Repeatability of Quantitative 18F-FET PET in Glioblastoma

To cite this article: Peter Ferjani et al 2021 Biomed. Phys. Eng. Express 7 035020

View the article online for updates and enhancements.
Repeatability of Quantitative 18F-FET PET in Glioblastoma

Peter Ferjančič1,6, Martin A Ebert2,3,4, Roslyn Francis2,3, Anna K Nowak2,3 and Robert Jeraj1,5,6

1 Department of Medical Physics, University of Wisconsin-Madison, Madison, WI, United States of America
2 Sir Charles Gairdner Hospital, Nedlands, Western Australia, Australia
3 Medical School, University of Western Australia, Crawley, Western Australia, Australia
4 5D Clinics, Perth, Western Australia, Australia
5 Faculty of Mathematics and Physics, University of Ljubljana, Ljubljana, Slovenia
6 First author: Peter Ferjančič, MS; Graduate student, Department of Medical Physics, Email: pferjancic@wisc.edu

Keywords: FET, PET, repeatability, SUV, limits of agreement, glioblastoma

Abstract

Purpose: O-(2-[18F]fluoroethyl)-L-tyrosine (FET), a PET radiotracer of amino acid uptake, has shown potential for diagnosis and treatment planning in patients with glioblastoma (GBM). To improve quantitative assessment of FET PET imaging, we evaluated the repeatability of uptake of this tracer in patients with GBM. Methods: Test-retest FET PET imaging was performed on 8 patients with histologically confirmed GBM, who previously underwent surgical resection of the tumour. Data were acquired according to the protocol of a prospective clinical trial validating FET PET as a clinical tool in GBM. SUVmean, SUVmax and SUV98% metrics were extracted for both test and retest images and used to calculate 95% Bland-Altman limits of agreement (LoA) on lesion-level, as well as on volumes of varying sizes. Impact of healthy brain normalization on repeatability of lesion SUV metrics was evaluated. Results: Tumour LoA were [0.72, 1.46] for SUVmean and SUVtotal, [0.79, 1.23] for SUVmax and [0.80, 1.18] for SUV98%. Healthy brain LoA were [0.80, 1.25] for SUVmean, [0.80, 1.25] for SUVmax and [0.81, 1.23] for SUV98%. Voxel-level SUV LoA were [0.76, 1.32] for tumour volumes and [0.80, 1.25] for healthy brain. When sampled over maximum volume, SUV LoA were [0.90, 1.12] for tumour and [0.92, 1.08] for healthy brain. Normalization of uptake using healthy brain volumes was found to improve repeatability, but not after normalization volume size of about 15 cm³. Conclusions Advances in Knowledge and Implications for Patient Care: Repeatability of FET PET is comparable to existing tracers such as FDG and FLT. Healthy brain uptake is slightly more repeatable than uptake of tumour volumes. Repeatability was found to increase with sampled volume. SUV normalization between scans using healthy brain uptake should be performed using volumes at least 15 cm³ in size to ensure best imaging repeatability.

Introduction

Gliomas are the most common malignant primary brain tumour, of which glioblastomas (GBM) constitute about 70% [1]. They are also the most aggressive type of brain tumour, with median survival being under 12 months [2, 3]. The current gold standard treatment for GBM consists of surgery followed by radiation therapy (RT) with concurrent and adjuvant chemotherapy [4]. While maximum resection of the tumour is usually attempted, functional impairment and quality of life have to be considered when determining the aggressiveness of resection [5]. Thus, for both surgical and RT planning, accurate delineation of disease volumes is highly important. Accurate quantification of a physiologically active tumour may allow high dose RT to be appropriately targeted, while minimizing risk to healthy tissue [6].

RT planning target volumes are derived using functional [7, 8] or CT-MRI fusion images [4]. A promising imaging tracer for imaging GBM is 18F-fluoro-ethyl-tyrosine (FET). It is an artificial amino acid PET tracer developed in the late 1990s that has preferential uptake in malignant cells with increased expression of amino acid transporters, but is not incorporated into proteins [9–12]. FET uses a
similar biological mechanism as the longer established 
$[^{11}C]-methyl-L$-methionine (MET), with the advantage of the longer lived $[^{18}F]$ isotope [13]. FET has been shown to have excellent performance for diagnosing primary brain tumours, including volumes not visible on MRI, as well as survival prediction [6, 10, 14]. In recent years, it is becoming increasingly recommended as a tracer for guiding treatment target definition [6, 15].

