European Journal of Cancer 75 (2017) 222–230

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.ejcancer.com

Original Research

A PET-based nomogram for oropharyngeal cancers

J. Castelli [a,b,c], A. Depeursinge [d,e], V. Ndoh [f], J.O. Prior [g], M. Ozsahin [a], A. Devillers [h], H. Bouchaab [a], E. Chajon [f], R. de Crevoisier [f], N. Scher [a], F. Jegoux [i], B. Laguerre [j], B. De Bari [a], J. Bourhis [a,*]

[a] *Radiotherapy Department, Lausanne University Hospital, Switzerland*
[b] *INSERM, U1099, Rennes, F-35000, France*
[c] *Université de Rennes 1, LTSI, Rennes, F-35000, France*
[d] *Ecole Polytechnique Fédérale de Lausanne, CH-1015, Lausanne, VD, Switzerland*
[e] *University of Applied Sciences Western Switzerland, 3960, Sierre, Switzerland*
[f] *Radiotherapy Department, Centre Eugene Marquis, Rennes, F-35000, France*
[g] *Nuclear Medicine and Molecular Imaging Department, Lausanne University Hospital, Switzerland*
[h] *Nuclear Medicine Department, Centre Eugene Marquis, Rennes, F-35000, France*
[i] *Head and Neck Department, CHU Rennes, Rennes, F-35000, France*
[j] *Oncology Department, Centre Eugene Marquis, Rennes, F-35000, France*

Received 27 September 2016; received in revised form 28 November 2016; accepted 14 January 2017

KEYWORDS
Oropharyngeal cancer;
Nomogram;
Prognostic score;
PET

Abstract ***Purpose:*** In the context of locally advanced oropharyngeal cancer (LAOC) treated with definitive radiotherapy (RT) (combined with chemotherapy or cetuximab), the aims of this study were: (1) to identify PET-FDG parameters correlated with overall survival (OS) from a first cohort of patients; then (2) to compute a prognostic score; and (3) finally to validate this scoring system in a second independent cohort of patients.

Materials and methods: A total of 76 consecutive patients (training cohort from Rennes) treated with chemoradiotherapy or RT with cetuximab for LAOC were used to build a predictive model of locoregional control (LRC) and OS based on PET-FDG parameters. After internal calibration and validation of this model, a nomogram and a scoring system were developed and tested in a validation cohort of 46 consecutive patients treated with definitive RT for LAOC in Lausanne.

Results: In multivariate analysis, the metabolic tumour volume (MTV) of the primary tumour and the lymph nodes were independent predictive factors for LRC and OS. Internal calibration showed a very good adjustment between the predicted OS and the observed OS at 24 months. Using the predictive score, two risk groups were identified (median OS 42 versus 14 months, $p < 0.001$) and confirmed in the validation cohort from Lausanne (median OS not reached versus 26 months, $p = 0.008$).

* *Corresponding author*: Department of Radiation Oncology, Centre Hospitalier Universitaire Vaudois (CHUV), Bugnon 46, CH-1011, Lausanne, Switzerland. Fax: +41 21 314 46 01.
E-mail address: Jean.Bourhis@chuv.ch (J. Bourhis).

http://dx.doi.org/10.1016/j.ejca.2017.01.018

Conclusions: This is the first report of a PET-based nomogram in oropharyngeal cancer. Interestingly, it appeared stronger than the classical prognostic factors and was validated in independent cohorts markedly diverging in many aspects, which suggest that the observed signal was robust.

1. Introduction

Head and neck cancers (HNC) are among the most common cancers world wide (5th leading cancer by incidence) [1]. The American Joint Committee on Cancer (AJCC) staging, based on the primary tumour extension and nodal spread, is generally used to estimate the prognosis and guide therapy [2]. Based on evidence-based medicine level 1 [3], chemoradiotherapy (CRT) is a standard treatment for non-resected or unresectable locally advanced HNC (LAHNC) [4–6]. Radiotherapy (RT) combined with cetuximab has been established as a potential alternative standard treatment, especially useful when concomitant chemotherapy cannot be used [7]. Despite these treatments, the prognosis of these cancers remains relatively poor and locoregional recurrence can occur in up to 40% patients, mostly occurring in the first 2 years after the treatment [8], suggesting a need to better identify patients with a worse prognosis.

^{18}F-fluorodeoxyglucose (^{18}F-FDG) positron emission tomography/computed tomography (PET/CT) allows to quantify the metabolic activity of a tumour (glycolysis) and has become a reference tool in oncology for staging, RT planning and monitoring tumour response in many cancers [9,10]. PET imaging allows a more accurate nodal staging of LAHNC [11,12] and could result in changing the therapeutic management in nearly 15% of patients [13]. A PET/CT performed at 2–3 months after the end of RT ± chemotherapy allows the identification of good responders and can be useful for decision-making of neck dissection for residual neck disease [14]; however, most of the available studies were based on visual analysis.

The maximum standard uptake value (SUV_{max}) corresponds to the maximal pixel value in the tumour. Thanks to its ease of use and relative robustness, it is one of the most widely used parameters in clinical practice. However, SUV_{max} is not representative of non-homogeneous overall tumour uptake. More recently, volumetric PET parameters, i.e. metabolic tumour volume (MTV) and total lesion glycolysis (TLG), have been correlated with clinical outcome [15–17]. Nonetheless, these parameters require to delineate the tumour. As PET imaging suffers from a low spatial resolution, along with a high noise background and partial volume effect, tumour delineation heavily depends on the chosen segmentation method. One of the most common methods consists of using an automatic threshold, although the threshold value of 42% has been suggested in many studies. However, there are no consistent data for using a specific threshold to compute MTV. Another point is the reproducibility of PET parameters between different scanners and/or institutions, as most of the studies published so far were monocentric.

In this context, the aims of our study were: (1) to identify PET parameters correlated with overall survival (OS) from a first cohort of patients (from Rennes Cancer Center, France); then (2) to create a prognostic scoring system; and (3) finally to validate this scoring system with a second independent cohort of patients (from Lausanne University Hospital, Switzerland).

