SUV_{peak} Performance in Lung Cancer: Comparison to Average SUV from the 40 Hottest Voxels

Eric Laffon[1–3], Irene A. Burger[4], Frederic Lamare[1], Henri de Clermont[1], and Roger Marthan[2,3]

[1]CHU de Bordeaux, Service de Médecine Nucléaire, Pessac, France; [2]Université de Bordeaux 2, Centre de Recherche Cardio-Thoracique, Bordeaux, France; [3]INSERM U 1045, Centre de Recherche Cardio-Thoracique, Bordeaux, France; and [4]Division of Nuclear Medicine, Department of Medical Radiology, University Hospital of Zurich, Zurich, Switzerland

The performance of an average SUV over a 1-mL-volume sphere within an ^{18}F-FDG–positive lesion resulting in the highest possible value (SUV_{peakW}) was compared with that of an average SUV computed from the 40 hottest voxels, irrespective of their location within the lesion (SUV_{max-40}). **Methods:** Dynamic PET performed in 20 lung cancer lesions yielded for each SUV metric its mean value, relative measurement error, and repeatability (MEr-R). **Results:** SUV_{peakW} mean value was significantly 9.66% lower than that of SUV_{max-40} ($P < 0.0001$). SUV_{peakW} and SUV_{max-40} MEr-R were significantly lower than the MEr-R of SUV_{max} (the hottest voxel): 9.35%–13.21% and 8.84%–12.49% versus 13.86%–19.59%, respectively, (95% confidence limit; $P < 0.0001$). Although being marginal, SUV_{peakW} MEr-R was not significantly greater than SUV_{max-40} MEr-R ($P = 0.086$). **Conclusion:** SUV_{max-40} is more likely to represent the most metabolically active portions of tumors than SUV_{peakW}, with close variability performance.

Key Words: oncology; ^{18}F-FDG PET; PERCIST criteria; SUV repeatability; treatment monitoring

J Nucl Med 2016; 57:85–88
DOI: 10.2967/jnumed.115.161968

PET imaging with ^{18}F-FDG is expected to play a major role in assessing whether a tumor is responding to therapy, allowing physicians then to quickly determine whether to continue, change, or abandon treatment, before morphologic changes can be detected. Because of limitations of anatomic tumor response metrics such as the RECIST, PERCIST has 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 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 SUV (SUV_{peakW}; $g{\cdot}mL^{-1}$) is therefore aimed at assessing the most aggressive portion of tumors with reduced statistical variability in comparison to that of the SUV_{max} (obtained from the voxel with the highest activity).

Several definitions of the SUV_{peak} have been proposed that can significantly affect its use for assessing treatment response (*2*). Variability of 2 arbitrary peak SUVs, defined as the average SUV over a small volume of interest centered on the SUV_{max} and encompassing neighboring voxels—that is, SUV_{peak}—has been recently reported in 2 studies with lung cancer patients (*3,4*). Although a different design was used, a PET dynamic acquisition involving 10 frames and 2 (test–retest) static acquisitions within a few days interval without treatment, respectively, a similar variability performance was found between SUV_{peak} and SUV_{max} in each study, showing that, in terms of variability performance, no advantage should be expected using SUV_{peak} rather than SUV_{max} for assessing response to treatment. However, the arbitrary SUV_{peak} that was used in these 2 studies, respectively, was not exactly the same as that defined by Wahl et al. with PERCIST, for which assessment software was not commercially available (*1*). Furthermore, an alternative quantitation tool with features similar to those of SUV_{peakW} has been recently proposed, which is an average SUV measurement obtained by pooling several hottest voxels regardless of their location within the ^{18}F-FDG–positive lesion—that is, SUV_{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_{max} and SUV_{peak} defined as SUV_{max} and its 26 neighboring voxels (*3*). In this previous study, the variability of SUV_{max} and SUV_{peak} were investigated within the same patients used for the current study. Because the tool enabling the assessment of SUV_{peakW} as defined by Wahl et al. with PERCIST (SUV_{peakW}) has become available, we performed further analysis of our data with the aim to compare the SUV_{peakW} variability performance with that of SUV_{max-40} (corresponding to a total hottest volume close to 1 mL).

MATERIALS AND METHODS

Patients

Twelve lung cancer patients (2 women, 10 men; average age, 63 y; age range, 43–78 y; 9 non–small cell lung cancer/3 small cell lung cancer) were included in the study, and 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 kg) and 169 cm (range, 157–179 cm), respectively. After 6-h fasting before the tracer injection, preinjection average plasma glucose concentration was 1.00 $g{\cdot}L^{-1}$ (range, 0.90–1.17 $g{\cdot}L^{-1}$).