While multiple publications exist exploring the use of FET [3, 7, 16], little analysis has been done regarding the uncertainties and repeatability of its uptake. Evaluation of a biomarker’s repeatability, defined as the variation in measurements when an experiment is repeated under the same conditions [17, 18], is necessary for accurate assessment of tumour presence and response, as demonstrated in many ways for other tracers [17, 19–25]. However, repeatability of FET still remains largely unexplored. Healthy tissue normalization techniques attempt to partially address this innate variability; however, no studies so far reported the impact of such normalization over varying volume sizes.

The goal of this work is to validate FET PET as a clinical tool in GBM for treatment response by performing the first test-retest analysis of FET PET images. Evaluating the repeatability of a PET tracer is a critical step in tracer validation for response assessment, because it allows for the accurate differentiation of random uptake fluctuations from statistically significant changes in uptake.

Additionally, this is the first work quantifying repeatability of SUV over variable volumes and the impact of uptake normalization on repeatability. Statistically, the variance of the mean of multiple data must be lower than the variance of a single measurement. Intuitively, the same rationale is used in visual image assessment; a single voxel of increased uptake would usually not be a cause for concern, but a larger volume uniformly exhibiting the same increase in uptake could be. Lesion volumes of different sizes might therefore be subject to different limits of repeatability. Understanding this relationship is essential for identifying differences in repeatability for lesions of different sizes, and the relationship remains unexplored until now.

### Materials and methods

#### Patient population and region of interest (ROI) definition

A cohort of patients with histologically confirmed GBM received test-retest FET PET/CT imaging at the Sir Charles Gairdner Hospital (SCGH, Perth, Western Australia) as part of a prospective study to evaluate FET for radiotherapy planning after surgical removal of gross tumour. The human participants in this study were enrolled in a prospective clinical study at the Sir Charles Gairdner Hospital (SCGH, Perth, Western Australia). The study was approved by the institutional review board under SCGH study number 2014-004, and all subjects signed an informed consent form. The study was registered at the Australian New Zealand Clinical trial registry under number ACTRN1261401114639.

The first FET PET/CT scan (test) was scheduled within 4 weeks before the start of RT, and the second scan (retest) was planned to be acquired 7 d later. Patients received intra-venous administration of 200 MBq ($\pm 10\%$) of FET and were subsequently imaged over the cranial region on a Siemens Biograph 16 PET/CT (Siemens Medical Solutions, Malvern, PA). A low dose CT for attenuation correction was acquired followed by a 30-minute list mode acquisition. The list mode data was processed into a 10-minute static image from 20–30 minutes post FET injection. There was an average of 1.9 min difference between test and retest scans between injection time and commencement of imaging. Retest images were rigidly registered to test using monomodal intensity-based registration (imregconfig) implemented in Matlab R2017b (Mathworks Inc., CA). Example axial slices of FET PET/CT data and their voxel-wise differences can be seen in figure 1.

Intracranial volume was segmented on baseline CT images. A bone mask was acquired using a threshold of 200 Hounsfield Units, followed by

![Figure 1. Example comparison of first baseline scan (‘test’, left), second baseline scan 7 d later (‘retest’, middle) and their voxel-level difference (right). Administered FET activity was 200 MBq and the image was acquired for 10 min at 20 min post injection. Disease delineations are indicated by a white contour for test and retest images. While the general shape of the lesion persists, considerable changes in uptake are present between the two scans.](image-url)
morphological closing using a spherical structure with a 10 mm radius. The largest connected region enclosed in this mask was morphologically closed, and the results were visually inspected to ensure quality of segmentation. Only uptake within the intracranial volume was considered for the purpose of this work.

