Review

Tumor Quantification in Clinical Positron Emission Tomography

Bing Bai[1], James Bading[2], Peter S Conti[1] ✉

1. Department of Radiology, University of Southern California, Los Angeles, CA;
2. Department of Cancer Immunotherapy and Tumor Immunology, City of Hope, Duarte, CA.

✉ Corresponding author: Peter S Conti, MD, PhD. Department of Radiology, University of Southern California, 2250 Alcazar Street, CSC/IGM 103, Los Angeles, CA 90033. Phone: 323 442 5940 Fax: 323 442 3253 Email: pconti@usc.edu.

Abstract

Positron emission tomography (PET) is used extensively in clinical oncology for tumor detection, staging and therapy response assessment. Quantitative measurements of tumor uptake, usually in the form of standardized uptake values (SUVs), have enhanced or replaced qualitative interpretation. In this paper we review the current status of tumor quantification methods and their applications to clinical oncology. Factors that impede quantitative assessment and limit its accuracy and reproducibility are summarized, with special emphasis on SUV analysis. We describe current efforts to improve the accuracy of tumor uptake measurements, characterize overall metabolic tumor burden and heterogeneity of tumor uptake, and account for the effects of image noise. We also summarize recent developments in PET instrumentation and image reconstruction and their impact on tumor quantification. Finally, we offer our assessment of the current development needs in PET tumor quantification, including practical techniques for fully quantitative, pharmacokinetic measurements.

Key words: Positron emission tomography, tumor quantification

INTRODUCTION

The chief advantages of radionuclide imaging in clinical oncology are its high tumor to non-tumor contrast and its ability to provide functional information about tumors. Compared with single photon emission computed tomography (SPECT), positron emission tomography (PET) has much higher sensitivity and substantially better and more uniform spatial resolution. Chiefly because of the better image resolution and the accuracy with which PET images can be corrected for photon attenuation in tissue, image intensity more closely depicts relative radiolabel concentration within the patient for PET than for SPECT. Thus, PET is said to be more "quantitative" than SPECT, i. e., better able to provide measurements of radiotracer distribution within the imaged subject. A recent example of PET accuracy is a study by Pryma et al., who compared PET-derived measurements of whole-tumor activity concentration with direct assay (i. e., weighing and well counting) and quantitative autoradiography of renal masses surgically resected from 26 patients injected with an ^{124}I-labeled antibody.[1] Highly positive correlations were found between PET and the ex-vivo measurements (Spearman rank correlation coefficients = 0.84 and 0.88 for well counting and autoradiographic measurements, respectively).

The most common application of PET in clinical oncology has been tumor detection and assessment of disease extent using PET combined with computed tomography (PET-CT) and ^{18}F-fluorodeoxyglucose (^{18}F-FDG). Qualitative, visual scan interpretation is highly effective in this context, and still predominates

over quantitative assessment.[2, 3] There are, however, numerous applications in which PET-derived measurements of tumor uptake have enhanced or replaced qualitative interpretation. Early PET studies used percent of administered dose per gram of tissue as a measure of tumor uptake. However when comparing between different patients this value is affected by the size of the patient.[4] A more meaningful measurement of the tumor uptake is standardized uptake value (SUV), which is the decay-corrected tumor activity concentration divided by injected activity per unit body weight, surface area or lean body mass.[5] Grgic et al. studied 140 patients who underwent ^{18}F-FDG/PET and subsequent pathologic assessment for solitary pulmonary nodules (SPNs).[6] They found that maximum single-voxel SUV (SUV_{max}) effectively distinguished malignant from benign SPNs, and that $SUV_{max} \geq 9.5$ predicted shorter survival (median 20 months) than lower SUV_{max} values (> 75 months). Schőder et al. reported on 97 patients with non-Hodgkin lymphoma (NHL) in whom whole-tumor mean SUV (SUV_{mean}) for ^{18}F-FDG was approximately 3 times higher in pathologically-confirmed aggressive NHL compared with indolent forms of the disease.[7]

Quantitative analysis is better than visual assessment for distinguishing effective from ineffective treatment early during the course of therapy.[3] For example, Lin et al. studied 92 patients with newly-diagnosed diffuse large B-cell lymphoma who were scanned with ^{18}F-FDG/PET before and mid-course course during chemotherapy.[8] Visual assessment predicted 2 year event-free survival (EFS) with an accuracy of 51% for FDG-positive patients and 79% for FDG-negative patients. On the other hand, only 21% of patients with SUV_{max} reduction < 66% had an EFS > 2 years, compared with 79% of patients with > 66% reduction (Figure 1). Fourteen patients considered positive on visual analysis could have been re-classified as good responders. SUV analysis of pre-therapy tumor uptake and changes in tumor uptake during therapy have also been shown to predict survival in non-small-cell lung cancer[9] and esophageal cancer.[10]

Tumor quantification has enabled other clinical PET applications besides tumor detection, response assessment and prognosis. In the category of tumor characterization, Cochet et al. used kinetic modeling of dynamic data acquired with PET during the first 2 minutes after bolus injection of ^{18}F-FDG to determine tumor blood flow in 40 patients with breast cancer.[11] The findings correlated positively with the expression of angiogenesis-related endothelial markers evaluated in tumor biopsy samples. Tumor oxygenation is a strong determinant of response to radio- and chemotherapy, and several PET radiopharmaceuticals have been developed which localize in areas of tissue hypoxia. Hugonnet and collaborators used ^{18}F-fluoromisonidazole to measure the effects of the anti-angiogenic drug sunitinib on tumor hypoxia in 53 patients with metastatic renal cell carcinoma.[12] They found that sunitinib reduced hypoxia in initially hypoxic tumors and did not induce it in initially non-hypoxic metastases. PET is frequently used to measure the relative expression of various receptors in tumors, especially when such receptors are potential therapeutic targets. O'Donaghue et al. found that tumor uptake of the humanized [124]I-huA33 antibody was directly related to A33 antigen expression in surgically resected tumors.[13] Peptides are also frequently used to image receptor expression. Because peptides are generally retained on receptors for much shorter intervals than are antibodies, they may require kinetic analysis. For example, Tomasi et al. employed ^{18}F-fluciclatide, dynamic PET and compartmental modeling to derive a parameter quantifying tumor $\alpha(v)\beta(3/5)$ integrin expression in six patients with metastatic breast cancer.[14]
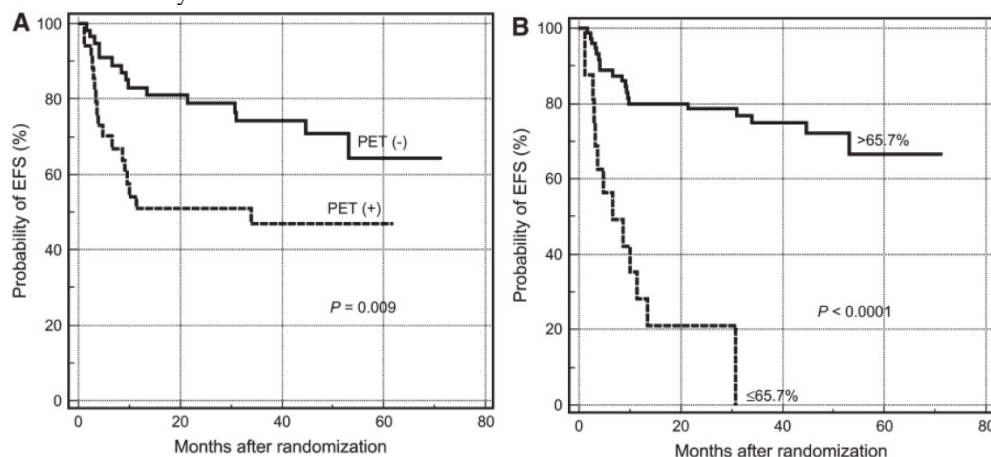

Fig 1. PET tumor quantification improves prediction of patient survival. Kaplan-Meier plots for estimating probability of event free survival (EFS) according to PET status at mid therapy. (A) Survival curves based on visual analysis. (B) Survival curves based on percentages of SUV_{max} reduction. Reprinted by permission of the Society of Nuclear Medicine from Lin et al.[8]

Measures of intra- and inter-tumoral variability of radiopharmaceutical uptake may carry information relevant to prognosis and treatment. Using a novel geometric comparison method to quantify the heterogeneity of tumor images, Eary et al. showed that intra-tumoral non-uniformity of ^{18}F-FDG uptake was a strong predictor of overall and disease-free survival in 238 patients with sarcoma, independent of patient age, sex, tumor subtype, tumor grade and SUV_{max}.[15] Inter-tumoral variability of target receptor expression in disseminated metastatic disease can strongly influence patient management. Kurland and collaborators used PET-derived measurements of ^{18}F-fluoroestradiol and ^{18}F-FDG uptake to characterize within- and between-patient variability of tumor estrogen receptor (ER) expression in 91 breast cancer patients with prior ER-positive biopsies.[16] For patients with multiple lesions, they found ^{18}F-fluoroestradiol SUV and ^{18}F-fluoroestradiol/^{18}F-FDG SUV ratios to be clustered around the patient's average value in most cases, suggesting that relative intra-patient homogeneity of ER expression is the rule rather than the exception in metastatic breast cancer.

It is reasonable to expect whole-body or "global" metabolic tumor burden as assessed with ^{18}F-FDG/PET-CT to be a better predictor of patient outcome than tumor uptake. Investigators have adopted two measures of metabolic tumor burden: metabolic tumor volume (MTV; global MTV = the sum over lesions of estimated volumes of increased uptake) and total lesion glycolysis (TLG; global TLG = the sum over lesions of the product of each lesion's MTV and the corresponding SUV_{mean} within that MTV).[17] Recently, Manohar et al. found pre-treatment MTV and TLG to be negatively correlated with progression-free survival in 51 patients who subsequently underwent chemotherapy for high-grade NHL, whereas neither SUV_{max} nor SUV_{mean} were significantly correlated with survival.[18] Liao et al. reported similar findings in 169 patients with newly diagnosed non-small cell lung cancer.[19]

As a final example of the value of tumor quantification, we note the use of PET-derived uptake measurements to predict radiation dose to tumors from internal radiotherapy. Single photon imaging of ^{111}In- and ^{131}I-labeled antibodies has routinely been used for organ/tissue dosimetry in radioimmunotherapy (RIT), whereas PET is much better suited than SPECT to support tumor dose estimation. For example, Carrasquillo and collaborators used PET to measure tumor uptake of ^{124}I-huA33 antibody in 25 patients with primary or metastatic colon cancer.[20] The measurement procedure was validated against direct assays of surgically resected tumors and could readily be used in estimating therapeutic dose in RIT with ^{131}I-huA33.

