Published in final edited form as:

Semin Ultrasound CT MR. 2010 December ; 31(6): 496–505. doi:10.1053/j.sult.2010.10.001.

PET/CT Standardized Uptake Values (SUVs) in Clinical Practice and Assessing Response to Therapy

Paul E. Kinahan, PhD^{*} and

Department of Radiology, University of Washington, Seattle, WA

James W. Fletcher, MD

Department of Radiology, Indiana University School of Medicine, Indianapolis, IN

Abstract

The use of standardized uptake values (SUVs) is now common place in clinical FDG-PET/CT oncology imaging, and has a specific role in assessing patient response to cancer therapy. Ideally, the use of SUVs removes variability introduced by differences in patient size and the amount of injected FDG. However, in practice there are several sources of bias and variance that are introduced in the measurement of FDG uptake in tumors and also in the conversion of the image count data to SUVs. The overall imaging process is reviewed and estimates of the magnitude of errors, where known, are given. Recommendations are provided for best practices in improving SUV accuracy.

1. The use of Standardized Uptake Values in FDG-PET Imaging

PET/CT imaging of cancer with combined positron emission tomography (PET) and x-ray computerized tomography (CT) scanners has become a standard component of diagnosis and staging in oncology^{1,2}. The use of the radiolabeled tracer 2-deoxy-2-[¹⁸F]fluoro-D-glucose (FDG) for oncology imaging accounts for the majority of all PET/CT imaging procedures since increased accumulation of FDG relative to normal tissue is a useful marker for many cancers³.

In addition to cancer detection and staging, PET/CT imaging is becoming more important as a quantitative monitor of individual response to therapy and an evaluation tool for new drug therapies. Changes in FDG accumulation have been shown to be useful as an imaging biomarker for assessing response to therapy⁴. However, FDG uptake in tumors is related in a complex manner to the proliferative activity of malignant tissue and to the number of viable tumor cells⁵.

There are several methods for measuring the rate and/or total amount of FDG accumulation in tumors. PET scanners are designed to measure the in vivo radioactivity concentration [kBq/ml], which is directly linked to the FDG concentration. Typically, however, it is the relative tissue uptake of FDG that is of interest. The two most significant sources of variation that occur in practice are the amount of injected FDG and the patient size. To

^{*}Corresponding author. Department of Radiology, University of Washington, 222 Fisheries Center (FIS), Box 357987, Seattle, WA 98195-7987, Phone: 206-543-0236, Fax: 206-543-8356, kinahan@uw.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

compensate for these variations, at least to first order, the standardized uptake value (SUV) is commonly used as a relative measure of FDG uptake⁶. The basic expression for SUV is

$$\text{SUV} = \frac{r}{(a'/w)} \quad [1]$$

where r is the radioactivity activity concentration [kBq/ml] measured by the PET scanner within a region of interest (ROI), a' is the decay-corrected amount of injected radiolabeled FDG [kBq], and w is the weight of the patient [g], which is used a surrogate for a distribution volume of tracer. If all the injected FDG is retained and uniformly distributed throughout the body, the SUV everywhere will be 1 g/ml regardless of the amount of FDG injected or patient size. SUVs are dimensionless under the assumption that 1 ml of tissue weights 1 gm. Both approaches are used in practice. The use of lean body mass for w has also been suggested to account for the lower uptake of FDG by adipose tissue⁷.

The use of SUVs as a measurement of relative tissue/organ uptake facilitates comparisons between patients, and has been suggested as a basis for diagnosis. However, the practice of using SUV thresholds for diagnosis is not widely accepted⁸. There are at least two general reasons for the inconsistent use of SUVs in practice. First is that accurate staging and diagnostic information do not have to depend upon accurate image quantification⁹, since the *relative* image content (i.e. image appearance) is often sufficient for such purposes. Second, measured SUVs have a large degree of variability due to physical and biological sources of error, as well as inconsistent and non-optimized image acquisition, processing and analysis¹⁰. More specifically, it has been repeatedly demonstrated that the use of SUV thresholds (e.g. SUV > 2.5), wherein a nodule or mass is characterized as benign or malignant using thresholds, is often invalid. As such, many benign infectious/inflammatory processes will have substantial FDG uptake with a high SUV value, and conversely, many indolent or slowly growing malignant processes may have minimal uptake, and low SUV values. This is not to say, however, that using SUV thresholds for diagnosis is not of any value. In circumstances where a nodule or tissue mass has uptake no greater than adjacent reference tissue and the pre-test likelihood of malignancy is low, the decision to develop a “watch and wait” strategy for management can often be safely adopted. In this situation, the very low false negative rate of negligible FDG uptake can assist with the decision to avoid unnecessary invasive procedures for tissue diagnosis. This has often been referred to as using FDG-PET as a “molecular imaging probe”¹¹. As such, FDG PET/CT can assist in the decision to avoid unnecessary invasive tissue biopsy as well as guide such a procedure to a tissue location where a valid diagnostic biopsy sample can be obtained.

2. The Role of SUV in Quantitative Imaging with PET/CT

There are three levels of relevance for the use of SUVs in PET/CT imaging as illustrated in Fig 1. The first, which is perhaps the most direct argument for accurate SUVs, is the use of PET in clinical research, clinical trials, and drug discovery. The importance of quantitative imaging depends on the objective of each study, but as a group these studies will directly benefit from PET measures that have well characterized variance and precision¹².

