University of Groningen

Quantification and Kinetic Analysis of Brain PET Studies
Mansor, Syahir

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2017

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):
Mansor, S. (2017). Quantification and Kinetic Analysis of Brain PET Studies. [s.l.]: [s.n.]

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Download date: 23-11-2019
Chapter 4

Experimental assessment of PET/CT repeatability

Syahir Mansor¹
Maqsood Yaqub¹
Dennis Heijtel²
Martin A. Lodge³
Ronald Boellaard¹,⁴

¹Department of Radiology & Nuclear Medicine, VU University Medical Center, Amsterdam, The Netherlands
²Philips Healthcare, Best, the Netherlands
³The Russell H. Morgan Department of Radiology and Radiological Science, Johns Hopkins University School of Medicine, Baltimore, MD
⁴Department of Nuclear Medicine and Molecular Imaging, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands

This chapter has been submitted to Medical Physics
Abstract

**Purpose:** In longitudinal oncological and brain PET/CT studies it is important to understand the repeatability of quantitative PET metrics in order to assess change in tracer uptake or binding. The present studies were performed in order to assess the aspects of PET/CT repeatability related to reconstruction settings, analysis methods, scan duration (or image noise) and position in the field of view. **Procedures:** Multiple (repeated) scans have been performed using a uniform $^{68}$Ge phantom, a NEMA image quality (IQ) phantom and a 3D Hoffman brain phantom filled with $^{18}$F solutions. Studies were performed with and without repositioning the phantom and all (12 replicate for IQ phantom and 10 replicate for Hoffman brain phantom) scans were performed using equal count statistics. For the $^{68}$Ge phantom, the coefficient of variation (COV%) and variance of the voxel values across the phantom were studied as function of scan statistics and reconstruction settings. For the NEMA IQ phantom we studied the maximum, peak and mean recovery coefficients (RC) in each sphere as function of experimental conditions (noise level, reconstruction settings and phantom repositioning). For the 3D Hoffman phantom the mean activity concentration was determined within several volumes of interest and activity recovery and its precision was studied as function of experimental conditions. **Results:** For all phantom studies voxel noise (expressed by variance and COV) and SUV$_{\text{max}}$ or SUV$_{\text{mean}}$ repeatability depended on reconstruction settings and frame duration, as expected. When exploring SUV$_{\text{max}}$, SUV$_{\text{peak}}$ or SUV$_{\text{mean}}$ of the spheres in the NEMA IQ phantom, it was observed that repeatability depended on phantom position with SUV$_{\text{max}}$ being most and SUV$_{\text{peak}}$ least sensitive to phantom repositioning. Moreover, the impact of phantom positioning on SUV metrics depended on sphere size. For the brain phantom, regional average SUVs were derived and these were only minimally affected by phantom repositioning. **Conclusion:** The repeatability of quantitative PET metrics depends on the combination of reconstruction settings, data analysis methods and scan duration (scan statistics). Moreover, repeatability was also affected by phantom repositioning but its impact depended largely on the data analysis method being used. The study suggest that for oncological PET studies use of SUV$_{\text{peak}}$ may be preferred over SUV$_{\text{max}}$ because SUV$_{\text{peak}}$ metric is likely less sensitive to patient positioning/tumor sampling effects.
4.1 Introduction