DIFFICULTIES IN TUMOR QUANTIFICATION WITH PET

Pharmacokinetics vs. Uptake. The radiotracer method is most powerful when used in conjunction with dynamic image acquisition and kinetic modeling. In a recent example, Williams and collaborators used dynamic PET and non-linear modeling to account for an observed blood glucose concentration dependence of tumor ^{18}F-FDG uptake in mice.[21] Cheebsumon et al. compared changes in tumor SUV with flux constants determined by kinetic modeling for 13 patients scanned pre- and post-therapy with ^{18}F-FDG/PET.[22] They concluded that, in some circumstances (e. g., when the blood clearance of ^{18}F-FDG varies significantly between baseline and follow-up scans), SUV analysis can yield very different results than the kinetic analysis "gold-standard".

It is generally conceded that dynamic imaging and kinetic analysis yield more and better information about tumor biology than single time point imaging and uptake measurement. Ideally, PET examinations should be interpreted from model-derived, biologically-indicative parametric images rather than images reflecting relative activity distribution.[23] Unfortunately, the kinetic approach is ill-suited for routine clinical application. It requires knowledge of tracer concentration in arterial blood over time and very intensive calculations to fit kinetic models to tissue time-activity curves and produce parametric images. Furthermore, the limited axial field of view of PET scanners (~ 20 cm) precludes large-area dynamic acquisitions. Thus, utilization of dynamic imaging/kinetic analysis is currently restricted to research applications and validation of protocols employing quasi steady-state, "semi-quantitative", measurements of tumor uptake.

Limitations Imposed by PET Instrumentation. PET scanning is based on the positron decay of certain isotopes, such as ^{18}F, ^{11}C and ^{64}Cu. A positron produced by a radioactive nucleus travels a short distance until it encounters an electron, at which point both the positron and electron disappear and their masses are converted into two 511 keV photons traveling in opposite directions. Many small detectors (usually scintillation crystals assembled as several rings) record photon pairs that interact at approximately the same time with a pair of detectors. After the scan, an image reconstruction algorithm is used to compute the tracer distribution image from the collected count data.

Several data corrections are necessary in order to obtain quantitatively accurate PET images. These include normalization (compensates for non-uniform detector sensitivity) and corrections for dead time, tissue attenuation, scattered and random coincidence events. A calibration factor converting image intensity to activity concentration is obtained by scanning a uniform phantom containing a known concentration of ^{18}F or [68]Ge. Over the last decade, standardization of the calibration procedures, as well as advances in PET instrumentation and data correction techniques, have minimized errors caused by detector non-uniformity, attenuation, scatter and random coincidences. However, the quantitative accuracy of PET images is still degraded due to the limited spatial resolution of the scanner and random counting noise in the raw data.

The spatial resolution of PET images is limited by several physical factors,[24] including positron range, photon pair non-colinearity, errors in photon localization caused by detector crystal penetration and scatter between detectors, and the finite size of the detector elements. The current achievable spatial resolution for whole body clinical scanners is about 4 mm.[25] Assuming the data are properly corrected and calibrated, the radiotracer concentration assigned to a given reconstructed voxel within a tumor image still may be biased because of two competing effects - "spill out" of activity into neighboring voxels and "spill in" of activity from neighboring voxels. The former effect reduces the measured activity concentration while the latter effect increases it. In addition, PET image voxels (dimensions usually a few mm per side) often subtend two or more different tissue types. As a result, the activity concentration in a given voxel is a weighted average of the concentrations in the constituent tissue types. Collectively, the distortions between image intensity and actual tissue activity distribution caused by limited spatial resolution and finite voxel size are referred to as "partial volume effects".

Noise in PET images is mainly caused by the finite number of photons detected. Factors limiting photon counts include the amount of activity injected, scan duration, attenuation within the body, detector sensitivity and the limited solid angle subtended by the detector array. As a result, Poisson-distributed, random noise is often the limiting factor in image quality and voxel based image analysis. In order to improve precision, activity concentration measurements are usually averaged over volumes of interest (VOIs) comprising multiple voxels. The effectiveness of this approach is, however, limited by spatial correlations among neighboring voxels.

Patient Motion. Like all radionuclear imaging, PET scanning is relatively slow and thus particularly susceptible to patient motion artifacts. While physical restraints and patient coaching can minimize problems due to voluntary movement, uncorrected cardiac motion and respiration inevitably add to image blurring in the thorax and upper abdomen. Combined PET-CT scanners were first introduced about a decade ago and now dominate in clinical practice. The coregistered CT image provides anatomical context and is used for PET attenuation correction. However, since the CT scan is very fast, there is a mismatch between the CT and PET images due to patient respiratory motion. The amount of mismatch and the effect on PET image values depend on the respiratory phase during which the CT scan is acquired. The mismatch causes error in the attenuation correction, which may degrade image quality and cause bias in the PET quantification. In a study of PET-CT images from 5 patients with lung lesions, Erdi and collaborators showed that SUV_{max} had a large dependence (up to 24%) on respiratory phase when CT images from different phases were used for attenuation correction.[26] Large changes in tumor location and size were also observed between PET images acquired at end of inhalation vs. end of exhalation. These problems can be mitigated by respiratory gating of PET data acquisition.

Discordance Between Tumor Uptake and Tumor Anatomy. Given the advent of combined PET-CT scanners, it might be supposed that tumor VOIs for SUV and tumor burden analysis would now be defined from coregistered anatomic images, and some investigators have used that approach.[27] However, tumor boundaries are often difficult to discern on CT, especially without the use of contrast material, which can interfere with CT-based attenuation corrections for PET.[28] Patient respiratory motion and movement between the CT and PET portions of a PET-CT examination can also limit the usefulness of CT-derived VOIs. Furthermore, metabolically active tumor is frequently confined to a sub-portion of the anatomic tumor, and, as previously noted, metabolic tumor volume can be a useful prognostic indicator. These factors have engendered an extensive effort to develop methods for estimating the 3-dimensional boundaries of intra-tumoral regions of radiotracer accumulation in PET images.

Standardization of PET Scan Acquisitions and Data Analysis. PET quantification is affected by many factors, as described in the section "Limitations Imposed by PET Instrumentation". In order to get meaningful results, variability must be minimized by implementing standard patient preparation, scanner quality control, scan acquisition, image processing and image analysis procedures.[29] It is also necessary to understand the variability caused by non-biological

effects, so that changes in SUV caused by patient biology can be distinguished from those introduced purely by instrumentation, statistical noise and observer subjectivity.

Multicenter studies suffer from variations in SUV measurement caused by inter-institutional differences in quality control, scanning and data analysis procedures. A PET phantom study conducted among 10 centers showed 10%-25% SUV variability of this type.[30] PET centers need to enforce a minimum performance standard for all instruments affecting quantification. This includes not only the PET scanner, but also ancillary equipment such as clocks, dose calibrators and injection devices.

PET scan protocols should be standardized to reduce systematic and statistical variability. This includes injected activity, time from injection to scan ("uptake time"), scan duration, and plasma glucose level. Figure 2 illustrates the significant differences in SUV values that may be caused by variations in uptake time.
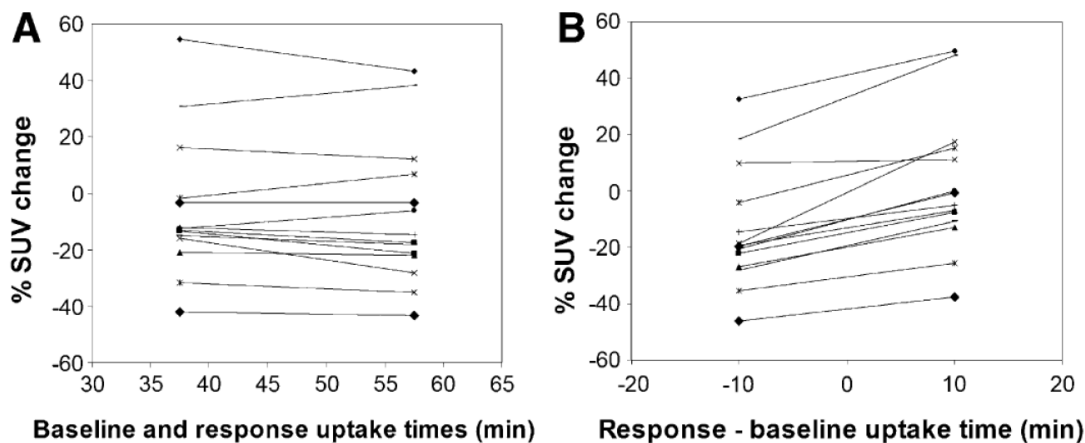

Fig 2. Effect of uptake time on measured changes in [18]**F-FDG SUV between pre-therapy baseline and intra-therapy follow-up scans.** Data were taken from 60 min dynamic ^{18}F-FDG PET studies. Each symbol and line represents data from a single subject. (A) Tracer uptake times were equal in response and baseline studies. (B) 10-min mismatch in uptake time between response and baseline studies. In B, baseline uptake period was set at 48 min, and response uptake time was set at 38 min or 58 min. When the uptake times differed by 10 min, there were significant differences in observed responses (paired t test, P<0.0001). Reprinted by permission of the Society of Nuclear Medicine from Boellaard et al.[29]

SUV AND ITS LIMITATIONS

Uptake as a Surrogate for Biologic Processes. Tumor activity concentration at time t after systemic injection may be a function of many factors including, at minimum, the arterial time course of the radiotracer, tumor perfusive blood flow, the blood and interstitial volumes of the tumor, and trans-vascular transport mechanisms. If the radiotracer targets a receptor species, uptake will depend on the radiotracer-receptor binding kinetics and may also be affected by the kinetics of radiolabel release following cellular internalization and metabolism of the receptor-radiotracer combination. For metabolic substrates such as ^{18}F-FDG, uptake is influenced by cell transport mechanisms and intracellular metabolism. If the radiotracer produces labeled, recirculating metabolites, the presence of those metabolites in tumor during the scan will also affect PET-derived measurements of tumor uptake.

