# A Systematic Review of the Factors Affecting Accuracy of SUV Measurements

Michael C. Adams[1]
Timothy G. Turkington[1,2]
Joshua M. Wilson[2]
Terence Z. Wong[2,3]

**OBJECTIVE.** There is growing interest in using PET/CT for evaluating early response to therapy in cancer treatment. Although widely available and convenient to use, standardized uptake value (SUV) measurements can be influenced by a variety of biologic and technologic factors. Many of these factors can be addressed with close attention to detail and appropriate quality control. This article will review factors potentially affecting SUV measurements and provide recommendations on ways to minimize when using serial PET to assess early response to therapy.

**CONCLUSION.** Scanner and reconstruction parameters can significantly affect SUV measurements. When using serial SUV measurements to assess early response to therapy, imaging should be performed on the same scanner using the same image acquisition and reconstruction protocols. In addition, attention to detail is required for accurate determination of the administered radiopharmaceutical dose.

PET using ^{18}F-FDG is routinely used for initial staging and follow-up of oncology patients [1]. FDG is a glucose analog that accumulates preferentially in malignant cells because of their higher glucose metabolism [2]. FDG PET has proven value in patients with lung cancer, melanoma, lymphoma, colorectal cancer, esophageal cancer, breast cancer, cervix cancer, and head and neck malignancies [3]. In general, higher-grade and less-differentiated tumors are associated with higher levels of FDG accumulation [4]. In addition, higher FDG accumulation is associated with poorer prognosis for some tumor types [5]. However, benign conditions such as granulomatous disease and chronic inflammatory processes also can be associated with high FDG accumulation [6]. Malignant processes are generally associated with higher FDG accumulation than is seen in inflammatory processes [7]; however, there is considerable overlap, and it may not be possible to distinguish malignant from benign uptake in some cases.

A major advantage of PET is the ability to quantify radiotracer accumulation. The most common parameter used to measure tracer accumulation in PET studies is the standardized uptake value (SUV). The SUV is a semiquantitative measure of normalized radioactivity concentration in PET images:

$$SUV = \frac{activity\ concentration\ in\ tissue}{injected\ activity\ /\ body\ size} \quad (1)$$

To measure SUV, a 2D or 3D region of interest (ROI) is positioned centrally within a target (i.e., tumor) using an interactive workstation. The measured radioactivity within the ROI is normalized to the average radioactivity concentration in the body, which is approximated as the injected dose divided by patient body size. Common body size measurements are based on the patient's body weight, lean body mass, or body surface area, with body weight being the most frequently used.

There are two common ways of reporting SUV: the mean or maximum SUV of all voxels within the ROI ($SUV_{mean}$ and $SUV_{max}$, respectively). $SUV_{mean}$ incorporates information from multiple voxels, making it less sensitive to image noise. However, measured $SUV_{mean}$ will vary depending on which voxels are included in the average, so it is sensitive to ROI definition and is subject to intra- and interobserver variability [8]. $SUV_{max}$ is the highest voxel value within the ROI, so it is independent of ROI definition (assuming the voxel with the highest activity concentration is included) but more susceptible to noise [9]. $SUV_{max}$ is most conveniently measured by surrounding the target lesion with a 3D ROI, taking care to avoid including extraneous regions of high activity, such as the

urinary bladder. Alternatively, 2D ROIs are drawn on multiple axial slices to determine the highest activity within the target. Currently, $SUV_{max}$ is most commonly used because it is less observer-dependent and more reproducible than $SUV_{mean}$.

Some groups have advocated a hybrid SUV measurement, $SUV_{peak}$, that includes a local average SUV value in a group of voxels surrounding the voxel with highest activity; the concept is to maintain the reproducibility of $SUV_{max}$ with improved statistics to reduce noise [10]. This introduces a tradeoff in which the $SUV_{peak}$ value is not likely to be as close to the physiologic radioactivity concentration as $SUV_{max}$ is in small lesions. Using $SUV_{peak}$ is a rough equivalent of performing extra smoothing on the image and then selecting the maximum smoothed pixel value; the effects of smoothing are discussed in detail later. Although currently under investigation, $SUV_{peak}$ has not been implemented in a standardized fashion. It should be noted that some articles refer to $SUV_{max}$ as $SUV_{peak}$, so readers should check the $SUV_{peak}$ definition an article uses.

Quantitative measurements such as SUV for interpreting FDG PET scans are limited because of the considerable overlap between SUV measurements in malignant and benign lesions. Most clinical PET interpretations are based on visual assessment of the FDG accumulation as well as the pattern of disease. PET scans after completion of therapy generally have changes that are evident on visual interpretation. As a consequence, slight variability in SUV measurements does not affect the clinical interpretation in PET performed for initial staging or follow-up after completion of therapy.

However, the recent interest in using FDG PET to evaluate early response to therapy has made quantitative measurements much more important. In this application, FDG PET scans are obtained early in the course of therapy to assess tumor response. Early response detected with PET may predict ultimate response to the therapy, whereas lack of response suggests that an alternate therapy should be considered. When evaluating PET for early response to therapy, changes may be subtle and not visually evident. Quantification such as SUV plays a more important role in this scenario; clinical studies to date indicate that most tumors responding to therapy show a 20–40% decrease in SUV early in the treatment course [11–15]. Therefore, reliable measurements are essential when evalu-

ating early response to therapy, and it is critically important to understand the variables that can affect SUV. Many biologic factors affect the FDG distribution in the body or influence how representative SUV is of malignancy. Technologic factors will affect how close the image measurement is to the physical FDG distribution. We will address biologic and technologic factors separately.

## Biologic Factors Affecting SUV
### Body Size Measurement

The most common method of measuring the patient's body size for the calculation of average radioactivity concentration is to use the total body weight. Ideally, using this method means that a region with the same affinity for FDG between individuals will also have the same SUV measurement. For example, if the same amount of FDG is injected into a heavy patient and a light patient, the measured radioactivity concentration in the heavier patient's tumor would be expected to be less than in an identical tumor in the lighter patient. However, the average radioactivity concentration over the entire body of the heavier patient would be reduced by a similar factor, so the SUV would be approximately equal for the two patients' tumors even if the measured tumor uptake is not.

However, the FDG distribution *is* weight dependent [16]. Heavier patients often have a higher body fat percentage, and white body fat is less metabolically active (i.e., takes up less FDG) than muscle tissue [16]. A thin patient with relatively more muscle will likely have a lower SUV for a given lesion because muscle competes for the same FDG as the lesion [17]. Comparison of SUVs among patients with different body compositions is flawed. Even comparison of SUVs between examinations of the same patient can be flawed if the patient has lost or gained weight—often, body composition will change over time.

Patients' weight can change during the course of treatment. For example, the Women's Healthy Eating and Living study examined more than 3,000 women undergoing treatment for breast cancer. The study accepted self-reported precancer weight, weighed patients at baseline before therapy, and then repeated weight measurements for 1–6 years after therapy. Approximately 45% had significant weight gain (defined as > 5% positive change) [18]. At 4 years after treatment, fewer than 5% of patients with significant weight gain had returned to precancer weight. A similar study analyzed 15 men and six women for 6 months

during therapy for non–small cell lung cancer and found a median 6.5% and 11% body weight decrease in men and women, respectively [19]. The more extreme weight changes within these populations could affect the validity of SUV change measurements.

Two common correction factors used to reduce SUV dependence on weight are SUV calculated using lean body mass ($SUV_{lbm}$) and SUV calculated using body surface area ($SUV_{bsa}$). In these calculations, the body size factor in equation 1 is replaced by lean body mass or body surface area. Lean body mass can be estimated from a patient's height. For example, one lean body mass formula is: 48 + 1.06 (height [cm] – 152) [16]. Body surface area can be calculated from a combination of height and weight [20].

Both $SUV_{bsa}$ and $SUV_{lbm}$ are less sensitive to patients' weight than the $SUV_{weight}$ [16, 20]. $SUV_{lbm}$ is closer in mean value to $SUV_{weight}$ than $SUV_{bsa}$ [21]. However, $SUV_{lbm}$ results can be inconsistent because there is no standard formula for calculating lean body mass [22].

A more physiologically meaningful measure from FDG PET images is the glucose metabolic rate (GMR) of a tissue. GMR is a more complex calculation than the SUV, and there is mixed evidence as to which SUV calculation method correlates best with GMR. Two studies reported that the best correlation with GMR is $SUV_{bsa}$ [23, 24], but a separate study found that $SUV_{lbm}$ had the best correlation with GMR [25]. Each of these three studies used different formulas to calculate the various SUVs and the GMR. Thus, a direct comparison of these results cannot be made, and it is unclear whether $SUV_{bsa}$ or $SUV_{lbm}$ correlates best with GMR. No study has reported $SUV_{weight}$ as having the best correlation.

