Impact of the Definition of Peak Standardized Uptake Value on Quantification of Treatment Response

Matt Vanderhoek[1], Scott B. Perlman[2], and Robert Jeraj[1]

[1]Department of Medical Physics, University of Wisconsin, Madison, Wisconsin; [2]Department of Radiology, Section of Nuclear Medicine, University of Wisconsin, Madison, Wisconsin

PET-based treatment response assessment typically measures the change in maximum standardized uptake value (SUV_{max}), which is adversely affected by noise. Peak SUV (SUV_{peak}) has been recommended as a more robust alternative, but its associated region of interest (ROI_{peak}) is not uniquely defined. We investigated the impact of different ROI_{peak} definitions on quantification of SUV_{peak} and tumor response. **Methods:** Seventeen patients with solid malignancies were treated with a multitargeted receptor tyrosine kinase inhibitor resulting in a variety of responses. Using the cellular proliferation marker 3′-deoxy-3′-^{18}F-fluorothymidine (^{18}F-FLT), whole-body PET/CT scans were acquired at baseline and during treatment. ^{18}F-FLT–avid lesions (~2/patient) were segmented on PET images, and tumor response was assessed via the relative change in SUV_{peak}. For each tumor, 24 different SUV_{peaks} were determined by changing ROI_{peak} shape (circles vs. spheres), size (7.5–20 mm), and location (centered on SUV_{max} vs. placed in highest-uptake region), encompassing different definitions from the literature. Within each tumor, variations in the 24 SUV_{peaks} and tumor responses were measured using coefficient of variation (CV), standardized deviation (SD), and range. For each ROI_{peak} definition, a population average SUV_{peak} and tumor response were determined over all tumors. **Results:** A substantial variation in both SUV_{peak} and tumor response resulted from changing the ROI_{peak} definition. The variable ROI_{peak} definition led to an intratumor SUV_{peak} variation ranging from 49% above to 46% below the mean (CV, 17%) and an intratumor SUV_{peak} response variation ranging from 49% above to 35% below the mean (SD, 9%). The variable ROI_{peak} definition led to a population average SUV_{peak} variation ranging from 24% above to 28% below the mean (CV, 14%) and a population average SUV_{peak} response variation ranging from only 3% above to 3% below the mean (SD, 2%). The size of ROI_{peak} caused more variation in intratumor response than did the location or shape of ROI_{peak}. Population average tumor response was independent of size, shape, and location of ROI_{peak}. **Conclusion:** Quantification of individual tumor response using SUV_{peak} is highly sensitive to the ROI_{peak} definition, which can significantly affect the use of SUV_{peak} for assessment of treatment response. Clinical trials are necessary to compare the efficacy of SUV_{peak} and SUV_{max} for quantification of response to therapy.

Key Words: PET; SUV_{peak}; treatment response

J Nucl Med 2012; 53:4–11

DOI: 10.2967/jnumed.111.093443

PET continues to gain importance as a tool to assess response to therapy. Typically, the change in standardized uptake value (SUV) is measured to quantify treatment response (*1*). Patients are then classified into different response categories based on the relative change in SUV. These categories include complete response, partial response, stable disease, and progressive disease. Such response classifications are often used to guide subsequent treatment decisions and can be predictive of clinical outcome (*2–4*).

Most response assessment studies measure the change in maximum SUV (SUV_{max}), a single-pixel value that is adversely affected by noise (*5–8*), which leads to uncertainty in the quantification of treatment response. Consequently, peak SUV (SUV_{peak}) has been suggested as a more robust alternative (*9*), defined as the average SUV within a small, fixed-size region of interest (ROI_{peak}) centered on a high-uptake part of the tumor (*9*). SUV_{peak} is illustrated in Figure 1. Because of its larger volume, SUV_{peak} is less affected by image noise than SUV_{max} (*6,7,10*) and therefore is expected to reduce uncertainties in the quantification of response to therapy.

There is a wide variety of SUV_{peak} definitions in the literature, and they differ in the shape, size, and location of the ROI_{peak} (Fig. 1). Shapes and sizes include square and cuboidal regions with side lengths ranging from 7 to 15 mm (*5,8,11–13*), as well as circular, cylindric, and spheric regions with diameters ranging from 9 to 17 mm (*6,7,9,14–16*). Locations include the tumor region with the highest radiotracer uptake, the tumor region yielding the greatest SUV_{peak}, and the tumor region containing the voxel of maximum uptake.

The definition of SUV_{peak} could significantly affect the quantification of treatment response. Altering the size, shape, or location of ROI_{peak} may affect the relative change in SUV_{peak} and ultimately the classification of response. Uncertainties in the quantification of response could have significant implications regarding treatment decisions and

Received May 19, 2011; revision accepted Oct. 13, 2011.

For correspondence or reprints contact: Robert Jeraj, Department of Medical Physics, Wisconsin Institutes for Medical Research, 1111 Highland Ave., Room 1005, Madison, WI 53705-2275.

