BJR

https://doi.org/10.1259/bjr.20170546

Received: 23 July 2017 | Revised: 11 November 2017 | Accepted: 23 November 2017

Cite this article as:
Nakajo M, Jinguji M, Shinaji T, Nakajo M, Aoki M, Tani A, et al. Texture analysis of ^{18}F-FDG PET/CT for grading thymic epithelial tumours: usefulness of combining SUV and texture parameters. *Br J Radiol* 2018; **90**: 20170546.

FULL PAPER

Texture analysis of ^{18}F-FDG PET/CT for grading thymic epithelial tumours: usefulness of combining SUV and texture parameters

[1]MASATOYO NAKAJO, MD, PhD, [1]MEGUMI JINGUJI, MD, PhD, [2]TETSUYA SHINAJI, MSc, [3]MASAYUKI NAKAJO, MD, PhD, [4]MASAYA AOKI, MD, PhD, [1]ATSUSHI TANI, MD, PhD, [4]MASAMI SATO, MD, PhD and [1]TAKASHI YOSHIURA, MD, PhD

[1]Department of Radiology, Kagoshima University, Graduate School of Medical and Dental Sciences, Kagoshima, Japan
[2]Department of Nuclear Medicine, University of Würzburg, Würzburg, Germany
[3]Department of Radiology, Nanpuh Hospital, Kagoshima, Japan
[4]Department of Thoracic Surgery, Kagoshima University, Graduate School of Medical and Dental Sciences, Kagoshima, Japan

Address correspondence to: Dr Masatoyo Nakajo
E-mail: *toyo.nakajo@dolphin.ocn.ne.jp*

Objective: To retrospectively investigate the standardized uptake value (SUV)-related and heterogeneous texture parameters individually and in combination for differentiating between low- and high-risk ^{18}Fluorone-fludeoxyglucose (^{18}F-FDG)-avid thymic epithelial tumours (TETs) with positron emission tomography (PET)/CT.

Methods: SUV-related and 6 texture parameters (entropy, homogeneity, dissimilarity, intensity variability, size-zone variability and zone percentage) were compared between 11 low-risk and 23 high-risk TETs (metabolic tumour volume >10.0 cm^3 and SUV ≥2.5). Diagnostic performance was evaluated by receiver operating characteristic analysis. The diagnostic value of combining SUV and texture parameters was examined by a scoring system.

Results: High-risk TETs were significantly higher in SUVmax (p = 0.022), entropy (p = 0.038), intensity variability (p = 0.041) and size-zone variability (p = 0.045) than low-risk TETs. Diagnostic accuracies of these 4 parameters, dissimilarity and zone percentage which also showed significance in receiver operating characteristic analysis ranged between 64.7 and 73.5% without significant differences in AUC (range; 0.71 to 0.75) ($p \geq 0.05$ each). Each parameter was scored as 0 (negative for high-risk) or 1 (positive for high-risk) according to each threshold criterion, then scores were summed [0 or 1 for low-risk TETs (median; 1); ≥2 for high-risk TETs (median; 4)]. The sensitivity, specificity and accuracy of detecting high-risk TETs were 100, 81.8 and 94.1%, respectively, with an AUC of 0.99.

Conclusion: The diagnostic performances of individual SUVmax and texture parameters were relatively low. However, combining these parameters can significantly increase diagnostic performance when differentiating between relatively large low- and high-risk ^{18}F-FDG-avid TETs.

Advances in knowledge: Combined use of SUVmax and texture parameters can significantly increase the diagnostic performance when differentiating between low- and high-risk TETs.

INTRODUCTION

Thymic epithelial tumours (TETs) are rare tumours account for 50% of the anterior mediastinal masses; they are the most frequent mediastinal tumours in the adult population, and thymomas and thymic carcinomas are the most frequent histological subtypes.[1] Their prognoses depend mainly on their Masaoka staging and World Health Organization (WHO) histological classification in addition to resectability.[1–5] In the WHO classification, TETs are classified into two major categories; five types of thymomas (Type A, AB, B1, B2 and B3) and thymic carcinomas. The WHO classification scheme correlates with invasiveness: Types A and AB are usually clinically benign and encapsulated, Type B has a greater likelihood of invasiveness (especially Type B3), and thymic carcinoma is almost always invasive. This scheme has been shown to reflect the clinical features of TETs and to correlate with prognosis.[2] The proportion of cases with invasive tumours and poor prognosis tended to increase according to the tumour type in the following orders among low-risk tumours (Types A, AB, and B1), and high-risk tumours (Types B2, B3, and thymic carcinomas).[3] Therefore, pre-operative predictions based on WHO histological subtypes, in addition to Masaoka staging system of TETs may help determine if tumours can be

treated with surgical resection alone or if they require pre- or postoperative adjuvant treatments such as chemotherapy and radiotherapy.[1,5]

The CT and MRI findings of TETs help in differentiating among various subtypes of the tumours.[6–9] Type A tumours have a higher prevalence of smooth contour and round shape than any other types of TETs.[8] On the other hand, thymic carcinomas show a higher prevalence of irregular contours, necrotic component, heterogeneous enhancement and lymphadenopathy than any other types of TETs on both CT and MRI.[7,9] However, there are many overlapping features, which usually make it difficult to arrive at a correct diagnosis.

Conversely, positron emission tomography (PET) is a molecular imaging technique using the radiotracer that can provide detailed pictures of what is happening inside the body at the molecular and cellular level. Glucose analogue ^{18}Fluorone-fludeoxyglucose (^{18}F-FDG) uptake represents glucose metabolic activity and is widely used as a tracer for PET/CT in oncology.[10] ^{18}F-FDG PET has recently emerged as a tool for grading histological types and staging of TETs.[11–16]

More recently, PET image texture analysis was proposed to characterize the heterogeneity of tumour ^{18}F-FDG uptake.[17] Tumour ^{18}F-FDG uptake is usually heterogeneous.[18–20] Tumour heterogeneity is associated with cellular proliferation, necrosis, hypoxia and angiogenesis, all of these factors being related with malignancy, response to therapy, disease progression and poorer prognosis in many cancers.[21,22] For example, ^{18}F-FDG PET texture features were studied for predicting treatment response and clinical outcome in non-small cell lung cancer[23] and oesophageal cancer,[24] and differentiating between FDG-avid benign and metastatic adrenal tumours.[25] However, to our knowledge, only one report has investigated ^{18}F-FDG PET texture features to grade malignancy of TETs.[26]

The present study was performed to investigate the standardized uptake value (SUV)-related and heterogeneous texture parameters individually and in combination for differentiating between low- and high-risk ^{18}F-FDG-avid TETs with PET/CT.