In some applications, the choice of $SUV_{weight}$, $SUV_{bsa}$, or $SUV_{lbm}$ may yield similar results. For example, one study [26] found no advantage for any method of calculating SUV when distinguishing benign from malignant lesions in patients with breast cancer.

*Bottom line*—In monitoring response to therapy in a patient with stable weight, all calculation methods will have similar percentage changes in value, and prediction of response to therapy will therefore not depend on the choice of body size measurement [27–29]. If $SUV_{weight}$ is used in monitoring response, then weight should be measured with the same scale at the PET facility—not on the basis of self-reported weight or from the patient chart.

*Blood Glucose Correction*

A patient's blood glucose level can affect SUV measurements because of the way FDG is processed by the cell. When a cell takes up FDG, an enzyme (hexokinase) phosphorylates it, creating FDG-6-phosphate that is trapped in the cell [30]. The FDG is in competition with glucose because hexokinase also phosphorylates glucose to form glucose-6-phosphate [31]. If a cell does not take up as much FDG as it would otherwise because of competitive inhibition by glucose, the SUV will be reduced.

A high patient serum glucose level before imaging can substantially decrease any SUV measurements [32]. Studies evaluating the usefulness of correcting for blood glucose have had conflicting results: some have found a benefit in normalizing SUV by blood glucose [25, 26, 33], some have found no benefit [29, 34, 35], and others have found that such glucose correction lowers the reproducibility of SUV measurements [28, 36]. It is likely that any attempt to measure blood glucose and correct for these differences introduces an error that is comparable to any existing error coming from such differences [37].

The National Cancer Institute's consensus recommendations on PET are to avoid correcting SUVs for blood glucose if the levels are within the reference range (< 200 mg/dL) [38]. In a PET protocol for multicenter trials in The Netherlands, Boellaard et al. [39] concur, recommending that a patient be rescheduled only if blood glucose is above twice the normal concentration.

*Bottom line*—It is good practice to measure the blood glucose level as a validity check to ensure that the patient's blood glucose levels are within the normal range. However, applying a correction factor is probably not warranted.

*Other Biologic Factors*

FDG will accumulate in tissue in proportion to the rate of glucose utilization [2]. The rate of FDG clearance will depend on the ratio of hexokinase to glucose-6-phosphatase in a cell. Normal and inflammatory tissue have relatively more glucose-6-phosphatase and exhibit faster clearance than most malignant tissue, which has less glucose-6-phosphatase and thus slower clearance [31]. Malignant tissues continue to accumulate more FDG over time compared with normal tissues. Thus, longer uptake times (time between injection and scanning) can lead to higher SUVs for malignant tissues compared with shorter uptake times [40]. This effect is reduced for uptake times exceeding 60 minutes.

Patient breathing can also affect measured SUVs [41], particularly in lesions in the lung bases or upper abdomen. This occurs because CT (used for attenuation correction during PET image reconstruction) can occur during a single breath-hold of the patient, but a PET acquisition for a given bed position takes minutes and is obtained while the patient is quietly breathing. If the diaphragm position in CT does not match the average position during PET, the attenuation correction may over- or undercorrect the radioactivity concentration, which would change the measured SUV. A more subtle effect is that the lung density will likely be different, on average (during PET), from during a particular respiratory phase (during breath-hold CT), and therefore the attenuation measured during CT produces an inaccurate correction [42], which can significantly affect the SUV calculation.

If PET images are acquired with the patient quietly breathing, the CT images are most accurately coregistered when single breath-hold CT is obtained at quiet end-expiration. When following lesions at the lung base or in the upper abdomen, it is important for the CT scans to be consistently obtained in this phase. It is essential for the technologists to explain to the patients the importance of the end-expiratory breath-hold because it is drastically different from routine dedicated CT. At Duke University Medical Center, it has been found that compliance greatly improves if the technologist goes through a rehearsal with the patient and confirms that the patient correctly follows the breathing instructions before the actual scanning. Aside from attenuation correction issues, any lesion that is moving will be measured inaccurately because of the effects of blurring. The implicit assumption is that the effects of blurring will be similar between the baseline and follow-up scans, but even this may not be true if breathing patterns are different during the two PET scans. Table 1 summarizes the biologic factors that affect SUV measurements in FDG PET.

*Bottom line*—Although many of these biologic factors cannot be controlled, it is possible to minimize the variability in SUV measurements by maintaining consistency between PET studies of the same patient. When performing follow-up PET studies, use the same uptake time as the baseline study. When following lesions in the lower thorax or upper abdomen, it is particularly important that the CT scans are consistently obtained at quiet end-expiration.

**Technologic Factors Affecting SUV**
*Interscanner Variability*

An important consideration in PET is whether images of the same patient have been acquired on different scanner models. Different manufacturers and scanner models have different physical properties as well as different acquisition and reconstruction options. Each scanner has a calibration factor to convert measured counts to radioactivity. The method and care of how this calibration is performed impact the underlying quantitative accuracy of the PET scanner. The American College of Radiology Imaging Network qualification of more than 100 PET scanners showed differences up to 6% among averages over all scanners of the same model [43]. In addition to the basic radioactivity calibration, other effects can play a substantial role.

One important difference between PET images from different scanners (and even from the same scanner used in different ways) is the spatial resolution. Image resolution affects the measured SUV of a small object because of the partial volume effect. There are two aspects of the partial volume effect: a voxel represents radioactivity from a volume larger than the voxel dimensions (and potentially multiple tissue types) and radioactivity from a very small region will be measured in a collection of neighboring voxels. Thus, a small source will show up in the final image as a larger, less intense source [38]. This causes underestimation of the original maximum or mean activity, especially for small sources. As shown in the simulated activity profiles in Figure 1, the larger spheres, blurred by the same amount, are affected less than the smaller spheres.

The spatial resolution and other factors vary from one PET scanner to the next, and different models from the same manufacturer may measure substantially different SUVs [44]. There also can be variation for a particular scanner, depending on the acquisition mode and image reconstruction and processing parameters, which will be illustrated further. A relatively recent development in commercial products is the implementation of a compensation for the inherent blurring in a PET scanner by modeling the point-spread function response [45]. Use of this feature, while improving the contrast of small lesions, would also give different SUV values than those obtained without this feature.

Despite standard acquisition and reconstruction parameters, biologic and other factors still play an important role. A recent

