

SCIENTIFIC REPORTS



OPEN

Development and External Validation of Nomograms for Predicting Survival in Nasopharyngeal Carcinoma Patients after Definitive Radiotherapy

Received: 01 June 2015
Accepted: 28 September 2015
Published: 26 October 2015

Lin Yang^{1,2,3,*}, Shaodong Hong^{1,2,3,*}, Yan Wang^{1,2,3,*}, Haiyang Chen⁴, Shaobo Liang⁵, Peijian Peng⁶ & Yong Chen^{1,2,3}

The distant metastasis free survival (DMFS) and overall survival (OS) differ significantly among individuals even within the same clinical stages. The purpose of this retrospective study was to build nomograms incorporating plasma EBV DNA for predicting DMFS and OS of nasopharyngeal carcinoma (NPC) patients after definitive radiotherapy. A total of 1168 non-metastatic NPC patients from two institutions were included to develop the nomograms. Seven and six independent prognostic factors were identified to build the nomograms for OS and DMFS, respectively. The models were externally validated by a separate cohort of 756 NPC patients from the third institutions. For predicting OS, the c-index of the nomogram was significantly better than that of the TNM staging system (Training cohort, $P=0.005$; validation cohort, $P=0.03$). The c-index of nomogram for DMFS in the training and validation set were both higher than that of TNM classification with marginal significance ($P=0.048$ and $P=0.057$, respectively). The probability of 1-, 3-, and 5-year OS and DMFS showed optimal agreement between nomogram prediction and actual observation. The proposed stratification of risk groups based on the nomograms allowed significant distinction between Kaplan-Meier curves for survival outcomes. The prognostic nomograms could better stratify patients into different risk groups.

Nasopharyngeal carcinoma (NPC) is one of the most common malignancies of head and neck in the Southeast Asia with an annual incidence of 15–50 cases per 100,000 persons, which is closely related to Epstein-Barr virus (EBV) infection¹. Radiotherapy or chemoradiotherapy is the main treatment for non-metastatic NPC, achieving a 5-year disease-free survival and overall survival (OS) of 84% and 75%, respectively^{2,3}. However, NPC has higher tendency for metastatic dissemination than other head and neck cancers⁴. Many patients eventually develop distant metastases after definitive radiotherapy and their

¹Sun Yat-sen University cancer center, 651 Dongfeng Road east, Guangzhou, China. ²State Key Laboratory of Oncology in Southern China, Guangzhou, China. ³Collaborative Innovation Center for Cancer Medicine, Guangzhou, China. ⁴The Six Affiliated Hospital of Sun Yat-sen University, Guangzhou, China. ⁵The First Hospital of Foshan, Foshan, China. ⁶The Fifth Affiliated Hospital of Sun Yat-sen University, Zhuhai, China. *These authors contributed equally to this work. Correspondence and requests for materials should be addressed to Y.C. (email: chenyon@sysucc.org.cn)

OS is very poor⁵. More intensive follow-up and treatment strategies might be needed for high risk patients. However, accurate prediction for OS and metastasis-free survival (DMFS) remains unavailable.

The seventh edition of the American Joint Committee on Cancer (AJCC) TNM staging system is the most widely used prognostic tool, in which non-metastatic NPC patients were stratified according to tumor size and invasion, and the extent of lymph node involvement. However, prediction of survival is far more complicated than TNM staging. For patients with equivalent TNM classification, there remains apparently heterogeneity of DMFS and OS⁶. Other independent prognostic factors could also significantly contribute to the prediction of clinical outcomes. For example, plasma EBV DNA copy number is closely related to tumor burden and could serve as a useful prognostic factor in NPC patients with different clinical stages⁷. Attempts are also made to reveal the prognostic significance of some laboratory index such as serum C-reactive protein (CRP), lactate dehydrogenase (LDH) and hemoglobin^{8,9}. However, until now no acknowledged and validated prognostic models are available.

The optimal follow-up strategies after definitive radiotherapy for NPC patients are still undefined. Identifying subgroups of patients at different risks for distant metastases could help determine the appropriate timing and imaging techniques in a more individualized manner. Also, more accurate prediction of OS could be of significance for both patients and clinicians in decision making.

Nomograms have been accepted as reliable and pragmatic prediction tools to quantify individual risk by incorporating a variety of important factors for oncological prognoses. In many types of cancers, nomograms have been proved to provide more precise prediction compared with traditional TNM classification¹⁰. However, nomograms for predicting DMFS and OS after definitive radiotherapy for non-metastatic NPC patients are rare. In this study, we hypothesized that nomograms combining T stage, N stage and objective laboratory index could generate more accurate prediction models for curative NPC patients.

Patients and Methods

Training cohort. The training cohort for nomogram development was derived from Sun Yat-sen University Cancer Center and the First Hospital of Foshan between October 2007 and December 2009. The inclusion criteria for the study are as follows: (i) pathological evidence of NPC; (ii) complete baseline clinical information and laboratory data; (iii) patients had received radical radiotherapy and (iv) complete follow-up data. Patients with distant metastasis at presentation were excluded. Ethical approvals were obtained from both institutions through their respective institutional review boards. Informed consent was granted a waiver due to the retrospective nature of the study. The study protocol was designed in accordance with the guidelines outlined in the Declaration of Helsinki and was approved by the Ethics Committee of Sun Yat-sen University Cancer Center and the First Hospital of Foshan, respectively.