To segment diseased volumes, an adaptive thresholding method was implemented, as it was previously shown that signal-to-background adaptive thresholding performs better in segmenting GBM than fixed threshold metrics \[10, 25\]. First, three spherical regions with the radius of 5 voxels (6.4 cm³) in the contralateral healthy brain were selected as ROIs for normal uptake reference. SUV mean values of these reference regions were calculated, and 1.6x SUV mean was taken as the threshold for lesion segmentation. This was done on both test and retest images and both segmentations were compared. Volumes greater than 1.1 cm³ were considered usable for analysis, as specified in the protocol of the SCGH study number 2014-004, introduced at the beginning of this section.

Original test and retest contours were used for whole-lesion comparisons, and a union of both volumes was used for the voxel-wise comparisons to have equal volumes on both time points. Healthy brain volume was defined as all intracranial volume, excepting the union of the test and retest disease contours dilated by a five-voxel margin. A schematic representation of this segmentation approach can be seen in figure 2.

**FET repeatability**

Contours on both test and retest imaging were used to extract tracer uptake in both tumour and healthy brain. Basic matching metrics were calculated: volume, Dice coefficient, and Hausdorff distance. \(\text{SUV}_{\text{mean}}\), \(\text{SUV}_{\text{max}}\), \(\text{SUV}_{\text{total}}\) and \(\text{SUV}_{98\%}\) were extracted for both volumes and both test and retest imaging.

\(\text{SUV}_{98\%}\) is the 98th quantile of SUV values in the ROI and was investigated as a measure of highest uptake robust to noise, a drawback of \(\text{SUV}_{\text{max}}\) \[26\].

Bland-Altman 95% Limits of Agreement (LoA) of SUV were calculated for log transformed values, as a way of accounting for increase in absolute differences that occurs with higher SUV \[18, 19\]. First, the differences \(d\) of log transformed SUV measurements on test \((M_A)\) and retest \((M_B)\) were calculated, as well as their bias \(B\), standard deviation \(\sigma\) and repeatability coefficient \(RC\).

\[
d = \log(M_B) - \log(M_A) = \log\left(\frac{M_B}{M_A}\right)
\]

\[
B = \text{mean}(d)
\]

\[
\sigma = \text{std}(d)
\]

\[
RC = 1.96 \times \sigma
\]

Finally, LoA were calculated

\[
\text{LoA} = [e^{(B-RC)}, e^{(B+RC)}]
\]

This provided a range in which a measurement could be found with a probability of 95%; e.g. LoA of [0.85, 1.15] would indicate that the ratio \(M_B/M_A\) is expected to fall between 0.85 and 1.15 with 95% frequency.

**SUV repeatability dependence on volume**

Repeatability of SUV values was evaluated on the healthy brain volume, and on the tumour volume, defined as the union of segmentations on test and retest imaging, and using neighborhoods of varying sizes.

For each voxel \(v_i\) within healthy or diseased brain, the mean voxel SUV of a neighborhood \(\text{SUV}_{\text{mean}}^{\text{NH}}\) was calculated. Bland–Altman 95% Limits of Agreement (LoA) were calculated for the test and retest \(\text{SUV}_{\text{mean}}^{\text{NH}}\) on a voxel level. To account for potential global SUV shift, bias \(B\) was set to 0 when calculating LoA for varying volumes. The neighborhoods of
For every size, were calculated for every voxel within their respective volumes. These neighborhood sizes ranged from 0.01 cm³ (single voxel) to 3.15 cm³ (335 voxels), which was the largest neighborhood size which could be fully contained within the larger lesion segmentations and was selected as the cutoff for this analysis. While larger neighborhoods can be used, SUV_{NH}^{\text{mean}} will be impacted by values outside the diseased volumes, obscuring the differences between healthy and diseased tissues.

A full list of neighborhood sizes considered is presented in Table 1. The X/Y label indicates that neighborhood voxels need to be within a cube of width X voxels, centered on the voxel of interest, and no more than Y voxels away from the voxel of interest. Examples of 3/1 and 5/2 neighborhoods can be seen in Figure 3. This way of defining neighborhoods was implemented in order to decrease the difference in the number of voxels between neighborhood sizes, compared to simply taking all voxels within distance Y.

