Improving the prediction of overall survival for head and neck cancer patients using image biomarkers in combination with clinical parameters

Zhai, Tian-Tian; van Dijk, Lisanne V; Huang, Bao-Tian; Lin, Zhi-Xiong; Ribeiro, Cássia O; Brouwer, Charlotte L; Oosting, Sjoukje F; Halmos, Gyorgy B; Witjes, Max J H; Langendijk, Johannes A

Published in:
Radiotherapy and Oncology

DOI:
10.1016/j.radonc.2017.07.013

Radiotherapy and Oncology 124 (2017) 256–262

Contents lists available at ScienceDirect

Radiotherapy and Oncology

journal homepage: www.thegreenjournal.com

Biomarkers in HNSCC

Improving the prediction of overall survival for head and neck cancer patients using image biomarkers in combination with clinical parameters

Tian-Tian Zhai [a,b,*], Lisanne V. van Dijk [a], Bao-Tian Huang [b], Zhi-Xiong Lin [b,*], Cássia O. Ribeiro [a], Charlotte L. Brouwer [a], Sjoukje F. Oosting [c], Gyorgy B. Halmos [d], Max J.H. Witjes [e], Johannes A. Langendijk [a], Roel J.H.M. Steenbakkers [a], Nanna M. Sijtsema [a]

[a] Department of Radiation Oncology, University Medical Center Groningen, University of Groningen, The Netherlands; [b] Department of Radiation Oncology, Cancer Hospital of Shantou University Medical College, China; [c] Department of Medical Oncology, University Medical Center Groningen, University of Groningen; [d] Department of Otolaryngology, University Medical Center Groningen, University of Groningen; and [e] Department of Maxillofacial Surgery, University Medical Center Groningen, University of Groningen, The Netherlands

ARTICLE INFO

Article history:
Received 28 March 2017
Received in revised form 25 June 2017
Accepted 13 July 2017
Available online 29 July 2017

Keywords:
Head and neck cancer
Nasopharyngeal cancer
Image biomarker
Overall survival
Prediction model

ABSTRACT

Purpose: To develop and validate prediction models of overall survival (OS) for head and neck cancer (HNC) patients based on image biomarkers (IBMs) of the primary tumor and positive lymph nodes (Ln) in combination with clinical parameters.

Material and methods: The study cohort was composed of 289 nasopharyngeal cancer (NPC) patients from China and 298 HNC patients from the Netherlands. Multivariable Cox-regression analysis was performed to select clinical parameters from the NPC and HNC datasets, and IBMs from the NPC dataset. Final prediction models were based on both IBMs and clinical parameters.

Results: Multivariable Cox-regression analysis identified three independent IBMs (tumor Volume-density, Run Length Non-uniformity and Ln Major-axis-length). This IBM model showed a concordance (c)-index of 0.72 (95%CI: 0.65–0.79) for the NPC dataset, which performed reasonably with a c-index of 0.67 (95%CI: 0.62–0.72) in the external validation HNC dataset. When IBMs were added in clinical models, the c-index of the NPC and HNC datasets improved to 0.75 (95%CI: 0.68–0.82; $p = 0.019$) and 0.75 (95%CI: 0.70–0.81; $p < 0.001$), respectively.

Conclusion: The addition of IBMs from the primary tumor and Ln improved the prognostic performance of the models containing clinical factors only. These combined models may improve pre-treatment individualized prediction of OS for HNC patients.

Head and neck cancer (HNC) accounts for about 0.65 million new cancer cases and 0.35 million cancer deaths worldwide every year [1]. Based on the Surveillance, Epidemiology, and End Results (SEER) data, the 5-year overall survival (OS) for HNC patients is approximately 60% [2]. The introduction of more intensified treatment regimens has resulted in improved OS rates, however the number of patients developing locoregional failure or distant metastases remains substantial [3,4]. To enable more personalized treatment approaches, risk stratification is becoming increasingly important [5]. Risk stratification in HNC requires new, robust and prognostic parameters to identify patients with different risk profiles for locoregional recurrence, distant metastasis and death [6–8].

In routine clinical practice, the TNM staging system is used to guide treatment decision-making often in combination with other classical prognostic factors such as performance status, tumor characteristics and age [9,10]. However, patients with similar prognostic factors may have different outcome [6,7] and thus new prognostic factors are needed to improve outcome prediction accuracy when added to prediction models based on classical prognostic factors only.

Recent studies have demonstrated the potential value of image biomarkers (IBMs), which are significantly associated with OS and complications in HNC, thoracic, pancreatic and colorectal cancer [11–13]. IBMs can be extracted from medical images and provide quantitative information regarding intensity, shape and textural characteristics of the region of interest [14–17]. By extracting

* Corresponding authors at: Department of Radiation Oncology, University Medical Center Groningen, P.O. Box 30001, 9300 RB Groningen, The Netherlands (T.-T. Zhai). Department of Radiation Oncology, Cancer Hospital of Shantou University Medical College, 515031 Shantou, China (Z.-X. Lin).
E-mail addresses: t.zhai@umcg.nl (T.-T. Zhai), zxlin5@qq.com (Z.-X. Lin).

http://dx.doi.org/10.1016/j.radonc.2017.07.013
0167-8140/

IBMs, the three-dimensional morphological tumor information can be transformed into multi-dimensional and mineable data [5,18]. Furthermore, IBMs enable decoding of a general prognostic phenotype existing in different cancer types, which may widen the scope of application [11].

