Noise Considerations for PET Quantification Using Maximum and Peak Standardized Uptake Value

Martin A. Lodge[1], Muhammad A. Chaudhry[1,2], and Richard L. Wahl[1,3]

[1]Division of Nuclear Medicine, Russell H. Morgan Department of Radiology and Radiological Sciences, Johns Hopkins University School of Medicine, Baltimore, Maryland; [2]Tawam Molecular Imaging Center, Al Ain, United Arab Emirates; [3]Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins, Johns Hopkins University School of Medicine, Baltimore, Maryland

In tumor response monitoring studies with ^{18}F-FDG PET, maximum standardized uptake value (SUV_{max}) is commonly applied as a quantitative metric. Although it has several advantages due to its simplicity of determination, concerns about the influence of image noise on single-pixel SUV_{max} persist. In this study, we measured aspects of bias and reproducibility associated with SUV_{max} and the closely related peak SUV (SUV_{peak}) using real patient data to provide a realistic noise context. **Methods:** List-mode 3-dimensional PET data were acquired for 15 min over a single bed position in twenty ^{18}F-FDG oncology patients. For each patient, data were sorted so as to form 2 sets of images: respiration-gated images such that each image had statistical quality comparable to a 3 min/bed position scan, and 5 statistically independent (ungated) images of different durations (1, 2, 3, 4, and 5 min). Tumor SUV_{max} and SUV_{peak} (12-mm-diameter spheric region of interest positioned so as to maximize the enclosed average) were analyzed in terms of reproducibility and bias. The component of reproducibility due to statistical noise (independent of physiologic and other variables) was measured using paired SUVs from 2 comparable respiration-gated images. Bias was measured as a function of scan duration. **Results:** Replicate tumor SUV measurements had a within-patient SD of 5.6% ± 0.9% for SUV_{max} and 2.5% ± 0.4% for SUV_{peak}. SUV_{max} had average positive biases of 30%, 18%, 12%, 4%, and 5% for the 1-, 2-, 3-, 4-, and 5-min images, respectively. SUV_{peak} was also biased but to a lesser extent: 11%, 8%, 5%, 1%, and 4% for the 1-, 2-, 3-, 4-, and 5-min images, respectively. **Conclusion:** The advantages of SUV_{max} are best exploited when PET images have a high statistical quality. For images with noise properties typically associated with clinical whole-body studies, SUV_{peak} provides a slightly more robust alternative for assessing the most metabolically active region of tumor.

Key Words: PET; standardized uptake value; SUV_{max}; SUV_{peak}; treatment monitoring

Standardized uptake value (SUV) analysis (*1,2*) of ^{18}F-FDG PET images is increasingly applied as a practical and effective method to characterize lesions and monitor response to therapy (*3,4*). SUVs have been used in PET for more than 20 y (*5*), but although significant efforts have been made to ensure more consistent patient preparation and data acquisition (*6,7*), there remain some differences in the practical implementation of the technique (*8*). One such area is the method of image analysis and, specifically, the influence of region-of-interest (ROI) definition (*9–11*).

A range of different ROI methods has been reported for SUV determination, including manual definition of tumor boundaries (*12*), automated and semiautomated tumor segmentation algorithms (*13*), and fixed-size ROIs that sample the tumor but do not attempt to conform to the precise tumor outline (peak SUV [SUV_{peak}]) (*14*). Another method, which has been widely adopted, involves determination of SUV using the maximum pixel within the tumor (maximum SUV [SUV_{max}]) (*15*). SUV_{max} has several attractive features, including the fact that it reflects the most metabolically active, and possibly most clinically significant, part of a potentially heterogeneous mass. SUV_{max} is also less susceptible to partial-volume effects than are other, more extended, ROIs (*16*). SUV_{max} can be determined without precise definition of the tumor boundaries and thus has a significant practical advantage. In addition, the maximum pixel value within a tumor can easily be obtained with most existing commercial workstations, making SUV_{max} particularly convenient. A recent review (*17*) found that SUV_{max} was by far the most widely used method of analyzing tumors in quantitative ^{18}F-FDG oncology studies, but despite this wide use, concern remains regarding the vulnerability of single-pixel measurements to image noise (*16*).

SUV_{max} can usually be measured with high reproducibility (*18*) when different readers review the same imaging study (interobserver reproducibility). However, a measure of reproducibility that is more relevant for the task of monitoring response using multiple sequential imaging studies is interstudy reproducibility. Replicate studies performed on the same patient within a short period using identical technique often show poorer interstudy reproducibility for SUV_{max} than do

SUVs based on the mean within larger ROIs (*10,19*). This quality can be attributed to the single-pixel nature of SUV_{max}. There is little volume averaging, making SUV_{max} more vulnerable to statistical noise in the image data. In addition, computer simulation (*9*) has shown that SUV_{max} is associated with an increasing positive bias as image noise increases. Use of an SUV index with poor interstudy reproducibility limits the ability to reliably quantify real changes in tumor metabolism. Additionally, noise-dependent bias presents a potential problem for the standardization of data collection in multicenter trials because the statistical quality of PET images varies considerably among sites, and data from different centers may not be directly comparable.

