

Variability of average SUV from several hottest voxels is lower than that of SUV_{max} and SUV_{peak}

E. Laffon · F. Lamare · H. de Clermont · I. A. Burger · R. Marthan

Received: 20 November 2013 / Revised: 17 April 2014 / Accepted: 6 May 2014 / Published online: 4 June 2014
© European Society of Radiology 2014

Abstract

Objectives To assess variability of the average standard uptake value (SUV) computed by varying the number of hottest voxels within an ¹⁸F-fluorodeoxyglucose (¹⁸F-FDG)-positive lesion. This SUV metric was compared with the maximal SUV (SUV_{max}: the hottest voxel) and peak SUV (SUV_{peak}: SUV_{max} and its 26 neighbouring voxels).

Methods Twelve lung cancer patients (20 lesions) were analysed using PET dynamic acquisition involving ten successive 2.5-min frames. In each frame and lesion, average SUV obtained from the $N=5, 10, 15, 20, 25$ or 30 hottest voxels (SUV_{max-N}), SUV_{max} and SUV_{peak} were assessed. The relative standard deviations (SDrs) from ten frames were calculated for each SUV metric and lesion, yielding the mean relative SD from 20 lesions for each SUV metric (SDr_N, SDr_{max} and SDr_{peak}), and hence relative measurement error and repeatability (MEr-R).

Results For each N , SDr_N was significantly lower than SDr_{max} and SDr_{peak}. SDr_N correlated strongly with N : $6.471 \times N^{-0.103}$ ($r=0.994$; $P<0.01$). MEr-R of SUV_{max-30} was 8.94–12.63 %

(95 % CL), versus 13.86–19.59 % and 13.41–18.95 % for SUV_{max} and SUV_{peak} respectively.

Conclusions Variability of SUV_{max-N} is significantly lower than for SUV_{max} and SUV_{peak}. Further prospective studies should be performed to determine the optimal total hottest volume, as voxel volume may depend on the PET system.

Key Points

- PET imaging provides functional parameters of ¹⁸F-FDG-positive lesions, such as SUV_{max} and SUV_{peak}.
- Averaging SUV from several hottest voxels (SUV_{max-N}) is a further SUV metric.
- Variability of SUV_{max-N} is significantly lower than SUV_{max} and SUV_{peak} variability.
- SUV_{max-N} should improve SUV accuracy for predicting outcome or assessing treatment response.
- An optimal total hottest volume should be determined through further prospective studies.

Keywords ¹⁸F-FDG PET · Uptake quantification · SUV repeatability · Peak SUV · Treatment response monitoring

E. Laffon (✉) · F. Lamare · H. de Clermont
Service de Médecine Nucléaire, Hôpital du Haut-Lévêque, CHU de Bordeaux, Avenue de Magellan, 33604 Pessac, France
e-mail: elaffon@u-bordeaux2.fr

E. Laffon · R. Marthan
Centre de Recherche Cardio-Thoracique, Université de Bordeaux 2, 33076 Bordeaux, France

E. Laffon · R. Marthan
INSERM U 1045, Centre de Recherche Cardio-Thoracique, 33076 Bordeaux, France

I. A. Burger
Division of Nuclear Medicine, Department Medical Radiology, University Hospital of Zurich, Ramistrasse 100, 8091 Zurich, Switzerland

Abbreviations

MEr	relative measurement error
MEr-R	relative measurement error and repeatability
SDr	relative standard deviation
R	repeatability
TV	tumour volume

Introduction

In quantitative ¹⁸F-fluorodeoxyglucose (¹⁸F-FDG) positron emission tomography (PET) imaging, several functional parameters of ¹⁸F-FDG-positive lesions can be extracted from images, thus allowing prognosis, staging and assessing the

treatment response in oncology [1]. The most widely used parameter is the maximal standard uptake value (SUV_{max} ; $g.ml^{-1}$) that is obtained from the voxel with the highest activity in a ^{18}F -FDG-positive lesion. It is equal to activity concentration ($kBq.ml^{-1}$) normalised to body activity concentration at injection (ID/W: injected dose to patient's weight; MBq/kg) [2]. Any SUV outcome should be considered with its measurement uncertainty. As an example, SUV repeatability (R; in %) represents the minimal relative change between two SUVs assessed at pre- and post-treatment examinations, which is required to consider a significant response to treatment [3, 4]. The lower the R value, the more accurate the assessment, which is the reason why lowering the R of the SUV is of major importance. It is well known that averaging SUV from several voxels leads to lowering R, and several averaging methods are available. A mean SUV can be obtained from outlining the ^{18}F -FDG-positive tissue and numerous segmentation methods have been proposed but, to date, no method has been recognised as a “gold standard” [5, 6]. SUV_{peak} represents the average SUV obtained over a fixed-size region of interest (ROI), which is usually defined as a 1-ml sphere (1.2-cm diameter) centred on a high-uptake part of the tumour [4]. However, various shapes, sizes and ROI locations have been proposed that can significantly affect the use of SUV_{peak} for assessing treatment response [7]. In a recent paper, Burger et al. [8] showed that average SUV obtained by pooling several (5–10) hottest voxels resulted in an improvement of R in comparison with that of SUV_{max} , i.e. the hottest voxel only. In comparison with SUV_{peak} , which involves several contiguous voxels belonging to a unique volume of interest (VOI), and hence are not mandatorily the hottest ones of the whole ^{18}F -FDG-positive lesion, this average SUV involves voxels that may be located in separate places of the ^{18}F -FDG-positive tissue, as is typically met when the uptake is heterogeneous (Fig. 1).