The second level, which is becoming more important, is the use of PET/CT in assessing response to therapy. This is often most clearly related to clinical trials, but considered from the viewpoint of individual patients. In some cases, such as Hodgkins lymphoma, quantitative PET/CT imaging may not actually be needed, as success can be defined by the complete absence of tracer uptake in the PET image following a course of standardized therapy¹³.

The utilization of PET/CT to assess response to therapy is increasing in the US related, in part, to the creation and subsequent favorable results of the National Oncologic PET Registry (NOPR)¹⁴ and has the potential to become much more widespread in routine clinical use⁴ as well as in clinical trials. Employment of RECIST criteria wherein response of tumors to therapy has been traditionally assessed by measurement of changes in size/dimension of the tumor(s) in CT images can have serious limitations^{15,16}. Changes in size as a result of therapy may take many months to develop and any opportunity to make early decisions about therapy success or failure is often unduly delayed or lost altogether. Additionally, many new cytostatic agents may achieve therapeutic success without any manifestation of change in size⁴. Independent measures of changes in metabolic activity via FDG PET/CT can provide an alternate approach to assess response to therapy -- often very early in the course of treatment. Thus SUVs are often reported for therapy monitoring scans, and this a recommended procedure.

The largest use is in current clinical practice. At this level, as discussed above, the primary requirement is the fidelity of the relative image appearance, independent of global scaling biases, and SUVs should not be relied upon exclusively for diagnostic scans. However, a change that has occurred in practice over the last few years is that SUVs are now routinely reported for known or suspected cancers as part of the image interpretation and reporting. Current recommendations are that tumor SUVs should be reported^{17-19,20}, thus SUV sources of bias and variance should be well understood and controlled in this type of implementation.

3. Determinants of PET/CT Quantitation Accuracy and Precision

As described in the comprehensive review by Boellaard²¹ there are a large number of potential sources of bias and variance in determining SUVs. To provide a simplified and integrated view of the error dependency for PET/CT SUVs the general structure outlined in Fig 2 is used²². The impact of each group will be considered in turn. Image artifacts (e.g. from PET scanner malfunctions or patient motion between the PET and CT scans) can impact SUV accuracy, but are not considered separately here.

3.1. Imaging Physics

The role of a PET scanner is to form an image of the spatially varying concentration of positron-emitters, typically ¹⁸F, which has a half-life of 110 minutes²³. The fundamental limitations on PET imaging are detector spatial resolution and total effective counts. Together these determine the overall achievable signal to noise ratio (SNR) in a PET image. Other data processing steps in Fig 2 can further reduce SNR, e.g. errors in scatter estimation or correction, by adding bias and/or variance to the measured data.

Spatial resolution is limited by the scintillator properties, the construction details of the detector units, and the detector electronics. With clinical scanners that are currently being manufactured, the best-case reconstructed image resolution is typically on the order of 4 to 7 mm full-width half-maximum (FWHM). This limited resolution leads to the well-known partial volume effect (PVE) or partial volume error²⁴, where for objects smaller than a few cm, the measured tracer concentration is less than the true tracer concentration value. This effect is illustrated in Fig 3 in this clinically relevant example where the ratio of measured/true SUV (called the recovery coefficient (RC)) is less than 100% for objects less than 30 mm in size. The partial volume effect causes an increasing reduction (i.e. increasing bias) in the measured SUV for objects of decreasing size.

It should be noted that the measured SUV in Fig 3 is only 65% to 85% of the true value for 2 cm spheres, depending on degree of image smoothing. In principle it is possible to correct

for partial volume errors if the true spatial distribution of the tracer is known (e.g. from CT). However, there are several significant challenges to partial volume correction methods²⁵, and such methods have not been widely adopted or made available by instrument manufacturers.

Total effective counts determines the noise in the measured data according to:

$$\text{Total effective counts} = (\text{scanner effective count rate [cps]}) \times (\text{scan duration [s]})$$

The effective count rate is less than the true event rate due to the impact of noise contributions from scattered and random events²⁶. The scanner effective count rate is also determined by the system sensitivity and live-time fraction at the clinical tissue activity levels used in practice. The presence of statistical noise (and limited spatial resolution) is visible in the PET image of Fig 3.

While the total effective counts are directly related to lesion detectability²⁷, the impact of total effective counts on SUV accuracy is probably not as important as other effects discussed below. For the typical total effective count levels used in Fig 3 (5 min scan per 16 cm bed position with a 10 mCi FDG injection 60 min prior to scanning), the SUV error bars are relatively small, with a coefficient of variation of approximately 3%²⁸. For example a 3% variation in SUV due to statistical noise is much less than the 25% variation caused by the partial volume effect of a 2 cm diameter lesion and other effects.

