Combined Assessment of Metabolic and Volumetric Changes for Assessment of Tumor Response in Patients with Soft-Tissue Sarcomas

Matthias R. Benz1, Martin S. Allen-Auerbach, Fritz C. Eilber1,2, Hui J.J. Chen3, Sarah Dry4, Michael E. Phelps4, Johannes Czernin1, and Wolfgang A. Weber1,5

1Department of Molecular and Medical Pharmacology, University of California Los Angeles, Los Angeles, California; 2Division of Surgical Oncology, University of California Los Angeles, Los Angeles, California; 3Department of Radiology, University of California Los Angeles, Los Angeles, California; 4Department of Pathology, University of California Los Angeles, Los Angeles, California; and 5Abteilung Nuklearmedizin, University of Freiburg, Freiburg, Germany

By allowing simultaneous measurements of tumor volume and metabolic activity, integrated PET/CT opens up new approaches for assessing tumor response to therapy. The aim of this study was to determine whether combined assessment of tumor volume and metabolic activity improves the accuracy of 18F-FDG PET for predicting histopathologic tumor response in patients with soft-tissue sarcomas. Methods: Twenty patients with locally advanced high-grade soft-tissue sarcoma (10 men and 10 women; mean age, 49 ± 17 y) were studied by 18F-FDG PET/CT before and after preoperative therapy. CT tumor volume (CTvol) was measured by delineating tumor borders on consecutive slices of the CT scan. Mean and maximum 18F-FDG standardized uptake value (SUVmax; single maximum pixel value) were determined. Two indices of total lesion glycolysis (TLG) were calculated by multiplying tumor volume by SUVmean and SUVmax, respectively) were determined. Two indices of total lesion glycolysis (TLG) were calculated by multiplying tumor volume by SUVmean (TLGmean) and SUVmax (TLGmax). Changes in CTvol, SUVmean, SUVmax, TLGmean, and TLGmax after chemotherapy were correlated with histopathologic tumor response (≥95% treatment-induced tumor necrosis). Accuracy for predicting histopathologic response was compared by receiver-operating-characteristic (ROC) curve analysis. Results: Baseline SUVmax, SUVmean, CTvol, TLGmean, and TLGmax were 11.22 g/mL, 2.84 g/mL, 544.1 mL, 1,619.8 g, and 8852.9 g, respectively. After neoadjuvant therapy, all parameters except CTvol showed a significant decline (ΔSUVmax = −51%, P < 0.001; ΔSUVmean = −40%, P < 0.001; ΔCTvol = −14%, P = 0.37; ΔTLGmean = −44%, P = 0.006; and ΔTLGmax = −54%, P = 0.001). SUV changes in histopathologic responders (n = 6) were significantly more pronounced than those in nonresponders (n = 14) (P = 0.001). Histopathologic response was well predicted by changes in SUVmean and SUVmax (area under ROC curve [AUC] = 1.0 and 0.98, respectively) followed by TLGmean (AUC = 0.77) and TLGmax (AUC = 0.74). In contrast, changes in CTvol did not allow prediction of treatment response (AUC = 0.48). Conclusion: In this population of patients with sarcoma, TLG was less accurate in predicting tumor response than were measurements of the intratumoral 18F-FDG concentration (SUVmax, SUVmean). Further evaluation of TLG in larger patient populations and other tumor types is necessary to determine the value of this conceptually attractive parameter for assessing tumor response.

Key Words: sarcoma; PET/CT; 18F-FDG; treatment monitoring; multimodality imaging

DOI: 10.2967/jnumed.108.053694

Several years ago, Larson et al. (1) proposed a semi-quantitative index of total tumor lesion glycolysis (TLG) for better describing tumor glucose use and their changes in response to treatment. This index combined volumetric tumor assessments with tumor 18F-FDG uptake. At that time PET/CT (2) was not routinely available, and the investigators derived both tumor volumes and average SUVs from whole-body PET images.

With the advent of PET/CT, changes in 18F-FDG uptake can now be determined accurately within an anatomically well-defined volume, which allows for establishing a potentially more robust index of TLG. The potential advantage of such an index is that volume measurements could be combined with glucose metabolic assessments of tumor responses.

Whether such measurements of TLG could provide a better readout of treatment effects than could maximum standardized uptake value (SUVmax; single maximum pixel value) (3,4) or mean SUV of a defined volume of interest (SUVmean) (5,6) is currently unknown. Intuitively, the TLG parameter is attractive, because it uses much more of the PET/CT information than does the intensity of 18F-FDG uptake in a single voxel.