Although many IBMs are significantly associated with outcome, it remains unclear to what extent the addition of IBMs improves the predictive power of models only consisting of classical prognostic factors, such as TNM staging and performance status. The aim of this study was to test whether the performance of prediction models for OS could be improved by the addition of IBMs compared to models based on solely classical prognostic factors for nasopharyngeal cancer (NPC) patients. Furthermore, the ability to generalize the prognostic value of IBMs for different tumor types was determined by externally validating this value for other HNC subtypes.

Materials and methods

Patient demographics and treatment

This retrospective study was composed of 289 consecutive NPC patients. Patients were treated with (chemo-)radiotherapy between January 2010 and June 2011 at the Cancer Hospital of Shantou University Medical College. All patients received a pre-treatment computed tomography (CT) scan (Philips Brilliance CT Big Bore Oncology Configuration, Cleveland, OH, USA; voxel size: $1.0 \times 1.0 \times 3.0$ mm; scan voltage: 120 kV; convolution kernel: Philips Healthcare's B) for radiotherapy planning. Patients were primarily treated with intensity-modulated radiotherapy (IMRT) and received a total dose of 70.4 Gy with fractions of 2.2 Gy in 6.5 weeks (5 fractions per week).

An additional set of 298 consecutive HNC patients (including 4.4% NPC patients) was treated with definitive radiotherapy, either combined or not, with chemotherapy or cetuximab at the University Medical Center Groningen between November 2007 and May 2013. For all patients, a pre-treatment CT-scan (Somatom Sensation Open, Siemens, Forchheim, Germany; voxel size: $1.0 \times 1.0 \times 2.0$ mm; scan voltage: 120 kV; convolution kernel: B30) was acquired for radiotherapy planning. Radiotherapy consisted of primarily three-dimensional conformal radiotherapy or IMRT to a total dose of 70 Gy with fractions of 2 Gy in 6–7 weeks (6 or 5 fractions per week).

Inclusion criteria were as follows: confirmed primary tumor with pathological diagnosis, standard contrast-enhanced planning CT-scan, treatment with curative intent, and OS data available.

Clinical parameters

All clinical parameters including age, gender, tumor location, treatment modality, human papilloma virus (HPV) status (only for oropharyngeal cancer (OPC)) and World Health Organization performance status (WHO PS) [19] were derived from medical records. Dose-volume information of the primary tumor (PT) and positive lymph nodes (pLN) was derived from the radiotherapy planning system (mean dose, $V50$, $V60$, $V70$, $V80$, $D_{90\%}$, $D_{95\%}$ and $D_{98\%}$). Tumor (T) and positive lymph node (N) stage were defined according to the 6th edition of the American Joint Committee on Cancer Staging Manual [10].

CT image biomarkers

The PT and pLN were delineated for the NPC and HNC datasets on the planning CT-scan by experienced head and neck radiation-oncologists. In-house software was used to extract the IBMs, developed using common formulas in Matlab R2014a (Mathworks, Natick, USA). Twenty-four CT intensity and 20 geometric IBMs were directly derived from every delineated structure (the PT, all pLN and the pLN with the largest volume). The intensity IBMs were obtained from the histogram of all voxel values, such as median of the voxels and entropy of the voxels. Geometric IBMs, such as volume, compactness and major axis length, were calculated from the three-dimensional shape and size of the contoured structures. Ninety textural CT IBMs from both the PT and the pLN with the largest volume were defined to quantify the heterogeneity of tissue. They were derived from three different matrices: the gray level co-occurrence matrix (GLCM) [15], gray level run-length matrix (GLRLM) [16] and gray level size-zone matrix (GLSZM) [17]. GLCM describes the gray level transition, GLRLM and GLSZM describe the directional and volumetric gray level repetition. They were calculated from the three-dimensional contoured structures. More details on feature extraction and used algorithms are described in our previous publication [20]. The lymph node IBMs from patients without lymph node metastasis were defined as 0.

Data analysis

The endpoint of this study was OS, defined as the time from the first day of radiotherapy to the date of death from any cause. Patients alive were censored at the date of last follow up. An overview of the analysis design is shown in Fig. 1.

Step 1: Clinical models

Potential clinical parameters that were considered for their prognostic ability in the NPC and HNC datasets included age (>median vs. ≤median), gender (female vs. male), T-stage (T3–T4 vs. T1–T2), N-stage (N2–N3 vs. N0–N1), treatment modality (RT with systemic treatment vs. RT only), WHO PS (1–3 vs. 0) and dose parameters (>median vs. ≤median). HPV status assessed by p16 immunohistochemistry and DNA polymerase chain reaction (OPC positive vs. others) was included in the analysis for the HNC dataset, as this is a strong risk factor for oropharyngeal cancer [10,21,22]. Due to the known difference in etiology between NPC and HNC, two multivariable clinical prediction models were created: one based on the NPC and the other on the HNC dataset.

Step 2: IBM model

IBM variables were pre-selected to reduce the probability of over-fitting. If the Pearson correlation between pairs of IBMs was larger than 0.80, then the IBM with the lower univariable association with OS was omitted from further analysis [23,24]. All pre-selected potential IBMs were analyzed for their prognostic power, using their median value (>median vs. ≤median) in the NPC dataset as the threshold value in the univariable analysis. After selection of the independent prognostic factors, the threshold values were optimized by testing the values around the median. A multivariable IBM model was developed based on the NPC dataset only. Finally, the thresholds of IBMs for the NPC dataset were used for the HNC dataset to externally validate the IBM model.