3.2 Patient status

The true tracer uptake in a patient is composed of two components: the first being the amount of tracer uptake (e.g. FDG) associated with the disease status (the signal of interest), which can be modified by the biophysiological status of the patient. One of the more important patient parameters is the blood glucose level, which has been shown to inversely-linearly affect SUVs²⁹. A prospective study by Crippa et al.³⁰ in eight patients showed that as blood glucose levels were increased from 92.4 ± 10.2 to 158 ± 13.8 mg/100 ml by glucose loading, the average SUV of 20 liver metastases decreased from 9.4 ± 5.7 to 4.3 ± 8.3 . Endogenous insulin levels can be equally important, but not as easily recognized as measurements are not routinely made. Patients who have failed to comply with caloric restriction will often have diffusely increased levels of FDG in muscle as a manifestation of insulin release following a meal. In this regard, insulin facilitates the availability of GLUT4 (a major membrane transporter of glucose in muscle) in the plasma membrane increasing uptake of FDG by muscle with less available for equilibration in tumor tissue. It is for these reasons that caloric intake and use of oral as well as injectable forms of insulin are strictly regulated in imaging protocols³¹.

Several other potential sources of bias are described by Huang²⁹, for example, that chemotherapy can result in impaired renal function, significantly reducing the clearance of plasma FDG through the kidney and thus increasing tumor SUV relative to an initial PET scan, even if there were no change in the underlying tumor glucose metabolic rate. The use of tracer kinetic modeling, which accounts for transport and biochemical reactions of FDG in tissue, can remove many of the effects of the confounding biological factors. However, tracer kinetic modeling methods are not yet clinically feasible.

The second component of the true tracer uptake is biological variability. The biological variability has been estimated in several test-retest studies^{7,32–35} at approximately 10% for scans repeated within a few days, which is a significantly larger source of variation than the typical scanner test-retest measurement variability of ~3% discussed above.

3.3 Scan Protocol

There are several components of a PET/CT scan protocol that can significantly effect SUV accuracy. These include (1) the uptake duration between injection and scan, (2) measuring the residual activity in the syringe, (3) accurate measurement of patient weight, (4) synchronization of clocks used for dose assays and scanning, (5) patient respiratory motion, and (6) correct data entry. Scan duration does not have a significant effect on SUV accuracy, as discussed above, except possibly for extremely short scans (e.g. less than 1–2 min per bed position). Scan mode (2D or 3D) does not by itself have a significant effect on SUV accuracy. As discussed below, however, the data correction methods (e.g. for scatter) can differ between data acquisition modes, and any differences in the accuracy of the data correction methods between data acquisition modes will impact SUVs.

Uptake duration typically linearly effects SUVs. An example of the affect of uptake duration on SUVs is shown in Fig 4. There is in general a linear increase in SUV with time over the typical range of uptake periods used in clinical practice. This general behavior has been shown for other types of cancer³⁶.

There are two implications from Fig 4: Tumor SUV will in general depend on the duration of the uptake time between FDG injection and scan start. In addition longitudinal or serial scans of the same patient will potentially have SUV changes introduced if the uptake period is not consistent. This was the motivation for a recommendation in a consensus report that duration periods remain consistent within ± 10 min³⁷.

Residual activity in the syringe after patient injection, if non-zero, directly affects the SUV since it is the net injected activity that should be used in Equation 1. Fig 5 presents the residual activity for 250 patients as function of the injected dose.

This illustrates variations in injected dose and post-injection residual activity. If both the injected dose and residual activity are accurately incorporated, then no error will be introduced into the SUV estimation. Ignoring the residual activity introduces a median SUV underestimation of approximately 2%. This underestimation, however, would be significantly larger in a few patients based on the data of Fig 5. A closely related source of error is when there is extravasation of the dose, which can trap a large fraction of the FDG in tissue near the injection site and reduce SUVs. In this case normalization to a reference tissue may be the best option.

Patient weight or lean body mass is used as the normalizing volume in Equation 1. If the patient weight is measured on an uncalibrated scale it may not be accurate. In addition previously measured or self-reported patient weights also may not be accurate.

Clock synchronization errors affect SUVs due to the radioactive decay corrections for ^{18}F and the need to maintain consistent uptake durations (Fig 4). Accurate timing information is needed for the decay correction that accounts for the delay between the two syringe measurement times, the injection time, and the scan start time as illustrated in Fig 6. Due to the 110 min half-life of ^{18}F , a 10 min drift in clock timing can lead to a ~6% SUV error from inaccurate decay corrections. As shown by Fig 4, changes in the uptake period can lead to variable amounts of bias, depending on the rate of FDG accumulation by the tumor.

Patient respiratory motion can lead to a ~25% (or more) under-estimation of SUV and a ~2-fold overestimation in size for small lesions near the diaphragm. These errors typically decrease with increased lesion size and/or decreased respiratory motion¹⁴. Respiratory effects can introduce additional errors in the form of image artifacts arise from positional mismatch of the PET data and the CT images used for attenuation correction³⁸. As for many

of the factors that influence SUV accuracy, changes in patient respiratory patterns between serial scans due to changes in the acquisition protocol (for example, patient anxiety, breathing pattern coaching by the technologist, or room ambiance) will confound the ability to measure true changes in SUV. In addition changes in respiratory patterns between the CT acquisition used for attenuation correction and the PET can also introduce artifacts, particularly in the lungs and near the diaphragm. Methods to compensate for respiratory effects are still currently under development and/or evaluation³⁹.