E-mail: rjeraj@wisc.edu

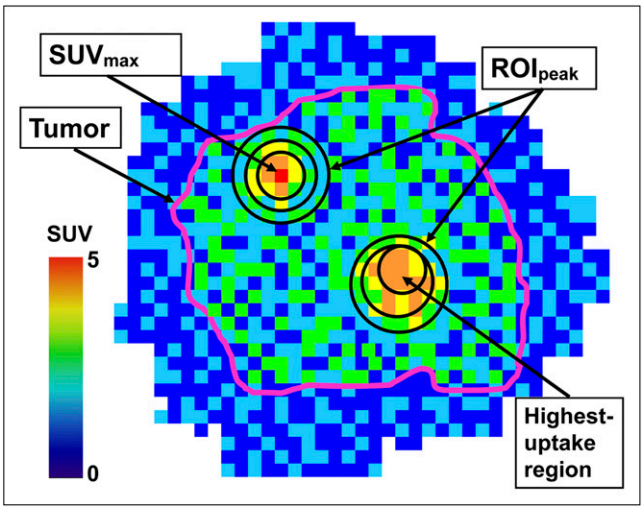

FIGURE 1. Multiple definitions of ROI_{peak}. In this schematic PET image of radiotracer uptake in tumor (purple outline), ROI_{peaks} of different sizes and in different locations are shown. Shape of ROI_{peak} can vary too (only circles are shown).

clinical prognoses. Furthermore, these uncertainties could influence the recommendation to use SUV_{peak} rather than SUV_{max} for response assessment. Consequently, we investigated the impact of different ROI_{peak} definitions on the quantification of SUV_{peak} and tumor response to therapy.

MATERIALS AND METHODS

Treatment and Imaging

Seventeen patients with advanced solid malignancies were treated with a multitargeted receptor tyrosine kinase inhibitor with antiproliferative and antiangiogenic effects. Malignancies included a diverse range of tumor types: renal cell carcinoma ($n = 7$), esophagus ($n = 2$), hepatocellular ($n = 2$), prostate ($n = 1$), sarcoma ($n = 1$), small cell lung ($n = 2$), thymus ($n = 1$), and uterine carcinosarcoma ($n = 1$). Response to therapy was measured using the PET radiotracer 3′-deoxy-3′-^{18}F-fluorothymidine (^{18}F-FLT). As a surrogate of cellular proliferation, ^{18}F-FLT is emerging as a promising candidate for chemotherapy response assessment as demonstrated in patients with lymphoma, breast cancer, and glioma (*17–23*). Patients were injected intravenously with 240 MBq (6.5 mCi) of ^{18}F-FLT and underwent whole-body PET/CT at baseline (pretreatment) and during treatment using a Discovery LS PET/CT scanner (GE Healthcare). ^{18}F-FLT was synthesized following the method described by Martin et al. with slight modifications (*24*). PET/CT began 47 ± 4 min after injection and extended inferiorly from the base of the skull to the distal femora. Acquisition time was 10 min per bed position. PET images were reconstructed on a 128 × 128 grid over a 50-cm field of view using the ordered-subset expectation maximization algorithm with 2 iterations, 28 subsets, a 5-mm gaussian loop (interiteration) filter, a 3-mm gaussian postprocessing filter, and CT attenuation correction. On average, patient weight changed only 1.5% between the 2 PET scans.

The study protocol was approved by the University of Wisconsin Health Sciences Institutional Review Board, the Scientific Review Board of the University of Wisconsin Carbone Comprehensive Cancer Center, and the University of Wisconsin Radiation Drug Research Committee. Written informed consent was obtained from each patient before enrollment in the study.

Quantification of SUV_{peak} and Tumor Response

PET activity concentrations (MBq/cm^3) were converted to standardized uptake values by dividing by the injected activity per patient mass. ^{18}F-FLT–avid lesions (~2/patient) were segmented on PET images by an experienced nuclear medicine physician. The location and number of lesions were as follows: lung, 14; mediastinum, 5; liver, 4; abdomen, 3; adrenal, 2; gastrointestinal tract, 2; pelvis, 1; gluteus, 1; uterus, 1; arm, 1; bone, 1. Tumor volumes ranged from 1 cm^3 to 530 cm^3, with an average of 66 cm^3.

For each tumor, 24 different SUV_{peaks} were determined by changing the region of interest (ROI_{peak}) used to measure SUV_{peak} (Fig. 2). The shape, size, and location of ROI_{peak} were varied as follows: for shape, circular (2-dimensional) or spheric (3-dimensional) ROIs were used; for size, ROI diameters of 7.5, 10, 12.5, 15, 17.5, or 20 mm, encompassing the range of fixed ROI lengths in the literature, were used; for location, the ROI was centered on SUV_{max} or was placed in the highest-uptake region. An example ROI is shown in Supplemental Figure 1, where a 12.5-mm-diameter circular ROI was placed in the highest-uptake region of a lung lesion (supplemental materials are available online only at http://jnm.snmjournals.org).

SUV_{peak} was determined automatically. First, an ROI (circular or spheric) was centered on each tumor voxel and the average SUV within the ROI was determined by weighting each voxel uptake value by the percentage of its volume contained within the ROI. The ROI location yielding the greatest average SUV was defined as the highest-uptake region of the tumor (Supplemental Fig. 1). In addition, an ROI (circular or spheric) was centered on SUV_{max} and the average SUV within the ROI was determined.

For each tumor, the 24 SUV_{peaks} were normalized to the mean intratumor SUV_{peak} (Eq. 1) and their variation was measured using the coefficient of variation (CV) and range.

$$\overline{SUV^n_{peak}} = \frac{\sum_{i=1}^{i=24} \left(SUV^n_{peak}\right)_i}{24} \qquad \text{Eq. 1}$$

Here, $\overline{SUV^n_{peak}}$ is the mean intratumor SUV_{peak} for an individual tumor n, and $(SUV^n_{peak})_i$ corresponds to an SUV_{peak} determined