A standardized data collection form was designed to retrieve all the relevant information on socio-demographic data (age, gender, smoking history, alcohol exposure, family history of malignant tumors and household registry), baseline laboratory data (plasma EBV DNA copy number, titers of IgA antibodies against EBV capsid antigen (EBV VCA-IgA, EA-IgA), serum calcium, serum magnesium, serum phosphorus, albumin (ALB), globulin (GLB), alanine transaminase (ALT), aspartate transaminase (AST), LDH, alkaline phosphatase (ALP), C-reactive protein (CRP), *et al.*), staging data (T stage based on the location, size and extension of the primary tumor; N stage based on the number and location of lymph node metastasis), therapeutic data (radiotherapeutic technique, radiation fractions and dosage, utility of chemotherapy). Clinical stage was classified according to the seventh edition of the AJCC/UICC TNM Staging System.

Validation cohort. To examine the generalizability of the model, an external validation cohort of 756 consecutive NPC patients who also received definitive radiotherapy were included from the Fifth Affiliated Hospital of Sun Yat-sen University between January 2007 and December 2010. Only patients with non-metastatic disease were included, and all the patients should have sufficient data to score all the variables in the established nomograms. The study protocol was designed in accordance with the guidelines outlined in the Declaration of Helsinki and was approved by the Ethics Committee of the Fifth Affiliated Hospital of Sun Yat-sen University.

Follow up. Distant metastasis was evaluated by physical examination, nasopharyngoscope, nasopharyngeal and neck magnetic resonance imaging (MRI), chest x-ray and/or CT, abdominal ultrasonography and bone scan every 6 months during the first three years after the completion of radiotherapy and annually thereafter. Survival follow-up was done via direct telecommunication or by referring to the clinic attendance records.

Statistical analysis. DMFS was defined as the time from definitive radiotherapy to the time of metastases or censored at the date of last follow-up. OS was defined as the time from diagnosis to the date of death from any cause or censored at the date of last follow-up. In the training dataset, survival curves for different variable values were plotted by Kaplan-Meier estimates and compared by using log-rank test. Variables achieving significant level of $P < 0.05$ were entered into multivariate analyses via the Cox proportional hazards model with forward stepwise procedures. Independent prognostic factors were determined if they had significant effect in the Cox model ($P < 0.05$). Statistical analyses for survival data were

performed by using SPSS 19.0 (SPSS, Chicago, IL). Nomograms were formulated to provide visualized risk prediction based on the results of multivariate analyses by R 2.14.1 (<http://www.r-project.org>) with the survival and rms packages. The nomogram was subjected to 1000 bootstrap resamples for interval validation and external validation to correct the concordance index (c-index) and explain variance for over-optimism. The performance of the nomograms and TNM staging system for prediction survival were measured by c-index, an equivalent variable of the area under curve (AUC) of the receiver operating characteristic curve for censored data. The maximum value of c-index is 1.0 indicating a perfect prediction model while 0.5 indicates a random chance to correctly predict outcome by the model. Comparisons between nomogram models and TNM staging were performed with the rcorr.cens in Hmisc in R. Calibration of the nomogram for 1-, 3-, and 5-year OS or DMFS were performed by comparing the predicted survival with the observed survival. During external validation of the nomograms, the total points for each patient were calculated according to the established nomograms and then Cox regression was performed using the total points as predictor in the validation cohort. In addition to numerically comparing the discrimination ability by c-index, we also attempted to demonstrate the independent discrimination ability of the nomograms beyond standard TNM classification. By grouping patients evenly into 3 risk groups in the training cohort according to the scores calculated by the nomograms, we determined the cut-off points of risk stratification and investigate its prediction role in different TNM stages with respective Kaplan-Meier survival curves. A two-sided *P* value of <0.05 was deemed significant. Details of R code for running nomograms could be assessed in supplementary information online.

Results

Patient characteristics and survival. A total of 1168 and 756 patients from the training and the external validation cohorts were included for analyses. Median follow-up for OS and DMFS in the training cohort were 70.0 months and 68.8 months, whereas the median follow-up for OS and DMFS in the validation dataset were 61.8 months and 60.25 months. Five-year events rates for OS and DMFS in the training cohort were 84.0% and 85.6%, and were 84.6% for OS and 83.5% for DMFS in the validation cohort. Details regarding patient characteristics are shown in Table 1.

Univariate analysis and multivariate analysis. For OS, the significant inferior prognostic factors included older age, male sex, smoking, elevated LDH, CRP and plasma EBV DNA, decreased albumin, elevated titers of EA-IgA, higher clinical T stage and N stage.

All significant variables were entered into Cox regression model. The results show that the following variables remained independently prognostic: age, gender, LDH, CRP, plasma EBV DNA, T stage and N stage. Likewise, for DMFS, independent prognostic factors were gender, LDH, CRP, plasma EBV DNA, T stage and N stage.

The detailed results of multivariate analyses are shown in Table 2.

Prognostic nomogram. The resulting coefficients from the Cox models were used to construct the nomograms for OS and DMFS (Fig. 1). Each subtype within the variables was assigned a score. By adding up the total score from all the variables and locating it to the total point scale, we could determine the probabilities of the outcomes by drawing a vertical line to the total score. N stage was the most important contributing factor both for OS and DMFS prediction. In the training cohort, the nomogram for OS had a bootstrap-corrected c-index of 0.76 (95% CI, 0.73–0.79) which was significantly better than that of TNM classification (0.65; 95% CI, 0.62–0.69; *P* = 0.005). The c-index of nomogram for DMFS (0.71; 95% CI, 0.68–0.75) was also significantly higher than that of TNM classification (0.64; 95% CI, 0.60–0.68; *P* = 0.048). In the external validation cohort, the c-index was 0.74 (95% CI, 0.69–0.78) for OS and 0.69 (95% CI, 0.65–0.74) for DMFS in nomograms, both of which were better than the c-index in TNM classification for OS (0.66; 95% CI, 0.61–0.70; *P* = 0.03) and DMFS (0.62; 95% CI, 0.57–0.66; *P* = 0.057), respectively. The results were shown in Table 3.