Correct data entry might seem redundant to include as a source of error, but as shown below, there are a typically 9 to 15 steps where the technologist has to make a measurement, record a value or enter a value. A mistake in any one of these steps can cause corresponding errors in estimated SUVs and is often the most frequent cause of erroneous SUVs.

3.4 Data Processing

There are two main components in the data processing step, the first is the set of necessary corrections for confounding effects. These include (in rough order of decreasing magnitude) attenuation, scattered and random coincidences, scanner deadtime, and detector efficiency variations. Although attenuation is the largest effect, correction used CT-based or other methods has been shown to be relatively accurate³⁸. In contrast scatter correction methods are more technically challenging, especially where there are large changes in density, such as in the lung or in bone. In these cases the bias introduced by inaccurate scatter correction is not well understood.

The second data processing component is image reconstruction, where the raw scanner data are transformed into an image of relative radioactivity concentration. The data corrections described above are applied before, or during, the image reconstruction process. The two main types of image reconstruction are the analytical and statistical/iterative methods⁴⁰. Analytical methods, such as filtered backprojection have well understood quantitative behaviors, but have fallen out favor due to the superior noise suppression properties of the statistical/iterative methods, such as ordered subsets expectation maximization (OSEM)⁴⁰.

Statistical/iterative methods are non-linear and have multiple parameters. This means that the methods are less predictable and less well-understood than analytical methods. The impact of the basic parameters, such as type and amount of smoothing, depend on the specific algorithm and implementation. However with respect to image smoothness, the well known trade-off of noise versus contrast still applies to statistical/iterative methods. Increased smoothing reduces noise but leads to decreased contrast, which in turn reduces SUV accuracy. Reduced image noise will also reduce the variability of SUVs, but as discussed this effect is already relatively small. These effects are illustrated in Figs 3 and 7.

In Fig 7 three levels of increasing smoothing are applied: 4, 7, and 10 mm FWHM, which cover most values used clinically. With increased smoothing there is an increase in the bias. The reduced noise due to increased smoothing is shown quantitatively by the narrower error bars in Fig 3, whose average values are approximately 3.5%, 2.5% and 2%²⁸.

There are several other image reconstruction parameters that will affect SUV values^{21,41} such as number of iterations, field of view and voxel dimensions, but a full analysis of these effects is beyond the scope of this review. It should also be noted that manufacturers will often update and/or enhance image reconstruction methods implemented on PET scanners. Recent developments, such as detector modeling and time-of-flight imaging, in principle improve image SNR, but their impact on SUV accuracy is less well understood⁴⁰.

3.5 Scanner Calibration

After reconstruction, the PET image is not yet in SUVs, but rather typically in units that are specific to the scanner, the reconstruction method, and the reconstruction parameters. To convert the image to SUV two calibration steps are needed. The first requires the estimation of the scanner calibration factor, which relates scanner units to kBq/ml. This estimation process varies by scanner, but is typically performed quarterly using a 20 cm diameter water-filled phantom using the process outlined in Fig 8.

Some scanners use a solid phantom containing $^{68}\text{Ge}/^{68}\text{Ga}$, which has a 270 day half life. The advantage of this process is that the sealed source phantom does not need filling each time. However a cross-calibration with the dose calibrator is still required.

The scanner calibration factor is then used to convert reconstructed images of patients from scanner units to radioactivity concentration [kBq/ml] as shown in Fig 9. The final step is the conversion of the image units to SUVs using Equation 1.

There are several sources of error that impact the bias and variance introduced into SUVs by these calibration steps. If there is a error in the scanner calibration factor, e.g. from an erroneous dose calibrator setting estimated during the process illustrated in Fig 8, it will affect all subsequent patient images. However in converting the images from kBq/ml to SUVs this error is potentially cancelled out if the same dose calibrator setting is used for patient scans (Fig 9). The impact of a scanner mis-calibration is illustrated in Fig 10.

Recent studies by Lockhart et al⁴² have shown that the variability of the calibration factors is approximately 4%, but only if there are no operator errors. The impact of operator errors on scanner calibration is difficult to assess, but anecdotal evidence suggests that at least 10% of all calibrations have measurable effects due to mistakes during the calibration process (Fig 8) and/or the patient-specific SUV calibration process (Fig 9). This observation is supported by Fig 11, showing Scanner calibration factors measured over several years using standard procedures with ^{18}F -FDG (as shown in Fig 8) and with fixed $^{68}\text{Ge}/^{68}\text{Ga}$ 20 cm cylinder source (270 d half-life).

Two operator errors were identified during this period as indicated. Sample standard deviations are:

- Standard procedures with ^{18}F -FDG (Fig 8): 6.1%
- Standard procedures with known operator errors removed: 4.1%
- Using the fixed $^{68}\text{Ge}/^{68}\text{Ga}$ source 3.2%

These results indicate that long term scanner calibration variations add approximately 3% variability, while operator interventions in the process increase this by an additional 1% to 3%. In addition some variability is introduced by the dose calibrators themselves⁴³. However, it is possible for a calibration error to add a larger bias to either a single patient study or to a group of studies.

3.6 Analysis Methods

After the images are reconstructed, there remains the impact of the choice of method used to analyze the tracer uptake in PET image after it has been converted to SUVs. In addition it is possible, albeit challenging, to measure the size of the lesion from the PET image alone.