METHODS AND MATERIALS

Patients

This retrospective study was approved by the institutional review board, and the need to obtain informed consent was waived. The inclusion criteria were: patients who underwent ^{18}F-FDG-PET/CT for evaluation of suspected mediastinal tumours from January 2011 to June 2016 and the pathologically diagnosed ^{18}F-FDG-avid TETs with metabolic tumour volume (MTV) of >10.0 cm^3 and SUV ≥2.5.

Imaging protocols

All patients were instructed to fast for ≥5 h before PET/CT, and the scans were performed using a Discovery 600M PET/CT (GE Medical Systems, Milwaukee, WI). ^{18}F-FDG was supplied as a 2-ml vial which contained 185 MBq at the assay date and time by an ^{18}F-FDG-delivery system (Nihon Med-Physics CO., Ltd., Tokyo, Japan). Thus, the amount of administrated ^{18}F-FDG could not be correctly fixed by per kilogram of the body mass. The administered ^{18}F-FDG dose ranged from 155.9 to 241.5 MBq [median, 200.5MBq, (interquartile range, 188.0–208.8 MBq)]. The mean plasma glucose level was 108 mg dl^{-1} (range, 88–160 mg dl^{-1}) just before ^{18}F-FDG injection. Image acquisition started 1 h after intravenous injection of ^{18}F-FDG, and the PET/CT images were obtained from brain to feet (acquisition time was 2.5 min per bed position, with 14 bed positions) after CT using a 16-slice CT scanner [slice thickness, 3.75 mm; pitch, 1.75 mm; 120 keV; auto mA (35–100 mA depending on patient body mass)]. Attenuation correction was performed using the CT data, and the acquired data were reconstructed using a three-dimensional ordered-subset expectation maximization algorithm (image matrix size, 192 × 192; 16 subsets, 2 iterations: VUE Point Plus). The reconstructed transaxial spatial resolution for PET was 5.1 mm full-width half-maximum (FWHM) in-plane.

Image analysis

All PET/CT images were analysed on an Advantage Windows Workstation (GE Healthcare, Milwaukee, WI). Two experienced radiologists in PET/CT, who knew the purpose of the study, but were blinded to clinical and pathological information interpreted the images in consensus to identify ^{18}F-FDG-avid tumours that were distinguishable from background uptake in the mediastinum. Another experienced radiologist in PET/CT obtained SUV-related parameters with reference to the interpreted results. Borders of the volume of interest (VOI) were set by manual adjustment to exclude adjacent physiological ^{18}F-FDG-avid structures. Tumour boundaries were then automatically contoured. MTV was defined as the total tumour volume (cm^3) with an SUV ≥2.5.[26] Thus, the area with SUV of <2.5 was not contained in MTV. SUVmax and SUVmean were the maximum and mean tissue concentrations in the structure delineated by the VOI divided by the activity injected per gram of body weight, respectively. Total lesion glycolysis (TLG) was

Figure 1. Flow chart of the study population. ^{18}F-FDG, ^{18}Fluorone-fludeoxyglucose; MTV, metabolic tumour volume; PET, positron emission tomography; SUV, standardized uptake value.

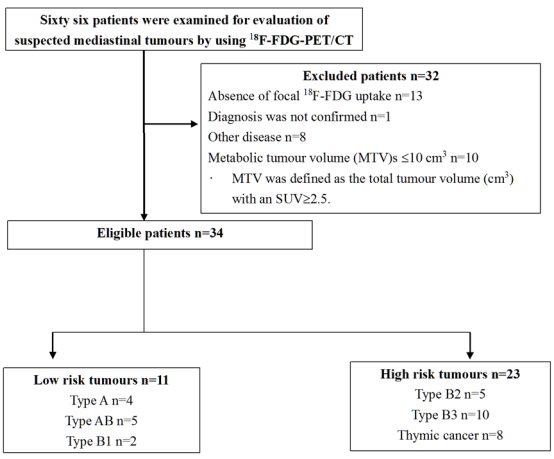


Table 1. Comparison of SUV-related and texture parameters among three histological classifications

Index	Low-risk thymoma (n = 11)			High-risk thymoma (n = 15)			Thymic cancer (n = 8)			p value
	Median	IQR	Range	Median	IQR	Range	Median	IQR	Range	
SUVmax	3.9	3.6–5.0	2.8–14.3	5.3	4.0–6.8	3.1–10.7	9.5	7.5–10.0	3.8–10.5	0.013
SUVmean	2.9	2.7–3.1	2.6–9.1	3.2	2.8–3.5	2.6–5.6	4.6	3.6–5.3	2.9–6.3	0.014
MTV (cm^3)	35.6	21.8–59.0	15.9–168.8	70.3	37.9–98.6	17.2–131.0	44.7	31.9–92.6	13.4–530.2	0.24
TLG	103.2	59.5–182.2	45.8–577.9	220.0	152.5–326.8	55.0–326.8	213.0	100.5–509.1	50.9–2863.1	0.22
Entropy	7.07	6.96–7.17	6.58–7.31	7.26	6.80–7.40	6.64–7.65	7.40	7.25–7.48	6.74–7.56	0.046
Homogeneity	0.212	0.202–0.221	0.194–0.264	0.207	0.174–0.234	0.134–0.241	0.193	0.166–0.203	0.139–0.264	0.20
Dissimilarity	8.12	7.23–8.38	5.61–9.21	8.15	7.44–10.82	6.74–12.11	9.64	8.55–11.97	5.92–14.39	0.070
IV	20.23	12.75–28.34	5.00–45.88	39.27	17.34–52.11	9.47–66.68	29.51	19.64–80.54	17.79–155.32	0.11
SZV	337.76	236.61–405.72	134.55–581.37	444.15	365.50–757.76	194.88–1063.74	403.41	284.99–1045.50	146.01–2778.96	0.13
ZP	0.27	0.23–0.30	0.10–0.41	0.21	0.16–0.28	0.08, 0.44	0.18	0.12–0.29	0.05–0.35	0.14

IQR, interquartile ranges; IV, intensity variability; MTV, metabolic tumour volume; SUV, standardized uptake value; SZV, size-zone variability; TLG, total lesion glycolysis; ZP, zone percentage.
p-values for comparison among three histological classifications (Kruskal–Wallis test).

defined as the SUVmean multiplied by the MTV. The software provided SUVmax, SUVmean and MTV values automatically.

