SUVpeak performance in Lung Cancer: 
Comparison to Average SUV from 40 Hottest Voxels.

E. Laffon 1, 2, 3 *, I.A. Burger 4, F. Lamare 1, H. de Clermont 1, R. Marthan 2, 3.

1 CHU de Bordeaux, Service de Médecine Nucléaire- F-33604 Pessac France.
2 Université de Bordeaux 2, Centre de Recherche Cardio-Thoracique, F-33076 Bordeaux France.
3 INSERM U 1045, Centre de Recherche Cardio-Thoracique, F-33076 Bordeaux France.
4 Division of Nuclear Medicine, Department Medical Radiology, University Hospital of Zurich, Ramistrasse 100, CH-8091 Zurich, Switzerland.

*Corresponding author:
Dr Eric LAFFON, Service de Médecine Nucléaire, Hôpital du Haut-Lévêque, Avenue de Magellan, 33604 PESSAC, France. Telephone: +33 5 57 65 68 38; Fax: +33 5 57 65 68 39; E-mail: elaffon@u-bordeaux2.fr

Word count: 3017.
ABSTRACT

The performance of an average standard uptake value (SUV) over a 1-mL-volume sphere within an 18F-FDG-positive lesion resulting in the highest possible value (SUV<sub>peakW</sub>), was compared to that of an average SUV computed from the 40 hottest voxels, irrespective of their location within the lesion (SUV<sub>max–40</sub>).

Methods: A PET dynamic acquisition performed in 20 lung cancer lesions yielded for each SUV metric its mean value, relative measurement error and repeatability (MEr–R).

Results: SUV<sub>peakW</sub> mean value was significantly 9.66% lower than that of SUV<sub>max–40</sub> (P<0.0001). SUV<sub>peakW</sub> and SUV<sub>max–40</sub> MEr–R were significantly lower than MEr-R of SUV<sub>max</sub> (the hottest voxel): 9.35–13.21%, 8.84–12.49% versus 13.86–19.59% respectively (95%-CL; P<0.0001). Although being marginal, SUV<sub>peakW</sub> MEr–R were not significantly greater than SUV<sub>max–40</sub> MEr–R (P=0.086).

Conclusions: SUV<sub>max–40</sub> is more likely to represent the most metabolically active portions of tumors than SUV<sub>peakW</sub>, with close variability performance.

Keywords: Oncology; 18F-FDG PET; PERCIST criteria; SUV repeatability; Treatment monitoring.
INTRODUCTION

$^{18}$F-fluorodeoxyglucose ($^{18}$F-FDG) positron emission tomography (PET) imaging is expected to play a major role in assessing whether a tumor is responding to therapy, allowing then to quickly determine whether to continue, change, or abandon treatment, before morphological changes can be detected. Because of limitations of anatomic tumor response metrics such as the Response Evaluation Criteria in Solid Tumors (RECIST) criteria, PERCIST (for PET Response Criteria in Solid Tumors) criteria have been proposed by Wahl et al. to quantitatively assess the metabolic tumor response with $^{18}$F-FDG PET (1). In particular, a major component of the proposed PERCIST criteria is the use of a 1-mL sphere (1.2-cm diameter) centered over the most active region of metabolically active tumors. The corresponding average standard uptake value ($\text{SUV}_{\text{peakW}; \ g.mL^{-1}}$) is therefore aimed at assessing the most aggressive portion of tumors with reduced statistical variability in comparison to that of the $\text{SUV}_{\text{max}}$ (obtained from the voxel with the highest activity).

