The prognostic value of CT radiomic features from primary tumours and pathological lymph nodes in head and neck cancer patients

Zhai, Tiantian

DOI:
10.33612/diss.111448998

CHAPTER 3

The prognostic value of CT-based image-biomarkers for head and neck cancer patients treated with definitive (chemo-)radiation

Published in: **Oral Oncology** 2019 August;95:178-186.

Tian-Tian Zhai[a,*], Johannes A. Langendijk[a], Lisanne V. van Dijk[a], Gyorgy B. Halmos[c], Max J.H. Witjes[d], Sjoukje F. Oosting[e], Walter Noordzij[f], Nanna M. Sijtsema[a], Roel J.H.M. Steenbakkers[a]

[a] Department of Radiation Oncology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands
[b] Department of Radiation Oncology, Cancer Hospital of Shantou University Medical College, Shantou, China
[c] Department of Otorhinolaryngology, Head and Neck Surgery, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands
[d] Department of Maxillofacial Surgery, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands
[e] Department of Medical Oncology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands
[f] Department of Nuclear Medicine and Molecular Imaging, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands
*Corresponding author

Abstract

Objectives

The aim of this study was to investigate whether quantitative CT image-biomarkers (IBMs) can improve the prediction models with only classical prognostic factors for local-control (LC), regional-control (RC), distant metastasis-free survival (DMFS) and disease-free survival (DFS) for head and neck cancer (HNC) patients.

Materials and Methods

The cohort included 240 and 204 HNC patients in the training and validation analysis, respectively. Clinical variables were scored prospectively and IBMs of the primary tumor and lymph nodes were extracted from planning CT-images. Clinical, IBM and combined models were created from multivariable Cox proportional-hazard analyses based on clinical features, IBMs, and both for LC, RC, DMFS and DFS.

Results

Clinical variables identified in the multivariable analysis included tumor-site, WHO performance-score, tumor-stage and age. Bounding-box-volume describing the tumor volume and irregular shape, IBM correlation representing radiological heterogeneity, and LN_major-axis-length showing the distance between lymph nodes were included in the IBM models. The performance of IBM LC, RC, DMFS and DFS models (c-index(validated):0.62, 0.80, 0.68 and 0.65) were comparable to that of the clinical models (0.62, 0.76, 0.70 and 0.66). The combined DFS model (0.70) including clinical features and IBMs performed significantly better than the clinical model. Patients stratified with the combined models revealed larger differences between risk groups in the validation cohort than with clinical models for LC, RC and DFS. For DMFS, the differences were similar to the clinical model.

Conclusion

For prediction of HNC treatment outcomes, image-biomarkers performed as good as or slightly better than clinical variables.

Introduction

Head and neck squamous cell carcinoma (HNSCC) is primarily managed by surgery and/or radiotherapy (RT) with or without systemic treatment. At present, the 5-year overall survival rate is around 60% [1]. However, 30%-50% of patients with locally advanced HNSCC still experience treatment failures, predominantly occurring at the site of the primary tumor, followed by regional failures and distant metastases [2]. Risk assessment of local control (LC), regional control (RC), distant metastasis-free survival (DMFS) and disease-free survival (DFS) of HNSCC patients becomes increasingly important to optimise treatment [3-7]. For HNSCC patients, molecular-based factors such as human papillomavirus (HPV) [3,4] and patient-specific factors such as age and World Health Organization performance-status (WHO PS) have been identified as prognostic clinical factors for LC, RC, DMFS and DFS [5-7]. However, these clinical factors are not sufficient for identifying patients that will benefit most from specific treatment strategies. For this purpose, more detailed information is required, including factors that reflect the characteristics of the whole tumor, such as tumor volume, shape and heterogeneity [8-16].

A wide variety of medical images is generated for diagnostic and staging purposes, such as TNM staging. In current clinical practice, these images are used for guiding treatment decision-making [3]. Image-biomarkers (IBMs) may also be extracted from these medical images, transforming image data into quantitative information that describes intensity, shape and textural characteristics of the whole tumor. IBMs can provide more spatial and textural information about tumor features than TNM staging [8-10]. For patients treated with primary non-surgical modalities, like (chemo)radiotherapy, where only limited pathological information is available, the use of image-biomarkers might improve the prediction of treatment outcomes and improve medical decision-making [9].

IBMs have demonstrated their value to predict treatment outcome and complications for patients with head and neck, lung, breast, pancreatic, and colorectal cancers [9,11-16]. In a previous study, we showed that the quantitative computed-tomography (CT) IBMs were good substitutes for the qualitative clinical variable N-stage, and that they improved the performance of multivariable prediction models for overall survival, compared to models consisting of clinical variables alone [17]. The next challenge is to find IBMs that allow for a better prediction of local-regional failure and distant

metastasis. Based on our previous study, we hypothesized that IBMs would provide similar or better predictive information than clinical variables for LC, RC, DMFS and DFS. The aim of this study was to investigate whether multivariable prediction models for LC, RC, DMFS and DFS, consisting of both clinical variables and IBMs perform better than prediction models with only classical prognostic factors.

Materials and Methods

Patient selection and treatment

This was a retrospective analysis in a prospective cohort study, which was composed of 707 consecutive non-surgically treated HNSCC patients. The tumors originated in the oral cavity, oropharynx, nasopharynx, hypopharynx or larynx and were primarily treated with definitive radiotherapy at the University Medical Center Groningen between July 2007 and December 2015. We excluded 202 patients without contrast-enhanced planning CT-scans, 45 patients with metal or motion artifacts in the region of the primary tumor (PT) or positive lymph nodes (LN), and 16 patients with previous neck dissection. Overall 444 patients with standard contrast-enhanced planning CT-scans (Somatom Sensation Open, Siemens, Forchheim, Germany; voxel size: 1.0×1.0×2.0 mm; scan voltage: 120kV; and convolution kernel: B30) were included. Overall 240 patients treated before June 2012 were enrolled in the training cohort and 204 patients treated after June 2012 in the validation cohort. All patients were treated with definitive three-dimensional conformal radiotherapy (3D-CRT), intensity-modulated radiotherapy (IMRT) or volumetric modulated arc therapy (VMAT) to a total dose of 70 Gy with fractions of 2 Gy in 6-7 weeks, with or without chemotherapy or cetuximab. Detailed radiation protocols have been published previously [17,18].