The aim of this work was to systematically assess the variability of the average SUV obtained from the $N=5, 10, 15, 20, 25$ or 30 hottest voxels (SUV_{max-N} ; $N \neq 1$), in ^{18}F -FDG-positive lesions of lung cancer patients, since the limitation of SUV_{max} has been shown before [8], but the direct impact of SUV_{max-N} on variability has not been assessed. SUV_{max-N} variability was compared to that of SUV_{max} ($N=1$) and SUV_{peak} , computed using the VOI encompassing the voxel corresponding to SUV_{max} and its 26 neighbouring voxels [9].

Materials and methods

Patients

This study received the approval of the Ethics Committee of the Teaching Hospital and all patients gave their informed consent before their inclusion in the study. Twelve lung cancer patients were included in the study and a total of 20 lesions

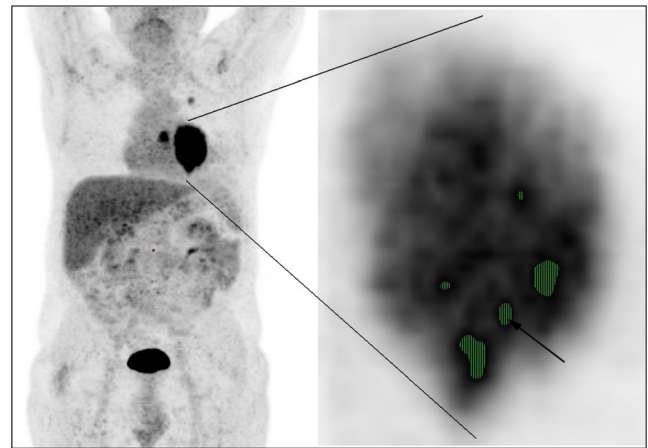


Fig. 1 *Left side:* PET maximal intensity projection of a patient presenting a lung squamous cell carcinoma of the left lower lobe, with a second malignant nodule of the left upper lobe and a subcarinal lymph node. *Right side:* Zoom of the main lesion showing a heterogeneous uptake: the set of *green areas* (as displayed by the AW workstation, GEMS) corresponds to the $N=30$ hottest voxels of the lesion ($SUV_{max-30}=15.5$ g/ml), while the *arrow* indicates the location of the hottest voxel ($SUV_{max}=18.0$ g/ml). The largest axis of the *green area* involving SUV_{max} is 6.5 mm. SUV_{peak} was centred on SUV_{max} and involved its 26 neighbouring voxels (0.66-ml volume; in-plane 1.16-cm diagonal line)

were investigated. Minimal lesion size, assessed with CT either in-plane or axial, was larger than 15 mm in order to minimise partial volume effects [10]. All patients fasted for at least 6 h before the ^{18}F -FDG injection and pre-injection average plasma glucose concentration was 1.00 $g.l^{-1}$ (range, 0.90–1.17). ^{18}F -FDG was administered intravenously for less than 1 min with a mean injected dose of 344 MBq (range, 229–460; assessed with a dose calibrator).

PET imaging and data processing

Patients underwent a static whole-body PET/CT for diagnosis purposes and a dynamic acquisition over the chest for the study purpose (one step, ten consecutive frames of 2.5 min each) that ranged within 60–110 min post-injection (Discovery ST PET/CT camera; General Electric Medical Systems, Milwaukee, USA; three-dimensional [3D] mode without septa; decay correction on). All PET images were reconstructed iteratively (FORE + OSEM; subsets, 32; iterations, 5; 3D postfilter of Hann, 0.9, 10.0); the voxel size was 2.73–3.27 mm (in-plane and axial respectively; FOV, 700×700 mm; matrix, 256×256 pixels) leading to a voxel volume of 0.0244 ml. For each examination, a low-dose unenhanced CT was performed for attenuation correction and used for anatomical localisation (pitch, 1.675; slice thickness, 3.75 mm; FOV, 500×500 mm; matrix, 512×512 pixels) leading to a voxel volume of 0.0036 ml.