The ideal radiotracer does not produce recirculating, labeled metabolites and is retained ("trapped") in tumor long-term relative to the time from injection to completion of the PET scan via a mechanism specific to the tracer's molecular target. In that case, at time t sufficiently long that freely-exchanging radiotracer has come to a quasi-equilibrium between blood and tumor, the activity concentration A in tumor can be expressed as[31]

$$A(t) = K_i \int_0^t C_B(\tau)\,d\tau + V_D C_B(t) \qquad \dots(1)$$

where K_i is referred to as the "flux constant" for radiotracer incorporation into the tumor, $C_B(t)$ is the activity concentration in arterial blood, and V_D is the distribution volume for free radiotracer within tumor. Under these conditions, the kinetics of tumor uptake can be described in terms of a 3-compartment model (blood, free radiotracer in tissue, and trapped radiotracer), from which the flux constant can be expressed as

$$K_i = \frac{K_1 k_3}{k_2 + k_3} \qquad \dots(2)$$

where K_1, k_2 and k_3 are compartmental model parameters representing, respectively, the product of tumor perfusive blood flow and blood-to-tumor ex-

traction fraction, the rate coefficient for transport from the tumor parenchymal space to blood, and the rate coefficient for the trapping mechanism. Note that $k_3/(k_2 + k_3)$ is the probability that a tracer molecule extracted from blood into tumor will be trapped. Note also that transport (K_1) and trapping (k_3) of the radiotracer may be affected by competition from natural substrates. From Eqs. 1 and 2, tumor SUV (normalized to body weight W) can be expressed as

$$SUV(t) = A(t)\frac{W}{D_{inj}(t)} = \left[\frac{K_1 k_3}{k_2 + k_3}\int_0^t C_B(\tau)d\tau + V_D C_B(t)\right]\frac{W}{D_{inj}(t)} \quad \ldots(3)$$

where $D_{inj}(t)$ is injected activity decayed to time t. From Eq. 3, it can be seen that in order for radiotracer uptake to be indicative of its molecular target (i. e., the trapping mechanism), k_3 cannot be $\gg k_2$, i. e., blood-to-tissue transport cannot be rate limiting for tumor uptake. If transport is not rate limiting, Eq. 3 can be used to articulate the conditions under which SUV may be an adequate substitute for kinetic analysis.

The objective of SUV analysis is to discern differences in molecular target expression and/or function in at least three different circumstances: (i) different tumors within the same patient; (ii) tumors in different patients; and (iii) the same tumor in different scans. For Circumstance i, it is clearly helpful for trapping to be sufficiently rapid and/or scan acquisition to be sufficiently long after injection that most of the tumor activity reflects trapped radiolabel, i.e.,

$$SUV(t) = \frac{K_1 k_3}{k_2 + k_3}\int_0^t C_B(\tau)d\tau \frac{W}{D_{inj}(t)} \quad \ldots(4)$$

Since the integral blood curve is common to all tumors within a given scan, SUV differences among tumors are due to differences in tumor kinetics, and sensitivity to the molecular target (i. e., k_3) is maximal. Under Circumstance ii (different patients), SUV comparisons may be confounded by differences in radiotracer blood clearance as manifested in the integral blood curve. Here it may be helpful to normalize tumor SUV to the SUV of a reference tissue or organ which also traps the radiotracer and thus for which SUV is also proportional to the integral blood curve.[3] For Circumstance iii (change in tumor SUV between scans 1 and 2), the parameter of interest is usually % change in SUV = 100·(SUV$_2$/SUV$_1$ – 1). Thus, as for Circumstance ii, it is necessary to account for potential differences in the integral blood curve between the two scans, and normalization within each scan to the SUV of a suitable reference organ may be helpful.

In principle, full kinetic analysis (compartmental modeling), which provides direct measurements of k_3, is always preferable to "semiquantitative" SUV analysis, which at best yields values that are proportional to flux constants. An intermediate approach is graphical methods such as Patlak analysis, which utilizes a linearized form of Eq 1:

$$\frac{A(t)}{C_B(t)} = K_i\frac{\int_0^t C_B(\tau)d\tau}{C_B(t)} + V_D \quad \ldots(5)$$

This type of analysis specifically accounts for blood clearance variations and the presence of unbound radiotracer, but requires measurements of arterial blood curves. Like SUV analysis, the relationship of Patlak analysis to the molecular target may be confounded by variations in tumor blood flow and transport (K_1 and k_2).

SUV has been compared with full kinetic analysis (FKA) most frequently for ^{18}F-FDG, for which the trapping mechanism is hexokinase-mediated monophosphorylation.[32] In an investigation of 40 patients with colon cancer metastatic to liver, Graham et al. found that, although strongly correlated with FKA, SUV was far more variable compared with K_i as determined by Patlak analysis.[31] Nonetheless, SUV was still highly effective in predicting patient survival. Krak and collaborators evaluated various analysis methods in 20 women with breast cancer scanned before and during chemotherapy.[33] They found SUV to be as well correlated as Patlak analysis with baseline FKA (r = 0.96 compared with 0.98), but less accurate in detecting changes in ^{18}F-FDG kinetics during therapy. Cheebsumon et al. found good concordance in baseline – therapy % changes between SUV and FKA in 7 patients (18 tumors) with lung cancer, while 6 patients (9 tumors) with gastrointestinal cancers showed large differences between SUV and FKA in measured % change.[22]

Normalization for body size. Other things being equal, tissue uptake expressed as a percentage of injected activity is inversely related to body size, and SUV analysis is meant to take that into account. Commonly used methods of size normalization include body weight (SUV$_{bw}$), ideal body weight (SUV$_{ibw}$), lean body mass (SUV$_{lbm}$) and body surface area (SUV$_{bsa}$). These normalization methods are defined as:[34]

$$\text{Ideal body weight} = 45.5 + 0.91 \cdot (\text{height} - 152)$$
$$\text{Lean body mass} = 1.07 \cdot (\text{weight}) - 148 \cdot (\text{weight/height})$$
$$\text{Body surface area} = \text{weight}^{0.425} \cdot \text{height}^{0.725} \cdot 0.007184 \quad \ldots(6)$$

where the units of weight, height and area are kg, cm and m^2, respectively.

Note that ideal body weight is a function of

height only, while lean body mass and body surface area depend on both weight and height. It has been found that SUV_{bw} is positively correlated with the body weight due to the higher percentage of fat tissue in heavy patients.[35, 36] (Fat has less ^{18}F-FDG uptake than other tissues when the patient is fasted, as in most clinical scans.) Sugawara et al. compared the alternative forms of SUV in 138 breast cancer patients.[34] They found that SUV_{bw} was positively correlated with weight, SUV_{ibw} was negatively correlated with weight, and neither SUV_{ibm} nor SUV_{bsa} were correlated with weight. Thus, it appears that SUV_{ibm} and SUV_{bsa} are more robust measurements. However, for a single patient with stable weight, percentage change in SUV is unaffected by normalization technique.

Methods based on the maximum voxel. The most widely used measures of tumor uptake have been SUV_{max}, the SUV value of the maximum intensity voxel within a defined subspace of the PET image matrix, and $SUV_{mean}(MVBT)$, the average SUV within a volume containing the maximum voxel and voxels with intensities ≥ a user-specified percentage of SUV_{max}. The choice of threshold for this maximum voxel-based thresholding (MVBT) technique may take account of adjacent background intensities; this is termed "adaptive thresholding".[37] SUV_{max} measurement is quick and highly reproducible. For small tumors, SUV_{max} provides partial compensation for underestimation due to the partial volume effect. Weaknesses of SUV_{max} include a noise-related, systematic positive bias that also depends on tumor size,[38] and a high susceptibility to precision loss due to random noise compared with other measures of tumor uptake.[39] Regardless of segmentation technique, $SUV_{mean}(MVBT)$ has better reproducibility than SUV_{max}.[40, 41] However, the bias and noise susceptibility of SUV_{max} carry over to $SUV_{mean}(MVBT)$.[3]

PERCIST. Wahl and collaborators proposed a comprehensive procedure for acquiring, analyzing and categorizing response from ^{18}F-FDG/PET-CT scans used in evaluating therapy response in patients with solid tumors.[3] The method is called PET Response Criteria in Solid Tumors (PERCIST). Regarding factors that affect tumor quantification, the PERCIST method prescribes an upper limit for patient serum glucose concentration, restricts variability of injected activity and uptake time, and requires CT-based attenuation correction for PET. Tumor uptake is measured in terms of SUV_{peak}, defined as average activity concentration within a 1 cm^3 spherical VOI centered on the "hottest focus" within the tumor image multiplied by the ratio of lean body mass (LBM) to injected activity decayed to time of scan. SUV_{peak} typically includes the maximum voxel, but is not necessarily centered on it. PERCIST, as currently defined, does not include partial volume correction, but recommends evaluating only tumors that are 2 cm or larger. Up to 5 tumors per scan are included in the quantitative assessment. An *ad hoc* compensation for statistical noise and variations in the integral blood curve is made by comparing tumor uptake with liver uptake and the amount of noise in the liver image. In order to be included, tumors must have $SUV_{peak}(LBM)$ ≥ 1.5 x liver $SUV_{mean}(LBM)$ + 2 SD, where $SUV_{mean}(LBM)$ and SD (the standard deviation of SUV(LBM) over voxels included in the VOI) are determined within a 3-cm-diameter VOI in the right hepatic lobe.

PERCIST has been widely adopted as the new standard for PET-based therapy response assessment. In one recent evaluation, Yanagawa et al. reported on the use of PERCIST to evaluate therapeutic response and prognosis in 51 patients with esophageal cancer.[10] Univariate analysis showed lymphatic invasion, venous invasion, resection level, pathologic stage and PERCIST, but not the standard RECIST CT-tumor size response criteria, to be significantly correlated with survival. In multivariate analysis, only venous invasion, resection level and PERCIST were significant independent predictors.