Clinical parameters

Clinical parameters including age, gender, TNM-stage, clinical stage, treatment modality and WHO PS were collected from our prospective data registration program. TNM and clinical stage were defined according to the 7th edition of the American Joint Committee on Cancer Staging Manual [3]. Tumor site was included in the analysis, and tumors originating in the oropharynx were further stratified by HPV status as HPV-positive, HPV-negative and HPV-unknown (Table 1). HPV-status was assessed by p16

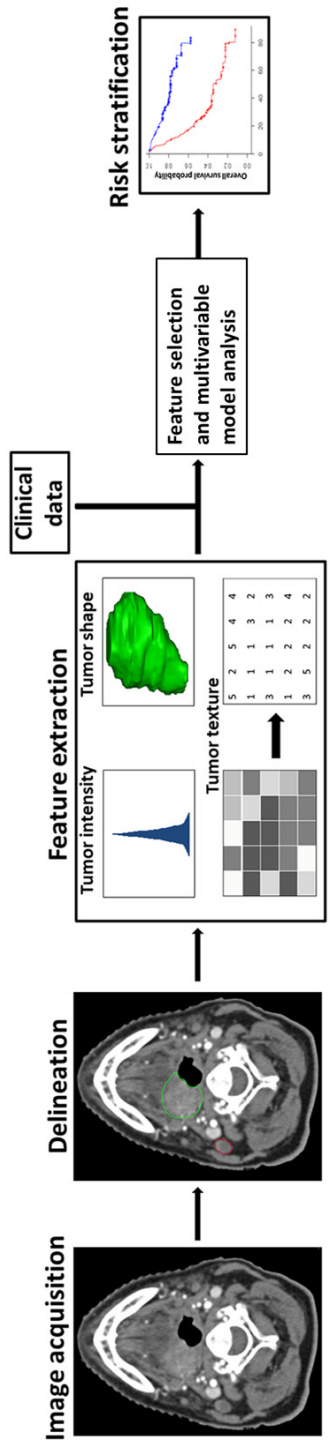


Fig. 1. The image-biomarkers (IBMs) extraction and analysis workflow.

immunohistochemistry followed by DNA polymerase chain reaction in cases of p16-positivity in OPC patients. Tumor volume was included in the analysis as a geometric IBM, and not as a clinical parameter.

CT image-biomarkers

An overview of the IBM extraction process and analysis is shown in Fig. 1. The IBMs were extracted using in-house developed Matlab based software (version R2014a; Mathworks, Natick, USA). For a more detailed description of the IBMs, we refer to previous work [11] and Supplementary A. All IBMs were reported complying with th
REMARK guidelines [18] and IBM formulas are in line with the "Image biomarker standardisation initiative" [10].

CT intensity and geometric IBMs

The primary tumor (PT) and pathological lymph node (LN) were delineated on the planning CT-scans by experienced head and neck radiation oncologists. Thirty-six intensity and 40 geometric IBMs were extracted from both the PT and LN. All IBMs from LN were marked as LN_IBMs. The intensity IBMs were obtained from the histogram of the voxel intensities of the delineated structures, e.g. mean represents the average voxel intensity and the skewness quantifies the degree of asymmetry around the mean value. The geometric IBMs, such as volume, bounding-box-volume and major-axis-length, were extracted from the three-dimensional (3D) contoured structures. The LN_IBMs from patients without lymph node metastasis were defined as 0.

CT textural IBMs

Forty-four textural CT IBMs describing the radiological heterogeneity of the PT tissue were derived from three different matrices: the gray level co-occurrence matrix (GLCM) [20], the gray level run-length matrix (GLRLM) [21] and the gray level size-zone matrix (GLSZM) [22]. Those matrices provide a statistical view of image texture based on the relationship between neighbouring pixels. GLCM IBMs describe the number of voxel transitions of certain gray levels, e.g. the IBM correlation is larger in case of larger areas of similar gray levels. GLRLM IBMs assess the number of directional gray level repetition, e.g. low run length non-uniformity means consecutive voxels with the same gray level are distributed homogeneously. The GLCM and GLRLM IBMs were computed from each 3D directional matrix and averaged over 13 directions. GLSZM quantifies the volumetric

gray level repetition, e.g. small zone emphasis depends on the occurrence of small zone. A higher value indicates to the fine texture and a lower value corresponds to the coarse texture. GLSZM IBMs were computed from a 3D matrix.

Endpoints

The endpoints were LC, RC, DMFS and DFS. The events of LC and RC were defined as recurrent or residual disease within or adjacent to the primary site and regional nodes, respectively. The events of DMFS were defined as distant metastasis. Events for DFS is defined as any events mentioned above or death due to any causes. Time to event was defined as the date from the first day of radiotherapy until the date of the event. Patients without failures were censored at the date of last follow-up. Patients received systematic follow-up every 3 months in the first year following treatment and every 6 months thereafter.

Data analysis

Step 1: Clinical models

Clinical factors that were considered as candidate predictors included categorical variables: gender (female vs. male), T-stage (T3-T4 vs. T1-T2), N-stage (N2-N3 vs. N0-N1), clinical stage (IV vs. I-III), treatment modality (radiotherapy with systemic treatment vs. radiotherapy only), WHO PS (1-3 vs. 0), tumor site combined with HPV status (nasopharynx vs. larynx vs. HPV-positive oropharynx vs. hypopharynx vs. oral cavity vs. HPV-negative oropharynx), and one continuous variable: age.

