Method to determine the standard deviation of SUV parameters

Giulia Maria Rita De Luca (✉ g.deluca@antoniusziekenhuis.nl)  
Sint Antonius Ziekenhuis  
https://orcid.org/0000-0002-7485-2914

Jan B.A. Habraken  
Sint Antonius Ziekenhuis

Original research

Keywords: Standard Uptake Value, PET quantification, variation in Standard Uptake Value

DOI: https://doi.org/10.21203/rs.3.rs-24705/v1

License: ☕️ This work is licensed under a Creative Commons Attribution 4.0 International License.  
Read Full License
Abstract

Some of the parameters used for the quantification of PET images are the Standardized Uptake Value (SUV)Max, SUVMean and SUVPeak. In order to assess the significance of an increasing or decreasing of these parameters for diagnostic purpose it is relevant to determine their standard deviation. In this study we present a method to determine the standard deviation of the SUV.

Our method is based on dividing an original dataset into subsets of shorter time length. The variation between the SUV parameters of the subsets is used to estimate the standard deviation of the of the original acquisition. This method was tested on images of a NEMA quality phantom with acquisition time of 150 s per bed position and foreground to background activity ratio of 10:1. This original dataset has been reconstructed with different reconstruction lengths, generating new data subsets. The SUVMax, Mean and Peak were calculated for each image in the subsets. Their standard deviation has been calculated per subset for the different spheres included in the phantom. The variation of each subset has then been used to estimate the expected variation between images at 150 s reconstruction length.

We report the largest standard deviation of the SUV parameters for the smallest sphere, and the smallest variation for the largest sphere. The expected variation at 150 s reconstruction length does not exceed 6% for the smallest sphere and 2% for the largest sphere. We also report a larger variation in SUVMax then in SUVMean and SUVPeak. This is in line with expectations that the standard deviation of the SUV Mean or SUVPeak parameter is lower, since the value of more voxels is included in the calculation, as opposed to the SUVMax, where a single voxel is decisive. With the presented method we are able to determine the standard deviation of SUV parameters and to evaluate the effect of parameter selection and lesion size on the standard deviation, and therefore to evaluate its relevance on the total variation of the SUV value between studies.

Introduction

Positron Emission Tomography (PET) has become an indispensable diagnostic tool over the last decades. Computed Tomography (CT) is added to the PET modality for the purpose of attenuation correction and furthermore PET-CT imaging provides a combined view of functional and morphological information.

The radio-ligand F\(^{18}\)-2-fluoro-2-deoxy-D-glucose (FDG) has ensured the success of PET-imaging. The glucose component of the molecule provides a higher uptake of FDG in malignant than in healthy cells [1], the fluor-18 component provides the detectability in the PET-CT system.

PET-CT images can be reported visually by the nuclear medicine physician, however an important advantage of PET imaging is that the uptake can be quantified in absolute measures.
Quantification of FDG PET enables the staging of cancer and the comparison of follow-up studies to track the evolution of cancer and response to tumour therapy [2].

The Standardized Uptake Value (SUV) is used as the standard unit for PET quantification [ref].

The proposed framework for PET Response Criteria in Solid Tumours (PERCIST) suggests to consider a 30% change in Standardized Uptake Value (SUV) as significant variation of tumour activity between images of the same patient [3]. The most used methods of the SUV parameter are the SUVMax and the SUVMean. In the SUVMax only the voxel with the highest uptake is considered, while in the SUVMean all the voxels in a certain region or volume are taken into account. The SUVMax has a low inter and intra observer variability but a high technical statistical variation. The SUVMean on the other hand has a lower technical variation but a higher inter and intra observer variability, since the borders of the volume are a determining factor of the result.

The SUVPeak is introduced as a “best of both worlds” parameter, it calculates the voxels in a limited volume around the voxel with the maximum value.

When comparing images of the same patient acquired at different moments in time, a certain degree of variability is unavoidable, such as patient preparation, biological variability and technical variability.