An Advantage 4.6 workstation (General Electric Medical Systems) was used for drawing in each dynamic frame, a VOI encompassing each ^{18}F -FDG-positive lesion, separately. A

histogram representing the percentage of all voxels included in an arbitrary VOI versus SUV was used (Fig. 2; histogramming bin size=0.032 g.ml⁻¹) allowing us to assess SUV_{max} and SUV_{max-N} for N=1 and 5, 10, 15, 20, 25 or 30 hottest voxels, respectively. On the histogram, two cursors (*vertical green dotted lines* in Fig. 2) were used: the first one was set at SUV_{max} and the second one was moved along the SUV axis. For each position of the second cursor, i.e. for each histogramming bin, the total hottest volume and the associated mean SUV value were calculated between the two cursors (Fig. 2). Hottest volumes of 0.0244×5, 10, 15, 20, 25 or 30 ml, corresponding to N=5, 10, 15, 20, 25 or 30 voxels, provided associated mean SUV values corresponding to SUV_{max-N} respectively. When several voxels were encompassed within the same histogramming bin, i.e. several voxels whose SUV value falls into the same SUV range, the exact total hottest volume for an arbitrary N value of 5, 10, 15, 20, 25 or 30, and hence SUV_{max-N}, was not always directly given by the software. A rule of three was therefore used to deduce the mean SUV value corresponding to SUV_{max-N}. SUV_{peak} calculated as the mean SUV value over the voxel corresponding to SUV_{max} and its 26 neighbouring voxels were assessed using the PMOD Software version 3.3 (PMOD Technologies, Zurich, Switzerland).

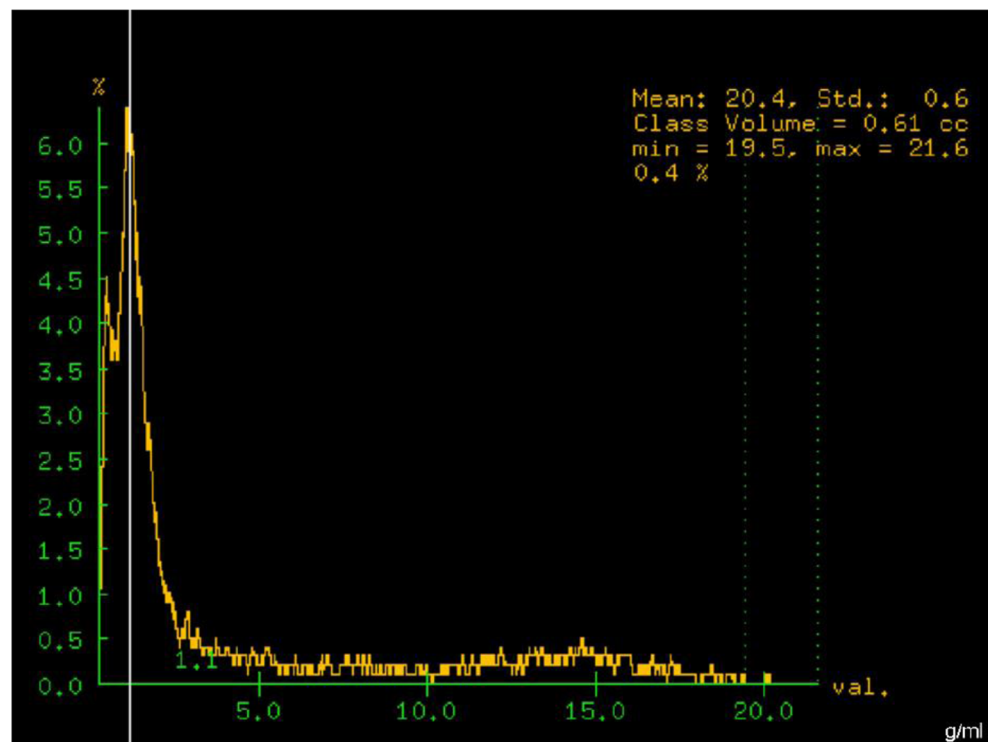
Statistical analysis

Statistical analysis by Bland and Altman was implemented over the lesion series to assess variability of

SUV_{max-N}, SUV_{max} and SUV_{peak}, involving the following steps [11, 12]: (1) for each ¹⁸F-FDG-positive lesion, mean and standard deviation (SD) of each SUV metric were assessed from the ten dynamic frames; (2) for each SUV metric, we found that SD was significantly related to magnitude (i.e. SUV mean) over the lesion series (Pearson coefficient), i.e. SD did not follow a normal law; we therefore transformed the data by using the relative standard deviation (SDr), i.e. a percentage, and for each SUV metric it was verified over the lesion series that SDr was not significantly related to magnitude (i.e. SUV mean) over the lesion series (Pearson coefficient); this feature then allowed us to calculate a mean SDr over the series (SDr_N, SDr_{max}, SDr_{peak}) respectively; (3) for each SUV metric, relative measurement error (MER), i.e. the relative difference between a single estimate of a parameter and its average “true” value, was computed as 1.96×mean SDr and 2.58×mean SDr with 95 % and 99 % confidence level (CL) respectively. Repeatability (R), i.e. the minimal relative change between two SUVs assessed from two successive examinations that is required to consider a significant difference, was computed as 2^{1/2}×MER; i.e. 2^{1/2}×1.96×mean SDr and 2^{1/2}×2.58×mean SDr with 95 % and 99 % CL respectively [11, 12].