[18F]fluorodeoxyglucose (18F-FDG) positron emission tomography and computed tomography (PET/CT) is being used for staging and tumor response assessment in oncology [1–7]. The analysis of 18F-FDG [8] uptake in tumors can be performed semi-quantitatively using the standard uptake value (SUV) rather than using visual assessment of relative change. Main drawback of using SUV is its sensitivity to various technical factors, such as image reconstruction settings [9] and region of interest delineation strategies [10–12]. The impact of different image acquisition and processing methods on SUV are well understood and various standardization efforts are made, especially in multi-center clinical trials, to mitigate these effects [13]. In order to yield a high reproducibility, standard operating procedure (SOP) or guidelines need to be followed that address patient preparation, image acquisition and processing, and data analysis and interpretation. For longitudinal studies, i.e. when quantitatively measuring tumor response to therapy, it is important to understand the repeatability of the quantitative metric being used to measure change in tracer uptake. Several repeatability studies have reported [14–16] repeatabilities ranging from 10 to 15% on average. This repeatability arises from several clinical and technical contributions, such as uncertainties in administered activity, variability in patient preparation and physiological condition (blood glucose level) etcetera, but also from image noise due to variability in scan statistics. Surprisingly, reported repeatabilities for FDG PET/CT studies seems to reach a plateau at 10-12% meaning that generally best repeatabilities reported are >10%. The question arises if the most optimal reported repeatabilities could have been limited by technical or experimental factors.

The aim of this study was, therefore, to experimentally evaluate PET/CT repeatability dependence on reconstruction settings, scan duration, image analysis methods and phantom positioning. The latter aspect was included to resemble the clinical condition in longitudinal settings where patients are not positioned in exactly the same manner and position for all imaging time points. To this end, several phantoms, that are typically used for quantitative performance assessments of PET/CT systems, were scanned repeatedly (n=12 for IQ phantom and n=10 for Hoffman brain phantom) with and without phantom repositioning, while keeping counting statistics equivalent between replicates. Additionally, the acquired data were reconstructed using various clinically applied reconstruction settings and data were analysed with commonly used quantitative metrics, such as SUV\(_{\text{max}}\), SUV\(_{\text{peak}}\) or SUV\(_{\text{mean}}\).
Experimental assessment of PET/CT repeatability

Materials and Methods

4.2 Materials and Methods

4.2.1 Phantom experiments

All phantom experiments were performed on an Ingenuity PET/CT scanner (Philips Healthcare, Cleveland, USA). All emission data were reconstructed using the vendor provided time of flight reconstruction method (BLOB-OS-TF), including all corrections needed for quantification such as scatter, random, normalisation and attenuation correction. A low dose CT, using vendor recommended setting, was used for attenuation correction. Moreover, reconstructions were performed both with and without point-spread-function (PSF) (1 iteration with 6mm of resolution regularization). The reconstructions generated emission images with a voxel size of $4 \times 4 \times 4$ mm$^3$ and matrix of $144 \times 144 \times 45$ for body mode acquisitions and a voxel size of $2 \times 2 \times 2$ mm$^3$ and matrix of $128 \times 128 \times 90$ for brain mode acquisitions (i.e. only in case of the 3D Hoffman phantom, as discussed below).

Three different types of phantoms were evaluated. First, a $^{68}$Ge filled phantom with a length of 19 cm and a 20 cm diameter (6000 mL) was studied. The phantom consist of uniformly distributed $^{68}$Ge in epoxy with a total activity of 11.91 MBq at the time of scanning. This phantom, was scanned in a fixed position for 120 minutes in order to assess average image quality/scan statistics for differently reconstructed frame durations (12 frames, each of 1, 2, 4, 5 and 10 min, so 12×1min, 12×2 min and so on).

Secondly, we performed studies using the NEMA NU2 Image Quality (IQ) phantom (Data Spectrum, Hillsborough, NC). This phantom is well known for its use in NEMA NU-2 IQ PET performance measurements and for its use in standardization of multi-center PET studies (EANM-EARL) [17]. The phantom consists of a large background volume (9400

---

**Figure 4.1** – Illustration of VOIs in grey matter for both brain hemispheres in (A) and VOIs in white matter in (B). Figure (C) shows two isolated VOI in white matter. These VOIs were used to assess RC for different brain structures and regions as function of experimental condition. (subfigure A; region 1: Putamen, region 2: Caudate, 3: Thalamus, 4: Frontal, 5: Temporal)
Materials and Methods