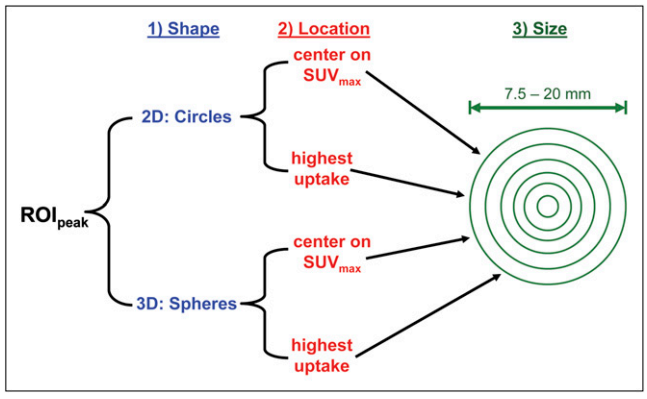

FIGURE 2. Varying ROI_{peak}. Shape, location, and size of ROI_{peak} were varied as shown to yield 24 different SUV_{peaks} for each tumor.

using 1 of the 24 ROI_{peak} definitions (the *ith* definition, e.g., 12.5-mm-diameter spheric ROI placed in the highest-uptake region).

For each ROI_{peak} definition, a population average SUV_{peak} (Eq. 2) was determined over all tumors.

$$\overline{(SUV_{peak})}_i = \frac{\sum_{n=1}^{N} (SUV_{peak}^n)_i}{N}. \quad \text{Eq. 2}$$

Here, $\overline{(SUV_{peak})}_i$ is the population average SUV_{peak} using the *ith* ROI_{peak} definition, n is an individual tumor, N is the total number of tumors, and $(SUV_{peak}^n)_i$ corresponds to an SUV_{peak} of an individual tumor determined using 1 of the 24 ROI_{peak} definitions (the *ith* definition). The variation of the 24 population average SUV_{peak} was measured using coefficient of variation and range.

Tumor response (R) during treatment was defined as the relative change in SUV_{peak} normalized to the baseline SUV_{peak} (Eq. 3).

$$R_i^n = \frac{(SUV_{peak}^{n,\,\text{during tx}})_i - (SUV_{peak}^{n,\,\text{baseline}})_i}{(SUV_{peak}^{n,\,\text{baseline}})_i} \times 100\%. \quad \text{Eq. 3}$$

Here, R_i^n is the response of an individual tumor (n) using 1 of the 24 ROI_{peak} definition (the *ith* definition).

For each tumor, the 24 different SUV_{peaks} gave rise to 24 different responses whose variation was measured using the standardized deviation (SD) and range. For each ROI_{peak} definition, a population average response (Eq. 4) was determined over all tumors.

$$\overline{R_i} = \frac{\sum_{n=1}^{N} R_i^n}{N}. \quad \text{Eq. 4}$$

Here, $\overline{R_i}$ is the population average response using the *ith* ROI_{peak} definition, n is an individual tumor, and N is the total number of tumors. The variation of the 24 population average responses was measured using SD and range.

Tumor response was also determined using SUV_{max} for comparison with response measured using SUV_{peak}. SUV_{max} can be considered as a special case of SUV_{peak} in the limit of a very small ROI_{peak} (single-voxel ROI_{peak}). For SUV_{max}, an equivalent diameter of 5 mm was derived by calculating the diameter of a sphere whose volume equaled the volume of the single voxel (65 mm^3) represented by SUV_{max}.

One-way ANOVA was used to test whether the ROI_{peak} definition resulted in statistically significant differences in SUV_{peak} and tumor response. The Levene test for equal variance was used, and means were compared with the Bonferroni test. Differences were considered statistically significant at an α-level of less than 0.05/24. Correlations between the variation in SUV_{peak} and tumor response and other tumor characteristics were tested using the Pearson correlation coefficient (r) and considered statistically significant at an α-level of less than 0.05.

RESULTS

Individual Tumors

SUV_{peak}. Within individual tumors, considerable variation in SUV_{peak} resulted from changing the ROI_{peak} definition. The variable ROI_{peak} definition led to an intratumor SUV_{peak} variation ranging from 49% above to 46% below the mean, resulting in a 17% CV. These intratumor variations in SUV_{peak} are highlighted for a retroperitoneal lesion in Supplemental Figure 2 and for all lesions in Supplemental Figure 3.

The size of ROI_{peak} caused more variation in intratumor SUV_{peak} than did the location or shape of ROI_{peak} (Supplemental Fig. 3). Within individual tumors, varying ROI_{peak} diameter resulted on average in a 14% CV associated with SUV_{peak}, compared with a CV of 9% and 5% when the location or shape, respectively, of ROI_{peak} was varied. In general, intratumor SUV_{peak} tended to decrease, but its variation tended to increase as the size of ROI_{peak} increased (Supplemental Fig. 2).

There was no significant correlation between tumor size and the variation in intratumor SUV_{peak} (Supplemental Fig. 3, tumors ordered by size). Furthermore, there was no significant correlation between intratumor uptake heterogeneity (measured by CV of tumor uptake) and the variation in intratumor SUV_{peak}.

Tumor Response. Within individual tumors, a substantial variation in tumor response resulted from changing the ROI_{peak} definition. Intratumor response ranged from 49% above to 35% below the mean, resulting in a 9% SD. These intratumor variations in response are highlighted for a retroperitoneal lesion in Figure 3 and for all lesions in Figure 4. Responses determined using SUV_{max} were within the range of responses quantified with SUV_{peak} in almost 70% of all tumors (Fig. 4).

