ORIGINAL ARTICLE

Harmonizing SUVs in multicentre trials when using different generation PET systems: prospective validation in non-small cell lung cancer patients

Charline Lasnon · Cédric Desmonts · Elske Quak ·
Radj Gervais · Pascal Do · Catherine Dubos-Arvis ·
Nicolas Aide

Abstract

Purpose We prospectively evaluated whether a strategy using point spread function (PSF) reconstruction for both diagnostic and quantitative analysis in non-small cell lung cancer (NSCLC) patients meets the European Association of Nuclear Medicine (EANM) guidelines for harmonization of quantitative values.

Methods The NEMA NU-2 phantom was used to determine the optimal filter to apply to PSF-reconstructed images in order to obtain recovery coefficients (RCs) fulfilling the EANM guidelines for tumour positron emission tomography (PET) imaging (PSF_{EANM}). PET data of 52 consecutive NSCLC patients were reconstructed with unfiltered PSF reconstruction ($PSF_{allpass}$), PSF_{EANM} and with a conventional ordered subset expectation maximization (OSEM) algorithm known to meet EANM guidelines. To mimic a situation in which a patient would undergo pre- and post-therapy PET scans on different generation PET systems, standardized uptake values (SUVs) for OSEM reconstruction were compared to SUVs for PSF_{EANM} and $PSF_{allpass}$ reconstruction.

Results Overall, in 195 lesions, Bland-Altman analysis demonstrated that the mean ratio between PSF_{EANM} and OSEM data was 1.03 [95 % confidence interval (CI) 0.94–1.12] and 1.02 (95 % CI 0.90–1.14) for SUV_{max} and SUV_{mean}, respectively. No difference was noticed when analysing lesions based on their size and location or on patient body habitus and image noise. Ten patients (84 lesions) underwent two PET scans for response monitoring. Using the European Organization for Research and Treatment of Cancer (EORTC) criteria, there was an almost perfect agreement between $OSEM_{PET1}/OSEM_{PET2}$ (current standard) and $OSEM_{PET1}/PSF_{EANM-PET2}$ or $PSF_{EANM-PET1}/OSEM_{PET2}$ with kappa values of 0.95 (95 % CI 0.91–1.00) and 0.99 (95 % CI 0.96–1.00), respectively. The use of $PSF_{allpass}$ either for pre- or post-treatment (i.e. $OSEM_{PET1}/PSF_{allpass-PET2}$ or $PSF_{allpass-PET1}/OSEM_{PET2}$) showed considerably less agreement with kappa values of 0.75 (95 % CI 0.67–0.83) and 0.86 (95 % CI 0.78–0.94), respectively.

Conclusion Protocol-optimized images and compliance with EANM guidelines allowed for a reliable pre- and post-therapy evaluation when using different generation PET systems. These data obtained in NSCLC patients could be extrapolated to other solid tumours.

Charline Lasnon and Cédric Desmonts contributed equally to this work.

Electronic supplementary material The online version of this article (doi:10.1007/s00259-013-2391-1) contains supplementary material, which is available to authorized users.

C. Lasnon · E. Quak · N. Aide
Nuclear Medicine Department, François Baclesse
Cancer Centre, Caen, France

C. Desmonts
Nuclear Medicine Department, Caen University Hospital,
Caen, France

R. Gervais · P. Do · C. Dubos-Arvis
Thoracic Oncology, François Baclesse Cancer Centre,
Caen, France

N. Aide (✉)
Service de Médecine Nucléaire, Centre François Baclesse,
Avenue Général Harris,
14076 Caen cedex 5, France
e-mail: n.aide@baclesse.fr

Introduction

Over the last 20 years, ^{18}F-fluorodeoxyglucose (FDG) positron emission tomography (PET) has played an increasing role in the management of non-small cell lung cancer (NSCLC) patients for staging [1] and restaging [2, 3]. More recently, ^{18}F-FDG PET has been used for response evaluation of chemotherapy and molecularly targeted therapies [4–6]. The standardized uptake value (SUV) is the most frequently used quantitative parameter in oncology [7]. When using SUV as a diagnostic [8, 9] or prognostic [10, 11] tool (i.e. single measurement) or for therapy monitoring (i.e. longitudinal studies) in multicentre trials or in sites equipped with multiple scanners, one needs to minimize the variability in semiquantitative measurements by harmonizing both patient preparation in the PET unit and acquisition and reconstruction parameters [12–14].