Experimental assessment of PET/CT repeatability

mL) with six hot spheres with inner diameters of 10, 13, 17, 22, 28 and 37 mm. The hot spheres and the background were filled with an $^{18}$F solution following EANM/EARL recommendation and resulted in sphere/background ratio’s of approximately 10:1. Two series of scans were performed. First, similar to the $^{68}$Ge phantom scans, the IQ phantom was filled once (2.36 kBq·ml$^{-1}$ in the background compartment and 22.53 kBq·ml$^{-1}$ in the spheres) and scanned in one fixed position for 120 minutes and the data were reconstructed using three different frame durations (2, 4 and 5 min for the first reconstructed frame). In order to keep scan statistics constant between all reconstructed images, frame duration was increased for each subsequent reconstructed frame to compensate for radioactive decay (i.e. yielding similar count statistics for each subsequent frame). Secondly, the IQ phantom was filled once (the sphere activity concentration equalled 28.63 kBq·ml$^{-1}$ and that of the background compartment equalled 3.08 kBq·ml$^{-1}$) and rescanned (both low dose CT and PET) 12 times while randomly repositioning the phantom ($\pm 20$ mm difference). Each of the acquisitions were reconstructed with frame durations set such that to yield the same count statistics as achieved with the first set of (stationary phantom) measurements.

Finally, we acquired data for the 3D Hoffman brain phantom (Data Spectrum, Hillsborough, NC). Similar to the IQ phantom experiment, the phantom was scanned in two series: one using the same phantom position over 120 minutes (125 MBq totally in the phantom at start scanning) and a series consisting of rescanning at 10 different phantom positions (114 MBq at start scanning). Similar to the IQ phantom experiment data were reconstructed with three different frame durations (2, 4 and 5 min for the first frame) and using increasing frame durations to compensate for radioactive decay (i.e. yielding similar count statistics for each frame). For the second series reconstructed scan durations were set such to yield the same count statistics as seen for the replicates in the first series.

4.2.2 Evaluation of parametric image statistics

Image noise was evaluated by generating parametric images of the mean ($\mu$), SD (square root of variance) and COV (SD/$\mu \times 100\%$) over the replicate measurements for all reconstructions applied. Repositioned phantom data were first rigidly realigned before generating parametric data using VINCI software (Max Planck Institute, Cologne, Germany). Parametric data were evaluated for each phantom as a function of initial frame duration (i.e. 2, 4, 5 min), reconstruction type (PSF or non-PSF) and presence of absence of phantom repositioning.

4.2.3 Regional assessments

Regional assessment of the experiments was performed using several automated (IQ Phantom) and manual image segmentation methods ($^{68}$Ge and 3D Hoffman phantom). For the $^{68}$Ge phantom study, regional average COV (%), activity concentration and variance were
calculated for a single central axial plane of the phantom within a large circular region of interest (ROI) of 15 cm diameter.

Segmentation or delineation of the spheres in the IQ phantom was performed using the EARL analysis tool which generated background corrected 50% of SUV\textsubscript{max} contours [17]. From these delineations we derived the maximum (SUV\textsubscript{max}), peak (SUV\textsubscript{peak}) and mean (SUV\textsubscript{mean}) uptake in each of the images. Note that the volume-of-interest analysis was performed on the original images without image registration to resemble clinical conditions as closely as possible. Next, we derived the recovery coefficient (RC\textsubscript{max}, RC\textsubscript{peak} and RC\textsubscript{mean}) by dividing observed max, peak and mean values by the expected activity concentration based on phantom filling parameters. RCs were derived for each sphere and for all acquired and reconstructed emission images. Finally RCs repeatabilities will be shown as a function of sphere size, data analysis method (max, peak and mean) and reconstruction methods for both stationary and repositioning phantom experiments.

Finally, for the 3D Hoffman brain phantom, several VOIs were drawn manually using the co-registered binary mask of grey and white matter of the phantom. For each hemi-sphere in total five different VOIs for grey and six VOIs for white matter of different sizes were drawn as shown in Fig. 4.1. VOI were chosen so as to obtain activity concentration estimates for both cortical and more deeply located brain structures. From these VOIs we derived the mean regional activity concentration and compared these with the actual activity concentration of the solution used to fill the phantom to produce the RC\textsubscript{mean}. For the repositioned phantom study, this VOI template was rigidly realigned onto the original phantom images (i.e. the images were not realigned).