The aim of this study was to determine whether combined assessment of volumetric and metabolic measurements (TLG) is superior to individual volumetric or metabolic measure-
ments for predicting histopathologic tumor response in patients with high-grade soft-tissue sarcomas. We selected high-grade soft-tissue sarcomas for this study because these tumors are generally large at diagnosis and demonstrate a high intratumoral heterogeneity in $^{18}$F-FDG uptake. Therefore, we hypothesized that maximum tumor $^{18}$F-FDG uptake, mean tumor $^{18}$F-FDG uptake, and TLG should differ in their ability to assess tumor response to therapy.

One motivation behind the current study was the need for standardizing quantitative PET measurements for treatment response assessments in multicenter trials and in clinical practice. Changes in $^{18}$F-FDG SUVs correlate well with tumor responses and patient survival across a variety of tumors (3,7–9). However, standardization of response assessment within and among institutions is lacking. Standardized metabolic response criteria have only been established and accepted in lymphoma (10). Thus, despite increasing evidence of the usefulness of $^{18}$F-FDG PET for treatment monitoring, treatment responses are still monitored almost exclusively by measuring changes in tumor size on CT.

**MATERIALS AND METHODS**

This study included patients with soft-tissue sarcoma who were enrolled prospectively at UCLA. Inclusion criteria were: age greater than 18 y, biopsy-proven soft-tissue sarcoma, a whole-body PET/CT study within 4 wk before the initiation of neoadjuvant therapy (baseline study), a second whole-body PET/CT study after completion of neoadjuvant therapy (posttreatment study), and tumor resection and histologic examination after the posttreatment PET/CT study. To ascertain that CT measurements are reliable and that SUV measurements are not subject to image noise, only patients with well-defined tumor borders on CT and a baseline SUVmax of greater than 2.5 were included. Patients with blood glucose levels of greater than 150 mg/dL were excluded (11). On the basis of these criteria, 20 patients were eligible for this analysis. Patient demographics are provided in Table 1.

Briefly, the study population consisted of 10 men and 10 women, with a mean age of 48.9 ± 17.4 y (range, 19–86 y). The most common site of disease was the extremity (n = 16, 80%). Six (30%) patients received systemic chemotherapy (ifosfamide/doxorubicin or gemcitabine/docetaxel) alone. Fourteen patients (70%) underwent additional external-beam radiation.

All participants gave written informed consent to participate in this study, which was approved by the UCLA Institutional Review Board.

**PET/CT Image Acquisition**

Patients were instructed to fast for at least 6 h before $^{18}$F-FDG PET to standardize blood glucose and insulin levels. Blood glucose levels were measured before the injection of $^{18}$F-FDG. All PET/CT studies were performed on a PET/CT scanner (Biograph Duo; Siemens) comprising an ECAT ACCEL (Siemens) and a dual-detector helical CT scanner. Images were acquired as a single whole-body protocol. For CT, all patients received 700–900 mL of oral contrast (Readi-Cat 2; EZEM) 1 h before the study. Intravenous contrast (iohexol [Omnipaque]; GE Healthcare) was administered to all patients at a rate of 2 mL/s, 30–40 s before imaging commenced. The CT acquisition parameters were 130 kVp, 120 mAs, 1-s rotation, 4-mm slice collimation, 1.3 pitch, and 8 mm/s bed speed. Patients were injected intravenously with 7.77 MBq/kg (0.21 mCi/kg) of $^{18}$F-FDG 77 ± 8.7 min before image acquisition. The PET emission scan duration per bed position was 1–5 min, depending on the patient’s body weight as previously described (12,13).

To minimize misregistration between the CT and the PET images, patients were instructed to use shallow breathing during the PET/CT image acquisition (14). The CT images were reconstructed using conventional filtered backprojection, at 3.4-mm axial intervals to match the slice separation of the PET data. PET images were reconstructed using iterative algorithms (ordered-subset expectation maximization, 2 iterations, 8 subsets) to a final image resolution of 8.8-mm full width at half maximum. A 5-mm full width at half maximum gaussian filter was applied after reconstruction.

To correct for the photon attenuation in PET, the previously published CT-based algorithm was applied (15).

**Image Analysis**

PET/CT scans were analyzed quantitatively by 1 observer unaware of clinical data and histopathologic response. Baseline and posttreatment PET/CT images were analyzed as follows. A soft-tissue CT window was used to display tumor images on a PET/CT display station (Mirada Workstation; Mirada) for delineation of the sarcomas. Because all tumors were well delineated on CT, tumor volumes were defined on the CT images without using the PET information. In contrast, tumors were frequently difficult to delineate on the posttherapeutic PET images. First, the tumor borders were manually delineated on consecutive axial slices of the CT scan (Fig. 1). Using these regions of interest, the following parameters were calculated: tumor volume in milliliters (CTvol), maximum $^{18}$F-FDG uptake (SUVmax) within this volume, and mean $^{18}$F-FDG uptake (SUVmean) within this volume. Finally, the TLG index was calculated as $\text{CTvol} \times \text{SUVmax}$ (TLGmean) and CTvol × SUVmax (TLGmax).