A weakness of PERCIST is its vagueness regarding VOI placement for SUV_{peak} evaluation. A recent report by Vanderhoek and collaborators suggests the resulting latitude in VOI definition may introduce variability in PERCIST response assessments.[42]

SUV Repeatability. SUV measurements are affected by various factors that cannot be entirely controlled. Knowledge of the resulting uncertainty is essential when SUV is used to identify malignancy,[6] differentiate between different cancer types,[7] predict patient outcome,[8] and especially when using SUV changes to assess therapy response.[3] There have been a number of empirical test-retest studies of the repeatability of ^{18}F-FDG tumor uptake measurements in humans. De Langen et al. recently reported on a meta-analysis of 5 repeatability studies that included various types of cancer and comprised repeat scans under nominally identical conditions performed at intervals of 1 day to 1 week.[41] Tumor segmentation was done by the MVBT technique using a threshold = 50% of SUV_{max}. Data from the study are shown in Figure 3. Percentage variability was inversely related to magnitude for both SUV_{max} and $SUV_{mean}(MVBT)$. $SUV_{mean}(MVBT)$ was found to be less variable than SUV_{max}, i. e., the 95% confidence limit (coefficient of repeatability, CR95 = 1.96 · SD) for real biologic change ranged from 37% to 21% between SUV_{max} = 2 and 20 and from 32% to 12% between SUV_{mean} = 2 and 16.
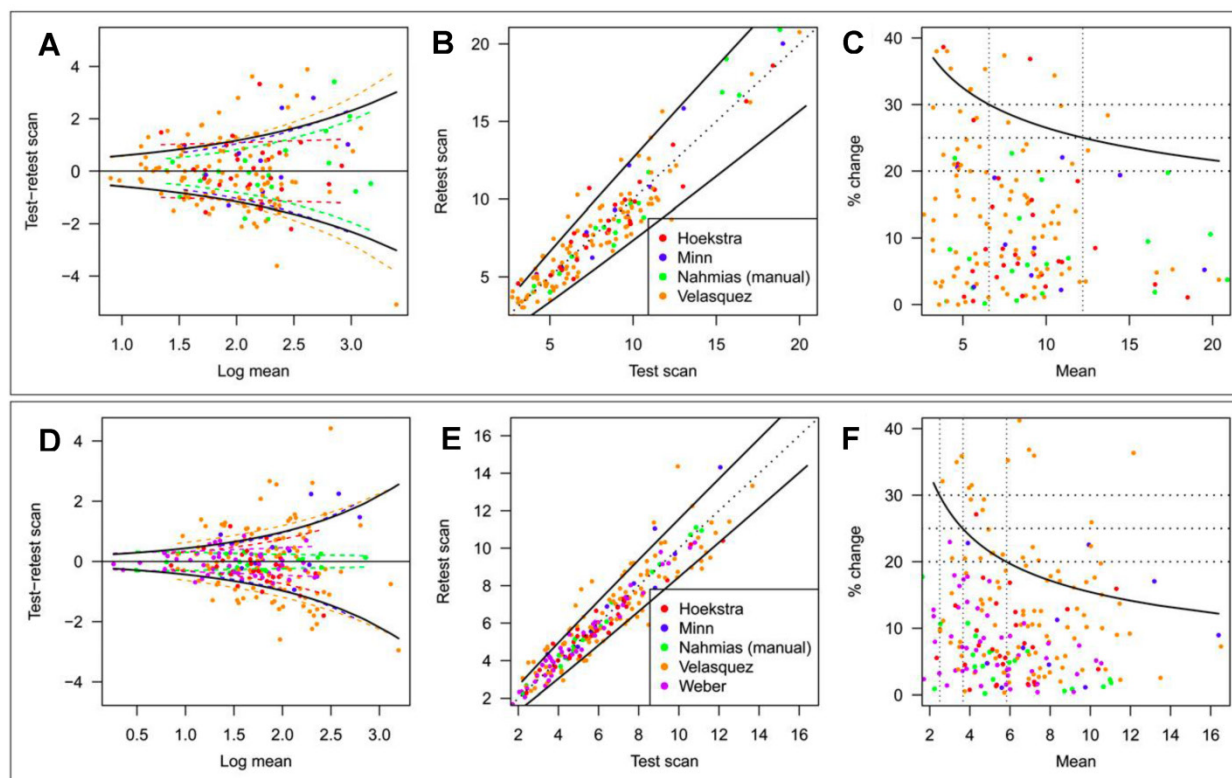
Fig 3. Repeatability of SUV$_{max}$ and SUV$_{mean}$(MVBT). Threshold for SUV$_{mean}$ = 50% of SUV$_{max}$. Estimated study-specific SD (colored dashed lines; study as a fixed effect) and overall SD (black solid lines; study as a random effect) of SUV$_{max}$ (A) and SUV$_{mean}$ (D). Test and retest scan values of SUV$_{max}$ (B) and SUV$_{mean}$ (E) plotted on original scale. Solid line is coefficient of repeatability (CR95). Relationship between CR95, expressed as percentage change, and level of SUV$_{max}$ (C) and SUV$_{mean}$ (F). Reprinted by permission of the Society of Nuclear Medicine from de Langen et al.[41]

NEW METHODS

Partial Volume Effect: Correction and Minimization. Even with accurate calibration, image-derived radiotracer concentration may differ from its true value because of the partial volume effect (PVE). For "hot" tumors, the effects of the PVE include reduced apparent tumor uptake and increased apparent tumor size. The magnitude of the effect depends on several factors, such as the size and shape of the tumor and the activity distribution in adjacent tissue. Thus, as the tumor size and shape change, the apparent uptake will change due to the PVE, even when the actual tumor uptake does not change.

Partial volume correction (PVC) methods can be divided into two categories: regional techniques and voxel-wise minimization techniques. For regional techniques, the image is divided into compartments (e. g., tumor and background), and the uptake in each compartment is assumed to be uniform. In the simplest of the regional techniques, recovery coefficients (RCs = measured values/true values) are determined from phantom scans employing fillable spheres of varying diameters and a range of different sphere-to-background activity concentration ratios. The RCs are then applied to the patient scan, taking account of tumor size and tumor:background contrast.[43] This approach assumes the tumor uptake is homogeneous and tumor metabolic size is known. In practice, it is not easy to estimate the true metabolic size of the tumor, and the shape of the tumor may be irregular, making it difficult to identify the appropriate RC value. An extension of the RC correction technique is the geometric transfer matrix (GTM) method developed by Rousset et al.[44] In GTM, the tumor image and its surroundings are divided into multiple regions, and the regional spread function (RSF) of each is calculated based on knowledge of the scanner point spread function (PSF). The set of RSF values is then used to compute a GTM for tumor plus background. The true concentrations of the regions are derived by inverting the GTM and multiplying by the measured regional activity concentrations. This method has been validated for PET brain studies using simulation and phantom data.[44, 45] For tumor imaging, the limitations of GTM include the homogeneity assumption and difficulties in registering the CT or magnetic resonance (MR) images with PET (e.g., due to patient motion). In addition, the anatomical and metabolic boundaries may not match, as seen e.g. in necrotic tumor. As a result, GTM has not been widely used in tumor studies.

Voxel-wise minimization seeks to recover higher-resolution images from blurred images or raw data. There are basically two approaches: image deconvolution and model-based image reconstruction. Deconvolution can significantly improve image resolution[46] and is attractive because it can readily be applied to reconstructed images. The only information required is the PSF of the reconstructed image, which is usually assumed to be Gaussian and is quantified from phantom scans.[47, 48] No anatomical or metabolic volume information is needed, and tumor uptake can be heterogeneous. Two deconvolution methods widely used in medical imaging are Richardson-Lucy [49, 50] and Van Cittert.[51] Using phantom and patient data, several groups have demonstrated that deconvolution methods can improve the accuracy of SUV without affecting test-retest variability.[51, 52] One potential problem is that PET image resolution is not uniform across the field-of-view (FOV). However for clinical images, the resolution is mostly determined by post-reconstruction filtering and other parameters such as number of iterations and number of subsets. As a result, for current PET scanners and reconstruction methods, it seems sufficient to use a spatially invariant PSF.

The other voxel-wise minimization method employs model-based reconstruction algorithms that incorporate the PSF in the forward model.[24, 53, 54] There are two approaches - one which models the PSF in the sinogram [24, 53] and the other in the image space.[54] The sinogram-based PSF is more accurate but requires more computation. As with deconvolution methods, a shift-invariant, image-based PSF is commonly used. Recently Lasnon et al. found that sinogram-based PSF reconstruction improved lesion detection sensitivity in nodal staging of non-small cell lung cancer (97% vs. 78% with non-PSF method).[55] Image-based PSF reconstruction has been shown to provide up to two-fold better precision in tumor SUV measurements than GTM or image deconvolution-based PVC (Figure 4).[51]

Both image deconvolution and PSF reconstruction tend to amplify image noise, which may increase SUV_{max}. In addition, these methods create image edge artifacts,[56] which may cause positive, tumor size-dependent bias in uptake measurements.[57, 58] In the study by Lasnon et al., detection specificity was lower for PSF than non-PSF reconstruction (58% vs. 71%), though the difference was not statistically significant. This decrease of specificity was caused by an apparent increase of SUV in disease-free nodes.[55]

Another interesting method for resolution recovery is to use coregistered, high resolution anatomical images (CT or MR) images to constrain PET image reconstruction. This approach is capable of improving the PET resolution without introducing edge artifacts.[59] The application of these methods to tumor imaging has been limited by a high computational cost as well as the difficulty of getting good spatially and temporally co-registered anatomical and functional data.