The calibration plots presented fair agreements between the nomogram prediction and actual observation for the 1-, 3-, and 5-year OS and excellent agreements for the 1-, 3-, and 5-year DMFS in both cohorts (Fig. 2).

The nomogram in stratifying risk of patients. We determined the cutoff values of nomogram-generated scores with which patients in the training cohort were evenly stratified into three risk groups. Each group represented a distinct prognosis (Table 4). This stratification could effectively discriminate the survival outcomes for the three proposed risk groups both in the training and validation cohorts (Fig. 3). Even within different TNM stages, the stratification could allow significant distinction among Kaplan-Meier curves for OS (see Supplementary Figure S1 online) and DMFS (see Supplementary Figure S2 online).

Discussion

The prediction of survival by TNM staging in NPC patients remains imperfect due to its simplicity and the heterogeneity of risk within the same stage. Developing prediction models with more precision that incorporate a variety of independent objective variables are urgently needed. To the best of our knowledge, this is one of the few studies to develop visualized, user-friendly and reliable prediction models for

| Characteristics | Training set | | | Testing set |
|-------------------------|---------------------|-------------|--------|---------------------|
| | Number of cases (%) | Univariable | | Number of cases (%) |
| | | OS | DMFS | |
| Age(years) | | <0.001 | 0.189 | |
| <45 | 602(51.5%) | | | 404(53.4%) |
| ≥45 | 566(48.5%) | | | 352(46.6%) |
| Gender | | <0.001 | 0.037 | |
| Male | 853(73%) | | | 555(73.4%) |
| Female | 315(27%) | | | 201(26.6%) |
| Smoking status | | <0.001 | 0.008 | |
| Absent | 705(60.4%) | | | 498(65.9%) |
| Present | 463(39.6%) | | | 258(34.1%) |
| Drinking status | | 0.211 | 0.183 | |
| Absent | 971(83.1%) | | | 526(69.6%) |
| Present | 197(16.9) | | | 230(30.4%) |
| Family history | | 0.254 | 0.273 | |
| Absent | 842(72.1%) | | | 576(76.2%) |
| present | 326(27.9%) | | | 180(23.8%) |
| T classification | | <0.001 | 0.001 | |
| 1+2 | 376(32.2%) | | | 240(31.7%) |
| 2 | 547(46.8%) | | | 358(47.4%) |
| 3 | 245(21.0%) | | | 158(20.9%) |
| N classification | | <0.001 | <0.001 | |
| 0 | 246(21.1%) | | | 162(21.4%) |
| 1 | 425(36.4%) | | | 305(40.3%) |
| 2 | 310(26.5%) | | | 201(26.6%) |
| 3 | 187(16.0%) | | | 88(11.6%) |
| Clinical Stage | | | | |
| I | 29(2.5%) | | | 28(3.7%) |
| II | 199(17%) | | | 120(15.9%) |
| III | 620(53.1%) | | | 413(54.6%) |
| IV | 320(27.4%) | | | 195(25.8%) |
| Treatment method | | | | |
| RT | 221(18.9%) | | | 182(24.1%) |
| Chemo-RT | 947(81.1%) | | | 574(75.9%) |
| Radiotherapy technology | | | | |
| IMRT+3DCRT | 496(42.5%) | | | 498(65.9%) |
| CRT | 672(57.5%) | | | 258(34.1%) |
| Radiation fractions | | | | |
| ≤34 | 602(51.5%) | | | 356(47.1%) |
| >34 | 566(48.5%) | | | 400(52.9%) |
| Radiation dosage(Gy) | | | | |
| ≤69 | 561(48.0%) | | | 392(51.9%) |
| >69 | 607(52.0%) | | | 364(48.1%) |
| Calcium | | 0.582 | 0.473 | |
| <2.46 | 554(47.4%) | | | 380(50.3%) |
| ≥2.46 | 614(52.6%) | | | 376(49.7%) |
| Magnesium | | 0.924 | 0.517 | |
| <0.9 | 633(54.2%) | | | 398(52.6%) |
| Continued | | | | |

