Eur J Nucl Med Mol Imaging (2015) 42:419–428
DOI 10.1007/s00259-014-2933-1

ORIGINAL ARTICLE

Zone-size nonuniformity of ^{18}F-FDG PET regional textural features predicts survival in patients with oropharyngeal cancer

Nai-Ming Cheng · Yu-Hua Dean Fang · Li-yu Lee · Joseph Tung-Chieh Chang · Din-Li Tsan · Shu-Hang Ng · Hung-Ming Wang · Chun-Ta Liao · Lan-Yan Yang · Ching-Han Hsu · Tzu-Chen Yen

Received: 27 May 2014 / Accepted: 2 October 2014 / Published online: 23 October 2014

Abstract
Purpose The question as to whether the regional textural features extracted from PET images predict prognosis in oropharyngeal squamous cell carcinoma (OPSCC) remains open. In this study, we investigated the prognostic impact of regional heterogeneity in patients with T3/T4 OPSCC.
Methods We retrospectively reviewed the records of 88 patients with T3 or T4 OPSCC who had completed primary therapy. Progression-free survival (PFS) and disease-specific survival (DSS) were the main outcome measures. In an exploratory analysis, a standardized uptake value of 2.5 (SUV 2.5) was taken as the cut-off value for the detection of tumour boundaries. A fixed threshold at 42 % of the maximum SUV (SUV_{max} 42 %) and an adaptive threshold method were then used for validation. Regional textural features were extracted from pretreatment ^{18}F-FDG PET/CT images using the grey-level run length encoding method and grey-level size zone matrix. The prognostic significance of PET textural features was examined using receiver operating characteristic (ROC) curves and Cox regression analysis.
Results Zone-size nonuniformity (ZSNU) was identified as an independent predictor of PFS and DSS. Its prognostic impact

Ching-Han Hsu and Tzu-Chen Yen contributed equally to this work.

Electronic supplementary material The online version of this article (doi:10.1007/s00259-014-2933-1) contains supplementary material, which is available to authorized users.

N.-M. Cheng · T.-C. Yen
Departments of Nuclear Medicine, Chang Gung Memorial Hospital and Chang Gung University, Taiyuan, Taiwan

N.-M. Cheng
Department of Nuclear Medicine, Chang Gung Memorial Hospital, Keelung, Taiwan

N.-M. Cheng · C.-H. Hsu
Department of Biomedical Engineering and Environmental Sciences, National Tsing Hua University, Hsinchu, Taiwan

Y.-H. D. Fang
Department of Electrical Engineering, Chang Gung University, Taiyuan, Taiwan

L.-y. Lee
Department of Pathology, Chang Gung Memorial Hospital, Chang Gung University College of Medicine, Taoyuan, Taiwan

J. T.-C. Chang · D.-L. Tsan
Department of Radiation Oncology, Chang Gung Memorial Hospital, Chang Gung University College of Medicine, Taoyuan, Taiwan

S.-H. Ng
Department of Diagnostic Radiology, Chang Gung Memorial Hospital, Chang Gung University College of Medicine, Taoyuan, Taiwan

H.-M. Wang
Division of Hematology/Oncology, Department of Internal Medicine, Chang Gung Memorial Hospital, Chang Gung University College of Medicine, Taoyuan, Taiwan

C.-T. Liao
Department of Otolaryngology-Head & Neck Surgery, Chang Gung Memorial Hospital, Chang Gung University College of Medicine, Taoyuan, Taiwan

L.-Y. Yang
Biostatistics Unit, Clinical Trial Center, Chang Gung Memorial Hospital, Taoyuan, Taiwan

T.-C. Yen (✉)
Department of Nuclear Medicine and Molecular Imaging Center, Chang Gung Memorial Hospital, Chang Gung University College of Medicine, Taipei, Taiwan
e-mail: yen1110@adm.cgmh.org.tw

was confirmed using both the SUV_{max} 42 % and the adaptive threshold segmentation methods. Based on (1) total lesion glycolysis, (2) uniformity (a local scale texture parameter), and (3) ZSNU, we devised a prognostic stratification system that allowed the identification of four distinct risk groups. The model combining the three prognostic parameters showed a higher predictive value than each variable alone.
Conclusion ZSNU is an independent predictor of outcome in patients with advanced T-stage OPSCC, and may improve their prognostic stratification.

Keywords Oropharyngeal carcinoma · HPV · FDG PET/CT · Texture analysis

Introduction

Oropharyngeal squamous cell carcinoma (OPSCC) is one of the most common types of head and neck cancer [1]. The major risk factors for OPSCC include alcohol consumption, tobacco smoking and human papillomavirus (HPV) infection [2, 3]. Interestingly, HPV-positive OPSCC patients generally have better clinical outcomes than HPV-negative subjects [4, 5]. Although advanced T-stage has been shown to portend a dismal prognosis [4, 6], the current staging system (mainly based on tumour extent and lymph node invasion) does not adequately stratify patients with regard to clinical outcome [7]. In order to improve the prognostic stratification of patients with advanced-stage OPSCC, Ang et al. have devised a prognostic system based on HPV status, smoking and clinical stage [4]. Although this system has been validated in areas in which HPV infection is prevalent [8], data from regions with a low rate of HPV-positive tumours are scarce.

Several studies have investigated the prognostic value in patients with OPSCC of ^{18}F-FDG PET parameters including maximum standard uptake value (SUV_{max}), metabolic tumour volume (MTV) and total lesion glycolysis (TLG), and textural analysis that characterizes ^{18}F-FDG uptake heterogeneity [9–11]. Local scale FDG PET textural features have been successfully used for predicting clinical outcome and treatment response in a variety of malignancies [12–14]. In this regard, we have previously shown that uniformity, derived using the normalized grey-level co-occurrence matrix (NGLCM) method, is a significant prognostic factor in patients with advanced OPSCC [13]. However, data on the regional scale PET textural features remain scarce [15, 16]. In this context, we investigated whether regional heterogeneity in intratumoral FDG distribution may predict outcomes and improve risk stratification in patients with advanced T-stage OPSCC.