**TABLE 1: Biologic Factors Affecting Standardized Uptake Value (SUV) Measurements**

| Factor | Description | Approximate Magnitude of Effect |
|---|---|---|
| Weight composition | Body fat has less FDG uptake than muscle; heavier patients often have higher SUV values because of higher body fat percentage | SUVs for blood in heavier patients were up to twice the SUVs in lighter patients [16] |
| Body size calculation | Body weight may change during treatment; body surface area or lean body mass may be more consistent | Study of lung cancer patients reported median 6.5% and 11% body weight decrease during treatment in men and women, respectively [19]; certain body surface area or lean body mass corrections remove any weight dependence of SUV |
| Blood glucose level | High blood glucose level competitively inhibits FDG uptake | After glucose loading, some SUVs decreased more than 50% [32], but no corrections are recommended unless a patient has twice the normal blood glucose level [38] |
| Postinjection uptake time | FDG cleared more rapidly by normal tissue than malignant tissue; up to a point, a longer scan may result in higher SUV | SUVs for high-grade tumors up to 30% larger for 4-h vs 1-h postinjection uptake times [40] |
| Respiratory motion | Breathing during PET vs CT may lead to incorrect attenuation correction and bias in SUVs | SUV changes up to 30% have been reported [41] |
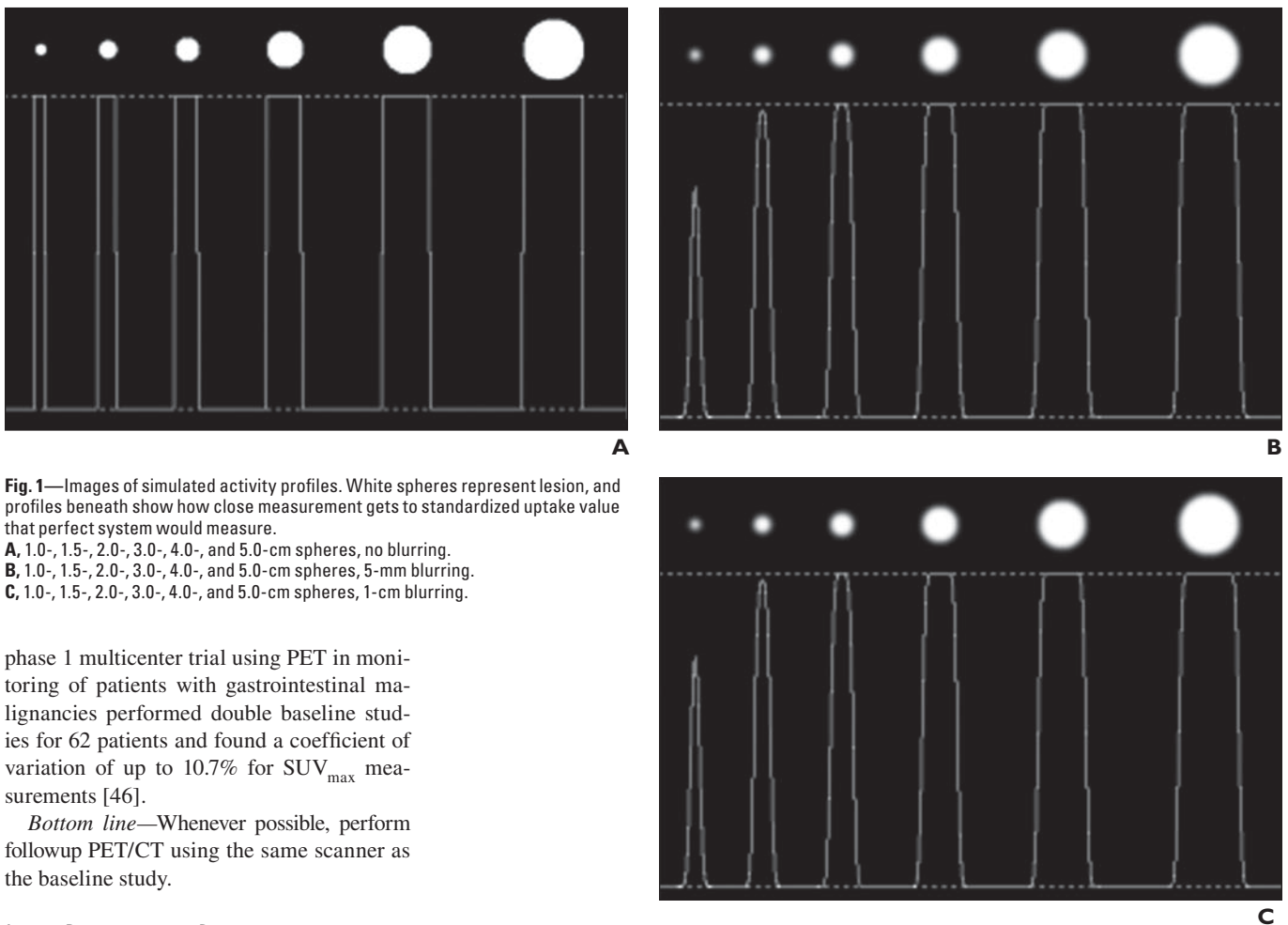
**Fig. 1**—Images of simulated activity profiles. White spheres represent lesion, and profiles beneath show how close measurement gets to standardized uptake value that perfect system would measure.
**A,** 1.0-, 1.5-, 2.0-, 3.0-, 4.0-, and 5.0-cm spheres, no blurring.
**B,** 1.0-, 1.5-, 2.0-, 3.0-, 4.0-, and 5.0-cm spheres, 5-mm blurring.
**C,** 1.0-, 1.5-, 2.0-, 3.0-, 4.0-, and 5.0-cm spheres, 1-cm blurring.

phase 1 multicenter trial using PET in monitoring of patients with gastrointestinal malignancies performed double baseline studies for 62 patients and found a coefficient of variation of up to 10.7% for $SUV_{max}$ measurements [46].

*Bottom line*—Whenever possible, perform followup PET/CT using the same scanner as the baseline study.

*Image Reconstruction Parameters*

A phantom study was performed to assess the impact of image reconstruction and processing parameters on SUV. A whole-body phantom was scanned with a Discovery 690 PET system (GE Healthcare), which has time-of-flight (TOF) capability. Fourteen 1.0-cm spheres and two 2.5-cm spheres were placed throughout the phantom. A solution of FDG and water was used to fill the spheres and phantom volume with a 6:1 sphere to background radioactivity concentration. Iterative reconstruction parameters were varied to evaluate their respective impact on measured SUV: image matrix size, postsmoothing, field of view (FOV) size, TOF versus non-TOF reconstruction, number of iterations, and image matrix placement.

All values are reported as the ratio of the measured value to the known value. For example, if an SUV of 6 was expected for a sphere and a value of 4 was measured, this is
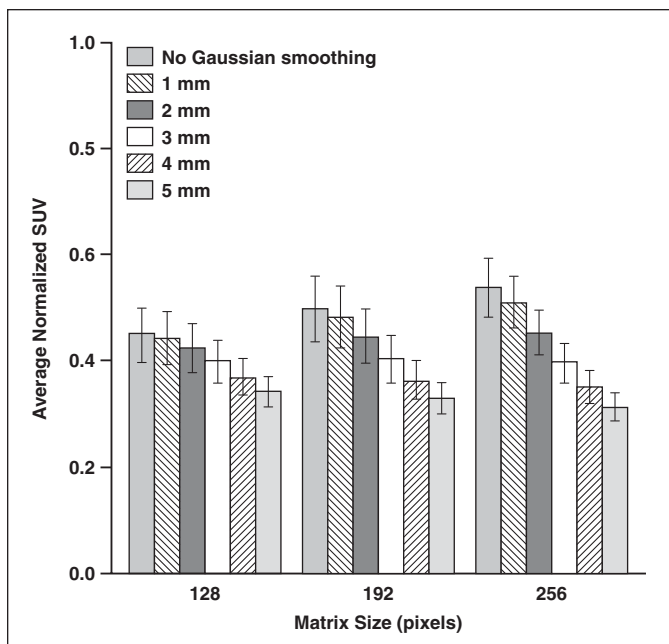
**Fig. 2**—Graph shows effect of image matrix size on standardized uptake value (SUV). Larger matrix sizes allow better-sampled images. Images with better sampling allow measurements that come closer to SUV that perfect system would measure.
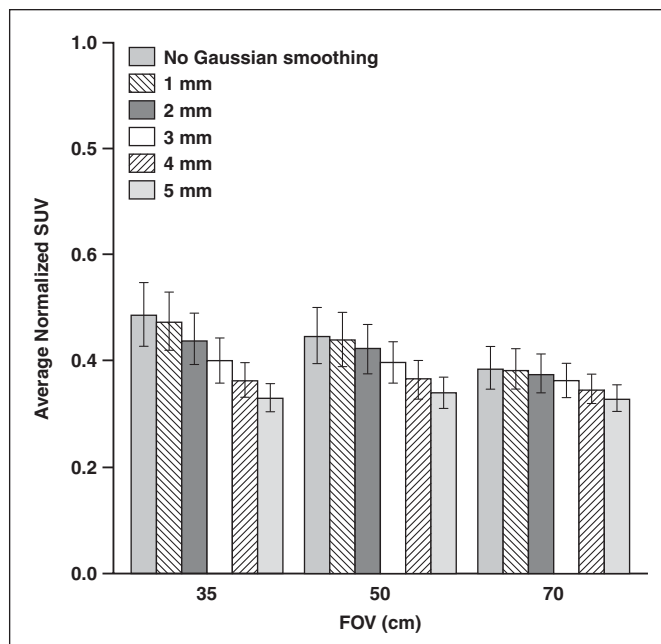


**Fig. 3**—Graph shows effect of field of view (FOV) size on standardized uptake value (SUV). Larger FOV sizes with constant image matrix size make larger voxels and thus lower-sampled images. Images with lower sampling may cause measurements that are further from SUV that perfect system would measure.

shown as 4 / 6 = 0.67. The primary point of these graphs is to illustrate the differences in values obtained, rather than the point that the measured values for small lesions are always lower than the actual values.

*1.0-cm spheres: matrix size and smoothing*—Three different image matrix sizes were tested for their impact on SUV measurements for 1.0-cm spheres: $128 \times 128$, $192 \times 192$, and $256 \times 256$ voxels. Figure 2 shows that using a larger matrix for a given FOV increased SUV measurements for 1.0-cm spheres. This is likely because larger matrix sizes for a constant FOV make each voxel smaller. Smaller voxels may yield higher spatial resolution but also increase the prob-

ability of sampling the peak of the lesion. A previous study by Westerterp et al. [47] had similar findings.

