# Impact of Different Standardized Uptake Value Measures on PET-Based 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; and [2]Department of Radiology, Section of Nuclear Medicine, University of Wisconsin, Madison, Wisconsin*

PET-based treatment response studies typically measure the change in the standardized uptake value (SUV) to quantify response. The relative changes of different SUV measures, such as maximum, peak, mean, or total SUVs ($SUV_{max}$, $SUV_{peak}$, $SUV_{mean}$, or $SUV_{total}$, respectively), are used across the literature to classify patients into response categories, with quantitative thresholds separating the different categories. We investigated the impact of different SUV measures on the quantification and classification of PET-based treatment response. **Methods:** Sixteen 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), we acquired whole-body PET/CT scans at baseline, during treatment, and after treatment. The highest $^{18}$F-FLT uptake lesions (~2/patient) were segmented on PET images. Tumor PET response was assessed via the relative change in $SUV_{max}$, $SUV_{peak}$, $SUV_{mean}$, and $SUV_{total}$, thereby yielding 4 different responses for each tumor at mid- and posttreatment. For each SUV measure, a population average PET response was determined over all tumors. Standard deviation (SD) and range were used to quantify variation of PET response within individual tumors and population averages. **Results:** Different SUV measures resulted in substantial variation of individual tumor PET response assessments (average SD, 20%; average range, 40%). The most extreme variation between 4 PET response measures was 90% in individual tumors. Classification of tumor PET response depended strongly on the SUV measure, because different SUV measures resulted in conflicting categorizations of PET response (ambiguous treatment response assessment) in more than 80% of tumors. Variation of the population average PET response was considerably smaller (average SD, 7%; average range, 16%), and this variation was not statistically significant. Differences in tumor PET response were greatest between $SUV_{mean}$ and $SUV_{total}$ and smallest between $SUV_{max}$ and $SUV_{peak}$. Variations of tumor PET response at midtreatment and posttreatment were similar. **Conclusion:** Quantification and classification of PET-based treatment response in individual patients were strongly affected by the SUV measure used to assess response. This substantial uncertainty in individual patient PET response was present despite the concurrent robustness of the population average PET response. Given the ambiguity of individual patient PET responses, selection of PET-based treatment response measures and their associated thresholds should be carefully optimized.

**Key Words:** PET; SUV; treatment response quantification; $SUV_{max}$; $SUV_{peak}$; $SUV_{mean}$; $SUV_{total}$

**J Nucl Med 2013; 54:1188–1194**
DOI: 10.2967/jnumed.112.113332

PET-based treatment response assessment studies typically measure the relative change in the standardized uptake value (SUV) to quantify response. SUV is the ratio of the tissue radioactivity concentration to the total injected activity per patient mass, lean body mass, or body surface area. Most studies measure the change in either the maximum SUV ($SUV_{max}$) (*1–3*) or the mean SUV ($SUV_{mean}$) (*4,5*) of the tumor following the recommendations of the European Organization for Research and Treatment of Cancer (EORTC) (*6*) and the Cancer Imaging Program of the National Cancer Institute (*7*). Others quantify the change in the peak SUV ($SUV_{peak}$) (*8–10*) as suggested by Wahl et al. in the recent recommendations for PET Response Criteria in Solid Tumors (PERCIST) (*11*), the most comprehensive criteria to date. Fewer SUV-based response studies measure the change in the total SUV ($SUV_{total}$), which is associated with the total metabolic burden in $^{18}$F-FDG PET (*4,12,13*).

The relative change in SUV can be used to classify patients into different PET-based treatment response categories: PET complete response, PET partial response (PR), PET stable disease, and PET progressive disease (PD), with quantitative thresholds separating the different response categories (e.g., greater than 30% increase in $SUV_{peak}$ for PET PD using PERCIST). Such response classifications are often used to guide subsequent treatment decisions and can be predictive of clinical outcome (*1,14,15*).

The SUV measure used for treatment response assessment may significantly affect the quantification of PET response. Studies have revealed minimal differences in PET-based response quantification averaged over many patients using different SUV measures (*8,16*). However, these studies have not examined differences in the quantification of PET response within individual patients using multiple SUV measures. It has been already demonstrated that inconsistent definition of a single SUV measure ($SUV_{peak}$) results in substantial variation ($\leq$50%) of individual tumor PET response (*17*). Therefore, it is highly likely that multiple SUV measures could result in different quantifications and classifications of PET response. For example, a patient's response might be classified as PET PD using $SUV_{max}$ but as PET PR using $SUV_{mean}$, even though both measures are recommended by the EORTC. Such ambiguities could cast confusion on subsequent

Received Aug. 30, 2012; revision accepted Feb. 24, 2013.
For correspondence or reprints contact: Robert Jeraj, Department of Medical Physics, Wisconsin Institutes for Medical Research, 1111 Highland Ave., Rm. 1005, Madison, WI 53705-2275.
E-mail: rjeraj@wisc.edu
Published online Jun. 17, 2013.

treatment decisions. Furthermore, the quantitative thresholds governing PET-based response categorization may strongly depend on the SUV measure used to gauge response, with different thresholds applying to different measures. The sensitivity of PET-based response assessment to different SUV measures could have significant clinical implications regarding the use of PET for quantification of treatment response. Consequently, we investigated the impact of different SUV measures on quantification and classification of PET-based treatment response.