Materials and methods

Patients

The study protocol was approved by the hospital Ethics Committee and informed consent was waived due to the retrospective nature of the study. We retrospectively reviewed the medical records of 88 patients with T3 or T4 stage OPSCC who had completed primary therapy and a follow-up of more than 20 months. The exclusion criteria were as follows: (1) M1 stage at diagnosis; (2) previous treatment for OPSCC at other institutions, and (3) previous history of malignancies without achieving complete remission. It was our treatment policy that every OPSCC patient received a complete staging work-up before therapy, which included MRI or CT scans of the head and neck region and ^{18}F-FDG PET/CT scan. Staging was performed according to the 2002 American Joint Committee on Cancer (AJCC) staging criteria [17]. The post-therapy status was assessed using the RECIST criteria 1.1 [18]. Patients were classified as having a complete response (CR), partial response (PR), stable disease (SD), or progressive disease (PD) according to follow-up CT or MRI scans 2 – 3 months after completing primary therapy by an expert radiologist who was blinded to the clinical data.

HPV typing and p16 immunohistochemistry

HPV typing and p16 immunohistochemistry were available in all participants. Between June 2006 and September 2010, HPV typing of primary tumour biopsy specimens was performed as previously described [19–21]. From September 2010, p16 immunohistochemistry was performed using the CINtec histology kit (MTM laboratories, Heidelberg, Germany) according to the manufacturer's protocol. High p16 expression was defined as strong intensity and staining of more than 60 % of the tumour cells [22]. HPV-positive tumours were defined as those having high p16 expression or positive results on HPV typing.

PET/CT acquisition

All patients underwent FDG PET/CT for staging before treatment. Participants were asked to fast for at least 6 h prior to examination. Images were obtained 50 min after intravenous injection of 370 – 555 MBq ^{18}F-FDG, depending on body weight. Whole-body PET emission scans were acquired using the same scanner (Discovery ST 16; GE Healthcare, Milwaukee, WI). PET images were reconstructed using CT-based attenuation correction.

PET/CT image analysis

Tumours were segmented using the PMOD 3.3 software package (PMOD Technologies Ltd, Zurich, Switzerland). Several segmentation methods have been proposed for the analysis of PET images in patients with head and neck malignancies, but no widely accepted guidelines exist. Although a fixed threshold of 50 % of the maximum SUV (SUV_{max} 50 %) may be unsuitable for determining tumour volume in the presence of heterogeneous lesions [23], this threshold has been used successfully in head and neck cancer [24, 25]. We selected this delineation method based on its availability and clinical usefulness. An experienced nuclear medicine physician (blinded to the clinical data but not to tumour location) retrospectively analysed all PET/CT images. The borders of volumes of interest (VOI) were set by manual adjustment to exclude adjacent physiological FDG-avid structures on PET/CT images. The VOIs were checked and validated by an independent senior nuclear medicine physician. The tumour boundaries were than automatically contoured based on the following three thresholds: (1) SUV 2.5, which has been previously used for texture analysis in oropharyngeal and oesophageal cancer [9, 13, 26, 27], (2) 42 % of SUV_{max}, which has been utilized for delineation of oropharyngeal cancer [11], and (3) an adaptive threshold method [28]. The background of the adaptive threshold was defined as the mean uptake of a 10-mm VOI sphere placed in the right lower neck where no lymph node lesions were evident in our study subjects. Two nuclear medicine physicians independently delineated the background and the uptake values were averaged.

TLG was calculated according to the following formula: TLG = mean SUV × MTV (cm^3) [29]. The grey-level run length encoding matrix (GLRLM) [30] and grey-level size zone matrix (GLSZM) [31] were used for assessing the regional textural features. The GLRLM indicates the number of voxel segments with the same intensity in a given direction. Because there are a total of 13 angular orientations in the three-dimensional model, 13 GLRLM values (one for each direction) were obtained for each tumour. The GLRLM features were derived from an average of 13 directions to make the rotation invariant and allow for different positions of the study patients.

Similar to GLRLM, the GLSZM estimates the number of adjacent voxels with the same intensity. A grey-level zone is a region of voxels with the same grey level value. The GLSZM includes different areas and does not require the calculation of different directions. A total of 22 different textural features were extracted from the GLRLM and GLSZM (Supplementary Table 1). Detailed information on the calculations of each textural feature has been previously reported [16, 30, 31]. The intensity of FDG uptake in the primary tumour was resampled to 16, 32 and 64 different values to reduce image noise and improve reproducibility [12, 32]. Therefore, we derived a total of 66 textural parameters. The textural features were analysed using a software package developed in-house (Chang-Gung Image Texture Analysis toolbox, CGITA) implemented under MATLAB 2012a (Mathworks Inc., Natick, MA).

Study outcomes

Progression-free survival (PFS) and disease-specific survival (DSS) were the main outcome measures. DSS was defined as the time from diagnosis to the time of OPSCC-related death. PFS was defined as the time from diagnosis to the time of recurrence, cancer death, or unequivocal radiological evidence of progression in patients who showed residual disease after primary therapy.

Statistical analysis

In an exploratory analysis, SUV 2.5 was taken as the cut-off value for the detection of tumour boundaries. The evaluation of textural parameters was based on a step-forward process. First, the parameters were examined in relation to PFS using receiver operating characteristic (ROC) curve analysis. The optimal cut-off values were identified as those that maximized the sum of sensitivity and specificity. PET parameters with an area under curve (AUC) significantly different from 0.5 were selected (as continuous variables) for further analyses in relation to PFS and DSS using univariate and multivariate Cox regression models. Multivariate Cox regression models were constructed to include as covariates HPV positivity, age, tobacco, alcohol consumption, T stage, N stage, AJCC stage and uniformity (given its significant predictive value for survival endpoints [13]). Although TLG and MTV have been shown to predict survival [10, 11], they were not included in the model because of their high collinearity with regional textural parameters. We therefore constructed separate multivariate Cox regression models (using PFS and DSS as endpoints) in the subgroups of patients with high and low TLG (or MTV). Significant parameters were further dichotomized and entered in the multivariate Cox regression models. Survival estimates were plotted using the Kaplan-Meier method and compared using the log-rank test. Significant parameters found using SUV 2.5 were further validated using the two other segmentation methods (i.e. SUVmax 42 % and adaptive threshold). The parameters derived using the different segmentation methods were tested as continuous variables in ROC curves analysis, and univariate and multivariate Cox regression analyses. Spearman's correlation (ρ) and intraclass correlation coefficients (ICC) were used to assess the associations among independent and interdependent parameters, respectively. Calculations were performed using PASW statistics 18 (SPSS Inc., Chicago, IL) for Windows. Two-tailed P values <0.05 were considered statistically significant. Bonferroni correction was applied to adjust P values for multiple comparisons.