Comparison between SDr_N, SDr_{max} and SDr_{peak} was achieved using two-tailed paired T-test. Two-tailed sign test was performed to assess whether SUV_{max-30} and SUV_{max} increased with time. P values of less than 0.05 were considered statistically significant.

Fig. 2 Histogram representing the percentage of all voxels included in a selected VOI (encompassing an arbitrary ¹⁸F-FDG-positive lesion in an arbitrary dynamic frame) versus SUV (g.ml⁻¹). For each lesion and in each of the ten dynamic frames, that histogram was used to obtain SUV_{max} (N=1) and SUV_{max-N} for the N=1, 5, 10, 15, 20, 25 or 30 hottest voxels (see text for details of the procedure). As an example, for a 0.61-ml volume, i.e. N=25, SUV_{max-25} is 20.4 g/ml (*Mean* displayed on the screen), SUV_{max} is 21.6 g/ml (maximal bin) and the minimal bin is 19.5 g/ml (*right and left vertical dotted green cursor* respectively); note that the SUV standard deviation, SD=0.6 g/ml, is also given between these limits



Results

Ten men and two women were included in the study (mean age, 63 years; range, 43–78) with 20 lesions in the thorax. Lesions were localised in the lung tissue ($n=13$), or in mediastinal lymph nodes ($n=7$; two patients presented two lymph nodes) and consisted of squamous cell carcinoma ($n=3$), small cell lung cancer ($n=3$), adenocarcinoma ($n=4$), large cell carcinoma ($n=2$). Minimal lesion size (assessed either in-plane or axial) ranged between 15.7 and 61.0 mm (mean, 29.4 mm). Patients’ mean weight and height were: 72 kg (range, 44–95 kg) and 169 cm (range, 157–179 cm), respectively.

For each SUV metric, Table 1 summarises the corresponding volume, the mean value (and range) over the lesion series, the correlation coefficient of the fit of SUV SDr versus mean SUV over the lesion series, mean SDr (and range) over the lesion series, MEr and R for 95 and 99 % CL. For each SUV metric, the correlation coefficient of the fit of SUV SDr versus mean SUV over the lesion series shows that SUV SDr was not significantly related to SUV magnitude ($r_0=0.44$; 95 % reliability; Table 1), therefore allowing calculation of mean SDr. Figure 3 graphically demonstrates this lack of correlation for SUV_{max-30} SDr versus mean SUV_{max-30} over the lesion series. A significant relationship was found between the correlation coefficient of the fit of the SDr of SUV_{max-N} versus mean SUV_{max-N} and N over the lesion series (Table 1; $N=5, 10, 15, 20, 25$ or 30 ; $y=-0.008x+0.400$; $r=0.94$; $P<0.01$; graph not shown).

Figure 4 shows a significant and a non-significant linear correlation with time for SUV_{max-30} and SUV_{max} in a typical lesion. Four and 11 lesions of a total of 20 showed a significant increase in SUV_{max-30} and SUV_{max} with time respectively (linear correlation; 95 % reliability). This result indicates that SUV_{max-30} significantly increases with time over the lesion series (two-tailed sign test, $P=0.012$), whereas no significant increase was found for SUV_{max} (two-tailed sign test, $P>0.75$).

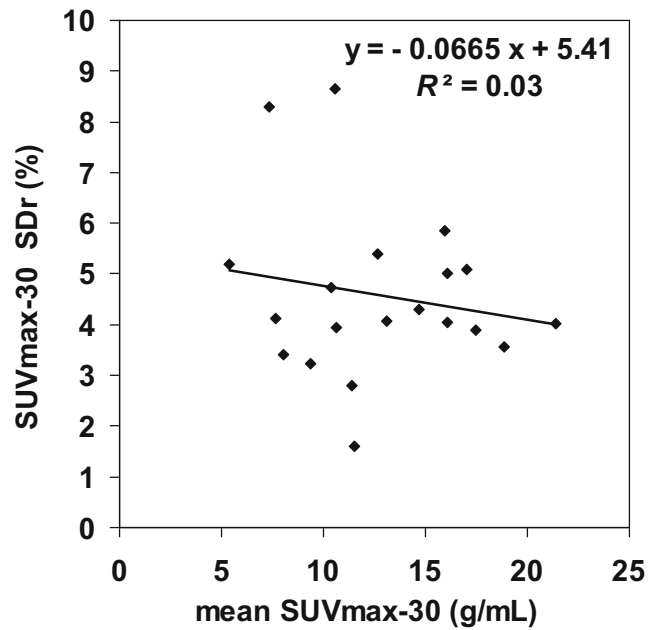


Fig. 3 SUV_{max-30} SDr versus mean SUV_{max-30} over the lesion series, which were obtained from ten values in each lesion. No significant correlation was found ($y=-0.0665x+5.41$; $r=0.17$), thus justifying calculation of SDr₃₀. For each SUV metric, such a lack of correlation was found over the lesion series

For all considered numbers of hottest voxels (N ranging from 5 to 30), SDr_N was found to be significantly lower than SDr_{max} and SDr_{peak} (two-tailed paired T -test, $P\leq0.004$). Figure 5 shows a strong correlation between SDr_N and N : $SDr_N=6.471 \times N^{-0.103}$ ($r=0.994$; $P<0.01$). No significant difference was observed between SDr_{max} and SDr_{peak} (two-tailed paired T -test, $P=0.64$).