2. Material and methods

All consecutive patients from Rennes Cancer Center and Lausanne University Hospital treated with definitive concurrent CRT or RT and cetuximab for locally advanced oropharyngeal carcinoma (LAOC) between January 2010 and December 2015 were retrospectively reviewed. Inclusion criteria were an age between 18 and 75 years, T3–4 or N+ stage, no surgery before RT, no history of cancer, a PET performed at least 8 weeks before RT, no metastasis at diagnosis and a minimal follow-up of 6 months.

The study enrolled a total of 122 patients (76 from Rennes and 46 from Lausanne). The main patient, tumour and treatment characteristics are shown in Table 1. All tumours were locally advanced, corresponding to T3–4 or N stage (stage III or IV, AJCC 7th edition).

2.1. Treatment and planning

All patients underwent intensity-modulated RT (IMRT) using volumetric modulated arc therapy (VMAT, Rennes) or helical tomotherapy (Lausanne). A total dose of 70 Gy 2 Gy/fraction/day, 35 fractions (Rennes) or 2.12Gy/fractions/day, 33 fractions (Lausanne) with a simultaneous integrated boost technique [18] was given in combination to concomitant chemotherapy [5,6] or cetuximab [7] if the patients were not fit for chemotherapy. The modality of planning and treatment were the same as previously published [19]. The study was approved by the institutional ethical committees (NCT02469922).

Table 1
Patient characteristics.

Characteristics	Training set cohort—Rennes (N = 76)	External validation set cohort—Lausanne (N = 46)	p-value
Mean age, years (SD)	59.2 (8.6)	63.3 (±9.17)	0.027
Gender, N (%)			0.34
Male	61 (80.3%)	40 (87%)	
Female	15 (19.7%)	6 (13%)	
T-classification, N (%)			0.005
T1	1 (1.4%)	6 (13%)	
T2	20 (26.3%)	16 (34.9%)	
T3	34 (44.7%)	18 (39.1%)	
T4	21 (27.6%)	6 (13%)	
N-classification, N (%)			0.49
N0	11 (14.5%)	3 (6.5%)	
N1	11 (14.5%)	10 (21.7%)	
N2	51 (67.1%)	29 (63.1%)	
N3	3 (3.9%)	4 (8.7%)	
GTV, cm^3	45.8 (±47.7)	25.6 (±26.7)	<0.001
p16			0.001
Positive	15 (19.8%)	17 (37%)	
Negative/ unknown	21 (27.6%)/40 (52.6%)	15(32.6%)/14 (30.4%)	
Chemotherapy, N (%)			0.058
Cisplatin[5]	51 (67.1%)	24 (52.2%)	
Carboplatin − 5FU[6]	9 (11.8%)	4 (8.7%)	
Cetuximab[7]	16 (21.1%)	18 (39.1%)	

GTV = Gross Tumour Volume.

2.2. PET/CT acquisition

All patients underwent FDG PET/CT for staging before treatment. For the training cohort (Rennes), the patients fasted at least 4 h before the injection of 4 Mbq/kg of (^{18}F)-FDG (Flucis). Blood glucose levels were checked before the injection of (^{18}F)-FDG. If not contra-indicated, intravenous contrast agents were administered before CT scanning. After a 60-min uptake period of rest, patients were imaged with the Discovery PET/CT imaging-system (General Electric Medical Systems, Milwaukee, WI, USA). First, a CT (120 kV, 80 mA, 0.8-s rotation time, slice thickness 3.75 mm) was performed from the base of the skull to the mid-thigh. PET scanning was performed immediately after acquisition of the CT. Images were acquired from the base of skull to the mid-thigh (3 min/bed position). PET images were reconstructed by using an ordered-subset expectation maximisation iterative reconstruction (OSEM) (two iterations, 28 subsets) and an iterative fully 3D (Discovery ST). CT data were used for attenuation calculation. A similar protocol was used in Lausanne; however, on a slightly more recent system, Discovery D690 TOF PET/CT (General Electric Healthcare, Milwaukee, WI, USA), which allowed shorter acquisition (2 min/bed position). PET images were reconstructed after time-of-flight and point-spread-function recovery corrections.

2.3. PET analysis

For each patient, gross tumour volume-tumour (GTV-T) and nodal GTV (GTV-N) were manually segmented on each PET/CT by the same radiation oncologist.

A set of quantitative parameters based on SUV histograms were extracted from GTV-T and GTV-N in PET images. SUV_{Max} was first computed from GTV-T as the maximum SUV in the delineated volume. Various metabolic volumes were subsequently defined based on two segmentation methods: (i) an absolute threshold of SUV (ranging from 2.5 g/ml to 8 g/ml, 0.5 g/ml steps) or (ii) a relative threshold of SUV_{Max} (30%, 35%, 40–60 (2% steps), 65% and 70%). Six metabolic intensity parameters were computed using the two segmentation methods at each threshold for both GTV-T and GTV-N. The four statistical moments of the intensity distribution, i.e. SUV_{Mean}, $SUV_{Variance}$, $SUV_{Skewness}$, $SUV_{Kurtosis}$, were computed. The latter are based on the assumption that SUVs are following normal distributions within the metabolic volumes. MTV was computed as the metabolic volume of the segmented region in millilitres. TLG was computed as $SUV_{Mean} \times$ MTV of the corresponding thresholded region. SUV_{Peak} was computed from GTV-T only. The latter was defined as the mean SUV inside a sphere of 1.2 cm centred on the position of SUV_{Max}. The intersection of the sphere and the metabolic region was used when the sphere was not fully included in the metabolic volume.

2.4. Statistical analysis

OS was calculated from the first day of RT to the date of death from any cause. Patients alive at the time of analysis were censored at the date of last follow-up. Locoregional control (LRC) was calculated from the first day of RT to the date of first recurrence in primary tumour and/or lymph node. Follow-up was calculated using a reverse Kaplan–Meier estimation [20]. Both LRC and OS estimations were computed using the Kaplan–Meier method and two-sided log-rank test was used to compare the groups.