Results

Patient characteristics

Between June 2006 and August 2012, a total of 88 consecutive patients with advanced T-stage OPSCC were identified (6 women and 82 men; median age at diagnosis 51.5 years) who were eligible for this study. The median follow-up times in the entire study cohort and in the subgroup of patients who survived were 32 months (range 5 – 92 months) and 57 months (range 20 – 92 months), respectively. The stage distribution was as follows: T4 (60 patients, 68.2 %), N2b-N3 (66 patients, 75.0 %), and IVa (50 patients, 56.8 %; Table 1). A total of 83 patients (94.3 %) received concurrent chemoradiotherapy, three patients (3.4 %) received bioradiotherapy, and the remaining two patients (2.3 %) received radiotherapy alone with curative intent. A total of 71 patients (80.7 %) were smokers. Quantitative smoking data (median 27.5, range 3 – 120 pack-years) were available in 54 patients (76.1 %). HPV typing was performed in 56 patients, whereas p16 was assessed by immunohistochemistry in 32 patients. There were 17 (19.3 %) HPV-positive patients, and of these, 11 (64.7 %) were smokers.

Clinical outcomes

The RECIST response was assessed by means of contrast-enhanced CT in 38 patients and by MRI in the remaining 50 patients. There were 58 (65.9 %), 12 (13.6 %), 5 (5.7 %) and 13 (14.8 %) patients who had CR, PR, SD and PD, respectively. Among the 30 patients who did not obtain CR, 7 underwent salvage surgery and 4 achieved long-term survival. At the time of analysis, a total of 49 patients (55.7 %) were dead. Of these, 36 died of cancer, 8 died of second primary cancers, 3 died of pneumonia, 1 died of cerebral vascular disease, and 1 died of acute myocardial infarction (Supplementary Figure 1). We identified 39 progression events: 21 were found on follow-up imaging, 15 were recurrence of OPSCC, and 3 were deaths from disease. Among the 17 HPV-positive patients, 2 (11.8 %) and 4 (23.5 %) had residual and recurrent disease, respectively. The risk profile of Ang et al. was determined in 83 patients, and 3 (3.6 %), 19 (22.9 %) and 61 (73.5 %) were in the low, intermediate and high risk groups, respectively.

Survival prediction

The results of ROC curve analysis indicated that zone-size nonuniformity (ZSNU) of 16, 32 and 64 bins, grey-level nonuniformity of 16 and 32 bins, short zones low grey-level emphasis of 64 bins, and run length nonuniformity of 32 and 64 bins were associated with PFS with P values less than 0.05. Univariate and multivariate Cox regression analyses revealed that all of the above parameters (the only exceptions being run length nonuniformity of 32 and 64 bins) were predictors of PFS (Supplementary Table 2) with P values less than 0.05.

In the subgroup of patients with high TLG values, ZSNU of 16, 32 and 64 bins retained its independent prognostic significance, but only ZSNU of 16 bins was significantly associated with DSS (Supplementary Table 2). A total of five analyses were used to select the regional textural parameters. Even after adjusting the P value for multiple testing, ZSNU of 16 bins retained its statistical significance. Similar results were found in the subgroup of patients with high MTV values (Supplementary Table 3). ZSNU (16 bins) and uniformity (either dichotomized or expressed as continuous variables) were significantly associated with PFS (Table 2) and DSS (Supplementary Table 4). Figure 1 shows the Kaplan-Meier estimates of PFS and DSS according to TLG, uniformity and ZSNU (16 bins).

In the subgroup of patients with low TLG values, the coefficients of the Cox model did not converge. Therefore, the multivariate Cox regression model was not applied, and Kaplan-Meier estimates were utilized. Notably, only uniformity was significantly associated with outcomes in the low TLG subgroup (Fig. 2). Similar results were evident in the low MTV subgroup.

Significantly lower uniformity and higher ZSNU (16 bins) were observed in patients who had residual or recurrent disease (Mann-Whitney U test, P=0.045 and 0.006, respectively). In contrast, less heterogeneity in tumour FDG distribution was associated with a better response to primary therapy. Accordingly, ZSNU (16 bins) values were significantly lower in patients with RECIST CR (Mann-Whitney U test, P= 0.032).

SUV_{max}, MTV and TLG were significantly associated with PFS (P=0.004, 0.014 and 0.006, respectively) in the ROC curve analysis. Tobacco, alcohol usage, high SUV_{max}, MTV and TLG were significantly associated with PFS in the univariate analysis (Table 1), with the highest hazard ratio found for high TLG. Moreover, ZSNU (16 bins) was significantly associated with both TLG and MTV (Spearman's ρ 0.791 and 0.787, respectively; both $P < 0.001$). However, uniformity was not significantly correlated with either MTV or TLG (Spearman's ρ 0.154 and 0.066; P=0.151 and 0.539, respectively). Uniformity and ZSNU (16 bins) were not significantly associated (ρ=−0.068, P=0.526). However, ZSNU of 32 and 64 bins were moderately correlated with uniformity (ρ=0.213 and 0.298; P=0.097 and 0.019, respectively) in the subgroup with high TLG values.

Validation data

MTV and TLG delineated using SUV_{max} 42 % and the adaptive threshold methods were significantly lower than those

Table 1 General characteristics and dichotomization of imaging parameters: results of univariate analysis of progression-free survival

Characteristic	No. (%) of patients	Hazard ratio (95 % confidence interval)	*P* value
HPV positivity	17 (19.3)	0.51 (0.20 – 1.30)	0.158
Age >50 years[a]	33 (37.5)	0.57 (0.28 – 1.14)	0.113
Sex			
Female	6 (6.8)	0.36 (0.05 – 2.60)	0.309
Male	82 (93.2)		
Tobacco use	71 (80.7)	3.39 (1.05 – 11.03)	0.042
Alcohol use	60 (68.2)	2.26 (1.04 – 4.92)	0.040
T stage			
T3	28 (31.8)	1.59 (0.78 – 3.27)	0.206
T4	60 (68.2)		
N stage			
N0-N2a	22 (25.0)	1.83 (0.81 – 4.15)	0.148
N2b-N3	66 (75.0)		
AJCC stage			
Stage III	13 (14.8)	1.50 (0.92 – 2.44)	0.106
Stage IVa	50 (56.8)		
Stage IVb	25 (28.4)		
PET parameters			
High SUV maximum	66 (75.0)	3.26 (1.44 – 7.40)	0.005
High MTV	65 (73.9)	5.35 (1.64 – 17.38)	0.005
High TLG	62 (70.5)	6.62 (2.04 – 21.54)	0.002
High uniformity	55 (62.5)	0.37 (0.20 – 0.70)	0.002
High zone size nonuniformity (16 bins)	56 (63.6)	4.11 (1.72 – 9.82)	0.001