The European Association of Nuclear Medicine (EANM) and the Society of Nuclear Medicine (SNM) have published guidelines [15, 16] regarding patient preparation, data acquisition, reconstruction parameters and definition of volume of interest (VOI) in or around the tumours. With regard to reconstruction parameters, the EANM guidelines, in line with the Netherlands protocol for standardization and quantification of ^{18}F-FDG PET studies in multicentre trials [17], provide recommendations based on an expected spatial resolution of the PET system equal to 7 mm. These recommendations include the use of the NEMA NU-2 phantom to check that activity concentration recoveries are concordant with those expected. Regarding quantitative analysis, SUV_{max} is currently the most frequently used quantitative parameter in oncological studies [18] despite being a suboptimal parameter due to noise-induced bias [19]. Therefore the EANM guidelines focus on getting comparable SUVs when using SUV_{max} in multicentre studies.

Hardware and software evolutions can lead to important device-dependent and reconstruction-dependent variations in quantitative values [20–22]. For instance, point spread function (PSF) reconstruction, which improves spatial resolution throughout the entire field of view, has recently become commercially available in clinical PET/CT systems. Our group has shown that, by improving activity recovery, especially for non-enlarged nodes, PSF reconstruction significantly improves the diagnostic performance of ^{18}F-FDG PET for nodal staging in NSCLC [23]. On average, PSF reconstruction increases SUV_{max} and SUV_{mean} by 48 and 28 %, respectively. As a result, recovery coefficient (RC) values obtained with PSF reconstruction are much higher than EANM's expected activity concentration recoveries as shown recently by Boellaard [24].

There is therefore a need for standardization of reconstruction protocols, keeping in mind that centres running PET systems with advanced reconstruction algorithms that participate in multicentre trials often wish to use their PET system with parameters chosen in order to achieve optimal lesion detection. A solution to optimize PET image quality for diagnostic purposes and at the same time to be able to use quantitative values within the framework of multicentre trials is the use of an additional filtering step [25] or to generate two sets of images: one to provide optimal diagnostic quality and a second one to meet quantitative harmonizing standards [24], with NEMA NU-2 phantom-based filtering chosen so that activity concentration recoveries are as close as possible to those recommended by EANM guidelines.

We aimed at prospectively evaluating such a strategy in NSCLC patients imaged on a PET/CT system equipped with PSF reconstruction. For that purpose, in order to mimic a situation in which a patient would undergo pre- and post-treatment scans on different generation PET systems, the same PET raw data were reconstructed with an ordered subset expectation maximization (OSEM) algorithm known to produce activity concentration recoveries meeting EANM requirements, PSF reconstruction for optimal tumour detection and PSF reconstruction with a filter optimized to fulfil EANM requirements. In addition, the potential impact of several confounding factors [tumour size, location and type as well as patient body mass index (BMI) and image noise] on the accuracy of our method was studied.

Materials and methods

Patient population

During 6 months, 52 patients referred to our institution for staging or restaging of a NSCLC were included in this study. The study was approved by the local Ethics Committee (ref A12-D24-VOL13, *Comité de protection des personnes Nord Ouest III*) waiving signed informed consent. Among these patients, ten underwent two PET examinations for the purpose of therapy monitoring. Patient demographics are described in Table 1.

Calibration and cross-calibration of the PET system

The calibration of the PET system was performed daily with a [68]Ge cylinder with a known radioactive concentration. This cross-calibration procedure was performed twice during the present study. A solution of ^{18}F-FDG (70.6 and 70.5 MBq, as assessed by the dose calibrator) was introduced into a cylindrical phantom with an exactly known volume and completed with water, which resulted in a solution with an exactly known concentration. A two-bed acquisition of the phantom was performed and images were reconstructed with attenuation

Table 1 Patient demographics

Characteristic	
Sex ratio (M/F)	7.7
Age (years)	
Range	46–80
Mean (SD)	63.9 (7.9)
Body habitus, n (%)	
BMI<25	22 (42.3)
BMI ≥ 25 to<30	22 (42.3)
BMI≥30	8 (15.4)
Histological diagnosis, n (%)	
Adenocarcinoma	26 (50.0)
Squamous cell carcinoma	18 (34.6)
Undifferentiated carcinoma	4 (7.7)
Large cell carcinoma	2 (3.9)
Adenosquamous carcinoma	1 (1.9)
Neuroendocrine carcinoma	1 (1.9)

and scatter correction identical to patient studies. Twelve VOIs were drawn on consecutive axial slices to determine the average activity concentration of ^{18}F-FDG within the phantom. The cross-calibration factor was calculated as the ratio of the calculated activity and the true activity. The cross-calibration factors were found to be 0.99 and 1.04.