Variation in intratumor response resulted in the ambiguous classification of individual tumors into multiple response categories (Table 1; Figs. 3 and 4). Different response thresholds were applied to classify tumors into response categories (e.g., +30% and –30% for progressive

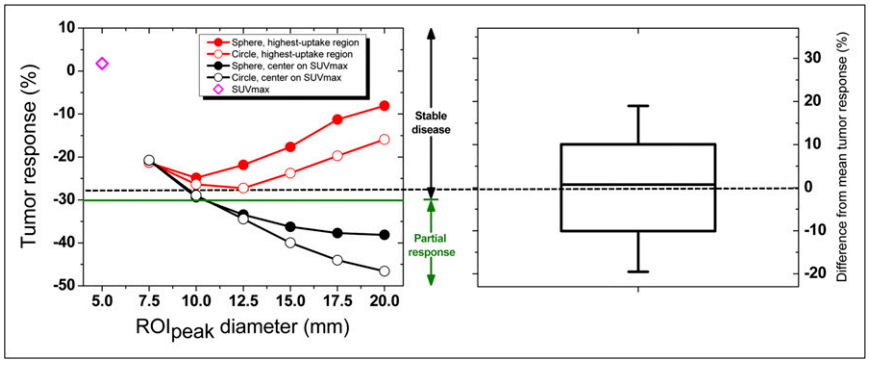

FIGURE 3. Variation in tumor response for retroperitoneal lesion (lesion 11). (Left) 24 different SUV_{peak} tumor responses (mean response, dashed line) arising from 24 different ROI_{peak} definitions. Response was ambiguously classified as either stable disease (above –30%, green line) or partial response (below –30%). SUV_{max} response is also shown. (Right) Box represents SD, whiskers show range, and solid line depicts median of SUV_{peak} response.

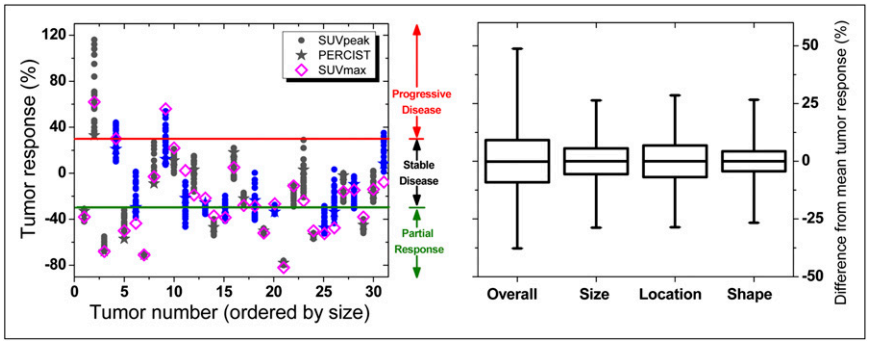

FIGURE 4. Intratumor variation in SUV_{peak} responses. (Left) Responses of 31 tumors (ordered from smallest to largest) arising from different ROI_{peak} definitions. Responses of 12 tumors (in blue) were ambiguously classified as either progressive disease/stable disease (+30%, solid red line) or stable disease/partial response (–30%, solid green line). SUV_{max} response and SUV_{peak} PERCIST response (1.25-cm-diameter sphere in highest-uptake region) are also shown. (Right) Overall variation in intratumor response and variation associated with changing size, location, and shape of ROI_{peak}. Boxes represent SD, whiskers show range, and solid line depicts median of response values.

disease/stable disease and stable disease/partial response thresholds, respectively, as recommended by PET Response Criteria in Solid Tumors [PERCIST]). When response thresholds of ±20%, ±30%, and ±40% were applied, 55%, 42%, and 32%, respectively, of all tumors suffered from an ambiguous response classification. In addition, response classifications using SUV_{peak} and SUV_{max} were compared (Table 1; Figs. 3 and 4).

The size, location, and shape of ROI_{peak} all caused similar variations in intratumor response (Fig. 4). Within individual tumors, varying ROI_{peak} size, location, and shape resulted on average in respective SDs of 5%, 7%, and 5% associated with tumor response.

In general, the magnitude of intratumor response was independent of the size of ROI_{peak}. However, the variation in intratumor response tended to increase as the size of ROI_{peak} increased (Fig. 3). The magnitude and variation of intratumor response were independent of ROI_{peak} shape and location.

A strong correlation was exhibited between tumor response (average of all 24 SUV_{peak} responses for each tumor) and variation in intratumor response ($r = 0.81$, $P < 0.001$, Figs. 4–6). Variation in intratumor response tended to increase as response increased (i.e., as response worsened from partial response to stable disease to progressive disease). There was no significant correlation between tumor size and variation in intratumor response (Fig. 4, tumors ordered by size). Furthermore, there was no significant correlation between intratumor uptake heterogeneity (measured by CV of tumor uptake) and variation in intratumor response.

Population Average

SUV_{peak}. Quantification of the population average SUV_{peak} was substantially affected by changing the ROI_{peak} definition. For different ROI_{peak} definitions, the population average SUV_{peak} ranged from 24% above to 28% below the mean, resulting in a 14% CV (Supplemental Fig. 4). Differences in SUV_{peak} (associated with the ROI_{peak} definitions) between the populations were statistically significant ($P < 0.001$).

The size of ROI_{peak} caused more variation in the population average SUV_{peak} than did the location or shape of ROI_{peak} (Supplemental Fig. 4). Varying ROI_{peak} diameter resulted in a 13% CV associated with the population average SUV_{peak}, compared with a CV of 7% and 4%, respectively, when ROI_{peak} location or shape was varied. Trends observed in the population average SUV_{peak} reflected trends associated with intratumor SUV_{peak}. As the size of ROI_{peak} increased, the population average SUV_{peak} tended to decrease and its variation increased (Supplemental Fig. 4).

Tumor Response. Tumor response during treatment averaged –21% but ranged as high as +116% and as low as −80% (Fig. 4). However, population average tumor response was not significantly affected by changing the ROI_{peak} definition. For different ROI_{peak} definitions, the population average tumor response ranged from only 3% above to 3% below the mean, resulting in a 2% SD (Fig. 7). Differences in response (associated with the ROI_{peak} definitions) between the populations were not statistically significant ($P = 1.00$).