| Characteristics | Training set | | | Testing set |
|------------------------------|---------------------|-------------|--------|---------------------|
| | Number of cases (%) | Univariable | | Number of cases (%) |
| | | OS | DMFS | |
| >=0.9 | 535(45.8%) | | | 358(47.4%) |
| Phosphorus | | 0.572 | 0.989 | |
| <1.15 | 563(48.2%) | | | 374(49.5%) |
| >=1.15 | 605(51.8%) | | | 382(50.5%) |
| WBC, ×10 ⁹ | | 0.188 | 0.302 | |
| <6.9 | 608(52.1%) | | | 413(54.6%) |
| >=6.9 | 560(47.9%) | | | 343(45.4%) |
| Neutrophil, ×10 ⁹ | | 0.787 | 0.804 | |
| <4.1 | 601(51.5%) | | | 400(52.9%) |
| >=4.1 | 567(48.5%) | | | 356(47.1%) |
| Neutrophil/WBC | | 0.714 | 0.724 | |
| <0.61 | 611(52.3%) | | | 413(54.6%) |
| >=0.61 | 557(47.7%) | | | 343(45.4%) |
| HGB, g/L | | 0.339 | 0.686 | |
| <143 | 596(51%) | | | 369(48.8%) |
| >=143 | 572(49%) | | | 387(51.2%) |
| GLB, g/L | | 0.661 | 0.524 | |
| <29 | 594(50.9%) | | | 380(50.3%) |
| >=29 | 574(49.1%) | | | 376(49.7%) |
| ALB, g/L | | 0.001 | 0.371 | |
| <45.6 | 585(50.6%) | | | 350(46.3%) |
| >=45.6 | 583(49.9%) | | | 406(53.7%) |
| ALT, U/L | | 0.635 | 0.135 | |
| <20.6 | 585(50.1%) | | | 358(47.4%) |
| >=20.6 | 583(49.9%) | | | 398(52.6%) |
| AST, U/L | | 0.054 | 0.132 | |
| <20.8 | 588(50.3%) | | | 377(49.9%) |
| >=20.8 | 580(49.7%) | | | 379(50.1%) |
| ALP, U/L | | 0.062 | 0.382 | |
| <66.7 | 591(50.6%) | | | 373(49.3%) |
| >=66.7 | 577(49.4%) | | | 383(50.7%) |
| LDH, U/L | | <0.001 | <0.001 | |
| <166 | 583(49.9%) | | | 379(50.1%) |
| >=166 | 585(50.1%) | | | 377(49.9%) |
| CRP, mg/L | | <0.001 | <0.001 | |
| <1.49 | 583(49.9%) | | | 381(50.4%) |
| >=1.49 | 585(50.1%) | | | 375(49.6%) |
| EBV-DNA, copies/ml | | <0.001 | <0.001 | |
| <3,760 | 614(52.6%) | | | 453(59.9%) |
| >=3760 | 554(47.4%) | | | 303(40.1%) |
| VCA-IgA | | 0.657 | 0.368 | |
| <1:320 | 591(56.5%) | | | 379(50.1%) |
| >=1:320 | 455(43.5%) | | | 377(49.9%) |
| EA-IgA | | <0.001 | 0.300 | |
| <=1:20 | 664(56.8%) | | | 382(50.5%) |
| >1:20 | 504(25.8%) | | | 374(49.5%) |
| Continued | | | | |

| Characteristics | Training set | | Testing set | |
|--------------------|---------------------|-------------|-------------|---------------------|
| | Number of cases (%) | Univariable | | Number of cases (%) |
| | | OS | DMFS | |
| Distant metastasis | | | | |
| Absent | 981(84%) | | 627(82.9%) | |
| Present | 187(16.0%) | | 129(17.1%) | |
| Living Status | | | | |
| Live | 952(81.5%) | | 632(83.6%) | |
| Dead | 216(18.5) | | 124(16.4%) | |

Table 1. Clinical and laboratorial characteristics of patient. Abbreviations: RT, radiotherapy; Chemo-RT, chemoradiotherapy; CRT, conventional radiotherapy; IMRT, intensity modulated radiation therapy; 3D-CRT, three dimensional conformal radiation therapy; WBC, White cell; HGB, hemoglobin; GLB, Globulin; ALB, Albumin; ALT, Alanine transaminase; AST, Asanine transaminase; ALP, Alkaline phosphatase; LDH, Lactate dehydrogenase; CRP, C-rective protein; EBV-DNA, Epstein-barr virus DNA; OS, overall survival; DMFS, disease metastasis free survival.

| Characteristics | OS | | | DMFS | | |
|--------------------|--------|-------------|--------------|--------|-------------|--------------|
| | P | 95% CI | Hazard Ratio | P | 95% CI | Hazard Ratio |
| Age(year) | <0.001 | 1.679–2.988 | 2.240 | | | |
| <45 | | | | | | |
| >=45 | | | | | | |
| gender | 0.001 | 0.392–0.801 | 0.560 | 0.030 | 0.476–0.964 | 0.677 |
| Male | | | | | | |
| Female` | | | | | | |
| LDH, U/L | 0.003 | 1.150–2.207 | 1.527 | 0.018 | 1.065–1.930 | 1.433 |
| <166 | | | | | | |
| >=166 | | | | | | |
| CRP, mg/L | <0.001 | 1.484–2.706 | 2.004 | 0.008 | 1.113–2.039 | 1.506 |
| <1.49 | | | | | | |
| >=1.49 | | | | | | |
| EBV-DNA, copies/ml | 0.017 | 1.063–1.878 | 1.413 | 0.018 | 1.065–1.958 | 1.444 |
| <3,760 | | | | | | |
| >=3760 | | | | | | |
| T classification | <0.001 | | | 0.028 | | |
| 1+2 | | | | | | |
| 3 | 0.470 | 0.799–1.627 | 1.140 | 0.401 | 0.812–1.682 | 1.169 |
| 4 | <0.001 | 1.491–3.124 | 2.159 | 0.011 | 1.124–2.509 | 1.679 |
| N classification | <0.001 | | | <0.001 | | |
| 0 | | | | | | |
| 1 | 0.110 | 0.913–2.440 | 1.492 | 0.085 | 0.937–2.750 | 1.650 |
| 2 | <0.001 | 1.546–4.093 | 2.516 | 0.001 | 1.424–4.177 | 2.439 |
| 3 | <0.001 | 3.122–8.319 | 5.096 | <0.001 | 3.152–9.168 | 5.376 |

Table 2. Selected Factors for building the model. Abbreviations: LDH, Lactate dehydrogenase; CRP, C-rective protein; EBV-DNA, Epstein-barr virus DNA; OS, overall survival; DMFS, disease-metastasis free survival.

