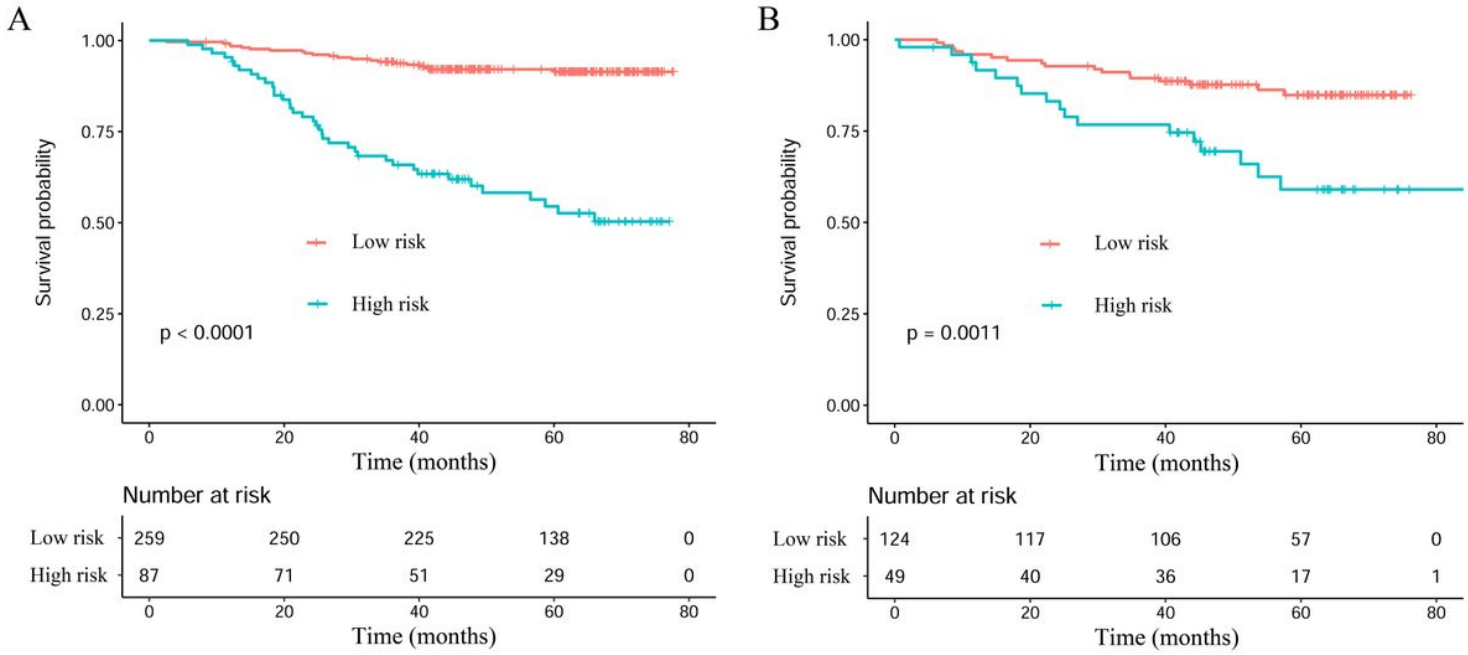

Figure 7

Kaplan–Meier analyses of OS according to the prognostic model risk score classifier in subgroups of NPC patients in training cohort and the validation cohort.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementFigure2.tif

- SupplementFigure1.tif
- SupplementaryFigure.docx