Discussion

According to the guidelines (version 1.0) of the European Association of Nuclear Medicine (EANM), assessment of

Table 1 SUV characteristics

Parameter ^a	SUV _{max}	SUV _{max-5}	SUV _{max-10}	SUV _{max-15}	SUV _{max-20}	SUV _{max-25}	SUV _{max-30}	SUV _{peak}
Volume (mL)	0.0244	0.1220	0.2440	0.3660	0.4880	0.6100	0.7320	0.6588
SUV mean (g/mL)	14.79	14.12	13.71	13.39	13.14	12.96	12.78	11.51
Range	6.61-23.18	6.18-22.54	5.85-22.16	5.67-21.92	5.55-21.51	5.45-21.56	5.35-21.41	4.79-18.85
Corr. coefficient	0.38	0.39	0.29	0.28	0.21	0.21	0.17	0.13
mean SDr	7.07	5.52	5.03	4.91	4.76	4.65	4.56	6.84
Range	2.97-15.28	1.96-10.36	1.90-9.46	1.73-9.50	1.51-8.71	1.56-8.63	1.60-8.66	3.93-12.77
MEr (95 % CL)	13.86	10.82	9.86	9.62	9.33	9.11	8.94	13.41
R (95 % CL)	19.58	15.29	13.93	13.60	13.19	12.88	12.63	18.95
MEr (99 % CL)	18.24	14.24	12.98	12.67	12.28	12.00	11.76	17.65
R (99 % CL)	25.81	20.15	18.36	17.92	17.37	16.97	16.64	24.97

^a mean SDr, MEr and R in %

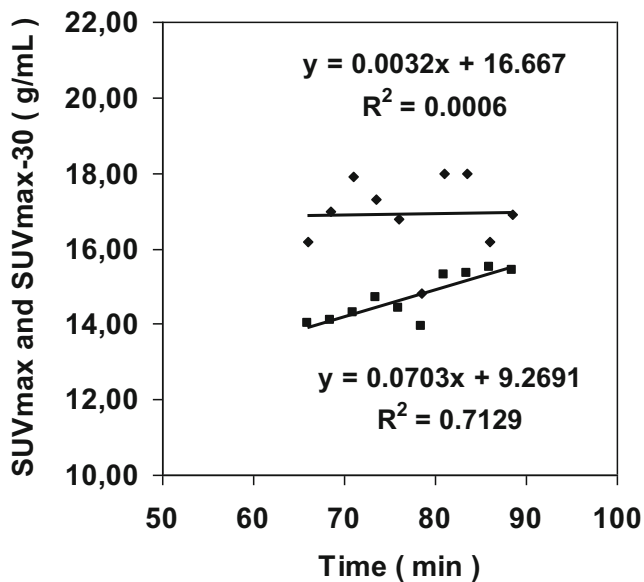


Fig. 4 SUV_{max} (diamonds) and SUV_{max-30} (squares) versus time in a typical lesion, showing a non-significant and a significant linear correlation respectively ($r=0.024$ and 0.844 respectively; 95 % reliability)

either SUV_{max} or SUV_{peak} , which is defined as the average SUV value of a 1-ml sphere (1.2-cm diameter) within the tumour that results in the highest value possible, is required for each clinically relevant ^{18}F -FDG-positive lesion [4, 13]. The present study showed that averaging SUV from several hottest voxels (i.e. SUV_{max-N} ; $N \neq 1$), instead of using one