OS and DMFS in non-metastatic NPC patients based on large database. The nomograms we established showed superior discrimination ability compared with traditional TNM staging and allowed risk scoring for individual patient.

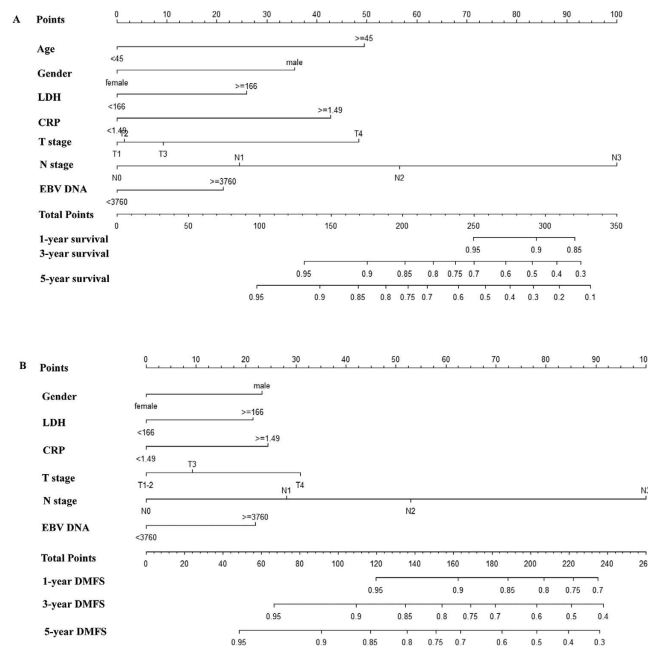


Figure 1. Nomograms of non-metastatic NPC patients after definitive radiotherapy for OS (A) and DMFS (B). LDH, Lactate dehydrogenase; CRP, C-reactive protein; EBV, Epstein-Barr virus; OS, overall survival; DMFS, disease-metastasis free survival.

| Model for Survival prediction | Training | | | Validation | | |
|-------------------------------|----------|-----------|-------|------------|-----------|-------|
| | C-index | 95% CI | P | C-index | 95% CI | P |
| Nomogram(OS) | 0.76 | 0.73–0.79 | 0.005 | 0.74 | 0.69–0.78 | 0.03 |
| TNM classification(OS) | 0.65 | 0.62–0.69 | | 0.66 | 0.61–0.70 | |
| Nomogram(DMFS) | 0.71 | 0.68–0.75 | 0.048 | 0.69 | 0.65–0.74 | 0.057 |
| T TNM classification(DMFS) | 0.64 | 0.60–0.68 | | 0.62 | 0.57–0.66 | |

Table 3. The c-index of OS and DMFS for multivariate model performance and TNM in the training set and the validation set. Abbreviations: OS, overall survival; DMFS, disease-metastasis free survival.

Previously, Cho *et al.* has developed two nomograms to predict the probability of complete response to radiotherapy and OS in non-metastatic NPC patients in Korea, respectively¹¹. However, the relatively small sample size and limited number of prognostic factors in this study call for larger studies for more accurate risk prediction. Also, establishing a prediction model for risk of distant metastasis is of clinical significance, which could help clinicians develop more individualized multidisciplinary treatment and follow-up strategies for NPC patients.

In the present study, the training cohort was obtained from two institutions in Southern China, both of which is located in the areas with high prevalence of NPC and receives a vast amount of NPC patients each year, guaranteeing the representativeness of NPC in endemic areas. Multivariable analyses show that variables include age, gender, LDH, CRP, plasma EBV DNA, T stage, and N stage could independently predict OS. For DMFS, the independent prognostic factors were the same with that for OS except that age did not remain significant. Based on the results of Cox regression model, we established the respective nomograms for predicting OS and DMFS with excellent discrimination ability (c-index for OS, 0.76; c-index for DMFS, 0.71). The established nomograms were found to override traditional TNM staging system in predicting OS and DMFS. More importantly, calibration curves show optimal agreements between prediction and actual observation of the studied outcomes, which guarantees the reliability of the established nomograms.

The nomogram models were further validated in an external cohort ($N=756$) from a third institution to avoid over-fitting of the models and determine their applicability. The c-index of the nomogram predictions for OS and DMFS in the validation cohort were 0.74 and 0.69, respectively, both of which were higher than that of TNM staging predictions. However, the superiority of the discrimination ability of nomogram over TNM classification for predicting DMFS in the validation cohort only had marginal significance ($P=0.057$). We think this phenomenon could be attributed to the relatively small sample

