Received:
13 July 2018

Revised:
22 September 2018

Accepted:
25 September 2018

<https://doi.org/10.1259/bjr.20180620>

Cite this article as:

Nakajo M, Jinguji M, Shinaji T, Tani A, Nakabeppu Y, Nakajo M, et al. ^{18}F -FDG-PET/CT features of primary tumours for predicting the risk of recurrence in thyroid cancer after total thyroidectomy: potential usefulness of combination of the SUV-related, volumetric, and heterogeneous texture parameters. *Br J Radiol* 2018; **91**: 20180620.

FULL PAPER

^{18}F -FDG-PET/CT features of primary tumours for predicting the risk of recurrence in thyroid cancer after total thyroidectomy: potential usefulness of combination of the SUV-related, volumetric, and heterogeneous texture parameters

¹MASATOYO NAKAJO, MD, PhD, ¹MEGUMI JINGUJI, MD, PhD, ²TETSUYA SHINAJI, MSc, ¹ATSUSHI TANI, MD, PhD, ³YOSHIKI NAKABEPPU, MD, PhD, ⁴MASAYUKI NAKAJO, MD, PhD, ⁵AKIHIRO NAKAJO, MD, PhD, ⁵SHOJI NATSUGOE, MD, PhD and ¹TAKASHI YOSHIURA, MD, PhD

¹Department of Radiology, Graduate School of Medical and Dental Sciences, Kagoshima University, Sakuragaoka, Kagoshima, Japan

²Department of Nuclear Medicine, University of Würzburg, Oberdürrbacher Str, Würzburg, Germany

³Department of Radiology, Kagoshima City Hospital, Uebara, Kagoshima, Japan

⁴Department of Radiology, Nanpuh Hospital, Nagata, Kagoshima, Japan

⁵Department of Digestive Surgery, Breast and Thyroid Surgery, Graduate School of Medical and Dental Sciences, Kagoshima University, Sakuragaoka, Kagoshima, Japan

Address correspondence to: Masatoyo Nakajo

E-mail: toyo.nakajo@dolphin.ocn.ne.jp

Objective: This retrospective study examined whether the primary tumour ^{18}F -FDG uptake features could predict the high-risk of recurrence in differentiated thyroid cancer (DTC) patients.

Methods: The enrolled 114 DTC patients underwent preoperative ^{18}F -FDG-PET/CT. SUVmax, SUVmean, metabolic tumour volume (MTV), total lesion glycolysis (TLG) and 6 texture parameters were obtained. Because the texture features can be confounded by the tumour volume effects, ^{18}F -FDG-avid tumour patients were divided into two groups (tumours with $\text{MTV} \leq 10.0 \text{ cm}^3$ and $>10.0 \text{ cm}^3$). Diagnostic performance for predicting the high-risk was evaluated by the area under the curve (AUC) by the ROC curve analysis.

Results: Eighty eight ^{18}F -FDG-avid tumours revealed more advanced-risk classification ($p = 0.015 \rightarrow 0.02$) than 26 ^{18}F -FDG-nonavid tumours, which yielded no high-risk patients. In the 44 $\text{MTV} > 10.0 \text{ cm}^3$ ^{18}F -FDG-avid

tumour patients, 8 high-risk patients revealed significantly higher SUVmax, SUVmean, MTV, TLG, intensity variability and size-zone variability, and lower zone percentage than 36 non-high-risk patients ($p < 0.001$ – 0.016). Their AUC (diagnostic accuracy) ranged between 0.77 (66%) and 0.92 (91%). When each parameter was scored as 0 (negative for high-risk) or 1 (positive for high-risk) according to each threshold criterion, and the 7 parameter summed score ≥ 5 was defined as high-risk, the accuracy was 93.2% (AUC: 0.98) in the $\text{MTV} > 10.0 \text{ cm}^3$ ^{18}F -FDG-avid tumour patients.

Conclusion: For primary $\text{MTV} > 10.0 \text{ cm}^3$ ^{18}F -FDG-avid DTCs, the combined use of SUV-related, volumetric, and texture parameters may be more useful to identify high-risk patients than the individual parameters.

Advances in knowledge: Combined use of SUV-related, volumetric, and texture parameters may be useful to identify high-risk DTC patients.

INTRODUCTION

Prognosis of differentiated thyroid cancer (DTC) patients who have either papillary or follicular thyroid carcinoma is more favorable than that of patients with other types of thyroid malignancies;¹ however, a recent study estimated the 10-year probability of death resulting from thyroid cancer to be 3%.² The American Thyroid Association (ATA) recommends the use of postoperative clinicopathological

characteristics for assessing the risk of recurrence and follow-up management of DTC patients.³ The ATA risk stratification system for predicting the recurrence of DTC could be a more practical method to identify the high-risk patients and provide more personalized management of the disease. If such parameters to predict the recurrence of DTC are present before therapy, they will contribute to follow-up management of DTC patients.