The analyses were performed as suggested in the TRIPOD statement [21].

In the first step, the analysis was performed only on the Rennes cohort. The association of the pretreatment parameters with LRC and OS was first assessed using univariate Cox analyses. Harrel's c-index was used to compare different models (c-index $\approx$ 0.5 $\rightarrow$ not predictive, c-index $\approx$ 1 $\rightarrow$ predictive) [22]. The c-index was used to determine the optimal SUV threshold giving the most predictive value for each PET parameter.

Factors with significance of p-value <0.1 and with highest c-index after univariate analyses were assessed

for multivariate Cox regression model using backward elimination. Variables were removed from the model if $p > 0.1$.

An internal validation on the patients from the training cohort (Rennes) was performed by the bootstrap method (1000 datasets constructed by random resampling with replacement from the original) [23]. This method was used to estimate the adjusted c-index and the 95% confidence interval (95% CI) of each parameter. Second, an internal calibration was performed to estimate the accuracy of the final model.

Based on this final model, a nomogram was built to estimate the individual OS probability at 18 and 24 months. β-Coefficient estimations from the final model were used to build a predictive score. Two prognostic risk groups were identified based on the estimated optimal cut-point by Hothorn and Lausen method [24]. Kaplan–Meier method was used to evaluate this score.

Finally, the multivariate Cox model and the prognostic scores were tested in the Lausanne validation cohort. Harrell's concordance index was used for the Cox model and Kaplan–Meier method for the prognostic score.

All analyses were performed using R software 3.2.4 (R Development CoreTeam; http://www.r-project.org).

2.5. Follow-up

A clinical evaluation was performed after RT every 3 months for the first 2 years and then every 6 months. Database was locked on 30th May 2016.

3. Results

3.1. Training cohort (Rennes)

For the training cohort (Rennes), the following parameters were associated with OS in univariate Cox analyses with a $p < 0.1$: N stage, GTV, MTV-N, MTV-T, SUVKurtosis_N, SUVKurtosis_T, SUVMean_N, SUVSkewness_N, SUVSkewness_T, SUVVariance_T, SUVVariance_N, TLG_N and TLG_T (Table 2).

In multivariate Cox analysis, the tumour MTV with a threshold of 35% (MTV_T_35) and the lymph node MTV with a threshold of 44% (MTV_N_44) were the two independent risk factors for OS (Table 3) ($p < 0.001$). The same parameters were also correlated with LRC ($p = 0.03$), with a hazard ratio of 1.01 and 1.043 for MTV_T_35 and MTV_N_44, respectively.

3.2. Internal validation and calibration of the final model for the training cohort

The c-index of the model was 0.69. After internal bootstrap validation, the adjusted c-index was estimated at 0.68. The 95% CI for the coefficient of the parameters

Table 2
Univariate cox analyses for overall survival in the training cohort (Rennes). For PET parameters, data are given only for absolute and relative thresholds with the highest c-index.

Parameters	HR [95% CI]	c-index	p
Gender	0.38 [0.13–1.08]	0.54	0.067
Chemotherapy (Platinum versus Cetuximab)	0.97 [0.67–1.42]	0.49	0.9
PS (0–1 versus 2)	1.56 [0.55–4.44]	0.52	0.39
Age	0.99 [0.95–1.04]	0.49	0.9
Tobacco	2.2 [0.68–7.43]	0.52	0.18
Alcohol	1.65 [0.73–3.79]	0.54	0.22
GTV (as continuous variable)	1.0 [0.99–1.01]	0.6	0.2
T-classification (T1–T2 versus T3–T4)	0.97 [0.47–2.02]	0.49	0.95
N-classification (N0–N1 versus N2–N3)	2.12 [0.95–4.69]	0.56	0.062
AJCC staging (stage III versus IV)	1.71 [0.71–4.11]	0.53	0.19
p16	0.3 [0.04–2.36]	0.53	0.17
SUV_{max}	0.98 [0.92–1.04]	0.52	0.57
MTV-N			
Absolute threshold (SUV = 4.5)	1.02 [1.006–1.03]	0.64	0.004
Relative threshold (SUV = 44%)	1.06 [1.03–1.09]	0.64	<0.001
MTV-T			
Absolute threshold (SUV = 2.5)	1 [0.99–1.02]	0.60	0.14
Relative threshold (SUV = 35%)	1.02 [1.002–1.04]	0.61	0.03
TLG N			
Absolute threshold (SUV = 2.5)	1.002 [1.001–1.004]	0.65	0.003
Relative threshold (SUV = 65%)	1.01 [1–1.01]	0.73	0.004
TLG T			
Absolute threshold (SUV = 2.5)	1 [0.99–1.001]	0.58	0.45
Relative threshold (SUV = 35%)	1 [0.99–1.002]	0.59	0.35
SUV Peak T			
Absolute threshold (SUV = 4.5)	0.98 [0.88–1.09]	0.53	0.8
Relative threshold (SUV = 56%)	0.97 [0.9–1.05]	0.58	0.57
SUV Mean N			
Absolute threshold (SUV = 2.5)	1.24 [1.03–1.49]	0.63	0.06
Relative threshold (SUV = 54%)	1.05 [0.99–1.11]	0.55	0.85
SUV Mean T			
Absolute threshold (SUV = 6.5)	0.98 [0.86–1.11]	0.57	0.79
Relative threshold (SUV = 35%)	0.96 [0.87–1.07]	0.54	0.53
SUV Kurtosis N			
Absolute threshold (SUV = 7)	1.32 [1.06–1.64]	0.58	0.01
Relative threshold (SUV = 60%)	1.27 [1.04–1.57]	0.74	0.02
SUV Kurtosis T			
Absolute threshold (SUV = 2.5)	0.72 [0.48–1.08]	0.57	0.11
Relative threshold (SUV = 70%)	1.07 [0.77–1.43]	0.58	0.6
SUV Skewness N			
Absolute threshold (SUV = 5.5)	0.97 [0.54–1.74]	0.6	0.9
Relative threshold (SUV = 58%)	2.5 [1.32–4.73]	0.65	0.03
SUV Skewness T			
Absolute threshold (SUV = 2.5)	0.35 [0.14–0.85]	0.61	0.02
Relative threshold (SUV = 30%)	0.67 [0.25–1.76]	0.55	0.42
SUV Variance N			
bsolute threshold (SUV = 2.5)	1.1 [1.03–1.18]	0.6	0.004
Relative threshold (SUV = 65%)	1.16 [1.02–1.33]	0.65	0.02
SUV Variance T			
Absolute threshold (SUV = 5.5)	0.97 [0.93–1.02]	0.56	0.24
Relative threshold (SUV = 46%)	0.94 [0.84–1.05]	0.57	0.28