SUVmean and Region of Interest (ROI) Definition—In an ideal case, where there was no resolution loss or uncertainty in boundary definition, simply computing the average SUV within a region of interest (ROI) would produce a reliable estimate of the mean SUV,

which is defined here as SUVmean. However, in practice, there are challenges imposed by image noise and the limited resolution of PET imaging as discussed in section 3.2. Both of these effects contribute to problems in defining the boundary of the region over which the average is to be computed. Numerous studies have attempted to define methods that are accurate and/or reproducible.

A comparison of several published methods was conducted by Nestle⁴⁴ in the context of radiotherapy planning, who concluded that “The different techniques of tumor contour definition by 18F-FDG PET in radiotherapy planning lead to substantially different volumes, especially in patients with inhomogeneous tumors”. In addition other studies, such as those by Ford et al⁴⁵ have shown that region definition depends on the reconstruction method, the amount of smoothing, and the lesion/background ratio. Finally, for regions less than approximately 3 cm, the partial volume effect leads to errors in the measured SUV. This is illustrated in Fig 3 and in Fig 12, where close inspection also reveals that the threshold needed for accurate lesion size determination depends on the size of the lesion.

For the reasons listed above, accurate measurement of SUVmean is only possible using the central region of relatively large lesion (e.g. the central 2 cm of the 5 cm lesion in Fig 5), and even then this is only accurate if the lesion has a uniform true SUV value. A second consequence of the challenges for measuring SUVmean described above is that measures of SUVmean are not as reproducible as the maximum SUV value.

SUVmax—Inspection of Fig 12 indicates that the maximum value of the measured value in the 2 cm lesion is a more accurate estimate of the true SUV than SUVmean. For this reason the use of the maximum SUV value, defined here as SUVmax, is becoming more common as indicated in Fig 13. In addition SUV max has a significantly improved reproducibility as compared to SUVmean, since the maximum value within an ROI is typically invariant with respect to small spatial shifts of the ROI.

A concern with the use of SUVmax is that it is basing a reported value for a lesion on perhaps only one pixel. Thus SUVmax is potentially biased and more noisy when compared to SUVmean and/or the true SUV. Recent results from Doot et al.²⁸ however have indicated both the bias and increased variance are less than might be expected, likely due to the noise correlations introduced during the image reconstruction process. In addition several patient studies have shown that SUVmax is a robust metric for assessing treatment response (e.g.⁴⁶).

Finally it should be noted that there are other metrics used for reporting SUVs from images, including the mean value within a fixed size ROI (regardless of lesion size)⁴⁷ and the integrated SUV within an SUV⁴⁸.

3.7 Summary

With reference to Fig 2, it is not yet possible to determine the overall bias or variance in SUV values. As noted above, however, several studies have estimated that the test-retest reproducibility of ~10% for tumor SUVs. These studies, however were monitored at an academic medical center, and so do not necessarily represent routine clinical practice. A second consideration is that the scans were repeated within a few days and so do not include longer term effects (e.g. scanner calibration variations). In addition it has been suggested that the variability when there is a change in SUV may be a log-normal distribution, rather than a normal distribution⁴⁹.

In a landmark study to assess SUV reproducibility at multiple centers, Velasquez et al.⁵⁰ performed test-retest scans of 62 patients at eight imaging centers. The test-retest 18F-FDG

PET studies were performed within 7 d (4.1 ± 2.6 d) of each other. Of the 62 patients who received repeat scans, only 41 scan sets (the 'QA dataset') were considered to have scan quality and/or patient information sufficient for accurate assessment of SUVs in both scans. The intrasubject coefficient of variation was 10%–12% for patients in the QA dataset, confirming the studies showing an approximate 10% reproducibility for carefully performed test-retest studies. However, this was obtained at the expense of removing 1/3 of the full dataset of patient studies. With the full dataset, there was an increase in SUVmax variability to approximately 16%. In addition it was noted that the largest effect on SUVmax variance was inconsistent uptake periods (e.g. Fig 4). For the subset of patients with differences of more than 15 min in the test and retest uptake durations, the SUVmax variability exceeded 21%. Variations in blood glucose concentration are potentially also problematic, but in this study essentially all patients had blood glucose concentrations under the recommended guidelines. This study was performed as part of a clinical trial incorporating PET imaging with detailed instructions. In addition as noted above short-term test-retest studies do not capture longer term sources of error. Recent studies by Doot et al. for multiple scanners at multiple time points have shown changes of $\pm 15\%$ for the same scanners at different time points⁵¹. Thus one can expect the actual variability of SUVmax in practice to be *greater* than 15% to 20%.

One alternative approach which has been suggested to assist reduction in variability of SUV measurements incorporates the use of "reference tissue" SUVmax values and normalization of lesion/target SUV measures to those of selected reference tissues. A number of tissues have been advocated as reference tissue including, blood pool, liver, lung and cerebellum. Perry et al evaluated the issue of which reference tissue was best for this purpose.⁵² Interpatient reproducibility and variance of multiple reference tissues was measured in 12 patients with at least 3 consecutive PET/CT studies over periods up to one year apart. They found that mediastinal blood pool showed the least inter-patient coefficient of variance of 0.17, followed by liver at 0.21 lung at 0.22 and cerebellum at 0.25. The patients were receiving various therapies for cancer during the period of measurement, and the advantage of this metric is that any influence of therapeutic intervention on SUVs would have been incorporated into the measurements.