![Figure 4.2](image)

**Figure 4.2** – (A) COV (%) for different frame durations and (B) mean concentration (circle symbol) and SD (cross symbol) as found in the \textsuperscript{68}Ge Phantom.
4.3 Results

4.3.1 Noise characteristic for the $^{68}$Ge Phantom

Fig. 4.2 shows the dependence of COV, mean, SD and variance as a function of different reconstructed frame durations. As expected a non-linear dependence of noise in relation to frame duration was seen but the average activity in the VOI did not depend on the chosen frame durations. The noise varies from 20% to 5% COV and seems to follow Poisson statistics, i.e. changing the frame duration by a factor of 2 results in a COV change of the square root of 2.

4.3.2 NEMA IQ phantom

Fig. 4.3 illustrates all recovery coefficients estimated for the IQ phantom for images with 5 min scan duration. In general, especially for smaller spheres, repositioning of the phantom increased variability of RC data compared with the stationary phantom data (Fig. 4.3). The variability and additional variability due to repositioning was larger when using $RC_{\text{max}}$ and/or using reconstructions that include PSF. Also, use of the maximum RC (Figs. 4.3A and 4.3B) resulted in overestimation of the recovery ($>1$) for the largest spheres (22, 28 and 37 mm in diameter) which increased if PSF reconstructions were used. The biases are smaller when using peak and mean RCs (Figs. 4.3C and 4.3D) and (Figs. 4.3E and 4.3F) respectively. In addition, TOF+PSF produced higher recoveries than TOF alone reconstruction, however in all cases peak and mean RCs were less overestimated compared with $RC_{\text{max}}$. In Fig. 4.4 recovery coefficients estimated for the IQ phantom for images with 2 min scan duration are shown. Although RC showed somewhat larger variabilities, as expected due to the lower count statistics of the 2 min versus 5 min data, similar trends were seen as compared to the 5 min data. In general, $RC_{\text{max}}$ were more sensitive to noise and phantom repositioning than the other quantitative metrics. Table 4.1 and 4.2 summarize the $F$-test significance for differences in repeatabilities between the stationary scan and repositioning phantom data for the various analysis methods. In general, differences between repositioned versus stationary phantom data were statistically significant ($p<0.05$, $F$ test) for small spheres (17, 13, and 10 mm). Fig. 4.5 shows parametric images that illustrate noise distribution (COV) in the images indicating that noise increases when using shorter frame durations (right columns), when the phantom data set is not stationary (rows B & D), and when PSF is used (C & D).
Results

Figure 4.3 – RC of NEMA IQ phantom data averaged over multiple phantom scans as a function of sphere diameter. Data were reconstructed with starting 5 min frame duration using TOF on the left column and TOF+PSF on the right column. Figure (A and B) represent (%) for max, (C and D), peak and (E and F) mean SUVs. Dotted lines correspond to the true RC based on the true activity within the phantom spheres. Boxes represent standard deviation (SD), whiskers show ranges, and solid line depicts median of the data.
Figure 4.4 – RC of NEMA IQ phantom data averaged over multiple phantom scans as a function of sphere diameter. Data were reconstructed with starting 2 min frame duration using TOF on the left column and TOF+PSF on the right column. Figure (A and B) represent RC (%) for max, (C and D), peak and (E and F) mean SUVs. Dotted lines correspond to the true RC based on the true activity within the phantom spheres. Boxes represent standard deviation (SD), whiskers show ranges, and solid line depicts median of the data.
Experimental assessment of PET/CT repeatability

Results

4.3.3 3D Hoffman brain phantom evaluation

Box plots in Fig. 4.6 demonstrate the mean RC for several grey matter regions drawn in the Hoffman brain phantom. There was no significant difference in RC variation between repositioned and stationary scans and when using shorter frame durations (data not shown). PSF based reconstructions yielded slightly higher RCs. Fig. 4.7 shows $R_{\text{mean}}$ for the white matter regions. For white matter similar results were seen except that in this case with PSF reconstruction the RC values were lower than those obtained without PSF. Finally, parametric images depicting noise distributions (COV) are given in Fig. 4.8. The noise (COV) in the images increases when using shorter frame durations (right columns) and when PSF is used (C & D).