HR = Hazard Ratio, CI = Confidence Interval, GTV = Gross Tumour Volume, SUV = Standard Uptake Value, MTV = Metabolic Tumour Volume, TLG = Total Lesion Glycolysis. Bold values refer to p-values <0.05.

Table 3
Parameters associated with overall survival in multivariate analysis in the training cohort (Rennes).

Parameters	Mean Value (SD) (in cm^3)	HR [95% CI] (per 1 cm^3)	p	95% CI (Bootstrap validation)
MTV_T_35	18 (±15.6)	1.021 [1.000–1.043]	0.052	[1.000–1.056]
MTV_N_44	4.52 (±9.7)	1.057 [1.028–1.087]	<0.001	[1.040–1.094]

HR = hazard ratio, CI = confidence interval, MTV_T_35 = metabolic tumour volume of the tumour computed with a relative threshold at 35% of SUV_{max}, MTV_N_44 = metabolic tumour volume of the lymph node computed with a relative threshold at 44% of SUV_{max}.

of the model are given in Table 3. Internal calibration showed a very good adjustment between the predicted and observed OS at 18 (Fig. 1) and 24 months (Figure E1).

3.3. Nomogram and prognostic score for the training cohort

Based on the final model, a nomogram was computed (Fig. 2). A prognostic score was calculated based on the β-parameter from the Cox model. A normalisation was applied to obtain a score ranging from 0 to 5. The estimated cut-point by Hothorn and Lausen method was 1.33 (Supplementary Figure E2). Based on this cut-off, two risk groups were identified. The median OS was 42 months (95% CI: 20–64) for the low-risk group versus 14 months (95% CI: 5–23) for the high-risk group ($p < 0.001$) (Fig. 3A). The same prognostic score was used to estimate the LRC. The median LRC was not reached for the low-risk group versus 10 months for the high-risk group ($p = 0.009$) (Fig. 3B).

3.4. Comparison between the training and the validation cohort

Median follow-up for the training cohort (Rennes) and validation cohort (Lausanne) were 38 (range, 2–80 months) and 23 months (range, 3–57 months), respectively ($p < 0.001$). The two populations differed notably concerning age (mean 59.2 versus 63.3 years [p = 0.02]), the tumour volume (GTV: 45.8 cm^3 versus 25.6 cm^3 [p < 0.001]) and p16 status (p16+: 18% versus 37%, [$p = 0.001$]) for Rennes and Lausanne, respectively. The use of cetuximab was slightly more frequent in the validation cohort ($p = 0.05$). In both cohorts, most of the patients were smokers (90–95%), with a performance status of 0 or 1.

At time of the analysis, 38 (50%) and six (13%) patients had died, while 26 (34.2%) and 8 (17.3%) patients had a locoregional recurrence for Rennes and Lausanne, respectively. The 2-year OS rate was 58% (95% CI: 46–70%) and 85% [74–99%] for Rennes and Lausanne, respectively ($p = 0.001$).

3.5. Evaluation of the final model and the prognostic score in the validation cohort (Lausanne)

The β-coefficients from the training model were applied to the validation cohort. The c-index was 0.76, higher than in the training cohort (0.69). The prognostic score was calculated for the validation cohort and the cut-off of 1.33 (obtained from the training cohort) was applied. The result confirmed the external validation of the model with a median OS not reached for the low-risk group versus 26 months (95% CI: 22–30) for the high-risk group ($p = 0.008$) (Fig. 3C). For the LRC, the same training model was applied to the validation cohort. The LRC at 18 months for the low-risk and the high-risk group were 96.1% and 63.1%, respectively ($p = 0.009$) (Fig. 3D).

4. Discussion

To our knowledge, this is the first study presenting a PET-based score allowing the prediction of the risk of death in LAOC patients. Even if there are other studies exploring the prognostic values of some PET parameters (SUV, MTV and TLG, etc.) in LAOC, none of them performed an external validation. We found that both MTV-T and MTV-N as continuous variables were major predictors of OS. Noteworthy, the classical clinical variables (T-classification, GTV [T and N], age, gender, etc) were less or not significantly predictive of patient outcome. Other established prognostic parameters (PS, AJCC staging and smoking status) were not significant in our study due to the lack of variability in the distribution of these parameters in both training and validation cohorts (more than 90% of smokers, all patients with PS 0 or 1). One limitation of our study was some missing data concerning p16 status which were available for only half of the patients, as p16 analysis was performed routinely in our centres only since 2013, and hence did not allow a full evaluation of this well-established prognostic parameter in our model. Prevalence of human papilloma virus (HPV) in oropharyngeal cancer was shown to be around 23% in Europe [25], being higher than in our training cohort (19%) and lower than in our validation cohort (37%). The p16 status was found to be highly predictive of treatment outcomes and survival in patients with oropharyngeal cancer [26]. However, smokers with p16-positive tumour seem to have a worse prognosis than those with p16-positive tumour without history of smoking. In our study, most of the patients were smokers which may explain the lack of significance of p16. However, despite the difference in the rate of p16 status (Table 1), the results of the external validation seem confirm that the good