4 Recommendations

While the overall variability of SUVs in practice is still not known, there are several steps that can reduce this variance, in particular for providing more reliable assessments of response to therapy. These include:

1. Monitoring scanner calibrations across time. Performance of manufacturer-recommended procedures for scanner calibration is necessary but not sufficient. This is particularly important when there are changes to scanner hardware or software.
2. Monitoring dose calibrator settings and all clock and timing systems used in the acquisition process (including internal PET scanner timing systems).
3. Measuring patient weight using a calibrated scale.
4. Performing serial patient studies at the same imaging center, using the same PET scanner, same display station, and the same protocols for preparation, acquisition, processing and analysis.
5. Ensuring repeat patient scans are acquired with uptake durations that are within ± 15 min of each other and with blood glucose concentrations within published guidelines³⁷.

6. Critical information should be tracked and recorded on a per-patient basis, including:
 - blood glucose concentration
 - injected and residual doses
 - uptake duration
 - untoward events (e.g. tracer extravasation, unusual respiratory motion, potential artifacts)
7. Monitoring average liver SUV or other reference tissues as a check that overall SUVs are within a normal range^{53, 52}.
8. Using consistent analysis and reporting procedures for SUVs.

Acknowledgments

We appreciate the encouragement of Paul Shreve for this work, and numerous helpful discussions with many individuals, including Ronald Boellaard, Robert Doot, Chi Liu, David Mankoff, Osama Mawlawi, Richard Wahl, Jeffery Yap, members of AAPM Task Group 145 and the RSNA Quantitative Imaging Biomarkers Alliance. The work was supported in part by NIH grants CA74135 and CA115870 and NCI Contract 24XS036-004.

References

1. Weber WA, Grosu AL, Czernin J. Technology Insight: advances in molecular imaging and an appraisal of PET/CT scanning. *Nat Clin Pract Oncol* 2008;5:160–170. [PubMed: 18253106]
2. Fletcher JW, Djulbegovic B, Soares HP, et al. Recommendations on the Use of 18F-FDG PET in Oncology. *J Nucl Med* 2008;49:480–508. [PubMed: 18287273]
3. 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. [PubMed: 15837727]
4. Weber WA. Assessing tumor response to therapy. *J Nucl Med* 2009;50 (Suppl 1):1S–10S. [PubMed: 19380403]
5. Higashi K, Clavo AC, Wahl RL. Does FDG uptake measure proliferative activity of human cancer cells? In vitro comparison with DNA flow cytometry and tritiated thymidine uptake. *J Nucl Med* 1993;34:414–419. [PubMed: 8441032]
6. Thie JA. Understanding the standardized uptake value, its methods, and implications for usage. *J Nucl Med* 2004;45:1431–1434. [PubMed: 15347707]
7. Nakamoto Y, Zasadny KR, Minn H, et al. Reproducibility of common semi-quantitative parameters for evaluating lung cancer glucose metabolism with positron emission tomography using 2-deoxy-2-[18F]fluoro-D-glucose. *Mol Imaging Biol* 2002;4:171–178. [PubMed: 14537140]
8. Keyes JWJ. SUV: standard uptake or silly useless value? *J Nucl Med* 1995;36:1836–1839. [PubMed: 7562051]
9. Coleman RE. Is quantitation necessary for oncological PET studies? *For. Eur J Nucl Med Mol Imaging* 2002;29:133–135. [PubMed: 11807615]
10. Boellaard R, Krak NC, Hoekstra OS, et al. 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. [PubMed: 15347719]
11. Gambhir SS. Molecular imaging of cancer with positron emission tomography. *Nat Rev Cancer* 2002;2:683–693. [PubMed: 12209157]
12. Hallett WA, Maguire RP, McCarthy TJ, et al. Considerations for generic oncology FDG-PET/CT protocol preparation in drug development. *IDrugs* 2007;10:791–796. [PubMed: 17968761]
13. Juweid ME, Stroobants S, Hoekstra OS, et al. Use of positron emission tomography for response assessment of lymphoma: consensus of the Imaging Subcommittee of International Harmonization Project in Lymphoma. *J Clin Oncol* 2007;25:571–578. [PubMed: 17242397]