Texture analysis

We chose the robust heterogeneity parameters; the second-order entropy, homogeneity, and dissimilarity as local parameters, and the intensity variability (IV), size-zone variability (SZV), and zone percentage (ZP) as regional parameters according to the other reports.[27–32] The details of calculating of these parameters were described in our previous report.[24] Entropy reflects irregularity in the grey level co-occurrence matrix, and the random distribution might show the high entropy values. Homogeneity measures the uniformity of the grey level co-occurrence matrix. If the image has little variation, then homogeneity is high, while the image has big variation then homogeneity is low. Dissimilarity is a measure that defines the variation of grey level pairs in an image. If the neighbouring pixels are very similar in their grey level values, then the dissimilarity in the grey level co-occurrence matrix is very low.

IV measures the similarity in pixel intensities and SZV measures the similarity in zone sizes throughout the image. If the grey level values are alike throughout the image, IV is small, and if the zone sizes are alike throughout the image, SZV is small. ZP measures the homogeneity and it is largest, when the size of the zones is 1 for all grey levels. Heterogeneous images tend to have larger zones of very different intensity pixels.

The intensity was rescaled using 64 discrete values. The rescaling was performed with the minimum to the maximum intensity range of each VOI to measure the heterogeneity of the intensity values of each tumour. The PET images were post-filtered with a Gaussian filter (FWHM: 5.1 mm), therefore, small variations can be regarded as representing heterogeneity rather than noise.[33] Because the texture features can be confounded by tumour volume effects in small volume tumours, especially those ≤10 cm^3, these texture analyses were only performed for MTVs >10 cm^3.[34] The program used to calculate the six texture parameters was implemented using Python computer language. The calculation program was executed on Mac OS X 10.11.3 with an Intel Core i7 (2.3 GHz) CPU and 4 GB memory.

Histological analyses and clinical staging

All clinical records including pathological reports were reviewed and the TETs were classified according to the WHO histological classification [thymoma types A, AB, B1, B2, B3 and thymic cancer (C)]. Types A, AB and B1 have been found to be less aggressive and have better prognoses than B2, B3 and C types.[2–4] Thus, all TETs were further grouped as low-risk (A, AB and B1) and high-risk (B2, B3 and C). The Masaoka clinical staging[35] was performed by pathological examination of tumour specimens and morphological imaging findings. Tumour size was presented as the maximum diameter on the transaxial CT image.

Scoring system

A scoring system was developed by combining imaging parameters which showed significance in receiver operating characteristics (ROC) analysis of diagnostic performance when

Table 2. Comparison of SUV-related and texture parameters and summed score between low- and high-risk thymic epithelial tumours

Index	Low-risk tumours (n = 11)			High-risk tumours (n = 23)			p value
	Median	IQR	Range	Median	IQR	Range	
SUVmax	3.9	3.6–5.0	2.8–14.3	6.4	4.2–9.5	3.1–10.7	0.022
SUVmean	2.9	2.7–3.1	2.6–9.1	3.0	3.0–4.7	2.6–6.3	0.034
MTV (cm^3)	35.6	21.8–59.0	15.9–168.8	54.6	33.2–92.8	13.4–530.2	0.11
TLG	103.2	59.5–182.2	45.8–577.9	220.0	117.9–390.5	50.9–2863.1	0.087
Entropy	7.07	6.96–7.17	6.58–7.31	7.29	6.90–7.44	6.64–7.65	0.038
Homogeneity	0.212	0.202–0.221	0.194–0.264	0.197	0.174–0.225	0.134–0.264	0.16
Dissimilarity	8.12	7.23–8.38	5.61–9.21	8.94	8.00–10.85	5.92–14.39	0.053
IV	20.23	12.75–28.34	5.00–45.88	37.33	18.78–59.54	9.47–155.32	0.041
SZV	337.76	236.61–405.72	134.55–581.37	425.50	363.27–757.76	146.01–2778.96	0.045
ZP	0.27	0.23–0.30	0.10–0.41	0.20	0.14–0.28	0.05–0.44	0.053
Summed score	1	0–1	0–2	4	3–4	2–5	<0.001

IQR, interquartile ranges; IV, intensity variability; MTV, metabolic tumour volume; SUV, standardized uptake value; SZV, size-zone variability; TLG, total lesion glycolysis; ZP, zone percentage.
p-values for comparison between low- and high-risk tumours (Mann–Whitney U test).

discriminating between low- and high-risk TETs. First, each parameter was scored as 0 (negative for high-risk) and 1 (positive for high-risk) using each threshold criterion. Second, their scores were summed, and the summed scores were used for differentiation between low-and high-risk TETs.

Statistical analysis

The Mann–Whitney U or Kruskal–Wallis test was used to assess differences in numerical variables between low- and high-risk TET groups or among the low-risk thymoma (A, AB and B1), high-risk thymoma (B2 and B3) and thymic cancer groups and clinical stages. When the p-values from the Kruskal–Wallis test were statistically significant, the non-parametric multiple comparison test was used to determine which group differed from the others. The Spearman rank test was used to examine the correlation between two quantitative variables. To examine the diagnostic performance of each parameter when discriminating between low- and high-risk TETs, ROC analysis was conducted, and the best cut-off point was determined by the Youden index.[36] The statistical significance of the difference between the areas under the ROC curves (AUCs) was analysed using the DeLong method.[37] Data are presented as medians and interquartile ranges. A $p < 0.05$ was considered statistically significant, and all p-values presented were two-tailed. The statistical analyses were performed using MedCalc statistical software (MedCalc, Mariakerke, Belgium) and SPSS® statistics 22 (IBM Corp., New York, NY; formerly SPSS Inc., Chicago, IL).