Step 3: Combined models

All clinical parameters from the NPC and HNC datasets and pre-selected IBMs from the NPC dataset were merged into multivariable analysis and the coefficients (β) of the features were refitted to the NPC dataset and HNC dataset respectively, to generate the combined IBM-NPC and IBM-HNC models.

Normal Q–Q probability plot, cumulative frequency (*P*–*P*) plot and the Kolmogorov–Smirnov test were used to test the normality of all potential clinical parameters and IBMs. The chi-square test was used to compare the rates and an independent sample t-test was used to compare normally distributed variables between different groups. Univariable Cox regression analysis was performed

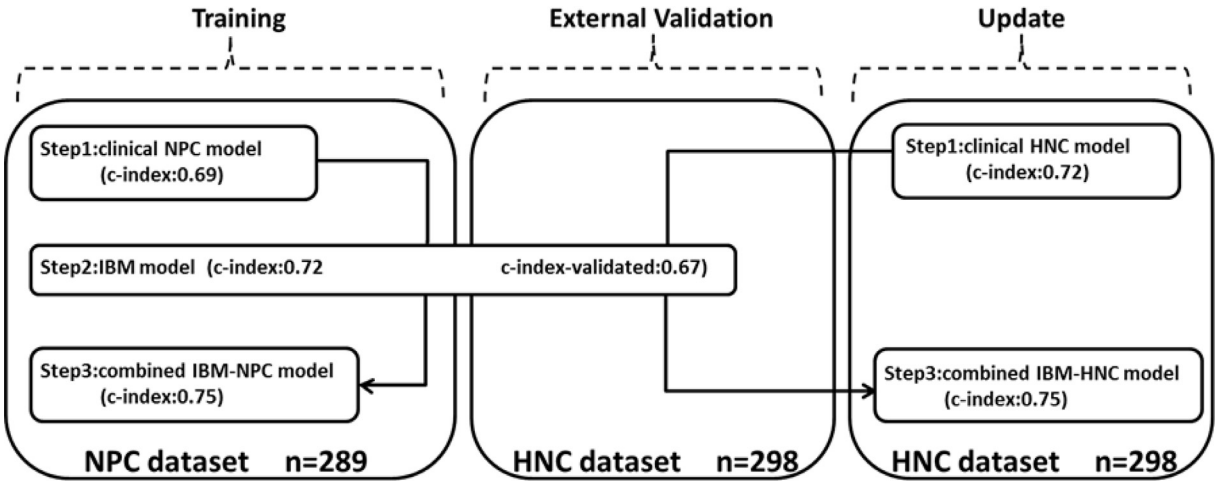

Fig. 1. Analysis workflow. Step 1: Two different clinical prediction models were created: for each of the NPC and HNC datasets. Step 2: an IBM model was created based on the NPC dataset and externally validated using the HNC dataset. Step 3: The combined models were created by combining the IBMs with the clinical parameters from NPC and HNC datasets separately. *Abbreviations:* NPC = nasopharyngeal cancer; c-index = concordance index; HNC = head and neck cancer; IBM = image biomarker.

to assess the risk factors for OS, and multivariable Cox proportional hazards regression analysis (forward selection: Likelihood ratio test, p-value < 0.05) was used for the development of the multivariable model. The concordance index (c-index) was determined to assess the models discriminative power. The z-score test (Package "compareC" in R) was used to test the difference between two c-indices. Internal validation was performed using bootstrap validation [25]. The differences in c-indices of the bootstrap model on the bootstrap sample and the original sample were calculated 1000 times. The optimism-corrected c-index was obtained by subtracting the average c-indices difference. The median of the linear predictor was defined as the threshold to separate the Kaplan–Meier survival curves: one curve showing patients with high hazard values (high risk) and the other with low hazard values (low risk). The Kaplan–Meier survival curves were compared using a log-rank test. Two tailed p-values < 0.05 were considered statistically significant. Statistical analysis was performed using the R software (version 3.2.1).

Results

The median follow up for the NPC patients was 37.6 months (range 2.4–58.6) and overall 64 deaths (22.1%) were observed. The median follow up for the HNC patients was 32.8 months (range: 1.6–89.7) and overall 126 deaths (42.3%) were observed.

Step 1: Clinical models

The clinical characteristics of the patients in the NPC and HNC datasets are listed in Table 1. All characteristics except gender were significantly different between the two datasets.

In the NPC dataset, the univariable analysis revealed significant associations between the minimum dose value of 98% of the primary tumor volume ($D_{98\%}$ of GTV-PT), age, N-stage and T-stage with OS (Table S1). In the multivariable analysis (Table S2), lower values of $D_{98\%}$ of GTV-PT (hazard ratio (HR): 0.37; 95% confidence interval (CI): 0.21–0.63), increasing age (HR: 2.14; 95%CI: 1.30–3.53) and N-stage (HR: 2.37; 95%CI: 1.26–4.45) were associated with worse OS. The multivariable clinical NPC model based on these three clinical features resulted in a c-index of 0.69 (95%CI: 0.61–0.76).