### Healthy brain normalization

In order to assess the impact of healthy brain normalization on uptake repeatability, repeatability of non-normalized lesion SUV_{mean} was compared to the repeatability of lesion SUV_{mean} normalized by mean uptake in a healthy brain reference region. A range of different reference region volumes following contralateral normal brain was used using the same neighborhood approach as in previous section (range: 0.01 cm³ to 22.7 cm³). Since healthy brain encompassed a much larger volume, larger reference regions could be used than within tumour contours. Therefore, additional crescent-shaped reference volumes were investigated, allowing for considerably larger reference volumes [27].

### Results

#### Patient population and ROI definition

Twenty-four patients were enrolled in the study, of whom nine consented to receive pre-treatment test-retest FET PET/CT scans, all of which were obtained within 6–9 d of each other. One patient had no visible increased uptake of FET in either test or retest; the other eight were selected for further analysis. Mean patient age was 53 years (range: 36–61) and there were 5 male and 3 female patients. All PET images were reconstructed using manufacturer recommended settings: PSF reconstruction with 3 iterations and 24 subsets (3i24s) and smoothed using a 4 mm Gaussian filter.

Lesions were segmented with a background uptake-based thresholding method (1.6 × background uptake). The average tumour segmentation threshold
was 1.62 g ml\(^{-1}\) [range: 1.08–2.66 g ml\(^{-1}\)]. Using these thresholds, lesion contours were generated with an average volume of 33.2 cm\(^3\) (range 2.6 – 149.1 cm\(^3\)). The volumes and basic matching metrics of these contours are shown in table 2. Lesion volume varied considerably between test and retest. While average lesion volume changed by only 6%, and was found not significant using a paired sample t-test, individual lesion volume change ranged from −57% to +87%. Average absolute volume difference was found to be 32%, and average Dice coefficient between test and retest lesion volumes was relatively low at 0.66, with poorest matching observed in smallest volumes of slightly increased uptake surrounding areas of surgical resection. For larger disease volumes, matching between test and retest was better, reaching Dice coefficient values of up to 0.90.

**FET repeatability**

The repeatability of FET SUV metrics was evaluated by comparing metric values between test and retest scans. The summary of this analysis can be found in table 3. A similar analysis was performed for the volumes of the healthy brain and shown in table 4. Healthy volume was defined outside joint contours of both test and retest imaging, therefore only one segmentation is used for both images. Because of that, SUV\(_{\text{total}}\) is SUV\(_{\text{mean}}\) multiplied by the healthy brain volume, and is not shown in the table below. For lesion-level SUV\(_{\text{mean}}\), SUV\(_{\text{max}}\), and SUV\(_{\text{98\%}}\) metrics, LoA were calculated using test and retest contours across all patients, the results of which are in table 5. Basic FET SUV metrics had limits of agreement around [0.80, 1.25] for both healthy and diseased intracranial volumes and for most SUV metrics.

**SUV repeatability dependence on volume**

Values of SUV\(_{\text{mean}}\) were calculated over a range of neighborhood volume sizes, and LoA were calculated for each patient for each of the neighborhoods considered. Resulting LoA are shown as the dashed gray lines in Figure 4. The mean LoA for SUV\(_{\text{NH}}\) across all patients was calculated for each neighborhood and the results are presented in Table 6 and demonstrated by the solid blue line in figure 4. Single voxel SUV in tumour was found to have LoA of [0.76, 1.32], but for the largest neighborhoods considered, the LoA narrowed to [0.90,1.12].

**Healthy brain normalization**

To explore the effect of healthy tissue uptake normalization on repeatability of lesion metrics, we calculated the LoA for lesion SUV\(_{\text{mean}}\), SUV\(_{\text{max}}\), and SUV\(_{\text{98\%}}\) under two conditions: calculating metrics from the

### Table 2. Tumour volumes for both test (\(^1\)) and retest (\(^2\)), volume differences, Dice coefficients, and Hausdorff distance for data used in the analysis.