RESULTS

Tumour characteristics

66 patients underwent ^{18}F-FDG-PET/CT for suspected mediastinal tumour during the defined period. Among them, 32 patients were excluded; 13 lacked focal ^{18}F-FDG uptake in the mediastinal tumours including 5 thymomas, 1 lacked a final diagnosis, 8 had other diseases, and 10 with a TET of MTV ≤10.0 cm^3. Finally, 34 patients (19 males, 15 females; mean age ± standard deviation, 66 ± 15 years; age range, 25–88 years) were eligible for analysis (Figure 1). Characteristics of these 34 individuals are shown in (Supplementary Table 1, Supplementary material available online) (26 TETs were diagnosed by surgical excision, 8 by percutaneous biopsy). There were 11 low-risk (4 A, 5 AB and 2 B1) and 23 high-risk (5 B2, 10 B3 and 8 C) TETs, and 17 Stage I, 5 Stage II, 1 Stage III, 5 Stage IVa and 6 Stage IVb patients. There was no significant difference in tumour size between the low-risk [60 mm, (interquartile range, 50.0–74.3)] and high-risk groups [80 mm, (interquartile range, 60.0–87.5)] (p = 0.22).

Parameter comparisons among histological types and clinical stages

The results among three (low-risk thymoma, high-risk thymoma and thymic cancer) groups, and between low- and high-risk TET groups are shown in Tables 1 and 2. Both the SUVmax and SUVmean among the three groups were significantly different (p = 0.013 and 0.014). Thymic cancers showed significantly higher SUVmax and SUVmean than low-risk thymomas (SUVmax: p = 0.010; SUVmean: p = 0.011); there was no significant difference in SUVmax and SUVmean between thymic cancers and high-risk thymomas (SUVmax: p = 0.19; SUVmean: p = 0.13), and between low- and high-risk thymomas (SUVmax: p = 0.49; SUVmean: p = 0.70). Neither MTV nor TLG was significantly different among the three groups. Entropy showed significant differences among the 3 groups (p = 0.046). Thymic cancers showed significantly higher entropy than low-risk thymomas (p = 0.040); there was no significant difference in entropy between thymic cancers and high-risk thymomas (p = 0.53), and between low- and high-risk thymomas (p = 0.48). No significant differences were seen in other texture parameters among the three groups ($p \geq 0.05$ each).

The high-risk TET group showed significantly higher SUVmax (p = 0.022), SUVmean (p = 0.034), entropy (p = 0.038), IV (p = 0.041) and SZV (p = 0.045) than the low-risk TET group. No significant differences were seen in other texture parameters between the low- and high-risk TET groups.

No significant differences were seen among the clinical stages for any of the parameters ($p \geq 0.071$) (Table 3).

Parameter diagnostic performances to differentiate between low- and high-risk TETs

The AUCs were 0.71 for dissimilarity and ZP, 0.72 for entropy, IV and SZV, 0.73 for SUVmean and 0.75 for SUVmax (p-value range; 0.013–0.031). The diagnostic performances of these parameters for detecting high-risk TETs using individual optimal cutoff values are summarized in Table 4. Their sensitivities, specificities, accuracies were similar and ranged from 56.5 to 78.3%, 63.6 to 100% and 64.7 to 73.5%, respectively. No significant differences were seen in AUC among them ($p \geq 0.05$ each).

Diagnostic performance by a scoring system

The scoring system was developed using six parameters (SUVmax, entropy, dissimilarity, IV, SZV and ZP) which showed statistical significances in ROC analysis. The relationship between summed scores and WHO histological type, and between summed scores and clinical stages are shown in Tables 5 and 6, respectively. The relationships between the individual parameter score and summed score of the individual patients are shown in Supplemental Table 2.

The summed scores were higher in the high-risk TET group than the low-risk TET group [4.0, (interquartile range, 3–4) *vs* 1.0 (interquartile range, 0–1); $p < 0.001$] (Table 2). When we applied a summed score 0 or 1 to low-risk TETs and ≥2 to high-risk TETs, the AUC was 0.99 [95% confidence interval (CI) (0.88–1.00)], and the sensitivity, specificity, and accuracy to detect high-risk TETs were 100% [23/23, 95% CI (85.2–100)], 81.8% [9/11, 95% CI (48.2–97.7)] and 94.1% [32/34, 95% CI (80.3–99.3)], respectively (Table 4). The AUC of the summed score was significantly higher than the AUC of each of the six parameters alone (p; 0.004–0.010).

The summed scores showed significant differences among the clinical stages (p = 0.016) (Table 3). The summed scores showed significantly higher in Stage IV than Stage I (p = 0.012); there was no significant difference in summed score between Stage I and Stage II (p = 0.49), between Stage I and Stage III (p = 0.22), between Stage II and Stage III (p = 0.41), between Stage II and Stage IV (p = 0.69) and between Stage III and Stage IV (p = 0.96). Significant correlations were found between the summed scores and WHO histological classification (ρ; 0.79, $p < 0.001$) and between the summed scores and clinical stages (ρ; 0.54, $p < 0.001$). Correlation coefficients between various parameters are shown in Supplemental Table 3. In general, correlations between two of the same category parameters [SUVmax and SUVmean (ρ; 0.93), MTV and TLG (ρ; 0.94), three local parameters and three regional parameters] were significant and high or moderate, but correlations between the different category parameters were not

Table 3. Comparison of SUV-related and texture parameters and summed score among Masaoka's clinical stages

	I (n = 17)			II (n = 5)			III (n = 1)	IV (n = 11)			p value
	Median	IQR	Range	Median	IQR	Range	Median	Median	IQR	Range	
SUVmax	4.0	3.9–5.3	2.8–14.3	5.5	3.5–7.8	3.3–9.7	9.6	7.8	5.6–9.5	5.5–10.5	0.086
SUVmean	2.9	2.8–3.3	2.6–9.1	3.5	2.7–4.2	2.7–5.4	5.6	3.5	3.2–5.0	2.7–6.3	0.071
MTV (cm^3)	45.4	31.9–89.5	15.9–168.8	26.2	19.0–168.5	13.4–530.2	87.3	49.0	33.2–89.9	17.2–102.4	0.65
TLG	147.4	89.2–359.2	45.8–577.9	70.7	54.9–841.5	50.9–2863.1	488.9	220.5	130.0–273.5	55.0–581.5	0.35
Entropy	7.13	6.92–7.25	6.58–7.40	7.38	7.10–7.51	6.93–7.55	6.65	7.39	7.18–7.50	6.73–7.65	0.071
Homogeneity	0.207	0.201–0.228	0.157–0.264	0.195	0.138–0.230	0.134–0.264	0.235	0.196	0.177–0.209	0.149–0.239	0.23
Dissimilarity	8.15	7.14–8.78	5.61–12.11	8.81	7.05–11.70	5.92–14.39	7.25	9.56	8.14–10.85	6.96–13.08	0.22
IV	27.23	15.60–45.96	5.00–66.68	20.70	17.62–66.83	14.40–155.32	41.16	21.68	18.05–58.94	9.47–88.78	0.87
SZV	401.23	298.61–610.18	134.55–1063.74	359.83	263.41–1027.85	146.01–2778.96	394.66	409.53	375.82–615.31	194.88–1460.05	0.90
ZP	0.24	0.19–0.29	0.10–0.41	0.25	0.17–0.37	0.05–0.44	0.08	0.21	0.15–0.28	0.09–0.37	0.29
Summed score	2	0.75–4.0	0–4	3	1.5–4	0–4	4	4	3.25–4.75	2–5	0.016