Univariable analysis was performed to assess risk factors for LC, RC, DMFS and DFS. In the training cohort, all factors were included in a multivariable Cox proportional hazard regression analysis (forward selection based on Likelihood ratio test, $p < 0.05$) to create multivariable clinical models. The entire process was repeated in 1000 bootstrap samples and only the most frequently selected variables were considered in the final clinical model.

Step 2: IBM models

Twenty patients from the training cohort were used for the evaluation of the inter- and intra-observer reproducibility. For each patient, the IBMs were extracted from two delineations by two radiation oncologists and two delineations within 6 months by one radiation oncologist. The interclass correlation coefficient (ICC) of IBMs larger than 0.70

were considered to be robust for delineation variation, and were included in the further analysis.

To reduce the probability of overfitting and multicollinearity, pre-selection was performed for IBMs. All IBMs were analyzed as continuous variables, if the Spearman rank correlation (ρ) between pairs of IBMs was > 0.80, then the IBM with the lower univariable association with the endpoint was excluded from further analysis [23,24]. After pre-selection, multivariable Cox proportional hazard regression analysis was used to develop multivariable IBM models.

The entire process of pre-selection and model development was repeated in 1000 bootstrap samples, and only the most frequently selected variables were included in the final IBM models. The same methodology was used for combined models described below.

Step 3: Combined models

All clinical factors and IBMs were included in the multivariable analysis to create combined models. The same bootstrapping methodology including pre-selection and model development described in step 2 was used to build final combined models.

Step 4: Model performance in the training and validation cohorts

The concordance index (c-index) was determined to assess the models' discriminative power and the z-score test was used to test the significance of c-index differences. Internal validation (bootstrapping) was used for the variable selection, and accordingly correction for optimism of coefficients and c-indexes according to the TRIPOD statement [25].

The performance of the corrected final clinical, IBM and combined models were then tested in the validation cohort. Patients in the validation cohort were stratified into two risk groups based on the models: a low-risk group with hazard values ≤ the median and a high-risk group with hazard values > the median. Kaplan-Meier curves were generated to analyze LC, RC, DMFS and DFS rates for the low- and high-risk groups and Log-rank tests were used to compare the differences.

The chi-square test was used to compare the categorical variables and an independent sample t-test was used to compare normally distributed variables between different groups. Two tailed p-values <0.05 were considered statistically significant. Statistical

analysis was performed using the R software (version 3.2.1). The R-package survival (version 2.41-3) was used for modeling.

Results

For the training cohort, the median follow-up times for LC, RC, DMFS and DFS were 48.2, 49.7, 50.3 and 45.2 months, respectively. Overall, 56 (23%) local recurrences, 38 (16%) regional recurrences, 34 (14%) distant-metastases and 137 (57%) events occurred for LC, RC, DMFS and DFS. For the validation cohort, the median follow-up times for LC,

Table 1 Characteristics of the head and neck squamous cell carcinoma patients in the training and validation cohorts

Characteristics	Training Cohort		Validation Cohort		p-Value
	n=240	%	n=204	%	
Age at diagnosis (mean ± SD, years)	62 ± 10		63 ± 9		0.351[b]
Gender					0.779[c]
Male	176	73.3	152	74.5	
Female	64	26.7	52	25.5	
T-stage[a]					0.839[c]
T1	27	11.3	25	12.3	
T2	70	29.2	52	25.5	
T3	73	30.4	67	32.8	
T4	70	29.2	60	29.4	
N-stage[a]					0.708[c]
N0	85	35.4	78	38.2	
N1	28	11.7	21	10.3	
N2	117	48.8	100	49.0	
N3	10	4.2	5	2.5	
Clinical stage[a]					0.451[c]
I	16	6.7	12	5.9	
II	41	17.1	27	13.2	
III	46	19.2	50	24.5	
IV	137	57.1	115	56.4	

Table 1 Characteristics of the head and neck squamous cell carcinoma patients in the training and validation cohorts-continued

Characteristics	Training Cohort		Validation Cohort		p-Value
	n=240	%	n=204	%	
Treatment modality					
RT only	123	51.3	114	55.9	
RT with systemic treatment	117	48.8	90	44.1	
WHO PS					0.913[c]
0	163	67.9	136	66.7	
1	65	27.1	56	27.5	
2	10	4.2	9	4.4	
3	2	0.8	3	1.5	
Tumor site (with HPV status)					0.078[c]
Oral cavity	14	5.8	13	6.4	
HPV-positive oropharynx	25	10.4	29	14.2	
HPV-negative oropharynx	57	23.8	50	24.5	
HPV-unknown oropharynx	3	1.3	8	3.9	
Nasopharynx	4	1.7	7	3.4	
Hypopharynx	37	15.4	16	7.8	
Larynx	100	41.7	81	39.7	

Abbreviations: T = tumor; N = lymph node; RT = radiotherapy; WHO PS = World Health Organization performance status; HPV = human papilloma virus status.
a According to the 7th edition of the AJCC/UICC staging system
b p-Value was calculated using the independent sample t-test
c p-Value was calculated using the chi-square test

RC, DMFS and DFS were 21.0, 21.1, 22.5 and 20.3 months. In total 35 (17%), 23 (11%), 20 (10%) and 68 (33%) events were observed for LC, RC, DMFS and DFS. The clinical characteristics of the training and validation cohorts in this study are listed in Table 1. No significant differences between the two datasets were found regarding the baseline characteristics.