[a] Age range 33 – 86 years

obtained using the SUV 2.5 threshold (Wilcoxon signed-ranks test, both P <0.001), with SUV_{max} 42 % yielding the lowest values. In contrast to the use of SUV 2.5, high ICCs were noted for MTV, TLG and ZSNU (16 bins) values using SUV_{max} 42 % (ICC 0.800, 0.934 and 0.926, respectively) and the adaptive threshold methods (ICC 0.973, 0.996 and

Table 2 Multivariate Cox regression analysis of progression-free survival rate

Characteristic	Dichotomized variables		Continuous variables	
	Hazard ratio (95 % confidence interval)	*P* value	Hazard ratio (95 % confidence interval)	*P* value
Textural parameters				
Uniformity	0.27 (0.14 – 0.53)	< 0.001	0.05 (0.01 – 0.26) [a]	< 0.001
Zone-size nonuniformity (16 bins)	4.38 (1.69 – 11.34)	0.002	1.64 (1.24 – 2.17) [b]	0.001
HPV positivity	0.65 (0.24 – 1.78)	0.400	0.80 (0.30 – 2.13)	0.654
Age >50 years	0.58 (0.28 – 1.22)	0.152	0.48 (0.23 – 1.00)	0.050
Tobacco use	2.96 (0.64 – 13.69)	0.164	2.36 (0.53 – 10.54)	0.262
Alcohol use	1.57 (0.60 – 4.10)	0.362	2.11 (0.78 – 5.74)	0.143
T stage				
T4 vs. T3	1.50 (0.66 – 3.39)	0.332	1.11 (0.48 – 2.58)	0.808
N stage				
N0-N2a vs. N2b-N3	1.82 (0.62 – 5.32)	0.275	2.15 (0.73 – 6.38)	0.167
AJCC stage				
Stage III vs. IVa vs. IVb	0.84 (0.43 – 1.62)	0.596	0.71 (0.35 – 1.42)	0.332

[a] Hazard ratio expressed per 0.1-unit increase of the variable

[b] Hazard ratio expressed per 10-unit increase of the variable

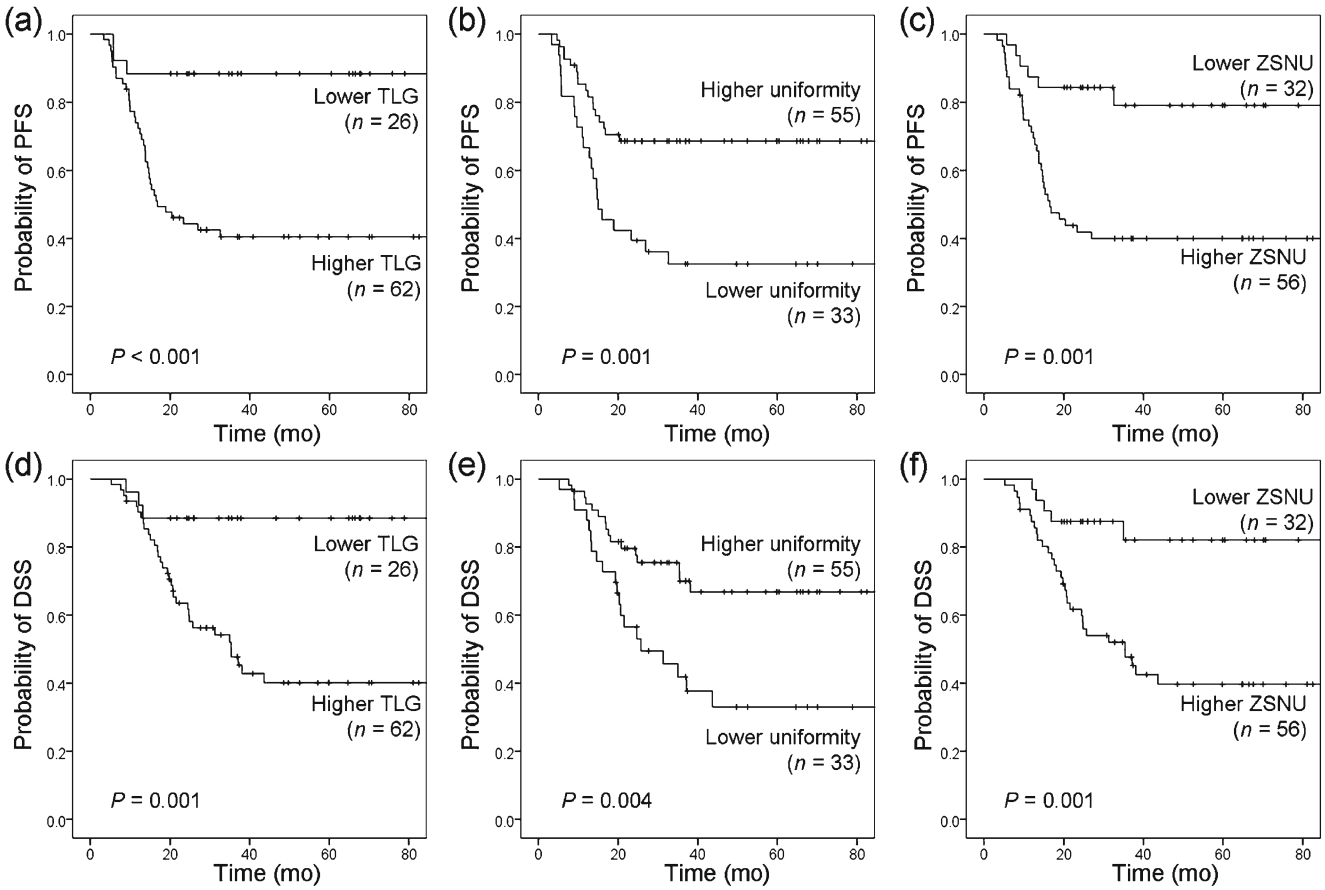

Fig. 1 Kaplan-Meier estimates of progression-free survival (PFS) and disease-specific survival (DSS) according to TLG, uniformity and zone-size nonuniformity (ZSNU) of 16 bins. *P* values (log-rank test) are also shown

0.926, respectively). Uniformity obtained using the adaptive threshold methods showed a good correlation (ICC 0.926), but a worse correlation (ICC 0.574) was found using SUV_{max} 42 %. The predictive value of ZSNU (16 bins) delineated using SUV_{max} 42 % and adaptive threshold methods for PFS (Table 3) and DSS was successfully validated (Supplementary Table 5). However, only uniformity derived using the adaptive threshold methods was statistically significant.