PET Imaging and Data Processing

^{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). Dynamic PET imaging was performed over the chest for

Received Jun. 5, 2015; revision accepted Sep. 8, 2015.
For correspondence contact: Eric Laffon, Service de Médecine Nucléaire, Hôpital du Haut-Lévèque, Avenue de Magellan, 33604 Pessac, France.
E-mail: elaffon@u-bordeaux2.fr
Published online Nov. 12, 2015.

the study purpose, without respiratory gating, within 60–110 min after injection (1 step, 10 consecutive frames of 2.5 min each), using a Discovery ST PET/CT device (GE Healthcare; 3-dimensional mode without septa; decay correction on). All PET images were reconstructed iteratively (Fourier rebinning plus ordered-subsets expectation maximization; subsets, 32; iterations, 5; 3-dimensional postprocessing filter of Hann, 0.9, 10.0), and the voxel size was $2.73 \times 2.73 \times 3.27$ mm (in-plane and axial, respectively; field of view, 700×700 mm; matrix, 256×256 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 anatomic localization (pitch, 1.675; slice thickness, 3.75 mm; field of view, 500×500 mm; matrix, 512×512 pixels) leading to a voxel volume of 0.0036 mL. Minimal lesion size was assessed with CT either in-plane or axial, which was always larger than 15 mm to minimize partial-volume effects (*6*).

An Advantage 4.6 workstation (GE Healthcare) was used for drawing in each dynamic frame a volume of interest encompassing each ^{18}F-FDG–positive lesion. The method to assess $SUV_{max\text{-}40}$ has been previously described in detail (*3*). Briefly, $SUV_{max\text{-}40}$ was obtained from the histogram representing the percentage of all voxels included in the volume of interest versus SUV. It is the averaged SUV from the 40 hottest voxels—that is, over a total hottest volume of 0.98 mL. SUV_{peakW} defined by Wahl et al. with PERCIST was obtained using the PET-VCAR application of the workstation (GE Healthcare).

Statistical Analysis

For each lesion, a mean SUV_{peakW} and a mean $SUV_{max\text{-}40}$ value and corresponding 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 SD (SDr) was not significantly related to magnitude, and a mean SDr was then calculated: $<SDr>_{peakW}$ and $<SDr>_{N=40}$ (*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 2 SUVs assessed from 2 successive scans that is required to consider a significant difference) were calculated as 1.96* <SDr> and $2^{1/2}$*1.96 <SDr> (95% confidence level [CL]), respectively.

Comparison between $<SDr>_{peakW}$, $<SDr>_{max\text{-}40}$, $<SDr>_{max}$, and $<SDr>_{peak}$, and between mean values over the lesion series of SUV_{peakW}, $SUV_{max\text{-}40}$, SUV_{max}, and SUV_{peak}, that is, $<SUV>_{peakW}$, $<SUV>_{max\text{-}40}$, $<SUV>_{max}$, and $<SUV>_{peak}$, was achieved using a 2-tailed paired *t* test. *P* values of less than 0.05 were considered statistically significant.