Recent work (*20*) has suggested that tumor SUVs are consistent over a range of noise levels, allowing scope for lower administered activities during follow-up PET studies. Although the intention to reduce patient radiation exposure is worthy, the detrimental effect of a loss of statistical image quality has also been noted (*21*). In the present study, we focused specifically on these noise issues, with particular reference to the SUV_{max} and SUV_{peak} metrics because they may be particularly sensitive to image statistical quality. Although SUV_{max} is known to be more vulnerable to noise than SUVs based on larger ROIs, the significance of this issue in the context of numerous other sources of variability and bias has not been established. Some investigators have found that the reproducibility of SUV_{max} was broadly similar to SUV metrics based on larger ROIs (*22*) or that reproducibility of SUV_{max} was only modestly worse (*23*). In the former report, data were acquired in a multicenter setting and the influence of different ROI methodologies may have been obscured by other effects. In the latter report, a series of phantom experiments was performed that provided valuable insight but may not have fully reflected the noise properties associated with clinical images. In the current paper, we measure aspects of bias and reproducibility for both SUV_{max} and SUV_{peak} using real patient data to provide a realistic noise context.

MATERIALS AND METHODS

Data Acquisition

A Discovery VCT (RX) (*24*) PET/CT system (GE Healthcare) was used to acquire image data for 20 patients with known or suspected malignancies in the chest or abdomen (lung, $n = 6$; liver, $n = 7$; pancreas, $n = 7$). Patients (mean weight $\pm$ SD, 74 $\pm$ 14 kg) were prepared according to a standard oncology protocol, and whole-body PET/CT data were acquired approximately 1 h after administration of 624 $\pm$ 83 MBq of ^{18}F-FDG. After completion of the whole-body study, additional image data were acquired over the tumor site using a single PET bed position (15 cm in the axial direction). These localized PET data were acquired in list mode, and an external camera system (Varian Medical Systems) was used to monitor respiratory motion. The list-mode PET data were acquired without septa for 15 min, starting 147 $\pm$ 37 min after injection. CT data were acquired over the same scan range using the following parameters: 64-slice multichannel CT, 120 kVp, approximately 150 mA, 0.5-s tube rotation, and pitch of 0.984. Institutional review board approval was obtained for a retrospective analysis of these single-bed-position data.

Phase-based respiration-gated PET images were reconstructed using a total of 5 gates over the respiratory cycle. PET images were reconstructed according to our current clinical oncology protocol, which involves 3-dimensional ordered-subset expectation maximization, 2 iterations, 21 subsets, a gaussian filter of 3 mm in full width at half maximum, model-based scatter correction (*25*), and CT-based attenuation correction (*26*). The standard 128 × 128 image matrix (4.7 × 4.7 × 3.3 mm voxel size) and a separate 256 × 256 (2.3 × 2.3 × 3.3 mm voxel size) matrix were used. As well as the respiratory-gated images, the 15-min list-mode data were sorted so as to form multiple (ungated) images of different statistical quality. Five sinograms of different durations (1, 2, 3, 4, and 5 min) were formed, along with an additional, summed, 15-min sinogram. The data were sorted in the following order (ignoring the respiratory triggers): 0–1, 1–3, 3–6, 6–10, 10–15, and 0–15 min. The first 5 frames were thus statistically independent, whereas the last 15-min dataset included all data and formed a low-noise reference. Images were reconstructed using the technique described above and a 128 × 128 image matrix.

Reproducibility

Test–retest studies (*10,19,27–30*) allow assessment of the overall interstudy reproducibility of the SUV technique, which includes numerous physiologic and methodologic sources of variability. To isolate the contribution of only statistical noise, we analyzed pairs of images extracted from the same respiration-gated image series. For each patient study, images from gates 3 and 4 were extracted from the respiration-gated series. Selection of these data was based on a visual assessment that these gates captured the tumors in approximately the same position (Fig. 1). Each image was effectively acquired for one fifth of the total acquisition time (1/5 × 15 min = 3 min) and thus had a statistical quality that was representative of typical whole-body protocols (3 min/bed position). Furthermore, because of the nature of the gated acquisition, any 2 images from the

FIGURE 1. Example images from representative patient study showing how respiration-gated images were used to obtain statistically independent SUV estimates for assessment of reproducibility. Panels A, B, C, D, and E show images derived from respiratory gates 1, 2, 3, 4, and 5, respectively. Red arrow in A indicates target lesion. Slice that contained SUV_{max} is indicated below images in panels A–E and, as expected, occurred in slightly different slices for each gate. Gates 3 and 4 reflected similar degrees of respiratory motion and were used to obtain 2 statistically independent SUV measurements. Main images are shown in same linear color scale, which was arbitrarily limited to a 0–8 g/mL range. Inset images are enlarged versions of target lesion, displayed in expanded color scale (0–22 g/mL), with maximum pixel highlighted in red.
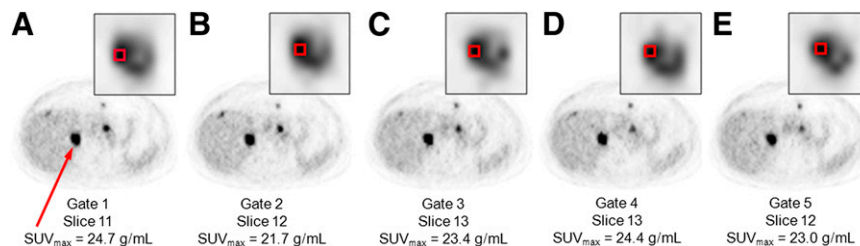
respiration-gated series were acquired over essentially the same period. Therefore, problems associated with tracer redistribution and radioactive decay were avoided. A quantitative assessment of the extent to which tumor SUVs derived from these 2 images were comparable and an estimate of their variability was determined in the following way.