## MATERIALS AND METHODS

### Treatment and Imaging

Sixteen patients with advanced solid malignancies were treated with sunitinib malate (Sutent; Pfizer), a multitargeted receptor tyrosine kinase inhibitor. Sunitinib has been demonstrated to increase objective response rate and progression-free survival (PFS) in patients with renal cell carcinoma (*18*) and gastrointestinal stromal tumors (*19*) and has shown significant antitumor activity in patients with metastatic breast cancer (*20*), non–small cell lung cancer (*21*), and neuroendocrine tumors (*22*). Malignancies in this study included a diverse range of tumor types: renal cell carcinoma ($n = 7$), esophageal ($n = 2$), hepatocellular ($n = 2$), prostate ($n = 1$), sarcoma ($n = 1$), small cell lung ($n = 1$), thymus ($n = 1$), and uterine carcino-sarcoma ($n = 1$). 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 (*23–25*). Patients were injected intravenously with approximately 240 MBq (6.5 mCi) of $^{18}$F-FLT and underwent whole-body PET/CT scans at baseline (pretreatment), during treatment, and after 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 (*26*). PET/CT imaging began 47 ± 4 min after injection and extended inferiorly from the base of the skull to the distal femora. Acquisition mode was 2-dimensional, and acquisition time was 10 min per bed position to minimize image noise. 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, 5-mm gaussian loop (interiteration) filter, 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 (UW) Health Sciences Institutional Review Board, the Scientific Review Board of the UW Carbone Comprehensive Cancer Center, and the UW Radiation Drug Research Committee. All patients signed a written informed consent form before enrollment in the study.

### Quantification of Tumor PET Response

PET activity concentrations (MBq/mL) were converted to SUVs by normalizing by the decay-corrected injected activity per patient mass. $^{18}$F-FLT–avid lesions (~2/patient) were segmented on PET images by an experienced nuclear medicine physician. Lesion boundaries were delineated on transverse images where uptake level was visually elevated above background. These segmentations were used to generate a 3-dimensional volume of interest (VOI) for each lesion. The location and number of lesions were as follows: lung, 11; mediastinum, 5; liver, 3; abdomen, 3; adrenal, 1; gastrointestinal, 2; pelvis, 1; gluteus, 1; uterus, 1; and arm, 1. Tumor volumes ranged from 1 to 530 mL, with an average volume of 66 mL.

For an individual lesion ($n$), $SUV_{total}$, $SUV_{mean}$, and $SUV_{max}$ were defined as follows:

$$SUV^n_{total} = \sum_{k=1}^{K} SUV^n_k \quad \text{Eq. 1}$$

$$SUV^n_{mean} = \frac{\sum_{k=1}^{K} SUV^n_k}{K} \quad \text{Eq. 2}$$

$$SUV^n_{max} = \max\left\{SUV^n_1, SUV^n_2, \ldots SUV^n_k, \ldots SUV^n_K\right\}. \quad \text{Eq. 3}$$

Here, $n$ is an individual tumor, $SUV^n_k$ is the SUV of an individual voxel ($k$) within the physician-delineated tumor VOI, and $K$ is the total number of voxels in the VOI. $SUV^n_{peak}$ was defined as the average SUV within a 1 $cm^3$ sphere centered in the highest uptake region of the tumor (*11*). With $^{18}$F-FLT PET, $SUV^n_{total}$ represents total lesion proliferation, which is similar to total lesion glycolysis with $^{18}$F-FDG PET.

$SUV_{max}$, $SUV_{peak}$, $SUV_{mean}$, and $SUV_{total}$ were calculated for individual tumors. $SUV_{peak}$ was determined automatically using an in-house MATLAB (The MathWorks, Inc.) script that computed the average SUV within a 1 $cm^3$ sphere centered in the highest uptake region of the tumor VOI. $^{18}$F-FLT PET–based tumor proliferative responses at time point $t$ (mid- or posttreatment) were quantified by the change in each SUV measure relative to baseline (Eqs. 4–7).

$$(R(t))^n_{SUV_{max}} = \frac{SUV^n_{max}(t) - SUV^n_{max}(\text{baseline})}{SUV^n_{max}(\text{baseline})} \times 100\% \quad \text{Eq. 4}$$

$$(R(t))^n_{SUV_{peak}} = \frac{SUV^n_{peak}(t) - SUV^n_{peak}(\text{baseline})}{SUV^n_{peak}(\text{baseline})} \times 100\% \quad \text{Eq. 5}$$

$$(R(t))^n_{SUV_{mean}} = \frac{SUV^n_{mean}(t) - SUV^n_{mean}(\text{baseline})}{SUV^n_{mean}(\text{baseline})} \times 100\% \quad \text{Eq. 6}$$

$$(R(t))^n_{SUV_{total}} = \frac{SUV^n_{total}(t) - SUV^n_{total}(\text{baseline})}{SUV^n_{total}(\text{baseline})} \times 100\%. \quad \text{Eq. 7}$$

Here, $n$ is an individual tumor, $SUV^n_{measure}(\text{baseline})$ is the baseline value of an SUV measure, $SUV^n_{measure}(t)$ is the value of the SUV measure at time point $t$ (mid- or posttreatment), and $(R(t))^n_{SUV_{measure}}$ is the $^{18}$F-FLT PET–based tumor proliferative response associated with an SUV measure at time point $t$. Unless otherwise noted, PET response is used subsequently to refer to $^{18}$F-FLT PET–based proliferative response in this study.