Phantom preparation

The phantom set is the International Electrotechnical Commission body phantom set, which consists of a torso cavity containing a 5-cm-diameter cylindrical insert filled with foam pellets with an average density of 0.30 g/ml positioned in the centre of the phantom to simulate lung tissue and six coaxial isocentred spheres with internal diameters of 10, 13, 17, 22, 28 and 37 mm. According to the EANM guidelines, the phantom was filled with a solution of ^{18}F-FDG (2.0 kBq/ml) and all of the spheres with a radioactivity concentration of 20.0 kBq/ml resulting in a lesion to background activity ratio equal to 10.

Patient studies

The weight and height of patients on the day of the PET examination were recorded. BMI was computed as follows and was used to separate overweight (BMI>25 to <30 kg/m^2) and obese patients (BMI≥30 kg/m^2) from low to normal weight patients (BMI<25 kg/m^2):

$$BMI = \frac{weight\,(kg)}{height^{\,2}(m^2)}$$

After a 15-min rest in a warm room, patients who had been fasting for 6 h were injected with ^{18}F-FDG. Mean (SD) injected activity was 4 (0.2) MBq per kg of body weight.

The delay (SD) between tracer injection and image acquisition was 62 (4) min, thus meeting EANM guidelines [15].

PET/CT acquisition and reconstruction parameters

All PET imaging studies were performed on a Biograph TrueV (Siemens Medical Solutions) with a 6-slice spiral CT component. The technical and performance characteristics of the PET component of the TrueV system can be found elsewhere [26].

CT acquisition was performed first, with the following parameters: 60 mAs, 130 kVp, pitch 1 and 6×2 mm collimation. Subsequently, the PET emission acquisition was performed in 3-D mode. Patients were scanned from the skull base to the mid-thighs. For low to normal weight and overweight to obese patients, the duration was 2 min 40 s and 3 min 40 s, respectively. For phantom scanning, two bed positions were performed. The duration of each bed position was set to 2 min 40 s and 10 min, as per EANM guidelines. In addition, phantom studies with durations of 1 min 40 s and 3 min 40 s were performed in order to study the impact of image noise on the accuracy of our method.

In our department, PET images are reconstructed with a PSF reconstruction algorithm (HD; TrueX, Siemens Medical Solutions; 3 iterations and 21 subsets) without filtering (PSF$_{allpass}$), as modelling the PSF during iterative reconstruction introduces correlations between neighbouring voxels in a manner similar to smoothing filters and thus has been shown to achieve maximal performance with little or no filtering [27].

For the purpose of this study, raw data were also reconstructed with the OSEM 3-D reconstruction algorithm (4 iterations and 8 subsets) and the PSF reconstruction algorithm (HD; TrueX, Siemens Medical Solutions; 3 iterations and 21 subsets) using a Gaussian filter and an increasing kernel ranging from 6 to 8 mm with a 0.5-mm increment. Only the PSF-reconstructed data without filtering were used for the purpose of diagnostic workup. The OSEM reconstruction parameters were chosen as recommended by the manufacturer. These parameters meet the EANM requirements regarding activity recoveries and they were recently used by another group with the same PET system [28]. For all reconstructions, matrix size was 168×168, resulting in a 4.07×4.07×4.07 mm voxel size. Scatter and attenuation corrections were applied.

PET/CT analysis

Phantom studies

Activity concentration RCs as a function of sphere (tumour) size were measured. RCs are defined as the ratio between measured and true activity concentration in a sphere. For that purpose, 3-D 50 % isocontour VOIs were drawn over each sphere for each set of reconstructed data and maximum and mean pixel values were recorded.

Patient analysis

The same reader (CL) analysed all PET data sets to extract PET quantitative values for OSEM and PSF reconstructions. Regions of interest (ROIs) were drawn over primary tumour lesions, mediastinal and hilar nodes considered to have pathologically increased uptake and metastatic lesions. ROIs were drawn on the axial slice on which lesions displayed the highest ^{18}F-FDG uptake, by means of a 50 % isocontour method.

The mean and maximum pixel values were extracted from each ROI and mean and maximum SUVs were computed as follows:

$$SUV = \frac{tumour\ activity\ (Bq/cc) \times body\ weight(g)}{injected\ dose\ (Bq)}$$

Finally, short axis size (cm), as determined on axial CT slices, was recorded for each mediastinal and hilar lymph node.