Figure 4.5 – Parametric COV (%) images over 12 frames for the stationary phantom data in (A and C) and over 12 different scan positions in (B and D). Images in row (A and B) were reconstructed using TOF, and in row (C and D) with TOF+PSF. First, second and third columns show the COV images obtained with 5, 4 and 2 min scan durations respectively.
Results

Figure 4.6 – RC (%) of Hoffman phantom data in different grey matter regions. Data were reconstructed using TOF (A) and TOF+PSF (B). RC for 5 min frame duration are shown. Boxes represent standard deviation (SD), whiskers show ranges, and solid line depicts median of the data.

Figure 4.7 – RC (%) of Hoffman phantom for different white matter regions. Data reconstructed using TOF are shown (A) and with TOF+PSF in (B). Data for 5 min frame durations are shown. Boxes represent standard deviation (SD), whiskers show ranges, and solid line depicts median of the data.
4.4 Discussion

4.4.1 $^{68}$Ge phantom experiment

The $^{68}$Ge phantom experiment was performed in order to verify the relationship between scan duration and the level of noise in a phantom with uniform structure/activity distributions. It was found that voxel variance can be explained by Poisson noise and that by changing the scan duration with a factor 2 changes voxel variance by a factor of 2 as well (i.e. COV (%) or SD changes with the square root of 2).

Table 4.1 – The significant $p$ values calculated by performing $F$-test between reposition and stationary phantom for different analysis method and reconstruction methods for each sphere sizes for 5 min frame/scan duration.

<table>
<thead>
<tr>
<th>Phantom</th>
<th>Sphere diameter (mm)</th>
<th>TOF (VOI$_{\text{max}}$)</th>
<th>TOF+PSF (VOI$_{\text{max}}$)</th>
<th>TOF (VOI$_{\text{mean}}$)</th>
<th>TOF+PSF (VOI$_{\text{mean}}$)</th>
<th>TOF (VOI$_{\text{peak}}$)</th>
<th>TOF+PSF (VOI$_{\text{peak}}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reposition vs Stationary</td>
<td>10</td>
<td>0.0003</td>
<td>0.00048</td>
<td>0.00086</td>
<td>0.00966</td>
<td>0.0529</td>
<td>0.0132</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>0.0022</td>
<td>0.00207</td>
<td>-</td>
<td>-</td>
<td>0.0359</td>
<td>0.034</td>
</tr>
<tr>
<td></td>
<td>17</td>
<td>0.0180</td>
<td>4.5x10^-8</td>
<td>0.00233</td>
<td>1.2x10^-8</td>
<td>0.00303</td>
<td>4.08x10^-8</td>
</tr>
<tr>
<td></td>
<td>22</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>28</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>37</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

4.4.2 NEMA IQ-phantom

The impact of phantom repositioning on RC precision can clearly be seen in Figs 4.3 and 4.4, especially in case of smaller spheres (<17mm diameter, table 4.1, $p<0.05$) for all analysis methods used. The figures show that, in particular for $R_{\text{C,max}}$, repeatability seems to be worse in case of phantom repositioning compared to stationary phantom data. Yet, use of regionally average values, such as $R_{\text{C,mean}}$ or $R_{\text{C,peak}}$ show less dependence on phantom position than $R_{\text{C,max}}$. Moreover, it was found that particularly $R_{\text{C,max}}$ shows upward bias with decreasing scan duration or worse scan statistics, as was shown before by Boellaard et al. [10] and Lodge et al. [18]. A possible strategy to reduce uncertainty caused by noise and repositioning could therefore be achieved by the use of SUV$_{\text{peak}}$ and this method might be the method of choice for tumor imaging in a clinical setting. Our findings are in good agreement with the study by Lodge et al. suggesting that the peak value is a more robust metric, not only experimentally [19] but also in clinical practice [18]. Moreover, as was shown by Makris et al. [20], SUV$_{\text{peak}}$ seems to be less dependent on differences in image resolution and might therefore be an
Discussion