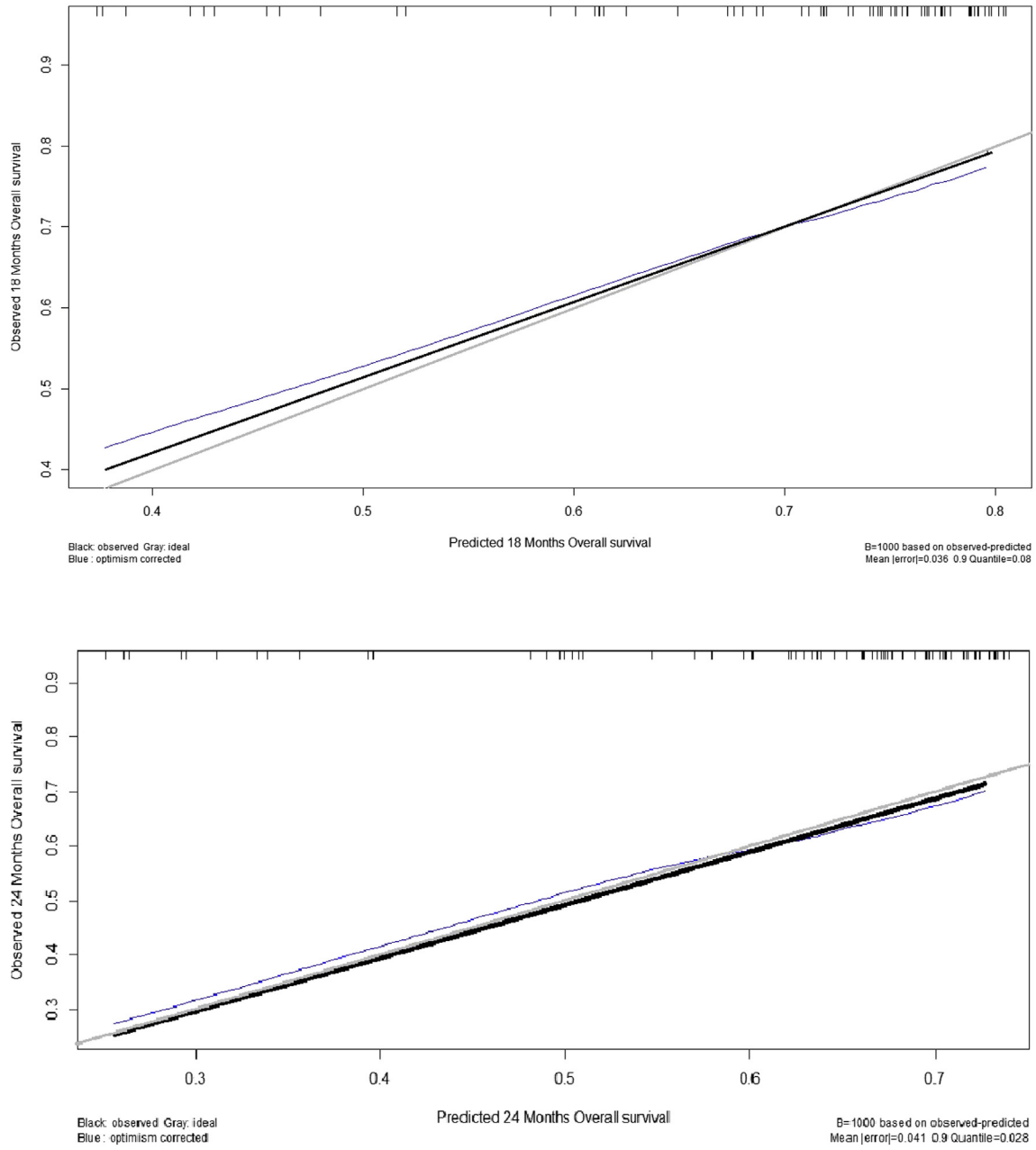

Fig. 1. Internal calibration of the final model for the training cohort (Rennes) at 18 and 24 months. Grey line is the ideal model, black line is the predicted survival and the blue-dotted line is the predicted survival corrected to avoid overfit. Both the 18- and 24-month survivals were nearly perfectly predicted. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

predictive value of our nomogram is not influenced by the p16 status.

The PET-based nomogram obtained from this score allowed the prediction of 18- and 24-month OS in this clinical setting. Three strengths of our study are noteworthy. First, we followed in our study the internationally accepted TRIPOD criteria to build our predicting factors. This is an important quality-assurance issue, reinforcing our results. Second, this is the first study showing the prognostic impact of MTV in an external validation-independent population of oropharyngeal cancer patients. Third, we used continuous parameters instead of dichotomised variables. Dichotomisation leads to loss of power, affects the ability to detect relationships and overestimates the magnitude of the effect.

PET volumetric parameters like MTV or TLG have been used to estimate the heterogeneity of the tumour FDG uptake. Limited data were available but it showed a higher predictive value of MTV compared with more classical parameters (TNM, SUV, GTV…) [27,28], which is consistent with our findings. Only two studies performed a validation on an independent dataset