Risk stratification for the prediction of PFS

To integrate the prognostic information provided by TLG, uniformity and ZSNU (16 bins), we constructed a stratification system for predicting PFS. To this aim, OPSCC patients were divided into subgroups according to the cut-off values of TLG, uniformity and ZSNU (16 bins). All of the *P* values for TLG, uniformity and ZSNU (16 bins) were lower than the Bonferroni-corrected threshold. We did not apply ZSNU of 16 bins in the lower TLG subgroup. Patients with a score of 0 had small and uniform tumours (i.e. TLG ≤110.67 g plus uniformity >0.135); patients with a score of 1 had only one risk factor (i.e. TLG ≤110.67 g and uniformity ≤0.135; or TLG >110.67 g plus uniformity >0.135 and ZSNU ≤16.22); patients with a score of 2 had two risk factors (i.e. TLG >110.67 g plus uniformity ≤0.135 or ZSNU >16.22); and patients with a score of 3 had all of the three risk factors (i.e. TLG >110.67 g plus uniformity ≤0.135 and ZSNU >16.22). The scoring system is summarized in Supplementary Figure 2. The resulting four risk groups were significantly different in terms of PFS and DSS (Fig. 3). Two patients with similar clinical characteristics but different scores and outcomes are presented in Supplementary Fig. 3.

The stratification system showed a high predictive value for both PFS and DSS, with AUCs of 0.81 and 0.80, respectively. AUCs for the three-variable system were significantly higher for both survival endpoints as compared with those for single parameters (TLG, uniformity or ZSNU of 16 bins) alone (PFS $P<0.001$, $P=0.003$ and $P=0.001$, respectively; DSS $P<0.001$, $P=0.004$ and $P=0.006$, respectively). Compared with the model based on TLG and uniformity described previously [13], the AUCs were higher for the system based on TLG, uniformity and ZSNU (difference for PFS, $P=0.050$; difference for DSS, $P=0.019$; Fig. 4). The results of the multivariate Cox regression analysis (83 patients) confirmed

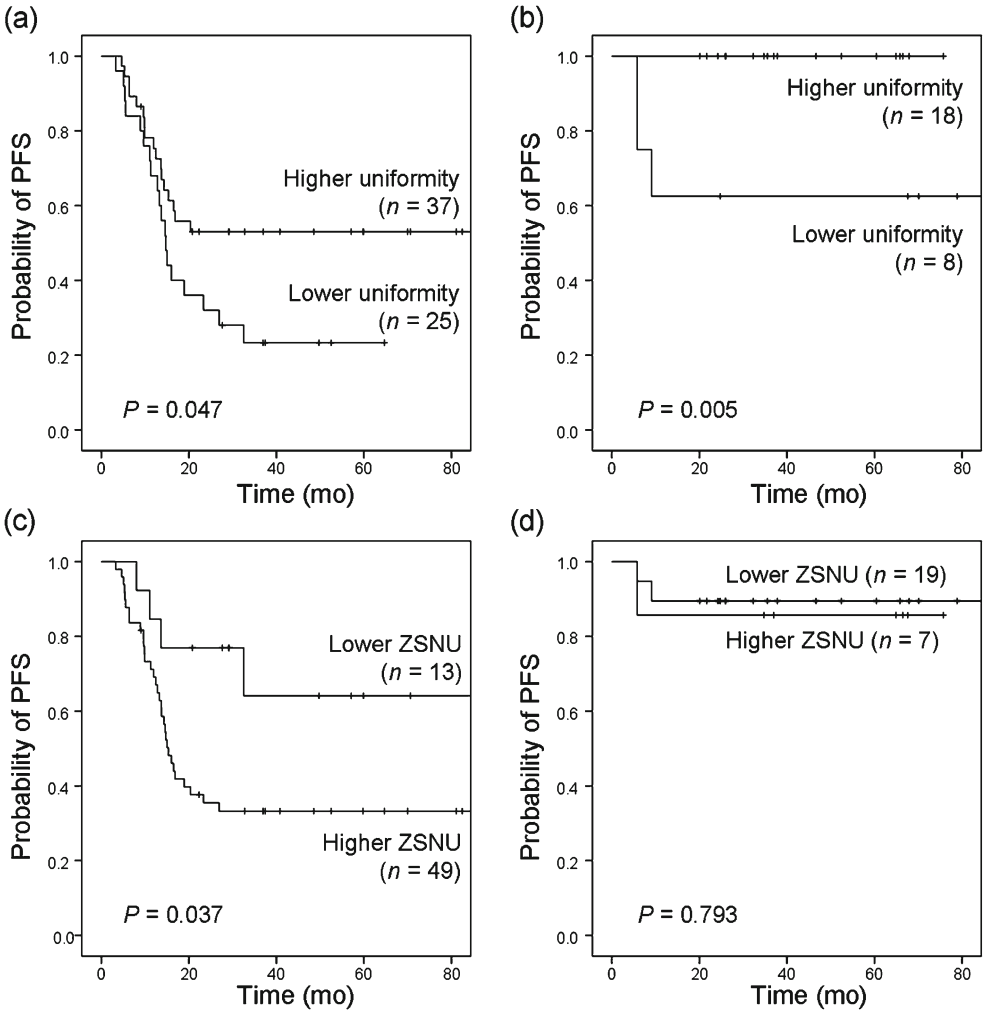

Fig. 2 Kaplan-Meier estimates of progression-free survival (PFS) according to uniformity in patients with high TLG (**a**) and low TLG (**b**), and according to zone-size nonuniformity (ZSNU) of 16 bins in patients with high TLG (**c**) and low TLG (**d**). *P* values (log-rank test) are also shown

the independent significance of our imaging-based stratification system for predicting PFS and DSS (Table 4). Notably, the system of Ang et al. provided only marginally significant predictions of PFS and DSS.