Step 1: Clinical models

Univariable analysis showed weak associations between tumor site (combined with HPV status) and endpoints. Therefore, the variable tumor site was analysed as a composite variable by combining different tumor sites. Similar associations with the endpoints were found for nasopharyngeal cancer, HPV-positive OPC and laryngeal cancer in this cohort, which is in line with the results of other studies [26-28]. Hence, tumor sites with

Table 2 Estimated coefficients (β) of clinical, IBM and combined models.

	Clinical models					IBM models					Combined models				
	β	Corrected β	HR	HR (95% CI)	p-Value	β	Corrected β	HR	HR (95% CI)	p-Value	β	Corrected β	HR	HR (95% CI)	p-Value
Local control (*LC*)															
Composite tumor site[a]	0.631	0.482	1.880	1.089-3.244	0.023										
WHO PS (1-3 vs. 0)	0.635	0.485	1.887	1.101-3.236	0.021						0.566	0.423	1.761	1.012-3.064	0.045
Correlation_GLCM*						3.843	2.713	46.640	3.858-564.0	0.003	3.100	2.316	22.206	1.735-284.2	0.017
Regional control (*RC*)															
WHO PS (1-3 vs. 0)	0.971	0.755	2.642	1.387-5.031	0.003										
Clinical stage (IV vs. I-III)	1.419	1.104	4.134	1.719-9.944	0.002										
Bounding-box-volume* (cm3)						0.003	0.003	1.003	1.002-1.004	<0.001	0.003	0.002	1.003	1.002-1.004	<0.001
LN_major-axis-length* (cm)						0.157	0.131	1.170	1.069-1.281	<0.001	0.157	0.124	1.170	1.069-1.281	<0.001
Distant metastasis-free survival (*DMFS*)															
Composite tumor site[a]	0.845	0.662	2.328	1.083-5.005	0.030						1.011	0.648	2.75	1.307-5.786	0.008
N-stage (N2-N3 vs. N0-N1)	1.366	1.071	3.920	1.572-9.774	0.003										
Bounding-box-volume* (cm3)						0.002	0.001	1.002	1.000-1.003	0.010					
LN_major-axis-length* (cm)						0.149	0.092	1.161	1.054-1.278	0.002	0.157	0.101	1.170	1.063-1.289	0.001
Disease-free survival (*DFS*)															
Composite tumor site[a]	0.807	0.713	2.241	1.582-3.173	<0.001						0.554	0.388	1.740	1.182-2.560	0.005
WHO PS (1-3 vs. 0)	0.528	0.466	1.695	1.198-2.398	0.003						0.377	0.264	1.458	1.016-2.093	0.041
Age	0.019	0.017	1.019	1.002-1.036	0.031						0.020	0.001	1.020	1.003-1.037	0.021
Bounding-box-volume* (cm3)						0.002	0.002	1.002	1.001-1.003	<0.001	0.002	0.001	1.002	1.001-1.003	<0.001
Correlation_GLCM*						2.153	1.992	8.613	1.570-47.26	0.013					

Abbreviations: IBM = image-biomarker; WHO PS = World Health Organization performance status; N = nodal; LN = lymph node; GLCM = grey level co-occurrence matrix ; HR = Hazard ratio; CI = confidence interval.

a hypopharynx, oral cavity and HPV-negative oropharynx vs. nasopharynx, larynx and HPV-positive oropharynx. *Image-biomarkers

worse treatment outcomes were grouped together to compare with the group of tumor sites which had favorable treatment outcomes (composite tumor site: hypopharynx, oral cavity and HPV-negative oropharynx vs. nasopharynx, larynx and HPV-positive oropharynx) [2,4,7,26-28]. Using this stratification, the composite tumor site was significantly associated with LC, RC, DMFS and DFS ($p = 0.006, 0.002, 0.001, <0.001$).

A number of other clinical parameters showed significant associations with outcomes in the univariable analysis. They are shown in Supplementary B.

In the multivariable clinical analysis (Table 2), composite tumor site was identified as an independent significant prognostic feature for LC, DMFS and DFS. Next to composite tumor site, WHO PS was associated with LC, and clinical stage and WHO PS were associated with RC. N-stage was a significant prognostic factor for DMFS, while WHO PS and age were associated with DFS.

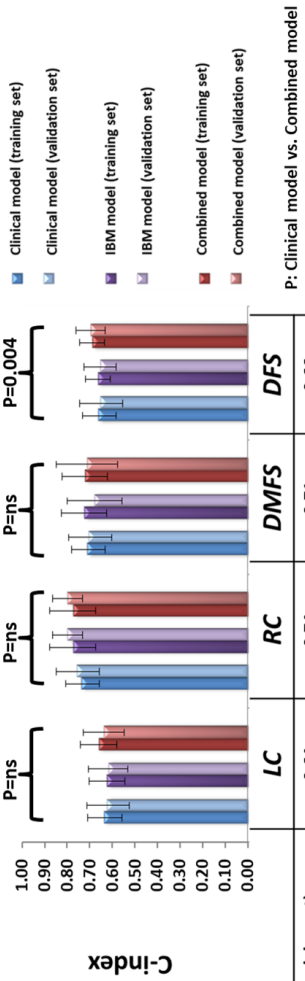
Step 2: IBM models

The average of the inter- and intra-observer agreement of all IBMs was 0.90 and 0.88, showing the stability of contouring was reasonably good. The ICC value for the inter-observer and intra-observer agreement was higher than 0.7 for 91% and 89% of the radiomic features and only the IBMs with an ICC > 0.7 were included in the further analysis.

According to the variable selection frequency plot (Supplementary C), the following IBMs were most frequently selected and significantly associated with the endpoints in the multivariable analyses for LC: correlation of GLCM; for RC and DMFS: bounding-box-volume and LN_major-axis-length, and for DFS: bounding-box-volume and correlation of GLCM (Table 2).