The goal is to control and minimize the inter and intra observer variability as well as the technical variability between images as much as possible. Studies shown that limiting the variation in image reconstruction, uptake time and scanner characteristics can limit the effect of technical variability to less than 10% of the SUV [4, 5, 6]. Knowledge of the significance of the difference between intra-patient quantifications parameters is crucial to provide a reliable interpretation of the data.

In this study we present a method to estimate the standard deviation of the SUV quantification. The basics of the method is that the PET acquisition is divided into a number of time-frames and that the variation between the quantification of the separate time-frames is used to estimate the standard deviation of the quantification of the total acquisition.

The method is described and validated and applied on a 150 s acquisition of the image quality phantom with a foreground to background activity ratio of 10:1 as example of application. The values and standard deviations of the SUVMax, SUVMean and SUVpeak of the several spheres in the phantom are presented and discussed.

Our method gives insight in the variation of the several SUV parameters in general, but can be used routinely as well to give an insight in the technical variation of a determined SUV quantification. Knowledge about the technical variation of the parameter enables a sharper definition on whether a change is significant or not.

Material And Methods
Image acquisition

A NEMA NU2–2007 image quality phantom was imaged on a GEMINI TF 64/TOF Performance 2010 (Philips Medical Systems International B.V.) according to the requirements for the EANM/EARL FDG-PET/CT accreditation [7]. The phantom was composed by a fillable torso compartment acting as background, by a cylindrical insert in the centre of the torso compartment and by 6 fillable spheres of different diameters (10 mm, 13 mm, 17 mm, 22 mm, 28 mm and 37 mm) placed around the central insert. The fillable torso compartment and the spheres have been filled with a solution of water and $^{18}$F-FDG. At the starting moment of the scan the activity concentration was 2,10 MBq/ml in the torso background compartment and 20,04 MBq/ml in the spheres, resulting in a sphere to background ratio of 9,6:1 (aim is 10:1). [8]

The original dataset was acquired with 150 s frame duration. The total acquisition time was 10 minutes. An attenuation corrected reconstruction was performed at different reconstruction lengths, varying from 4 s to 30 s, generating as many images as possible per subset, without using the same coincidences by varying the starting time of the reconstruction. For example, for the first subset (4 s reconstruction length), the first image was reconstructed using the coincidences recorded between 0 and 4 seconds, the second image by using the coincidences recorded between 5 and 8 seconds and so on, varying the starting moment of the reconstruction, generating a total of 37 images. The longest frame length was 30 s, generating a subset of 5 images. A total of 14 subsets was generated, of respectively 4s, 6s, 8s, 10s, 12s, 15s, 17s, 19s, 20s, 22s, 24s 26s, 28s and 30s acquisition length.

The Philips reconstruction software automatically corrected each reconstruction for the decay of $^{18}$F (half-life of 109.7 minutes [9]), compensating the time difference between the start of the study and the start of the reconstruction by using an opportune scaling factor.

Image analysis

The datasets were analysed using a Python 3.7.0 script (default, Jun 28 2018, 08:04:48) [MSC v.1912 64 bit (AMD64)]. Different SUV parameters were calculated in each image of the subsets:

- the maximum in the central 2D plane of each sphere, defined as SUVMean 2D;
- the maximum of each 3D sphere, defined as SUVMean 3D;
- the average value in the central 2D plane of each sphere, defined as SUVMean 2D;
- the average value of each 3D sphere, defined as SUVMean 3D;
- the average value within a 1 cm$^3$ sphere centred in the maximum value of the sphere [10], defined as SUVPeak.

The SUVMean was calculated by using a ROI of dimensions according to the specifications of the diameters of the spheres, without using for example a thresholding technique on pixel values or a percentage of the maximum value.
The SUV values were calculated per each image in a subset. The SUV values population has been tested for normality with a Kolmogorov-Smirnov test and all subsets matched the characteristics of a normal distribution. We report data until reconstruction length 30 s because the subsequent subset (32 s), composed by 4 images, did not test for normality with a Kolmogorov-Smirnov test. The SUV parameters of the different images in a subset were averaged and their standard deviation was calculated. The variation of the SUV parameters was calculated as the standard deviation divided by their average value multiplied by 100.