SUV$_{max}$ was measured for the target lesion in both images extracted from each patient's respiratory series. A large, spheric ROI was defined so as to encapsulate the entire tumor, and SUV$_{max}$ was determined (PMOD Technologies Ltd.). In addition to generating SUV data using the maximum pixel value (SUV$_{max}$), we also used the SUV$_{peak}$ methodology. SUV$_{peak}$ was determined by averaging the image data within a 12-mm-diameter spheric ROI (strictly volume of interest) that was positioned within the tumor so as to maximize the enclosed average. In general, this ROI included the maximum pixel but was not constrained to do so. This analysis was performed separately for the PET images reconstructed with 128×128 and 256×256 image matrices. In this way, the component of SUV reproducibility that can be attributed to statistical noise was estimated under 4 conditions: SUV$_{max}$, 128×128 image matrix; SUV$_{max}$, 256×256 image matrix; SUV$_{peak}$, 128×128 image matrix; and SUV$_{peak}$, 256×256 image matrix.

For each of these 4 reconstruction and analysis conditions, the difference between corresponding SUV measurements was recorded for each patient and reproducibility was assessed using the Bland–Altman approach (*31*). To reflect the way that SUVs are commonly used in response-monitoring studies, we selected 1 measurement as the baseline and present the difference between the 2 measurements as a percentage of this baseline value. The relative difference *d* is thus defined as

$$d = \frac{\text{SUV}_{\text{Gate 4}} - \text{SUV}_{\text{Gate 3}}}{\text{SUV}_{\text{Gate 3}}} \times 100, \qquad \text{Eq. 1}$$

where the SUV subscripts, in this case, denote the respiration-gated images from which the SUVs were determined. Reproducibility (strictly, repeatability, but we retain the word *reproducibility* for compatibility with previous reports) was quantified using 3 related metrics. The within-patient SD was the SD of *d* over all patients divided by $\sqrt{2}$ and reflects variability in a single measurement. The coefficient of repeatability was defined as $1.96 \times \sqrt{2} \times$ within-patient SD and reflects the 95% limits of agreement that can be expected for the difference between 2 repeated measurements. For comparison with some previous reports, we also quote the mean absolute percentage difference, which is the absolute value of *d* averaged over all patients.

Bias

SUV bias was determined as a function of image statistical quality using the ungated images of different durations (1, 2, 3, 4, 5, and 15 min). For each image, both SUV$_{max}$ and SUV$_{peak}$ were determined using the ROI methods described above. In addition, a third SUV measurement was made using a large, 36-mm-diameter, spheric ROI centered on the tumor. In this case, SUV$_{mean}$ was determined using the average of the pixels encompassed by this ROI. Whereas the 2 smaller ROIs may be vulnerable to image noise, these effects were expected to be substantially suppressed by volume averaging with the 36-mm ROI. The purpose of this large ROI was to serve as a quality assurance tool to confirm that bias seen with the smaller ROIs was due to statistical effects associated with image sampling and not systematic bias in the underlying image data. To account

for variations in the absolute magnitude of the tumor SUVs between the different patients, the SUVs determined from the low-noise 15-min data were used to normalize each of the corresponding SUVs from the shorter-scan-duration images. Mean bias was then estimated for each ROI method by averaging the normalized SUV data from all patients at each scan duration.

RESULTS

Reproducibility

Two patient studies were excluded from the reproducibility analysis, one because the respiratory gating failed and the other because the tumor was not evaluable because of low tracer uptake (SUV$_{max}$, 1.8 g/mL; SUV$_{peak}$, 1.2 g/mL). Mean SUV$_{max}$ for all tumors was 9.1 ± 6.3 g/mL for respiratory gate 3 and 9.0 ± 6.3 g/mL for respiratory gate 4, using the 128×128 images. For the 256×256 images, mean SUV$_{max}$ was 9.3 ± 6.2 g/mL and 9.3 ± 6.5 g/mL for gates 3 and 4, respectively. Mean SUV$_{peak}$ was 6.9 ± 4.8 g/mL for respiratory gate 3 and 6.8 ± 4.9 g/mL for respiratory gate 4, using the 128×128 images. For the 256×256 images, mean SUV$_{peak}$ was 6.8 ± 4.6 g/mL and 6.7 ± 4.7 g/mL for gates 3 and 4, respectively. In each case, paired *t* tests indicated no statistically significant differences ($P > 0.23$) between the SUV measurements obtained from the 2 respiration-gated images, supporting our use of these data as statistically independent replicates. Reproducibility of the tumor SUV measurements is shown in Figure 2 for the various cases considered. Table 1 quantifies the mean absolute percentage difference, the within-patient SD, and the repeatability, which is approximately equal to the 95% limits of agreement indicated in Figure 2. Within-patient SD of SUV$_{max}$ was $5.6\% \pm 0.9\%$ for the 128×128 image matrices and $6.5\% \pm 1.1\%$ for the 256×256 image matrices. Compared with SUV$_{max}$, SUV$_{peak}$ gave rise to improved within-patient SD: $2.5\% \pm 0.4\%$ for the 128×128 image matrices and $2.4\% \pm 0.4\%$ for the 256×256 image matrices.