Size, location, and shape of ROI_{peak} all caused minimal variations in population average tumor response (Fig. 7), as all SDs were less than 2%. The magnitude and variation of population average tumor response were independent of the size, shape, and location of ROI_{peak} (Fig. 7).

Tumor Subgroup Analysis

Variation in SUV_{peak} and tumor response was determined using all 35 tumors assessed in this study. In addition, the results were recalculated on 2 different tumor subgroups. The first subgroup, in which lesions were in regions without significant background activity ($n = 23$), was studied in order to reduce the chance that elevated background activity was incorrectly included in ROI_{peak}. Consequently, abdominal, hepatic, renal, and bone lesions were excluded from this group. The second subgroup was one in which lesions were larger than 20 mL ($n = 12$), which is approximately 5 times larger than the largest ROI_{peak} (4.2 mL, 20-mm diameter). Studying this subgroup ensured that ROI_{peak} was completely inside the tumor boundaries and that no background activity was incorrectly included in ROI_{peak}. Results for the 2 tumor subgroups were almost identical to results determined using all tumors (Table 2).

TABLE 1
Tumor Response Classification Using Different Response Thresholds

Response thresholds	Response metric	Total tumors	Partial-response tumors	Stable-disease tumors	Progressive-disease tumors	Partial-response/stable-disease tumors	Stable-disease/progressive-disease tumors	Partial-response/stable-disease/progressive-disease tumors	Ambiguous response (%)
±20%	SUV_{peak}	31	12	1	1	10	6	1	55
	SUV_{max}	31	18	9	4	NA	NA	NA	NA
±30%	SUV_{peak}	31	9	8	1	9	4	0	42
	SUV_{max}	31	13	15	3	NA	NA	NA	NA
±40%	SUV_{peak}	31	7	14	0	7	3	0	32
	SUV_{max}	31	9	20	2	NA	NA	NA	NA

NA = not applicable.

FIGURE 5. Variation in intratumor response vs. response. Variation in intratumor response tended to increase as response increased (i.e., as response worsened from partial response to stable disease to progressive disease).

DISCUSSION

Individual Tumor Response Versus Population Average Response

The region of interest used to determine SUV_{peak} can have a profound effect on its quantification and on the response of individual tumors. On average, different ROI_{peak} definitions resulted in intratumor variations of approximately 17% and 9% for SUV_{peak} and tumor response, respectively, and these variations ranged as high as 50%. This degree of variation can lead to different categorizations of response (Figs. 3 and 4) using criteria such as PERCIST (*9*). With PERCIST, such ambiguous response categorizations

FIGURE 6. Small vs. large variation in intratumor SUV_{peak} response. ^{18}F-FLT PET/CT images of periaortic lesion (left, lesion 24) and pelvic tumor (right, lesion 30) at baseline (top) and during treatment (bottom). Lesions are indicated by white circles. Periaortic lesion responded well, exhibiting fairly uniform reduction of ^{18}F-FLT uptake in higher-uptake regions. Consequently, there was little variation in SUV_{peak} response. In contrast, pelvic tumor responded poorly, with heterogeneous response in higher-uptake regions, resulting in large variation in SUV_{peak} response.

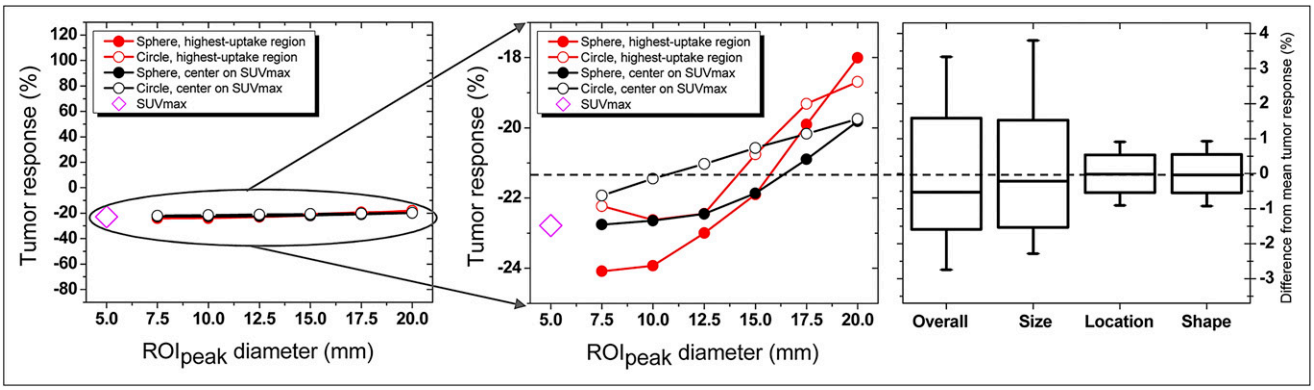

FIGURE 7. Variation in average tumor response in population. (Left and middle) For each ROI_{peak} definition, tumor response was averaged over all tumors, resulting in population average response (mean response, dashed line). SUV_{max} response is also shown. (Right) Overall variation in population average response and variation associated with changing size, location, and shape of ROI_{peak}. Boxes represent SD, whiskers show range, and solid line depicts median. Variation in population average response is extremely small (~2%).

arose in over 40% of the tumor responses assessed in this study (Fig. 4; Table 1). Ambiguous response categorization of tumors increased with narrower response criteria (e.g., ±20%) but was reduced using broader response criteria (e.g., ±40%), similar to that of the MUNICON phase II trial (Table 1) (*23*). The sensitivity of response quantification to the ROI_{peak} definition reveals the need to optimize PET metrics (such as SUV_{peak}) for quantitative response assessment in individual patients. An ambiguous response classification underscores the necessity for a unique, consistent, standard region of interest with associated criteria that can accurately assess response.