The 4 different SUV measures ($SUV_{max}$, $SUV_{peak}$, $SUV_{mean}$, and $SUV_{total}$) yielded 4 different PET responses for each tumor at midtreatment and at posttreatment. At each time point, a mean PET response for each tumor was determined (mean intratumor PET response, Eq. 8), and the variation of the 4 PET responses about the mean PET response was quantified using SD and range.

$$\overline{(R(t))^n} = \frac{(R(t))^n_{SUV_{max}} + (R(t))^n_{SUV_{peak}} + (R(t))^n_{SUV_{mean}} + (R(t))^n_{SUV_{total}}}{4} \quad \text{Eq. 8}$$

Here, $n$ is an individual tumor and $\overline{(R(t))^n}$ is the mean intratumor PET response at time point $t$ (mid- or posttreatment).

In addition, a population average PET response (Eqs. 9–12) was determined for each SUV measure by averaging the PET responses of all tumors at midtreatment or at posttreatment.

$$\overline{(R(t))_{SUV_{max}}} = \frac{\sum_{n=1}^{N} (R(t))^n_{SUV_{max}}}{N} \quad \text{Eq. 9}$$

$$\overline{(R(t))_{SUV_{peak}}} = \frac{\sum_{n=1}^{N} (R(t))^n_{SUV_{peak}}}{N} \quad \text{Eq. 10}$$

$$\overline{(R(t))_{SUV_{mean}}} = \frac{\sum_{n=1}^{N} (R(t))^n_{SUV_{mean}}}{N} \quad \text{Eq. 11}$$

$$\overline{(R(t))_{SUV_{total}}} = \frac{\sum_{n=1}^{N} (R(t))^n_{SUV_{total}}}{N} \quad \text{Eq. 12}$$

Here, $n$ is an individual tumor, $N$ is the total number of tumors, and $\overline{(R(t))_{SUV_{measure}}}$ is the population average PET response associated with an SUV measure at time point $t$. Variation of the population average PET responses was measured using SD and range.

One-way ANOVA was used to test whether the changes in the different SUV measures resulted in statistically significant differences in tumor PET responses. Means were compared using Tukey honestly significant difference test. Differences were considered statistically significant at an α-level less than 0.05, after adjustment for multiple comparisons. Correlations between the variation of tumor PET response and other tumor characteristics were tested using the Pearson correlation coefficient ($r$) and considered statistically significant at an α-level less than 0.05.

### Association of PET Response with Clinical Endpoint

A Cox proportional hazards survival regression was used to associate the change in each SUV measure at each imaging time point (mid- and posttreatment) with the clinical endpoint, PFS. PFS was defined as the time to disease progression, either radiographic progression on CT or clinical progression of symptoms related to disease. Time to disease progression ranged from 2 to 22 mo, with a mean of 7.3 mo. Hazard ratio, covariate coefficient, and survivor function along with $\chi^2$ statistic and $P$ value were determined for each SUV measure at each imaging time point. Statistical significance was achieved at an α-level less than 0.05.

## RESULTS

### Individual Tumors

PET responses of individual tumors were sensitive to the SUV measure used to quantify the response. On average, different SUV measures resulted in substantial variation of individual tumor PET response (average SD, 20%; average range, 40%; Figs. 1–3). In individual tumors, the most extreme variation between SUV response measures was 90% (largest SD, 44%). On average, differences in tumor PET response were greatest between $SUV_{mean}$ and $SUV_{total}$ (average difference, 28%) and smallest between $SUV_{max}$ and $SUV_{peak}$ (average difference, 13%). Results at midtreatment and at posttreatment were similar (Figs. 2 and 3).

Variation of individual tumor PET response is highlighted for a uterine tumor in Figure 1. Pre- to midtreatment, all 4 SUV measures decreased by different amounts ($R_{SUV_{max}}$ = –48%; $R_{SUV_{peak}}$ = –35%; $R_{SUV_{mean}}$ = –55%; and $R_{SUV_{total}}$ = –77%) yet all SUV response measures indicated a PET PR (using PERCIST thresholds for different PET response categories). Pre- to posttreatment, there was wide variation associated with the changes of the different SUV measures. $SUV_{max}$ and $SUV_{peak}$ increased ($R_{SUV_{max}}$ = +42%; $R_{SUV_{peak}}$ = +53%) whereas $SUV_{mean}$ and $SUV_{total}$ decreased ($R_{SUV_{mean}}$ = –34%; $R_{SUV_{total}}$ = –22%), resulting in multiple PET response classifications of this uterine tumor. $SUV_{max}$ and $SUV_{peak}$ indicated PET PD, $SUV_{total}$ indicated PET stable disease, and $SUV_{mean}$ indicated PET PR. Similar ambiguous categorizations of tumor PET response arose in more than 80% of tumors assessed in this study (Fig. 2).