Table 3 Receiver operating characteristic (ROC) curve, univariate and multivariate Cox regression analyses of progression-free survival in segmentations used for validation

	ROC curve analysis		Univariate Cox analysis		Multivariate Cox analysis[a]	
	AUC (range)	*P* value	Hazard ratio (95 % confidence interval)	*P* value	Hazard ratio (95 % confidence interval)[b]	*P* value
42 % of SUV_{max}						
Uniformity	0.48 (0.36 – 0.61)	0.785	0.90 (0.33 – 2.44)[b]	0.830	0.50 (0.17 – 1.44)[b]	0.198
Zone-size nonuniformity (16 bins)	0.68 (0.57 – 0.80)	0.003	1.31 (1.10 – 1.56)[c]	0.003	1.44 (1.12 – 1.83)[c]	0.004
Adaptive threshold						
Uniformity	0.32 (0.21 – 0.44)	0.004	0.16 (0.04 – 0.70)[b]	0.015	0.10 (0.02 – 0.50)[b]	0.005
Zone-size nonuniformity (16 bins)	0.70 (0.59 – 0.81)	0.001	1.43 (1.16 – 1.77)[c]	0.001	1.53 (1.17 – 2.02)[c]	0.002

[a] Analysis with covariables as shown in Table 2

[b] Expressed per 0.1-unit increase of the variable

[c] Expressed per 10-unit increase of the variable

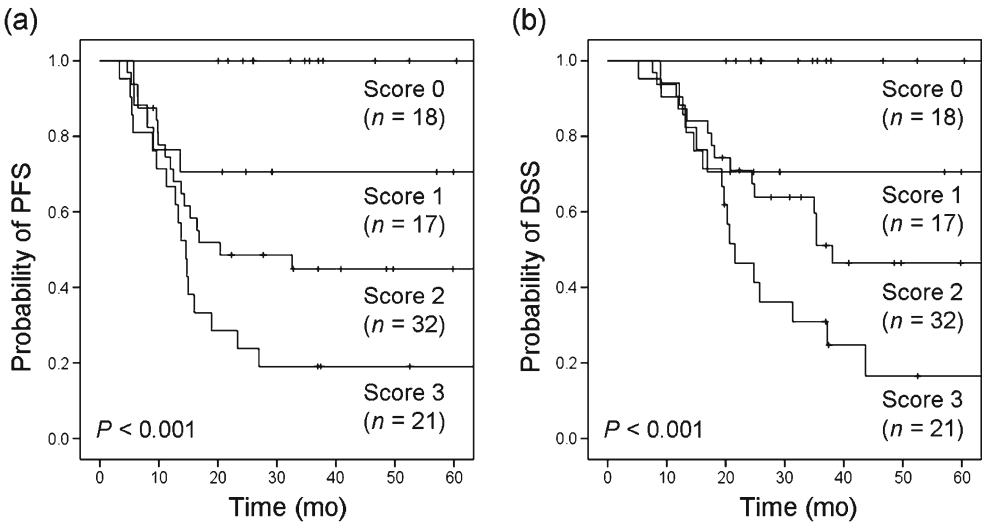

Fig. 3 Risk stratification model for predicting 5-year survival rates based on tumour TLG, uniformity and zone-size nonuniformity (ZSNU) of 16 bins. There were significant differences between prognostic score groups in terms of both PFS (**a**) and DSS (**b**). *P* values (log-rank test) are also shown

Discussion

Evidence suggests that intratumour heterogeneity plays a crucial role in the proliferation, vascular supply and metabolism of different solid malignancies [33], and may facilitate the selection of drug-resistant cells [34]. In the present study, we characterized intratumour heterogeneity in FDG PET images by textural analysis. The main findings indicated that uniformity and ZSNU (16 bins) were significantly associated with survival of OPSCC patients. The predictive value of uniformity and ZSNU can be explained by their capacity to differentiate residual or recurrent disease in OPSCC. Notably, ZSNU (16 bins) was able to identify patients who showed CR, which may account for the higher predictive ability of the three-variable prognostic model (that included ZSNU of 16 bins) as compared with each prognostic factor alone. Uniformity and ZSNU are derived from NGLCM and GLSZM, respectively. Basically, NGLCM indicates how frequently, denoted as $C(i,j)$, a voxel of resampled intensity i is close to another voxel of intensity j. Uniformity is the sum of the squares of the $C(i,j)$ values. For heterogeneous images, the resulting value of uniformity is small. GLSZM is a higher-order texture matrix that estimates the number of grey-level zones of different sizes. A grey-level zone is a region of voxels showing the same grey-level value. ZSNU is the sum of the squares of grey-level zones normalized to the total number of zones. For heterogeneous images, ZSNU is expected to be larger due to the uneven zone distribution.

Different MTV values were noted when different segmentation methods were used. However, ZSNU (16 bins) was significantly associated with survival independently of the three different delineation methods used. This finding may have been due, at least in part, to the normalization step in the ZSNU equation. Indeed, this would have reduced the impact of MTV changes and higher ICC values among the three different delineation methods. In contrast, the calculation of uniformity does not include a normalization step. Consequently, the use of SUVmax 42 % as a threshold was associated with a lower degree of correlation and a lack of association with survival was noted.

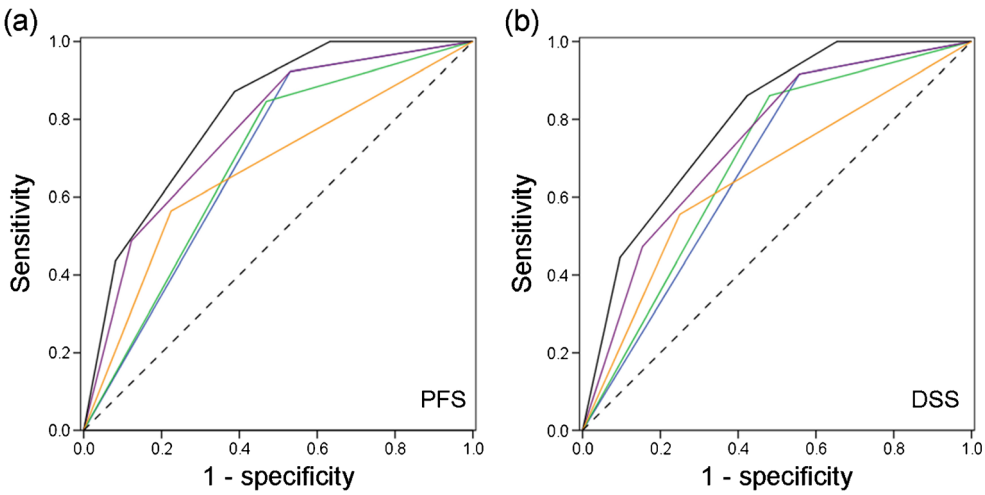

Fig. 4 ROC curves and AUCs for the identification of disease progression (**a**) and OPSCC-related death (**b**) obtained from the risk stratification model based on TLG, uniformity and ZSNU of 16 bins (*black*). The AUC of the three-variable prognostic model was significantly larger than the AUCs obtained using TLG (*blue*), uniformity (*orange*), ZSNU (*green*) alone or the combination of TLG and uniformity (*purple*). The *dashed line* is the reference line