Bias

Figure 3 shows the effect of decreasing scan duration (degrading image statistical quality) on tumor SUV measurements. Relative to the low-noise 15-min images, SUV$_{max}$ was on average biased by factors of 1.30, 1.18, 1.12, 1.04, and 1.05 for the 1-, 2-, 3-, 4-, and 5-min images, respectively. Expressed as percentages, SUV$_{max}$ had positive biases of 30%, 18%, 12%, 4%, and 5%, whereas SUV$_{peak}$ was also biased but to a lesser extent: 11%, 8%, 5%, 1%, and 4% for the 1-, 2-, 3-, 4-, and 5-min images, respectively. The difference between the bias for SUV$_{max}$ and SUV$_{peak}$ was significant for the 1-, 2-, 3-, and 4-min images (paired *t* test, $P < 0.001$) but was not significant for the 5-min images ($P = 0.19$). In individual patient studies, the bias observed with SUV$_{max}$ and SUV$_{peak}$ could be substantially higher than the average. The SD of the bias was 26%, 14%, 8%, 6%, and 9% with SUV$_{max}$ and 16%, 12%, 7%, 3%, and 8% with SUV$_{peak}$ for the 1-, 2-, 3-, 4-, and 5-min images, respectively. Average bias was lower with SUV$_{mean}$ obtained from the 36-mm ROI
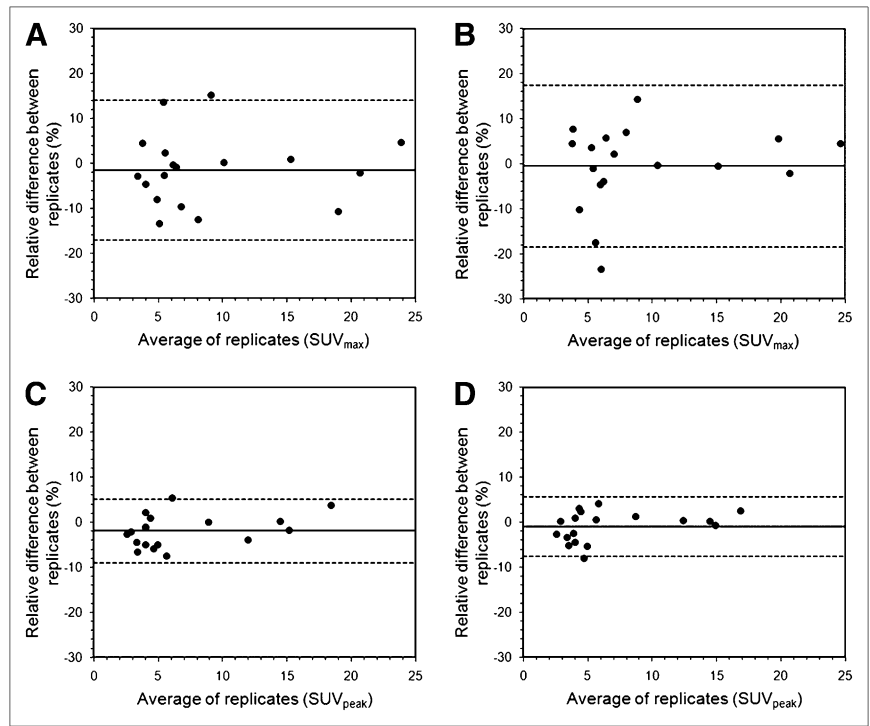
FIGURE 2. Relative differences between replicate tumor SUV measurements made under 4 different conditions: SUV_{max}, 128 × 128 image matrix (A); SUV_{max}, 256 × 256 image matrix (B); SUV_{peak}, 128 × 128 image matrix (C); and SUV_{peak}, 256 × 256 image matrix (D). Relative difference (*d*) is defined as difference between 2 measurements, expressed as percentage of baseline measurement. Dashed lines indicate 95% limits of agreement and are equal to mean of *d* ± 1.96 × SD of *d*.

(3%, 3%, 2%, 0%, and 0%, respectively) and was not significantly different from zero (single-sample *t* test, $P > 0.05$).