Glucose analog F-18-fluorodeoxyglucose (^{18}F -FDG) is widely used as a tracer of positron emission tomography (PET)/computed tomography (CT) in oncology.⁴ ^{18}F -FDG-PET/CT-based volumetric imaging parameters such as metabolic tumour volume (MTV) and total lesion glycolysis (TLG) were recently demonstrated to predict the lymph node metastasis in incidentally detected DTC patients.⁵

^{18}F -FDG tumour uptake is usually nonhomogeneous across the tumour due to necrosis, cell proliferation, microvessel density, and hypoxia,^{6–8} and the tumour heterogeneity was reported to be associated with disease progression, therapeutic response, and malignant behavior of the tumour.⁹ PET image texture analysis was proposed to characterize the heterogeneity of tumour ^{18}F -FDG uptake.^{10–13}

Nevertheless, to our knowledge, no study has assessed the use of ^{18}F -FDG-PET/CT features including ^{18}F -FDG-avidity, and volumetric and texture parameters to predict the high-risk of recurrence in DTC. If such pretreatment ^{18}F -FDG related parameters in the primary tumour are present to predict the high-risk of recurrence in DTC, they may help optimize the follow-up management of DTC patients.

The present study examined whether ^{18}F -FDG-avidity, standardized uptake value (SUV)-related, volumetric, and texture parameters of the primary tumour could predict the postoperative high-risk DTC patients.

METHODS AND MATERIALS

Patients

This retrospective study was approved by the institutional review board and the requirement for the written informed consent was waived off. We reviewed the clinical records of 115 primary thyroid cancer patients who underwent ^{18}F -FDG-PET/CT study before surgery in the period from January 2011 to December 2016. Medullary or anaplastic thyroid cancer patients were excluded from the study.

Imaging protocols

All patients were instructed to fast for ≥ 5 h prior to PET/CT using a Discovery 600M PET/CT system (GE Medical Systems, Milwaukee, WI). Image acquisition started 1 h after iv injection of ^{18}F -FDG (140–260 MBq). The mean plasma glucose level just before ^{18}F -FDG injection, was 103 (range, 74–161) mg dl^{-1} . Whole-body PET was performed (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]) from brain to feet. Emission data were reconstructed using a three-dimensional ordered-subset expectation maximization algorithm (16 subsets, 2 iterations: VUE Point Plus) to 192×192 matrices using the CT data for attenuation correction. The reconstructed transaxial spatial resolution for PET was 5.1 mm full-width half-maximum (FWHM) in-plane.

All patients underwent total thyroidectomy with neck dissection for suspicious or biopsy-proven metastatic cervical

lymphadenopathy between 2 and 111 (mean, 30 ± 20) days after ^{18}F -FDG-PET/CT examination.

Image analysis

PET/CT images were analysed on the Advantage Windows Workstation (GE Healthcare, Milwaukee, WI). Two radiologists experienced in PET/CT who knew the purpose of the study, but blinded to clinical and pathological information, interpreted the images by consensus to identify ^{18}F -FDG-avid tumours that were distinguishable from background uptake. Another radiologist experienced in PET/CT obtained the SUV-related and volumetric parameters with reference to the interpreted results. Borders of the volume of interest (VOI) were set by manual adjustment to exclude the adjacent physiological ^{18}F -FDG-avid structures. Tumour boundaries were then automatically contoured. MTV was the total tumour volume (cm^3) with an $\text{SUV} \geq 2.5$, which was used for the volumetric analysis in DTC with ^{18}F -FDG-PET/CT.⁵ Thus, the intra tumour areas with $\text{SUV} < 2.5$ were not also included in MTV. SUV_{max} and SUV_{mean} 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. TLG was measured as $\text{MTV} \times \text{SUV}_{\text{mean}}$. The software (Advantage Windows Workstation; GE Healthcare, Milwaukee, WI) automatically provided the SUV_{max} , SUV_{mean} , and MTV values.

Textural analysis

We chose robust heterogeneity parameters and the calculation methods of these parameters were previously described in detail^{14,15}: three local parameters (entropy, homogeneity, and dissimilarity) and three regional parameters (intensity variability [IV], size-zone variability [SZV], and zone percentage [ZP]). The intensity was rescaled using 64 discrete values with the minimum to maximum intensity range of each VOI. PET images were postfiltered with a Gaussian filter (FWHM: 5.1 mm). Thus, the small variations can be regarded as representing the heterogeneity rather than the noise.¹⁶ Because PET image heterogeneity quantification using textural features can be confounded by the tumour volume effects in small volume tumours, especially those $\leq 10 \text{ cm}^3$,¹⁷ patients were divided into two groups (tumours with $\text{MTV} \leq 10.0 \text{ cm}^3$ and those with $\text{MTV} > 10.0 \text{ cm}^3$). The program used to calculate these texture parameters was implemented using Python computer language. The calculation program was executed on Mac OS $\times 10.11.3$ with an Intel Core i7 (2.3 GHz) CPU and 4 GB memory.