Unlike individual tumor responses, population average response was relatively insensitive to the definition of ROI_{peak} used to measure response (Fig. 7), as is consistent with the findings of Krak et al. (*6*). The small variation (only 2%) in population average response occurred because the magnitude of individual tumor responses was independent of the ROI_{peak} definition. Therefore, because of an averaging effect, variation was reduced when determining population average response and might be reduced even further as more tumors are included in the population average. This robustness of population average response points to the strength of PET for accurate quantification of the average response to therapy.

Effects of Different Factors on Variation in Intratumor Response

The variation in intratumor response correlated strongly with tumor response (Fig. 5). Tumors that responded well (i.e., partial response, tumor response < –30%) exhibited significantly less variation in intratumor SUV_{peak} response than did tumors that responded poorly (i.e., stable disease or progressive disease, tumor response > –30%). Well-responding tumors seemed to exhibit a response more uniform than the heterogeneous response of poorly responding tumors (Fig. 6). Thus, SUV_{peak}–based response was considerably more sensitive to the ROI_{peak} definition for poorly responding tumors than for well-responding tumors.

Surprisingly, neither tumor size nor tumor uptake heterogeneity had a significant effect on the variation in either intratumor response or SUV_{peak}. This finding suggests that the characteristics (size, heterogeneity, etc.) of only the high-uptake regions encompassed by ROI_{peak}, not those of the entire tumor, directly affect the variation in tumor response and SUV_{peak}. Though not investigated, partial-volume effects

TABLE 2
Comparison of All Tumors with Different Tumor Subgroups

		Individual tumors				Population average			
		SUV_{peak}		Response		SUV_{peak}		Response	
Tumor group	*n*	Range	Average CV	Range	Average SD	Range	CV	Range	SD
All tumors	35	−46% to +49%	17%	−35% to +49%	9%	−28% to +24%	14%	−3% to +3%	2%
Tumors in regions without significant background activity*	23	−46% to +49%	17%	−35% to +49%	9%	−29% to +26%	15%	−4% to +3%	2%
Tumors > 20 cm^3	12	−43% to +49%	16%	−18% to +36%	8%	−27% to +22%	14%	−4% to +7%	3%

*Tumors in abdominal, hepatic, renal, and bone regions were excluded because normal, elevated ^{18}F-FLT uptake was present in these areas.

tend to reduce uptake heterogeneity and therefore are expected to reduce the variation in both SUV_{peak} and response. Thus, a greater variation in both SUV_{peak} and response should result from partial-volume correction of the PET data.

The variable ROI_{peak} definition led to a variation in intratumor response that was about half that of SUV_{peak}. Tumor response was determined via normalization by baseline SUV_{peak}, in effect canceling out some of the variation in SUV_{peak}, which may explain the reduced variation in intratumor response. Most of the variation in SUV_{peak} was due to the size of ROI_{peak}, as is consistent with the findings of Boellaard et al. (*5*). As expected, variation in both intratumor response and SUV_{peak} increased as the size of ROI_{peak} increased.

For each ROI_{peak} definition, population average SUV_{peak} preserved the trends caused by the size, shape, and location of ROI_{peak}. Consequently, the variation in population average SUV_{peak} was approximately equal to the variation in intratumor SUV_{peak}. This result is in contrast to tumor response, in which the variation in intratumor response for different ROI_{peak} definitions (9%) was much larger than the variation in population average response (2%). For tumor response, there were no significant trends caused by size, shape, or location of ROI_{peak}, resulting in very little variation in population average response due to an averaging effect.

The wide variation in both intratumor response and SUV_{peak} stemmed from changes to the size, shape, and location of ROI_{peak}, reflecting the range of different ROI_{peak} definitions found in the literature. Therefore, a wide variation in intratumor response is expected under normal, realistic conditions. It is likely that an even greater variation would occur because of errors during image analysis for response assessment. For example, improper localization of ROI_{peak} in an average- or low-uptake region of a tumor at baseline would result in a measured tumor response that is artificially large, leading to a more extreme variation in intratumor response.

^{18}F-FLT, rather than ^{18}F-FDG, was selected as a radiotracer in this study because of the antiproliferative nature of the molecular targeted therapy. Furthermore, ^{18}F-FLT may be more effective for assessment of treatment response than is ^{18}F-FDG (*21,25,26*). However, imaging of tumors using both ^{18}F-FLT and ^{18}F-FDG has revealed a somewhat higher SUV and broader SUV range with ^{18}F-FDG than with ^{18}F-FLT (*17,27,28*). Thus, compared with ^{18}F-FLT, ^{18}F-FDG is expected to result in a greater variation in both SUV_{peak} and tumor response due to different ROI_{peaks}.

SUV_{peak} was determined using body weight (SUV^{BW}_{peak}) and not lean body mass (SUV^{LBM}_{peak}, as recommended by PERCIST). However, on average, patient weight changed only 1.5% between the 2 PET scans, and this weight change would result in an approximate difference of only 0.6% between response determined using SUV^{BW}_{peak} and SUV^{LBM}_{peak}. Consequently, in this study, approximately the same variation in SUV_{peak} and tumor response is expected using either SUV^{BW}_{peak} or SUV^{LBM}_{peak}.