It is also worth pointing out that complete resolution recovery cannot be achieved in either the deconvolution or voxel-based approaches. This is due to the resolution and noise limitations of PET raw data.[24]
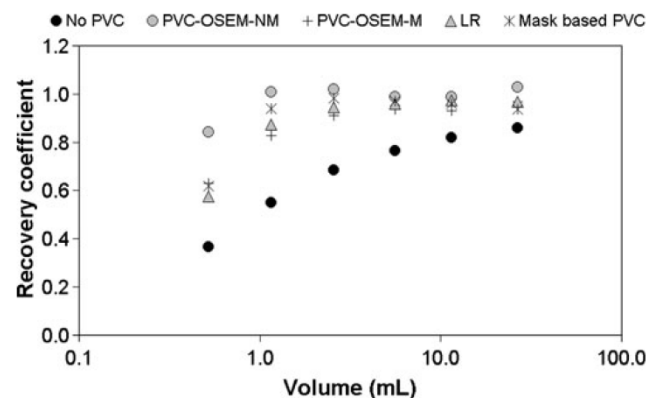

Fig 4. Image reconstruction-based (OSEM-NM, OSEM-M) vs. image deconvolution-based (△LR) and GTM (Mask-based) partial volume correction. Shown are recovery coefficients for six hot spheres positioned in the NEMA NU2 image quality phantom. Reprinted with permission from Hoetjes et al.[51]

Metabolic Tumor Volume and Corresponding SUV_{mean}. Tumor uptake of ^{18}F-FDG is generally regarded to be a measure of tumor viable cell density.[3] Heterogeneity of uptake may be manifested in PET images of lesions with diameters larger than about 2.5 x image full width at half maximum (FWHM), and portions of the anatomic tumor volume may appear not to accumulate the radiotracer. Thus there is interest in defining the boundary or boundaries of the "metabolically active" subregion(s) of a tumor for the purpose both of measuring the viable tumor volume and obtaining its SUV_{mean}. Segmentation of the metabolic tumor volume (MTV) has most often been done manually by visual inspection, or "semi-automatically" within a user-defined volume using MVBT or background-adaptive MVBT techniques.[37] Manual segmentation is highly operator dependent, and techniques based on the maximum voxel reflect the statistical uncertainty and tumor size-dependent biases of SUV_{max}.[38] Due to the lack of robust segmentation techniques, PERCIST recommends that MTV be defined as the combination of voxels with SUV(LBM) ≥ liver SUV_{mean}(LBM) + 3 liver SDs within a 3 cm diameter ROI placed around the

hottest portion of the tumor image.[3]

Various tumor segmentation approaches are being developed which do not depend on the maximum voxel. These techniques are mathematically sophisticated and use information from the tumor-background interface region of the PET image to estimate the tumor boundary. Perhaps the most conceptually straightforward of the new segmentation techniques estimates a gradient image in the tumor-background interface region and then equates the tumor boundary with the locus of the maximum gradient.[60] Li et al. developed a locally-adaptive region-growing technique that exploits the rapid increase of VOI size when the growth process crosses the tumor-background interface.[61] Hofheinz and collaborators recently reported on an iterative technique that combines local background correction and adaptive thresholding in the boundary region.[62] Another approach employs fuzzy logic and spatial context to identify boundaries comprising the voxels most likely to lie on the tumor-background boundary.[63, 64] Among the new techniques, fuzzy logic appears to be the most robust for small tumors and tumors with heterogeneous uptake.[65, 66] All of the segmentation techniques cited here are reasonably effective for isolated tumors with high contrast relative to adjacent tissue. However, there has been relatively little testing of the newer techniques for geometries in which the segmented tumor lies close to other structures (tumor or non-tumor) with high uptake.

Several studies have examined the repeatability of ^{18}F-FDG based MTV measurements in clinical applications. Frings et al. used background-adaptive and non-adaptive MVBT techniques to analyze images of 34 lesions in 11 patients with non-small cell lung cancer.[67] For the best technique (background-adaptive 50% threshold) and tumors with average diameter $\geq$ 2.0 cm, the repeatability coefficient for MTV was 37%. Heijmen and collaborators compared adaptive MVBT and fuzzy locally adaptive Bayesian (FLAB) segmentation techniques in scans of 20 patients with hepatic metastases from colon cancer.[68] Repeatability coefficients were 45% and > 85% for FLAB and MVBT, respectively, suggesting the newer FLAB technique may provide substantial improvement in MTV precision.

The new segmentation techniques of course engender corresponding measurements of SUV_{mean}. However, accurate segmentation of the MTV does not guarantee accurate estimation of its mean SUV. For tumors that appear hot relative to their immediate surroundings, the PVE causes image intensity to underestimate tumor activity concentration inside and near the boundary of the active region. Thus, the accuracy of SUV_{mean} is inversely related to image

FWHM. Bhatt et al. recently reported on a technique that combines segmentation by iterative histogram thresholding with partial volume correction by iterative deconvolution.[69] Another, simpler approach for tumors with images that appear homogeneous except for the PVE is to erode the VOI boundary within the segmented boundary by a distance based on image resolution.[70, 71]

Metabolic Tumor Burden. Given that ^{18}F-FDG uptake reflects tumor viable cell density, the product $SUV_{mean}(^{18}F\text{-FDG}) \cdot MTV$ for a given lesion is a measure of the total number of viable cells in that tumor and is termed "total lesion glycolysis" (TLG).[17] Global TLG, i. e., the sum of TLG values for individual tumors, is thus a measure of whole-body viable or "metabolic" tumor burden. Some studies have suggested that both global MTV and global TLG may be better predictors of patient survival than is tumor uptake. In a recent example, Muralidharan et al. used a locally-adaptive region-growing segmentation technique to evaluate colorectal hepatic metastases in 30 patients.[72] Both MTV and TLG were significantly (negatively) correlated with patient recurrence-free and overall survival (OS), while SUV_{max} and SUV_{mean} were not. Liao et al. compared pre-treatment ^{18}F-FDG/PET-CT with OS for 169 patients with inoperable non-small cell lung cancer.[19] Tumors were segmented using the gradient technique. Global MTV, TLG and SUV_{max} (i. e., the highest-valued single voxel for all tumors) were statistically and comparably associated with overall survival, while global SUV_{mean} was not significantly correlated with OS. Manohar and colleagues evaluated MTV and TLG according to the PERCIST prescription in 51 patients with NHL.[18] Both MTV and TLG showed weak but significant correlations with progression-free and overall survival, while neither SUV_{max} nor SUV_{mean} were significantly correlated with survival.

It may be noted that, in principle, anatomic tumor segmentation can provide measurements of TLG that are equivalent to those obtained with metabolic tumor segmentation. The MTV occupies a subvolume of, or the same volume as the anatomic tumor volume (ATV). Total activity within the MTV and ATV is the same. Therefore, $TLG = MTV \cdot SUV_{mean}(MTV) = ATV \cdot SUV_{mean}(ATV)$. The development of MTV segmentation has been driven by the inadequacies of CT-derived anatomic tumor volumes. MR imaging is superior to CT for visualization of tumor boundaries within regions of soft tissue. When tumor is visualized in high contrast to its surroundings, anatomic segmentation is far simpler, and presumably more accurate and precise, than metabolic segmentation. Therefore, it is reasonable to anticipate that, with the advent of PET-MR scanners, anatomic tumor seg-

mentation will find a role in the measurement of metabolic tumor burden.

Quantification of Intra-Tumoral Uptake Heterogeneity. Intra-tumoral uptake of ^{18}F-FDG is, in general, not uniform, and the nonuniformity may change as the tumor progresses. For example, in early stage sarcoma, tumors tend to have highly uniform uptake, while in late stage sarcoma, uptake is characteristically reduced in the central region relative to the tumor boundary. Intra-tumoral heterogeneity can be caused by cell proliferation, blood flow, hypoxia and necrosis.[73]

Quantification of intra-tumoral heterogeneity is useful in monitoring and predicting response to therapy. Several methods have been proposed, including shape analysis, texture feature extraction and cumulative SUV-volume histogram (CSH).[73-75] A simple ellipsoidal shape model was applied to human sarcoma by Eary et al.,[15] where the deviation of the tumor uptake pattern from the model was used as a measure of heterogeneity. Later, the same investigators elaborated their approach in the form of an empirical tubular model with varying uptake from core to boundary.[74] Texture features (global, regional and local) can be extracted from PET images.[75] The regional and local features, such as coarseness and local entropy, are useful in predicting response to therapy. In the CSH method, the percentage of total tumor volume above a specified percentage of SUV_{max} is plotted, and the area under the CSH curve (AUC) is used as a measure of heterogeneity. Higher AUC corresponds to a lower degree of heterogeneity (Figure 5). Note that the PVE and image noise both contribute to the heterogeneity of tumor images, so PVC and/or noise reduction may be necessary prior to heterogeneity analysis.

Several simulation and patient studies have demonstrated that measurement of intratumoral heterogeneity is effective in predicting treatment response in lung cancer,[73] sarcoma[74] and esophageal cancer. [75] Tixier et al. found pre-treatment heterogeneity measurements to be more sensitive than SUV_{max} or SUV_{mean} in differentiating subsequent non-, partial- and complete-responders among 41 patients with esophageal cancer.[75] The ability to accurately predict response from baseline FDG scans avoids potential errors in the interpretation of intra- or post-therapy ^{18}F-FDG scans due to therapy-induced inflammation.[75]

Accounting for Random Image Noise. As previously noted, knowledge of the repeatability or precision of PET-derived measurements of tumor SUV, MTV and TLG is highly important in the application of those parameters to patient management. Factors contributing to the uncertainty in SUV measurements include, at minimum: (i) short-term variations in patient physiology and metabolism; (ii) imprecision in measurements of patient weight and height, as well as injected activity; (iii) patient motion; (iv) the inherent randomness of radioactive decay and photon detection as amplified through various corrections to raw data and the image reconstruction process; and (v) voxel size and inter-voxel correlations. Measurements of SUV_{mean} are also affected by random variability in tumor segmentation, which is the source of uncertainty in MTV measurements. Test-retest studies, such as those previously cited for ^{18}F-FDG/PET-CT, are useful for generalized characterizations of intra-patient repeatability for a given radiopharmaceutical and scan procedure. However, the utility of PET quantification might be greatly enhanced if uncertainty could be estimated for individual SUV and MTV measurements.
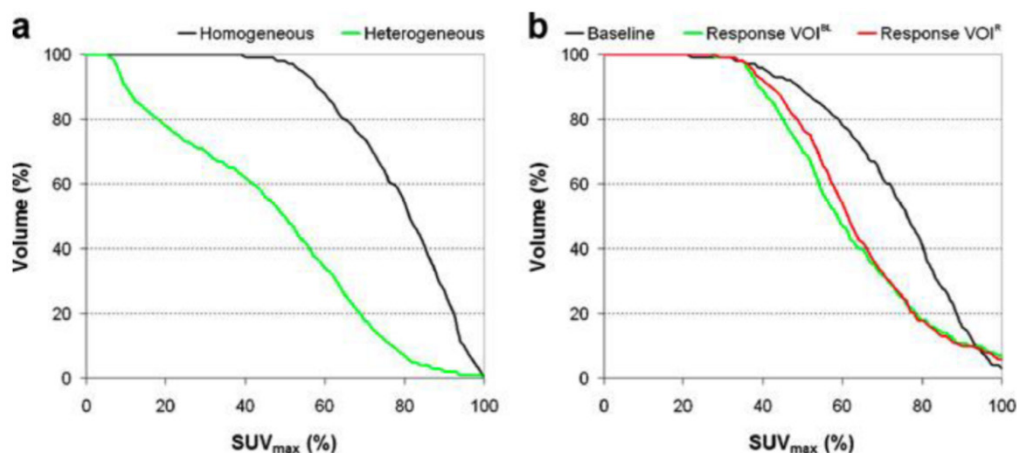