14. Hillner BE, Siegel BA, Shields AF, et al. The impact of positron emission tomography (PET) on expected management during cancer treatment: findings of the National Oncologic PET Registry. *Cancer* 2009;115:410–418. [PubMed: 19016303]
15. Husband JE, Schwartz LH, Spencer J, et al. Evaluation of the response to treatment of solid tumours - a consensus statement of the International Cancer Imaging Society. *Br J Cancer* 2004;90:2256–2260. [PubMed: 15150551]
16. Michaelis LC, Ratain MJ. Measuring response in a post-RECIST world: from black and white to shades of grey. *Nat Rev Cancer* 2006;6:409–414. [PubMed: 16633367]
17. Boellaard R, O'Doherty MJ, Weber WA, et al. FDG PET and PET/CT: EANM procedure guidelines for tumour PET imaging: version 1.0. *Eur J Nucl Med Mol Imaging* 2010;37:181–200. [PubMed: 19915839]
18. Delbeke D, Coleman RE, Guiberteau MJ, et al. Procedure guideline for tumor imaging with 18F-FDG PET/CT 1.0. *J Nucl Med* 2006;47:885–895. [PubMed: 16644760]
19. Fukukita H, Senda M, Terauchi T, et al. Japanese guideline for the oncology FDG-PET/CT data acquisition protocol: synopsis of Version 1.0. *Ann Nucl Med* 2010;24:325–334. [PubMed: 20401547]
20. Rohren, EM. PET/CT Reporting in Radiation Therapy Planning and Response Assessment. *Seminars in Ultrasound, CT and MRI*; 2010.
21. Boellaard R. Standards for PET Image Acquisition and Quantitative Data Analysis. *J Nucl Med* 2009;50:11S–120. [PubMed: 19380405]
22. 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. [PubMed: 19956382]
23. Cherry, SR.; Sorenson, JA.; Phelps, ME. *Physics in Nuclear Medicine*. Philadelphia, PA: Saunders; 2003.
24. Soret M, Bacharach SL, Buvat I. Partial-Volume Effect in PET Tumor Imaging. *J Nucl Med* 2007;48:932–945. [PubMed: 17504879]
25. Tylski P, Stute S, Grotus N, et al. Comparative Assessment of Methods for Estimating Tumor Volume and Standardized Uptake Value in 18F-FDG PET. *J Nucl Med* 2010;51:268–276. [PubMed: 20080896]
26. Strother SC, Casey ME, Hoffman EJ. Measuring PET scanner sensitivity: relating countrates to imagesignal-to-noise ratios using noise equivalents counts. *IEEE Transactions on Nuclear Science* 1990;37:783–788.
27. Kinahan, PE.; Cheng, PM.; Alessio, AM., et al. A quantitative approach to a weight-based scanning protocol for PET oncology imaging. 2005 IEEE Nuclear Science Symposium Conference Record; 2005. p. 1886-1890.
28. Doot, RK.; Scheuermann, JS.; Christian, PE., et al. Medical Physics. 2010. Instrumentation factors affecting variance and bias of quantifying tracer uptake with PET/CT.
29. Huang SC. Anatomy of SUV. Standardized uptake value. *Nucl Med Biol* 2000;27:643–646. [PubMed: 11091106]
30. Crippa F, Gavazzi C, Bozzetti F, et al. The influence of blood glucose levels on [18F]fluorodeoxyglucose (FDG) uptake in cancer: a PET study in liver metastases from colorectal carcinomas. *Tumori* 1997;83:748–752. [PubMed: 9349315]
31. Roy FN, Beaulieu S, Boucher L, et al. Impact of intravenous insulin on 18F-FDG PET in diabetic cancer patients. *J Nucl Med* 2009;50:178–183. [PubMed: 19164226]
32. Krak NC, Boellaard R, Hoekstra OS, et al. 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. [PubMed: 15791438]
33. Minn H, Zasadny KR, Quint LE, et al. Lung cancer: reproducibility of quantitative measurements for evaluating 2-[F-18]-fluoro-2-deoxy-D-glucose uptake at PET. *Radiology* 1995;196:167–173. [PubMed: 7784562]
34. Nahmias C, Wahl LM. Reproducibility of standardized uptake value measurements determined by 18F-FDG PET in malignant tumors. *J Nucl Med* 2008;49:1804–1808. [PubMed: 18927325]