There was no significant correlation between tumor size and the variation of individual tumor PET response (Fig. 2, tumors ordered by size). Furthermore, there was no significant correlation between the degree of PET response (i.e., PET PD, PET stable disease, or PET PR) and the variation of individual tumor PET response.

For each SUV measure at each response time point, individual tumor PET responses were tested for strength of association with the clinical endpoint of PFS (Table 1). PET response determined posttreatment using $SUV_{total}$ ($R(\text{posttreatment})_{SUV_{total}}$) was significantly associated with PFS ($P$ = 0.046). In addition, PET response determined posttreatment using $SUV_{peak}$ ($R(\text{posttreatment})_{SUV_{peak}}$) was marginally associated with PFS ($P$ = 0.071). However, all other PET response measures failed to achieve statistically significant association with PFS. In general, posttreatment PET response was more strongly associated with PFS than midtreatment PET response.

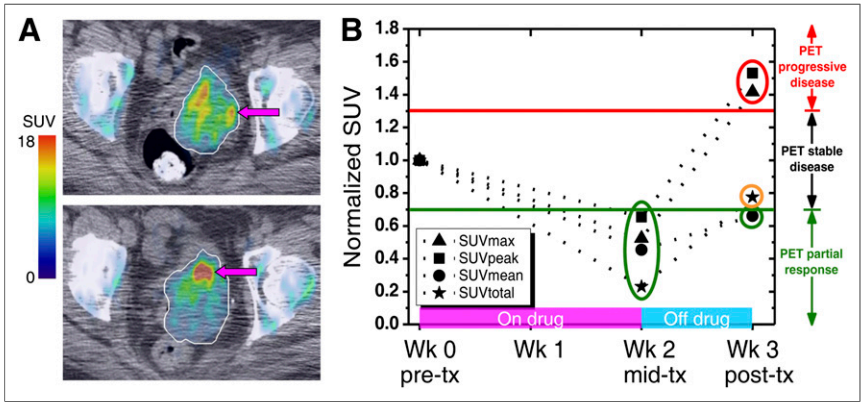


**FIGURE 1.** Variation of individual tumor PET response using different SUV measures. (A) $^{18}$F-FLT PET/CT images of uterine tumor (white outline), pretreatment (top), and posttreatment (bottom). After treatment, $SUV_{max}$ (arrows) increased by 40% whereas $SUV_{mean}$ decreased by 35%. (B) SUV measures (normalized to baseline) changed throughout therapy. Midtreatment, decreases of SUV measures varied but all measures indicated partial PET response (below –30%, green line). Posttreatment, wide variation of changes of SUV measures resulted in multiple, ambiguous PET response classifications, including PET PR, PET stable disease, and PET PD (above +30%, red line). mid-tx = midtreatment; post-tx = posttreatment; pre-tx = pretreatment; wk = week.

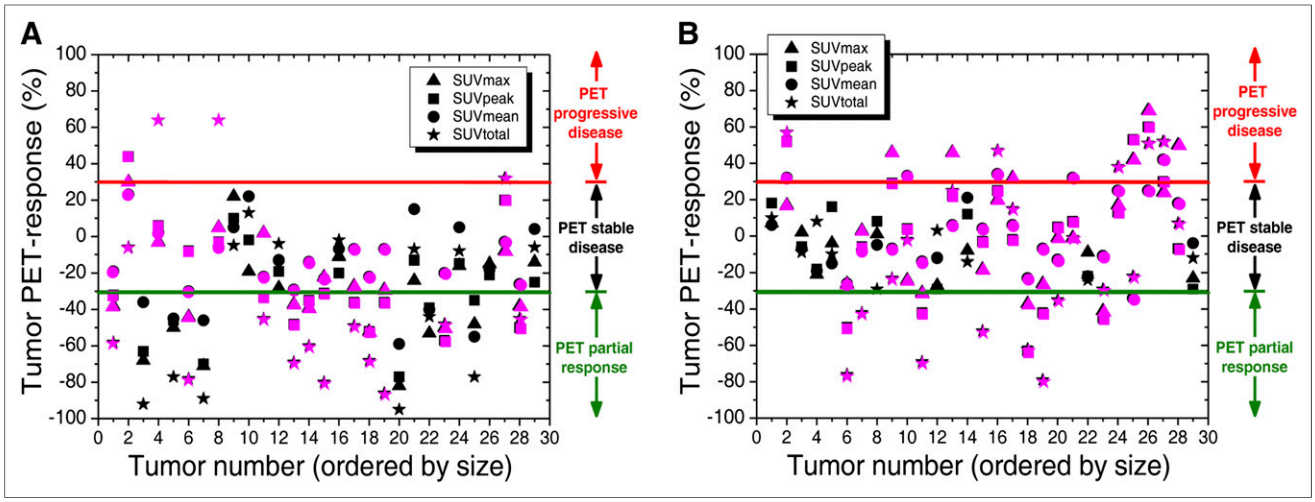


**FIGURE 2.** Variation of individual tumor PET responses pre- to midtreatment (A) and pre- to posttreatment (B) using different SUV measures. Substantial variation of PET response quantification resulted in classification of individual tumors into multiple response categories (green line, PET PR/PET stable disease cutoff, –30%; red line, PET PD/PET stable disease cutoff, +30%). Such ambiguous PET response categorization (magenta) occurred in more than 80% of tumors in this study.