*1.0-cm spheres: FOV and smoothing*—Three different FOV sizes were tested: 35, 50, and 70 cm. The image matrix size was kept constant at $128 \times 128$ voxels. Figure 3 shows that larger FOVs and the same matrix size will cause lower SUV measurements for 1.0-cm spheres. Larger FOVs for the same matrix size make each voxel larger, decreasing sampling, which is similar to the preceding result. Measurements are more likely to underestimate the true SUV when the FOV is larger unless the larger FOV is accompanied by a larger image matrix.

*1.0-cm spheres: TOF versus non-TOF iterations*—TOF is an important technology in PET that accounts for photon arrival time differences to improve the statistical quality of PET data [48]. TOF reconstruction versus non-TOF reconstruction for the same raw data was evaluated.

$SUV_{max}$ measurements were made on the 14 1.0-cm spheres located in the phantom. Each sphere had the same expected SUV. However, there is some natural variation in measurements of SUV. If there is more noise in the image, then larger variation in SUV measurements of the identical spheres will occur. To assess variation in SUV measurements, an SD across all SUV measurements of the 14 spheres was calculated.
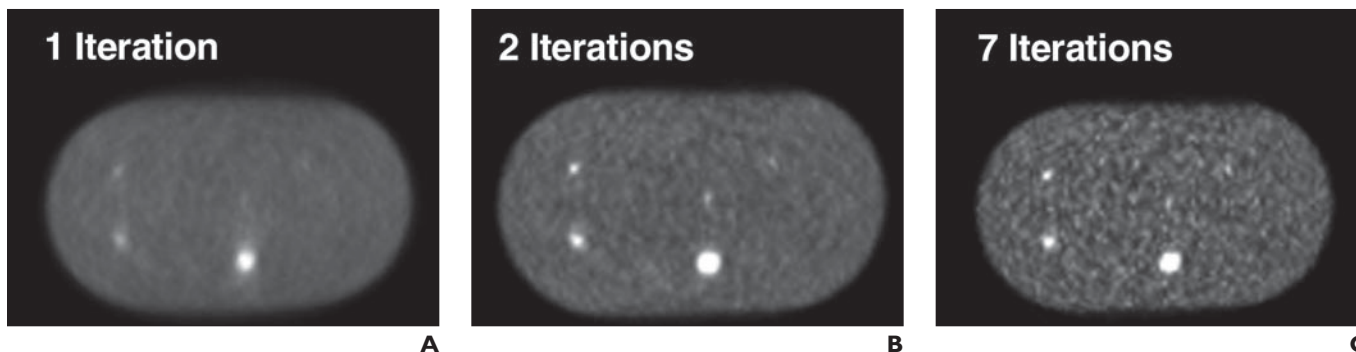


**Fig. 4**—Effect of reconstruction algorithm iterations on image quality.
**A–C,** Each image is from same slice of image stack, and each is reconstructed with different number of iterations of image reconstruction algorithm: one iteration (**A**), two iterations (**B**), and seven iterations (**C**).
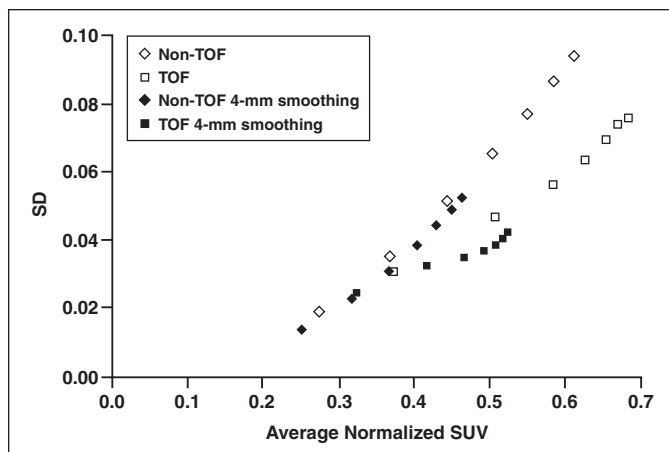
**Fig. 5**—Graph shows comparison between time-of-flight (TOF) and non-TOF reconstructed images of standardized uptake value (SUV) variability and measured SUV. Fourteen identical 1.0-cm spheres were measured. This graph shows SD of sphere measurements versus normalized SUV. Each point represents number of reconstruction iterations. Higher SDs indicate more image noise. Ideally, SUVs could be measured that were close to what perfect system would measure (normalized SUV of 1) with low image noise.
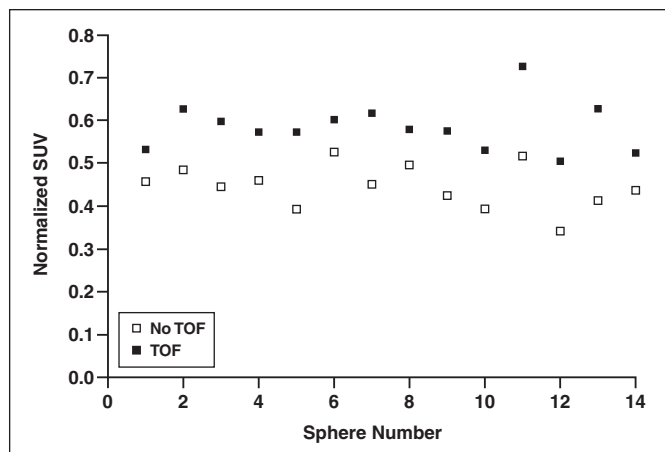
**Fig. 6**—Graph shows comparison between time-of-flight (TOF) and non-TOF reconstructed images of measured SUVs for identical spheres. This graph shows individual normalized SUV measured for each identical sphere in phantom when three iterations in iterative reconstruction were used. In this case, TOF-reconstructed images had universally higher normalized SUVs than non-TOF images.

Figure 4 shows the impact of iterations on image quality and the trade-off between noise and resolution. After one iteration of the reconstruction algorithm, SUV measurements of lesions in the image substantially underestimate the expected SUV because of low image resolution, but the image has low noise. After seven iterations, there will be better recovery of the expected SUV, but the image has much higher noise. This noise would be expected to yield larger variation in SUV measurements of the 14 identical spheres.

Variation of measured $SUV_{max}$ values is expected because of image noise and because the fourteen 1.0-cm spheres were in

slightly different axial positions and therefore placed differently relative to the image slices. Figure 5 shows SD versus the average normalized SUV across the 1.0-cm spheres as a function of iterations (from one to seven). The diamonds denote no postsmoothing, and squares denote 4 mm of postsmoothing. This figure illustrates that larger variation and larger SDs of SUV occur in noisier images. For a given number of iterations, the TOF compared with non-TOF reconstruction has higher $SUV_{max}$ values for comparatively little noise. Figure 6 compares normalized SUVs for each 1.0-cm sphere for three iterations of TOF versus non-TOF reconstruction.

$SUV_{max}$ measurements for TOF-reconstruction are higher and have less noise. As shown in previous studies [49], increasing iterations increases image noise. Figure 7 is a clinical example, showing the effect of TOF versus non-TOF reconstruction on SUV measurements of lesions.

*1.0-cm spheres: image matrix placement*—When reconstructing an image, one can center the reconstruction image matrix in a different place relative to the center of the FOV (i.e., shifting the image matrix). Each acquisition of the same patient will place any lesions in a different place relative to the image matrix, and this can affect SUV
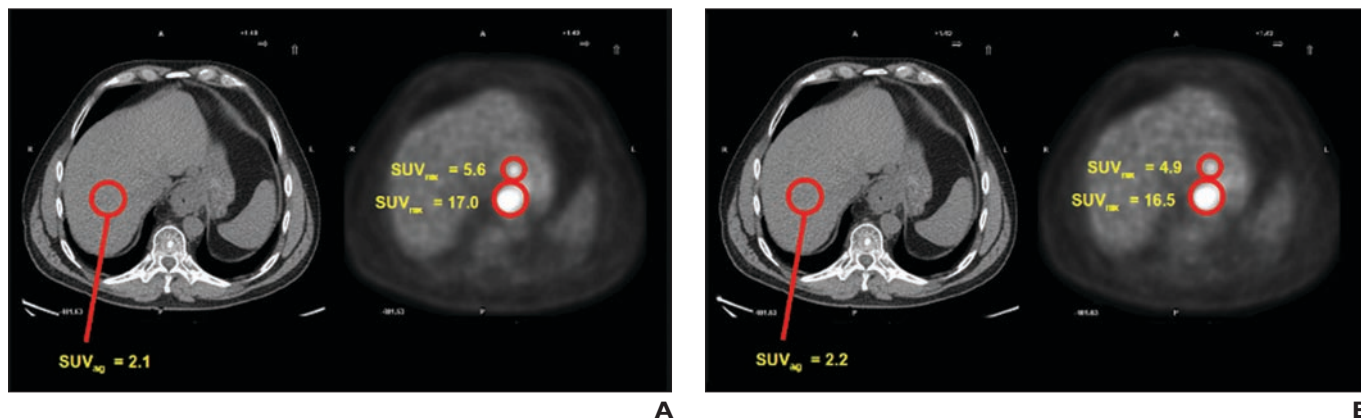