IQR, interquartile ranges; IV, intensity variability; MTV, metabolic tumour volume; SUV, standardized uptake value; SZV, size-zone variability; TLG, total lesion glycolysis; ZP, zone percentage.
p-values for comparison among stages (Kruskal–Wallis test).

Table 4. Diagnostic performances of SUVs and texture parameters and summed score to detect high-risk thymic epithelial tumours

Parameter	Threshold value criterion	Sensitivity (%)	Specificity (%)	Accuracy (%)	AUC
SUVmax	>5.5	56.5 (13/23) 34.5–76.8[a]	90.9 (10/11) 58.7–99.8[a]	67.6 (23/34) 49.5–82.6[a]	0.75 0.57–0.88[a]
SUVmean	>3.3	56.5 (13/23) 34.5–76.8[a]	90.9 (10/11) 58.7–99.8[a]	67.6 (23/34) 49.5–82.6[a]	0.73 0.55–0.87[a]
Entropy	>7.22	65.2 (15/23) 42.7–83.6[a]	90.9 (10/11) 58.7–99.8[a]	73.5 (25/34) 55.6–87.1[a]	0.72 0.54–0.86[a]
Dissimilarity	>9.21	47.8 (11/23) 26.8–69.4[a]	100 (11/11) 71.5-100[a]	64.7 (22/34) 46.5–80.3[a]	0.71 0.53–0.85[a]
IV	>34.56	56.5 (13/23) 34.5–76.8[a]	90.9 (10/11) 58.7–99.8[a]	67.6 (23/34) 49.5–82.6[a]	0.72 0.54–0.86[a]
SZV	>354.49	78.3 (18/23) 56.3–92.5[a]	63.6 (7/11) 30.8–89.1[a]	73.5 (25/34) 55.6–87.1[a]	0.72 0.54–0.86[a]
ZP	≤0.213	65.2 (15/23) 42.7–83.6[a]	90.9 (10/11) 58.7–99.8[a]	73.5 (25/34) 55.6–87.1[a]	0.71 0.53–0.85[a]
Summed score	≥2	100 (23/23) 85.2-100[a]	81.8 (9/11) 48.2–97.7[a]	94.1 (32/34) 80.3–93.3[a]	0.99 0.88–1.00[a]

IV, intensity variability; SUV, standardized uptake value; SZV, size-zone variability; ZP, zone percentage.
[a]95% confidence interval.

significant or low. Concerning six parameters which showed statistical significances in ROC analysis, SUVmax had no significant correlations with entropy, dissimilarity and SZV, and significant but low correlations with IV (ρ; 0.38) and ZP (ρ; −0.53). Entropy had no significant correlations with IV and SZV, and significant but low correlations with dissimilarity (ρ; 0.60) and ZP (ρ; 0.46). Dissimilarity had significant, but low correlations with IV (ρ; −0.54), SZV (ρ; −0.51) and ZP (ρ; 0.51). IV had significant and moderate correlations with SZV (ρ; 0.80) and ZP (ρ; −0.78). The correlation between SZV and ZP was significant, but low (ρ; −0.55).

Representative images of low- and high-risk TETs are shown in Figures 2 and 3, respectively.

DISCUSSION

The WHO histological classification correlates with invasiveness and prognosis and has important implications for treatment of TETs.[2,4] We examined the diagnostic performances of SUV-related and texture parameters for differentiation between low- and high-risk ^{18}F-FDG-avid TETs. Herein, SUVmax and SUVmean were significantly higher in thymic cancers than low- and high-risk thymomas and higher in the high-risk than the low-risk TET group, but they were not significantly different between low- and high-risk thymomas. Discordant results were reported previously about the utility of SUVmax and/or SUVmean for differentiation between two of these TETs or groups probably due to small study populations and different proportions of histological types in individual studies.[11–13,15,16,26] Treglia et al[38] performed a meta-analysis of SUVmax and demonstrated statistically significant differences in SUVmax between low- and high-risk thymomas, and between low-risk or high-risk thymomas and thymic cancers, although a SUVmax cut-off for discriminating between groups could not be defined due to the large overlap of these semi-quantitative values among the different TETs.

There has been controversy regarding the utility of SUVmax to predict clinical stage. Herein, there was no significant difference in SUVmax among the clinical stages, in concordance with previous reports.[14,15] Others have reported a significant correlation between SUVmax and clinical stages[11] and significant

Table 5. The number of each histological type of thymic epithelial tumours which were classified by the summed score

Histological type	Score 0	Score 1	Score 2	Score 3	Score 4	Score 5	Total
A	3	0	1	0	0	0	4
AB	1	3	1	0	0	0	5
B1	1	1	0	0	0	0	2
B2	0	0	2	0	3	0	5
B3	0	0	1	2	7	0	10
Thymic cancer	0	0	0	2	3	3	8
Total	5	4	5	4	13	3	34

Table 6. The number of each clinical stage of thymic epithelial tumours which were classified by the summed score

Clinical stage	Score 0	Score 1	Score 2	Score 3	Score 4	Score 5	Total
I	4	4	3	1	5	0	17
II	1	0	1	1	2	0	5
III	0	0	0	0	1	0	1
IV	0	0	1	2	5	3	11
Total	5	4	5	4	13	3	34

differences between Stage IV and Stage I or II, but not between Stage III and IV.[16]

Two reports are available regarding MTV and TLG in TETs; MTV above an SUV of 3.5 was significantly higher in Type B3 thymomas than other type thymomas and more related to advanced stage than SUVmax.[39] MTV and TLG with an SUV threshold of 2.5 were not related to histological type, whereas only TLG correlated with clinical staging.[40] Herein, although MTV and TLG with an SUV ≥2.5 were higher in high-risk TETs than low-risk ones, the difference was not significant. There were no significant differences in MTV and TLG among the clinical stages.