Univariable analysis showed that $D_{98\%}$ of GTV-PT, WHO PS, N-stage, T-stage and HPV-status were significantly associated with OS (Table S1) in the HNC dataset. The multivariable clinical HNC model (Table S2) including WHO PS (HR: 3.46; 95%CI: 2.41–4.97), $D_{98\%}$ of GTV-PT (HR: 0.51; 95%CI: 0.35–0.74) and T-stage (HR: 1.93; 95%CI: 1.28–2.91) as independent prognostic factors for OS, had a c-index of 0.72 (95%CI: 0.67–0.78).

Table 1
Baseline characteristics of the patients in the NPC and HNC datasets.

Characteristic	NPC dataset (n = 289) (%)	HNC dataset (n = 298) (%)	p-Value
Gender			0.489[b]
Male	203 (70.2)	217 (72.8)	
Female	86 (29.8)	81 (27.2)	
Age at diagnosis (median ± SD) (year)	51 ± 12	62 ± 11	<0.001[c]
T stage[a]			<0.001[b]
T1	22 (7.6)	22 (7.4)	
T2	49 (17.0)	103 (34.6)	
T3	104 (36.0)	90 (30.2)	
T4	114 (39.4)	83 (27.9)	
N stage[a]			<0.001[b]
N0	28 (9.7)	117 (39.3)	
N1	60 (20.8)	24 (8.1)	
N2	158 (54.7)	142 (47.7)	
N3	43 (14.9)	15 (5.0)	
Treatment modality			<0.001[b]
RT only	48 (16.6)	161 (54.0)	
RT with systemic treatment	241 (83.4)	137 (46.0)	
$D_{98\%}$ of GTV-PT (median ± SD) (Gy)	69.60 ± 2.30	68.22 ± 3.10	<0.001[c]
WHO PS			
0	241 (83.4)	192 (64.4)	<0.001[b]
1–3	48 (16.6)	106 (35.6)	
HPV			
OPC negative	–	68 (22.8)	
OPC positive	–	29 (9.7)	
Not OPC	–	201 (67.4)	
Tumor site			
Nasopharynx	289 (100.0)	13 (4.4)	
Oropharynx	–	97 (32.6)	
Hypopharynx	–	26 (8.7)	
Larynx	–	120 (40.3)	
Oral cavity	–	29 (9.7)	
Others	–	13 (4.4)	

Abbreviations: NPC = nasopharyngeal caner; HNC = head and neck cancer; T = tumor; N = lymph node; RT = radiotherapy; $D_{98\%}$ of GTV-PT = minimum dose value of 98% volume of primary tumor; WHO PS = World Health Organization performance status; HPV = human papilloma virus; OPC = oropharyngeal cancer.
[a] According to the 6th edition of the AJCC/UICC staging system.
[b] p-Value was calculated using the chi-square test.
[c] p-Value was calculated using the independent samples t-test.

The Kaplan–Meier curves of clinical NPC and HNC models are shown in Fig. 2A and B. The probability of OS after 2 and 3 years was calculated for the two datasets, and shown in Table S3.

Step 2: IBM model

Forty-eight of the 312 IBMs were pre-selected from the NPC dataset by testing the inter-variable correlation. In the univariable analysis, nine of the pre-selected IBMs were significantly associated with OS (Table S4). Multivariable analysis revealed three independent features: Volume-density (HR, 0.44; 95%CI: 0.26–0.74), Run Length Non-uniformity (RLN) (HR, 2.98; 95%CI, 1.73–5.14) from the PT and Major-axis-length from all pLN (threshold: 55 mm; HR: 2.11; 95%CI, 1.29–3.46) (Table S5). The optimal threshold was the median unless stated otherwise. A multivariable IBM model was developed based on the three IBMs and resulted in