### Population Average

The use of different SUV measures resulted in small variation of the population average PET response (average SD, 7%; average range, 16%; Figs. 3 and 4). There was slightly greater variation of the population average PET response at midtreatment (SD, 8%; range, 20%) than at posttreatment (SD, 6%; range, 12%). Differences in population average PET responses were greatest between $SUV_{mean}$ and $SUV_{total}$ (average difference, 16%) and smallest between $SUV_{max}$ and $SUV_{peak}$ (average difference, 2%). Differences between the populations of PET response associated with each SUV measure were not statistically significant.

The minimal variation of the population average PET response is shown in Figures 3 and 4. Pre- to midtreatment, all 4 SUV measures decreased by similar amounts ($\overline{R_{SUV_{max}}}$ = –28%; $\overline{R_{SUV_{peak}}}$ = –27%; $\overline{R_{SUV_{mean}}}$ = –16%; and $\overline{R_{SUV_{total}}}$ = –36%), and accordingly almost all measures classified the midtreatment population average PET response as PET stable disease ($SUV_{total}$ response fell slightly below the PET PR/PET stable disease cutoff). Pre- to posttreatment, changes of the different SUV measures varied even less ($\overline{R_{SUV_{max}}}$ = 3%; $\overline{R_{SUV_{peak}}}$ = 0%; $\overline{R_{SUV_{mean}}}$ = 2%; and $\overline{R_{SUV_{total}}}$ = –10%), and consequently all SUV response measures indicated PET stable disease.

## DISCUSSION

The SUV measure used to determine treatment response had a dramatic effect on the quantification of PET response. On average, different SUV measures caused a 20% variation of individual tumor PET response, and this variation ranged as high as 90%. Large variation can lead to different categorizations of PET response using established response criteria where fixed thresholds separate different PET response categories (e.g., EORTC response criteria (*6*) or PERCIST (*11*)). One such case is illustrated in Figure 1 where the posttreatment PET response (week 3) was classified either as PET PD, PET stable disease, or PET PR, depending on the SUV measure used to quantify the response. Such ambiguous PET response categorizations arose in more than 80% of the tumor PET responses assessed in this study (Fig. 2). These ambiguities remained using either the EORTC or PERCIST thresholds (which are slightly different) that

**TABLE 1**
Association of SUV Response Measures with PFS

| SUV response measure | Assessment time point | Hazard ratio | Covariate coefficient* | SE | $\chi^2$ statistic | *P* |
|---|---|---|---|---|---|---|
| $\Delta SUV_{max}$ | Midtx | 0.990 | −0.010 | 0.007 | 2.3 | 0.132 |
| $\Delta SUV_{peak}$ | Midtx | 0.991 | −0.009 | 0.007 | 1.6 | 0.205 |
| $\Delta SUV_{mean}$ | Midtx | 0.994 | −0.006 | 0.009 | 0.4 | 0.551 |
| $\Delta SUV_{total}$ | Midtx | 0.998 | −0.002 | 0.004 | 0.1 | 0.705 |
| $\Delta SUV_{max}$ | Posttx | 1.011 | 0.011 | 0.007 | 2.5 | 0.105 |
| $\Delta SUV_{peak}$ | Posttx | 1.011 | 0.011 | 0.006 | 3.3 | 0.073 |
| $\Delta SUV_{mean}$ | Posttx | 1.014 | 0.014 | 0.009 | 2.1 | 0.145 |
| $\Delta SUV_{total}$ | Posttx | 1.010 | 0.010 | 0.005 | 4.0 | 0.048 |

*Covariates are SUV response measures.
Δ = change; Midtx = midtreatment; Posttx = posttreatment.

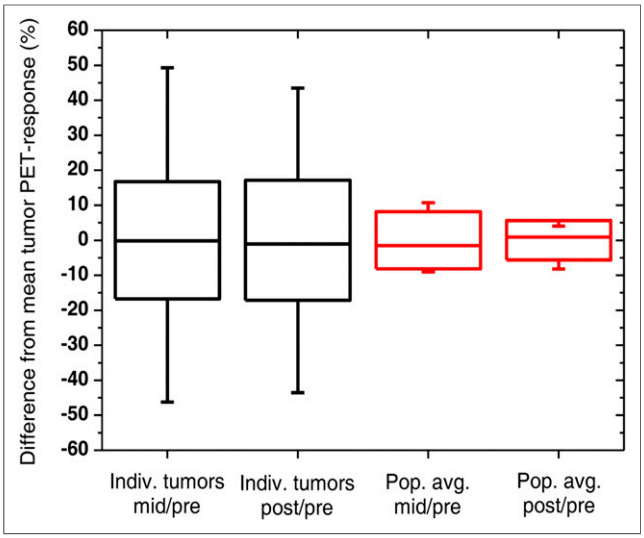


**FIGURE 3.** Variation of individual tumor PET responses (black) and population average PET responses (red) at mid- and posttreatment arising from different SUV measures. There is substantial variation of individual tumor PET response but much smaller variation of population average PET response, which is expected to be further reduced as more tumors are included in average. Boxes represent SD, whiskers show range, and solid lines depict median. Indiv = individual; mid = midtreatment; pop avg = population average; post = posttreatment; pre = pretreatment.

separate the different PET response categories. This sensitivity of PET response quantification to the SUV measure reveals the need to optimize PET imaging metrics for quantitative response assessment in individual patients.