In our measurements we can assume a random sampling model, with no correlations, for independent and identically distributed random measurements. The different subsets do not differ in activity nor voxel dimensions and the quantification of the SUV parameters has been done by using the same ROI dimension. We can therefore describe the ratio of the standard deviations SD of two independent repetitions of PET measurements as:

\[
\frac{SD_1}{SD_2} = \left(\frac{RL_1}{RL_2}\right)^{\frac{1}{2}}
\]

(1)

With SD1 and SD2 being the standard deviations and RL1 and RL2 being the reconstruction lengths. [5].

By using the measured variation of the SUV values in a subset as SD1 and the length of the reconstruction of the specific subset as RL1, it is possible to estimate the variation SD2 between repetitions of scans at the total acquisition time RL2=150 s.

Since we divided our acquisition into 14 different subsets, we could calculate 14 different estimations of the SD2, using the described method above. We validated our method by testing whether the value of the estimated SD2 was independent of the acquisition length of the images in the subset.

**Results**

The method that we describe in this paper for estimating the variation of SUV parameters between PET images includes three main steps:

- acquisition of a dataset of a specific length
- generation and reconstruction of subsets
- estimation of the variation in the original dataset by using the subsets, according to Formula 1.

As an example for this method we used a 150 s acquisition for the original dataset, then we divided it in 14 subsets and estimated the SUV parameters. The aimed foreground to backgroud activity ratio was 10:1.

The SUVMax 2D and 3D, the SUVMean 2D and 3D and the SUVPeak were calculated for the 6 spheres sphere in each image of a subset. Their values were averaged and their standard deviation was
estimated. In this way it was possible to plot the recovery curves based on each parameter, with their standard deviation, for each reconstruction length. The results are shown in Fig. 1.

It is noteworthy that at the larger spheres, for the SUVMax and the SUVPeak parameter, there is a bigger variation, where the shorter acquisition lengths tend to have a higher value.

Fig. 2 to 6 show the variation of the measured SUV parameters. The percentage of the variation of the SUV is plotted as a function of the reconstruction length for the different spheres.

Using Formula 1 and the standard deviation of the subsets with different reconstruction lengths we derived the expected SD at reconstruction length 150 s. The estimated variation of the SUVMax 2D and 3D, SUVMean 2D and 3D and SUVPeak at 150 s reconstruction length are plotted in Fig. 7 to 11 for spheres of different diameters. Each figure reports the average value (red line) plus and minus the standard deviation (blue lines). The dots represent the estimated variation of the SUV at 150 s reconstruction length as a function of the reconstruction length of the subset used for the estimation.

From Fig. 7 to 12 can be seen that the acquisition length of the images in the subset has no structural effect on the estimated standard deviation at 150 seconds. The average estimated standard deviation, denoted by the red-line in Fig. 7 to 12 is summarized in Table 1:

| Sphere Diameter (mm) | Est. var. SUVMax 2D | Est. var. SUVMax 3D | Est. var. SUVMean 2D | Est. var. SUVMean 3D | Est. var. SUVPeak |
|----------------------|----------------------|----------------------|----------------------|----------------------|------------------|
| d_{sphere}=10mm      | 5.6±1.1%             | 5.5±1.1%             | 5.0±1.0%             | 5.0±0.9%             | 5.5±1.1%         |
| d_{sphere}=13mm      | 3.8±1.2%             | 3.8±1.2%             | 3.7±1.1%             | 3.6±0.9%             | 3.6±0.8%         |
| d_{sphere}=17mm      | 3.4±0.8%             | 3.4±0.8%             | 2.8±0.5%             | 2.5±0.6%             | 2.8±0.6%         |
| d_{sphere}=22mm      | 3.3±0.7%             | 3.4±0.7%             | 2.2±0.4%             | 1.6±0.3%             | 2.9±0.7%         |
| d_{sphere}=28mm      | 2.9±0.9%             | 2.6±0.6%             | 1.7±0.4%             | 1.3±0.3%             | 2.9±0.9%         |
| d_{sphere}=37mm      | 1.6±0.3%             | 1.9±0.6%             | 0.9±0.3%             | 0.8±0.2%             | 1.7±0.6%         |