To our knowledge, there is one study that has previously investigated the ability of ^{18}F-FDG imaging heterogeneity features to predict the malignant nature of TETs.[26] They examined the relationships of various local- and regional-scale indices with the WHO histological classification and showed that some regional-scale indices were independent of SUVmax when discriminating between TET grades, suggesting the complementary value of SUVmax and heterogeneity indices in differentiating TET subgroups.

In our study, high-risk TETs showed significantly higher entropy, IV and SZV, as well as SUVmax, than low-risk TETs. Although there was significant difference in neither dissimilarity nor ZP between low- and high-risk TETs, these two parameters were significant in ROC analysis for discriminating between low- and high-risk TETs. Moreover, these six parameters showed similar diagnostic accuracies in spite of non-significant, low or moderate correlations between individual two parameters and demonstrated a difference on different aspects of tumour heterogeneity of TETs. Thus, combining these parameters may increase the diagnostic accuracy. Although SUVmean was significantly higher in high-risk TETs than in low-rtsk TETs, SUVmax and SUVmean were the same in the diagnostic performance and their correlation was very high (ρ; 0.93). Thus, we used SUVmax alone for scoring.

Although no significant differences in any of the SUV-related and texture parameters were seen among the clinical stages, the summed scores showed significant differences among the clinical stages (p = 0.016, Table 3), and significant correlations between the summed scores and clinical stages (p < 0.001). However, the correlation coefficient was low (0.54), and there were overlap of the summed score among the clinical stage (Table 6). Thus, it might be difficult to predict the individual clinical stage by combing these parameters. On the other had, the summed score was significantly higher in the high-risk TET group than the low-risk TET group (p < 0.001) and the correlation coefficient was high (0.79) between the two groups. When using the summed score, sensitivity and accuracy improve prominently because of decreasing the false negative cases (Table 4). Consequently, the diagnostic performance increased significantly. These results also suggest that the high-risk TETs have more different aspects of tumour heterogeneity than the low-risk TETs.

The study limitations were as follows: most of the texture features considered to quantify intratumoral heterogeneity were found to be significantly correlated with tumour volume. However, Hatt et al[34] have suggested that heterogeneity quantification in PET images using texture features can be confounded by tumour volume effects in small volume tumours, especially those ≤10 cm^3. As mentioned above, we only studied ^{18}F-FDG-avid TETs with MTV >10.0 cm^3 which resulted in a small study population consisted of 11 low-risk and 23 high-risk TETs. Therefore, the obtained results were only applicable to relatively large ^{18}F-FDG-avid TETs (larger than about 2.7 cm in spherical MTV diameter). In fact, any parameters could not predict high-risk TETs and clinical stages in 5 low-risk and 10 high-risk TETs with MTV ≤10 cm^3 (Supplemental Tables 4 and 5). We did not correct the partial volume effect. This effect typically

Figure 2. A 40-year-old male with type AB thymoma transaxial (a), coronal (b) and sagittal (c) ^{18}F-FDG PET/CT images. Green lines represent the borders of the VOI. The summed score is 1.

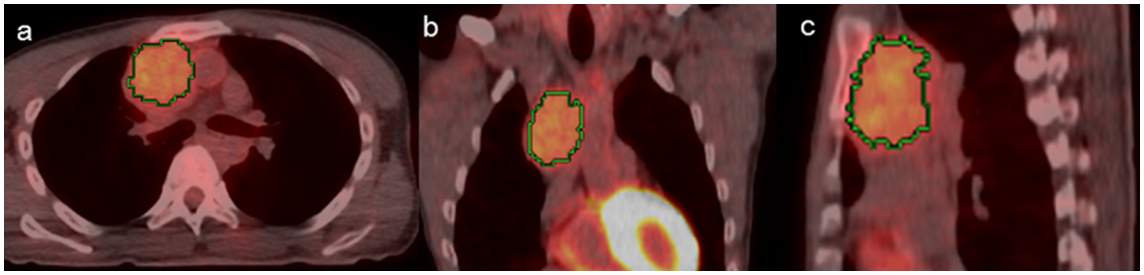
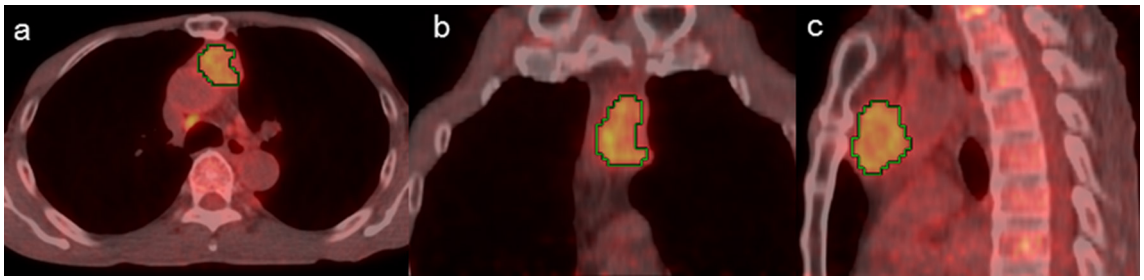


Figure 3. A 78-year-old male with thymic cancer transaxial (a), coronal (b) and sagittal (c) ^{18}F-FDG PET/CT images. Green lines represent the borders of the VOI. The summed score is 3. ^{18}F-FDG, ^{18}Fluorone-fludeoxyglucose; PET, positron emission tomography; VOI, volume of interest.



occurs whenever the tumour size is less than three-fold the FWHM of the reconstructed image resolution and probably, results in underestimates of the true maximum uptake value especially in patients with small tumour.[41] In our study, the reconstructed transaxial spatial resolution for PET was 5.1 mm FWHM and the size in diameter was larger than about 2.7 cm in spherical MTV, so the partial volume effect might not significantly affect the results. Although we did not examine the pathological correlates of these texture parameters, which are largely unknown and remain to be clarified in future, our results suggest that high-risk TETs have more different aspects of microscopic (local) and macroscopic (regional) tumour heterogeneity than low-risk TETs which was shown by the difference in summed scores between low- and high-risk TETs.