Table 4 Multivariate Cox regression analysis of progression-free survival (PFS) and disease-specific survival (DSS) rates

Prognostic model	PFS		DSS	
	Hazard ratio (95 % confidence interval)	*P* value	Hazard ratio (95 % confidence interval)	*P* value
TLG, uniformity, ZSNU	4.35 (1.49 – 12.68)	0.007	5.12 (1.51 – 17.28)	0.009
TLG, uniformity[a]	0.48 (0.13 – 1.73)	0.261	0.38 (0.09 – 1.58)	0.182
Risk profile of Ang et al.	2.16 (0.99 – 4.72)	0.053	2.10 (0.96 – 4.60)	0.062

[a] See Cheng et al. [13] for detailed information

Our study successfully replicated the results of a recent investigation [35] showing that MTV and TLG are significantly associated with ZSNU, but not with uniformity. The previous study [35] suggested the use of at least 32 bins for textural analysis due to the higher degree of correlation with the 64 bin data. Our present study showed a higher correlation for ZSNU (32 bins) (Spearman's ρ=0.972, P<0.001) than for ZSNU (64 bins). However, ZSNU (16 bins) was a significant predictor of survival, especially in the subgroup of patients with high TLG values. In these subjects, moderate correlations of ZSNU 32 and 64 bins with uniformity were noted (Spearman's ρ=0.213 and 0.298, P=0.097 and 0.019, respectively). In this regard, the choice of the optimal number of bins of texture parameters for survival analysis should be guided by the interaction and interdependency among the parameters. In presence of highly correlated data, the use of advanced techniques (e.g. machine learning models) [36] may be required. Further research is needed to address this important issue.

No clinical variable was found to be significantly associated with outcomes in the current study. There are at least two reasons for this finding. First, most of our patients (61 patients, 73.5 %) were classified as being in the high-risk group (Ang et al. [4]). Because only patients with the most severe OPSCC were included in our study, common clinical risk factors could not be expected to predict outcomes in a reliable manner. Nonetheless, our PET image-based scoring system was independently associated with both PFS and DSS. Another point that merits consideration is the high incidence (64.7 %) of tobacco use in our HPV-positive patients, which could have weakened the favourable impact of HPV infections on PFS and DSS.

Some caveats inherent in this study merit comment. The exact biological mechanisms that may link uniformity and ZSNU remain unclear. Further studies are required. Because of the retrospective nature of our investigation, we cannot exclude the presence of a selection bias. Consequently, our results need to be confirmed and validated in future prospective investigations. Another point is the use of a single PET/CT scanner in our study. Further research is needed to investigate how the use of different scanners may affect textural analysis and its prognostic significance. Finally, our model was constructed using the SUV threshold of 2.5. The role and impact of other tailored segmentation methods [37, 38] for survival prediction and model construction deserve further investigation.

Conclusion

Our results demonstrate that ZSNU, a regional textural parameter, may serve as an independent predictor of PFS and DSS in advanced T-stage OPSCC. Moreover, a prognostic scoring system that integrates TLG, uniformity and ZSNU derived from pretreatment PET images may be clinically useful for improving risk stratification and clinical management in this group of patients.

Conflicts of interest None.

Financial support This study was supported by the grant NMRPG102-2221-E-182-074-MY2 from the National Science Council, Taiwan. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

1. Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin DM. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. Int J Cancer. 2010;127:2893–917.
2. Applebaum KM, Furniss CS, Zeka A, Posner MR, Smith JF, Bryan J, et al. Lack of association of alcohol and tobacco with HPV16-associated head and neck cancer. J Natl Cancer Inst. 2007;99:1801–10.
3. Bouvard V, Baan RSK, Grosse Y, Secretan B, El Ghissassi F, Benbrahim-Tallaa L, et al. A review of human carcinogens - Part B: biological agents. Lancet Oncol. 2009;10:321–2.
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. Fakhry C, Westra WH, Li S, Cmelak A, Ridge JA, Pinto H, et al. Improved survival of patients with human papillomavirus-positive head and neck squamous cell carcinoma in a prospective clinical trial. J Natl Cancer Inst. 2008;100:261–9.