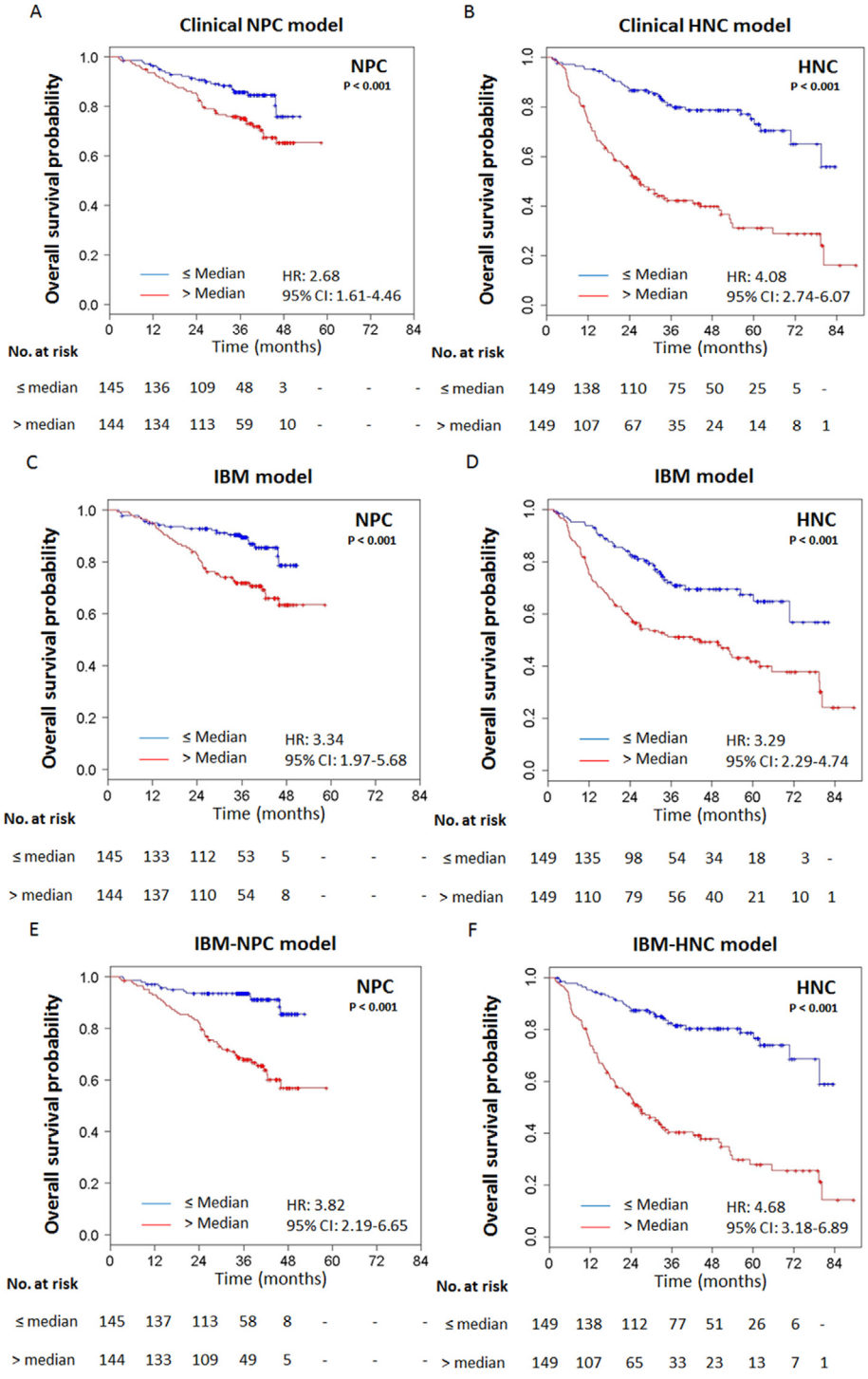

Fig. 2. Overall survival stratified by risk groups for the NPC and HNC datasets. The survival curve separation and hazard ratio (>median vs. ≤median) between different risk groups are shown based on the clinical model of NPC dataset (A) and HNC dataset (B), IBM model in NPC dataset (C) and HNC dataset (D), the combined IBM-NPC model (E) in the NPC dataset and the combined IBM-HNC model in the HNC dataset (F). *Abbreviations:* NPC = nasopharyngeal cancer; HNC = head and neck cancer; IBM = image biomarker; HR = hazard ratio; CI = confidence interval.

a c-index of 0.72 (95%CI: 0.65–0.79). External validation using the HNC dataset resulted in a c-index of 0.67 (95%CI: 0.62–0.72) (Fig. 2C and D). The subgroup multivariable analysis using patient data with positive lymph nodes only revealed the same three significant IBMs with similar model performance.

Step 3: Combined models

For the NPC dataset, RLN, Volume-density, lymph nodes' Major-axis-length, $D_{98\%}$ of GTV-PT and age were identified as independent prognostic factors. However, N-stage was no longer significantly related to OS. Compared to the clinical prediction model, the combined model showed a significantly improved c-index of 0.75 (95% CI: 0.68–0.82; $p = 0.019$) (Fig. 2E). For the HNC dataset, WHO PS, RLN, lymph nodes' Major-axis-length, T-stage, Volume-density and $D_{98\%}$ of GTV-PT were identified as independent prognostic factors. Compared to the clinical HNC model, the combined IBM-HNC model showed better performance with an increased c-index of 0.75 (95%CI: 0.70–0.81; $p < 0.001$) (Fig. 2F and Table S6). The coefficients and performance of the two combined models are depicted in Table 2.

Discussion

Non-invasive tools to predict treatment outcome could add value in the guidance of individual therapeutic strategies for HNC patients [26]. In this study, two PT IBMs (Volume-density and RLN) and one pLN IBM (Major-axis-length) were extracted from CT images of an NPC dataset. These IBMs performed well for the external HNC dataset and significantly improved the prognostic performance of the models based on the clinical parameters only, both in NPC and HNC populations. This study is the first to investigate not only IBMs of the primary tumor site but IBMs of the pathologic lymph nodes as potential prognostic factors.

Although NPC falls under the category of HNC, the prognosis of patients with NPC is more favorable than HNC of other sites [2,7]. Furthermore, NPC has a distinct etiology compared to other HNC sites. The risk factors for NPC (Epstein-Barr virus, salted fish, preserved foods and genetic components) are also different from those of HNC (HPV, smoking and alcohol) [27]. Moreover, the geographical distribution and incidence of NPC are different from other HNC [28–31]. Therefore, two separate clinical models for both the NPC and the HNC datasets were developed and the combined models were fitted to the respective cohort without external validation.

Volume-density refers to the tumor volume bounded by the smallest cube containing the tumor (Fig. 3A and B). The smaller the Volume-density, the more irregular the tumor shape, indicating tumor growth patterns [25,32,33]. A more invasive, irregular-shaped tumor requires a larger bounding cube. In Fig. 3A, the tumor with a smaller Volume-density is in more irregular shape.