Standards of reference

Primary lesions and regional nodal metastases were all pathologically diagnosed using the resected surgical specimens. The standard of reference for metastasis was based on the follow-up information including the measurements of serum thyroglobulin (Tg) levels and imaging examinations using neck ultrasound, CT, ^{131}I scan, or ^{18}F -FDG-PET/CT, which were repeated every 3–6 months. Patients were considered to have no clinical evidence of disease at the final follow-up if they had no abnormal serum Tg levels, and no structural evidence of the disease. Recurrence was defined as the abnormal elevation in serum Tg level or any evidence of the disease on imaging examination that

was detected during the follow-up period. The mean duration of follow-up was 26 ± 19 (range: 1–72) months. Patients were staged according to the eighth edition of the AJCC/UICC staging system.¹⁸

Risk stratification

Each patient was risk-stratified using the ATA risk of recurrence stratification system (low-, intermediate-, or high-risk of recurrence):³ Low-risk patients revealed complete tumour resection, no aggressive tumour histology, no invasion of locoregional tissues, no vascular invasion, and no metastases. Intermediate-risk patients revealed microscopic invasion into the perithyroidal tissue, cervical lymph node metastases, aggressive histology, or vascular invasion. High-risk patients revealed macroscopic gross extrathyroidal extension, incomplete tumour resection, or distant metastases.

Scoring system

A scoring system was developed by combining the significant imaging parameters using the receiver operating characteristics (ROC) curve analysis of diagnostic performance when discriminating between the low- or intermediate (non-high)-risk and high-risk for recurrence. First, each parameter was scored as 0 (negative for high-risk) or 1 (positive for high-risk) using each threshold criterion. Second, their scores were summed, and the summed scores were used for differentiation between the non-high-risk and high-risk patients.

Statistical analysis

The Mann-Whitney U test was used to assess the between-group differences with respect to the quantitative variables. The chi-squared test or Fisher's exact test was used to compare the categorical data. The Spearman-rank test was used to examine the correlation between the two quantitative variables. Areas under the ROC curve (AUCs) were calculated for individual parameters to assess the accuracy in predicting high-risk patients and the best cut-off point was determined by the Youden index.¹⁹ Differences between the AUCs were assessed for statistical significance using the DeLong method.²⁰

A value of $p < 0.05$ was considered as statistically significant; all p values presented are two-sided. Data were presented as medians and interquartile ranges. The statistical analyses were performed using the MedCalc Statistical Software (MedCalc Software, Mariakerke, Belgium).

RESULTS

^{18}F -FDG-avidity of primary tumours

One medullary thyroid cancer patient was excluded. Eventually, 114 patients (33 males, 81 females; mean age $\pm$ standard deviation; 58 ± 17 years [range, 15–89]) were eligible for analysis. The relationship between the clinicopathological factors and primary tumour ^{18}F -FDG uptake is summarized in Table 1. Twenty six and 88 patients revealed ^{18}F -FDG-nonavid and avid tumours, respectively. The maximum pathological diameter was significantly larger in the ^{18}F -FDG-avid tumours than the ^{18}F -FDG-nonavid tumours (21 mm, [interquartile range, 15–34] vs 13 mm [interquartile range, 6–20]; $p < 0.001$). The ^{18}F -FDG-avid

group revealed more advanced ATA-risk classification than the ^{18}F -FDG-nonavid one ($p = 0.02$). None of the ^{18}F -FDG-nonavid tumour patients were classified as high-risk. The ^{18}F -FDG-avidity was not significantly associated with age, sex, histopathology, lymph node metastases, or stage. Among 10 high-risk patients, 9 patients showed papillary carcinoma and 1 patient showed poorly differentiated carcinoma. Among 95 intermediate-risk patients, 91 patients showed papillary carcinoma, 2 patients showed follicular carcinoma and 2 patients showed poorly differentiated carcinoma. All 9 low risk patients showed papillary carcinoma. No significant correlation was observed between histology and ATA risk of recurrence stratification by the chi-squared test ($p = 0.58$).

Comparison of ^{18}F -FDG-avid tumour PET parameters between non-high-risk and high-risk patients (Table 2)

Of 88 ^{18}F -FDG-avid tumour patients, 10 and 78 were classified as high-risk and non-high-risk, respectively (Table 2). The high-risk patients revealed significantly higher SUVmax ($p = 0.012$), SUVmean ($p = 0.008$), MTV ($p < 0.001$), TLG ($p < 0.001$), IV ($p = 0.001$), and SZV ($p = 0.001$), and significantly lower ZP ($p = 0.003$) than the non-high-risk patients.

Of 44 MTV $\leq 10.0 \text{ cm}^3$ ^{18}F -FDG-avid tumour patients, 42 and 2 were non-high-risk and high-risk, respectively (Table 2). The statistical analysis was unable to be performed because of only 2 high-risk patients; one with MTV of 8.3 cm^3 and multiple lung metastases, and the other with MTV of 7.4 cm^3 and macroscopic invasion of the right parathyroid and recurrent laryngeal nerve.

Of 44 MTV $> 10.0 \text{ cm}^3$ ^{18}F -FDG-avid tumour patients, 8 and 36 were high-risk and intermediate-risk, respectively (Table 2). The high-risk patients revealed significantly higher SUVmax ($p = 0.016$), SUVmean ($p = 0.004$), MTV ($p < 0.001$), TLG ($p < 0.001$), IV ($p = 0.004$), and SZV ($p = 0.005$), and significantly lower ZP ($p = 0.008$) than the intermediate-risk patients.