Experimental assessment of PET/CT repeatability

attractive method in multicentre studies. A possible drawback of using SUV\textsubscript{peak} might be the lower recovery for smaller spheres/tumors when the size of the peak VOI may be equal to or larger than that of the sphere/tumor such that background activity is included within the VOI. The latter may hamper its application for very small tumors and the use of SUV\textsubscript{peak} in a longitudinal setting, e.g. to measure treatment response, therefore warrants further exploration.

**Figure 4.8** – Parametric COV (%) images over 10 frames for the stationary phantom experiment in (A and C) and over 10 different scan positions in (B and D). Images in row (A and B) were reconstructed using TOF, and in (C and D) with TOF+PSF. First, second and third columns consist of 5, 4 and 2 min frame/scan durations respectively.

The choice of acquisition settings and reconstruction algorithm can also heavily affect the quantitative precision. As expected, shorter scans (i.e 2 min scan duration) tend to provide overestimated RC\textsubscript{max} which is consistent with the finding by Boellaard et al. [10] and
Akamatsu et al. [19]. Furthermore, data in this study showed an increase in RC variability from 20 to 30% when using reconstructions that include PSF for both repositioned and stationary data. Even in the stationary phantom study, recoveries varied with reconstruction setting which is in agreement with Armstrong et al. [21].

The parametric COV images (Fig. 4.5) support the RC findings regarding their dependence on various factors tested: image noise increases when using shorter frame durations (right columns); when the phantom data set is not stationary (rows B & D) and when using PSF reconstructions (C & D). It should be noted that parametric variability seen at the edges of the spheres might have been affected by co-registration errors in combination with the sphere walls. Yet the general trend in variability changes as function of scan duration, reconstruction method and phantom positioning are still clearly visible.

### 4.4.3 Hoffman brain phantom

The Hoffman brain phantom consists of a complex structure that mimic the structures of the human brain. The measurement of tracer uptake in small brain structures such as the caudate and putamen can be hampered by partial volume effects. The inclusion of the PSF in the reconstruction increases $RC_{\text{mean}}$ up to 5-10% compared to those seen without PSF kernel. This pattern is consistent with that by Shao et al. [22]. In contrast, $RC_{\text{mean}}$ in white matter regions was reduced by 2-5% when using PSF. These results can be expected as the use of PSF within the reconstruction results in improved spatial resolution and should therefore results in higher recoveries in grey matter structures and lower ones for white matter. On the other hand, use of PSF may introduce Gibbs artefacts which in turn could lead to activity concentration overestimations [23].

Statistical analysis showed a significant difference between repositioned and stationary phantoms scans for both grey and white matter VOIs. However, the differences were very small (<5%) and likely not clinically relevant. The low sensitivity of RC variability for phantom repositioning likely results from the use of mean regional values. This was also observed in the NEMA IQ phantom, where $SUV_{\text{mean}}$ seems to be less sensitive to phantom (re-)positioning than $SUV_{\text{max}}$. Therefore spatially averaging data over an extended volume of interest seems to mitigate the effects of phantom repositioning and/or (voxel) sampling of the phantom. Although the distribution of the radiotracer in the Hoffman brain phantom is assumed to be uniform within grey and white matter regions, the distribution in a real human brain might exhibit larger variations. Therefore, it cannot yet be ruled out that in clinical data the effect of patient position on regional average values might be larger. Yet, we believe that when analysing brain studies using regionally averaged metrics the impact of patient position across multiple longitudinal scans might be minimal (<5%), although more work is needed to fully substantiate these finding.
4.4.4 Future perspectives