**Fig. 7**—Effect of time-of-flight (TOF) reconstruction on standardized uptake value (SUV) measurements in 58 year-old man with newly diagnosed esophageal cancer. PET images were processed with and without TOF. Mean SUV in liver is similar in two cases; however, $SUV_{max}$ values measured in primary esophageal mass and adjacent gastrohepatic lymph node are higher on TOF images. Acquisition protocol was 63-minute uptake, 150 s/bed position.
**A,** TOF reconstruction: 3D ordered-subset expectation maximization (OSEM), 16 subsets, and two iterations. Postsmoothing: 6.4-mm full width at half maximum (FWHM) Gaussian smoothing.
**B,** Non-TOF reconstruction: 3D OSEM, 24 subsets, and two iterations. Postsmoothing: 6.4-mm FWHM Gaussian smoothing.

**Fig. 8**—Image matrix shift.
**A** and **B,** Effect of image matrix shift: no image matrix shift (**A**) and shifted image matrix (**B**). Each square represents an image voxel, and the blurred sphere near the center of the image represents a lesion. If lesion is not centered in voxel, then lower standardized uptake values will be measured for it.
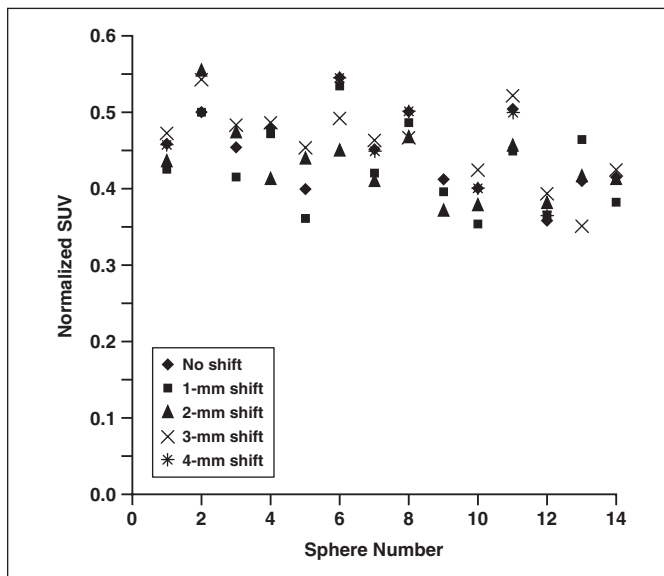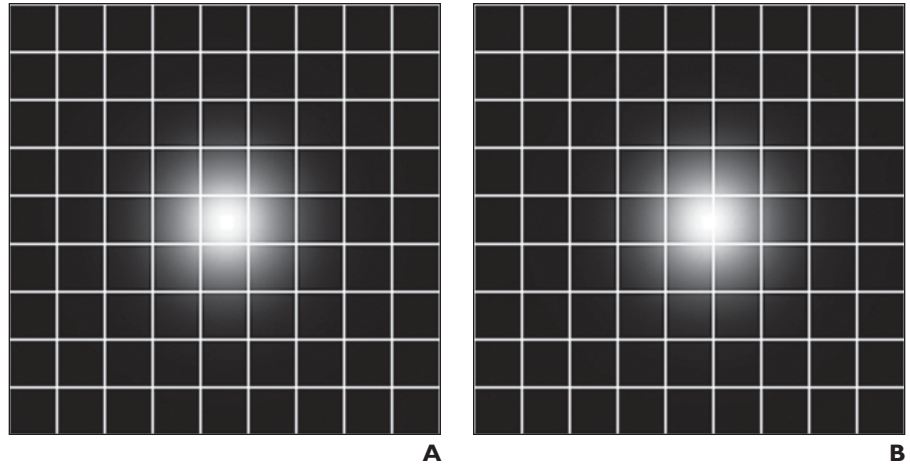


A



B



**Fig. 9**—Image matrix shift and standardized uptake value (SUV) with no smoothing. Graph shows measured SUV value for each image shift for each identical 1.0-cm sphere in phantom. No smoothing was done on images.
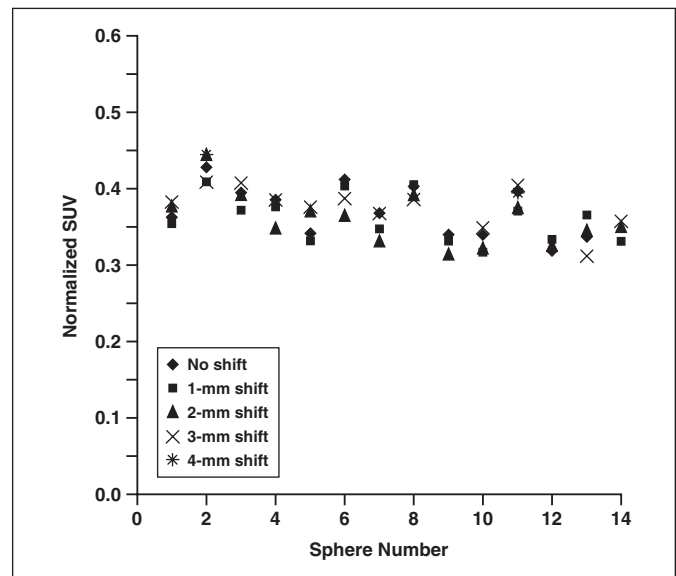


**Fig. 10**—Image matrix shift and standardized uptake value (SUV) with 4-mm smoothing. Graph shows measured SUV value for each image shift for each identical 1.0-cm sphere in phantom. For all images, 4 mm of smoothing was used.

measurements. Because only one acquisition was performed for this phantom study, the image matrix was shifted instead to illustrate the same effect.

Figure 8 shows the effect of an image matrix shift. In both images, the white grid represents the image matrix, with each square representing a voxel. The white gradient sphere represents a hot lesion in which the intensity corresponds to a number of counts. With no shift, the hot lesion happens to fall in the center of one of the voxels. Because the area of the lesion with the highest number of counts is centered within one voxel, the highest possible SUV can be recorded. However, if the image matrix is shifted by less than a voxel, a lesion that was centered in a voxel

will now be off center. Even if an ROI fully contains the lesion, no voxel in this ROI contains as many counts as when centered. Therefore, a lower $SUV_{max}$ is measured.

The position of the image matrix is arbitrary. However, as this illustration shows, if the matrix happens to be positioned such that a lesion's highest count density is within a voxel, then a higher SUV could be measured. In this phantom study, as in the body, some spheres (lesions) happened to be favorably aligned and some were not. Thus, even identical spheres naturally had some variability in recorded SUV, as identical lesions in the human body would.

Image shifts from 1 to 4 mm were tested, and a normalized SUV was calculated for each sphere with no postsmoothing (Fig.

9) and 4-mm postsmoothing (Fig. 10). When no smoothing was applied, shifting the image matrix center had a substantial effect on SUV.

When smoothing was applied, the variability between the maximum and minimum values for a sphere was reduced, depending on the amount of shift. This is to be expected because there is less possible variation in the number of counts from different areas of the lesion after smoothing. Thus, the value for each voxel will not be as sensitive to the voxel's placement over the lesion.