Diagnostic performance of PET parameters to predict high-risk patients

About the 88 ^{18}F -FDG-avid tumour patients, the AUC and diagnostic accuracy to predict the high-risk patients were 0.75 ($p = 0.008$) and 71.6% for SUVmax, 0.76 ($p = 0.009$) and 83.0% for SUVmean, 0.87 ($p < 0.001$) and 90.9% for MTV, 0.85 ($p < 0.001$) and 87.5% for TLG, 0.82 ($p < 0.001$) and 85.2% for IV, 0.82 ($p < 0.001$) and 93.2% for SZV, and 0.79 ($p < 0.001$) and 81.8% for ZP, respectively (Table 3). The AUCs of MTV ($p = 0.019$) and TLG ($p = 0.016$) were significantly higher than that of SUVmax.

The above seven significant parameters were used for scoring. The summed scores were significantly higher in the 10 high-risk patients than the 78 non-high-risk patients (6.0, [interquartile range, 5–7] vs 0.0 [interquartile range, 0–2]; $p < 0.001$). Use of the summed score ≥ 5 as the criteria to identify the high-risk patients, was associated with 80% sensitivity, 96.2% specificity, and 94.3% accuracy (AUC: 0.85; Table 3). No significant differences were observed in AUC between the summed score and each of the seven parameters.

Table 1. Clinicopathological factors and primary tumour ^{18}F -FDG uptake in 114 patients with primary thyroid cancer

Factors	No.	Primary tumour FDG uptake		<i>p</i> value
		Visible	Nonvisible	
No. of cancers		88	26	
Age (y) ^a		60 (49, 75)	56 (45, 65)	0.11
Size (mm) ^a		21 (15, 34)	13 (6, 20)	<0.001
Sex				
Male	33	21	12	0.05
Female	81	67	14	
Histopathology				0.68
Papillary	109	84	25	
Follicular	2	2	0	
Poorly differentiated	3	2	1	
Nodal metastasis				0.21
Negative	21	14	7	
Positive	93	74	19	
AJCC stage				0.17
I	30	19	11	
II	1	1	0	
III	38	30	8	
IV	45	38	7	
ATA risk classification				0.02
Low	9	4	5	
Intermediate	95	74	21	
High	10	10	0	

Interquartile range values are shown in parentheses.

^aData are median value.

About the 44 MTV > 10.0 cm³ ^{18}F -FDG-avid tumour patients, the AUCs ($p = 0.001$ for all) and diagnostic accuracy to predict high risk were 0.77 and 65.9% for SUVmax, 0.83 and 79.5% for SUVmean, 0.92 and 86.4% for MTV, 0.91 and 79.5% for TLG, 0.83 and 84.1% for IV, 0.82 and 90.9% for SZV, and 0.80 and 72.7% for ZP, respectively (Table 4). No significant differences were observed in AUC between these parameters. The summed scores were significantly higher in the 8 high-risk patients than the 36 intermediate-risk patients (6.5, [interquartile range, 5–7] vs 1.0 [interquartile range, 0–3]; $p < 0.001$). Use of the summed score ≥ 5 as the criterion for high-risk patients was associated with 100% sensitivity, 91.7% specificity, and 93.2% accuracy (AUC: 0.98; Table 4). The AUC for the summed score was significantly higher than that for SUVmax ($p = 0.003$), SUVmean ($p = 0.047$), and ZP ($p = 0.005$), whereas no significant differences were observed between the AUC for the summed score and those for each of MTV, TLG, IV, and SZV. The relationship between the summed scores and ATA risk classification for all and MTV > 10.0 cm³ ^{18}F -FDG-avid tumours are summarized in Table 5.

About the seven AUC significant parameters, very high (≥ 0.90) and significant ($p < 0.001$) correlations were observed between

SUVmax and SUVmean, MTV and TLG, MTV and IV, MTV and SZV, TLG and IV, IV and SZV in all 88 ^{18}F -FDG-avid tumours (Supplementary Table 1), and between SUVmax and SUVmean, MTV and TLG, and MTV and IV in the 44 MTV > 10.0 cm³ ^{18}F -FDG-avid tumours (Supplementary Table 2), respectively.

During follow-up, 8 and 106 patients were alive with recurrence, and without recurrence, respectively. The recurrence rate was 0% for low-risk (0/9), 6% for intermediate-risk (6/95), and 20% for high-risk patients (2/10) ($p = 0.19$).

Representative images of the intermediate- and high-risk patients are presented in Figures 1 and 2, respectively.

DISCUSSION

We examined the potential use of various ^{18}F -FDG uptake features of primary tumours to identify the DTC patients who are at high-risk of postoperative recurrence.

^{18}F -FDG-AVIDITY OF PRIMARY THYROID CANCER

We noted 26 ^{18}F -FDG-nonavid and 88 ^{18}F -FDG-avid primary tumours. The ^{18}F -FDG-avidity was not related to age, sex,

Table 2. Comparison of PET parameters between low- or intermediate-risk and high-risk for recurrence on ^{18}F -FDG visible primary thyroid cancers