Ambiguous PET-based treatment response categorization of individual tumors illustrates the shortcomings of relying on a single SUV measure to quantify response as well as the somewhat arbitrary thresholds to categorize the response. These problems are even more alarming because these PET response classifications are often used to guide subsequent treatment decisions. Ambiguous PET response assessment could muddle the intricate process of determining the need for further therapy. Assessment of treatment response using multiple SUV measures may offer a more complete characterization of response. Moreover, it is likely that some combination of SUV measures may provide a more comprehensive picture of treatment response and would be more informative and potentially more predictive of clinical outcome.

Differences in tumor PET response between $SUV_{mean}$ and $SUV_{total}$ were approximately twice as large as those between $SUV_{max}$ and $SUV_{peak}$. These larger differences are likely due to the inherent variability associated with manual VOI tumor delineation, even by an experienced nuclear medicine physician, which strongly affects $SUV_{mean}$ (normalized by tumor volume) and $SUV_{total}$ (integrated over tumor volume). Automated approaches to VOI tumor definition would reduce variability and improve the reproducibility and objectivity of tumor PET response using volume-sensitive metrics such as $SUV_{mean}$ and $SUV_{total}$. This improvement is illustrated by the smaller differences in tumor PET response between $SUV_{max}$ and $SUV_{peak}$, both of which were determined in an automated manner.

Variation of PET response within individual tumors is not surprising because different SUV measures assess different tumor characteristics. In PET imaging, $SUV_{max}$ and $SUV_{peak}$ measure the tumor region of most intense proliferation using $^{18}$F-FLT (or most intense metabolism with $^{18}$F-FDG) whereas $SUV_{mean}$ and $SUV_{total}$ assess overall proliferation in the tumor. Tumors tend to be heterogeneous so the average response of the entire tumor may be different from the response of one particular subregion. The uterine tumor in Figure 1 highlights this phenomenon. Posttreatment, $SUV_{mean}$ decreased by 35% whereas $SUV_{max}$ increased by 40%, implying that overall tumor proliferation decreased despite an increase in the most intense proliferative activity of the tumor. Using SUV, the heterogeneity and complexity of such responses can be captured only with multiple SUV measures or histograms of tumor voxel SUV. It is also quite possible that alternative, non-SUV measures may be better suited for PET-based assessment of treatment response (*27,28*). Furthermore, complex responses reveal the risk of relying on one or even multiple SUV measures for PET response assessment. Visual readings of PET examinations by trained nuclear medicine physicians are vital to fully understand treatment response. Physicians examine changes in tumor size, extent, uptake, and other characteristics that may support or contradict SUV-based response assessment.

Different SUV measures assess different tumor characteristics. Consequently, it is likely that each SUV measure will have its own unique threshold for PET response classification. For example, the PET PR/PET stable disease threshold for the change in $SUV_{max}$ may be different from that of $SUV_{peak}$. However, currently, the thresholds for PERCIST ($\pm$30% based on $SUV_{peak}$) and EORTC response criteria ($\pm$25% based on $SUV_{max}$ and $SUV_{mean}$) are quite similar even though these criteria use different SUV measures for response assessment. This study illustrates the danger of using a generic one-size-fits-all threshold for different SUV measures. Assessment of different aspects of the underlying tumor physiology will likely result in different response thresholds for different SUV measures. Furthermore, the study demonstrates that there are different uncertainties associated with different SUV metrics. For example, $SUV_{max}$ is a single pixel value that is adversely affected by image noise whereas $SUV_{mean}$ is quite sensitive to the delineation of tumor volume (*8,11,29–31*). PET response thresholds specific to each SUV measure must account for the sensitivity of each measure to uncertainties due to image noise, partial-volume effects, tumor motion, tumor contouring, and other scan acquisition and reconstruction parameters. Clearly, the unique test–retest repeatability and underlying tumor physiology associated with each SUV measure should both factor into SUV measure–specific response thresholds.

The considerable variation of quantification of PET response of individual tumors using different SUV measures underscores the pressing need for systematic selection of those measures that are most effective for assessment of treatment response. Ideally, these measures should be predictive of clinical outcome and robust to imaging uncertainties. As an example, SUV response measures in this study were correlated with a clinical endpoint using a Cox proportional hazards model. Despite small patient numbers, the posttreatment change in $SUV_{total}$ was identified as significantly associated with PFS. Larger clinical trials are necessary to establish the superiority of specific PET measures (SUV or non-SUV) for quantification of response to therapy. These trials should determine and compare the correlation of different PET response measures with clinical outcome. Combinations of PET response measures could also be explored to ascertain whether they offer improved predictive power over individual measures. Further-

more, these trials should investigate the sensitivities of these measures to a variety of imaging factors including image noise, scan acquisition and image reconstruction parameters, partial-volume effects, tumor motion, and others. Ultimately, the most predictive and robust PET measures (or combination of measures) should be selected for quantification of treatment response.