RESULTS

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

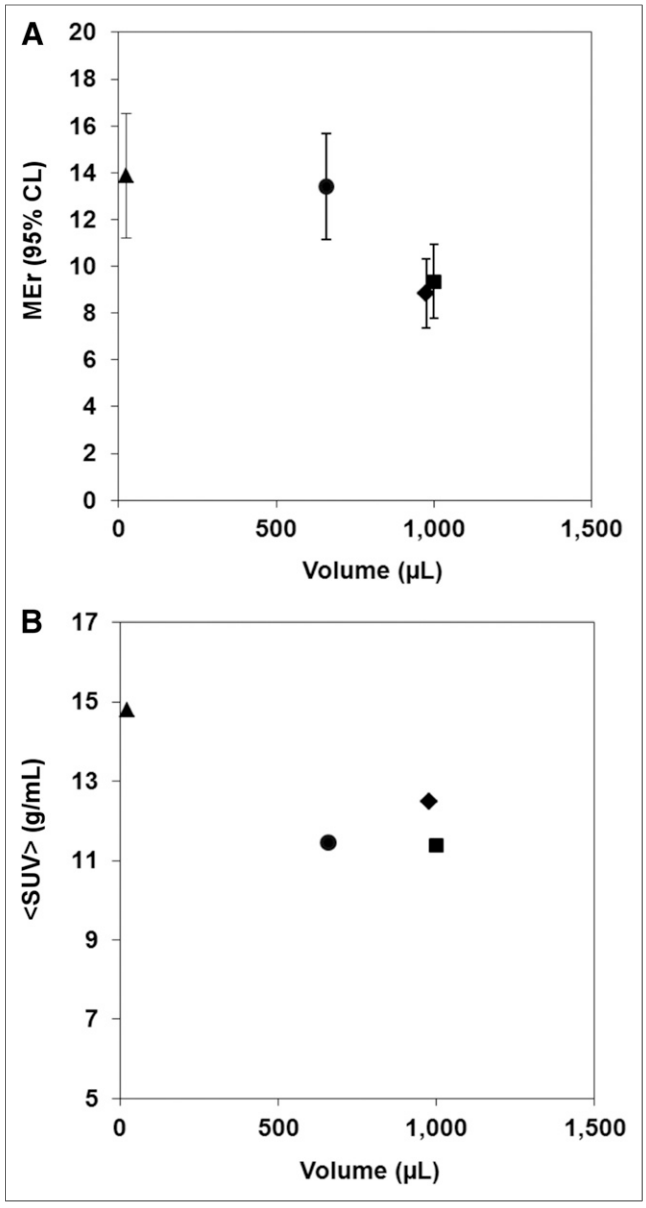

FIGURE 1. (A) MEr comparison of SUV_{peakW} (■) and of $SUV_{max\text{–}40}$ (♦), involving also comparison with MEr of SUV_{max} (▲) and of SUV_{peak} (●) previously published (*3*). Bars represent 95% CLs. Repeatability (R) can be obtained by multiplying MEr by √2. (B) Comparison of $<SUV>_{max}$ (▲), $<SUV>_{peak}$ (●), $<SUV>_{peakW}$ (■), and $<SUV>_{max\text{-}40}$ (♦) over lesion series.

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

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

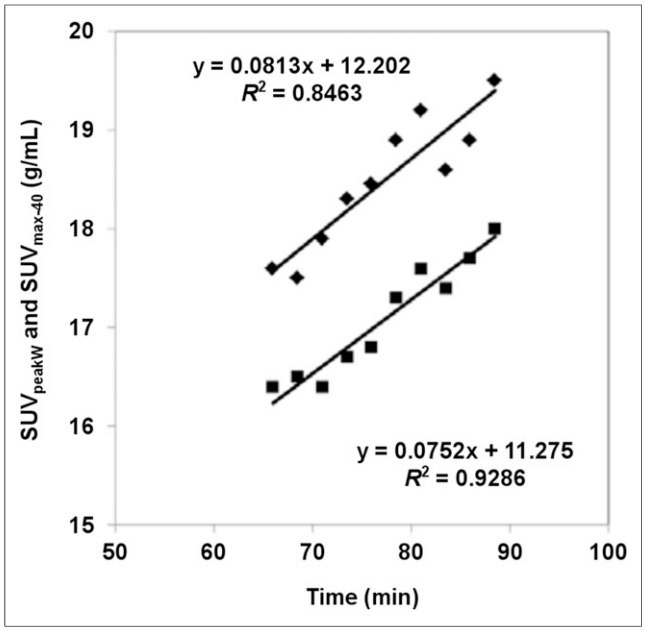

FIGURE 2. SUV_{peakW} (■) and SUV_{max-40} (♦) versus time in typical lesion, showing significant linear correlation (r = 0.96 and 0.92, respectively; 95% reliability).

DISCUSSION

^{18}F-FDG PET imaging in oncology is in need of robust methods enabling the reliable assessment of 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 $SUV_{max,}$ which is obtained from the hottest voxel, Wahl et al. have proposed the use of SUV_{peakW} to reduce SUV outcome variability. SUV_{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 SUV_{peakW} with that of SUV_{max-40}, that is, pooling 40 hottest voxels (total hottest volume of 0.98 mL), irrespective of their location within the lesion. SUV_{peakW} was significantly 9.66% lower (on average) than SUV_{max-40}, and both were significantly lower than SUV_{max} (Fig. 1). SUV_{peakW} and SUV_{max-40} showed close variability performance that was significantly better than that of SUV_{max} (95% CL; $P < 0.0001$; Fig. 1A). Therefore, we suggest that SUV_{max-40} might be superior to SUV_{peakW} for assessing the most metabolically active portions of tumors, with close variability performances for both metrics.