DISCUSSION

The statistical limitations of SUV, and in particular SUV_{max}, have been appreciated for some time (*9*). However, recent interest in the role of PET for monitoring tumor response to treatment has generated renewed interest in the topic (*32*), partly because the reproducibility of the method imposes a minimum change in SUV that is required to indicate a statistically significant change in the tumor. Overall SUV reproducibility includes components due to biologic and protocol issues, but much recent work (*23,33,34*) has focused on the instrument and analysis components of reproducibility. These studies used phantom experiments that approximated the noise environment encountered in clinical imaging and assessed bias and reproducibility in single-center and multicenter settings. The data presented in the current paper augment these studies by measuring aspects of reproducibility and bias in real patient images, thus accurately reflecting the statistical quality that is encountered in the clinical environment. Because the numerous factors that influence image statistical quality, and

their variability between patients, are hard to accurately capture with current phantom designs, the use of real patient data in the present study is significant.

In this study, we found the within-patient SD for tumor SUV_{max} to be 5.6% ± 0.9% under conditions typical of whole-body oncology protocols. Comparing this value with previously published data for the overall reproducibility of SUV_{max} is slightly complicated by the use of different metrics, but despite this complication, the literature is quite consistent. The mean absolute percentage difference between successive SUV_{max} measurements has been reported by 3 studies to be 11.3% ± 8.0% (*29*), 13% ± 12% (*10*), and 16.1% ± 10.5% (*35*). The higher value in the last study may be due to the fact that the measurements were made on 2 different scanner systems: one PET/CT and the other PET only. Because the mean absolute percentage difference approximates the within-patient SD (Table 1), these data are in good agreement with 2 other publications, which quoted 11%–12% (*22*) and 11.8% (within-patient SD, $16.7\%/\sqrt{2} = 11.8\%$) (*30*). Direct comparison of these data with those of Nahmias and Wahl (*19*) is not possible because their results are presented in absolute SUV units, as opposed to a relative change. However,

TABLE 1
Reproducibility of Tumor SUV Measurements (*n* = 18) for 4 Combinations of ROI Definition and Image Matrix Size

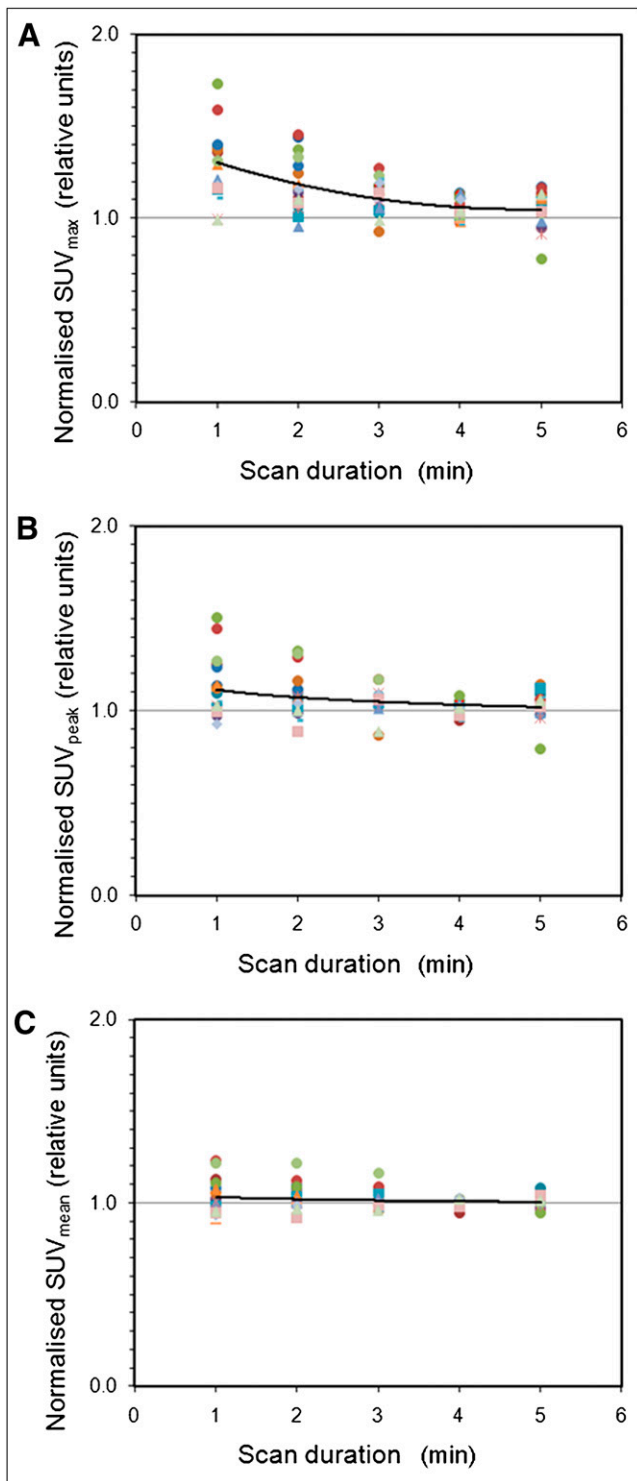| Combination	Mean absolute difference (%)	Within-patient SD (%)	Repeatability (%)
SUV_{max}, 128 × 128	6.1 ± 5.2	5.6 ± 0.9	15.6 ± 2.6
SUV_{max}, 256 × 256	6.6 ± 6.2	6.5 ± 1.1	18.0 ± 3.0
SUV_{peak}, 128 × 128	3.3 ± 2.3	2.5 ± 0.4	7.0 ± 1.2
SUV_{peak}, 256 × 256	2.7 ± 2.2	2.4 ± 0.4	6.6 ± 1.1