Unlike individual tumors, the population average PET response was relatively insensitive to the SUV measure used to quantify the response. On average, different SUV measures caused only a 7% variation in the population average PET response. This is consistent with the findings of Krak et al. and Yap et al. who demonstrated a high correlation of PET-based treatment responses using different SUV measures averaged over many tumors and patients (*8,16*). Because of an averaging effect, this variation is expected to be further reduced as more tumors are included in the population average. The minimal variation resulted in almost all SUV measures categorizing the population average PET response as PET stable disease at mid- and posttreatment (Fig. 4). This robustness of the population average highlights the strength of PET imaging for quantification of the average response to therapy. Using large numbers of patients, the population average PET response could be applied to establish clinically validated thresholds for more accurate response classification.

$^{18}$F-FLT, rather than $^{18}$F-FDG, was selected as a radiotracer in this study because of the antiproliferative nature of the molecule-targeted therapy. Furthermore, $^{18}$F-FLT may be more effective for PET-based assessment of treatment response than $^{18}$F-FDG (*32–34*). Imaging of tumors using both $^{18}$F-FLT and $^{18}$F-FDG has revealed somewhat higher SUV and broader SUV range with $^{18}$F-FDG than with $^{18}$F-FLT (*23,35,36*). Thus, compared with $^{18}$F-FLT, $^{18}$F-FDG is expected to result in similar if not greater variation of tumor PET response using different SUV measures.

PERCIST thresholds were applied to the $^{18}$F-FLT PET imaging response data in this study. However, PERCIST and EORTC response criteria are both based on $^{18}$F-FDG PET imaging studies.

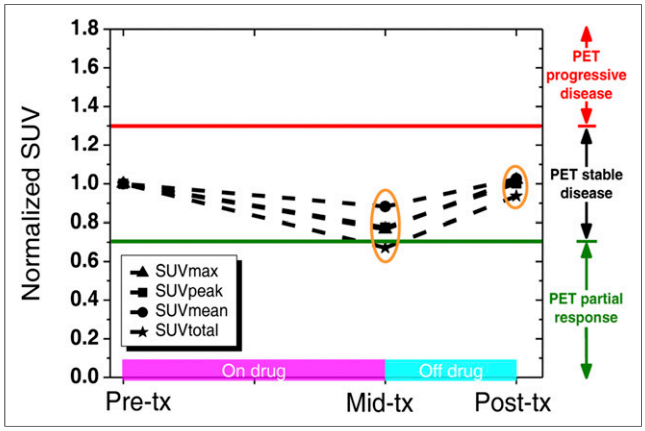


**FIGURE 4.** Quantification of population average PET response using different SUV measures. Both at mid- and posttreatment, there was minimal variation of changes of different SUV measures averaged over all tumors (normalized to baseline), and almost all measures indicated PET stable disease. PET response category thresholds are indicated by green line (–30%, PET PR/PET stable disease cutoff) and red line (+30%, PET PD/PET stable disease cutoff). mid-tx = midtreatment; post-tx = posttreatment; pre-tx = pretreatment.

The PET response thresholds (percentage change in SUV measure) are slightly more stringent (larger) for PERCIST than for EORTC to better account for the uncertainties and variability associated with PET imaging (*11,37*). Minimally, PET response thresholds (e.g., percentage ± 30% in PERCIST) must be greater than these uncertainties for PET response data to achieve a meaningful level of significance. These uncertainties plague PET imaging regardless of the specific radiotracer being imaged (*37*). Consequently, in this study, the PERCIST thresholds were applied to the $^{18}$F-FLT PET imaging response data to account for the associated uncertainties and variability. Furthermore, uncertainties are likely to be similar for $^{18}$F-FDG PET and $^{18}$F-FLT PET because the $^{18}$F radionuclide is common to both radiotracers. In addition, PET response thresholds of ±30% (as in PERCIST) are supported by a variety of other $^{18}$F-FLT PET–based response assessment studies (*25,38,39*). Ultimately, future and more refined PET response criteria may depend on the specific response metric, disease, radiotracer, imaging time point, and other relevant factors.

In this study, all SUV measures were determined using body weight ($SUV^{BW}$) and not lean body mass ($SUV^{LBM}$, recommended by PERCIST). However, on average, patient weight changed only 1.5% among the 3 PET scans, which would result in approximately 0.6% difference between PET response determined using $SUV^{BW}$ and $SUV^{LBM}$. Consequently, in this study, approximately the same variation of tumor PET response is expected using either $SUV^{BW}$ or $SUV^{LBM}$.

## CONCLUSION

PET-based quantification of treatment response was affected substantially by the SUV measure used to assess response. Different SUV measures resulted in a 20% variation of individual tumor PET response, and this variation ranged as high as 90%. Consequently, classification of individual tumor PET response strongly depended on the SUV measure, because different SUV measures resulted in different categorizations of response in more than 80% of tumors. This substantial uncertainty in individual patient PET response was present despite the concurrent robustness of the population average PET response. Given these uncertainties, PET-based quantification of treatment response should be optimized for accurate response assessment in individual patients. Clinical trials are necessary to select the most predictive, robust SUV measures (or combinations of measures) and associated response thresholds that should be used for assessment of treatment response.