For patients who underwent a post-therapeutic examination, the post-therapeutic status of each lesion was determined by using European Organization for Research and Treatment of Cancer (EORTC) criteria [29, 30]. SUV_{max}, recorded as described above, was used. The changes in SUV_{max} between the PET1 and PET2 scans were recorded for all lesions. The percentage change in SUV_{max} allowed classification into the following groups:

- Complete metabolic response (CMR): complete resolution of ^{18}F-FDG uptake in the tumour volume (indistinguishable from surrounding normal tissue)
- Partial metabolic response (PMR): at least 25 % reduction in tumour uptake
- Stable metabolic disease (SMD): less than 25 % increase or less than 25 % decrease in tumour ^{18}F-FDG SUV and no visible increase in extent of tumour uptake
- Progressive metabolic disease (PMD): greater than 25 % increase in ^{18}F-FDG tumour SUV within the tumour

Statistical analysis

The first step of the analysis was to determine the optimal filter settings for PSF reconstruction to meet EANM harmonizing standards. For that purpose, for all sets of reconstructed data, RCs for all spheres were compared to EANM expected values by means of the root mean square error (RMSE) method. The kernel size that minimizes the RMSE when compared to EANM expected values was selected as the optimal filter for PSF reconstruction on our PET/CT system. RMSE were computed with R, a freeware statistical package (http://www.r-project.org/foundation/).

Quantitative data extracted from clinical PET/CT examinations are presented as mean (standard deviation, SD). In all