Changing FOV size for a constant image matrix causes the voxel dimensions to change and their alignment relative to the image features (e.g., lesions, organ boundaries) to change as well. When the voxels are larger,
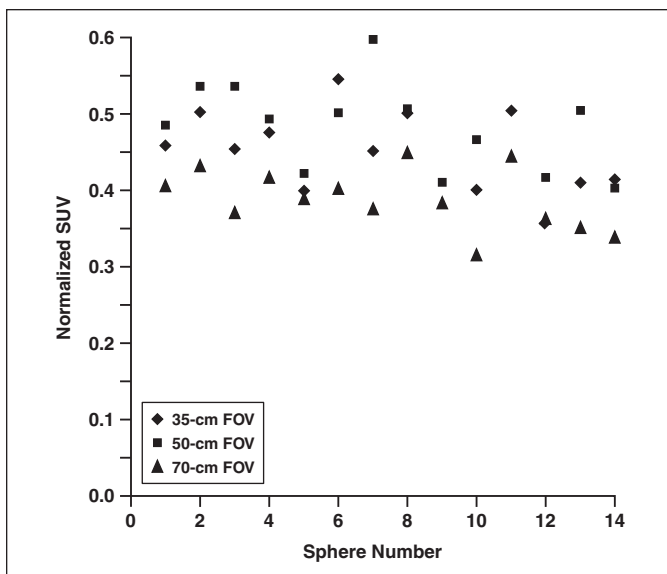
**Fig. 11**—Graph shows field of view (FOV) size and standardized uptake value (SUV) measurements for identical spheres. Larger FOV sizes with same image matrix size have lower sampling of raw data. In this case, 70-cm FOV resulted in SUV values for identical 1.0-cm spheres in phantom that are lower than measurements using 35- or 50-cm FOV.

follow-up studies. Smaller targets are more subject to variability in measurements. Smoothing or local averaging techniques can reduce the variability of measurements at the expense of underestimating the true maximum value.

*Injected Radioactivity*

SUV calculations are based on decay-corrected radioactivity. Therefore, if the dose calibrator and PET scanner clocks are not synchronized, the calculated decay time will be incorrect, introducing an error in the SUV calculation. Assuming there are no other errors in SUV and that the timing mismatch is relatively small compared with the half-life of $^{18}$F, there is an approximately linear relationship between timing mismatch and error in SUV. Table 3 presents estimates of the error in SUV resulting from a timing mismatch. A large timing mismatch between the scanner and dose calibrator clocks could result in considerable errors in SUV measurements.

If there is an error in calibration of the measured count rate to the true radioactivity concentration between the PET scanner and the dose calibrator, the SUV will be affected because that includes both the injected activity and the measured radioactivity concentration. Errors up to 10% have been found [50]. Furthermore, if there is activity still in the syringe after administration of the radiotracer, this may also cause error in SUV.

*Bottom line*—The radioactivity in the injection syringe should be accurately measured.

SUVs are more sensitive to placement effects. Figure 11 shows this variability in SUV measurement for each sphere as FOV changes, with no postsmoothing. Figure 12 is a clinical example showing the effect on SUV of changing the FOV size from 50 to 70 cm. This substantial change in SUV is likely due to placement effects as well.

*2.5-cm spheres*—All of the reconstruction parameters tested had impact on SUV measurements for small 1.0-cm spheres. These effects are greatly diminished for larger 2.5-cm spheres. Table 2 summarizes this change.

The results from Table 2 indicate that in a clinical setting, reconstruction parameter changes would have greater impact on SUV measurements for smaller lesions than larger lesions. However, small 1.0-cm lesions are clinically meaningful—early detection of cancer necessitates evaluation of small lesions. At a later time, a tumor may be larger and diffuse but contain necrotic regions. Thus, small volumes can be highly metabolically active and clinically relevant.

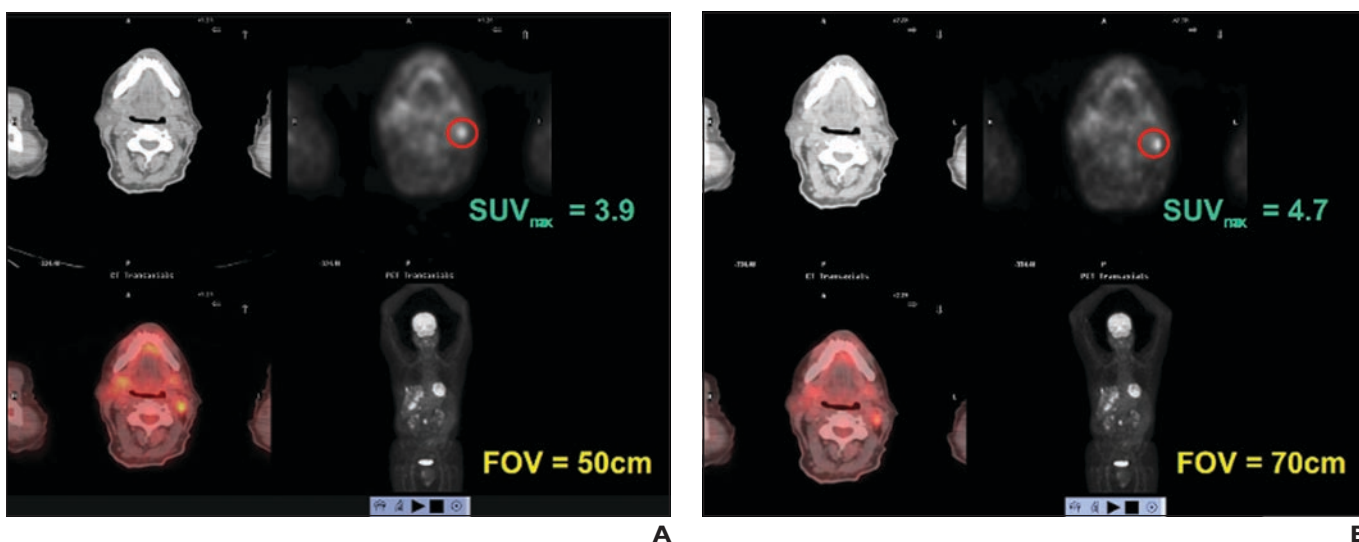*Bottom line*—The same FOV and reconstruction parameters should be used for baseline and



**Fig. 12**—Effect of field of view (FOV) change on standardized uptake value (SUV) measurement of lesion in 61-year-old man with metastatic melanoma.
**A** and **B**, Initial images reconstructed with 50-cm FOV (**A**) were obtained. Images were also reconstructed with 70-cm FOV (**B**) to include area of primary disease in left forearm. Different FOVs with same image matrix size will change placement of image matrix voxels relative to lesions in image. This can substantially affect SUV measurements of lesions. Acquisition protocol was 66-minute uptake and 150 s/bed position. Parameters for 50-cm FOV reconstruction (**A**): 2D ordered-subset expectation maximization (OSEM) and 20 subsets with two iterations. Postsmoothing: 8-mm full width at half maximum (FWHM) Gaussian smoothing. Parameters for 70-cm FOV reconstruction (**B**): 2D OSEM and 20 subsets with two iterations. Postsmoothing: 8-mm FWHM Gaussian smoothing.

**TABLE 2: Largest Percentage Differences in Standardized Uptake Value (SUV) Due to Reconstruction Parameters**

| Reconstruction Parameter (4-mm smoothing) | Specific Comparison | Sphere Diameter (cm) | Largest % Difference in Average Normal SUV |
|---|---|---|---|
| Field of view | 35 cm | 1.0 | 6.0 |
|  | 50 cm | 2.5 | 2.6 |
| Matrix size | 128 voxels | 1.0 | 5.3 |
|  | 256 voxels | 2.5 | 2.5 |
| Shift of matrix center | 2 vs 1 mm | 1.0 | 11.0 |
|  | 3 vs 1 mm | 2.5 | 3.1 |

**TABLE 3: Percentage Error in Standardized Uptake Value (SUV) From Timing Mismatch**

| Timing Mismatch (minutes) | SUV Error (%) |
|---|---|
| 2 | 1.3 |
| 4 | 2.5 |
| 8 | 5.1 |
| 10 | 6.3 |
| 15 | 9.5 |

**TABLE 4: Technologic Factors Affecting Standardized Uptake Value (SUV) Measurement**

| Factor | Description | Approximate Magnitude of Effect |
|---|---|---|
| Scanner variability | Different detector crystal dimensions, random correction options, TOF capability, other physical factors may change SUV | SUV differences up to 22.6% found even with standard imaging protocol between scanners [50] |
| Reconstruction parameter changes | Reconstruction parameters such as matrix size, FOV, and TOF reconstruction could substantially change SUV for small lesions | SUV differences up to 12% for 1.0-cm spheres in phantom study |
| Calibration error between scanner and dose calibrator | Improper calibration between scanner and dose calibrator can result in errors in SUV | SUV changes up to 10% [50] |
| Timing mismatch between scanner and dose calibrator | Error in calculation of decay time between injection and scanning time will affect calculated SUV; for timing errors small compared with the half-life of FDG, the error in SUV is linearly dependent on the timing error | SUV error up to 5% for 8-minute timing error |
| Use of contrast material for PET/CT | Incorrect attenuation correction will cause differences in SUV between contrast-enhanced studies and unenhanced studies | SUV differences up to 5.9% found [42] |
| Interobserver variability | Different observers may draw different ROIs, which will select different voxels for measurement | Variation up to 17% for some $SUV_{mean}$ measurements [9] |

Note—TOF = time-of-flight, FOV = field of view, ROIs = regions of interest.