Furthermore, it was verified if the variation measured per subset was in accordance with the expected variation as calculated by using the other measured standard deviations. This calculation provided a 68% confidence interval were the measured standard deviation is to be expected. The results are shown in Fig. 13 to 17 for different sphere diameters with measured data as already reported in Fig. 2 to 6.

**Discussion**

We present a method to estimate the standard deviation of a SUV quantification. The method divides the acquisition into timeframes and estimates the standard deviation of the total acquisition using the standard deviation between the timeframes. The method is tested on a 150 s acquisition with a foreground to background activity ratio of 10:1, showing that there is no structural effect on the estimated standard deviation from the acquisition length of the time-frames (Fig. 6 to 11).
Furthermore to validate the method it was shown that the standard deviation of the subsets can successfully estimate the calculated standard deviation of the other subsets (Fig. 12 to 17). We show how subsets of an original scan can be used to estimate the variation between images at different reconstruction lengths and we compared the estimated and the measured results. The measured data are generally in accordance with the estimation performed by used the other subsets and Formula 1. Because of this we could use the available measured variations to estimate the expected variation at reconstruction length 150 s.

In Fig. 1 the calculated values of the SUV parameters are shown for all spheres and for all acquisition lengths of the timeframes in the subset. It was noted that for the larger spheres the SUVMax as well as the SUVPeak values are significantly higher for the shorter acquisition lengths. This can be explained by the fact that at the larger spheres the chance is bigger that a single voxel or group of voxels will have a higher value due to the a higher statistical variation, as is the case with shorter acquisition lengths. This effect occurs with the SUVMax and the SUVPeak, but not at the SUVMean where all the voxels in the regions are used for the calculation.

In table 1 the result of the estimated standard deviation of the SUV quantifications are reported. We report significant differences in variation between images for different sphere dimensions. Typically ranging from 5–1%, for the 10 mm sphere to the 37 mm. This is in accordance with what is reported for simulated data [11]. The dependency of the variation of the SUV quantification of the lesion reason might be a reason to implement our method routinely to give the physician insight to the variation of the SUV quantification. Our maximum expected variation between images, for any estimated parameter, does not exceed the 6% for the smallest object (sphere of diameter 10 mm) and the 2% for the largest object (sphere of diameter 37 mm). This provides an indication of the contribution of the technical variation when the same scanner is used, with equal reconstruction length and activity, and can be compared with the variation measured in FDG PET test-retest studies reporting a typical variation of approximately 10% [14, 15, 16]. In our study the variation is smaller perhaps due the fact that we do not have to deal with other factors, as for example repositioning of the patient or phantom and reinjection of the activity, as in test-retest studies. Other factors introducing variation of the technical components could be reconstruction protocols, analysis methods and scan duration. [17]

We also report a variation between the different SUV parameters (Max, Mean and Peak) for the same sphere diameter. Because the SUVMax only uses a single-pixel evaluation, it is more susceptible to noise and shows a higher variation between images. When a larger ROI is used to estimate the SUV, such as for SUVMean and SUVPeak, the variation reduces. Our definition of SUVMean was based on the knowledge of the objects that were measured because the ROI was defined as a sphere of diameter equal to the diameter of the imaged sphere. This is not always possible during the analysis of images for diagnosis purposes. In that case another definition of SUVMean must be used and the variation between measurements might be expected to increase. The SUVPeak shows a lower variation between images than the SUVMax and it has the advantage to have an univocal definition, independent from the dimension of the lesion or of the object being imaged. This is in accordance with what suggested by
PERCIST and reported in literature [10, 12, 13]. Other factors present in clinical practice, as glucose blood levels, velocity of FDG uptake in the lesions or weight recording can increase the SUV variation in diagnostic images of patients. [18, 19, 20]