<table>
<thead>
<tr>
<th>Patient</th>
<th>Vol.(^1) (cm(^3))</th>
<th>Vol.(^2) (cm(^3))</th>
<th>Vol. diff (cm(^3))</th>
<th>Vol. diff (%)</th>
<th>Dice coeff.</th>
<th>Haus. Dist. (cm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>47.82</td>
<td>56.53</td>
<td>8.70</td>
<td>18</td>
<td>0.90</td>
<td>0.14</td>
</tr>
<tr>
<td>2</td>
<td>2.96</td>
<td>5.43</td>
<td>2.48</td>
<td>84</td>
<td>0.39</td>
<td>0.37</td>
</tr>
<tr>
<td>3</td>
<td>117.66</td>
<td>149.06</td>
<td>31.41</td>
<td>27</td>
<td>0.62</td>
<td>0.36</td>
</tr>
<tr>
<td>4</td>
<td>32.67</td>
<td>16.21</td>
<td>−16.47</td>
<td>−50</td>
<td>0.66</td>
<td>0.39</td>
</tr>
<tr>
<td>5</td>
<td>29.08</td>
<td>29.34</td>
<td>0.26</td>
<td>1</td>
<td>0.83</td>
<td>0.17</td>
</tr>
<tr>
<td>6</td>
<td>2.59</td>
<td>1.10</td>
<td>−1.49</td>
<td>−57</td>
<td>0.28</td>
<td>0.43</td>
</tr>
<tr>
<td>7</td>
<td>5.29</td>
<td>5.30</td>
<td>0.01</td>
<td>0</td>
<td>0.83</td>
<td>0.10</td>
</tr>
<tr>
<td>8</td>
<td>13.74</td>
<td>17.20</td>
<td>3.45</td>
<td>25</td>
<td>0.81</td>
<td>0.15</td>
</tr>
<tr>
<td>Mean</td>
<td>31.48</td>
<td>35.02</td>
<td>3.55</td>
<td>6</td>
<td>0.66</td>
<td>0.26</td>
</tr>
</tbody>
</table>

### Table 3. SUV metrics for tumour volumes: SUV mean, max, total, and 98th percentile. Secondary column labels indicate whether data refers to test (T) or retest (R) image.

<table>
<thead>
<tr>
<th>Patient</th>
<th>SUV(_{\text{mean}}) (g/ml)</th>
<th>SUV(_{\text{max}}) (g/ml)</th>
<th>SUV(_{\text{total}}) (g/ml)</th>
<th>SUV(_{\text{98%}}) (g/ml)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>T</td>
<td>R</td>
<td>T</td>
<td>R</td>
</tr>
<tr>
<td>1</td>
<td>2.72</td>
<td>2.91</td>
<td>4.78</td>
<td>5.40</td>
</tr>
<tr>
<td>2</td>
<td>1.25</td>
<td>1.15</td>
<td>1.56</td>
<td>1.34</td>
</tr>
<tr>
<td>3</td>
<td>1.87</td>
<td>1.46</td>
<td>4.06</td>
<td>3.80</td>
</tr>
<tr>
<td>4</td>
<td>2.60</td>
<td>3.65</td>
<td>5.48</td>
<td>5.33</td>
</tr>
<tr>
<td>5</td>
<td>2.01</td>
<td>2.07</td>
<td>4.07</td>
<td>4.02</td>
</tr>
<tr>
<td>6</td>
<td>2.60</td>
<td>2.71</td>
<td>3.23</td>
<td>3.00</td>
</tr>
<tr>
<td>7</td>
<td>1.84</td>
<td>1.64</td>
<td>2.66</td>
<td>2.43</td>
</tr>
<tr>
<td>8</td>
<td>2.32</td>
<td>2.77</td>
<td>3.35</td>
<td>4.04</td>
</tr>
<tr>
<td>Mean</td>
<td>2.33</td>
<td>2.65</td>
<td>3.54</td>
<td>4.10</td>
</tr>
</tbody>
</table>
original lesion SUV values in the PET image, and calculating metrics using lesion SUV values that have first been divided by the mean SUV of healthy brain reference region volumes of varying sizes. Both smaller reference volumes of neighborhoods centered on points, as well as larger crescent-shaped reference regions were considered. The results are shown in figure 5.