Several definitions of the peak SUV have been proposed that can significantly affect its use for assessing treatment response (2). Variability of two arbitrary peak SUVs, defined as the average standard uptake value over a small volume of interest (VOI) centered on the $\text{SUV}_{\text{max}}$ and encompassing neighbouring voxels, i.e. $\text{SUV}_{\text{peak}}$, has been recently reported in two studies with lung cancer patients (3,4). Although using a different design, a PET dynamic acquisition involving ten frames and two (test-retest) static acquisitions within a few days interval without treatment, respectively, a similar variability performance was found between $\text{SUV}_{\text{peak}}$ and $\text{SUV}_{\text{max}}$ in each study, showing that, in terms of variability performance, no advantage should be expected by using $\text{SUV}_{\text{peak}}$ rather than $\text{SUV}_{\text{max}}$ for assessing response to treatment. However, the arbitrary $\text{SUV}_{\text{peak}}$ that was used in these two studies, respectively, was not exactly the same as that defined by Wahl et al. with PERCIST criteria, of which assessment software was not
commercially available (1). Furthermore, an alternative quantitation tool with similar features to those of SUV\textsubscript{peakW} has been recently proposed, which is an average SUV measurement obtained by pooling several hottest voxels regardless of their location within the \textsuperscript{18}F-FDG-positive lesion, i.e. SUV\textsubscript{max–N} when N voxels are pooled (3,5). It has been shown that its use resulted in a significantly lower variability than that of SUV\textsubscript{max} and SUV\textsubscript{peak} defined as SUV\textsubscript{max} and its 26 neighbouring voxels (3). In this previous study the variability of SUV\textsubscript{max} and SUV\textsubscript{peak} were investigated within the same patients used for the current study. Since the tool enabling the assessment of SUV\textsubscript{peakW} as defined by Wahl et al. with PERCIST criteria (SUV\textsubscript{peakW}) has become available, we performed further analysis of our data with the aim to compare the SUV\textsubscript{peakW} variability performance to that of SUV\textsubscript{max–40} (corresponding to a total hottest volume close to 1 mL).

MATERIALS AND METHODS

Patients

Twelve lung cancer patients (2 females; 10 males; 63-years-old on average; range, 43-78; 9/3 non-small/small cell lung cancer) were included in the study and a total of 20 lesions were investigated (lung tissue lesions, n=13; mediastinal lymph nodes, n=7). This retrospective study received the approval of the Ethics Committee of the Teaching Hospital and the requirement to obtain informed consent was waived. Patients’ mean weight and height were: 72 kg (range, 44–95) and 169 cm (range, 157–179). After 6-hours fasting before the tracer injection, preinjection average plasma glucose concentration was 1.00 g.L\textsuperscript{-1} (range, 0.90-1.17).

PET Imaging and Data Processing
\(^{18}\text{F-FDG}\) was administered intravenously for less than 1 minute with a mean injected dose of 344 MBq (range, 229–460; assessed with a dose calibrator). Dynamic PET imaging was performed over the chest for the study purpose, without respiratory gating, within 60–110 min post-injection (1 step, 10 consecutive frames of 2.5-min each), using a Discovery ST PET/CT device (General Electric Medical Systems, Milwaukee, WI, USA; 3D mode without septa; decay-correction on). All PET images were reconstructed iteratively (Fourier-Rebinning + Ordered-Subsets-Expectation-Maximization; subsets: 32; iterations: 5; 3D postfilter of Hann: 0.9, 10.0), and the voxel size was 2.73x2.73x3.27 mm\(^3\) (in-plane and axial respectively; FOV: 700x700 mm\(^2\); matrix: 256x256 pixels) leading to a voxel volume of 0.0244 mL. Unenhanced CT transmission imaging was performed before the PET imaging for attenuation correction and used for anatomical localization (pitch: 1.675; slice thickness: 3.75 mm; FOV: 500x500 mm\(^2\); matrix: 512x512 pixels) leading to a voxel volume of 0.0036 mL. Minimal lesion size was assessed with CT either in-plane or axial, that was always larger than 15mm in order to minimize partial volume effects (6).

An "Advantage 4.6" workstation (General Electric Medical Systems) was used for drawing in each dynamic frame a VOI encompassing each \(^{18}\text{F-FDG}\)-positive lesion. The method to assess SUV\(_{\text{max-40}}\) has been previously described in details (3). Briefly, SUV\(_{\text{max-40}}\) was obtained from the histogram representing the percentage of all voxels included in the VOI versus SUV. It is the averaged SUV from the 40 hottest voxels, i.e. over a total hottest volume of 0.98 mL. SUV\(_{\text{peakW}}\) defined by Wahl et al. with PERCIST criteria was obtained by using the “PET-VCAR” application of the workstation (General Electric Medical Systems).