Step 3: Combined models

The coefficients and variables of the final combined models are depicted in Table 2. Composite tumor site and correlation of GLCM were selected with a comparable frequency as independent prognostic factors for the combined LC and DFS models (Frequency plot in Supplementary C). Correlation of GLCM was found to be significantly associated with composite tumor site ($p < 0.001$, logistic regression analysis), therefore any one of these could be included in the model, and performed similarly. For this study, only the feature with the larger frequency was included in the combined model.



P: Clinical model vs. Combined model	*LC*	*RC*	*DMFS*	*DFS*
Clinical model (training set)	0.64	0.74	0.71	0.66
Clinical model (validation set)	0.62	0.76	0.70	0.66
IBM model (training set)	0.62	0.78	0.73	0.66
IBM model (validation set)	0.62	0.80	0.68	0.65
Combined model (training set)	0.66	0.78	0.72	0.69
Combined model (validation set)	0.64	0.80	0.71	0.70

Fig. 2. Prediction performance of clinical, IBM and combined models.
Abbreviation: LC: local control; RC: regional control; DMFS: distant metastasis-free survival; DFS: disease-free survival; IBM: image-biomarker.

Therefore, correlation of GLCM was included in the LC-model and composite tumor site in the DFS-model. No clinical variables were selected into the combined RC model. The combined RC and DMFS models containing bounding-box-volume and LN_major-axis-length showed better performance than models containing clinical stage and N-stage for RC and DMFS.

Step 4: Model performance and external validation

The performances of the clinical, IBM and combined models in the training and validation

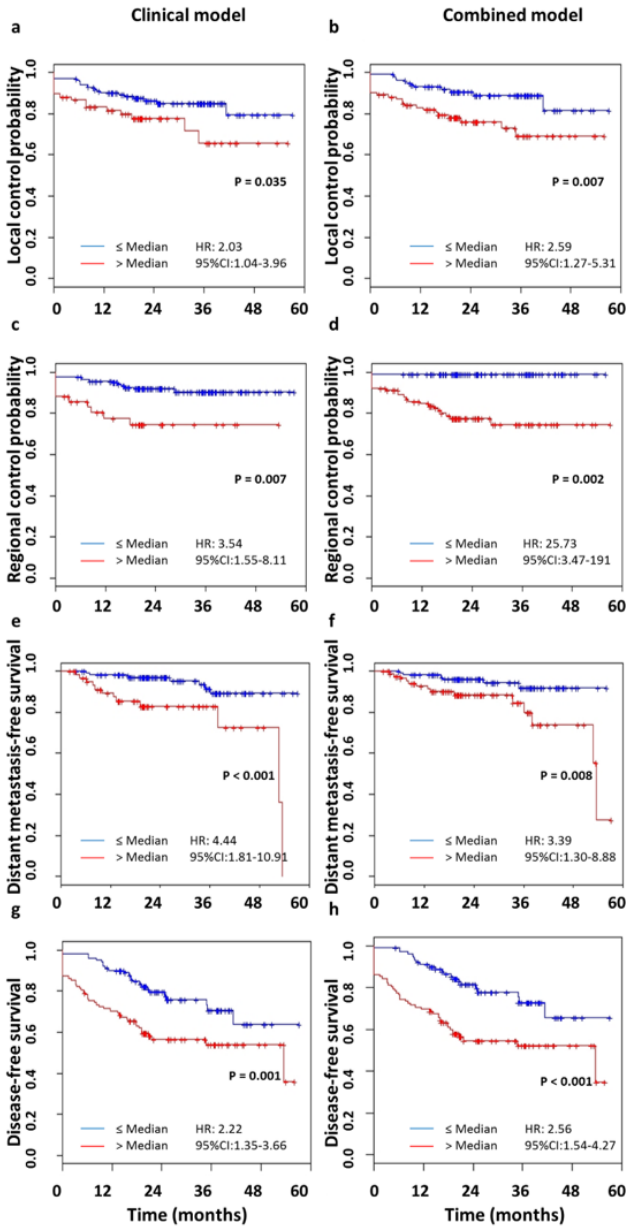


Fig. 3. Kaplan-Meier curves of high (hazard values > median) and low (hazard values ≤ median) risk groups stratified by clinical and combined models. When the patients are stratified with the combined models (b,d,f,h), the LC, RC, DMFS and DFS curves separation and hazard ratio (> median vs. ≤ median) between different risk groups are larger than or similar as clinical models (a,c,e,g).
Abbreviation: LC: local control; RC: regional control; DMFS: distant metastasis-free survival; DFS: disease-free survival; HR = hazard ratio; CI = confidence interval

cohorts are outlined in Fig. 2. The clinical models' prediction performances in the training cohort were as follows, with the 95% confidence-interval (CI)), LC: 0.64 (0.56-0.71), RC: 0.74 (0.64-0.83), DMFS: 0.71 (0.62-0.81) and DFS: 0.66 (0.61-0.72). The performances of IBM models were comparable to that of the clinical models. The combined models, performed as good as or significantly better than the clinical models, with c-indexes of 0.66 (0.58-0.74, z-score test: p = 0.23) for LC, 0.78 (0.67-0.88, p = 0.13) for RC, 0.72 (0.62-0.82, p = 0.36) for DMFS and 0.69 (0.64-0.74, p = 0.004) for DFS in the training cohort and performed well in the validation cohort with c-indexes of 0.64 (0.55-0.73), 0.80 (0.73-0.87), 0.71 (0.58-0.85) and 0.70 (0.63-0.76) for LC, RC, DMFS and DFS, respectively.