Fig. 1 Recovery coefficients for mean (**a**) and maximum (**b**) values for OSEM 3-D reconstruction algorithm, PSF reconstruction algorithm without filtering (PSF$_{allpass}$) and PSF reconstruction algorithm with a 7-mm Gaussian filter (PSF$_{EANM}$). Corresponding NEMA NU-2 transverse images through the hot spheres (**c**). Phantom images are scaled on the same maximum value
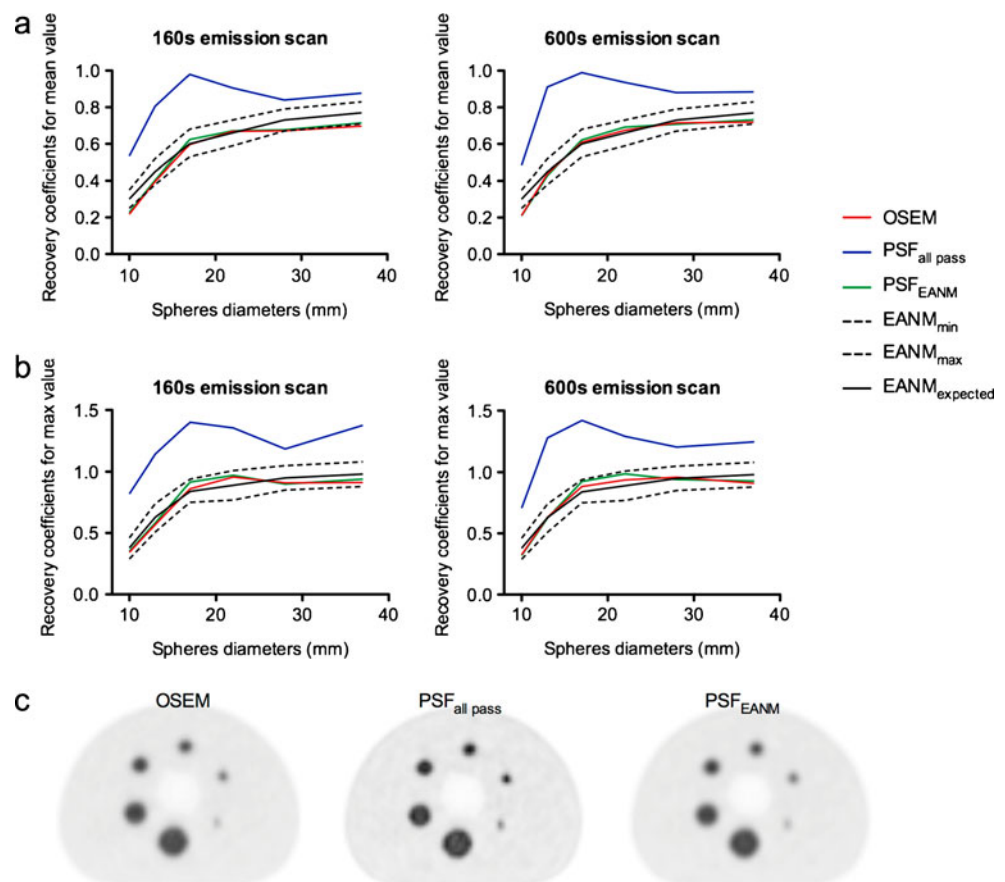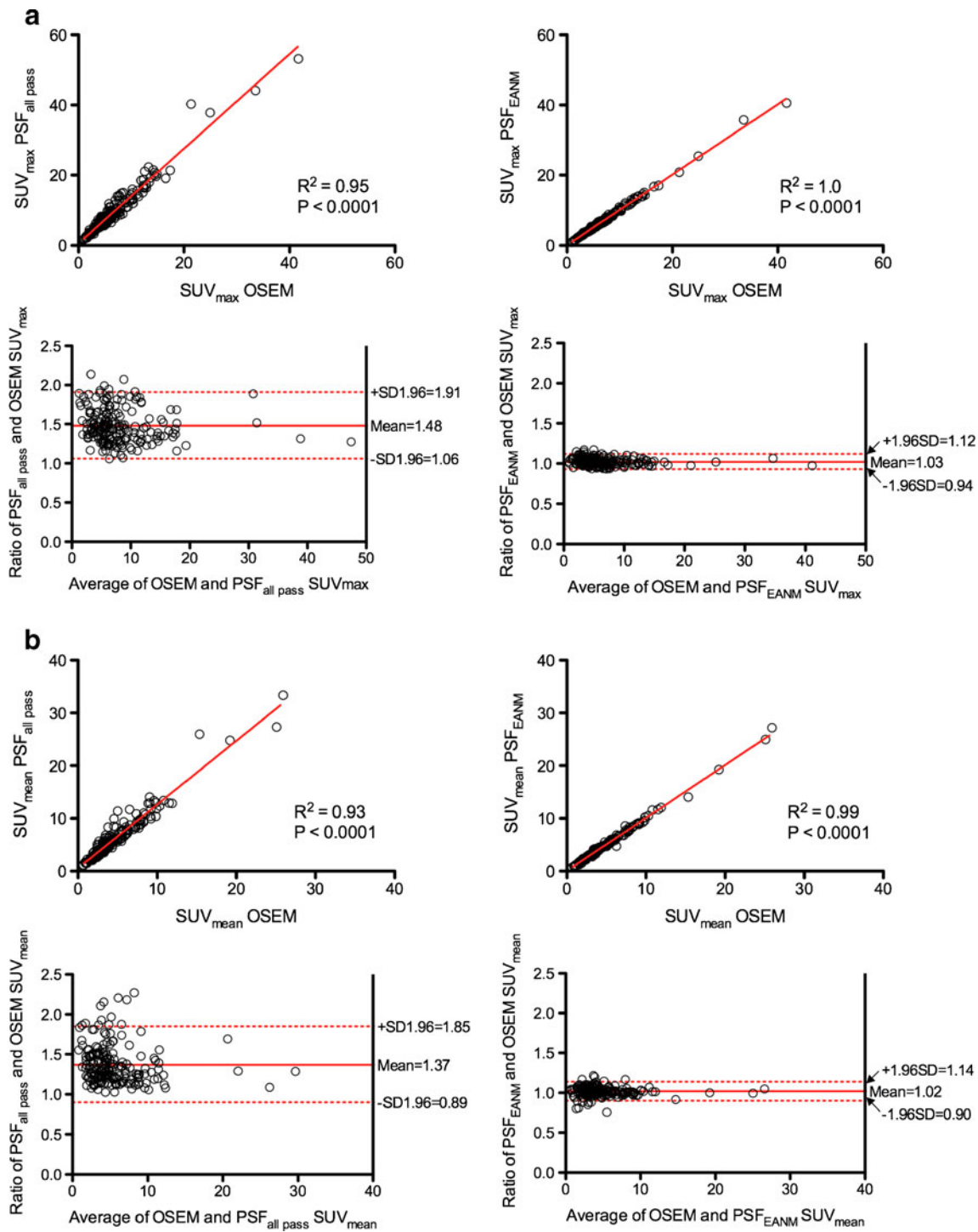
Fig. 2 Relationship between quantitative values extracted from $PSF_{allpass}$ or PSF_{EANM} and OSEM images, assessed using linear regression analysis and Bland-Altman plots for SUV_{max} (**a**) and SUV_{mean} (**b**)

statistical tests, a two-tailed p value of less than 0.05 was considered statistically significant. The ratios between PSF_{EANM} and OSEM quantitative values (SUV_{mean}, SUV_{max}), according to lesion size, location and type (heterogeneous vs homogeneous uptake), BMI (low to normal weight vs overweight vs obese patients) and acquisition time per bed position

(2 min 40 s vs 3 min 40 s) were compared using the Mann–Whitney test for unpaired samples and the Kruskal-Wallis test to compare multiple groups. The relationship between $PSF_{allpass}$ or PSF_{EANM} and OSEM quantitative values was assessed using a linear regression analysis and Bland-Altman plots [31]. In the subset of ten patients that underwent two