**Statistical Analysis**
For each lesion, a mean $\text{SUV}_{\text{peakW}}$ and a mean $\text{SUV}_{\text{max–40}}$ value and corresponding standard deviation (SD) were computed from 10 measurements performed in each of the 10 frames of the dynamic PET imaging. For each SUV metric, it was verified over the lesion series that the relative standard deviation (SDr) was not significantly related to magnitude, and a mean SDr was then calculated: $<\text{SDr}>_{\text{peakW}}$ and $<\text{SDr}>_{\text{N=40}}$ (7,8). For each SUV metric, MEr (i.e. the relative difference between a single estimate of a parameter and its average "true" value), and R (i.e. the minimal relative change between two SUVs assessed from two successive scans that is required to consider a significant difference) were calculated as $1.96*<\text{SDr}>$ and $2^{1/2}*1.96<\text{SDr}>$ (95%-CL), respectively.

Comparison between (i) $<\text{SDr}>_{\text{peakW}}$, $<\text{SDr}>_{\text{max–40}}$, $<\text{SDr}>_{\text{max}}$, $<\text{SUV}>_{\text{peak}}$, and between (ii) mean values over the lesion series of $\text{SUV}_{\text{peakW}}$, $\text{SUV}_{\text{max–40}}$, $\text{SUV}_{\text{max}}$, $\text{SUV}_{\text{peak}}$, i.e. $<\text{SUV}>_{\text{peakW}}$, $<\text{SUV}>_{\text{max–40}}$, $<\text{SUV}>_{\text{max}}$, $<\text{SUV}>_{\text{peak}}$, was achieved using a two-tailed paired T-test. $P$ values of less than 0.05 were considered statistically significant.

**RESULTS**

Because SDr of $\text{SUV}_{\text{peakW}}$ and $\text{SUV}_{\text{max–40}}$ was not significantly related to SUV magnitude over the lesion series ($r = 0.25$ and 0.13 respectively; 95% reliability), $<\text{SDr}>_{\text{peakW}}$ and $<\text{SDr}>_{\text{max–40}}$ were calculated: 4.77 and 4.61%, respectively. MEr–R of $\text{SUV}_{\text{peakW}}$ and $\text{SUV}_{\text{max–40}}$ were found to be 9.35–13.21% and 8.84–12.49%, respectively (95%-CL). Although on the borderline, MEr–R of $\text{SUV}_{\text{peakW}}$ were not significantly greater than MEr–R of $\text{SUV}_{\text{max–40}}$ ($P=0.086$). MEr–R of $\text{SUV}_{\text{peakW}}$ and $\text{SUV}_{\text{max–40}}$ were found to be significantly lower than those of $\text{SUV}_{\text{max}}$, and $\text{SUV}_{\text{peak}}$: 13.86–19.59% and 13.41–18.95% respectively ($P<0.0001$; Figure 1A)(3). Figure 1B shows $<\text{SUV}>_{\text{peak}}$ and $<\text{SUV}>_{\text{max–40}}$: 11.39 and 12.49 g/mL (range: 4.58–19.18 and 5.21–21.17):
the former was significantly 9.66% (on average) lower than the latter ($P<0.0001$). $\text{SUV}_{\text{peakW}}$ and $\text{SUV}_{\text{max-40}}$ were significantly lower, with 29.85 and 18.41%, respectively, than $\text{SUV}_{\text{max}}$ ($\text{SUV}_{\text{max}}=14.79$ g/mL; range: 6.61-23.18; $P<0.0001$). $\text{SUV}_{\text{peakW}}$ was not found to be significantly different from $\text{SUV}_{\text{peak}}$: 11.39 g/mL versus 11.45 g/mL ($P=0.47$).

Sixteen lesions of a total of 20 showed a significant increase with time in both $\text{SUV}_{\text{peakW}}$ and $\text{SUV}_{\text{max-40}}$ (linear correlation; 95% reliability), indicating that both $\text{SUV}_{\text{peakW}}$ and $\text{SUV}_{\text{max-40}}$ significantly increase with time over the lesion series ($P=0.012$; two-tailed sign test). Figure 2 shows in a typical lesion that, whatever the time point, $\text{SUV}_{\text{peakW}}$ outcomes are significantly lower than those of $\text{SUV}_{\text{max-40}}$ ($P=0.002$, two-tailed sign test).

No significant correlation was found between $\text{SUV}_{\text{peakW}}$, $\text{SUV}_{\text{max-40}}$ and minimal lesion size assessed with CT (either in-plane or axial).