**Histopathology**

All specimens were analyzed in a standardized fashion as previously described (16), with masking of PET and CT data. Each specimen was bisected along the greatest diameter, and the

### Table 1

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>n = 20</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (y)</td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>50</td>
</tr>
<tr>
<td>Range</td>
<td>19–86</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>10 (50%)</td>
</tr>
<tr>
<td>Female</td>
<td>10 (50%)</td>
</tr>
<tr>
<td>Site</td>
<td></td>
</tr>
<tr>
<td>Extremity</td>
<td>16 (80%)</td>
</tr>
<tr>
<td>Retroperitoneum/abdomen</td>
<td>1 (5%)</td>
</tr>
<tr>
<td>Chest/trunk</td>
<td>3 (15%)</td>
</tr>
<tr>
<td>Neoadjuvant therapy</td>
<td></td>
</tr>
<tr>
<td>Chemotherapy</td>
<td>6 (30%)</td>
</tr>
<tr>
<td>Chemo-/radiotherapy</td>
<td>14 (70%)</td>
</tr>
<tr>
<td>Pathologic necrosis</td>
<td></td>
</tr>
<tr>
<td>≥95% (responder)</td>
<td>6 (30%)</td>
</tr>
<tr>
<td>&lt;95% (nonresponder)</td>
<td>14 (70%)</td>
</tr>
</tbody>
</table>

Table 1: Clinical, Pathologic, and Treatment Characteristics

- **Clinical, Pathologic, and Treatment Characteristics**: Age (y) (range 19–86 y); Sex (male/female 10/10); Site (extremity, retroperitoneum/abdomen, chest/trunk); Neoadjuvant therapy (chemotherapy, chemo-/radiotherapy); Pathologic necrosis (≥95% responder, <95% nonresponder)

- **PET/CT Image Acquisition**: Patients instructed to fast for at least 6 h before $^{18}$F-FDG PET; Blood glucose levels measured before injection; PET/CT scanner (Biograph Duo; Siemens) comprising an ECAT ACCEL (Siemens) and a dual-detector helical CT scanner.

- **Image Analysis**: PET/CT scans analyzed quantitatively by 1 observer; Parameters calculated: tumor volume (CTvol), maximum $^{18}$F-FDG uptake (SUVmax), mean $^{18}$F-FDG uptake (SUVmean), TLGmean, and TLGmax.

- **Histopathology**: Specimens analyzed in standardized fashion; Masking of PET and CT data; Specimen bisected along greatest diameter.
perimeter of the tumor was defined. Along with additional randomly sampled areas, the entire cross-sectional area of the bisected tumor was partitioned into 2.0-cm² (average) blocks and processed for histologic examination. Histopathologic response to therapy was assessed in a manner previously established for soft-tissue sarcomas and quantified as percentage of pathologic necrosis (16). For the purpose of this study, patients with no less than 95% pathologic necrosis (<5% viable tumor cells) were classified as histopathologic responders as previously described (16).

Statistical Analysis
Quantitative data for tumor volume, SUVmean, SUVmax, and TLG are expressed as mean ± 1 SD and range. Intra- and inter-individual comparisons of absolute values and changes were made by the Wilcoxon signed rank and the Mann–Whitney tests, respectively. The SD in the treatment-induced changes of the various studied parameters describes the SD across patients. To determine the diagnostic accuracy of ¹⁸F-FDG PET and CT in identifying responders, receiver-operating-characteristic (ROC) curve analysis was used (17). Histopathology was used as the reference standard. The area under the ROC curve (AUC) provides a measure for the accuracy of a diagnostic test. It ranges from 0.5 (random guessing) to 1.0 (perfect test).

RESULTS
Histopathology
On the basis of excised tumor tissue, 6 patients (30%) were classified as histopathologic responders (<95% necrosis). The mean percentage tumor necrosis was 62.9% ± 30.7%, ranging from 10% to 99%.

PET/CT Findings
The imaging findings are listed in Tables 2 and 3. Baseline SUVmax, SUVmean, CTvol, TLGmean, and TLGmax averaged 11.22 ± 8.23 g/mL, 2.84 ± 1.46 g/mL, 544.14 ± 543.50 mL, 1,619.77 ± 1,837.03 g, and 8,852.85 ± 14,103.66 g, respectively.