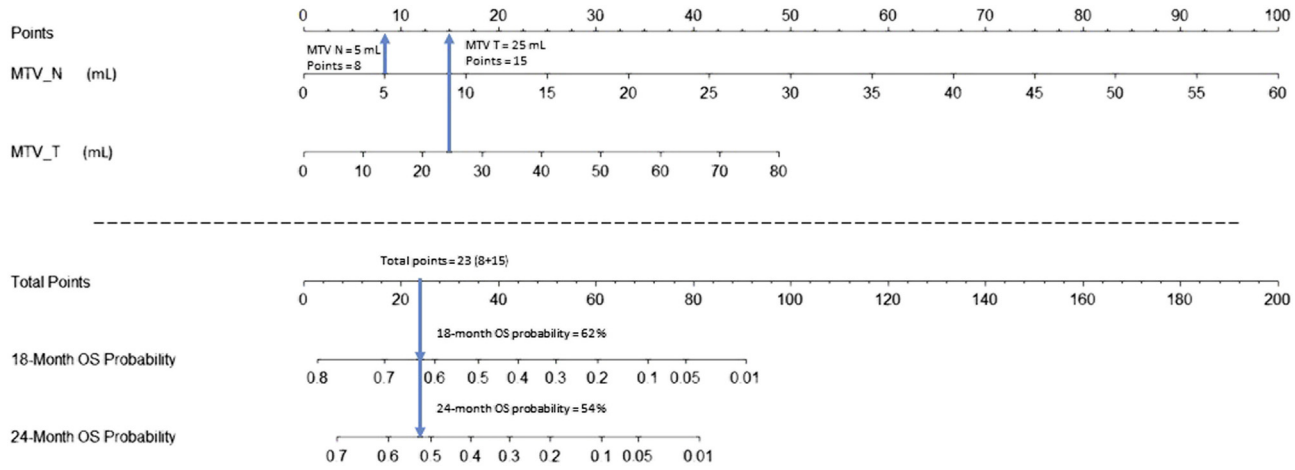

Fig. 2. Nomogram to predict the overall survival (OS) at 18 and 24 months. For each PET parameter, the corresponding points is obtained by drawing a line upward from the corresponding values to the 'Points' line. The total points for each patient is obtained by summing the points for each of the individual factors in the nomogram and is plotted on the 'Total points' line. A line is drawn down to read the corresponding predictions of 18- and 24-month LRC and OS. An example is given: an MTV-T of 25 mL corresponding to 15 points and an MTV-N of 5 mL to 8 points. The total score is 23, corresponding to 18-month and 24-month OS probabilities of 62% and 54%. MTV_N = metabolic tumour volume of the lymph node computed with a threshold = 44% of the SUV_{max}. MTV_T = metabolic tumour volume of the tumour computed with a threshold = 35% of the SUV_{max}.

[29,30]. La *et al.* [31] included 85 patients and showed that an increase of MTV of 17 cm^3 (from the 25th to the 75th percentile) was significantly correlated with an increased risk of death (hazard ratio [HR] 2.1). The authors validated their results on a dataset of 83 patients treated in the same institution after the original dataset [30]. Despite these interesting results, the possibility of adopting their model in the clinical practice is limited by the monocentric nature of their study and by the definition of the cut-off of MTV based on population-dependent characteristics (from the 25th to the 75th percentile). Based on the results of a previous study [32], Hofheinz *et al.* [29] used a cut-off of 58.7 ml for TLG. They showed a correlation of TLG only with better disease-free survival (DFS; HR 3.01, $p = 0.048$) but not with OS (HR 2.02, $p = 0.22$). After adjusting the cut-off at a value of 141 ml, TLG was also correlated with OS (HR 3.32, $p = 0.016$). Such methodologies and findings underline the difficulty in identifying a cut-off, which may be tested and reproducibly validated on an external dataset of patients. In our study, we choose to evaluate the MTV as continuous variables and to perform an external validation without modifying the model and the score obtained from the training dataset. Important differences in age, use of cetuximab, T-classification, p16 status, tumour and nodal volume (GTV) and PET/CT scanner between the training and the validation cohorts were observed. Despite these differences, the very good predictive performance obtained with the training cohort was confirmed (and even higher) for the validation cohort. These data strongly suggest that this new scoring system seem to be robust and could be further proposed and tested for patients' selection in clinical trials to identify patients with a high risk of locoregional failure and death, potentially candidates for treatment intensification, for instance, by dose escalation with dose painting in the MTV.

Noteworthy, the reproducibility of the MTV or TLG is considered to be limited by the initial definition of these parameters, which is based on a threshold of SUV, absolute (all pixel with SUV value > threshold) or relative (all pixel with SUV value > threshold % of SUV_{Max}). In the study by Schinagl *et al.* [33], four thresholds (2.5%, 40%, 50% and an adaptive threshold based on liver uptake) were compared for 77 patients treated by RT with or without chemotherapy. The authors found that 40% MTV was the strongest predictor of DFS and OS. However, also all the other thresholds were correlated with OS and DFS. Same results were reported on a population of 118 patients using three thresholds of MTV (2, 2.5 and 3) [34]. In our study, we evaluated 11 different absolute thresholds (from 2.5 to 8) and 15 relative thresholds (from 30% to 70%). All relative thresholds between 30% and 60% were correlated with OS, confirming the robustness of the MTV as a predictor factor, regardless the threshold chosen. However, based on p-value and c-index, the relative thresholds of 35% for the tumour and 44% for the lymph nodes were the best predictors of OS. The tumour MTV was nearly significant in the final model ($p = 0.052$). However, the final model (combination of both tumour and lymph node MTV) was highly significant ($p < 0.001$) and using tumour MTV increased the c-index of the model to 0.69 versus 0.64 for lymph node MTV alone. The combination of these two parameters probably takes into account the risk of death by local