35. Weber WA, Ziegler SI, Thodtmann R, et al. Reproducibility of metabolic measurements in malignant tumors using FDG PET. *J Nucl Med* 1999;40:1771–1777. [PubMed: 10565769]
36. Lowe VJ, DeLong DM, Hoffman JM, et al. Optimum scanning protocol for FDG-PET evaluation of pulmonary malignancy. *J Nucl Med* 1995;36:883–887. [PubMed: 7738668]
37. Shankar LK, Hoffman JM, Bacharach S, et al. Consensus Recommendations for the Use of 18F-FDG PET as an Indicator of Therapeutic Response in Patients in National Cancer Institute Trials. *J Nucl Med* 2006;47:1059–1066. [PubMed: 16741317]
38. Kinahan PE, Hasegawa BH, Beyer T. X-ray-based attenuation correction for positron emission tomography/computed tomography scanners. *Semin Nucl Med* 2003;33:166–179. [PubMed: 12931319]
39. Nehmeh SA, Erdi YE. Respiratory motion in positron emission tomography/computed tomography: a review. *Semin Nucl Med* 2008;38:167–176. [PubMed: 18396177]
40. Tong, S.; Alessio, AM.; Kinahan, PE. Imaging in Medicine. 2010. Image reconstruction for PET/CT scanners: past achievements and future challenges.
41. Adams MC, Turkington TG, Wilson JM, et al. A Systematic Review of the Factors Affecting Accuracy of SUV Measurements. *Am J Roentgenol* 2010;195:310–320. [PubMed: 20651185]
42. Lockhart CM, MacDonald LR, Alessio A, et al. Minimizing instrument calibration error to reduce the effect of variability on PET/CT SUV measurements. *J Nucl Med* 2009;50:235. (abstract).
43. Zimmerman B, Kinahan P, Galbraith W, et al. Multicenter comparison of dose calibrator accuracy for PET imaging using a standardized source. *J Nucl Med* 2009;50:472. (abstract).
44. Nestle U, Kremp S, Schaefer-Schuler A, et al. Comparison of Different Methods for Delineation of 18F-FDG PET-Positive Tissue for Target Volume Definition in Radiotherapy of Patients with Non-Small Cell Lung Cancer. *J Nucl Med* 2005;46:1342–1348. [PubMed: 16085592]
45. Ford EC, Kinahan PE, Hanlon L, et al. Tumor delineation using PET in head and neck cancers: threshold contouring and lesion volumes. *Med Phys* 2006;33:4280–4288. [PubMed: 17153406]
46. Benz MR, Allen-Auerbach MS, Eilber FC, et al. Combined assessment of metabolic and volumetric changes for assessment of tumor response in patients with soft-tissue sarcomas. *J Nucl Med* 2008;49:1579–1584. [PubMed: 18794268]
47. Ott K, Fink U, Becker K, et al. Prediction of response to preoperative chemotherapy in gastric carcinoma by metabolic imaging: results of a prospective trial. *J Clin Oncol* 2003;21:4604–4610. [PubMed: 14673049]
48. Larson SM, Erdi Y, Akhurst T, et al. Tumor Treatment Response Based on Visual and Quantitative Changes in Global Tumor Glycolysis Using PET-FDG Imaging. The Visual Response Score and the Change in Total Lesion Glycolysis. *Clin Positron Imaging* 1999;2:159–171. [PubMed: 14516540]
49. Thie JA, Hubner KF, Smith GT. The diagnostic utility of the lognormal behavior of PET standardized uptake values in tumors. *J Nucl Med* 2000;41:1664–1672. [PubMed: 11037996]
50. Velasquez LM, Boellaard R, Kollia G, et al. Repeatability of 18F-FDG PET in a multicenter phase I study of patients with advanced gastrointestinal malignancies. *J Nucl Med* 2009;50:1646–1654. [PubMed: 19759105]
51. Doot R, Allberg K, Kinahan P. Errors in serial PET SUV measurements. *J Nucl Med* 2010;51:126P.
52. Perry K, Tann M, Miller M. Which reference tissue is best for semiquantitative determination of FDG activity? *J Nucl Med* 2008;49:425P-a.
53. Paquet N, Albert A, Foidart J, et al. Within-patient variability of (18)F-FDG: standardized uptake values in normal tissues. *J Nucl Med* 2004;45:784–788. [PubMed: 15136627]
54. Beaulieu S, Kinahan P, Tseng J, et al. SUV varies with time after injection in (18)F-FDG PET of breast cancer: characterization and method to adjust for time differences. *J Nucl Med* 2003;44:1044–1050. [PubMed: 12843218]
55. Wahl RL, Jacene H, Kasamon Y, et al. From RECIST to PERCIST: Evolving Considerations for PET Response Criteria in Solid Tumors. *J Nucl Med* 2009;50:122S–1150. [PubMed: 19403881]

Fig 1.
Roles of SUVs in Quantitative Imaging with PET/CT.

Fig 2.
Error dependency for PET/CT SUVs.

Fig 3.

Left: Transverse PET and CT images of a modified NEMA image quality phantom containing six spheres with diameters from 10 to 37 mm. There is a 4:1 sphere:background ratio of FDG concentration in each sphere. Right: Recovery coefficient (ratio of measured/true SUV) as a function of object size for the six spheres. Error bars are based on 20 repeated 5 min scans at typical clinical tissue activity values²⁸. (Reprinted with permission from.)

Fig 4.
Average breast cancer tumor SUV values versus time from injection for 20 patients⁵⁴.
(Reprinted with permission from.)

Fig 5.
Injected dose and post-injection residual activity for 250 patients. Data courtesy of Dr Osama Mawlawi, MD Anderson Cancer Center.

Fig 6.
Timing of measurement and injection steps needed for accurate decay injection steps.

Fig 7.
Effect of increased smoothing on SUV quantitation. S1 and S2 are 1 cm diameter spheres with a true SUV of 4.0 in a torso phantom. Increased smoothing increases the bias of the measured SUV of the small spheres and reduces noise.

Fig 8.
Steps in establishing scanner calibration factor.

Fig 9. Steps in generating SUV images showing the dependency on patient weight, decay corrected net injected dose, and scanner calibration factor.

Fig 10.
Impact of incorrect scanner calibration on patient SUV values for a large lung lesion (box).
Both images are scaled to their maximum values.

Fig 11. Scanner calibration factors measured over a 3.6 year period using standard procedures with ^{18}F -FDG and with fixed $^{68}\text{Ge}/^{68}\text{Ga}$ source.

Fig 12.

Left: Image of a simulated 33 cm diameter cylindrical phantom containing two cylindrical lesions of 2 and 5 cm diameter. The SUVs of the background and lesions are 1.0 and 2.0. Right: Profile is from a reconstructed image including the spatial volume errors and reconstruction smoothing effects. The amount of signal loss in the measured SUV depends on lesion size and position within the lesion.

Fig 13.

Use of SUVmax versus SUV mean (and other approaches in scientific publications. Data from Wahl et al.⁵⁵ (Reprinted with permission from.)