Implications for Treatment Response Assessment

Currently, most response assessment studies use SUV_{max}, although recently SUV_{peak} has been recommended as a more robust alternative (*9*). Patient-specific response quantification is subject to significant uncertainty because of the different ROI_{peak} definitions, and therefore SUV_{peak} requires further study to optimize its use for quantification of response in individual patients. Though stemming from different causes, the uncertainties associated with SUV_{peak} and SUV_{max} are comparable (*6*). Moreover, the noise uncertainty associated with SUV_{max} continues to be reduced because of the increased counts with 3-dimensional PET acquisition, the current standard on most scanners. A correlation between SUV_{max} and SUV_{peak} responses has been demonstrated (*6,29*), and in this study, SUV_{max} response was within the range of responses quantified with SUV_{peak} in almost 70% of all tumors (Fig. 4). Nevertheless, despite this correlation, there can be substantial differences between SUV_{max} and SUV_{peak} responses in individual tumors. For example, response quantification using the PERCIST-recommended SUV_{peak} (1.25-cm-diameter sphere in highest-uptake region) was 45% smaller than that of SUV_{max} in tumor 9 (Fig. 4), resulting in different response categorizations. Such differences underscore the need to establish the relative predictive power of SUV_{peak} versus SUV_{max} for response assessment. Consequently, the recent recommendation in favor of SUV_{peak} over SUV_{max} should be approached with caution (*9*). It must first be determined whether SUV_{peak} or SUV_{max} is best suited for treatment response assessment.

Clinical trials are necessary to establish the superiority of SUV_{peak} or SUV_{max} for quantification of response to therapy. These trials should investigate the sensitivities of SUV_{peak} and SUV_{max} to a variety of factors, including image noise, scan acquisition and image reconstruction parameters, partial-volume effects, tumor motion, and others. Furthermore, the clinical utility of either SUV_{peak} or SUV_{max} for response quantification will strongly depend on its correlation with patients' clinical outcomes. Ultimately, the most robust and predictive SUV measure should be selected for quantification of treatment response.

It is probably not feasible to compare all definitions of SUV_{peak} with SUV_{max}, within the context of a larger clinical trial. Rather, a standard ROI_{peak} should be carefully selected to determine SUV_{peak}. ROI_{peak} should be large enough to prevent SUV_{peak} from suffering from noise, partial-volume effects, and other sensitivities that plague SUV_{max}. However, ROI_{peak} should not be so large that it includes substantial uptake heterogeneity and voxels that lie outside the tumor. These considerations lend support to the 1.2-cm-diameter sphere recommended by PERCIST as a standard definition of ROI_{peak} (for 2-cm or larger diameter tumors). This size is in the middle of the range of ROI_{peak} definitions found in the literature.

Identification of a suitable SUV measure for response quantification requires clinical trials. After these trials, thresholds for the different response categories (complete

response, partial response, stable disease, and progressive disease) can be established using population average response data in which the uncertainties are small. Unique thresholds may be established for specific diseases and their associated therapies. The size of the thresholds will need to exceed the overall uncertainty associated with the selected SUV measure (SUV_{peak} or SUV_{max}). Subsequently, the SUV measure could be quantified in individual patients to gauge their response to therapy using the established response thresholds as a guide.

CONCLUSION

Quantification of individual tumor response with SUV_{peak} is sensitive to the region of interest used to determine SUV_{peak}. Changes to the size, shape, and location of ROI_{peak} result in substantial variation ($\leq$50%) in both SUV_{peak} and tumor response for individual tumors. These considerable uncertainties in SUV_{peak} and tumor response call into question recommendations favoring SUV_{peak} over SUV_{max} for quantification of treatment response. Clinical trials are necessary to compare the efficacy of SUV_{peak} and SUV_{max} for quantification of response to therapy.

DISCLOSURE STATEMENT

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

ACKNOWLEDGMENTS

This work was financially supported by NIH grant R01 CA136927. We acknowledge the PET technologists Chris Jaskowiak and Mark McNall for scanning patients after hours as well as the University of Wisconsin Cyclotron Research Group for producing the ^{18}F-FLT used in the study. No other potential conflict of interest relevant to this article was reported.