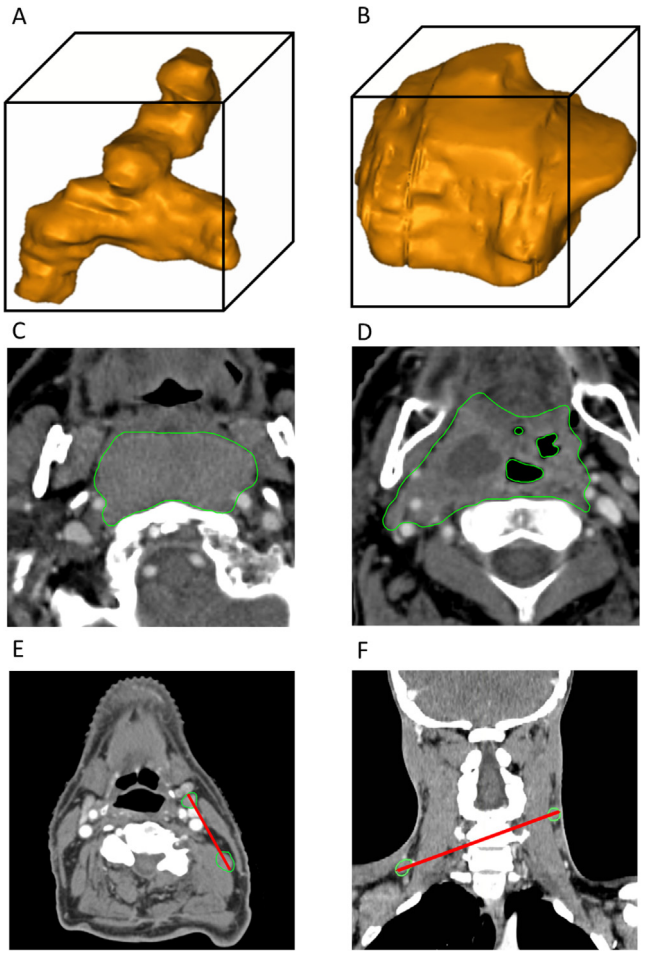

Fig. 3. Examples of patients with low (A) and high (B) values of Volume-density of the tumor. Examples of patients with low (C) and high (D) values of Run Length Non-uniformity of the tumor. Examples of Major-axis-length in patients with two unilateral (E) and two bilateral (F) positive lymph nodes.

Irregularly shaped tumors were associated with worse OS, which may reflect the increase of their aggressive behavior [1,25,32,33].

The RLN is derived from the GLRLM, which describes the frequency of consecutive voxels with the same gray level value. RLN is low when runs are equally distributed along the run length in contoured structures. A higher RLN value indicates PT heterogeneity, our results showed it was associated with worse survival for the NPC and HNC datasets (Fig. 3C and D). This observation is supported by other studies [11,34,35]. Aerts et al. quantified the intra-tumor heterogeneity using another feature "Gray level Non-Uniformity", which had a high correlation (0.79) with RLN [11] in

Table 2
Estimated coefficients of the combined models.

	Combined IBM-NPC model				Combined IBM-HNC model			
	β	Corrected β	p-Value	HR (95%CI)	β	Corrected β	p-Value	HR (95%CI)
Run length non-uniformity	0.69	0.47	0.028	2.00 (1.08–3.69)	0.37	0.24	0.099	1.45 (0.93–2.25)
Volume-density	−0.88	−0.60	0.001	0.42 (0.24–0.71)	−0.41	−0.27	0.023	0.66 (0.46–0.94)
Major-axis-length	0.77	0.52	0.003	2.16 (1.31–3.57)	1.33	0.86	<0.001	3.79 (2.00–7.19)
$D_{98\%}$ of GTV-PT	−0.84	−0.57	0.008	0.43 (0.23–0.81)	−0.44	−0.29	0.032	0.65 (0.43–0.96)
Age	0.56	0.38	0.029	1.76 (1.06–2.91)	–	–	–	–
T stage	–	–	–	–	0.55	0.36	0.020	1.74 (1.09–2.76)
WHO PS	–	–	–	–	1.36	0.88	<0.001	3.88 (2.66–5.66)

Abbreviations: IBM = image biomarker; NPC = nasopharyngeal caner; HNC = head and neck cancer; HR = Hazard ratio; CI = confidence interval; $D_{98\%}$ of GTV-PT = minimum dose value of 98% volume of primary tumor; T = tumor; WHO PS = World Health Organization performance status.

our study. Multiple subclonal populations coexist in the tumor and their evolution causes the intratumor heterogeneity. This process has been shown at cellular and genetic levels [36–39]. However, intratumor heterogeneity, can be underestimated if determined from a single or limited tumor-biopsy sample [36,40]. Medical imaging could be a better tool for detecting the spatial dimension of tumor features. We quantified IBMs by extracting features from the complete tumor volume, in other words IBMs reflect the overall tumor feature. It would be even possible to distinguish different regions within the tumor based on the IBMs. The association between IBMs and cellular and genetic information requires further investigation [11,38,41].