FIGURE 3. Effect of varying scan duration on tumor SUV_{max} (A), SUV_{peak} (B), and SUV_{mean} (C). Each colored data series represents normalized SUVs for particular patient. Normalization is relative to corresponding SUV measurement obtained from low-noise 15-min image. Solid lines join averages for all patients. Increasing positive bias is seen for SUV_{max} and, to lesser extent, SUV_{peak} as scan duration decreases (image statistical quality degrades).

their 95% confidence intervals of ± 2.23 SUV units (within-patient SD, 2.23/2.77 = 0.80) and a mean SUV_{max} of approximately 8 SUV units indicate reproducibility results that are consistent with the previously mentioned publications.

The within-patient SD of 5.6% $\pm$ 0.9% for tumor SUV_{max} measured in the present study is lower than the literature values because the reports mentioned above include variability due to multiple sources, not simply image noise. These factors include differences in patient preparation, plasma glucose levels, and tracer uptake periods, as well as potentially real changes in tumor metabolism between studies performed on separate days. In addition, technical errors related to such things as scanner calibration and clock synchronization may also contribute. It is worth noting that for SUV_{max}, the component of variability that can be attributed to image noise accounts for approximately half the overall variability. Image statistical quality is therefore not a negligible consideration, at least when uptake measurements are derived from single-pixel SUV_{max}. Although the previously reported values of 11%–13% for overall within-patient SD may seem relatively low, they imply 95% limits of agreement for the difference between repeated measurements of around $\pm 30\%$ (2.77 $\times$ 11% = 30%). In other words, repeated SUV_{max} measurements that differ by up to 30% should be expected simply from measurement error. The excellent interobserver reproducibility that has been reported (18) for SUV_{max} should not be confused with the within-patient SD, which better reflects the variability that is encountered in response-monitoring studies involving sequential imaging.

SUV_{peak} provides a mechanism for improving reproducibility for SUV measurements of the most metabolically active tumor region. The component of the overall within-patient SD due to image noise was reduced from 5.6% $\pm$ 0.9% with SUV_{max} to 2.5% $\pm$ 0.4% with SUV_{peak} (128 $\times$ 128 image matrix). SUV_{peak} is by no means a new proposal, and its use predates by many years the adoption of the term *SUV_{peak}*. In this work, we have implemented SUV_{peak} using a fixed-size 12-mm-diameter spheric ROI (17), positioned so as to maximize the enclosed average. Compared with SUV_{max}, larger bias due to the partial-volume effect is expected for small tumors, and this is clearly a limitation of the SUV_{peak} approach. However, greater volume averaging with SUV_{peak} was seen to improve reproducibility and offers a slightly more robust alternative to SUV_{max}. Achieving this advantage in clinical practice requires consistent placement of the peak ROI, something that is not trivial if performed manually. Fortunately, the inclusion of automated SUV_{peak} algorithms in the software of many commercial vendors promises to make this index more widely available and potentially as convenient to use as SUV_{max}. Another potential advantage of SUV_{peak} over SUV_{max} suggested by the data in Figure 2 and Table 1 may be that the reproducibility of SUV_{peak} is less affected by changes in pixel size. If confirmed, this property could have advantages for multicenter studies, in which images from different sites are likely to have pixels of different sizes.

In addition to limiting the reproducibility of SUV measurements, image noise also has the potential to introduce bias. Figure 3 provides clinical data confirming the potential for significant positive bias when the maximum pixel value is used to characterize PET uptake measurements. This trend is a consequence of the way SUV_{max} is defined. In a region of uniform tracer accumulation, statistical noise gives rise to a range of nonuniform pixel values. When one is considering the mean within an extended ROI, these pixels tend to average out, resulting in an unbiased estimate of the underlying signal (not withstanding other sources of error). SUV_{max}, however, consistently takes the highest pixel value and therefore tends to overestimate the underlying average. Figure 3A shows a mean positive bias for SUV_{max} of 30% ± 26% for 1-min acquisitions. SUV_{peak}, in contrast, was biased by only 11% ± 16% for the same 1-min images (Fig. 3B). Noise-dependent bias of SUV_{max} has been previously reported in relation to computer simulations (9), experimental phantoms (20), and respiration-gated patient studies (36). Murray et al. (20) noted this bias effect in phantom studies with a time-of-flight PET system but did not observe it in their patient data. A possible explanation might be that, although their phantom images were statistically independent, their patient images may not have been and a potentially misleading correlation between SUVs may have resulted.