PET/CT examinations for therapy monitoring purposes, levels of agreement between the different types of reconstruction were evaluated using the kappa statistic. The use of OSEM reconstruction both for pre- and post-therapeutic PET examination ($OSEM_{PET1}/OSEM_{PET2}$) was used as the "current standard" to determine the post-treatment status of each lesion. This was compared to the use of PSF_{EANM} reconstruction either for pre-therapeutic PET evaluation ($PSF_{EANM-PET1}/OSEM_{PET2}$) or for post-therapeutic PET evaluation ($OSEM_{PET1}/PSF_{EANM-PET2}$), to the use of $PSF_{allpass}$ reconstruction either for pre-therapeutic PET evaluation ($PSF_{allpass-PET1}/OSEM_{PET2}$) or for post-therapeutic PET evaluation ($OSEM_{PET1}/PSF_{allpass-PET2}$) and to the use of PSF_{EANM} reconstruction for both pre- and post-therapeutic PET evaluation ($PSF_{EANM-PET1}/PSF_{EANM-PET2}$). Kappa values were reported using the benchmarks of Landis and Koch [32] (0.81–1 almost perfect agreement, 0.61–0.8 substantial agreement, 0.41–0.6 moderate agreement and 0.21–0.4 fair agreement). For the kappa estimates, 95 % confidence intervals were calculated using bootstrapping. Graphs and analyses were carried out using the GraphPad software and VassarStats (http://vassarstats.net/).

Results

Phantom data

As shown in Fig. 1, the OSEM 3-D reconstruction algorithm RCs for mean and maximum values fulfilled the EANM recommendations for both the 160-s and the 600-s emission scan. It is noticeable that for mean values (Fig. 1a), the OSEM RCs of the smallest spheres were slightly below the proposed minimum EANM specification. As expected, RCs for mean and maximum values of the PSF reconstruction algorithm without filtering were above the maximum EANM specifications whatever the duration of the emission scans, especially for the smallest hot spheres. When considering maximum values (Fig. 1b), with the exception of the 10-mm sphere, $PSF_{allpass}$ RCs were even greater than 1.0. This can be explained by the fact that PSF modelling results in overshoot along the edge. This artefact (the so-called Gibbs artefact [21, 33, 34]) was visible for the largest sphere for $PSF_{allpass}$ reconstruction and was partially corrected for by applying the Gaussian filters. When using shorter acquisition times, there were higher noise levels, which in combination with the Gibbs artefact led to less accurate (overestimated) measurements, especially for the maximum pixel value. The application of Gaussian filters with an increasing kernel during PSF reconstruction allowed for RCs to be more consistent with the EANM recommendations. When calculating the RMSE, the kernel size that minimized the error

Fig. 3 Impact of the size of the lesion (**a**), the BMI (**b**), the location of ▶ the lesion (**c**), tumour homogeneity (**d**) and emission scan duration (**e**) on the ratio between PSF_{EANM} PET and OSEM PET quantitative values (*left panels* SUV_{max}, *right panels* SUV_{mean}). Note that 30 lesions were not measurable and are therefore not included in the "per size" analysis (**a**)

compared to EANM expected values was the kernel of 7 mm (supplementary material). This kernel size of 7 mm was then selected as the optimal filter for PSF reconstruction (PSF_{EANM}).

An evaluation of the potential impact of image noise on the accuracy of our method was performed in a second experiment by scanning the phantom for 1 min 40 s, 2 min 40 s and 3 min 40 s. As expected, the RC values for $PSF_{allpass}$ reconstruction were higher for the shortest acquisition, due to noise in the reconstructed images (supplementary material: Fig. 1). However, calculation of the RMSE showed that our strategy performed well even when image noise was higher (supplementary material: Table 4).

Clinical data

Validation of the PSF_{EANM} reconstruction to overcome reconstruction-dependent variability

A total of 52 consecutive patients with NSCLC were included, for whom clinical data are summarized in Table 1. Among these patients, 36 were referred for initial staging of NSCLC and 16 for restaging of NSCLC recurrence.