After therapy, these parameters decreased to 4.58 ± 3.34* (1.24–14.90) g/mL, 1.52 ± 0.74* (0.43–3.03) g/mL, 532.58 ± 715.35 mL, 1,619.77 ± 1,837.03 g, and 8,852.85 ± 14,103.66 g, respectively. Changes in SUVmax (P < 0.001), SUVmean (P < 0.001), TLGmean

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Baseline study</th>
<th>Posttreatment study</th>
<th>Changes (%)</th>
<th>Responder</th>
<th>Nonresponder</th>
</tr>
</thead>
<tbody>
<tr>
<td>SUVmax</td>
<td>11.22 ± 8.23 (2.52–31.08)</td>
<td>4.58 ± 3.34* (1.24–14.90)</td>
<td>−77.9 ± 6.9</td>
<td>−39.5 ± 20.2*</td>
<td></td>
</tr>
<tr>
<td>SUVmean</td>
<td>2.84 ± 1.46 (1.05–6.72)</td>
<td>1.52 ± 0.74* (0.43–3.03)</td>
<td>−70.3 ± 8.5</td>
<td>−6.9 ± 19.5*</td>
<td></td>
</tr>
<tr>
<td>CTvol</td>
<td>544.14 ± 543.50 (18.60–1,913.60)</td>
<td>532.58 ± 715.35 (8.7–2,663.10)</td>
<td>−7.5 ± 49.5</td>
<td>−16.3 ± 48.0</td>
<td></td>
</tr>
<tr>
<td>TLGmean</td>
<td>1,619.77 ± 1,837.03 (39.80–6,815.65)</td>
<td>704.37 ± 631.80* (10.53–2,978.30)</td>
<td>−70.5 ± 23.3</td>
<td>−32.7 ± 48.1</td>
<td></td>
</tr>
<tr>
<td>TLGmax</td>
<td>8,852.85 ± 14,103.66 (58.96–5,9474.69)</td>
<td>2,625.82 ± 3,585.97* (16.27–14,316.82)</td>
<td>−77.6 ± 18.6</td>
<td>−43.2 ± 44.9</td>
<td></td>
</tr>
</tbody>
</table>

*P values < 0.05.
Data presented are mean ± 1 SD, with range in parentheses.

FIGURE 1. Contrast-enhanced CT images acquired before and after treatment are shown (axial cuts at level of proximal thigh [top] and coronal images [bottom]). (A) Patient with soft-tissue sarcoma located in right proximal thigh (arrow). (B) Regions of interest were placed manually around all axial cuts that included tumor on baseline and follow-up scans. (C) This volume of interest was then used to determine SUVmax (1) and SUVmean (2) within tumor volume.
The data were then stratified by histopathologic response (Fig. 2). Changes in SUVmax and SUVmean (P = 0.001) were significantly more pronounced in responders than in nonresponders. In contrast, TLGmean (P = 0.06), TLGmax (P = 0.09), and CTvol (P = 0.87) changes did not differ significantly between responders and nonresponders. Changes in all measured parameters are listed in Table 2.

**ROC Curve Analysis**

SUVmean and SUVmax predicted treatment responses (AUC = 1.0 and 0.98, respectively) more accurately than did TLGmean (AUC = 0.77) and TLGmax (AUC = 0.74). CTvol failed to predict treatment responses, with an AUC of 0.48 (Fig. 3).

**Correlation Between SUVmax and SUVmean**

A high correlation was found for changes in the single maximum pixel value (SUVmax) and changes in the global tumor 18F-FDG uptake (SUVmean) from the baseline to the posttreatment scan (Pearson correlation coefficient = 0.92, P < 0.001; Fig. 4). In contrast, the correlation between SUVmax and SUVmean at the pretreatment and posttreatment scans was only 0.598 and 0.579, respectively.

**DISCUSSION**

In this group of patients with high-grade soft-tissue sarcoma, changes in SUVmax and SUVmean were superior to those of TLG for predicting histopathologic tumor responses. Changes in CT-based tumor volumes failed to discriminate between responders and nonresponders. These findings have several implications. First, both SUVmean and SUVmax are well suited to identify responders to neoadjuvant therapy among patients with sarcoma. Second, changes in tumor volume by CT were not useful for identifying treatment responders, casting more doubts on the usefulness of response evaluation criteria in solid tumors (RECIST) (18) in sarcoma treatment evaluations.
As a consequence, combining CTvol changes with lesion glycolysis diminished the response information that can be derived from glucose metabolic imaging alone.