After injection into the patient, residual activity within the syringe, needle, and any tubing should be measured and subtracted from the original syringe activity to determine the net activity administered to the patient. If there are multiple dose calibrators at the PET facility, they should be calibrated accurately.

*Other Factors*

Another factor that can impact SUV measurements is use of CT contrast material in a PET/CT study. The contrast-enhanced CT may be used to perform attenuation correction on the PET images, but use of contrast material can affect this correction. SUVs from contrast-enhanced PET/CT studies can differ from unenhanced PET/CT studies by up to 5.9% [42], even on a system that compensates for the presence of contrast material.

Variability between observers in ROI placement and size can be substantial, especially for $SUV_{mean}$ measurements [9]. Interobserver variability in determining change in $SUV_{max}$ of 16.7% ± 36.2% has been found [51] due to different ROI placement with-

in the pre- and posttherapy images. Use of screen-saves or other documentation may allow more reproducibility in defining ROIs between pre- and posttherapy images.

Finally, it is important to compare only SUVs from the same type of acquisitions if monitoring a patient for response to therapy. Figure 13 shows images from a neck protocol and a skull base to midthigh protocol of the same supraclavicular node on one patient. Substantially different SUVs are measured.

Table 4 shows a list of technologic factors that affect SUV. As a quality control measure, it may be useful to record a patient's liver SUV for each PET study. The liver SUV does not vary greatly between different scans of the same patient [52]. If a liver SUV was substantially different from previous values, it might indicate a flaw in scanner calibration or function.

**Conclusions**

The SUV formula has three variables: the measurement of the radioactivity concentration in tissue, the injected radioactivity, and the body size. Most factors that change SUV

measurements affect the measurement of the radioactivity concentration. Biologic factors, such as patient blood glucose level, uptake time of the tracer, and respiratory motion, can make a substantial impact on SUV measurements. Technologic factors, such as interscanner variability, image acquisition and reconstruction parameters, and interobserver variability, can make a difference as well. Thus, it is important to keep as many of these factors as possible the same between baseline and follow-up studies of a patient.

SUV measurements also depend on the accurate measurement of injected radioactivity and choice of body size measurement. Calibration of the clocks on a PET scanner and dose calibrator is important for reducing error in SUV based on the injected radioactivity. Different body size measurements can significantly change SUV measurements; however, this can be controlled by using the same body size measurement between baseline and follow-up studies.

Although the current considerations are based on PET with FDG, these general con-
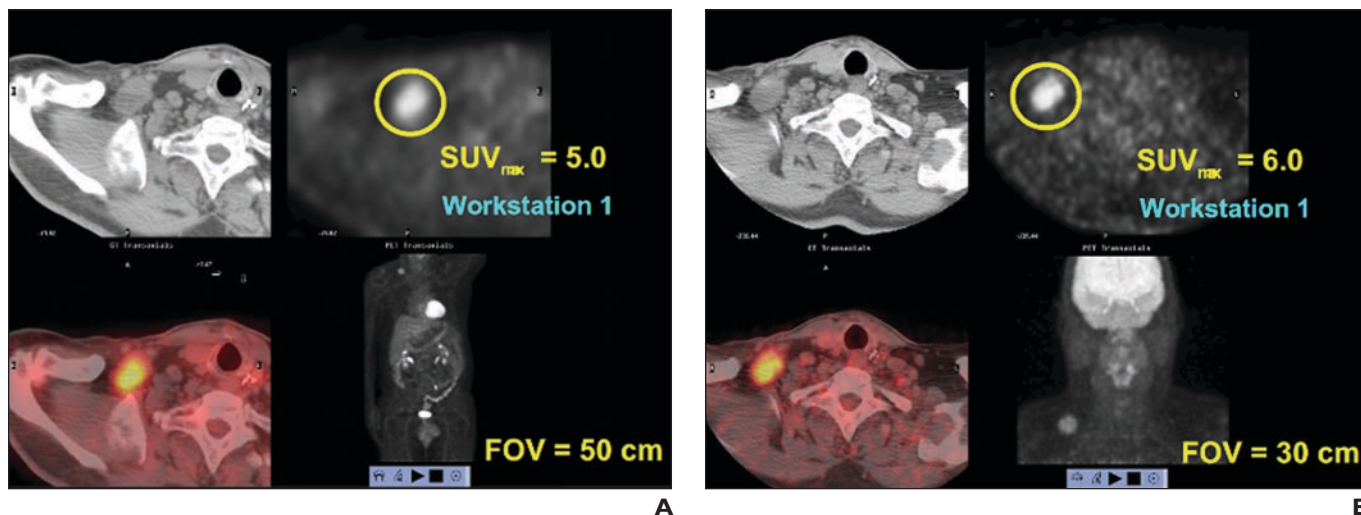
**Fig. 13**—Different acquisitions of same lesion in 67-year-old man with metastatic medullary thyroid cancer.
**A** and **B**, Enlarged right supraclavicular node is identified on both standard body images (**A**) and dedicated neck images (**B**). Skull-base to midthigh acquisition will have different field of view than acquisition from neck up. Both acquisitions are of same supraclavicular node; however, different SUVs are measured. Protocol for skull base to midthigh (**A**): 73-minute uptake, 150 s/bed position, 2D ordered-subset expectation maximization (OSEM), 20 subsets with two iterations. Postsmoothing: 8-mm full width at half maximum (FWHM) Gaussian smoothing. Protocol for neck (**B**): 97-minute uptake, 360 s/bed position, 2D OSEM, and 28 subsets with three iterations. Postsmoothing: 5-mm FWHM Gaussian smoothing.

cepts will also have applicability in the quantification of other PET tracers. The following are specific recommendations for using SUV for determining early response to therapy:

1. Whenever possible, obtain baseline and follow-up PET/CT on the same scanner.
2. Measure residual activity in the syringe and injection tubing to accurately determine the administered dose.
3. A minimum uptake time of 60 minutes is recommended for oncology patients. For follow-up PET, use the same uptake time that was used for the baseline examination (± 10 minutes).
4. Maintain the same acquisition technique and reconstruction parameters (2D vs 3D, FOV and image matrix size, ± TOF, number of iterations, subsets, smoothing) for baseline and subsequent examinations. Consider using the same CT protocol (± IV contrast administration) for attenuation correction of the PET images.
5. For quality assurance of measurements, obtain serum glucose before each PET, and record average SUV in the liver as an additional quality assurance mechanism.
6. Weigh every patient before imaging at the PET facility using a calibrated scale.
7. Maintain calibration of dose calibrators, and synchronize dose calibrator clocks with the scanner clocks.
8. Use screen-saves or other documentation to improve reproducibility in defining ROIs between baseline and follow-up studies.

SUV measurements are currently the most convenient method for quantitatively evaluating changes in metabolic activity. However, it is important to understand the limitations of these measurements and to minimize the effects from variables that can be controlled. This will become increasingly important as PET is used to determine response to therapy and is incorporated in therapeutic decision-making.