Another method to estimate the variation between images would be to acquire a dataset with long acquisition time in comparison with the acquisition time used for diagnostic and generate subsets of the long acquisition. This could be a more direct way to measure the variation, possibly less susceptible to low photon statistics as we report in images reconstructed with a short reconstruction length. A similar approach has been shown in [5] for SUVMax and Mean for reconstruction length of 5 minutes with variation between 11.2–1.2% depending on the filter, type of acquisition (2D or 3D) and parameter (Max or Mean) used.

For this study we worked with a foreground to background activity ratio of 10:1. In order to further verify the method with other uptakes it would be possible to repeat the study with other ratio’s, as for example 5:1 and 2.5:1. Furthermore, the acquisitions could be repeated after a certain amount of hours in order to analyse the variation with other levels of noise.

**Conclusion**

Our method to estimate the standard deviation of a SUV quantification is feasible and, for the described example of PET acquisition, presents results in variation compatible with what present in literature. The method shows that the variation of SUVMax, SUVMean and SUVPeak varies as a function of the dimension of the objects in the imaged phantom and ranges between 5 and 6% for the smallest sphere (diameter of 10 mm) and between 0,9 and 2% for the largest sphere (diameter of 37 mm). SUV values based on a single pixel (SUVMax 2D and 3D) have a larger variation between images than SUV values based on a larger ROI (SUVMean 2D and 3D and SUVPeak) due to their higher susceptibility to noise. Our method could be used routinely to provide insight into the variation of a SUV quantification.

**Declarations**

**Availability of data and materials**

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

**Acknowledgements**

The authors thank Ronald Boellaard, Amsterdam University Medical Centres, Location VUMC, for reading the manuscript and providing valuable feedback. The authors thank Olga Beregovaya, St. Antonius hospital, for her help with the measurements and the reconstructions.

**Funding**
Author information

Department of Medical Physics, St. Antonius Hospital, Nieuwegein, Netherlands

Contributions

The authors read and approved the final manuscript.

Corresponding author

Correspondence to G. M. R. De Luca, g.deluca@antoniusziekenhuis.nl

Ethics declarations

Not applicable

Consent for publication

Not applicable

Competing interests

The authors declare that they have no competing interests.