Parameters	All visible tumours (<i>n</i> = 88)						<i>p</i> value
	Low- or intermediate-risk (<i>n</i> = 78)			High-risk (<i>n</i> = 10)			
	Median	IQR	Range	Median	IQR	Range	
SUVmax	15.0	8.4, 23.6	3.6, 95.0	28.3	21.2, 35.3	4.7, 50.7	0.012
SUVmean	5.0	3.8, 7.0	2.5, 25.6	9.2	5.2, 12.2	3.1, 19.3	0.008
MTV (cm ³)	8.4	3.9, 19.1	0.5, 95.5	50.2	37.2, 93.9	7.4, 245.4	<0.001
TLG	47.6	19.1, 118.4	1.6, 2444.8	451.6	175.4, 1436.7	22.9, 2993.9	<0.001
Entropy	6.82	6.40, 7.24	4.94, 7.70	7.24	6.81, 7.37	6.43, 7.54	0.064
Homogeneity	0.18	0.15, 0.22	0.08, 0.35	0.23	0.22, 0.25	0.14, 0.27	0.077
Dissimilarity	11.8	9.4, 14.2	3.9, 23.5	8.6	8.3, 10.0	6.0, 13.8	0.079
Intensity variability	5.5	3.4, 9.9	1.1, 36.3	20.9	7.3, 32.0	4.0, 80.4	0.001
Size-Zone variability	92.7	54.1, 168.3	13.1, 554.1	386.2	144.2, 751.3	50.4, 1569.4	0.001
Zone percentage	0.33	0.17, 0.52	0.01, 0.77	0.08	0.04, 0.12	0.03, 0.50	0.003
	Visible tumours with MTV ≤ 10.0 cm ³ (<i>n</i> = 44)						
	Low- or intermediate-risk (<i>n</i> = 42)			High-risk (<i>n</i> = 2) ^a			
	Median	IQR	Range	Median	Range		
SUVmax	9.5	6.7, 18.6	3.6, 48.3	10.4	4.7, 16.7	NA	
SUVmean	4.1	3.6, 5.7	2.5, 17.7	4.1	3.1, 5.0	NA	
MTV (cm ³)	4.1	2.8, 5.8	0.5, 9.2	7.9	7.4, 8.3	NA	
TLG	19.4	10.6, 34.0	1.6, 65.5	32.2	22.9, 41.5	NA	
Entropy	6.67	6.26, 6.96	4.94, 7.51	6.91	6.44, 7.37	NA	
Homogeneity	0.15	0.14, 0.18	0.08, 0.27	0.17	0.14, 0.19	NA	
Dissimilarity	14.0	12.9, 16.0	8.6, 23.5	13.0	12.2, 13.8	NA	
Intensity variability	3.4	2.9, 4.5	1.1, 6.7	4.3	4.0, 4.6	NA	
Size-Zone variability	60.8	47.6, 78.4	13.1, 146.5	91.2	50.4, 132.1	NA	
Zone percentage	0.48	0.38, 0.58	0.04, 0.77	0.42	0.35, 0.50	NA	
	Visible tumours with MTV > 10.0 cm ³ (<i>n</i> = 44)						
	Intermediate-risk (<i>n</i> = 36)			High-risk (<i>n</i> = 8)			
	Median	IQR	Range	Median	IQR	Range	
SUVmax	20.0	13.8, 28.2	5.5, 95.0	30.9	24.5, 39.7	21.2, 50.7	0.016
SUVmean	6.5	4.7, 7.6	3.5, 25.6	10.1	8.2, 13.8	5.2, 19.3	0.004
MTV (cm ³)	20.1	14.3, 31.8	10.6, 95.5	73.8	41.8, 98.7	37.2, 245.4	<0.001
TLG	123.2	87.6, 200.2	38.2, 2444.8	744.5	213.1, 1348.9	200.7, 2993.9	<0.001
Entropy	7.19	6.61, 7.33	5.45, 7.70	7.24	6.99, 7.39	6.43, 7.54	0.438
Homogeneity	0.21	0.19, 0.24	0.15, 0.35	0.23	0.22, 0.25	0.22, 0.27	0.094
Dissimilarity	9.5	8.4, 10.6	3.9, 14.2	8.6	7.7, 8.7	6.0, 10.0	0.055
Intensity variability	10.6	8.1, 16.7	5.3, 36.3	26.4	17.2, 32.8	7.3, 80.4	0.004
Size-Zone variability	170.4	127.8, 279.4	46.0, 554.1	602.2	248.9, 788.4	144.2, 1569.4	0.005
Zone percentage	0.18	0.09, 0.29	0.01, 0.59	0.06	0.03, 0.09	0.03, 0.12	0.008

IQR, interquartile range; MTV, metabolic tumour volume; NA, not available; TLG, total lesion glycolysis; IQR, interquartile range;

Table 3. Diagnostic performance of ^{18}F -FDG PET parameters for predicting the high-risk for recurrence on all ^{18}F -FDG visible primary thyroid cancers