We acknowledge several limitations in our present work. The list-mode data were acquired 147 ± 37 min after ^{18}F-FDG administration, and thus significant additional radioactive decay of the tracer (additional decay factor, 0.58) would be expected, compared with the more conventional oncology start time of 60 min. Our protocol attempted to compensate for this additional decay via the higher ^{18}F-FDG activities that were administered. To approximate a typical patient administration of 370 MBq, an activity of 638 MBq would be required (370 MBq/0.58). In the present study, an average of 624 ± 83 MBq was administered, suggesting that the effect of delayed scanning may have been adequately compensated. Another limitation of our protocol was the use of nongated CT for attenuation correction of the respiration-gated PET series. Ideally, the CT data would have been gated in a similar way to the PET, allowing more accurate attenuation correction. This approach was not adopted because of the increased patient radiation dose that would have resulted. At least for the abdominal lesions, we believe that this may not have been a major limitation, because although respiratory motion can be significant in the abdomen, attenuation differences between abdominal organs are small and errors due to slightly misaligned CT are expected to be minimal. A further limitation is the fact that we do not present data for the various different tumor segmentation algorithms that have been proposed. Although we recognize this limitation, it was our intention to focus on SUV_{max} and SUV_{peak}, because they are widely used metrics that may be particularly vulnerable to image noise. Finally, the data presented in this report are strictly applicable only to the scanner model and protocol that were used. Although similar trends are expected on other scanner systems, the magnitude of the effects may differ if different acquisition and reconstruction protocols are used.

Although the issue of statistical noise and its effect on SUV_{max} has been previously explored (9,10), the subject bears reexamination in light of moves toward lower-activity protocols (20) and shorter data acquisitions (37). Although both developments are welcome in principle, the potential for increased image noise should not be overlooked when SUV_{max} is to be used. Given the current interest in tumor quantification and the fact that SUV_{max} has become the quantitative metric of choice for many centers, additional data on the influence of image noise in real patient studies is timely. This report serves as a reminder of these statistical limitations and we hope will contribute to improved accuracy and reproducibility of quantitative PET studies.

CONCLUSION

The advantages of SUV_{max} are best exploited when PET images have a high statistical quality. For images with noise properties typically associated with clinical whole-body studies, SUV_{peak} provides a slightly more robust alternative for assessing the most metabolically active region of tumor.

DISCLOSURE STATEMENT

The costs of publication of this article were defrayed in part by the payment of page charges. Therefore, and solely to indicate this fact, this article is hereby marked "advertisement" in accordance with 18 USC section 1734.

REFERENCES

1. Huang S-C. Anatomy of SUV. *Nucl Med Biol.* 2000;27:643–646.
2. Thie JA. Understanding the standardized uptake value, its methods, and implications for usage. *J Nucl Med.* 2004;45:1431–1434.
3. Lin C, Itti E, Haioun C, et al. Early ^{18}F-FDG PET for prediction of prognosis in patients with diffuse large B-cell lymphoma: SUV-based assessment versus visual analysis. *J Nucl Med.* 2007;48:1626–1632.
4. Weber WA, Petersen V, Schmidt B, et al. Positron emission tomography in non-small-cell lung cancer: prediction of response to chemotherapy by quantitative assessment of glucose use. *J Clin Oncol.* 2003;21:2651–2657.
5. Strauss LG, Clorius JH, Schlag P, et al. Recurrence of colorectal tumors: PET evaluation. *Radiology.* 1989;170:329–332.
6. Shankar LK, Hoffman JM, Bacharach S, et al. Consensus recommendations for the use of ^{18}F-FDG PET as an indicator of therapeutic response in patients in National Cancer Institute trials. *J Nucl Med.* 2006;47:1059–1066.
7. Boellaard R. Standards for PET image acquisition and quantitative data analysis. *J Nucl Med.* 2009;50(suppl):11S–20S.
8. Adams MC, Turkington TG, Wilson JM, Wong TZ. A systematic review of the factors affecting accuracy of SUV measurements. *AJR.* 2010;195:310–320.
9. Boellaard R, Krak NC, Hoekstra OS, Lammertsma AA. Effects of noise, image resolution, and ROI definition on the accuracy of standard uptake values: a simulation study. *J Nucl Med.* 2004;45:1519–1527.

10. Krak NC, Boellaard R, Hoekstra OS, Twisk JWR, Hoekstra CJ, Lammertsma AA. Effects of ROI definition and reconstruction method on quantitative outcome and applicability in a response monitoring trial. *Eur J Nucl Med Mol Imaging.* 2005;32: 294–301.

11. Westerterp M, Pruim J, Oyen W, et al. Quantification of FDG PET studies using standardised uptake values in multi-centre trials: effects of image reconstruction, resolution and ROI definition parameters. *Eur J Nucl Med Mol Imaging.* 2007; 34:392–404.

12. Dimitrakopoulou-Strauss A, Hoffmann M, Bergner R, et al. Prediction of short-term survival in patients with advanced nonsmall cell lung cancer following chemotherapy based on 2-deoxy-2-[F-18]fluoro-D-glucose-positron emission tomography: a feasibility study. *Mol Imaging Biol.* 2007;9:308–317.

13. Zaidi H, El Naqa I. PET-guided delineation of radiation therapy treatment volumes: a survey of image segmentation techniques. *Eur J Nucl Med Mol Imaging.* 2010;37:2165–2187.