Fig 5. Quantification of intra-tumoral uptake heterogeneity. Cumulative SUV-volume histograms for a diagnostic lung tumor study (a) and lung tumor response study (b). VOIs of response scans were either defined on the baseline scan (VOI^{BL}) or on the response scan (VOI^{R}). Reprinted with permission from van Velden et al.[73]

It is, in fact, possible to calculate exact count- and image reconstruction/image matrix-related SUV uncertainties (items iv and v above) from the raw count data when images are reconstructed by the inherently-linear filtered back projection (FBP) method.[76] Such calculations are impractical on a routine basis, and Carson et al. derived an approximation technique for estimating SUV variances that requires only knowledge of average single-voxel variance and inter-voxel covariance within the VOI.[77]

FBP has been superseded in current practice by the iterative, non-linear, expectation maximization (EM) method, for which analytic calculation of SUV statistical noise is not feasible.[71] In the absence of such calculations, empirical estimates may be useful. For example, PERCIST requires that a tumor have $SUV_{peak}(LBM) \geq 1.5 \cdot$ liver $SUV_{mean}(LBM) + 2$ SD in order to be used in the assessment of treatment response. Inclusion of measured liver image noise (i. e., single voxel SD) in the formula is meant to compensate for variations in tumor image noise due to differences in factors common to liver and tumors (i. e., scanner sensitivity, patient size, injected activity, time from injection to scan, scan duration per bed position, and systemic blood clearance of ^{18}F-FDG). The PERCIST prescription does not, however, compensate for variations in tumor image noise due to differences in tumor uptake.

Local noise variance in EM-reconstructed images is proportional to the expected value of local image intensity (i. e., the average of an infinite number of repeat measurements under identical conditions).[78] This suggests it may be possible to estimate count- and image reconstruction/image matrix-related contributions to tumor SUV uncertainty from tumor image intensity and the image intensity and noise characteristics of other, relatively large and uniform "background" regions within the PET image set, such as liver. The fundamental requirement of the approach is that the repeatability of SUV_{mean} measurements in tumors be reliably related to the repeatability of SUV_{mean} measurements in multiple VOIs placed within the background region. If that were true, the variance of tumor SUV_{mean} could be expressed in terms of single voxel variance and inter-voxel covariance for the background region.[79] Bading et al. used this approach to model the bias and noise characteristics of SUV_{max} in a region of uniform activity concentration.[80] It may be possible to extend the technique to tumors via the relationship between local image intensity and noise variance that characterizes EM image reconstruction.

New PET Imaging Technologies and Their Effects on Tumor Quantification. PET technology has greatly improved during the last decade. Major developments include new radiation detection methods, better data correction methods, new image reconstruction algorithms and hybrid imaging techniques. Current PET scanners are usually combined in tandem with CT scanners, and PET-MR is also becoming available. Most new clinical PET scanner models developed in the last five years use time-of-flight (TOF) information to improve image signal-to-noise ratio (SNR). As a result, PET images today have higher spatial resolution, less noise and better quantitative accuracy than previously. It is important to understand how these new technologies affect tumor quantification in order to properly interpret the results and avoid pitfalls.

Although TOF-PET was first introduced in the 1980s,[81] commercial TOF-PET scanners have become available only in the last few years.[25, 82] With faster detector materials and electronics, timing resolution is now 300-500 pico seconds, which corresponds to location uncertainty of 4.5-7.5 cm. The benefits of TOF, which increase with patient size, include better contrast recovery, higher SNR and faster reconstruction convergence.[83] TOF images have better lesion detection performance than conventional PET images[84] and produce more accurate and more reproducible SUV values.[85] It has also been shown that cumulative benefit is obtained by combining TOF with PSF modeling.[84]

Another recent change in clinical PET is the widening use of iterative reconstruction algorithms with PSF modeling of intrinsic scanner response.[24, 53] It has been demonstrated in phantom and patient studies that PSF modeling improves image resolution and noise properties,[53, 79] which translates to better tumor detection.[84] One pitfall, however, is that PSF modeling introduces edge artifacts into the images,[56] which tend to add positive bias to SUV values.[57, 58] For tumor size between 1 and 2 cm, the error in SUV_{max} can be as much as 40%.[57] Possible solutions to this problem include filtering[58] and under-modeling the system by using a narrower-than-actual PSF.[56, 86]

The combination of PET with MR promises to have a major impact on the utility of PET tumor quantification, as well as its accuracy and precision. Biologic information can now be obtained simultaneously via the tracer-PET method and functional MR (diffusion, blood flow, spectroscopy),[87] a development that promises to greatly expand the use of quantitative imaging for both modalities. The improvement in soft tissue contrast provided by MR will also impact the quality and analysis of PET images. The co-registered anatomical information from MR can be incorporated in PET Image reconstruction to improve image quality and reduce the PVE,[59] thus improving the accuracy and precision of SUV measurements.

Another benefit of PET-MR is better delineation of tumor boundaries, which enhances the potential utility of anatomic, whole-tumor average SUV and tumor functional burden measurements.

DISCUSSION AND CONCLUSIONS

Current Status of Tumor Quantification. The casual use of PET-derived measurements of tumor uptake was correctly criticized early in the history of the field.[88] Since then, a great deal of progress has been made in understanding the limitations, as well as the advantages, of such measurements. The fidelity of PET images to the actual activity distribution within the patient has improved with more accurate corrections to raw count data and better methods of image reconstruction, including PSF modeling and inclusion of TOF data. As a result, image resolution has improved, thus reducing quantification errors due to the PVE. Image noise was reduced when single photon counting and calculation replaced delayed measurement and subtraction as the method of choice for random coincidence noise correction. Scanner sensitivity greatly increased with the development of accurate methods for estimating scatter noise, which obviated the need for inter-detector shielding and thus permitted 3-D data acquisition. However, in clinical practice this improvement has primarily been used to reduce scan duration rather than image noise. Tumor detection and PET attenuation correction have improved with PET-CT, but CT is not routinely used for tumor segmentation due to its low soft tissue contrast. Thus there is current emphasis on the development of techniques for segmenting tumor metabolic volume, a situation that may reverse with the advent of PET-MR. Sophisticated software tools have been developed for PVC, tumor segmentation and parameterization of uptake heterogeneity, but these are not widely available. As a result, SUV_{max} and SUV_{max}-based thresholding remain the predominant techniques for measurements of tumor uptake. The repeatability of SUV_{max}-based uptake measurements has been characterized, and it is relatively poor. Alternative PET-derived parameters (including global MTV, global TLG and various measures of uptake heterogeneity) are being developed and show promise for predicting treatment response and patient survival. PERCIST has been generally accepted as the standard protocol for solid tumor response assessment with ^{18}F-FDG/PET-CT.

It should also be noted that much progress has been made in quantitative SPECT,[89] for which the primary application is patient-specific dosimetry for internal radionuclide therapy. Hybrid SPECT-CT supports accurate attenuation correction and volume of interest definition. Algorithms have been developed that permit accurate compensation for scatter and collimator-detector response within the context of iterative image reconstruction. SPECT-CT currently provides activity measurements that are accurate to better than 10% in larger organs. However, with spatial resolution on the order of 20 mm FWHM, quantification of tumor uptake is still problematic and heavily dependent on PVC. Accuracy of 10% or better has been reported in simulations with spherically-shaped tumor surrogates, but tumor quantification with SPECT remains to be validated in the clinical setting.

Development Needs. Given that PERCIST is now an "industry standard", there is a pressing need to characterize the accuracy and reproducibility of SUV_{peak} measurements as well as measurements of SUV_{peak} change during therapy.[42] The PERCIST prescription for MTV determination should be compared with alternative segmentation techniques.

Another acute need, and the area of greatest current focus, is for methods of tumor segmentation and SUV_{mean} calculation that are accurate, have good reproducibility and are sufficiently facile for routine use in the clinical research setting. The accuracy of SUV_{mean} measurements is degraded by the PVE; thus the development of practical and accurate techniques for partial volume correction or minimization is also among the most pressing needs. New methods based on the tumor-background interface appear very promising for segmentation of tumor functional volumes, and more studies are needed to characterize their accuracy and reproducibility. Once validated, the algorithms must be packaged as user-friendly applications and made widely available to the PET community.

In the absence of facile alternatives, SUV_{max} and SUV_{max}-based thresholding techniques continue to be widely used. Thus there is a continuing need for methods to quantify the associated noise- and tumor size-dependent biases of the SUV_{max} technique.

The utility of SUV and MTV measurements could be greatly enhanced if the individual measurements were accompanied by error bars. Thus practical methods are needed for estimating the variances in these measurements, or at least the count- and image reconstruction/image matrix-related component of such uncertainty.

As PET-MR comes into routine use, the utility of anatomic tumor segmentation should be re-examined in comparison with metabolic tumor segmentation. To the extent that TLG becomes the parameter of choice in predicting patient outcome, MR-based anatomic tumor segmentation may become preferable to PET-based metabolic tumor segmentation because of the accuracy and precision conferred by the higher

spatial resolution and tissue contrast of MR images.

Heterogeneity of ^{18}F-FDG intra-tumoral uptake, and perhaps other PET radiopharmaceuticals as well, carries information that bears on tumor aggressiveness and patient outcome. Access to this information requires development and distribution of facile techniques that succinctly and understandably characterize the degree and spatial correlations of the heterogeneity.