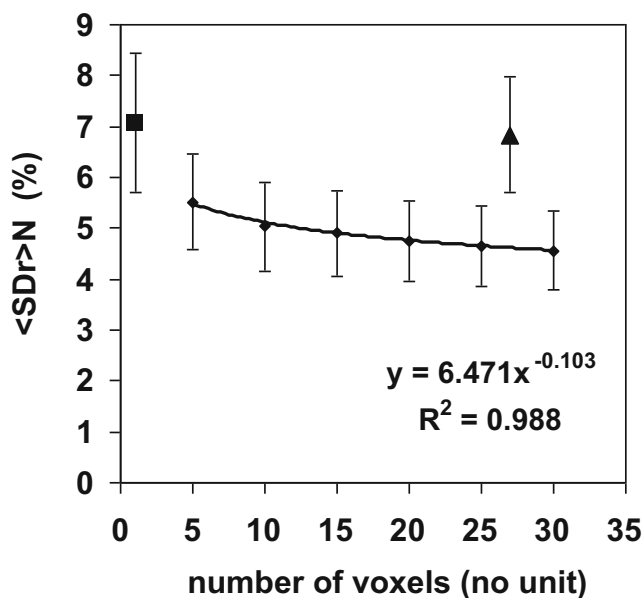


Fig. 5 SDr_N versus N showing a strong correlation: $SDr_N = 6.471 \times N^{-0.103}$ ($r=0.994$; $P<0.01$). For $N=5, 10, 15, 20, 25$ or 30 (diamonds), SDr_N was found to be significantly lower than SDr_{max} (square; $N=1$) and SDr_{peak} (triangle; $N=27$) (two-tailed paired T -test, $P \leq 0.004$; bars represent 95 % confidence limits). No significant difference was found between SDr_{max} and SDr_{peak} (two-tailed paired T -test, $P=0.64$). Note that when SUV_{max} is added to the fit, i.e. $N=1$, the equation of the further fit is: $SDr_N = 6.932 \times N^{-0.127}$ ($r=0.993$; $P<0.01$)

unique voxel (i.e. SUV_{max} ; $N=1$) results in a significantly lower variability of about 1/3 for $N=30$. This result is in very good agreement with the previous study by Burger et al. [8], whose results were obtained with $N=5-10$ in patients with malignant chest tumours: the ratios of SDr_5/SDr_{max} and of SDr_{10}/SDr_{max} (from ratios of mean percentage changes in reference [8]) were found to be $0.78 (=5.52/7.07)$ and $0.71 (=5.03/7.07)$ in the current study versus $0.81 (=6.60/8.17)$ and $0.77 (=6.30/8.17)$ in the study by Burger et al. Furthermore, irrespective of the N value, the current study showed that SUV_{max-N} also has a better variability performance than SUV_{peak} . Various shape, size and ROI location have been proposed in the literature and SUV_{peak} was calculated in this study from SUV_{max} and its 26 neighbouring voxels, corresponding to a 0.66-ml volume (in-plane 1.16-cm diagonal line) that is slightly lower than a 1-ml sphere (1.2-cm diameter) [4, 7, 9]. The better variability performance of SUV_{max-N} in comparison with SUV_{peak} is very likely related to the feature that SUV_{max-N} involves the hottest voxels not mandatorily close to each other in a ^{18}F -FDG-positive lesion, whereas SUV_{peak} is obtained over 27 adjacent voxels. As a consequence, SUV_{peak} includes voxels that may not be the hottest ones, as is typically encountered in a lesion showing a heterogeneous uptake (Fig. 1) or in an excavated lesion. In contrast, SUV_{max-N} takes into consideration the heterogeneity of the lesion uptake, a feature whose relevance has been recently shown [14, 15].

We suggest that the improvement in SUV accuracy and repeatability obtained by pooling several hottest voxels (SUV_{max-N}) could be useful for predicting outcome or for assessing treatment response, besides the use of other functional parameters such as the tumour volume (TV) and the total lesion glycolysis (TLG) [16–18]. Besides those volume-based parameters, baseline SUV_{max} value (i.e. the hottest SUV value) remains relevant in PET quantification, as recently demonstrated for prognostic stratification of lung cancer treated with erlotinib, thus opening up the interest for SUV_{max-N} (i.e. average of N hottest SUV values) [19]. This 3D pooling better represents the most aggressive parts of the tumour than SUV_{max} does, which is related to one voxel only ($N=1$; even if SUV_{max} is influenced by its neighbouring voxels due to the limited spatial resolution of PET systems), or SUV_{peak} , which is also fixed to one specific location within the VOI (Fig. 1). Moreover, it should be noted that SUV_{max-N} variability is slightly lower than that of a mean SUV obtained with a fixed 40 % of the SUV_{max} threshold, which does not only involve the hottest voxels [20]: $MEr-R$ of 8.94–12.63 % versus 9.72–13.74 % (95 % confidence level), for SUV_{max-30} and the latter respectively. Furthermore, SUV_{max-N} calculation can be easily implemented in current clinical practice and no intra- or inter-observer variability may be expected. Indeed, when a selected ^{18}F -FDG-positive lesion is separated from other ^{18}F -FDG-positive tissues, the histogram representing the percentage of all voxels included in that arbitrary VOI versus SUV (or any

list of the SUV voxel values sorted by descending order), is definitely established for the highest SUV values. This is particularly true when the total hottest volume that is averaged (in our study $N=30$ corresponds to a total hottest volume of 0.732 ml) is much lower than the whole lesion volume and Fig. 1 illustrates this feature. For the lowest SUV values of the histogram, not involved in the $SUV_{\max-N}$ calculation, there is a potential variability that depends on the number of included background voxels, in other words, that depends on the delineation method used.