The Major-axis-length of pLN is defined as the longest distance between any two voxels of the all pLN. This could either be for one pLN, or any two unilateral or bilateral pLNs (Fig. 2E and F). Our results demonstrated that the longer length (>55 mm) was associated with an increased risk of death. N-stage did not increase prediction performance for the combined IBM-NPC model. N1 stage represents patients with only unilateral metastasis. In this situation, patients with N1 would have much shorter Major-axis-length compared to patients with N2 and N3. In fact, we found that Major-axis-length from the entire pLN had a linear correlation of 0.67 with the clinical variable N-stage. However, in the univariable analysis, the c-index of Major-axis-length was larger than that of the N-stage (0.60 vs. 0.57). Therefore, we expect that Major-axis-length would be a good substitute for N-stage to improve survival prediction in the future.

To explore the general prognostic cancer phenotype and enlarge the scope of the IBM model application, Aerts et al. developed prognostic models based on IBMs from a lung cancer dataset and validated them against HNC datasets (c-index: 0.69) [11]. We have shown that the NPC IBM model performed well against the HNC dataset (c-index: 0.67), supporting the generalization of prognostic IBMs.

When IBMs were added in clinical models, the c-index of the NPC and HNC datasets significantly improved from 0.69 to 0.75 ($p = 0.019$) and from 0.72 to 0.75 ($p < 0.001$), respectively. Fig. 2 showed that the Kaplan–Meier curve separation and hazard ratio between high-low risk patients are larger if the patients are stratified with the combined IBM-NPC and IBM-HNC models (Fig. 2E and F) than with the separate clinical (Fig. 2A and B) and IBM (Fig. 2C and D) models. The exact survival probability difference is shown in Table S3. For example, two year survival difference between low and high risk groups improved from 10.8% with the clinical model to 14.2% with the combined model in the NPC group.

Standardization of image acquisition and reconstruction between different institutions are necessary for sharing quantitative image analyses. Different CT slice thickness and convolution kernel may influence the IBM model accuracy. Although the IBM model was validated externally for the HNC dataset, the clinical models and also the combined models were only fitted to the respective cohort without external validation, external validation of the combined models is the subject of future study. Moreover, a more systematic and thorough analysis using different endpoints, such as disease free survival and locoregional control, will also need to be investigated in the future.

Conclusion

In conclusion, the addition of image biomarkers from the primary tumor and positive lymph nodes improved the performance of clinical prediction models significantly for both NPC and HNC datasets. This addition could facilitate the pretreatment individualized prediction of survival for head and neck cancer patients.

Conflict of interest statement

The authors state that the research presented in this manuscript is free of conflicts of interest.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.radonc.2017.07.013.

References

[1] Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-tieulent J, Jemal A. Global cancer statistics, 2012. CA Cancer J Clin 2015;65:87–108.
[2] Howlader N, Noone AM, Krapcho M, et al. SEER cancer statistics review, 1975–2013, based on November 2015 SEER. http://seer.cancer.gov/csr/1975_2013/.
[3] Pignon JP, le Maître A, Maillard E, Bourhis J, MACH-NC Collaborative Group. Meta-analysis of chemotherapy in head and neck cancer (MACH-NC): an update on 93 randomised trials and 17,346 patients. Radiother Oncol 2009;92:4–14.
[4] Blanchard P, Baujat B, Holostenco V, et al. Meta-analysis of chemotherapy in head and neck cancer (MACH-NC): a comprehensive analysis by tumour site. Radiother Oncol 2011;100:33–40.
[5] Aerts HJ. The potential of radiomic-based phenotyping in precision medicine: a review. JAMA Oncol 2016;2:1636–42.
[6] Jia WH, Huang QH, Liao J, et al. Trends in incidence and mortality of nasopharyngeal carcinoma over a 20–25 year period (1978/1983–2002) in Sihui and Cangwu counties in southern China. BMC Cancer 2006;6:178.
[7] Argiris A, Karamouzis MV, Raben D, Ferris RL. Head and neck cancer. Lancet 2008;371:1695–709.
[8] Liu MT, Chang TH, Lin JP, Huang CC, Wang AY. Prognostic factors affecting the outcome of nasopharyngeal carcinoma. Jpn J Clin Oncol 2003;33:501–8.
[9] Ma J, Mai HQ, Hong MH, et al. Is the 1997 AJCC staging system for nasopharyngeal carcinoma prognostically useful for Chinese patient populations? Int J Radiat Oncol Biol Phys 2001;50:1181–9.
[10] Pfister DG, Spencer S, Brizel DM, et al. Head and neck cancers, version 1.2015. J Nat Compr Cancer Networks 2015;13:847–55.
[11] Aerts HJ, Velazquez ER, Leijenaar RT, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat Commun 2014;5:4006.
[12] Grove O, Berglund AE, Schabath MB, et al. Quantitative computed tomographic descriptors associate tumor shape complexity and intratumor heterogeneity with prognosis in lung adenocarcinoma. PLoS One 2015;10:e0118261.
[13] Huang YQ, Liang CH, He L, et al. Development and validation of a radiomics nomogram for preoperative prediction of lymph node metastasis in colorectal cancer. J Clin Oncol 2016;34:2157–64.
[14] Cui Y, Song J, Pollom E, et al. Quantitative analysis of (18)F-fluorodeoxyglucose positron emission tomography identifies novel prognostic imaging biomarkers in locally advanced pancreatic cancer patients treated with stereotactic body radiation therapy. Int J Radiat Oncol Biol Phys 2016;96:102–9.
[15] Haralick R, Shanmugan K, Dinstein I. Textural features for image classification. IEEE Trans Syst Man Cybern 1973;3:610–21.
[16] Tang X. Texture information in run-length matrices. IEEE Trans Image Process 1998;7:1602–9.
[17] Thibault G, Fertil B, Navarro C, et al. Texture indexes and gray level size zone matrix application to cell nuclei classification. Pattern Recognit Inf Process 2009:140–5.
[18] Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. Radiology 2016;278:563–77.
[19] Oken MM, Creech RH, Tormey DC, et al. Toxicity and response criteria of the Eastern Cooperative Oncology Group. Am J Clin Oncol 1982;5:649–55.
[20] van Dijk LV, Brouwer CL, van der Schaaf A, et al. CT image biomarkers to improve patient-specific prediction of radiation-induced xerostomia and sticky saliva. Radiother Oncol 2017;122:185–91.
[21] Ang KK, Harris J, Wheeler R, et al. Human papillomavirus and survival of patients with oropharyngeal cancer. N Engl J Med 2010;363:24–35.
[22] Kano M, Kondo S, Wakisaka N, et al. The influence of human papillomavirus on nasopharyngeal carcinoma in Japan. Auris Nasus Larynx 2016. S0385-8146:30223-1.
[23] Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc B 1995;57:289–300.
[24] van der Schaaf A, Xu CJ, van Luijk P, Van't Veld AA, Langendijk JA, Schilstra C. Multivariate modeling of complications with data driven variable selection: guarding against overfitting and effects of data set size. Radiother Oncol 2012;105:115–21.
[25] Moons KGM, Altman DG, Reitsma JB, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): explanation and elaboration. Ann Intern Med 2015;162:W1–W73.
[26] Emerick KS, Leavitt ER, Michaelson JS, Diephuis B, Clark JR, Deschler DG. Initial clinical findings of a mathematical model to predict survival of head and neck cancer. Otolaryngol Head Neck Surg 2013;149:572–8.