Finally, there is a compelling need to increase the use of fully quantitative kinetic analysis in clinical PET. PET will not realize its full potential in oncology or other diseases until scans are routinely interpreted in terms of model-derived parametric images that relate directly to the biologic abnormalities that underlie disease. That requires advances in PET instrumentation such as increasing axial field of view (mainly a function of cost) and automated methods for determining arterial time-activity curves, as well as increasing the speed with which kinetic models are fitted to single-voxel time-activity curves. The current development of 4-D image reconstruction methods[90] will be helpful in reducing noise in the parametric images reconstructed. Regarding tumor quantification, accurate estimation of single-voxel activity concentrations (which requires time-dependent partial volume correction) and their statistical uncertainties over the course of a dynamic scan is also a major challenge.

Competing Interests

The authors have declared that no competing interest exists.

References

1. Pryma DA, O'Donoghue JA, Humm JL, Jungbluth AA, Old LJ, Larson SM, et al. Correlation of in vivo and in vitro measures of carbonic anhydrase IX antigen expression in renal masses using antibody [124]I-cG250. J Nucl Med. 2011;52:535-40.
2. Juweid ME, Stroobants S, Hoekstra OS, Mottaghy FM, Dietlein M, Guermazi A, 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;25:571-8.
3. 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-50S.
4. Woodard HQ, Bigler RE, Freed B, Russ G. Expression of tissue isotope distribution. J Nucl Med. 1975;16:958-9.
5. Strauss LG, Conti PS. The applications of PET in clinical oncology. J Nucl Med. 1991;32:623-48.
6. Grgic A, Yuksel Y, Groschel A, Schafers HJ, Sybrecht GW, Kirsch CM, et al. Risk stratification of solitary pulmonary nodules by means of PET using ^{18}F-fluorodeoxyglucose and SUV quantification. Eur J Nucl Med Mol Imag. 2010;37:1087-94.
7. Schoder H, Noy A, Gonen M, Weng L, Green D, Erdi YE, et al. Intensity of [18]fluorodeoxyglucose uptake in positron emission tomography distinguishes between indolent and aggressive non-Hodgkin's lymphoma. J Clin Oncol. 2005;23:4643-51.
8. Lin C, Itti E, Haioun C, Petegnief Y, Luciani A, Dupuis J, et al. Early ^{18}F-FDG PET for prediction of prognosis in patients with diffuse large B-cell lymphoma: SUV-based assessment versus visual analysis. J Nucl Med. 2007;48:1626-32.
9. Sasaki R, Komaki R, Macapinlac H, Erasmus J, Allen P, Forster K, et al. [18F]fluorodeoxyglucose uptake by positron emission tomography predicts outcome of non-small-cell lung cancer. J Clin Oncol. 2005;23:1136-43.
10. Yanagawa M, Tatsumi M, Miyata H, Morii E, Tomiyama N, Watabe T, et al. Evaluation of response to neoadjuvant chemotherapy for esophageal cancer: PET response criteria in solid tumors versus response evaluation criteria in solid tumors. J Nucl Med. 2012;53:872-80.
11. Cochet A, Pigeonnat S, Khoury B, Vrigneaud JM, Touzery C, Berriolo-Riedinger A, et al. Evaluation of breast tumor blood flow with dynamic first-pass ^{18}F-FDG PET/CT: comparison with angiogenesis markers and prognostic factors. J Nucl Med. 2012;53:512-20.
12. Hugonnet F, Fournier L, Medioni J, Smadja C, Hindie E, Huchet V, et al. Metastatic renal cell carcinoma: relationship between initial metastasis hypoxia, change after 1 month's sunitinib, and therapeutic response: an ^{18}F-fluoromisonidazole PET/CT study. J Nucl Med. 2011;52:1048-55.
13. O'Donoghue JA, Smith-Jones PM, Humm JL, Ruan S, Pryma DA, Jungbluth AA, et al. [124]I-huA33 antibody uptake is driven by A33 antigen concentration in tissues from colorectal cancer patients imaged by immuno-PET. J Nucl Med. 2011;52:1878-85.
14. Tomasi G, Kenny L, Mauri F, Turkheimer F, Aboagye EO. Quantification of receptor-ligand binding with [^{18}F]fluciclatide in metastatic breast cancer patients. Eur J Nucl Med Mol Imag. 2011;38:2186-97.
15. Eary JF, O'Sullivan F, O'Sullivan J, Conrad EU. Spatial heterogeneity in sarcoma ^{18}F-FDG uptake as a predictor of patient outcome. J Nucl Med. 2008;49:1973-9.
16. Kurland BF, Peterson LM, Lee JH, Linden HM, Schubert EK, Dunnwald LK, et al. Between-patient and within-patient (site-to-site) variability in estrogen receptor binding, measured in vivo by ^{18}F-fluoroestradiol PET. J Nucl Med. 2011;52:1541-9.
17. Larson SM, Erdi Y, Akhurst T, Mazumdar M, Macapinlac HA, Finn RD, 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;2:159-71.
18. Manohar K, Mittal BR, Bhattacharya A, Malhotra P, Varma S. Prognostic value of quantitative parameters derived on initial staging ^{18}F-fluorodeoxyglucose positron emission tomography/computed tomography in patients with high-grade non-Hodgkin's lymphoma. Nucl Med Commun. 2012;33:974-81.
19. Liao S, Penney BC, Wroblewski K, Zhang H, Simon CA, Kampalath R, et al. Prognostic value of metabolic tumor burden on ^{18}F-FDG PET in nonsurgical patients with non-small cell lung cancer. Eur J Nucl Med Mol Imag. 2012;39:27-38.
20. Carrasquillo JA, Pandit-Taskar N, O'Donoghue JA, Humm JL, Zanzonico P, Smith-Jones PM, et al. [124]I-huA33 antibody PET of colorectal cancer. J Nucl Med. 2011;52:1173-80.
21. Williams SP, Flores-Mercado JE, Port RE, Bengtsson T. Quantitation of glucose uptake in tumors by dynamic FDG-PET has less glucose bias and lower variability when adjusted for partial saturation of glucose transport. EJNMMI Research. 2011;2:6.
22. Cheebsumon P, Velasquez LM, Hoekstra CJ, Hayes W, Kloet RW, Hoetjes NJ, et al. Measuring response to therapy using FDG PET: semi-quantitative and full kinetic analysis. Eur J Nucl Med Mol Imag. 2011;38:832-42.
23. Eary JF, Mankoff DA, Spence AM, Berger MS, Olshen A, Link JM, et al. 2-[C-11]thymidine imaging of malignant brain tumors. Cancer Research. 1999;59:615-21.
24. Qi J, Leahy RM, Cherry SR, Chatziioannou A, Farquhar TH. High-resolution 3D Bayesian image reconstruction using the microPET small-animal scanner. Phys Med Biol. 1998;43:1001-13.
25. Jakoby BW, Bercier Y, Conti M, Casey ME, Bendriem B, Townsend DW. Physical and clinical performance of the mCT time-of-flight PET/CT scanner. Phys Med Biol. 2011;56:2375-89.
26. Erdi YE, Nehmeh SA, Pan T, Pevsner A, Rosenzweig KE, Mageras G, et al. The CT motion quantitation of lung lesions and its impact on PET-measured SUVs. J Nucl Med. 2004;45:1287-92.
27. Benz MR, Allen-Auerbach MS, Eilber FC, Chen HJ, Dry S, Phelps ME, et al. Combined assessment of metabolic and volumetric changes for assessment of tumor response in patients with soft-tissue sarcomas. J Nucl Med. 2008;49:1579-84.
28. Ahmadian A, Ay MR, Bidgoli JH, Sarkar S, Zaidi H. Correction of oral contrast artifacts in CT-based attenuation correction of PET images using an automated segmentation algorithm. Eur J Nucl Med Mol Imag. 2008;35:1812-23.
29. Boellaard R. Need for standardization of ^{18}F-FDG PET/CT for treatment response assessments. J Nucl Med. 2011;52:93S-100S.
30. Fahey FH, Kinahan PE, Doot RK, Kocak M, Thurston H, Poussaint TY. Variability in PET quantitation within a multicenter consortium. Med Phys. 2010;37:3660-6.
31. Graham MM, Peterson LM, Hayward RM. Comparison of simplified quantitative analyses of FDG uptake. Nucl Med Biol. 2000;27:647-55.
32. Hawkins RA, Choi Y, Huang SC, Messa C, Hoh CK, Phelps ME. Quantitating tumor glucose metabolism with FDG and PET. J Nucl Med. 1992;33:339-44.
33. Krak NC, van der Hoeven JJ, Hoekstra OS, Twisk JW, 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 Imag. 2003;30:674-81.
34. Sugawara Y, Zasadny KR, Neuhoff AW, Wahl RL. Reevaluation of the standardized uptake value for FDG: variations with body weight and methods for correction. Radiology. 1999;213:521-5.

35. 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;189:847-50.

36. 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.

37. Cheebsumon P, Yaqub M, van Velden FH, Hoekstra OS, Lammertsma AA, Boellaard R. Impact of [18F]FDG PET imaging parameters on automatic tumour delineation: need for improved tumour delineation methodology. Eur J Nucl Med Mol Imag. 2011;38:2136-44.

38. 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;45:1519-27.

39. Boucek JA, Francis RJ, Jones CG, Khan N, Turlach BA, Green AJ. Assessment of tumour response with 18F-fluorodeoxyglucose positron emission tomography using three-dimensional measures compared to SUVmax--a phantom study. Phys Med Biol. 2008;53:4213-30.

40. Schwartz J, Humm JL, Gonen M, Kalaigian H, Schoder H, Larson SM, et al. Repeatability of SUV measurements in serial PET. Med Phys. 2011;38:2629-38.

41. de Langen AJ, Vincent A, Velasquez LM, van Tinteren H, Boellaard R, Shankar LK, et al. Repeatability of 18F-FDG uptake measurements in tumors: a metaanalysis. J Nucl Med. 2012;53:701-8.

42. Vanderhoek M, Perlman SB, Jeraj R. Impact of the definition of peak standardized uptake value on quantification of treatment response. J Nucl Med. 2012;53:4-11.