14. Benz MR, Czernin J, Allen-Auerbach MS, et al. FDG-PET/CT imaging predicts histopathologic treatment responses after the initial cycle of neoadjuvant chemotherapy in high-grade soft-tissue sarcomas. *Clin Cancer Res.* 2009;15:2856–2863.

15. Borst GR, Belderbos JSA, Boellaard R, et al. Standardised FDG uptake: a prognostic factor for inoperable non-small cell lung cancer. *Eur J Cancer.* 2005;41: 1533–1541.

16. Soret M, Bacharach SL, Buvat I. Partial-volume effect in PET tumor imaging. *J Nucl Med.* 2007;48:932–945.

17. 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(suppl):122S–150S.

18. Benz MR, Evilevitch V, Allen-Auerbach MS, et al. Treatment monitoring by ^{18}F-FDG PET/CT in patients with sarcoma: interobserver variability of quantitative parameters in treatment-induced changes in histopathologically responding and nonresponding tumors. *J Nucl Med.* 2008;49:1038–1046.

19. Nahmias C, Wahl LM. Reproducibility of standardized uptake value measurements determined by ^{18}F-FDG PET in malignant tumors. *J Nucl Med.* 2008;49: 1804–1808.

20. Murray I, Kalemis A, Glennon J, et al. Time-of-flight PET/CT using low-activity protocols: potential implications for cancer therapy monitoring. *Eur J Nucl Med Mol Imaging.* 2010;37:1643–1653.

21. McDermott G, Iball G, Scarsbrook A. The potential pitfalls of low-activity protocols in PET/CT imaging. *Eur J Nucl Med Mol Imaging.* 2011;38:185–187.

22. Velasquez LM, Boellaard R, Kollia G, et al. Repeatability of ^{18}F-FDG PET in a multicenter phase 1 study of patients with advanced gastrointestinal malignancies. *J Nucl Med.* 2009;50:1646–1654.

23. Doot RK, Scheuermann JS, Christian PE, Karp JS, Kinahan PE. Instrumentation factors affecting variance and bias of quantifying tracer uptake with PET/CT. *Med Phys.* 2010;37:6035–6046.

24. Kemp BJ, Kim C, Williams JJ, Ganin A, Lowe VJ. NEMA NU 2-2001 performance measurements of an LYSO-based PET/CT system in 2D and 3D acquisition modes. *J Nucl Med.* 2006;47:1960–1967.

25. Ollinger JM. Model-based scatter correction for fully 3D PET. *Phys Med Biol.* 1996;41:153–176.

26. Kinahan PE, Townsend DW, Beyer T, Sashin D. Attenuation correction for a combined 3D PET/CT scanner. *Med Phys.* 1998;25:2046–2053.

27. Minn H, Zasadny KR, Quint LE, Wahl RL. Lung cancer: reproducibility of quantitative measurements for evaluating 2-[F-18]-fluoro-2-deoxy-D-glucose uptake at PET. *Radiology.* 1995;196:167–173.

28. Weber WA, Ziegler SI, Thodtmann R, Hanauske A-R, Schwaiger M. Reproducibility of metabolic measurements in malignant tumors using FDG PET. *J Nucl Med.* 1999;40:1771–1777.

29. Nakamoto Y, Zasadny KR, Minn H, Wahl RL. Reproducibility of common semiquantitative parameters for evaluating lung cancer glucose metabolism with positron emission tomography using 2-deoxy-2-[^{18}F]fluoro-D-glucose. *Mol Imaging Biol.* 2002;4:171–178.

30. Hatt M, Cheze-Le Rest C, Aboagye EO, et al. Reproducibility of ^{18}F-FDG and 3′-deoxy-3′-^{18}F-fluorothymidine PET tumor volume measurements. *J Nucl Med.* 2010;51:1368–1376.

31. Bland JM, Altman DG. Measuring agreement in method comparison studies. *Stat Methods Med Res.* 1999;8:135–160.

32. Kinahan PE, Doot RK, Wanner-Roybal M, et al. PET/CT assessment of response to therapy: tumor change measurement, truth data, and error. *Transl Oncol.* 2009;2:223–230.

33. Fahey FH, Kinahan PE, Doot RK, Kocak M, Thurston H, Poussaint TY. Variability in PET quantitation within a multicenter consortium. *Med Phys.* 2010;37:3660–3666.

34. Schwartz J, Humm JL, Gonen M, et al. Repeatability of SUV measurements in serial PET. *Med Phys.* 2011;38:2629–2638.

35. Kamibayashi T, Tsuchida T, Demura Y, et al. Reproducibility of semi-quantitative parameters in FDG-PET using two different PET scanners: influence of attenuation correction method and examination interval. *Mol Imaging Biol.* 2008;10:162–166.

36. Liu C, Alessio A, Pierce L, et al. Quiescent period respiratory gating for PET/CT. *Med Phys.* 2010;37:5037–5043.

37. Halpern BS, Dahlbom M, Quon A, et al. Impact of patient weight and emission scan duration on PET/CT image quality and lesion detectability. *J Nucl Med.* 2004;45:797–801.