CONCLUSIONS

The diagnostic performances of individual SUVmax and texture parameters were relatively low. However, combining these parameters can significantly increase diagnostic performance when differentiating between relatively large low- and high-risk ^{18}F-FDG-avid TETs.

REFERENCES

1. Venuta F, Anile M, Diso D, Vitolo D, Rendina EA, De Giacomo T, et al. Thymoma and thymic carcinoma. *Eur J Cardiothorac Surg* 2010; **37**: 13–25. doi: https://doi.org/10.1016/j.ejcts.2009.05.038
2. Okumura M, Ohta M, Tateyama H, Nakagawa K, Matsumura A, Maeda H, et al. The World Health Organization histologic classification system reflects the oncologic behavior of thymoma: a clinical study of 273 patients. *Cancer* 2002; **94**: 624–32. doi: https://doi.org/10.1002/cncr.10226
3. Okumura M, Miyoshi S, Fujii Y, Takeuchi Y, Shiono H, Inoue M, et al. Clinical and functional significance of WHO classification on human thymic epithelial neoplasms: a study of 146 consecutive tumors. *Am J Surg Pathol* 2001; **25**: 103–10.
4. Chen G, Marx A, Chen WH, Yong J, Puppe B, Stroebel P, et al. New WHO histologic classification predicts prognosis of thymic epithelial tumors: a clinicopathologic study of 200 thymoma cases from China. *Cancer* 2002; **95**: 420–9. doi: https://doi.org/10.1002/cncr.10665
5. Jeong YJ, Lee KS, Kim J, Shim YM, Han J, Kwon OJ. Does CT of thymic epithelial tumors enable us to differentiate histologic subtypes and predict prognosis? *AJR Am J Roentgenol* 2004; **183**: 283–9. doi: https://doi.org/10.2214/ajr.183.2.1830283
6. Marom EM, Milito MA, Moran CA, Liu P, Correa AM, Kim ES, et al. Computed tomography findings predicting invasiveness of thymoma. *J Thorac Oncol* 2011; **6**: 1274–81. doi: https://doi.org/10.1097/JTO.0b013e31821c4203
7. Sadohara J, Fujimoto K, Müller NL, Kato S, Takamori S, Ohkuma K, et al. Thymic epithelial tumors: comparison of CT and MR imaging findings of low-risk thymomas, high-risk thymomas, and thymic carcinomas. *Eur J Radiol* 2006; **60**: 70–9. doi: https://doi.org/10.1016/j.ejrad.2006.05.003
8. Tomiyama N, Johkoh T, Mihara N, Honda O, Kozuka T, Koyama M, et al. Using the World Health Organization classification of thymic epithelial neoplasms to describe CT findings. *AJR Am J Roentgenol* 2002; **179**: 881–6. doi: https://doi.org/10.2214/ajr.179.4.1790881
9. Inoue A, Tomiyama N, Fujimoto K, Sadohara J, Nakamichi I, Tomita Y, et al. MR imaging of thymic epithelial tumors: correlation with World Health Organization classification. *Radiat Med* 2006; **24**: 171–81. doi: https://doi.org/10.1007/s11604-005-1530-4
10. von Schulthess GK, Steinert HC, Hany TF. Integrated PET/CT: current applications and future directions. *Radiology* 2006; **238**: 405–22. doi: https://doi.org/10.1148/radiol.2382041977
11. Sung YM, Lee KS, Kim BT, Choi JY, Shim YM, Yi CA. 18F-FDG PET/CT of thymic epithelial tumors: usefulness for distinguishing and staging tumor subgroups. *J Nucl Med* 2006; **47**: 1628–34.
12. Terzi A, Bertolaccini L, Rizzardi G, Luzzi L, Bianchi A, Campione A, et al. Usefulness of 18-F FDG PET/CT in the pre-treatment evaluation of thymic epithelial neoplasms. *Lung Cancer* 2011; **74**: 239–43. doi: https://doi.org/10.1016/j.lungcan.2011.02.018
13. Nakajo M, Kajiya Y, Tani A, Yoneda S, Shirahama H, Higashi M, et al. ^{18}FDG PET for grading malignancy in thymic epithelial tumors: significant differences in ^{18}FDG uptake and expression of glucose transporter-1 and hexokinase II between low and high-risk tumors: preliminary study. *Eur J Radiol* 2012; **81**: 146–51. doi: https://doi.org/10.1016/j.ejrad.2010.08.010
14. Endo M, Nakagawa K, Ohde Y, Okumura T, Kondo H, Igawa S, et al. Utility of 18FDG-PET for differentiating the grade of malignancy in thymic epithelial tumors. *Lung Cancer* 2008; **61**: 350–5. doi: https://doi.org/10.1016/j.lungcan.2008.01.003
15. Viti A, Bertolaccini L, Cavallo A, Fortunato M, Bianchi A, Terzi A. 18-Fluorine fluorodeoxyglucose positron emission tomography in the pretreatment evaluation of thymic epithelial neoplasms: a metabolic

biopsy confirmed by Ki-67 expression. *Eur J Cardiothorac Surg* 2014; **46**: 369–74. doi: https://doi.org/10.1093/ejcts/ezu030