**DISCUSSION**

$^{18}$F-FDG PET imaging in oncology is in need of robust methods enabling to reliably assess treatment efficacy. The most aggressive portions of tumors are acknowledged to be the most critically important parts for this purpose (1). In this context, besides $\text{SUV}_{\text{max}}$, which is obtained from the hottest voxel, Wahl et al. have proposed the use of $\text{SUV}_{\text{peakW}}$ in order to reduce SUV outcome variability. $\text{SUV}_{\text{peakW}}$ is the average SUV obtained from a 1-mL sphere within the tumor that results in the highest possible value. In a series of lung cancer patients, the present study compared the performance of $\text{SUV}_{\text{peakW}}$ with that of $\text{SUV}_{\text{max-40}}$, i.e. pooling 40 hottest voxels (total hottest volume of 0.98 mL), irrespective of their location within the lesion. $\text{SUV}_{\text{peakW}}$ was significantly 9.66% lower (on average) than $\text{SUV}_{\text{max-40}}$, and both were significantly lower than $\text{SUV}_{\text{max}}$ (Figure 1). $\text{SUV}_{\text{peakW}}$ and $\text{SUV}_{\text{max-40}}$ showed close variability performance that was significantly better than that of $\text{SUV}_{\text{max}}$ (95%-CL; $P<0.0001$; Figure 1A).
Therefore, we suggest that $\text{SUV}_{\text{max-40}}$ might be superior to $\text{SUV}_{\text{peakW}}$ for assessing the most metabolically active portions of tumors, with close variability performances for both metrics.

$\text{SUV}_{\text{peakW}}$ and $\text{SUV}_{\text{max-40}}$ also showed variability performance that was significantly better than that of an arbitrary $\text{SUV}_{\text{peak}}$, defined as $\text{SUV}_{\text{max}}$ and its 26 neighbouring voxels (3). This result is consistent with that of Weber et al. in lung cancer patients that used a different VOI centered on the $\text{SUV}_{\text{max}}$ and reported similar variability performance between $\text{SUV}_{\text{peak}}$ and $\text{SUV}_{\text{max}}$ (4). Furthermore, it is noteworthy that the current study used a PET dynamic acquisition involving ten frames (equivalent to ten sequential static acquisitions) that ruled out origins of SUV variability such as changes in plasma glucose level, injected dose and positioning, in comparison with the test-retest study of Weber et al. We therefore suggest that the design of the current study, which takes into consideration the patient dose, is relevant to compare the performance of different SUV metrics.

Some results published by Lodge et al. about the comparison between $\text{SUV}_{\text{max}}$ and $\text{SUV}_{\text{peakW}}$ performance are consistent with those of the current study, despite major differences in study design such as investigated malignancy (including lung, liver and pancreas, instead of lung only), injection-acquisition time delay (147 ± 37, instead of 60–110 min), acquisition (respiration-gated from 15-min list-mode data, including only 2 phases, instead of 10-frames dynamic acquisition)(9). In particular, for a 256x256 image matrix, Lodge et al. also reported that $\text{SUV}_{\text{peakW}}$ was significantly lower than $\text{SUV}_{\text{max}}$, of 35.77% on average, a finding comparable to 29.85% obtained in the present study. However, although $\text{SUV}_{\text{peakW}}$ repeatability (R, i.e. the minimal relative change between two SUVs assessed from two successive scans that is required to consider a significant difference) was found to be significantly lower than $\text{SUV}_{\text{max}}$ R in each study, there is a twofold discrepancy about $\text{SUV}_{\text{peakW}}$ R between Lodge’s and the current study: 6.65 versus 13.21%, respectively (95%-CL). For comparison, $\text{SUV}_{\text{max}}$ R is found similar:
18.02% versus 19.59%, respectively (95%-CL). We suggest that this discrepancy in SUV_{peakW} R may be related to a different study design. In particular, further studies are warranted for investigating the potential role of respiratory gating for further reduction of SUV_{peakW} R.