## References

1. Ben-Haim S, Ell P. ^{18}F-FDG PET and PET/CT in the evaluation of cancer treatment response. *J Nucl Med* 2009; 50:88–99
2. Kelloff GJ, Hoffman JM, Johnson B, et al. Progress and promise of FDG-PET imaging for cancer patient management and oncologic drug development. *Clin Cancer Res* 2005; 11:2785–2808
3. Weber WA, Wieder H. Monitoring chemotherapy and radiotherapy of solid tumors. *Eur J Nucl Med Mol Imaging* 2006; 33:27–37
4. Jadvar H, Alavi A, Gambhir SS. ^{18}F-FDG uptake in lung, breast, and colon cancers: molecular biology correlates and disease characterization. *J Nucl Med* 2009; 50:1820–1827
5. Eary JF, O'Sullivan F, Powitan Y, et al. Sarcoma tumor FDG uptake measured by PET and patient outcome: a retrospective analysis. *Eur J Nucl Med Mol Imaging* 2002; 29:1149–1154
6. Shreve PD, Anzai Y, Wahl RL. Pitfalls in oncologic diagnosis with FDG PET imaging: physiologic and benign variants. *RadioGraphics* 1999; 19:61–77
7. Hellwig D, Graeter TP, Ukena D, et al. ^{18}F-FDG PET for mediastinal staging of lung cancer: which SUV threshold makes sense? *J Nucl Med* 2007; 48:1761–1766
8. Lucignani G. SUV and segmentation: pressing challenges in tumour assessment and treatment. *Eur J Nucl Med Mol Imaging* 2009; 36:715–720
9. Benz MR, Evilevitch V, Allen-Auerbach MS, et al. Treatment monitoring by ^{18}F-FDG PET/CT in patients with sarcomas: interobserver variability of quantitative parameters in treatment-induced changes in histopathologically responding and nonresponding tumors. *J Nucl Med* 2008; 49: 1038–1046
10. 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
11. Holdsworth CH, Badawi RD, Manola JB, et al. CT and PET: early prognostic indicators of response to imatinib mesylate in patients with gastrointestinal stromal tumor. *AJR* 2007; 189:1450; [web]W324–W330
12. Francis RJ, Byrne MJ, van der Schaaf AA, et al. Early prediction of response to chemotherapy and survival in malignant pleural mesothelioma using a novel semiautomated 3-dimensional volume-based analysis of serial ^{18}F-FDG PET scans. *J Nucl Med* 2007; 48:1449–1458
13. Weber WA. PET for response assessment in oncology: radiotherapy and chemotherapy. *Br J Radiol Suppl* 2005; 28:42–49
14. Lordick F, Ott K, Krause B, et al. PET to assess early metabolic response and to guide treatment of adenocarcinoma of the oesophagogastric junction: the MUNICON phase II trial. *Lancet Oncol* 2007; 8:797–805
15. Strobel K, Skalsky J, Steinert HC, et al. S-100B and FDG-PET/CT in therapy response assess-

ment of melanoma patients. *Dermatology* 2007; 215:192–201

16. Zasadny KR, Wahl RL. Standardized uptake values of normal tissues at PET with 2-[fluorine-18]-fluoro-2-deoxy-D-glucose: variations with body weight and a method for correction. *Radiology* 1993; 189:847–850

17. Keyes JW. SUV: standard uptake or silly useless value? *J Nucl Med* 1995; 36:1836–1839

18. Saquib N, Flatt SW, Natarajan L, et al. Weight gain and recovery of pre-cancer weight after breast cancer treatments: evidence from the Women's Healthy Eating and Living (WHEL) study. *Breast Cancer Res Treat* 2007; 105:177–186

19. Harvie MN, Campbell IT, Thatcher N, Baildam A. Changes in body composition in men and women with advanced nonsmall cell lung cancer (NSCLC) undergoing chemotherapy. *J Hum Nutr Diet* 2003; 16:323–326

20. Kim CK, Gupta NC, Chandramouli B, Alavi A. Standardized uptake values of FDG: body surface area correction is preferable to body weight correction. *J Nucl Med* 1994; 35:164–167

21. Sugawara Y, Zasadny KR, Neuhoff AW, Wahl RL. Reevaluation of the standardized uptake value for FDG: variations with body weight and methods for correction1. *Radiology* 1999; 213: 521–525

22. Erselcan T, Turgut B, Dogan D, Ozdemir S. Lean body mass-based standardized uptake value, derived from a predictive equation, might be misleading in PET studies. *Eur J Nucl Med Mol Imaging* 2002; 29:1630–1638

23. Graham MM, Peterson LM, Hayward RM. Comparison of simplified quantitative analyses of FDG uptake. *Nucl Med Biol* 2000; 27:647–655

24. Izquierdo-Garcia D, Davies JR, Graves MJ, et al. Comparison of methods for magnetic resonance-guided [18-F]fluorodeoxyglucose positron emission tomography in human carotid arteries: reproducibility, partial volume correction, and correlation between methods. *Stroke* 2009; 40:86–93

25. Krak NC, van der Hoeven JJM, Hoekstra OS, Twisk JWR, van der Wall E, Lammertsma AA. Measuring [(18)F]FDG uptake in breast cancer during chemotherapy: comparison of analytical methods. *Eur J Nucl Med Mol Imaging* 2003; 30:674–681

26. Avril N, Bense S, Ziegler SI, et al. Breast imaging with fluorine-18-FDG PET: quantitative image analysis. *J Nucl Med* 1997; 38:1186–1191

27. 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 1]:122S–150S

28. Paquet N, Albert A, Foidart J, Hustinx R. Within-patient variability of ^{18}F-FDG: standardized uptake values in normal tissues. *J Nucl Med* 2004; 45:784–788

29. Stahl A, Ott K, Schwaiger M, Weber WA. Comparison of different SUV-based methods for monitoring cytotoxic therapy with FDG PET. *Eur J Nucl Med Mol Imaging* 2004; 31:1471–1478

30. Castell F, Cook GJR. Quantitative techniques in 18FDG PET scanning in oncology. *Br J Cancer* 2008; 98:1597–1601

31. Acton P, Zhuang H, Alavi A. Quantification in PET. *Radiol Clin N Am* 2004; 42:1055–1062

32. Lindholm P, Minn H, Leskinen-Kallio S, Bergman J, Ruotsalainen U, Joensuu H. Influence of the blood glucose concentration on FDG uptake in cancer: a PET study. *J Nucl Med* 1993; 34:1–6

33. Wong CY, Thie J, Parling-Lynch KJ, et al. Glucose-normalized standardized uptake value from ^{18}F-FDG PET in classifying lymphomas. *J Nucl Med* 2005; 46:1659–1663

34. Vriens D, de Geus-Oei L, van Laarhoven HW, et al. Evaluation of different normalization procedures for the calculation of the standardized uptake value in therapy response monitoring studies. *Nucl Med Commun* 2009; 30:550–557

35. Sanghera B, Emmott J, Wellsted D, Chambers J, Wong W. Influence of N-butylscopolamine on SUV in FDG PET of the bowel. *Ann Nucl Med* 2009; 23:471–478

36. Hadi M, Bacharach SL, Whatley M, et al. Glucose and insulin variations in patients during the time course of a FDG-PET study and implications for the "glucose-corrected" SUV. *Nucl Med Biol* 2008; 35:441–445

37. 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

38. 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

39. Boellaard R, Oyen WJG, Hoekstra CJ, et al. The Netherlands protocol for standardisation and quantification of FDG whole body PET studies in multi-centre trials. *Eur J Nucl Med Mol Imaging* 2008; 35:2320–2333

40. Lowe VJ, DeLong DM, Hoffman JM, Coleman RE. Optimum scanning protocol for FDG-PET evaluation of pulmonary malignancy. *J Nucl Med* 1995; 36:883–887

41. Erdi YE, Nehmeh SA, Pan T, et al. The CT motion quantitation of lung lesions and its impact on PET-measured SUVs. *J Nucl Med* 2004; 45:1287–1292

42. Bunyaviroch T, Turkington T, Wong T, Wilson J, Colsher J, Coleman R. Quantitative effects of contrast enhanced CT attenuation correction on PET SUV measurements. *Mol Imaging Biol* 2008; 10: 107–113

43. Scheuermann JS, Saffer JR, Karp JS, Levering AM, Siegel BA. Qualification of PET scanners for use in multicenter cancer clinical trials: The American College of Radiology Imaging Network experience. *J Nucl Med* 2009; 50:1187–1193

44. 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

45. Tong S, Alessio AM, Kinahan PE. Noise and signal properties in PSF-based fully 3D PET image reconstruction: an experimental evaluation. *Phys Med Biol* 2010; 55:1453–1473

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

47. 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

48. Kadrmas DJ, Casey ME, Conti M, Jakoby BW, Lois C, Townsend DW. Impact of time-of-flight on PET tumor detection. *J Nucl Med* 2009; 50: 1315–1323

49. Jaskowiak CJ, Bianco JA, Perlman SB, Fine JP. Influence of reconstruction iterations on ^{18}F-FDG PET/CT standardized uptake values. *J Nucl Med* 2005; 46:424–428

50. Geworski L, Knoop BO, de Wit M, Ivancevic V, Bares R, Munz DL. Multicenter comparison of calibration and cross calibration of PET scanners. *J Nucl Med* 2002; 43:635–639

51. Jacene HA, Leboulleux S, Baba S, et al. Assessment of interobserver reproducibility in quantitative ^{18}F-FDG PET and CT measurements of tumor response to therapy. *J Nucl Med* 2009; 50:1760–1769

52. Chin BB, Green ED, Turkington TG, Hawk TC, Coleman RE. Increasing uptake time in FDG-PET: standardized uptake values in normal tissues at 1 versus 3 h. *Mol Imaging Biol* 2009; 11:118–122