16. Matsumoto I, Oda M, Takizawa M, Waseda R, Nakajima K, Kawano M, et al. Usefulness of fluorine-18 fluorodeoxyglucose-positron emission tomography in management strategy for thymic epithelial tumors. *Ann Thorac Surg* 2013; **95**: 305–10. doi: https://doi.org/10.1016/j.athoracsur.2012.09.052
17. El Naqa I, Grigsby P, Apte A, Kidd E, Donnelly E, Khullar D, et al. Exploring feature-based approaches in PET images for predicting cancer treatment outcomes. *Pattern Recognit* 2009; **42**: 1162–71. doi: https://doi.org/10.1016/j.patcog.2008.08.011
18. Pugachev A, Ruan S, Carlin S, Larson SM, Campa J, Ling CC, et al. Dependence of FDG uptake on tumor microenvironment. *Int J Radiat Oncol Biol Phys* 2005; **62**: 545–53. doi: https://doi.org/10.1016/j.ijrobp.2005.02.009
19. van Baardwijk A, Bosmans G, van Suylen RJ, van Kroonenburgh M, Hochstenbag M, Geskes G, et al. Correlation of intra-tumour heterogeneity on 18F-FDG PET with pathologic features in non-small cell lung cancer: a feasibility study. *Radiother Oncol* 2008; **87**: 55–8. doi: https://doi.org/10.1016/j.radonc.2008.02.002
20. van Velden FH, Cheebsumon P, Yaqub M, Smit EF, Hoekstra OS, Lammertsma AA, et al. Evaluation of a cumulative SUV-volume histogram method for parameterizing heterogeneous intratumoural FDG uptake in non-small cell lung cancer PET studies. *Eur J Nucl Med Mol Imaging* 2011; **38**: 1636–47. doi: https://doi.org/10.1007/s00259-011-1845-6
21. Asselin MC, O'Connor JP, Boellaard R, Thacker NA, Jackson A. Quantifying heterogeneity in human tumours using MRI and PET. *Eur J Cancer* 2012; **48**: 447–55. doi: https://doi.org/10.1016/j.ejca.2011.12.025
22. Lu X, Kang Y. Hypoxia and hypoxia-inducible factors: master regulators of metastasis. *Clin Cancer Res* 2010; **16**: 5928–35. doi: https://doi.org/10.1158/1078-0432.CCR-10-1360
23. 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. doi: https://doi.org/10.2967/jnumed.112.107375
24. Nakajo M, Jinguji M, Nakabeppu Y, Nakajo M, Higashi R, Fukukura Y, et al. Texture analysis of 18F-FDG PET/CT to predict tumour response and prognosis of patients with esophageal cancer treated by chemoradiotherapy. *Eur J Nucl Med Mol Imaging* 2017; **44**: 206–14. doi: https://doi.org/10.1007/s00259-016-3506-2
25. Nakajo M, Jinguji M, Nakajo M, Shinaji T, Nakabeppu Y, Fukukura Y, et al. Texture analysis of FDG PET/CT for differentiating between FDG-avid benign and metastatic adrenal tumors: efficacy of combining SUV and texture parameters. *Abdom Radiol* 2017; **42**: 2882–9. doi: https://doi.org/10.1007/s00261-017-1207-3
26. Lee HS, Oh JS, Park YS, Jang SJ, Choi IS, Ryu JS. Differentiating the grades of thymic epithelial tumor malignancy using textural features of intratumoral heterogeneity via (18)F-FDG PET/CT. *Ann Nucl Med* 2016; **30**: 309–19. doi: https://doi.org/10.1007/s12149-016-1062-2
27. Hatt M, Tixier F, Cheze Le Rest C, Pradier O, Visvikis D. Robustness of intratumour ^{18}F-FDG PET uptake heterogeneity quantification for therapy response prediction in oesophageal carcinoma. *Eur J Nucl Med Mol Imaging* 2013; **40**: 1662–71. doi: https://doi.org/10.1007/s00259-013-2486-8
28. Galavis PE, Hollensen C, Jallow N, Paliwal B, Jeraj R. Variability of textural features in FDG PET images due to different acquisition modes and reconstruction parameters. *Acta Oncol* 2010; **49**: 1012–6. doi: https://doi.org/10.3109/0284186X.2010.498437
29. 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. doi: https://doi.org/10.2967/jnumed.111.099127
30. Gebejes A, Huertas R. Texture characterization based on grey-level co-occurrence matrix. *Proceedings ICTIC* 2013; **2**: 375–8.
31. 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. doi: https://doi.org/10.2967/jnumed.110.082404
32. Cheng NM, Fang YH, Lee LY, Chang JT, Tsan DL, Ng SH, et al. Zone-size nonuniformity of 18F-FDG PET regional textural features predicts survival in patients with oropharyngeal cancer. *Eur J Nucl Med Mol Imaging* 2015; **42**: 419–28. doi: https://doi.org/10.1007/s00259-014-2933-1
33. 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. doi: https://doi.org/10.2967/jnumed.113.129858
34. Hatt M, Majdoub M, Vallières M, Tixier F, Le Rest CC, Groheux D, et al. 18F-FDG PET uptake characterization through texture analysis: investigating the complementary nature of heterogeneity and functional tumor volume in a multi-cancer site patient cohort. *J Nucl Med* 2015; **56**: 38–44. doi: https://doi.org/10.2967/jnumed.114.144055
35. Masaoka A, Monden Y, Nakahara K, Tanioka T. Follow-up study of thymomas with special reference to their clinical stages. *Cancer* 1981; **48**: 2485–92. doi: https://doi.org/10.1002/1097-0142(19811201)48:11<2485::AID-CNCR2820481123>3.0.CO;2-R
36. Youden WJ. Index for rating diagnostic tests. *Cancer* 1950; **3**: 32–5. doi: https://doi.org/10.1002/1097-0142(1950)3:1<32::AID-CNCR2820030106>3.0.CO;2-3
37. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988; **44**: 837–45. doi: https://doi.org/10.2307/2531595
38. Treglia G, Sadeghi R, Giovanella L, Cafarotti S, Filosso P, Lococo F. Is (18)F-FDG PET useful in predicting the WHO grade of malignancy in thymic epithelial tumors? A meta-analysis. *Lung Cancer* 2014; **86**: 5–13. doi: https://doi.org/10.1016/j.lungcan.2014.08.008
39. Benveniste MF, Moran CA, Mawlawi O, Fox PS, Swisher SG, Munden RF, et al. FDG PET-CT aids in the preoperative assessment of patients with newly diagnosed thymic epithelial malignancies. *J Thorac Oncol* 2013; **8**: 502–10. doi: https://doi.org/10.1097/JTO.0b013e3182835549
40. Park SY, Cho A, Bae MK, Lee CY, Kim DJ, Chung KY. Value of 18F-FDG PET/CT for Predicting the World Health Organization malignant grade of thymic epithelial tumors: focused in volume-dependent parameters. *Clin Nucl Med* 2016; **41**: 15–20. doi: https://doi.org/10.1097/RLU.0000000000001032
41. Moon SH, Choi JY, Lee HJ, Son YI, Baek CH, Ahn YC, et al. Prognostic value of 18F-FDG PET/CT in patients with squamous cell carcinoma of the tonsil: comparisons of volume-based metabolic parameters. *Head Neck* 2013; **35**: 15–22. doi: https://doi.org/10.1002/hed.22904