As our results demonstrate that the best variability performance is achieved for $N=30$ in comparison with other SUV metric, the use of $SUV_{\max-30}$ might be recommended in current clinical practice. Moreover, it can be noted that $N=30$ is close to the total number of 27 neighbouring voxels used for the computation of SUV_{peak} [9]. However, this study has been conducted with specific reconstruction parameters (voxel volume of 0.0244 ml) and $N=30$ corresponds to a ^{18}F -FDG-positive volume of 0.732 ml (Table 1). It is worth noting that $N=40$ corresponds to a total ^{18}F -FDG-positive volume of 1 ml in our study, i.e. the volume of the sphere recommended by Wahl et al. [4], and $N=40$ could be recommended as well, with predicted MER and R for $SUV_{\max-40}$ of 8.67–12.26 % (95 % CL; Fig. 5; $SDr_N=6.471 \times N^{-0.103}$; $r=0.994$). However, increasing the N number of hottest voxels in the $SUV_{\max-N}$ calculation may be relevant to continue lowering its variability, but at the detriment of its inherent nature to represent the most “aggressive” parts of the tumour (Fig. 1). In this connection the present study may be considered, as in the knowledge phase of the Rapid Learning methodology [21], and further studies are required to give a decision on the optimal total hottest volume to be used ($N \times$ voxel volume) rather than on the optimal number of voxels (N), as the specific reconstruction parameters of each PET system may be different. Moreover, it should be noted that the malignant pathology could also play a role in the determination of the optimal total hottest volume to be used [22].

The improvement in SUV accuracy and repeatability obtained by pooling several hottest voxels could also be applied to the assessment of ^{18}F -FDG-positive TV obtained with fixed percentage of the SUV_{\max} thresholds [23]. For different percentage thresholds, it has been shown that TV variability correlated with SUV_{\max} variability. Because $SUV_{\max-N}$ variability is lower than that of SUV_{\max} , assessing ^{18}F -FDG-positive TV created by using a fixed percentage of the $SUV_{\max-N}$ threshold, instead of an SUV_{\max} threshold, might result in a lower TV variability. This suggestion also appears relevant for other functional parameters requiring tumour outlining such as the TLG, which is the product of TV with associated mean SUV [16, 20].

The current study presents some limitations. First, SUV variability was assessed by using a single 25-min dynamic

PET acquisition, therefore providing data over a ± 12.5 (=25/2)-min time window around a mean injection-acquisition time delay. As a result, temporal changes in SUV occurring during a 12.5-min time duration may provide a slight overestimation of the SUV variability. Indeed, $SUV_{\max-30}$ was found to significantly increase with time over the lesion series, a feature that is well-established in ^{18}F -FDG-positive tumours (two-tailed sign test, $P=0.012$), and no significant increase was found for SUV_{\max} , a feature that may be related to its greater variability than that of $SUV_{\max-30}$ (Fig. 4). We suggest that this slight overestimation is more clinically acceptable than an underestimation [20]. Furthermore, we also suggest that test-retest studies, involving two baseline examinations repeated on two different days with the same injection-acquisition time delay, are warranted for comparison with the present one. Indeed, although the patient radiation dose should be taken into consideration, test-retest studies allow assessing other origins of SUV variability than those investigated with the present study design, such as changes in plasma glucose level, in injected dose and in positioning [24]. Second, the SUV values investigated in the present study were higher than those of Burger et al.’s: 6.61–23.18 $\text{g}\cdot\text{ml}^{-1}$ versus 3–16 $\text{g}\cdot\text{ml}^{-1}$ (Fig. 2 in reference [8]) respectively. In a recent meta-analysis de Langen et al. [25] have shown that the lower the SUV_{\max} , the greater the repeatability percentage (and hence the greater the SDR). Moreover, minimal lesion size was larger than 15 mm in order to minimise partial volume effects [10]. Therefore, it is suggested that a full separate analysis is required to more precisely assess $SUV_{\max-N}$ variability in small-size lesions and in lesions showing faint ^{18}F -FDG uptake, the former requiring correction for partial volume effect and respiratory gating.

To conclude, this study shows that averaging SUV from several hottest voxels ($SUV_{\max-N}$) significantly improves the SUV variability performance in comparison with that of SUV_{\max} and SUV_{peak} , of about 1/3 for $N=30$. This averaging can be very easily implemented clinically and therefore we recommend the use of $SUV_{\max-N}$ in current clinical practice to improve the accuracy of SUV values for predicting outcome or for assessing treatment response. However, further studies are warranted to determine the optimal total hottest volume to be used.

Acknowledgements The scientific guarantor of this publication is Prof. Roger Marthan. The authors of this manuscript declare no relationships with any companies, whose products or services may be related to the subject matter of the article. The authors state that this work has not received any funding. No complex statistical methods were necessary for this paper. Institutional Review Board approval was obtained. Written informed consent was obtained from all subjects (patients) in this study. Some study subjects or cohorts have been previously reported in references 20 and 23 of the revised paper.

Methodology: retrospective, as PET images had already been acquired [20, 23], diagnostic or prognostic study, performed at one institution.