Overall, 195 ROIs were drawn over 64 (32.8 %) primary tumour lesions, 91 (46.7 %) mediastinal and hilar nodes considered to have pathologically increased uptake and 40 (20.5 %) visceral and bone metastatic lesions. The mean (SD) number of lesions per patient, all types combined, was 3.8 (3.6). Among the 91 analysed nodes, 45 (49.4 %) had a short axis less than 1 cm [mean (SD) short axis, 0.80 (0.13)], whereas 46 (50.6 %) had a short axis 1 cm or greater [mean (SD) short axis, 1.46 (0.43)]. The mean SUV_{mean} (SD) for OSEM, PSF_{EANM} and PSF reconstruction were 4.70 (3.43), 4.77 (3.46) and 6.24 (4.30), respectively. The mean SUV_{max} (SD) for OSEM, PSF_{EANM} and PSF reconstruction were 6.60 (4.95), 6.71 (4.97) and 9.52 (6.85), respectively. Linear regression and Bland-Altman analysis are shown in Fig. 2. As expected, a good correlation was found between quantitative values extracted from the PSF and OSEM reconstructions, with an r^2 greater than 0.90 for both SUV_{max} and SUV_{mean} values. As shown in the Bland-Altman analysis, PSF reconstruction increased SUV_{max} and SUV_{mean} by 48 and 37 %, respectively. An even better correlation was found between PSF_{EANM} and OSEM reconstruction with r^2 equal to 1.0 for SUV_{max} and close to 1.0 for SUV_{mean} (0.99). Bland-Altman analysis demonstrated that the mean ratios between PSF_{EANM} and OSEM quantitative values
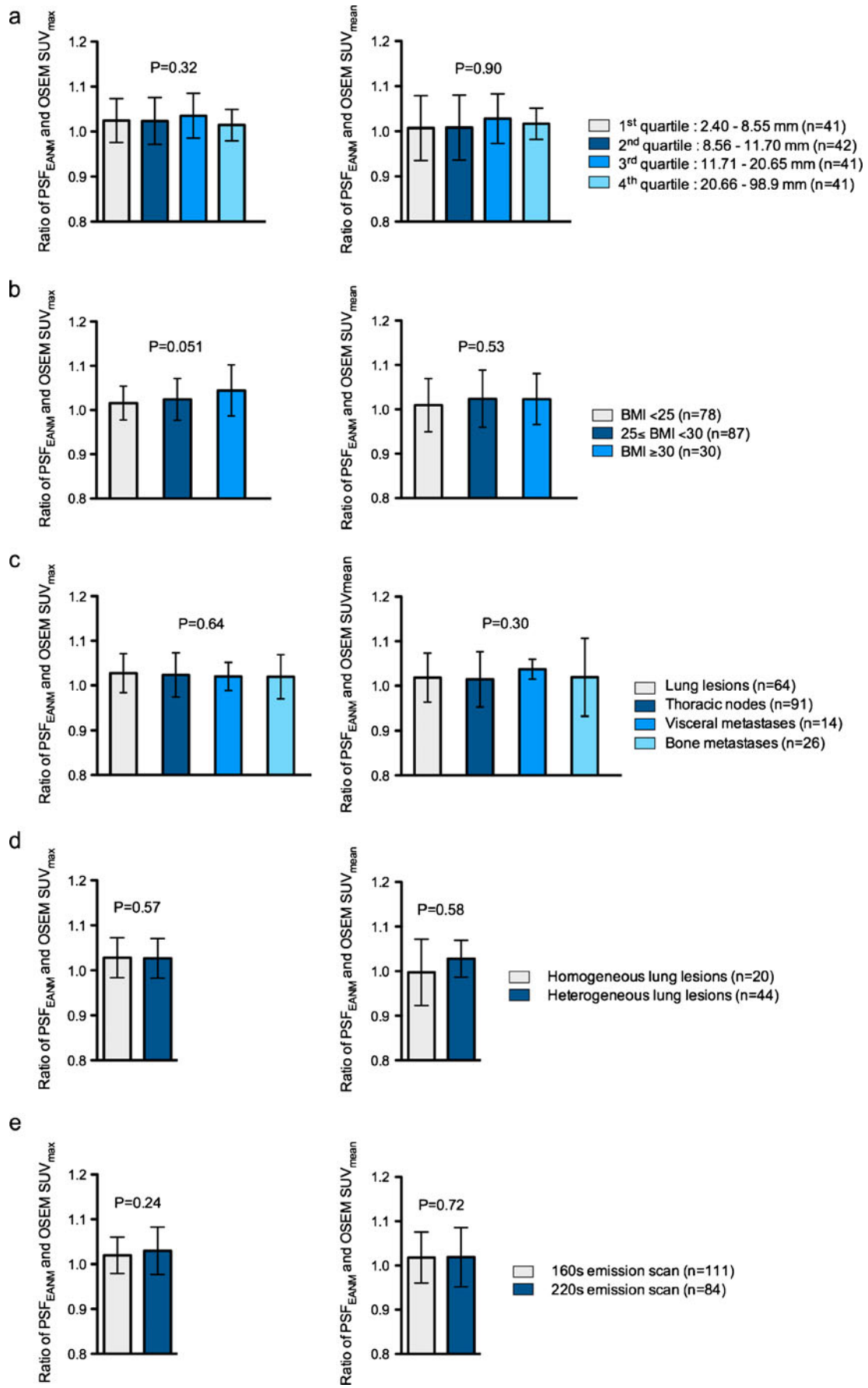
were 1.03 and 1.02 for SUV_{max} and SUV_{mean}, respectively, with very narrow 95 % limits of confidence in both cases. Amongst the 195 analysed lesions, Bland-Altman plots identified 8 outliers for the SUV_{max} values for which the ratios of SUV_{max} PSF_{EANM} and SUV_{max} OSEM were all above the upper limit of the confidence interval. These outliers corresponded to one tumour, five nodes (four mediastinal nodes and one hilar node) and two bone metastases. For the SUV_{mean} values, Bland-Altman plots identified 14 outliers of which 9 had a ratio below the lower limit of the confidence interval (2 tumours, 2 mediastinal nodes, 2 hilar nodes, 2 bone metastases and 1 lung metastasis) and 5 above the upper limit of the confidence interval (1 tumour, 1 hilar node and 3 bone metastases).