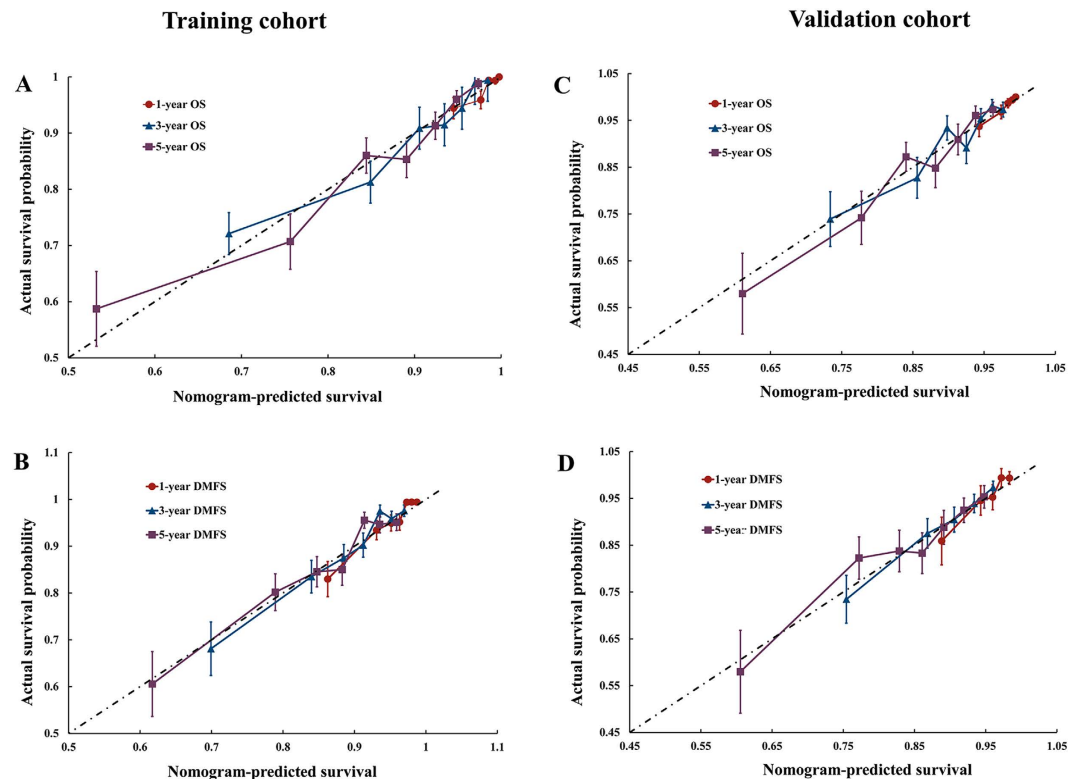


Figure 2. Calibration plots of OS at 1, 3, 5 years (A,C) and DMFS at 1, 3, 5 year (B,D) in training cohort (left) and validation cohort (right). Nomogram-predicted OS and DMFS are plotted on the x-axis; actual OS and DMFS are plotted on the y-axis. Dashed lines along the 45-degree line through the origin point represent the perfect calibration models in which the predicted probabilities are identical to the actual probabilities. OS, overall survival; DMFS, disease-metastasis free survival.

size in the validation cohort or the existence of other potential predictors for distant metastasis, which deserves further investigation in the future.

Based on the nomograms, we further stratify patients into three distinct risk groups for OS and DMFS. Surprisingly, even within the same TNM category, the proposed risk groups could still significantly discriminate the survival outcomes except for the discrimination of OS in the early stage NPC in the validation cohort ($P = 0.636$). One feasible contributor to this insignificance is the small sample size of patients with early stage in this cohort. All in all, these results imply that the nomograms could provide satisfactory discrimination ability beyond TNM staging systems and might be good substitutes for traditional TNM classification.

Age at diagnosis was a strong prognostic factor for OS. Increasing age adversely affected OS in NPC patients after definitive radiotherapy¹². The increased risk of comorbidities, less tolerance to intensive therapies and the declined immune function with increasing age might account for the inferior OS in the elder patients¹³. This results call for the involvement of multidisciplinary approaches apart from oncological aspects to further improve the outcome of elderly patients. Also, clinical trials with special attention to older patients are warranted.

We also found that females had better prognoses than male patients. This phenomenon might be explained by the fact that testosterone could negatively affect the immune function in men, whereas female hormones might have an immune enhancing role¹⁴. However, the mechanisms underlying gender differences in the prognosis of NPC could not be fully demonstrated unless solid biological studies are available.

The link between inflammation and cancer is well-established¹⁵. CRP is a non-specific, acute phase marker of inflammation which has been proved to be associated with inferior survival of numerous malignancies^{16,17}. In this study we also found that CRP has moderate contribution to the nomogram prediction of OS and DMFS. In the future, manipulating the inflammatory status and the immune function of NPC patients might be a promising strategy to further improve their clinical outcomes.

In the present study, we further confirmed that elevated serum LDH is inversely related to patients' survival. Consistently, another retrospective study showed that high pretreatment LDH was correlated with poorer 4-year OS and DMFS in NPC⁹. The link between LDH and survival might rely on cancer hypoxia¹⁸. Hypoxia is a characteristic property of NPC due to rapid cell proliferation, high metabolic demands and impaired angiogenesis. In response to hypoxic stress, anaerobic glycolysis becomes the main

| Characteristics | OS | | DMFS | |
|------------------------|---------|---------------------|--------|-----------------------|
| | Score | Estimated 5-year OS | Score | Estimated 5-year DMFS |
| Age(year) | | | | |
| <45 | 0 | | | |
| >=45 | 50 | | | |
| Gender | | | | |
| Male | 36 | | 23 | |
| Female` | 0 | | 0 | |
| LDH, U/L | | | | |
| <166 | 0 | | 0 | |
| >=166 | 26 | | 21 | |
| CRP, mg/L | | | | |
| <1.49 | 0 | | 0 | |
| >=1.49 | 43 | | 24 | |
| EBV-DNA, copies/ml | | | | |
| <3,760 | 0 | | 0 | |
| >=3760 | 21 | | 22 | |
| T classification | | | | |
| 1+2 | 0 | | 0 | |
| 3 | 8 | | 9 | |
| 4 | 47 | | 31 | |
| N classification | | | | |
| 0 | 0 | | 0 | |
| 1 | 25 | | 28 | |
| 2 | 57 | | 53 | |
| 3 | 100 | | 100 | |
| Total prognostic score | | | | |
| Training cohort | | | | |
| Low risk | <119 | 96.2 | <80 | 95.3 |
| Middle risk | 119–177 | 87.7 | 80–122 | 87.4 |
| High risk | >177 | 66.8 | >122 | 73.0 |
| Validation cohort | | | | |
| Low risk | <119 | 94.3 | <80 | 92.4 |
| Middle risk | 119–17 | 84.4 | 80–12 | 85.0 |
| High risk | >177 | 69.5 | >122 | 69.8 |