Fig. 3. shows the Kaplan-Meier curves for low- and high-risk groups in the validation cohort, stratified according to their median hazard value. When patients were stratified with the combined models (Fig. 3b,d,f,h), the differences between the curves were larger than with all clinical models (Fig. 3a,c,e,g) except for DMFS. The actual probabilities of LC, RC, DMFS and DFS at 2-year in low- and high-risk groups are shown in Supplementary D. Using combined models resulted in a more distinct risk group classification than using models based on clinical variables only for LC, RC and DFS. For example, the 2-year LC difference between the low- and high-risk groups improved from 8.6 % with the clinical model to 14.3 % with the combined model, and the hazard ratio improved from 2.0 to 2.6.

Discussion

This study showed a detailed analysis on the different patterns of failure by evaluating LC, RC, DMFS and DFS for HNC patients primarily treated with radiotherapy. Firstly, all clinical variables were explored thoroughly to build optimal clinical models for comparison [29]. Secondly, the quantitative IBMs were included in the models and provided similar information as qualitative clinical variables. Finally, combined models showed slightly better performance or as good performance as clinical and IBM models in this study.

Clinical models were developed based on clinical parameters alone. HPV status, as a confirmed significant prognostic factor in OPC patients, was combined with tumor site in the analysis [4,27,28] (Table 1). However, no strong associations between tumor site and endpoints were found in the univariable analysis. In order to obtain the optimal

clinical models, the composite tumor sites with similar associations with endpoints were grouped together, and showed significant associations with LC, RC, DMFS and DFS [2,4,7,26]. The other clinical prognostic factors (WHO PS, clinical stage, N-stage and age) in our study are in line with those found by other investigators [27-33].

Several IBMs were identified as independent prognostic factors in the final IBM models, two geometric and one textural. These included: bounding-box-volume, the LN_major_axis_length and correlation of GLCM. With only one or two IBMs, the IBM models performed as well as the clinical models.

The bounding-box-volume refers to the volume of the smallest cube that encloses all pixels of the contoured tumor. Generally, a larger bounding-box-volume indicates a more invasive, irregular-shaped and larger tumor, and was indeed associated with worse RC and DFS in our study. IBM models with bounding-box-volume as the only variable could achieve c-indexes of 0.77 and 0.66 in predicting RC and DFS, which were already comparable with the performances of clinical models. When bounding-box-volume was added to clinical models, it performance better than clinical stage in modelling RC and significantly improved the DFS model (p=0.004, Table 2).

Since tumor volume was a prognostic factor for overall survival for HNSCC patients with advanced stage, tumor volume was also included in the analysis as a geometric IBM [5,8,34,35]. Tumor volume was a significant prognostic factor in the univariable analysis for RC, DMFS and DFS. It was strongly correlated with bounding-box-volume ($\rho = 0.95$), but performed slightly worse than bounding-box-volume in predicting RC, DMFS and DFS in this study. The difference between the bounding-box-volume and tumor volume is that bounding-box-volume includes information on both tumor volume and shape (Fig. 4a and b). The tumor volumes in Fig. 4a and Fig. 4b were similar, but the bounding-box-volume of Fig. 4b had an irregular shape, and was twice as large as that of Fig. 4a. Tumor volume depends on the doubling time of tumor cells, while tumor shape is caused by invasive growth patterns. Our result suggests that the bounding-box-volume is more relevant for the RC and DFS than tumor volume.

LN_major-axis-length describes the largest distance between any two voxels of the positive lymph node(s) (Fig. 4c and d), which is the most selected IBM of RC and DMFS IBM models (Supplementary C). When the patient has only one lymph node, LN_major-