## DISCLOSURE

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. This work was financially supported by NIH grant R01 CA136927. No other potential conflict of interest relevant to this article was reported.

## ACKNOWLEDGMENTS

We acknowledge the PET technologists Chris Jaskowiak and Mark McNall for scanning patients after hours as well as the UW Cyclotron Research Group for producing the $^{18}$F-FLT used in the study.

## REFERENCES


1. 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.
2. Hutchings M, Loft A, Hansen M, et al. FDG-PET after two cycles of chemotherapy predicts treatment failure and progression-free survival in Hodgkin lymphoma. *Blood.* 2006;107:52–59.
3. Kidd EA, Siegel BA, Dehdashti F, Grigsby PW. The standardized uptake value for F-18 fluorodeoxyglucose is a sensitive predictive biomarker for cervical cancer treatment response and survival. *Cancer.* 2007;110:1738–1744.
4. Benz MR, Allen-Auerbach MS, Eilber FC, et al. Combined assessment of metabolic and volumetric changes for assessment of tumor response in patients with soft-tissue sarcomas. *J Nucl Med.* 2008;49:1579–1584.
5. Tiling R, Linke R, Untch M, et al. $^{18}$F-FDG PET and $^{99m}$Tc-sestamibi scintimammography for monitoring breast cancer response to neoadjuvant chemotherapy: a comparative study. *Eur J Nucl Med.* 2001;28:711–720.
6. Young H, Baum R, Cremerius U, et al. Measurement of clinical and subclinical tumour response using [$^{18}$F]-fluorodeoxyglucose and positron emission tomography: review and 1999 EORTC recommendations. European Organization for Research and Treatment of Cancer (EORTC) PET Study Group. *Eur J Cancer.* 1999;35:1773–1782.
7. Shankar LK, Hoffman JM, Bacharach S, et al. Consensus recommendations for the use of $^{18}$F-FDG PET as an indicator of therapeutic response in patients in National Cancer Institute Trials. *J Nucl Med.* 2006;47:1059–1066.
8. 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.
9. 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.
10. 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.
11. 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.
12. Campbell JM, Wong CO, Muzik O, Marples B, Joiner M, Burmeister J. Early dose response to yttrium-90 microsphere treatment of metastatic liver cancer by a patient-specific method using single photon emission computed tomography and positron emission tomography. *Int J Radiat Oncol Biol Phys.* 2009;74:313–320.
13. Hong D, Lunagomez S, Kim EE, et al. Value of baseline positron emission tomography for predicting overall survival in patient with nonmetastatic esophageal or gastroesophageal junction carcinoma. *Cancer.* 2005;104:1620–1626.
14. 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.
15. 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.
16. 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):1767.
17. 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.
18. Motzer RJ, Hutson TE, Tomczak P, et al. Sunitinib versus interferon alfa in metastatic renal-cell carcinoma. *N Engl J Med.* 2007;356:115–124.
19. Demetri GD, van Oosterom AT, Garrett CR, et al. Efficacy and safety of sunitinib in patients with advanced gastrointestinal stromal tumour after failure of imatinib: a randomised controlled trial. *Lancet.* 2006;368:1329–1338.
20. Burstein HJ, Elias AD, Rugo HS, et al. Phase II study of sunitinib malate, an oral multitargeted tyrosine kinase inhibitor, in patients with metastatic breast cancer previously treated with an anthracycline and a taxane. *J Clin Oncol.* 2008;26:1810–1816.
21. Socinski MA, Novello S, Brahmer JR, et al. Multicenter, phase II trial of sunitinib in previously treated, advanced non-small-cell lung cancer. *J Clin Oncol.* 2008;26:650–656.
22. Kulke MH, Lenz HJ, Meropol NJ, et al. Activity of sunitinib in patients with advanced neuroendocrine tumors. *J Clin Oncol.* 2008;26:3403–3410.
23. 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.
24. 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.
25. 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.
26. 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.
27. El Naqa I, Grigsby P, Apte A, et al. Exploring feature-based approaches in PET images for predicting cancer treatment outcomes. *Pattern Recognit.* 2009;42:1162–1171.
28. Galavis PE, Hollensen C, Jallow N, Paliwal B, Jeraj R. Variability of textural features in FDG PET images due to different acquisition modes and reconstruction parameters. *Acta Oncol.* 2010;49:1012–1016.
29. 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.
30. 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.
31. 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.
32. 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.
33. 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.
34. 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.
35. 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.
36. 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.
37. Boellaard R. Standards for PET image acquisition and quantitative data analysis. *J Nucl Med.* 2009;50(suppl 1):11S–20S.
38. Sohn HJ, Yang YJ, Ryu JS, et al. [$^{18}$F]fluorothymidine positron emission tomography before and 7 days after gefitinib treatment predicts response in patients with advanced adenocarcinoma of the lung. *Clin Cancer Res.* 2008;14:7423–7429.
39. Zander T, Scheffler M, Nogova L, et al. Early prediction of nonprogression in advanced non-small-cell lung cancer treated with erlotinib using [$^{18}$F]fluorodeoxyglucose and [$^{18}$F]fluorothymidine positron emission tomography. *J Clin Oncol.* 2011;29:1701–1708.