Table 4. Point Assignment from nomograms and Prognostic Score. Abbreviations: LDH, Lactate dehydrogenase; CRP, C-rective protein; EBV-DNA, Epstein-barr virus DNA; OS, overall survival; DMFS, disease-metastasis free survival.

energy source for cancer cells, resulting in up-regulation of LDH and activation of the hypoxia-inducible factor (HIF) pathway and ultimately impairment of immune response and survival¹⁹.

EBV infection is associated with increased risk for NPC in endemic areas²⁰. Several studies have substantiated that circulating EBV DNA originates from the EBV-infected primary tumor cells which reflect the overall tumor load and tumor metabolic activity²¹. Plasma EBV DNA level (either prior to treatment or during follow-up) has been proved to be a significant marker in the prognostication of both non-disseminated and disseminated NPC²². However, most studies simply investigated the prognostic value of plasma EBV DNA as a single variable and lack solid validation, whereas the present study for the first time incorporated plasma EBV DNA into prognostic models and achieved more accurate discrimination of survival in NPC patients, further confirming the prognostic value of plasma EBV DNA in NPC patients after curative radiotherapy.

Of course, our nomograms have some limitations. First, the nomograms only include basic clinical and laboratory data. However, the present study aimed to build reliable prediction models. Objective variables are therefore the most ideal factors to be included, while subjective variables might negatively

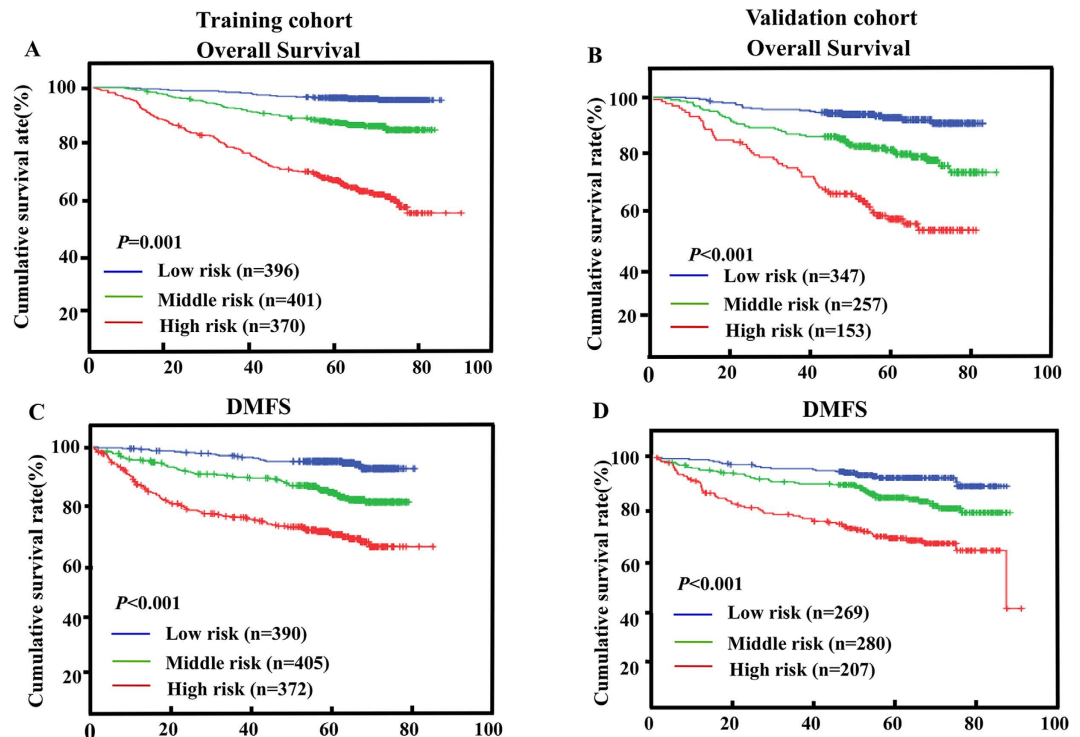


Figure 3. Kaplan–Meier curves of risk group stratification for OS (A,B) and DMFS (C,D). Nomogram risk group stratifications for the 33 and 66 percentiles are shown for the training cohort (left) and for the validation cohort (right). OS, overall survival; DMFS, disease-metastasis free survival.

impact the models due to inevitable bias. Second, the study was conducted retrospectively and selection bias might exist. However, we have included a relatively large training cohort to build the nomograms and externally validated them. The results consistently show the satisfactory performance of the established models. Of course, additional validation of these nomograms by prospective datasets could be useful. Overall, the established nomograms predicting OS and DMFS in NPC patients after definitive radiotherapy provide practical tools for individualized prognostication. Practitioners as well as patients could readily assess these nomograms and immediately apply them to predicting the risks of distant metastasis and death. By doing this, tailored post-treatment follow-up and/or adjuvant therapy could be feasible and improved survival might be achieved.