Parameter	Threshold value criterion	Sensitivity (%)	Specificity (%)	Accuracy (%)	AUC
SUVmax	>20.9	80.0 (8/10) 44.4–97.5 ^a	70.5 (55/78) 59.1–80.3 ^a	71.6 (63/88) 61.0–81.0 ^a	0.75 0.64–0.83 ^a
SUVmean	>7.7	70.0 (7/10) 34.8–93.3 ^a	84.6 (66/78) 74.7–91.8 ^a	83.0 (73/88) 73.4–90.1 ^a	0.76 0.66–0.85 ^a
MTV(cm ³)	>34.6	80.0 (8/10) 44.4–97.5 ^a	92.3 (72/78) 84.0–97.1 ^a	90.9 (80/88) 82.9–96.0 ^a	0.87 0.78–0.93 ^a
TLG	>168.6	80.0 (8/10) 44.4–97.5 ^a	88.5 (69/78) 79.2–94.6 ^a	87.5 (77/88) 78.7–93.6 ^a	0.85 0.75–0.91 ^a
Intensity variability	>14.3	70.0 (7/10) 34.8–93.3 ^a	87.2 (68/78) 77.7–93.7 ^a	85.2 (75/88) 76.1–91.9 ^a	0.82 0.72–0.89 ^a
Size-zone variability	>554.1	40.0 (4/10) 12.2–73.8 ^a	100 (78/78) 95.4–100 ^a	93.2 (82/88) 85.7–97.5 ^a	0.82 0.72–0.89 ^a
Zone percentage	≤0.122	80.0 (8/10) 44.4–97.5 ^a	82.1 (64/78) 71.7–91.8 ^a	81.8 (72/88) 72.2–89.2 ^a	0.79 0.69–0.87 ^a
Scoring system	≥5	80 (8/10) 63.1–100 ^a	96.2 (75/78) 89.2–99.2 ^a	94.3 (83/88) 87.2–98.1 ^a	0.85 0.76–0.92 ^a

AUC, area under the curve; MTV, metabolic tumour volume; TLG, total lesion glycolysis;

^a95% confidence interval.

histopathology, nodal metastasis and stage; however, it was significantly related to the tumour size and risk classification. Choi et al²¹ examined the clinical and pathological variables that were associated with false-negative preoperative ^{18}F -FDG-PET/CT findings in the patients with primary papillary thyroid cancers. They found that the tumour size influenced ^{18}F -FDG positivity; however, they observed no significant association of ^{18}F -FDG positivity with age, gender, or lymph node metastasis; these findings are consistent with our results. No studies have examined the use of ^{18}F -FDG-avidity to predict the high-risk of recurrence in DTC patients. In our study, the ^{18}F -FDG-avid

tumour group revealed more advanced ATA-risk classification as compared to the ^{18}F -FDG-nonavid group. Moreover, no patient with ^{18}F -FDG-nonavid tumour was classified as high-risk for recurrence. Thus, the ^{18}F -FDG-nonavid tumours may have a low tendency for recurrence.

^{18}F -FDG PARAMETERS TO PREDICT HIGH-RISK PATIENTS

Among the 88 ^{18}F -FDG-avid primary tumour patients, two SUV-related (SUVmax and SUVmean), two volumetric (MTV and TLG), and three regional texture (IV, SZV, and ZP)

Table 4. Diagnostic performance of ^{18}F -FDG PET parameters for predicting the high-risk for recurrence on ^{18}F -FDG visible primary thyroid cancers with MTV > 10.0 cm³

Parameter	Threshold value criterion	Sensitivity (%)	Specificity (%)	Accuracy (%)	AUC
SUVmax	>20.9	100 (8/8) 63.1–100 ^a	58.3 (21/36) 40.8–74.5 ^a	65.9 (29/44) 50.1–79.5 ^a	0.77 0.62–0.89 ^a
SUVmean	>7.7	87.5 (7/8) 47.3–99.7 ^a	77.8 (28/36) 60.8–89.9 ^a	79.5 (35/44) 64.7–90.2 ^a	0.83 0.68–0.92 ^a
MTV (cm ³)	>34.6	100 (8/8) 63.1–100 ^a	83.3 (30/36) 67.2–93.6 ^a	86.4 (38/44) 72.6–94.8 ^a	0.92 0.80–0.98 ^a
TLG	>168.6	100 (8/8) 63.1–100 ^a	75.0 (27/36) 57.8–87.9 ^a	79.5 (35/44) 64.7–90.2 ^a	0.91 0.78–0.98 ^a
Intensity variability	>18.2	75.0 (6/8) 34.9–96.8	86.1 (31/36) 70.5–95.3 ^a	84.1 (37/44) 69.9–93.4 ^a	0.83 0.69–0.93 ^a
Size-zone variability	>554.1	50.0 (4/8) 15.7–84.3 ^a	100 (36/36) 90.3–100 ^a	90.9 (40/44) 78.3–97.5 ^a	0.82 0.68–0.92 ^a
Zone percentage	≤0.122	100 (8/8) 63.1–100 ^a	66.7 (24/36) 49.0–81.4 ^a	72.7 (32/44) 57.2–85.0 ^a	0.80 0.65–0.91 ^a
Scoring system	≥5	100 (8/8) 63.1–100 ^a	91.7 (33/36) 77.5–98.2 ^a	93.2 (41/44) 81.3–98.6 ^a	0.98 0.88–1.00 ^a

AUC, area under the curve; MTV, metabolic tumour volume; TLG, total lesion glycolysis;

^a95% confidence interval.

Table 5. The number of patients with low- or intermediate-risk and high-risk for recurrence classified by the summed score

ATA risk classification	Score 0	Score 1	Score 2	Score 3	Score 4	Score 5	Score 6	Score 7	Total
All visible tumours ($n = 88$)									
Low- or intermediate-risk	46	11	9	6	3	1	2	0	78
High-risk	2	0	0	0	0	2	2	4	10
Total	48	11	9	6	3	3	4	4	88
Visible tumours with MTV $> 10.0 \text{ cm}^3$ ($n = 44$)									
Intermediate-risk	14	8	4	4	3	1	2	0	36
High-risk	0	0	0	0	0	3	1	4	8
Total	14	8	4	4	3	4	3	4	44

parameters were significantly different between the non-high-risk and high-risk patients.