REFERENCES

1. Weber WA. PET for response assessment in oncology: radiotherapy and chemotherapy. *Br J Radiol.* 2005;28(suppl):42–49.
2. Hawkins DS, Schuetze SM, Butrynski JE, et al. [^{18}F]fluorodeoxyglucose positron emission tomography predicts outcome for Ewing sarcoma family of tumors. *J Clin Oncol.* 2005;23:8828–8834.
3. Ott K, Weber WA, Lordick F, et al. Metabolic imaging predicts response, survival, and recurrence in adenocarcinomas of the esophagogastric junction. *J Clin Oncol.* 2006;24:4692–4698.
4. Stroobants S, Goeminne J, Seegers M, et al. ^{18}FDG-positron emission tomography for the early prediction of response in advanced soft tissue sarcoma treated with imatinib mesylate (Glivec). *Eur J Cancer.* 2003;39:2012–2020.
5. Boellaard R, Krak NC, Hoekstra OS, Lammertsma AA. Effects of noise, image resolution, and ROI definition on the accuracy of standard uptake values: a simulation study. *J Nucl Med.* 2004;45:1519–1527.
6. Krak NC, Boellaard R, Hoekstra OS, Twisk JW, Hoekstra CJ, Lammertsma AA. Effects of ROI definition and reconstruction method on quantitative outcome and applicability in a response monitoring trial. *Eur J Nucl Med Mol Imaging.* 2005;32:294–301.
7. Nahmias C, Wahl LM. Reproducibility of standardized uptake value measurements determined by ^{18}F-FDG PET in malignant tumors. *J Nucl Med.* 2008;49:1804–1808.
8. Nakamoto Y, Zasadny KR, Minn H, Wahl RL. Reproducibility of common semiquantitative parameters for evaluating lung cancer glucose metabolism with positron emission tomography using 2-deoxy-2-[^{18}F]fluoro-D-glucose. *Mol Imaging Biol.* 2002;4:171–178.
9. Wahl RL, Jacene H, Kasamon Y, Lodge MA. From RECIST to PERCIST: evolving considerations for PET response criteria in solid tumors. *J Nucl Med.* 2009;50(suppl 1):122S–150S.
10. Westerterp M, Pruim J, Oyen W, et al. Quantification of FDG PET studies using standardised uptake values in multi-centre trials: effects of image reconstruction, resolution and ROI definition parameters. *Eur J Nucl Med Mol Imaging.* 2007;34:392–404.
11. Brun E, Kjellen E, Tennvall J, et al. FDG PET studies during treatment: prediction of therapy outcome in head and neck squamous cell carcinoma. *Head Neck.* 2002;24:127–135.
12. Minn H, Zasadny KR, Quint LE, Wahl RL. Lung cancer: reproducibility of quantitative measurements for evaluating 2-[F-18]-fluoro-2-deoxy-D-glucose uptake at PET. *Radiology.* 1995;196:167–173.
13. Wahl RL, Zasadny K, Helvie M, Hutchins GD, Weber B, Cody R. Metabolic monitoring of breast cancer chemohormonotherapy using positron emission tomography: initial evaluation. *J Clin Oncol.* 1993;11:2101–2111.
14. Choi NC, Fischman AJ, Niemierko A, et al. Dose-response relationship between probability of pathologic tumor control and glucose metabolic rate measured with FDG PET after preoperative chemoradiotherapy in locally advanced non-small-cell lung cancer. *Int J Radiat Oncol Biol Phys.* 2002;54:1024–1035.
15. Römer W, Hanauske AR, Ziegler S, et al. Positron emission tomography in non-Hodgkin's lymphoma: assessment of chemotherapy with fluorodeoxyglucose. *Blood.* 1998;91:4464–4471.
16. Schelling M, Avril N, Nahrig J, et al. Positron emission tomography using [^{18}F] fluorodeoxyglucose for monitoring primary chemotherapy in breast cancer. *J Clin Oncol.* 2000;18:1689–1695.
17. Buck AK, Halter G, Schirrmeister H, et al. Imaging proliferation in lung tumors with PET: ^{18}F-FLT versus ^{18}F-FDG. *J Nucl Med.* 2003;44:1426–1431.
18. Chen W, Delaloye S, Silverman DH, et al. Predicting treatment response of malignant gliomas to bevacizumab and irinotecan by imaging proliferation with [^{18}F] fluorothymidine positron emission tomography: a pilot study. *J Clin Oncol.* 2007;25:4714–4721.
19. Herrmann K, Wieder HA, Buck AK, et al. Early response assessment using 3′-deoxy-3′-[^{18}F]fluorothymidine-positron emission tomography in high-grade non-Hodgkin's lymphoma. *Clin Cancer Res.* 2007;13:3552–3558.
20. Kenny L, Coombes RC, Vigushin DM, Al-Nahhas A, Shousha S, Aboagye EO. Imaging early changes in proliferation at 1 week post chemotherapy: a pilot study in breast cancer patients with 3′-deoxy-3′-[^{18}F]fluorothymidine positron emission tomography. *Eur J Nucl Med Mol Imaging.* 2007;34:1339–1347.
21. Pio BS, Park CK, Pietras R, et al. Usefulness of 3′-[F-18]fluoro-3′-deoxythymidine with positron emission tomography in predicting breast cancer response to therapy. *Mol Imaging Biol.* 2006;8:36–42.
22. Buck AK, Kratochwil C, Glatting G, et al. Early assessment of therapy response in malignant lymphoma with the thymidine analogue [^{18}F]FLT. *Eur J Nucl Med Mol Imaging.* 2007;34:1775–1782.
23. Lordick F, Ott K, Krause BJ, et al. PET to assess early metabolic response and to guide treatment of adenocarcinoma of the oesophagogastric junction: the MUNICON phase II trial. *Lancet Oncol.* 2007;8:797–805.
24. Martin SJ, Eisenbarth JA, Wagner-Utermann U, et al. A new precursor for the radiosynthesis of [^{18}F]FLT. *Nucl Med Biol.* 2002;29:263–273.
25. Dittmann H, Dohmen BM, Kehlbach R, et al. Early changes in [^{18}F]FLT uptake after chemotherapy: an experimental study. *Eur J Nucl Med Mol Imaging.* 2002;29:1462–1469.
26. Been LB, Suurmeijer AJ, Cobben DC, Jager PL, Hoekstra HJ, Elsinga PH. [^{18}F] FLT-PET in oncology: current status and opportunities. *Eur J Nucl Med Mol Imaging.* 2004;31:1659–1672.
27. Yap CS, Czernin J, Fishbein MC, et al. Evaluation of thoracic tumors with ^{18}F-fluorothymidine and ^{18}F-fluorodeoxyglucose-positron emission tomography. *Chest.* 2006;129:393–401.
28. Kasper B, Egerer G, Gronkowski M, et al. Functional diagnosis of residual lymphomas after radiochemotherapy with positron emission tomography comparing FDG- and FLT-PET. *Leuk Lymphoma.* 2007;48:746–753.
29. Yap J, Locascio T, Tanaka Y, Syrkin L, Van Den Abbeele A. Impact of variations in SUV methods for assessing cancer response using FDG-PET [abstract]. *J Nucl Med.* 2011;52(suppl 1):392P.