References

1. Visvikis D, Hatt M, Tixier F, Cheze-Le Rest D (2012) The age of reason for FDG PET image-derived indices. *Eur J Nucl Med Mol Imaging* 39:1670–1672
2. Boellaard R (2009) Standards for PET image acquisition and quantitative data analysis. *J Nucl Med* 50:11S–20S
3. Wahl RL, Zasadny K, Helvie M, Hutchins GD, Weber B, Cody R (1993) Metabolic monitoring of breast cancer chemohormonotherapy using positron emission tomography: initial evaluation. *J Clin Oncol* 11:2101–2111
4. Wahl RL, Jacene H, Kasamon Y, Lodge MA (2009) From RECIST to PERCIST: evolving considerations for PET response criteria in solid tumours. *J Nucl Med* 50:122S–150S
5. Visser EP, Boerman OC, Oyen WJG (2010) SUV: from silly useless value to smart uptake value. *J Nucl Med* 51:173–175
6. Tyłski P, Stute S, Grotus N et al (2010) Comparative assessment of methods for estimating tumor volume and standardized uptake value in ^{18}F -FDG PET. *J Nucl Med* 51:268–276
7. Vanderhoek M, Perlman SB, Jeraj R (2012) Impact of the definition of peak standardized uptake value on quantification of treatment response. *J Nucl Med* 53:4–11
8. Burger IA, Huser DM, Burger C, von Schulthess GK, Buck A (2012) Repeatability of FDG quantification in tumor imaging: averaged SUVs are superior to SUVmax. *Nucl Med Biol* 39:666–670
9. Hatt M, Visvikis D, Albarghach NM, Tixier F, Pradier O, Cheze-Le Rest D (2011) Prognostic value of ^{18}F -FDG PET image-based parameters in oesophageal cancer and impact of tumour delineation methodology. *Eur J Nucl Med Mol Imaging* 38:1191–1202
10. Soret M, Bacharach SL, Buvat I (2007) Partial-volume effect in PET tumor imaging. *J Nucl Med* 48:932–945
11. Bland JM, Altman DG (1996) Statistics notes: measurement error proportional to the mean. *BMJ* 313:106–108
12. Bland JM, Altman DG (1996) Statistics notes: measurement error. *BMJ* 313:744–746
13. Boellaard R, O'Doherty MJ, Weber WA et al (2010) FDG PET and PET/CT: EANM procedure guidelines for tumour PET imaging: version 1.0. *Eur J Nucl Med Mol Imaging* 37:181–200
14. Tixier F, Hatt M, Le Rest CC, Le Pogam A, Corcos L, Visvikis D (2012) Reproducibility of tumor uptake heterogeneity characterization through textural feature analysis in ^{18}F -FDG PET. *J Nucl Med* 53:693–700
15. Hatt M, Tixier F, Cheze Le Rest C, Pradier O, Visvikis D (2013) Robustness of intratumour ^{18}F -FDG PET uptake heterogeneity quantification for therapy response prediction in oesophageal carcinoma. *Eur J Nucl Med Mol Imaging* 40:1662–1671
16. Larson SM, Erdi Y, Akhurst T et al (1999) 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* 2:159–171
17. Chen HH, Chiu NT, Su WC, Guo HR, Lee BF (2012) Prognostic value of whole-body total lesion glycolysis at pretreatment FDG PET/CT in non-small cell lung cancer. *Radiology* 264:559–566
18. Soussan M, Chouahnia K, Maisonobe J-A et al (2013) Prognostic implications of volume-based measurements on FDG PET/CT in stage III non-small-cell lung cancer after induction chemotherapy. *Eur J Nucl Med Mol Imaging* 40:668–676
19. Scheffler M, Zander T, Nogova L et al (2013) Prognostic impact of [^{18}F]fluorothymidine and [^{18}F]fluoro-D-glucose baseline uptakes in patients with lung cancer treated first-line with erlotinib. *PLoS ONE* 8:e53081
20. Laffon E, de Clermont H, Lamare F, Marthan R (2013) Variability of total lesion glycolysis by ^{18}F -FDG-positive tissue thresholding in lung cancer. *J Nucl Med Technol* 41:186–191
21. Lambin P, Roelofs E, Reymen B et al (2013) Rapid learning health care in oncology—an approach towards decision support systems enabling customised radiotherapy. *Radiother Oncol* 109:159–164
22. Ambrosini V, Nicolini S, Caroli P (2012) PET/CT imaging in different types of lung cancer: an overview. *Eur J Radiol* 81:988–1001
23. Laffon E, de Clermont H, Marthan R (2013) Variability of (^{18}F)-FDG-positive lung lesion volume by thresholding. *Eur Radiol* 23:1131–1137
24. Boellaard R (2011) Need for standardization of ^{18}F -FDG PET/CT for treatment response assessments. *J Nucl Med* 52:93S–100S
25. de Langen AJ, Vincent A, Velasquez LM et al (2012) Repeatability of ^{18}F -FDG uptake measurements in tumors: a meta-analysis. *J Nucl Med* 53:701–708