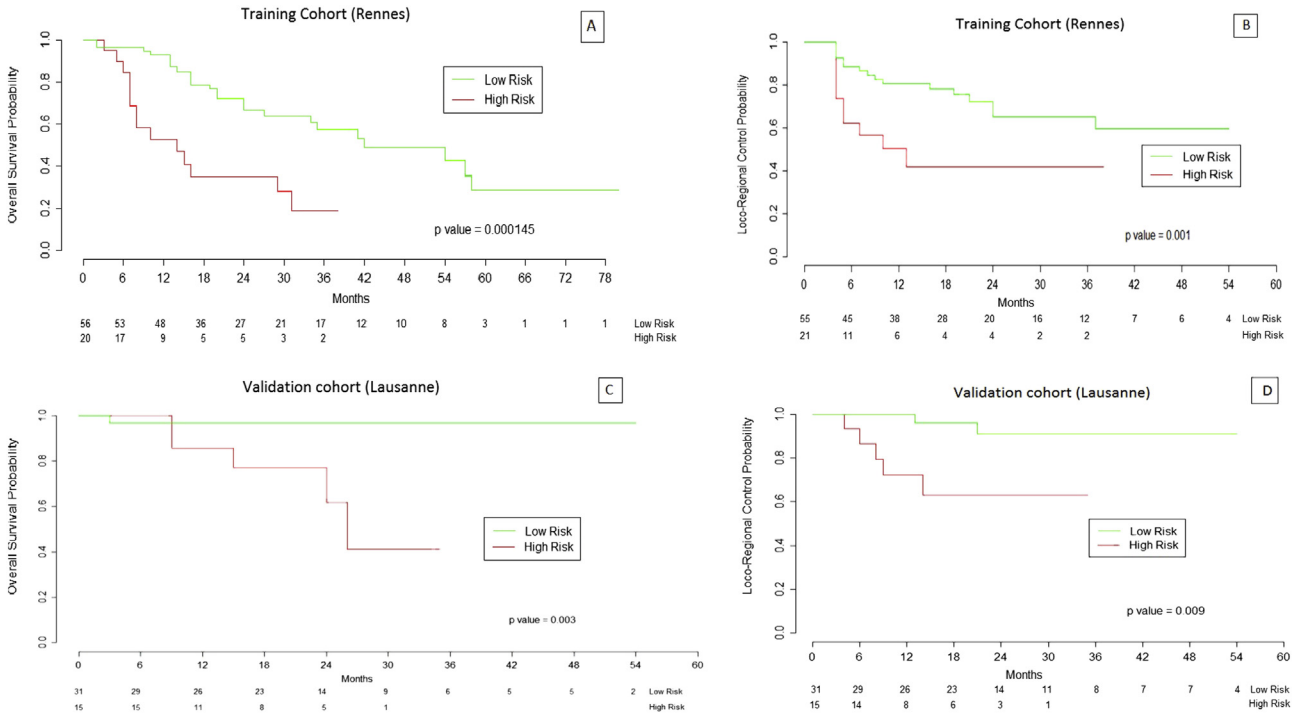

Fig. 3. Kaplan–Meier curves of overall survival and locoregional control for the training cohort (Rennes, A and B) and the validation cohort (Lausanne, C and D) according to the predictive score group (optimal cut-off defined by the Hothorn & Lausen method). High risk: score >1.33, low risk: score ≤1.33. Total score = 5.

relapse (primary tumour) and by metastasis (lymph node), both with a different weight in the final model.

5. Conclusion

MTV as a continuous variable was a strong prognostic factor for OS in LAOC patients treated with CRT or RT + cetuximab. We defined, and successfully validated on an independent dataset, a PET-based predictive score and nomogram that need to be further tested in larger prospective series to define their potential interest for tailoring the therapeutic approach.

Contributors

All authors contributed to the design and concept of the study. JC, EC, AD, MO, JB, HB, FJ, BL and J.O.P were responsible for patients' treatment and care. JC, AD, VN and NS acquired and analysed the data, with all authors contributing to data interpretation. JC and JB drafted and revised the manuscript for content, with all authors contributing to writing the text. All authors approved the final manuscript.

Funding sources

This work was partly supported by the Swiss National Science Foundation with grant agreement PZ00P2_154891 (A. Depeursinge).

Conflict of interest statement

None declared.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.ejca.2017.01.018.

References

[1] Parkin DM, Bray F, Ferlay J, Pisani P. Global cancer statistics, 2002. CA Cancer J Clin 2005;55(2):74–108.
[2] Edge SB, Compton CC. The American Joint Committee on Cancer: the 7th edition of the AJCC cancer staging manual and the future of TNM. Ann Surg Oncol 2010;17(6):1471–4.
[3] National Comprehensive Cancer Network. Head and neck cancers (Version 1. 2016). In.
[4] Pignon JP, Bourhis J, Domenge C, Designe L. Chemotherapy added to locoregional treatment for head and neck squamous-cell carcinoma: three meta-analyses of updated individual data. MACH-NC Collaborative Group. Meta-Analysis of Chemotherapy on Head and Neck Cancer. Lancet 2000;355(9208):949–55.
[5] Bernier J, Domenge C, Ozsahin M, Matuszewska K, Lefebvre JL, Greiner RH, et al. Postoperative irradiation with or without concomitant chemotherapy for locally advanced head and neck cancer. N Engl J Med 2004;350(19):1945–52.
[6] Bourhis J, Sire C, Graff P, Gregoire V, Maingon P, Calais G, et al. Concomitant chemoradiotherapy versus acceleration of radiotherapy with or without concomitant chemotherapy in locally advanced head and neck carcinoma (GORTEC 99-02): an