Normalization using single voxels or very small reference volumes resulted in poorer repeatability than using non-normalized values across all metrics. Normalization did improve repeatability for the tumour metrics starting at reference volumes of about 1 cm³, and continued to improve it for all metrics up to about 5 cm³. For SUV$_{\text{max}}$ and SUV$_{95}$ no improvement in repeatability was observed for larger volumes. For SUV$_{\text{mean}}$, repeatability continued to improve up to reference volumes of about 15 cm³, but no improvement was observed for volumes larger than that. No improvement was observed even when using much larger crescent shaped reference volumes of up to 200 cm³. Results for SUV$_{\text{total}}$ are just SUV$_{\text{mean}}$ results scaled by healthy brain volume and are therefore not explicitly shown.

**Discussion**

**Patient population and basic repeatability**

This work presents the first prospective study quantifying the repeatability of FET PET in patients with
glioblastoma. Evaluating the repeatability of a PET tracer allows for the accurate differentiation of random uptake fluctuations from statistically significant changes in uptake and is therefore a critical step in tracer validation for response assessment. Knowing the magnitude of this variability is especially important in quantitative disease assessment to treatment, as it allows differentiation of significant changes to random fluctuations.

Segmented lesion volumes varied considerably between test and retest, with an average absolute change of 32%. For reference, MRI-based 3D volume repeatability has found variation in tumour volumes of approximately 10% [28], when comparing gross tumour volumes before surgical resection. This large variation in volumes can be partially explained with the fact that thresholding as a segmentation method is sensitive to gradual changes in uptake. Even so, this is an important consideration for any FET quantitative analysis as it highlights considerable variability in FET uptake. This variability could also FET-based radiation therapy planning, although target expansion to PTV is expected to somewhat reduce these effects.

### Basic FET repeatability

Healthy brain measurements were found to be slightly more repeatable than those of tumour volumes. SUVmean of tumour volumes was found to be least repeatable, although that could be caused by the segmentation method used. This result is comparable to previous studies exploring repeatability in brain that found SUVmean repeatability of 20%-30% for FDG [22] and 18–24% for FLT PET uptake variability [20].

### SUV repeatability dependence on volume

Dependence of SUV repeatability on volume size was evaluated. Averaging SUV over larger neighborhoods improved repeatability, however it was only for the largest neighborhoods when SUViNH was calculated over hundreds of voxels. This result indicates that when evaluating whether a significant change of SUV occurred within a volume of interest, the size of the volume should be taken into account as well.

### Healthy brain normalization

LoA were also compared between non-normalized whole-volume SUV metrics and metrics where uptake was normalized using healthy brain reference region uptake. Normalization using a single voxel did not improve repeatability for any metric compared to using non-normalized values, but repeatability did

---

Table 6. LoA of SUViNH, dependency on neighborhood (Nhood) size.

<table>
<thead>
<tr>
<th>Nhood size (cm³)</th>
<th>Nhood size (vox)</th>
<th>Mean LoA (tumour)</th>
<th>Mean LoA (healthy)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.01</td>
<td>1</td>
<td>[0.76, 1.32]</td>
<td>[0.80, 1.25]</td>
</tr>
<tr>
<td>0.09</td>
<td>7</td>
<td>[0.78, 1.28]</td>
<td>[0.83, 1.21]</td>
</tr>
<tr>
<td>0.33</td>
<td>27</td>
<td>[0.81, 1.24]</td>
<td>[0.85, 1.17]</td>
</tr>
<tr>
<td>0.41</td>
<td>33</td>
<td>[0.81, 1.23]</td>
<td>[0.86, 1.16]</td>
</tr>
<tr>
<td>1.00</td>
<td>81</td>
<td>[0.84, 1.19]</td>
<td>[0.88, 1.13]</td>
</tr>
<tr>
<td>1.52</td>
<td>123</td>
<td>[0.86, 1.17]</td>
<td>[0.89, 1.12]</td>
</tr>
<tr>
<td>3.11</td>
<td>251</td>
<td>[0.88, 1.14]</td>
<td>[0.91, 1.09]</td>
</tr>
<tr>
<td>3.18</td>
<td>257</td>
<td>[0.88, 1.13]</td>
<td>[0.92, 1.09]</td>
</tr>
<tr>
<td>4.15</td>
<td>335</td>
<td>[0.90, 1.12]</td>
<td>[0.92, 1.08]</td>
</tr>
</tbody>
</table>