References

1. Warburg O. On the origin of cancer cells. Science. 1956;123:309–14.
2. Hoekstra CJ, et al., “Monitoring response to therapy in cancer using [18F]-2-fluoro-2-deoxy-D-glucose and positron emission tomography: an overview of different analytical methods”, European Journal of Nuclear Medicine June 2000 vol 27(6).
3. Walh RL. From RECIST to PERCIST: Evolving Considerations for PET Response Criteria in Solid Tumors. J Nucl Med May. 2009;50(1):122S–150S.
4. Boellaard R. Standards for PET image acquisition and quantitative data analysis. J Nucl Med. 2009;50(suppl 1):11S–20S.
5. Doot RK, et al. Instrumentation factors affecting variance and bias of quantifying tracer uptake with PET/CT. Med Phys. 2010;37:6035–46.
6. Fahey FH, et al. Variability in PET quantitation within a multicenter consortium. Med Phys. 2010;37:3660–6.
7. Available online. http://earl.eanm.org/cms/website.php?id=/en/projects/fdg_pet_ct_accreditation.htm.
8. Kaalep A et al., “EANM/EARL FDG-PET/CT accreditation - summary results from the first 200 accredited imaging systems”, Eur J Nucl Med Mol Imaging. 2018 Mar vol 45(3) 412–422.
9. Keverling Buisman AS, “Handboek Radionucliden”, BetaText v.o.f., 2015.
10. Joo Hyun O, “Practical PERCIST. A Simplified Guide to PET Response Criteria in Solid Tumors 1.0”. Radiology August. 2016;280(2):576–84.
11. Boellaard R. Effects of Noise, Image Resolution, and ROI Definition on the Accuracy of Standard Uptake Values: A Simulation Study. J Nucl Med. 2004;45:1519–27.
12. Abgral R, et al. Comparison of prognostic value of tumor SUL-peak and SUV-max on pretreatment FDG-PET/CT in patients with HNSCC. J Nucl Med May. 2013;54(2):513.
13. Julyan PJ. SUVpeak: a new parameter for quantification of uptake in FDG PET. Nucl Med Com April. 2004;25(4):407.
14. Lodge MA. “Repeatability of SUV in Oncologic 18F-FDG PET”, J Nucl Med. Apr 2017 58(4) 523–532.
15. Weber WA, et al. Reproducibility of metabolic measurements in malignant tumors using FDG PET. J Nucl Med. 1999;40:1771–7.
16. Nahmias C, et al. Reproducibility of standardized uptake value measurements determined by 18F-FDG PET in malignant tumors. J Nucl Med. 2008;49:1804–8.
17. Syahir M, et al. Impact of PET/CT system, reconstruction protocol, data analysis method, and repositioning on PET/CT precision: An experimental evaluation using an oncology and brain phantom. Med Phys. 2017;44:vol 12 6413–24.
18. Kumar V, et al. Variance of Standardized Uptake Values for FDG-PET/CT Greater in Clinical Practice than Under Ideal Study Settings. Clin Nucl Med. 2013 Mar;38(3):175–82.
19. Lodge MA, et al. A PET study of 18FDG uptake in soft tissue masses. Eur J Nucl Med. 1999;26:22–30.
20. Fahey FH, et al. Variability in PET quantitation within a multicenter consortium. Med Phys. 2010;37:3660–6.

Figures
Figure 1

SUVMax 2D and 3D, SUVMean 2D and 3D and SUVPeak for spheres with different volume for different reconstruction lengths.
Figure 2

Variation of the SUVMax 2D as a function of the reconstruction length for spheres of different diameter.
Figure 3

Variation of the SUVMax 3D as a function of the reconstruction length for spheres of different diameter.
Figure 4

Variation of the SUVMean 2D as a function of the reconstruction length for spheres of different diameter.
Figure 5

Variation of the SUVMean 3D as a function of the reconstruction length for spheres of different diameter.
Figure 6

Variation of the SUVPeak as a function of the reconstruction length for spheres of different diameter.
Figure 7

Estimated variation of the SUV parameters between repetitions of images with reconstruction length 150 s for a sphere of 10 mm.
**Figure 8**

Estimated variation of the SUV parameters between repetitions of images with reconstruction length 150 s for a sphere of 13 mm.
Figure 9

Estimated variation of the SUV parameters between repetitions of images with reconstruction length 150 s for a sphere of 17 mm.
Figure 10

Estimated variation of the SUV parameters between repetitions of images with reconstruction length 150 s for a sphere of 22 mm.
Figure 11

Estimated variation of the SUV parameters between repetitions of images with reconstruction length 150 s for a sphere of 28 mm.
Figure 12

Estimated variation of the SUV parameters between repetitions of images with reconstruction length 150 s for a sphere of 37 mm.
Figure 13

Variation of the SUVMax 2D as a function of the reconstruction length for spheres of different diameter with the expected variation range calculated according to Formula 1 and the other subsets.
Figure 14

Variation of the SUVMax 3D as a function of the reconstruction length for spheres of different diameter with the expected variation range calculated according to Formula 1 and the other subsets.
Figure 15

Variation of the SUVMean 2D as a function of the reconstruction length for spheres of different diameter with the expected variation range calculated according to Formula 1 and the other subsets.
Figure 16

Variation of the SUVMean 3D as a function of the reconstruction length for spheres of different diameter with the expected variation range calculated according to Formula 1 and the other subsets.
Figure 17

Variation of the SUVPeak as a function of the reconstruction length for spheres of different diameter with the expected variation range calculated according to Formula 1 and the other subsets.