open-label phase 3 randomised trial. Lancet Oncol 2012;13(2):145–53.
[7] Bonner JA, Harari PM, Giralt J, Cohen RB, Jones CU, Sur RK, et al. Radiotherapy plus cetuximab for locoregionally advanced head and neck cancer: 5-year survival data from a phase 3 randomised trial, and relation between cetuximab-induced rash and survival. Lancet Oncol 2010;11(1):21–8.
[8] Chajon E, Lafond C, Louvel G, Castelli J, Williaume D, Henry O, et al. Salivary gland-sparing other than parotid-sparing in definitive head-and-neck intensity-modulated radiotherapy does not seem to jeopardize local control. Radiat Oncol 2013;8:132.
[9] Fletcher JW, Djulbegovic B, Soares HP, Siegel BA, Lowe VJ, Lyman GH, et al. Recommendations on the use of ^{18}F-FDG PET in oncology. J Nucl Med 2008;49(3):480–508.
[10] Gambhir SS, Czernin J, Schwimmer J, Silverman DH, Coleman RE, Phelps ME. A tabulated summary of the FDG PET literature. J Nucl Med 2001;42(5 Suppl.):1S–93S.
[11] Kyzas PA, Evangelou E, Denaxa-Kyza D, Ioannidis JP. ^{18}F-fluorodeoxyglucose positron emission tomography to evaluate cervical node metastases in patients with head and neck squamous cell carcinoma: a meta-analysis. J Natl Cancer Inst 2008;100(10):712–20.
[12] Yoo J, Henderson S, Walker-Dilks C. Evidence-based guideline recommendations on the use of positron emission tomography imaging in head and neck cancer. Clin Oncol (R Coll Radiol) 2013;25(4):e33–66.
[13] Lonneux M, Hamoir M, Reychler H, Maingon P, Duvillard C, Calais G, et al. Positron emission tomography with [^{18}F]fluorodeoxyglucose improves staging and patient management in patients with head and neck squamous cell carcinoma: a multicenter prospective study. J Clin Oncol 2010;28(7):1190–5.
[14] Castaldi P, Rufini V, Bussu F, Micciche F, Dinapoli N, Autorino R, et al. Can "early" and "late" ^{18}F-FDG PET-CT be used as prognostic factors for the clinical outcome of patients with locally advanced head and neck cancer treated with radiochemotherapy? Radiother Oncol 2012;103(1):63–8.
[15] Schwartz DL, Harris J, Yao M, Rosenthal DI, Opanowski A, Levering A, et al. Metabolic tumor volume as a prognostic imaging-based biomarker for head-and-neck cancer: pilot results from Radiation Therapy Oncology Group protocol 0522. Int J Radiat Oncol Biol Phys 2015;91(4):721–9.
[16] Moon SH, Choi JY, Lee HJ, Son YI, Baek CH, Ahn YC, et al. Prognostic value of volume-based positron emission tomography/computed tomography in patients with nasopharyngeal carcinoma treated with concurrent chemoradiotherapy. Clin Exp Otorhinolaryngol 2015;8(2):142–8.
[17] Cacicedo J, Navarro A, Del Hoyo O, Gomez-Iturriaga A, Alongi F, Medina JA, et al. Role of [^{18}F] fluorodeoxyglucose PET/CT in head and neck oncology: the point of view of the radiation oncologist. Br J Radiol 2016:20160217.
[18] Mohan R, Wu Q, Manning M, Schmidt-Ullrich R. Radiobiological considerations in the design of fractionation strategies for intensity-modulated radiation therapy of head and neck cancers. Int J Radiat Oncol Biol Phys 2000;46(3):619–30.
[19] Castelli J, Simon A, Rigaud B, Lafond C, Chajon E, Ospina JD, et al. A Nomogram to predict parotid gland overdose in head and neck IMRT. Radiat Oncol 2016;11:79.
[20] Schemper M, Smith TL. A note on quantifying follow-up in studies of failure time. Control Clin Trials 1996;17(4):343–6.
[21] Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. Br J Cancer 2015;112(2):251–9.
[22] Harrell FE, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med 1996;15(4):361–87.
[23] Efron B, Gong G. A leisurely look at the bootstrap, the jackknife, and cross-validation. Am Stat 1983;37(1):36–48.
[24] Hothorn T, Lausen B. On the exact distribution of maximally selected rank statistics. Comput Stat Data Anal 2003;43(2):121–37.
[25] Kreimer AR, Clifford GM, Boyle P, Franceschi S. Human papillomavirus types in head and neck squamous cell carcinomas worldwide: a systematic review. Cancer Epidemiol Biomarkers Prev 2005;14(2):467–75.
[26] 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(1):24–35.
[27] Romesser PB, Lim R, Spratt DE, Setton J, Riaz N, Lok B, et al. The relative prognostic utility of standardized uptake value, gross tumor volume, and metabolic tumor volume in oropharyngeal cancer patients treated with platinum based concurrent chemoradiation with a pre-treatment [(18)F] fluorodeoxyglucose positron emission tomography scan. Oral Oncol 2014;50(9):802–8.
[28] Akagunduz OO, Savas R, Yalman D, Kocacelebi K, Esassolak M. Can adaptive threshold-based metabolic tumor volume (MTV) and lean body mass corrected standard uptake value (SUL) predict prognosis in head and neck cancer patients treated with definitive radiotherapy/chemoradiotherapy? Nucl Med Biol 2015;42(11):899–904.
[29] Hofheinz F, Lougovski A, Zophel K, Hentschel M, Steffen IG, Apostolova I, et al. Increased evidence for the prognostic value of primary tumor asphericity in pretherapeutic FDG PET for risk stratification in patients with head and neck cancer. Eur J Nucl Med Mol Imaging 2015;42(3):429–37.
[30] Tang C, Murphy JD, Khong B, La TH, Kong C, Fischbein NJ, et al. Validation that metabolic tumor volume predicts outcome in head-and-neck cancer. Int J Radiat Oncol Biol Phys 2012;83(5):1514–20.
[31] La TH, Filion EJ, Turnbull BB, Chu JN, Lee P, Nguyen K, et al. Metabolic tumor volume predicts for recurrence and death in head-and-neck cancer. Int J Radiat Oncol Biol Phys 2009;74(5):1335–41.
[32] Apostolova I, Steffen IG, Wedel F, Lougovski A, Marnitz S, Derlin T, et al. Asphericity of pretherapeutic tumour FDG uptake provides independent prognostic value in head-and-neck cancer. Eur Radiol 2014;24(9):2077–87.
[33] Schinagl DA, Span PN, Oyen WJ, Kaanders JH. Can FDG PET predict radiation treatment outcome in head and neck cancer? Results of a prospective study. Eur J Nucl Med Mol Imaging 2011;38(8):1449–58.
[34] Yabuki K, Shiono O, Komatsu M, Sano D, Nishimura G, Takahashi M, et al. Predictive and prognostic value of metabolic tumor volume (MTV) in patients with laryngeal carcinoma treated by radiotherapy (RT)/concurrent chemoradiotherapy (CCRT). PLoS One 2015;10(2):e0117924.