[27] Rodriguez CP, Adelstein DJ. Survival trends in head and neck cancer: opportunities for improving outcomes. Oncologist 2010;15:921–3.
[28] Patel VJ, Chen NW, Resto VA. Racial and ethnic disparities in nasopharyngeal cancer survival in the United States: a SEER study. Otolaryngol Head Neck Surg 2017;156:122–31.
[29] Sethi S, Ali-fehmi R, Franceschi S, et al. Characteristics and survival of head and neck cancer by HPV status: a cancer registry-based study. Int J Cancer 2012;131:1179–86.
[30] Dayyani F, Etzel CJ, Liu M, Ho C, Lippman SM, Tsao AS. Meta-analysis of the impact of human papillomavirus (HPV) on cancer risk and overall survival in head and neck squamous cell carcinomas (HNSCC). Head Neck Oncol 2010:2–15.
[31] Zang J, Li C, Zhao LN, et al. Prognostic model of death and distant metastasis for nasopharyngeal carcinoma patients receiving 3DCRT/IMRT in nonendemic area of China. Medicine (Baltimore) 2016;95:e3794.
[32] Chen LL, Nolan ME, Silverstein MJ, et al. The impact of primary tumor size, lymph node status and other prognostic factors on the risk of cancer death. Cancer 2009;115:5071–83.
[33] Qin L, Wu F, Lu H, Wei B, Li G, Wang R. Tumor volume predicts survival rate of advanced nasopharyngeal carcinoma treated with concurrent chemoradiotherapy. Otolaryngol Head Neck Surg 2016;155:598–605.
[34] Chicklore S, Goh V, Siddique M, Roy A, Marsden PK, Cook GJR. Quantifying tumor heterogeneity in 18F-FDG PET/CT imaging by texture analysis. Eur J Nucl Med Mol Imaging 2013;40:133–40.
[35] Tixier F, Le Rest CC, Hatt M, et al. Intratumor heterogeneity characterized by textural features on baseline 18F-FDG PET images predicts response to concomitant radiochemotherapy in esophageal cancer. J Nucl Med 2011;52:369–78.
[36] Gerlinger M, Rowan AJ, Horswell S, et al. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. N Engl J Med 2012;366:883–92.
[37] Zhang XC, Xu C, Zhang B, et al. Tumor evolution and intratumor heterogeneity of an oropharyngeal squamous cell carcinoma revealed by whole-genome sequencing. Neoplasia 2013;15:1371–8.
[38] Balluff B, Hanselmann M, Heeren RM. Mass spectrometry imaging for the investigation of intratumor heterogeneity. Adv Cancer Res 2017;134:201–30.
[39] Alizadeh AA, Aranda V, Bardelli A, et al. Toward understanding and exploiting tumor heterogeneity. Nat Med 2015;21:846–53.
[40] Lipinski KA, Barber LJ, Davies MN, Ashenden M, Sottoriva A, Gerlinger M. Cancer evolution and the limits of predictability in precision cancer medicine. Trends Cancer 2016;2:49–63.
[41] O'Connor JP, Rose CJ, Waterton JC, Carano RA, Parker GJ, Jackson A. Imaging intratumor heterogeneity: role in therapy response, resistance, and clinical outcome. Clin Cancer Res 2015;21:249–57.