Close variability performance of SUV_{max–40} and SUV_{peakW}, which was found to be significantly lower than that of SUV_{max}, is related to the fact that both methodologies are based on the same strategy, i.e. averaging SUV from several voxels to lower its variability. However, a significantly lower performance of SUV_{peakW} was found for reporting the hottest part(s) of tumors, in comparison with SUV_{max–40}. This finding may be related to the fact that the hottest voxels in an $^{18}$F-FDG-positive lesion are not mandatorily close to each other, and a 1-mL sphere unavoidably includes some voxels that are not the hottest ones. In other words, the spatial resolution of the SUV_{peakW} metric is much lower than that of the SUV_{max–40} metric, which is only limited by the voxel size of the PET system used. Furthermore, it should be noted that (i) SUV_{max–40} metric can be easily implemented in current clinical practice, (ii) low intra- or inter-observer variability was reported (5), and (iii) SUV_{max–40} metric may be normalized either to body weight (as in the current study) or to lean body mass as well (10,11).

The current study presents some limitations. First, even if it was performed using clinical patient data to provide a realistic SUV variability context, SUV_{peakW} and SUV_{max–40} range did not involve small-size lesions and lesions showing faint $^{18}$F-FDG uptake lesions: minimal lesion size was larger than 15mm in order to minimize partial volume effects (6), and uptake range was 4.58–19.18 and 5.21–21.17 g/mL for SUV_{peakW} and SUV_{max–40} respectively. Nevertheless, we suggest that, unlike SUV_{peakW}, SUV_{max–N} metric may be considered as an adjustable tool that is suitable to report $^{18}$F-FDG uptake in lesions of smaller size and of lower uptake than those of the current study. Indeed, reducing the total hottest volume to be reported, i.e. lowering the number of hottest voxels to be pooled (but keeping it greater than 1), will always lower variability.
percentage in comparison with that of $SUV_{\text{max}}$ (one voxel)\(^{(3)}\). This suggestion is supported by Hasenclever et al. in interim PET performed in lymphoma patients, who used an arbitrary $SUV_{\text{peak}}$ metric involving $SUV_{\text{max}}$ and 3-hottest-adjacent voxels assessed in a target lesion \(^{(12)}\). Therefore, we suggest that further studies are warranted to determine the optimal total hottest volume to be reported depending on the clinical situation and on the specific reconstruction parameters of each PET system \(^{(2,12,13)}\). Second, $SUV_{\text{peakW}}$ and $SUV_{\text{max-40}}$ were found to significantly increase with time over the lesion series (Figure 2). We suggest that this correlation versus time of both SUVs does not alter the conclusion of the present study. For instance in a typical lesion, Figure 2 shows that, whatever the time point, $SUV_{\text{peakW}}$ outcomes are significantly lower than those of $SUV_{\text{max-40}}$ \(P = 0.002\), two-tailed sign test).

**CONCLUSION**

This study showed that variability performance of $SUV_{\text{max-40}}$ and $SUV_{\text{peakW}}$ are close and both superior to $SUV_{\text{max}}$ and $SUV_{\text{peak}}$. Furthermore $SUV_{\text{max-40}}$ might be superior to $SUV_{\text{peakW}}$ for assessing the most metabolically active, and hence the most aggressive, portions of tumors. Comparison between $SUV_{\text{peakW}}$ and $SUV_{\text{peak}}$ performance suggests that $SUV_{\text{peak}}$ may be ruled out as a reliable tool for PET quantification.
REFERENCES


Figure 1. A) MEr comparison of $\text{SUV}_{\text{peakW}}$ (square) and of $\text{SUV}_{\text{max-40}}$ (rhomb), involving also comparison with MEr of $\text{SUV}_{\text{max}}$ (triangle) and of $\text{SUV}_{\text{peak}}$ (circle) previously published (3). Bars represent 95% confidence limits. Repeatability (R) can be obtained by multiplying MEr by $\sqrt{2}$; B) Comparison of $\langle \text{SUV} \rangle_{\text{max}}$ (triangle), $\langle \text{SUV} \rangle_{\text{peak}}$ (circle), $\langle \text{SUV} \rangle_{\text{peakW}}$ (square) and $\langle \text{SUV} \rangle_{\text{max-40}}$ (rhomb) over the lesion series.
Figure 2. SUV<sub>peakW</sub> (square) and SUV<sub>max-40</sub> (rhomb) versus time in a typical lesion, showing a significant linear correlation (r = 0.96 and 0.92 respectively; 95% reliability).