SUV_{peakW} and SUV_{max-40} also showed variability performance that was significantly better than that of an arbitrary SUV_{peak}, defined as SUV_{max} and its 26 neighboring voxels (*3*). This result is consistent with that of Weber et al. in lung cancer patients that used a different volume of interest centered on the SUV_{max} and reported similar variability performance between SUV_{peak} and SUV_{max} (*4*). Furthermore, the current study used a PET dynamic acquisition involving 10 frames (equivalent to 10 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–restest 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 SUV_{max} and SUV_{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), and acquisition (respiration-gated from 15-min list-mode data, including only 2 phases, instead of 10-frame dynamic acquisition) (*9*). In particular, for a 256 × 256 image matrix, Lodge et al. also reported that SUV_{peakW} was significantly lower than SUV_{max}, 35.77% on average, a finding comparable to the 29.85% obtained in the present study. However, although SUV_{peakW} repeatability (R—that is, the minimal relative change between 2 SUVs assessed from 2 successive scans that is required to consider a significant difference) was found to be significantly lower than SUV_{max} R in each study, there was a 2-fold discrepancy about SUV_{peakW} R between Lodge's and the current study: 6.65% versus 13.21%, respectively (95% CL). For comparison, SUV_{max} R was found to be 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.

The 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—that is, averaging SUV from several voxels to lower its variability. However, a significantly lower performance of SUV_{peakW} was found for reporting the hottest parts of the 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 limited only by the voxel size of the PET system used. Furthermore, the SUV_{max-40} metric can be easily implemented in current clinical practice, low intra- or interobserver variability was reported (*5*), and 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: minimal lesion size was larger than 15 mm 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—that is, 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_{max} (1 voxel) (*3*). This suggestion is supported by Hasenclever et al. in interim PET performed in lymphoma patients, who used an arbitrary SUV_{peak} metric involving SUV_{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_{peakW} and SUV_{max-40} were found to significantly increase with

time over the lesion series (Fig. 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_{peakW} outcomes are significantly lower than those of $SUV_{max\text{-}40}$ ($P = 0.002$, 2-tailed sign test).

CONCLUSION

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

REFERENCES

1. Wahl RL, Jacene H, Kasamon Y, Lodge MA. From RECIST to PERCIST: evolving considerations for PET response criteria in solid tumors. *J Nucl Med.* 2009;50(suppl1):122S–50.
2. Vanderhoek M, Perlman SB, Jeraj R. Impact of the definition of peak standardized uptake value on quantification of treatment response. *J Nucl Med.* 2012;53: 4–11.
3. Laffon E, Lamare F, de Clermont H, Burger IA, Marthan R. Variability of average SUV from several hottest voxels is lower than that of SUVmax and SUVpeak. *Eur Radiol.* 2014;24:1964–1970.
4. Weber WA, Gatsonis CA, Mozley PD. Repeatability of ^{18}F-FDG PET-CT in advanced non-small cell lung cancer: prospective assessment in two multicenter trials. *J Nucl Med.* 2015;56:1137–1143.
5. Burger IA, Huser DM, Burger C, von Schulthess GK, Buck A. Repeatability of FDG quantification in tumor imaging: averaged SUVs are superior to SUVmax. *Nucl Med Biol.* 2012;39:666–670.
6. Soret M, Bacharach SL, Buvat I. Partial-volume effect in PET tumor imaging. *J Nucl Med.* 2007;48:932–945.
7. Bland JM, Altman DG. Statistics notes: measurement error proportional to the mean. *BMJ.* 1996;313:106–108.
8. Bland JM, Altman DG. Statistics notes: measurement error. *BMJ.* 1996;313:744–746.
9. Lodge MA, Chaudhry MA, Wahl RL. Noise considerations for PET quantification using maximum and peak standardized uptake value. *J Nucl Med.* 2012;53: 1041–1047.
10. Boellaard R, Delgado-Bolton R, Oyen WJG, et al. FDG PET/CT: EANM procedure guidelines for tumor imaging: version 2.0. *Eur J Nucl Med Mol Imaging.* 2015;42:328–354.
11. Jhaveri K, Linden H. Measuring tumor metabolism by ^{18}F-FDG predicts outcome in a multicenter study: a step off in the right direction. *J Nucl Med.* 2015;56: 1–2.
12. Hasenclever D, Kurch L, Mauz-Körholz, et al. qPET: a quantitative extension of the Deauville scale to assess response in interim FDG-PET scans in lymphoma. *Eur J Nucl Med Mol Imaging.* 2014;41:1301–1308.
13. Ambrosini V, Nicolini S, Caroli P. PET/CT imaging in different types of lung cancer: an overview. *Eur J Radiol.* 2012;81:988–1001.