6. Sedaghat AR, Zhang Z, Begum S, Palermo R, Best S, Ulmer KM, et al. Prognostic significance of human papillomavirus in oropharyngeal squamous cell carcinomas. Laryngoscope. 2009;119:1542–9.
7. Wu Y, Posner MR, Schumaker LM, Nikitakis N, Goloubeva O, Tan M, et al. Novel biomarker panel predicts prognosis in human papillomavirus-negative oropharyngeal cancer: an analysis of the TAX 324 trial. Cancer. 2011;118:1811–7.
8. Granata R, Miceli R, Orlandi E, Perrone F, Cortelazzi B, Franceschini M, et al. Tumor stage, human papillomavirus and smoking status affect the survival of patients with oropharyngeal cancer: an Italian validation study. Ann Oncol. 2012;23:1832–7.
9. Cheng NM, Chang JT, Huang CG, Tsan DL, Ng SH, Wang HM, et al. Prognostic value of pretreatment (18)F-FDG PET/CT and human papillomavirus type 16 testing in locally advanced oropharyngeal squamous cell carcinoma. Eur J Nucl Med Mol Imaging. 2012;39:1673–84.
10. Dibble EH, Alvarez AC, Truong MT, Mercier G, Cook EF, Subramaniam RM. 18F-FDG metabolic tumor volume and total glycolytic activity of oral cavity and oropharyngeal squamous cell cancer: adding value to clinical staging. J Nucl Med. 2012;53:709–15.
11. Lim R, Eaton A, Lee NY, Setton J, Ohri N, Rao S, et al. 18F-FDG PET/CT metabolic tumor volume and total lesion glycolysis predict outcome in oropharyngeal squamous cell carcinoma. J Nucl Med. 2012;53:1506–13.
12. Tixier F, Le Rest CC, Hatt M, Albarghach N, Pradier O, Metges JP, 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.
13. Cheng NM, Fang YH, Chang JT, Huang CG, Tsan DL, Ng SH, et al. Textural features of pretreatment 18F-FDG PET/CT images: prognostic significance in patients with advanced T-stage oropharyngeal squamous cell carcinoma. J Nucl Med. 2013;54:1703–9.
14. Cook GJ, Yip C, Siddique M, Goh V, Chicklore S, Roy A, et al. Are pretreatment 18F-FDG PET tumor textural features in non-small cell lung cancer associated with response and survival after chemoradiotherapy? J Nucl Med. 2013;54:19–26.
15. Hatt M, Tixier F, Cheze Le Rest C, Pradier O, Visvikis D. Robustness of intratumour 18F-FDG PET uptake heterogeneity quantification for therapy response prediction in oesophageal carcinoma. Eur J Nucl Med Mol Imaging. 2013;40:1662–71.
16. Yang F, Thomas MA, Dehdashti F, Grigsby PW. Temporal analysis of intratumoral metabolic heterogeneity characterized by textural features in cervical cancer. Eur J Nucl Med Mol Imaging. 2013;40:716–27.
17. O'Sullivan B, Shah J. New TNM staging criteria for head and neck tumors. Semin Surg Oncol. 2003;21:30–42.
18. Therasse P, Arbuck SG, Eisenhauer EA, Wanders J, Kaplan RS, Rubinstein L, et al. New guidelines to evaluate the response to treatment in solid tumors. European Organization for Research and Treatment of Cancer, National Cancer Institute of the United States, National Cancer Institute of Canada. J Natl Cancer Inst. 2000;92:205–16.
19. Huang SL, Chao A, Hsueh S, Chao FY, Huang CC, Yang JE, et al. Comparison between the hybrid capture II test and an SPF1/GP6+ PCR-based assay for detection of human papillomavirus DNA in cervical swab samples. J Clin Microbiol. 2006;44:1733–9.
20. Lin CY, Chen HC, Lin RW, You SL, You CM, Chuang LC, et al. Quality assurance of genotyping array for detection and typing of human papillomavirus. J Virol Methods. 2007;140:1–9.
21. Luo CW, Roan CH, Liu CJ. Human papillomaviruses in oral squamous cell carcinoma and pre-cancerous lesions detected by PCR-based gene-chip array. Int J Oral Maxillofac Surg. 2007;36:153–8.
22. Al-Swiahb JN, Huang CC, Fang FM, Chuang HC, Huang HY, Luo SD, et al. Prognostic impact of p16, p53, epidermal growth factor receptor, and human papillomavirus in oropharyngeal cancer in a betel nut-chewing area. Arch Otolaryngol Head Neck Surg. 2010;136:502–8.
23. Hatt M, Cheze-le Rest C, van Baardwijk A, Lambin P, Pradier O, Visvikis D. Impact of tumor size and tracer uptake heterogeneity in (18)F-FDG PET and CT non-small cell lung cancer tumor delineation. J Nucl Med. 2011;52:1690–7.
24. 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;74:1335–41.
25. 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;83:1514–20.
26. Chang KP, Tsang NM, Liao CT, Hsu CL, Chung MJ, Lo CW, et al. Prognostic significance of 18F-FDG PET parameters and plasma Epstein-Barr virus DNA load in patients with nasopharyngeal carcinoma. J Nucl Med. 2012;53:21–8.
27. Tan S, Kligerman S, Chen W, Lu M, Kim G, Feigenberg S, et al. Spatial-temporal [18F]FDG-PET features for predicting pathologic response of esophageal cancer to neoadjuvant chemoradiation therapy. Int J Radiat Oncol Biol Phys. 2013;85:1375–82.
28. Nestle U, Kremp S, Schaefer-Schuler A, Sebastian-Welsch C, Hellwig D, Rübe C, et al. Comparison of different methods for delineation of 18F-FDG PET-positive tissue for target volume definition in radiotherapy of patients with non-small cell lung cancer. J Nucl Med. 2005;46:1342–8.
29. Larson SM, Erdi Y, Akhurst T, Mazumdar M, Macapinlac HA, Finn RD, et al. Tumor treatment response based on visual and quantitative changes in global tumor glycolysis using PET-FDG imaging. The visual response score and the change in total lesion glycolysis. Clin Positron Imaging. 1999;2:159–71.
30. Loh HH, Leu JG, Luo RC. The analysis of natural textures using run length features. IEEE Trans Ind Electron. 1988;35:323–8.
31. 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. In: Krasnoproshin V, Ablameyko S, Sadykhov R, editors. Pattern Recognition and Information Processing: Proceedings of the Tenth International Conference, 19–21 May 2009, Minsk, Belarus. Minsk: Belarusian State University; 2009. p. 140–5.
32. Tixier F, Hatt M, Le Rest CC, Le Pogam A, Corcos L, Visvikis D. Reproducibility of tumor uptake heterogeneity characterization through textural feature analysis in 18F-FDG PET. J Nucl Med. 2012;53:693–700.
33. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011;144:646–74.
34. Gerlinger M, Swanton C. How darwinian models inform therapeutic failure initiated by clonal heterogeneity in cancer medicine. Br J Cancer. 2010;103:1139–43.
35. Orlhac F, Soussan M, Maisonobe JA, Garcia CA, Vanderlinden B, Buvat I. Tumor texture analysis in 18F-FDG PET: relationships between texture parameters, histogram indices, standardized uptake values, metabolic volumes, and total lesion glycolysis. J Nucl Med. 2014;55:414–22.
36. Zhang H, Tan S, Chen W, Kligerman S, Kim G, D'Souza WD, et al. Modeling pathologic response of esophageal cancer to chemoradiation therapy using spatial-temporal 18F-FDG PET features, clinical parameters, and demographics. Int J Radiat Oncol Biol Phys. 2014;88:195–203.
37. Schinagl DA, Span PN, van den Hoogen FJ, Merkx MA, Slootweg PJ, Oyen WJ, et al. Pathology-based validation of FDG PET segmentation tools for volume assessment of lymph node metastases from head and neck cancer. Eur J Nucl Med Mol Imaging. 2013;40:1828–35.
38. Arens AI, Troost EG, Hoeben BA, Grootjans W, Lee JA, Grégoire V, et al. Semiautomatic methods for segmentation of the proliferative tumour volume on sequential FLT PET/CT images in head and neck carcinomas and their relation to clinical outcome. Eur J Nucl Med Mol Imaging. 2014;41:915–24.