The volumetric parameters have a potential to become valuable imaging biomarkers for the clinical staging, evaluation of therapeutic response, and prognostic assessment of patients with various cancers.^{22,23} Because they reflect the metabolic information in the entire tumour, they may be more accurate for tumour characterization than a single voxel measure such as SUVmax. Kim et al⁵ reported that the preoperative MTV and TLG with the SUV threshold of 2.5 were significant predictors of lateral lymph node metastasis in papillary thyroid carcinoma patients.

Three regional texture parameters (IV, SZV, and ZP) were used to extract regional tumor uptake characteristics, representing regional heterogeneity, such as variation of intensity between regions and in the size and alignment of homogeneous areas. IV measures the similarity in pixel intensities throughout the image and it is small, if the grey level values are alike throughout the image. SZV measures the similarity in zone sizes and it is small, if the zone sizes are alike throughout the image. 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 higher IV, SZV and lower ZP. On the other hand, the co-occurrence features such as entropy, homogeneity and dissimilarity analyse interrelationships between pairs of voxels, corresponding to the characterization of local heterogeneity.^{15,24} Our results showed that the regional, but not local parameters were significantly different between the

non-high-risk and high risk patients, suggesting that regional (macroscopic) rather than local (microscopic) heterogeneity may be more related to postoperative recurrence in DTC. Such a tendency was also noted in patients with esophageal cancer treated by chemoradiotherapy.^{15,24}

Because the textural features can be confounded by tumour volume effects in small volume tumours, we also performed these textural analyses for the tumours with MTV $\leq 10.0 \text{ cm}^3$ and MTV $> 10.0 \text{ cm}^3$, separately. In patients with MTV $> 10.0 \text{ cm}^3$, the above seven parameters were significantly different between the non-high-risk and high-risk patients; however, the statistical analysis was unable to be performed due to only 2 high-risk patients in 44 MTV $\leq 10.0 \text{ cm}^3$ tumour patients. Further studies will need to clarify the significance of these parameters to predict the high-risk patients with smaller tumours.

The seven parameters were significantly different between the non-high-risk and high-risk patients; however, these revealed a wide range of sensitivity, specificity, and accuracy in both all and MTV $> 10.0 \text{ cm}^3$ ^{18}F -FDG-avid tumours. As presented in the Supplemental Tables, although some regional parameters (IV or SZV) revealed a close correlation with the volumetric parameters (MTV or TLG), the other parameters reported low or moderate correlation with each other, which suggests that they could provide complementary information. Therefore, we examined whether the combined use of SUV-related, volumetric, and texture parameters could improve the diagnostic performance.

Figure 1. A 34-year-old female with DTC with an intermediate risk for recurrence. Transaxial (a), coronal (b), and sagittal (c) ^{18}F -FDG PET/CT images. Green lines represent the borders of the VOI. The SUVmax, SUVmean, MTV, TLG, IV, SZV and ZP were 5.5, 3.6, 39.8 cm^3 , 143.3, 18.0, 505.7 and 0.33, respectively. Among these 7 parameters, MTV alone was positive for high-risk for recurrence according to each threshold criterion. Thus, the summed score is 1.

Figure 2. A 85-year-old female with DTC with a high risk for recurrence. Transaxial (a), coronal (b), and sagittal (c) ^{18}F -FDG PET/CT images. Green lines represent the borders of the VOI. The SUVmax, SUVmean, MTV, TLG, IV, SZV and ZP were 44.0, 15.3, 93.9 cm^3 , 1436.7, 33.7, 742.3 and 0.047, respectively. These 7 parameters were all positive for high-risk for recurrence according to each threshold criterion. Thus, the summed score is 7.

The 88 ^{18}F -FDG-avid primary tumour patients showed no significant differences in AUC between the summed score and each parameter; however, in the MTV > 10.0 cm³ tumour patients, the summed score revealed the highest AUC value (0.98) with significant differences for each of certain parameters [SUVmax ($p = 0.003$), SUVmean ($p = 0.047$), and ZP ($p = 0.005$)], which suggests that the combined use of these parameters may potentially improve the diagnostic performance to predict the high-risk of recurrence in DTC patients, and may help optimize the follow-up management of DTC patients. Our results are in accordance with those recently reported by Nakajo et al that demonstrated the potential utility of combining such parameters for differentiation between ^{18}F -FDG-avid benign and metastatic adrenal tumours²⁵ and for grading thymic epithelial tumours.²⁶

Our study has the following limitations. We did not examine the pathological correlates of these texture parameters, which are largely unknown and remain to be clarified in future. Although the recurrence rate was higher in the high-risk patients (20%)

than the non-high-risk patients (5.8%), the follow-up period was insufficient for prognostic assessment; however, the ATA recurrence staging system is an established system for predicting the recurrence in DTC patients.²⁷

CONCLUSIONS

The ^{18}F -FDG-nonavid primary DTCs may have a low tendency for postoperative recurrence. For primary MTV > 10.0 cm³ ^{18}F -FDG-avid DTCs, the combined use of primary tumour SUV-related, volumetric, and texture parameters has a potential to significantly increase the ability to identify patients who are at a high-risk of postoperative recurrence.