A second unexpected finding of our study was that SUVmean and SUVmax provided a similar accuracy for differentiation of histopathologically responding and nonresponding tumors. The average tumor volume in this study was more than 500 mL. As shown in Figure 1, the intensity of $^{18}$F-FDG uptake within these large tumor volumes was highly heterogenous. In fact, the poor correlation between SUVmax and SUVmean at the time of the baseline scan and at the posttreatment scan reflects the heterogeneity of these tumors at both time points. Nevertheless, changes in SUVmax and SUVmean from the baseline to the follow-up scan were closely correlated (Pearson correlation coefficient $= 0.92; P < 0.001$). This correlation suggests that neoadjuvant therapy has led to similar relative changes in all voxels. Therefore, the use of SUVmax to assess treatment response is as accurate as measurements of changes of the average $^{18}$F-FDG uptake (SUVmean).

The assessment of tumor response to neoadjuvant chemotherapy or chemoradiotherapy by anatomic imaging techniques remains challenging. RECIST, which define partial response as a 30% decrease in the sum of the largest diameter of the target lesions, have failed in several cancers to assess treatment responses accurately (19–21). This failure of CT-based tumor size measurements to separate reliably between responding and nonresponding tumors may be caused by, among others, necrosis, edema, and intratumoral hemorrhage, all of which present as mass and cannot reliably be differentiated from viable tumors. Thus, on the basis of the lack of significant size changes tumors that are actually responding may be classified incorrectly as nonresponding. As evidenced in the current and other studies, this problem is not limited to monitoring the effects

![ROC curve for assessment of histopathologic response by changes in SUVmean, SUVmax, TLGmean, TLGmax, and CTvol.](image)

**FIGURE 3.** ROC curve for assessment of histopathologic response by changes in SUVmean, SUVmax, TLGmean, TLGmax, and CTvol.

![Graphs show correlations between SUVmax and SUVmean at baseline (A) and at follow-up (B) (Pearson correlation coefficient $= 0.598$ and 0.579, respectively). (C) Relative changes in SUVmax and SUVmean from baseline to posttreatment scan were closely correlated (Pearson correlation coefficient $= 0.92$).](image)

**FIGURE 4.** Graphs show correlations between SUVmax and SUVmean at baseline (A) and at follow-up (B) (Pearson correlation coefficient $= 0.598$ and 0.579, respectively). (C) Relative changes in SUVmax and SUVmean from baseline to posttreatment scan were closely correlated (Pearson correlation coefficient $= 0.92$).
of cytostatic treatments (22) but also occurs with established chemotherapy and chemoradiotherapy regimens.

In the original publication of the TLG, Larson et al. (1) proposed that the SUV of a lesion should be multiplied by its volume on PET. Tumor volume on PET was obtained by including voxels with an activity concentration above a certain threshold (I,9,23). In pretreated, homogeneous tumors with high 18F-FDG uptake, this approach is feasible and reproducible. However, we found that in posttherapeutic scans the contrast between tumor 18F-FDG uptake and background or surrounding normal tissue is frequently low. However, more sophisticated segmentation algorithms may be developed that work in the PET image space. This development, in turn, results in underestimated tumor volumes and underestimated tumor mean activity concentration within this volume. For these reasons, we decided to use the CT information to outline tumor borders.

Limitations to this approach exist, because CT-based volume measurements cannot provide information about tumor viability. This is most likely the reason why the introduction of the TLG index derived by PET/CT was less predictive of percentage tissue necrosis than was SUVmax or SUVmean alone. A poor correlation between changes in tumor size and histopathologic response or patient outcome has been described in soft-tissue sarcoma previously (19,24,25). Future studies will have to clarify whether the combined analysis of metabolic and volumetric changes by PET/CT (i.e., TLG) may improve tumor response assessment in other cancers. For example, in esophageal cancer, changes in tumor volume in response to neoadjuvant therapy were correlated with histopathologic tumor response (26). In this situation, a combined analysis of tumor size and volume may improve the accuracy of response assessment.

CONCLUSION

Our study indicates that in soft-tissue sarcomas, changes in SUVmax are as accurate for assessment of tumor response as more sophisticated parameters, such as TLG. The high correlation between changes in SUVmax and changes in SUVmean suggest that decreases and increases in SUVmax are representative of changes in average tumor 18F-FDG uptake. Therefore, a sophisticated delineation of tumor extension may not be necessary for response assessment by 18F-FDG PET.

REFERENCES

9. Hoekstra CJ, Stroobants SG, Smit EF, et al. Prognostic relevance of response evaluation using 18F-FDG PET as a predictor of percentage tissue necrosis than was SUVmax.