43. Geworski L, Knoop BO, de Cabrejas ML, Knapp WH, Munz DL. Recovery correction for quantitation in emission tomography: a feasibility study. Eur J Nucl Med Mol Imag. 2000;27:161-9.

44. Rousset OG, Ma Y, Evans AC. Correction for partial volume effects in PET: principle and validation. J Nucl Med. 1998;39:904-11.

45. Frouin V, Comtat C, Reilhac A, Gregoire MC. Correction of partial-volume effect for PET striatal imaging: fast implementation and study of robustness. J Nucl Med. 2002;43:1715-26.

46. DiFilippo F, Patel S, Asosingh K, Erzurum S. Small-animal imaging using clinical positron emission tomography/computed tomography and super-resolution. Mol Imag. 2012;11:210-9.

47. Brix G, Doll J, Bellemann ME, Trojan H, Haberkorn U, Schmidlin P, et al. Use of scanner characteristics in iterative image reconstruction for high-resolution positron emission tomography studies of small animals. Eur J Nucl Med Mol Imag. 1997;24:779-86.

48. Reader AJ, Julyan PJ, Williams H, Hastings DL, Zweit J. EM algorithm system modeling by image-space techniques for PET reconstruction. IEEE Trans Nucl Sci. 2003;50:1392-7.

49. Richardson WH. Bayesian-based iterative method of image restoration. J Optical Society of America. 1972;62:55-9.

50. Lucy L. An iterative technique for the rectification of observed distributions. The Astronomical Journal. 1974;79:745.

51. Hoetjes NJ, van Velden FHP, Hoekstra OS, Hoekstra CJ, Krak NC, Lammertsma AA, et al. Partial volume correction strategies for quantitative FDG PET in oncology. Eur J Nucl Med Mol Imag. 2010;37:1679-87.

52. Teo BK, Seo Y, Bacharach SL, Carrasquillo JA, Libutti SK, Shukla H, et al. Partial-volume correction in PET: validation of an iterative postreconstruction method with phantom and patient data. J Nucl Med. 2007;48:802-10.

53. Panin VY, Kehren F, Michel C, Casey M. Fully 3-D PET reconstruction with system matrix derived from point source measurements. IEEE Trans Med Imag. 2006;25:907-21.

54. Mourik J, Lubberink M, Klumpers U, Comans EF, Lammertsma AA, Boellaard R. Partial volume corrected image derived input functions for dynamic PET brain studies: methodology and validation for [11C] flumazenil. Neuroimage. 2008;39:1041.

55. Lasnon C, Hicks RJ, Beauregard JM, Milner A, Paciencia M, Guizard AV, et al. Impact of point spread function reconstruction on thoracic lymph node staging with 18F-FDG PET/CT in non small cell lung cancer. Clin Nucl Med. 2012;37:971-6.

56. Snyder DL, Miller MI, Thomas LJ, Politte DG. Noise and edge artifacts in maximum-likelihood reconstructions for emission tomography. IEEE Trans Med Imag. 1987;6:228-38.

57. Bai B, Esser PD. The effect of edge artifacts on quantification of positron emission tomography. IEEE Nuclear Science Symposium and Medical Imaging Conference Record. 2010;: 2263-66.

58. Tong S, Alessio AM, Thielemans K, Stearns C, Ross S, Kinahan PE. Properties and mitigation of edge artifacts in PSF-based PET reconstruction. IEEE Trans Nucl Sci. 2011;58:2264-75.

59. Bai B, Li Q, Leahy RM. MR guided PET image reconstruction. Seminars in Nuclear Medicine. 2013;43:30-44.

60. Geets X, Lee JA, Bol A, Lonneux M, Gregoire V. A gradient-based method for segmenting FDG-PET images: methodology and validation. Eur J Nucl Med Mol Imag. 2007;34:1427-38.

61. Li H, Thorstad WL, Biehl KJ, Laforest R, Su Y, Shoghi KI, et al. A novel PET tumor delineation method based on adaptive region-growing and dual-front active contours. Med Phys. 2008;35:3711-21.

62. Hofheinz F, Potzsch C, Oehme L, Beuthien-Baumann B, Steinbach J, Kotzerke J, et al. Automatic volume delineation in oncological PET: evaluation of a dedicated software tool and comparison with manual delineation in clinical data sets. Nuklearmedizin. 2012;51:9-16.

63. Hatt M, Cheze le Rest C, Turzo A, Roux C, Visvikis D. A fuzzy locally adaptive Bayesian segmentation approach for volume determination in PET. IEEE Trans Med Imag. 2009;28:881-93.

64. Belhassen S, Zaidi H. A novel fuzzy C-means algorithm for unsupervised heterogeneous tumor quantification in PET. Med Phys. 2010;37:1309-24.

65. Hatt M, Cheze Le Rest C, Albarghach N, Pradier O, Visvikis D. PET functional volume delineation: a robustness and repeatability study. Eur J Nucl Med Mol Imag. 2011;38:663-72.

66. Hatt M, Cheze-le Rest C, van Baardwijk A, Lambin P, Pradier O, Visvikis D. Impact of tumor size and tracer uptake heterogeneity in 18F-FDG PET and CT non-small cell lung cancer tumor delineation. J Nucl Med. 2011;52:1690-7.

67. Frings V, de Langen AJ, Smit EF, van Velden FH, van Tinteren H, et al. Repeatability of metabolically active volume measurements with 18F-FDG and 18F-FLT PET in non-small cell lung cancer. J Nucl Med. 2010;51:1870-7.

68. Heijmen L, de Geus-Oei LF, de Wilt JH, Visvikis D, Hatt M, Visser EP, et al. Reproducibility of functional volume and activity concentration in 18F-FDG PET/CT of liver metastases in colorectal cancer. Eur J Nucl Med Mol Imag. 2012;12:1858-67.

69. Bhatt R, Adjouadi M, Goryawala M, Gulec SA, McGoron AJ. An algorithm for PET tumor volume and activity quantification: without specifying camera's point spread function (PSF). Med Phys. 2012;39:4187-202.

70. Bading JR, Frankel P, Zhan J, Mas J, Raubitschek A. A recessed central volume of interest (VOI) technique for tumor quantitation in PET. J Nucl Med. 2009;50 (Suppl. 2):284P.

71. Carson RE, Yan Y, Chodkowski B, Yap TK, Daube-Witherspoon ME. Precision and accuracy of regional radioactivity quantitation using the maximum likelihood EM reconstruction algorithm. IEEE Trans Med Imag. 1994;13:526-37.

72. Muralidharan V, Kwok M, Lee ST, Lau L, Scott AM, Christophi C. Prognostic ability of 18F-FDG PET/CT in the assessment of colorectal liver metastases. J Nucl Med. 2012;53:1345-51.

73. van Velden FH, Cheebsumon P, Yaqub M, Smit EF, Hoekstra OS, Lammertsma AA, et al. Evaluation of a cumulative SUV-volume histogram method for parameterizing heterogeneous intratumoural FDG uptake in non-small cell lung cancer PET studies. Eur J Nucl Med Mol Imag. 2011;38:1636-47.

74. O'Sullivan F, Wolsztynski E, O'Sullivan J, Richards T, Conrad EU, Eary JF. A statistical modeling approach to the analysis of spatial patterns of FDG-PET uptake in human sarcoma. IEEE Trans Med Imag. 2011;30:2059-71.

75. Tixier F, Le Rest CC, Hatt M, Albarghach N, Pradier O, Metges JP, et al. Intratumor heterogeneity characterized by textural features on baseline 18F-FDG PET images predicts response to concomitant radiochemotherapy in esophageal cancer. J Nucl Med. 2011;52:369-78.

76. Huesman RH. A new fast algorithm for the evaluation of regions of interest and statistical uncertainty in computed tomography. Phys Med Biol. 1984;29:543-52.

77. Carson RE, Yan Y, Daube-Witherspoon ME, Freedman N, Bacharach SL, Herscovitch P. An approximation formula for the variance of PET region-of-interest values. IEEE Trans Med Imag. 1993;12:240-50.

78. Barrett HH, Wilson DW, Tsui BM. Noise properties of the EM algorithm: I. theory. Phys Med Biol. 1994;39:833-46.

79. 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;55:1453-73.

80. Bading JR, Frankel P, Bertilsson H, Lopatin G, Raubitschek A. Statistical modeling of the PET maximum voxel. J Nucl Med. 2011;52 (Suppl. 1):484P.

81. Mullani NA, Ficke DC, Hartz R, Markham J, Wong G. System design of fast PET scanners utilizing time-of-flight. IEEE Trans Nucl Sci. 1981;28:104-8.

82. Karp JS, Surti S, Daube-Witherspoon ME, Muehllehner G. Benefit of time-of-flight in PET: experimental and clinical results. J Nucl Med. 2008;49:462-70.

83. Surti S, Karp S, Popescu LM, Daube-Witherspoon E, Werner M. Investigation of time-of-flight benefit for fully 3-DPET. IEEE Trans Med Imag. 2006;25:529-38.

84. 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;50:1315-23.

85. Adams MC, Turkington TG, Wilson JM, Wong TZ. A systematic review of the factors affecting accuracy of SUV measurements. Am J Roentgenology. 2010;195:310-20.

86. Watson CC. Estimating effective model kernel widths for PSF reconstruction in PET. IEEE Nuclear Science Symposium and Medical Imaging Conference Record. 2011;:2368-74.

87. Ng TSC, Bading JR, Park R, Sohi H, Procissi D, Colcher D, et al. Quantitative, simultaneous PET/MRI for intratumoral imaging with an MRI-compatible PET scanner. J Nucl Med. 2012;53:1102-9.

88. Keyes Jr JW. SUV: Standard uptake or silly useless value? J Nucl Med. 1995;36:1836-1839.

89. Dewaraja YK, Frey EC, Sgouros G, Brill AB, Robertson P, Zanzonico PB, et al. MIRD Pamphlet No. 23: quantitative SPECT for patient-specific 3-dimensional dosimetry in internal radionuclide therapy. J Nucl Med. 2012;53:1310-25.

90. Wang G, Qi J. Generalized algorithms for direct reconstruction of parametric images from dynamic PET data. IEEE Trans Med Imag. 2009;28:1717-26.