References

1. Cao, S. M., Simons, M. J. & Qian, C. N. The prevalence and prevention of nasopharyngeal carcinoma in China. *Chin J Cancer*. **30**, 114–119 (2011).
2. Lai, S. Z. *et al.* How does intensity-modulated radiotherapy versus conventional two-dimensional radiotherapy influence the treatment results in nasopharyngeal carcinoma patients? *Int J Radiat Oncol Biol Phys*. **80**, 661–668 (2011).
3. Lee, A. W. *et al.* Treatment results for nasopharyngeal carcinoma in the modern era: the Hong Kong experience. *Int J Radiat Oncol Biol Phys*. **61**, 1107–1116 (2005).
4. Li, J. X., Huang, S. M., Wen, B. X. & Lu, T. X. Prognostic factors on overall survival of newly diagnosed metastatic nasopharyngeal carcinoma. *Asian Pac J Cancer Prev*. **15**, 3169–3173 (2014).
5. Hui, E. P. *et al.* Lung metastasis alone in nasopharyngeal carcinoma: a relatively favorable prognostic group. A study by the Hong Kong Nasopharyngeal Carcinoma Study Group. *Cancer*. **101**, 300–306 (2004).
6. Wei, W. I. & Sham, J. S. Nasopharyngeal carcinoma. *Lancet*. **365**, 2041–2054 (2005).
7. Leung, S. F. *et al.* Plasma Epstein-Barr viral deoxyribonucleic acid quantitation complements tumor-node-metastasis staging prognostication in nasopharyngeal carcinoma. *J Clin Oncol*. **24**, 5414–5418 (2006).
8. Xia, W. X. *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*. **49**, 2152–2160 (2013).
9. Zhou, G. Q. *et al.* Baseline serum lactate dehydrogenase levels for patients treated with intensity-modulated radiotherapy for nasopharyngeal carcinoma: a predictor of poor prognosis and subsequent liver metastasis. *Int J Radiat Oncol Biol Phys*. **82**, e359–365 (2012).
10. Liang, W. *et al.* Development and validation of a nomogram for predicting survival in patients with resected non-small-cell lung cancer. *J Clin Oncol*. **33**, 861–869 (2015).
11. Cho, J. K. *et al.* Development and external validation of nomograms predictive of response to radiation therapy and overall survival in nasopharyngeal cancer patients. *Eur J Cancer*. **51**, 1303–1311 (2015).
12. Sham, J. Prognostic factors in nasopharyngeal carcinoma investigated by computer tomography—an analysis of 659 patients by Teo *et al.* *Radiother Oncol*. **25**, 216–217 (1992).
13. Sham, J. S. & Choy, D. Prognostic factors of nasopharyngeal carcinoma: a review of 759 patients. *Br J Radiol*. **63**, 51–58 (1990).

14. Garden, A. S. *et al.* Outcomes and patterns of care of patients with locally advanced oropharyngeal carcinoma treated in the early 21st century. *Radiat Oncol.* **8**, 21 (2013).
15. Balkwill, F. & Mantovani, A. Cancer and inflammation: implications for pharmacology and therapeutics. *Clin Pharmacol Ther.* **87**, 401–406 (2010).
16. Pepys, M. B. & Hirschfield, G. M. C-reactive protein: a critical update. *J Clin Invest.* **111**, 1805–1812 (2003).
17. Scott, H. R. *et al.* The systemic inflammatory response, weight loss, performance status and survival in patients with inoperable non-small cell lung cancer. *Br J Cancer.* **87**, 264–267 (2002).
18. Eigentler, T. K. *et al.* Number of metastases, serum lactate dehydrogenase level, and type of treatment are prognostic factors in patients with brain metastases of malignant melanoma. *Cancer.* **117**, 1697–1703 (2011).
19. Harris, A. L. Hypoxia—a key regulatory factor in tumour growth. *Nat Rev Cancer.* **2**, 38–47 (2002).
20. Macdonald, M. R. *et al.* Role of Epstein-Barr virus in fine-needle aspirates of metastatic neck nodes in the diagnosis of nasopharyngeal carcinoma. *Head Neck.* **17**, 487–493 (1995).
21. Raab-Traub, N. & Flynn, K. The structure of the termini of the Epstein-Barr virus as a marker of clonal cellular proliferation. *Cell.* **47**, 883–889 (1986).
22. An, X. *et al.* Plasma Epstein-Barr virus DNA level strongly predicts survival in metastatic/recurrent nasopharyngeal carcinoma treated with palliative chemotherapy. *Cancer.* **117**, 3750–3757 (2011).

Acknowledgements

This work was supported by Natural Science Foundation of Guangdong Province (No: 303040790002). The funding agency had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author Contributions

L.Y. and Y.C. designed the overall project. L.Y. analyzed the data and wrote the manuscript; S.H. and Y.W. collected and analyzed the data; H.C. and S.L. performed the research; P.P. did the statistical analysis. All the authors (L.Y., S.H., Y.W., H.C., S.L., P.P. and Y.C.) have read and approved the final manuscript.

Additional Information

Supplementary information accompanies this paper at <http://www.nature.com/srep>

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Yang, L. *et al.* Development and External Validation of Nomograms for Predicting Survival in Nasopharyngeal Carcinoma Patients after Definitive Radiotherapy. *Sci. Rep.* **5**, 15638; doi: 10.1038/srep15638 (2015).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>