---

Figure 5. SUV LoA width dependence on reference region volume size. Dark blue lines represent values using smaller spherical reference regions, light blue lines represent values normalized using larger crescent-shaped reference regions and black the reference non-normalized values, also shown in table 5. For all three SUV metrics, normalization improved repeatability, demonstrated by the fact that the vertical distance between the two blue lines is narrower than the reference black lines. As seen on the bottom three plots with expanded X-axis, using much larger reference volumes did not seem to improve repeatability of lesion SUV metrics.
improve with increasing normalization volumes, up to a point. For \( \text{SUV}_{\text{max}} \) and \( \text{SUV}_{95\%} \), repeatability continued to improve up to about 5 cm\(^3\), after which the LoA plateaued with no noticeable improvement. For \( \text{SUV}_{\text{mean}} \) the repeatability still kept improving up to about 15 cm\(^3\), and plateaued after that. Using much larger, crescent shaped reference volumes ranging up to 200 cm\(^3\) also did not improve repeatability compared to these volumes. This is an important consideration for future studies using FET PET/CT.

**Limitations**

This analysis was performed on eight patients with data suitable for analysis, a number that could not be increased more, because patient accrual ceased. The patient scans were also acquired in the weeks post-surgery, which could introduce undesired effects in FET uptake and may not be representative of repeatability in other clinical scenarios such as pre-operatively or in assessment of recurrence. Most lesions appeared larger on retest images seven days after test, although these changes were found not to be statistically significant in the current population. Finally, the images were registered to each other using direct rigid registration, which was deemed appropriate for the constrained intracranial volumes, but it could still affect the exact results.

**Conclusions**

The repeatability of FET PET/CT derived SUV metrics was assessed for the first time for both lesion and healthy brain volumes in patients with glioblastoma. FET uptake limits of agreement (LoA) were found to be around [0.80, 1.25] for whole-volume metrics. SUV LoA dependence on volume size was found for both healthy brain and tumour volumes. Uptake normalization by healthy tissue uptake should be performed for comparing SUV tumour metrics, and the reference volume should be at least around 15 cm\(^3\) in volume.

These results can be used to approximate criteria for \(^{18}\text{F}\)-FET PET evaluation of treatment response in patients with GBM, although the results should be validated in a larger population size.

**Disclosures**

This work was supported by a Cancer Council WA research grant and University of Wisconsin Carbone Cancer Center Support Grant P30 CA014520. Robert Jeraj is the co-founder and CSO of AIQ Solutions. No other potential conflict of interest relevant to this article was reported.

**Acknowledgments**

We thank the patients who volunteered their time, and we thank the imaging technologists who acquired the data. We also thank the members of the ICRU committee for Prescribing, Recording and Reporting Radiation Treatments for their comments and insight.

**Data availability statement**

The data generated and/or analysed during the current study are not publicly available for legal/ethical reasons but are available from the corresponding author on reasonable request.

**Declarations**

Robert Jeraj is the co-founder and CSO of AIQ Solutions.

**ORCID iDs**

Peter Ferjancič: https://orcid.org/0000-0001-9604-9439

Robert Jeraj: https://orcid.org/0000-0002-2192-2931

**References**

[16] Debus C et al 2018 Feasibility and robustness of dynamic 18F-FET PET based tracer kinetic models applied to patients with recurrent high-grade glioma prior to carbon ion irradiation Sci. Rep. 8 14760
[20] Lodge M A et al 2017 Repeatability of 18F-FLT PET in a multicenter study of patients with high-grade glioma J. Nucl. Med. 58 393–8
[23] Hatt M et al 2010 Reproducibility of 18F-FDG and 3'-Deoxy-3'-18F-fluorothymidine PET tumor volume measurements J. Nucl. Med. 51 1368–76
[27] Unterrainer M et al 2017 Towards standardization of 18F-FET PET imaging: Do we need a consistent method of background activity assessment? EJNMMI Res. 7 1–8