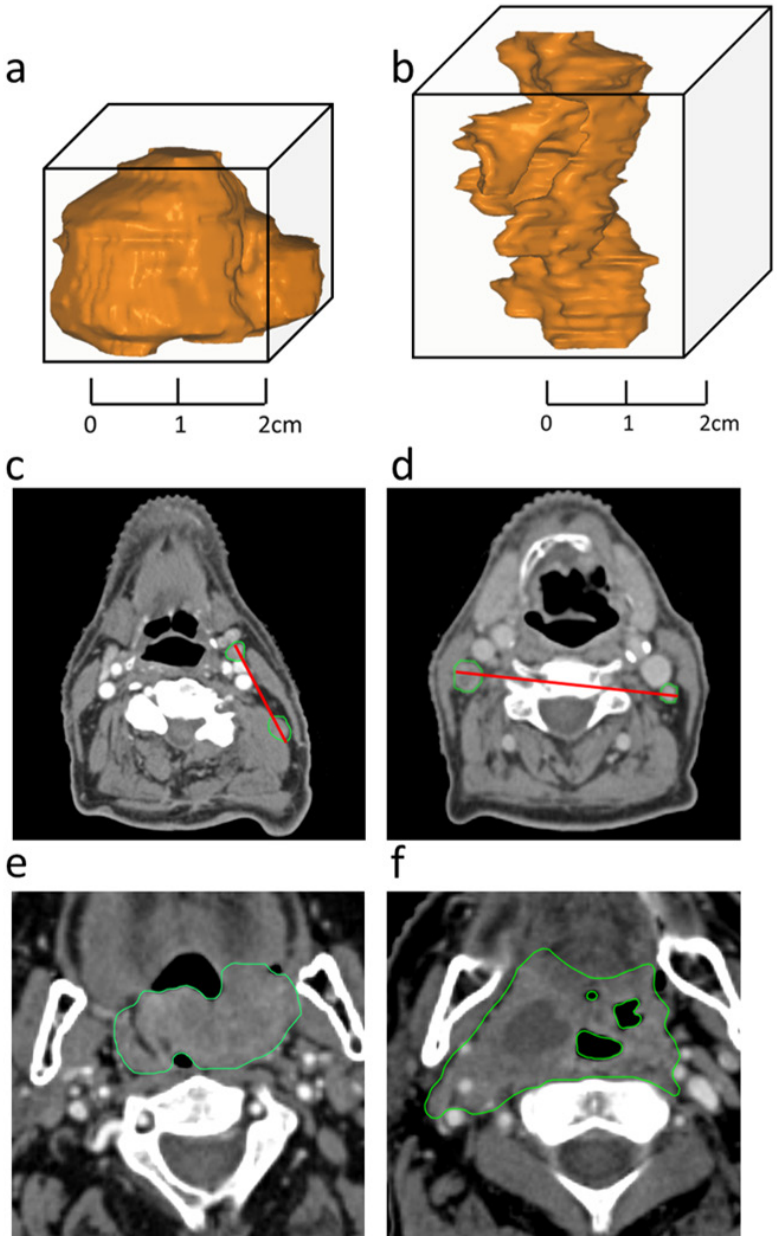


Fig. 4. Examples of patients with low (a,c,e) and high (b,d,f) values of bounding-box-volume, LN_major-axis-length and correlation of GLCM.
Abbreviation: GLCM: gray level co-occurrence matrix.

axis-length represents the size of the LN. When the patient has more than 1 lymph node, it means the largest distance between two distant lymph nodes. A larger LN_major-axis length is not only related to the size of the pathological lymph nodes but also to the distance between them, which might indicate more aggressive behavior of the tumor cells. LN_major-axis-length showed a spearman rank correlation of 0.80 with N-stage. However, LN_major-axis-length performed better than N-stage in the prediction of RC and DMFS, and N-stage did not add prognostic information to the model with LN_major-axis-length. This observation was supported by our previous study on overall survival, in which LN_major-axis-length had a stronger association with overall survival than N-stage [17]. The LN_major-axis length reflects not only N-stage information (size and location) but also patterns of growth. Therefore, it is advantageous to use LN_major-axis length as a surrogate for N-stage.

Correlation was selected as a prognostic IBM for LC and DFS. A lower correlation value indicates a higher radiological homogeneity, which was associated with improved LC and DFS in our study (Fig. 4e and f). This observation is supported by the findings of Haralick et al., showing that correlation was suitable for distinguishing heterogeneous and homogeneous materials [17,20,36]. There are other IBMs which are also used to quantify the intra-tumor heterogeneity and homogeneity on a millimeter scale, such as run length non-uniformity and gray level non-uniformity. These IBMs demonstrate prognostic performance for overall survival in HNC [9,17]. The intra-tumor heterogeneity is caused by multiple coexisting sub-clonal populations in the tumor [36,37]. The association between cellular information and IBMs may explain the performance of IBMs in survival analysis. Further research is necessary to investigate the underlying mechanisms and to identify IBMs representing tumor radiological heterogeneity for different endpoints. Our study found that the correlation of GLCM was significantly associated with the composite tumor site ($p < 0.001$). This association could be explained by the higher homogeneous tissue density in HPV-positive OPC, laryngeal and nasopharyngeal cancers, compared with other tumor sites. This hypothesis should be explored and these models might be improved with the addition of subgroup analysis in different head and neck tumor locations.

The advantage of using IBMs is that they provide quantitative and objective information

compared with traditional medical image analysis. Furthermore, on the basis of non-invasive medical imaging, IBMs could be used to assess the characteristics of tumor tissue. In this study, it was shown that CT IBMs performed as well as clinical features in predicting treatment outcomes. It is expected that IBMs extracted from more advanced imaging techniques would provide more tumor-specific information. Therefore, more studies on IBMs from advanced imaging techniques is recommended, to improve risk stratification for HNSCC. However, the variations in imaging protocols, segmentation, feature selection and modelling between different institutes may reduce the reproducibility, robustness and clinical utility. Standardization of the extraction of IBMs and report guidelines have been proposed and must be followed [29]. Furthermore, validation of IBMs on large external datasets is needed to ensure widespread acceptance.

The prediction models we described can be used to identify patients with a high risk of recurrence and metastasis prior to definitive (chemo-)radiation. Close imaging follow up for high risk patients after treatment can be suggested in cases when salvage surgery is applicable. Furthermore, dose intensification and more aggressive adjuvant treatment are options for high risk patients. These options should be investigated in order to guide future personalized strategies aimed at improving treatment outcome. Whether this model is applicable to HNC patients treated with surgery is not discussed in this paper.

Conclusion

Models containing quantitative image-biomarkers describing the volume, irregular shape and radiological heterogeneity of the tumor and the distance between lymph nodes performed as good as clinical variables in predicting treatment outcomes for HNC patients. These image-biomarkers are worth exploring in future studies to determine whether they can improve the clinical approaches currently employed.