The data in this study were all collected on the same PET/CT system using a specific vendor provided reconstruction method and vendor predefined reconstruction settings or protocols (which can only be marginally changed by the user). The results shown in this study may not be directly translated to those seen on other systems where different acquisition and reconstruction methods and settings are applied and it is therefore of interest to further explore the repeatability or precision dependence on several experimental or technical factors on different systems. Yet, several observations made here, such as precision dependence on scan statistics/duration, data analysis methods and reconstructions settings, are consistent with those published elsewhere and may therefore be assumed to be generically applicable (or at least the observed trends). In our work we extended earlier studies by including the effects of phantom position in order to resemble the clinical conditions encountered in longitudinal studies. We found that phantom position and thereby tumor voxel sampling variations particularly affected the precision of SUV$_{\text{max}}$ analysis for small spheres, while use of regionally averaged values by SUV$_{\text{peak}}$ or SUV$_{\text{mean}}$ seem to be able to mitigate these uncertainties (in part). The latter can be understood easily as averaging data over multiple voxels will automatically mitigate some of the voxel sampling effect. In particular, use of a fixed size VOI, such as VOI$_{\text{peak}}$, generates uptake values that can be expected to be less influenced by voxel size provided fractional voxel coverage by the VOI$_{\text{peak}}$ is taken into account appropriately, as was the case in this study.

Table 4.2 – The significant $p$ values calculated by performing $F$-test between different analysis methods for 5 min frame/scan duration base on different reconstruction corresponding to different sphere sizes

<table>
<thead>
<tr>
<th>Sphere diameter (mm)</th>
<th>Reposition TOF</th>
<th>Reposition TOF+PSF</th>
<th>Stationary TOF</th>
<th>Stationary TOF+PSF</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>VOI$_{\text{max}}$</td>
<td>VOI$_{\text{peak}}$</td>
<td>VOI$_{\text{max}}$</td>
<td>VOI$_{\text{peak}}$</td>
</tr>
<tr>
<td></td>
<td>VS</td>
<td>VS</td>
<td>VS</td>
<td>VS</td>
</tr>
<tr>
<td>10</td>
<td>0.0039</td>
<td>-</td>
<td>0.0020</td>
<td>-</td>
</tr>
<tr>
<td>13</td>
<td>0.0005</td>
<td>-</td>
<td>8.7 × 10^{-5}</td>
<td>-</td>
</tr>
<tr>
<td>17</td>
<td>-</td>
<td>-</td>
<td>0.0107</td>
<td>-</td>
</tr>
<tr>
<td>22</td>
<td>-</td>
<td>-</td>
<td>0.0208</td>
<td>-</td>
</tr>
<tr>
<td>28</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.0250</td>
</tr>
<tr>
<td>37</td>
<td>-</td>
<td>0.0065</td>
<td>-</td>
<td>0.0032</td>
</tr>
</tbody>
</table>
4.5 Conclusion

Repeatability or precision of quantitative tracer uptake values depends on scan duration, data analysis methods, reconstruction settings and phantom (re-)positioning. The latter effect was most pronounced in an oncological experimental phantom setting for smaller spheres (<15mm diameter) when using SUV$_{\text{max}}$. For other phantoms and when using either fixed sized VOIs (SUV$_{\text{peak}}$ in the IQ phantom) or using regionally averaged activity concentration data (brain phantom) the impact of phantom position on quantitative precision is minimized. As in longitudinal studies it is impossible to exactly put the patient in same position in the PET/CT system, it would be preferred to quantify tracer uptake using methods that are resistant to patient positioning. The use of SUV$_{\text{peak}}$ in an oncological setting may therefore be a good alternative to SUV$_{\text{max}}$, but its use for smaller lesions needs to be further studied due to the lower recoveries seen for spheres smaller than 15 mm diameter.
Bibliography