ACKNOWLEDGMENT

The authors would like to thank Prof Chihaya Koriyama, the statistical specialist who belongs to Department of Epidemiology and Preventive Medicine, Kagoshima University, for the validation of the statistical analysis.

REFERENCES

- Sherman SI. Thyroid carcinoma. *Lancet* 2003; **361**: 501–11. doi: [https://doi.org/10.1016/S0140-6736\(03\)12488-9](https://doi.org/10.1016/S0140-6736(03)12488-9)
- Yang L, Shen W, Sakamoto N. Population-based study evaluating and predicting the probability of death resulting from thyroid cancer and other causes among patients with thyroid cancer. *J Clin Oncol* 2013; **31**: 468–74. doi: <https://doi.org/10.1200/JCO.2012.42.4457>
- American Thyroid Association (ATA) Guidelines Taskforce on Thyroid Nodules and Differentiated Thyroid Cancer, Cooper DS, Doherty GM, Haugen BR, Hauger BR, Kloos RT, et al. Revised American Thyroid Association management guidelines for patients with thyroid nodules and differentiated thyroid cancer. *Thyroid* 2009; **19**: 1167–214. doi: <https://doi.org/10.1089/thy.2009.0110>
- 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>
- Kim BH, Kim SJ, Kim K, Kim H, Kim SJ, Kim WJ, et al. High metabolic tumor volume and total lesion glycolysis are associated with lateral lymph node metastasis in patients with incidentally detected thyroid carcinoma. *Ann Nucl Med* 2015; **29**: 721–9. doi: <https://doi.org/10.1007/s12149-015-0994-2>
- 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>
- van Baardwijk A, Bosmans G, van Suylen RJ, van Kroonenburgh M, Hochstenbag M, Geskes G, et al. Correlation of intratumour heterogeneity on ^{18}F -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>
- 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>
- 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>
- 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>
- Cook GJ, O'Brien ME, Siddique M, Chicklore S, Loi HY, Sharma B, et al. Non-small cell lung cancer treated with erlotinib: heterogeneity of (^{18}F) -FDG uptake at PET-association with treatment response and prognosis. *Radiology* 2015; **276**: 883–93. doi: <https://doi.org/10.1148/radiol.2015141309>
- 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>
- Hyun SH, Kim HS, Choi SH, Choi DW, Lee JK, Lee KH, et al. Intratumoral heterogeneity of (^{18}F) -FDG uptake predicts survival in patients with pancreatic ductal adenocarcinoma. *Eur J Nucl Med Mol Imaging* 2016; **43**: 1461–8. doi: <https://doi.org/10.1007/s00259-016-3316-6>
- Hatt M, Tixier F, Cheze Le Rest C, Pradier O, Visvikis D, Hatt M. 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>
- Tixier F, Le Rest CC, Hatt M, Albarghach N, Pradier O, Metges JP, et al. Intratumor heterogeneity characterized by textural features on baseline ^{18}F -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>

16. Orlhac F, Soussan M, Maisonneuve 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>
17. 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>
5. The American Joint Committee on Cancer: the 8th edition of the AJCC cancer staging system. *Gastric Cancer* 2016; **21**((4)): 643–52. doi: <https://doi.org/doi:10.1007/s10120-017-0779-5>
19. Youden WJ. Index for rating diagnostic tests. *Cancer* 1950; **3**: 32–5. doi: [https://doi.org/10.1002/1097-0142\(1950\)3:1<1::AID-CNCR2820030106>3.0.CO;2-3](https://doi.org/10.1002/1097-0142(1950)3:1<1::AID-CNCR2820030106>3.0.CO;2-3)
20. 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>
21. Choi JW, Yoon YH, Yoon YH, Kim SM, Koo BS. Characteristics of primary papillary thyroid carcinoma with false-negative findings on initial (18)F-FDG PET/CT. *Ann Surg Oncol* 2011; **18**: 1306–11. doi: <https://doi.org/10.1245/s10434-010-1469-2>
22. 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. doi: <https://doi.org/10.2967/jnumed.111.099531>
23. Van de Wiele C, Kruse V, Smeets P, Sathekge M, Maes A. Predictive and prognostic value of metabolic tumour volume and total lesion glycolysis in solid tumours. *Eur J Nucl Med Mol Imaging* 2013; **40**: 290–301. doi: <https://doi.org/10.1007/s00259-012-2280-z>
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. Nakajo M, Jinguji M, Shinaji T, Nakajo M, Aoki M, Tani A, et al. Texture analysis of ¹⁸F-FDG PET/CT for grading thymic epithelial tumours: usefulness of combining SUV and texture parameters. *Br J Radiol* 2018; **91**: 20170546. doi: <https://doi.org/10.1259/bjr.20170546>
27. Tuttle RM, Tala H, Shah J, Leboeuf R, Ghossein R, Gonen M, et al. Estimating risk of recurrence in differentiated thyroid cancer after total thyroidectomy and radioactive iodine remnant ablation: using response to therapy variables to modify the initial risk estimates predicted by the new American Thyroid Association staging system. *Thyroid* 2010; **20**: 1341–9. doi: <https://doi.org/10.1089/thy.2010.0178>