Online supplementary materials

https://doi.org/10.1016/j.oraloncology.2019.06.020

REFERENCES

1. Howlader N, Noone AM, Krapcho M, Miller D, Bishop K, Altekruse SF, et al. SEER cancer statistics review, 1975–2013, http://seer.cancer.gov/csr/1975_2013/. [based on November 2015 SEER].
2. Pagh A, Grau C, Overgaard J. Failure pattern and salvage treatment after radical treatment of head and neck cancer. Acta Oncol 2016;55(5):625-32.
3. Pfister DG, Spencer S, Brizel DM, Burtness B, Busse PM, Caudell JJ, et al. Head and neck cancers, version 1.2015. J Natl Compr Canc Netw 2015;13:847–55.
4. Ang KK, Harris J, Wheeler R, Weber R, Rosenthal DI, Nguyen-Tân PF, et al. Human papillomavirus and survival of patients with oropharyngeal cancer. N Engl J Med 2010;363:24-35.
5. Cadoni G, Giraldi L, Petrelli L, et al. Prognostic factors in head and neck cancer: a 10-year retrospective analysis in a single-institution in Italy. Acta Otorhinolaryngol Ital. 2017; 37(6):458-466.
6. Langius JAE, Bakker S, Rietveld DHF, Kruizenga HM, Langendijk JA, Weijs PJM, et al. Critical weight loss is a major prognostic indicator for disease-specific survival in patients with head and neck cancer receiving radiotherapy.Br J Cancer 2013; 109(5):1093-9.
7. Regueiro CA, Aragón G, Millán I, Valcárcel FJ, de la Torre A, Magallón R. Prognostic factors for local control, regional control and survival in oropharyngeal squamous cell carcinoma. Eur J Cancer. 1994;30A(14):2060-7.
8. Aerts HJ. The potential of radiomic-based phenotyping in precision medicine. JAMA Oncol. 2016; 2(12):1636-42.
9. Aerts HJ, Velazquez ER, Leijenaar RT, Parmar C, Grossmann P, Carvalho S, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat Commun 2014;5:4006
10. Zwanenburg A, Leger S, Vallières M, Löck S. Image biomarker standardisation initiative feature definitions. arXiv:1612.07003 2016.
11. van Dijk LV, Brouwer CL, van der Schaaf A, Burgerhof JGM, Beukinga RJ, Langendijk JA, et al. CT image biomarkers to improve patient-specific prediction of radiation induced xerostomia and sticky saliva. Radiother Oncol. 2017;122(2): 185-191.
12. Grove O, Berglund AE, Schabath MB, Aerts HJ, Dekker A, Wang H, 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, Tian J, Liang CS, Chen X, 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, Alagappan M, Shirato H, Chang DT, 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. Park H, Lim Y, Ko ES, Cho HH, Lee JE, Han BK, et al. Radiomics Signature on Magnetic Resonance Imaging: Association with Disease-Free Survival in Patients with Invasive Breast Cancer. Clin Cancer Res 2018 1;24(19):4705-14.
16. Elhalawani H, Kanwar A, Mohamed ASR, White A, Zafereo J, Wong A, et al. Investigation of radiomic signatures for local recurrence using primary tumor texture analysis in oropharyngeal head and neck cancer patients. Scientific Reports 2018;24;8(1):1524.
17. Zhai TT, van Dijk LV, Huang BT, Lin ZX, Ribeiro CO, Brouwer CL, et al. Improving the prediction of overall survival for head and neck cancer patients using image biomarkers in combination with clinical parameters. Radiother Oncol. 2017;124:256-262.
18. van der Laan HP, Christianen ME, Bijl HP, Schilstra C, Langendijk JA. The potential benefit of swallowing sparing intensity modulated radiotherapy to reduce swallowing dysfunction: An in silico planning comparative study. Radiother Oncol 2012;103:76-81.
19. McShane L, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM. Reporting recommendations for tumor MARKer prognostic studies (REMARK). Eur J Cancer 2005; 41, 1690-1696.
20. Haralick R, Shanmugan K, Dinstein I. Textural features for image classification. IEEE Trans Syst Man Cybern 1973;3:610–21.
21. Tang X. Texture information in run-length matrices. IEEE Trans Image Process 1998;7:1602–9.
22. Thibault G, Fertil B, Navarro C, Pereira S, Cau P, Levy N, et al. Texture indexes and gray level size zone matrix application to cell nuclei classification. Pattern Recognit Inf Process 2009:140–5.
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, Ioannidis JPA, Macaskill P, Steyerberg EW, 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. Baatenburg de Jong RJ, Hermans J, Molenaar J, Briaire JJ, le Cessie S. Prediction of survival in patients with head and neck cancer. Head Neck. 2001 Sep;23(9):718.

27. O'Sullivan B, Huang SH, Siu LL, Waldron J, Zhao H, Perez-Ordonez B, et al. Deintensification in candidate subgroups in human papillomavirus-related oropharyngeal cancer according to minimal risk of distant metastasis. J Clin Oncol. 2013;10;31(5):543-50.
28. Lassen P, Primdahl H, Johansen J, Kristensen CA, Andersen E, Andersen LJ, et al. Impact of HPV-associated p16-expression on radiotherapy outcome in advanced oropharynx and non-oropharynx cancer. Radiother Oncol. 2014;113(3):310-6.
29. Lambin P, Leijenaar RTH, Deist TM, Peerlings J, de Jong EEC, van Timmeren J, et al. Radiomics: the bridge between medical imaging and personalized medicine. Nat Rev Clin Oncol. 2017;14(12)749-762.
30. Wang JR, Habbous S, Espin–Garcia O, Chen D, Huang SH, Simpson C, et al. Comorbidity and performance status as independent prognostic factors in patients with head and neck squamous cell carcinoma. Head Neck. 2016;38: 736–742.
31. Hall SF, Groome PA, Irish J, O'Sullivan B. Towards further understanding of prognostic factors for head and neck cancer patients: The example of hypopharyngeal cancer. The Laryngoscope 2009;119: 696–702.
32. Rietbergen MM, Witte BI, Velazquez ER, Snijders PJ, Bloemena E, Speel EJ, et al. Different prognostic models for different patient populations: validation of a new prognostic model for patients with oropharyngeal cancer in Western Europe. Br J Cancer.2015;112:1733-6.
33. Vallières M, Kay-Rivest E, Perrin LJ, Liem X, Furstoss C, Aerts HJ, et al. Radiomics strategies for risk assessment of tumour failure in head-and-neck cance. Sci Rep. 2017;7:10117.
34. Chen LL, Nolan ME, Silverstein MJ, Mihm MC, Sober AJ, Tanabe KK, 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.
35. 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.
36. Gerlinger M, Rowan AJ, Horswell S, Math M, Larkin J, Endesfelder D, 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, Zhang B, Zhao D, Li Y, et al. Tumor evolution and intratumor heterogeneity of an oropharyngeal squamous cell carcinoma revealed by whole-genome sequencing. Neoplasia 2013;15:1371–8.