As shown in Fig. 3, the ratios between PSF_{EANM} and OSEM quantitative values (SUV_{max} and SUV_{mean}) were not different according to the size of the lesion. The mean ratio (SD) for SUV_{max} values and SUV_{mean} values (SD) ranged from 1.01 (0.04) (4th quartile) to 1.04 (0.05) (3rd quartile) and from 1.01 (0.07) (1st quartile) to 1.03 (0.06) (3rd quartile), respectively. Similarly, there was no significant difference according to the BMI, the location of the lesion or the type of lesion (homogeneous versus heterogeneous). When analysing the ratios between PSF_{EANM} and OSEM quantitative values for SUV_{max} values according to BMI, there was a trend towards higher ratios ($p=0.051$) in obese patients.

An example of OSEM, $PSF_{allpass}$ and PSF_{EANM} reconstructions is shown in Fig. 4.

### The use of PSF_{EANM} quantitative values for therapy monitoring

Among the series of 52 consecutive patients, 10 patients underwent both a pre- and post-therapy PET evaluation with

an average time between the first and the second PET scan of 72.6±34.6 days (Table 2).

Overall, 84 lesions were evaluated post-treatment: 12 (14.3 %) primary tumour lesions, 41 (48.8 %) mediastinal and hilar nodes and 31 (36.9 %) visceral and bone metastatic lesions. When OSEM reconstruction was used for interpreting both pre- and post-therapeutic PET examinations ($OSEM_{PET1}$/ $OSEM_{PET2}$), 37 lesions were considered to have had a CMR, 28 a PMR, 13 were stable and 6 had progressed. These results were then compared to several scenarios (Fig. 5, Table 3) when using PSF_{EANM}, OSEM or $PSF_{allpass}$ for either the pre- or post-therapeutic PET examination or both. $OSEM_{PET1}$/$OSEM_{PET2}$ was regarded as the standard of reference. All lesions considered, there was almost perfect agreement between $OSEM_{PET1}$/ $OSEM_{PET2}$ and $OSEM_{PET1}$/$PSF_{EANM-PET2}$, $PSF_{EANM-PET1}$/$OSEM_{PET2}$ or $PSF_{EANM-PET1}$/$PSF_{EANM-PET2}$ with kappa values higher than 0.90. In addition, the associated 95 % confidence intervals virtually matched the almost perfect range of kappa values. When analysing tumours, nodes or visceral and bone metastases separately, the strength of agreement was also considered to be very good. There were four cases (4.8 %) of disagreement (two nodes and two metastatic lesions) in which $OSEM_{PET1}$/$PSF_{EANM-PET2}$ diagnosed stable disease, whereas $OSEM_{PET1}$/$OSEM_{PET2}$ identified partial response. When $PSF_{EANM-PET1}$/$OSEM_{PET2}$ was used, there was only one disagreement (1.2 %) that occurred in a node, coming to a conclusion of stable disease, whereas $OSEM_{PET1}$/$OSEM_{PET2}$ identified partial response. With $PSF_{EANM-PET1}$/$PSF_{EANM-PET2}$, there were two cases of disagreement that occurred in nodes, coming to a conclusion of stable disease, whereas $OSEM_{PET1}$/$OSEM_{PET2}$ identified partial response.

Importantly, when $PSF_{allpass}$ reconstruction was used either for the pre- or post-therapeutic examination ($OSEM_{PET1}$/ $PSF_{allpass-PET2}$ or $PSF_{allpass-PET1}$/$OSEM_{PET2}$), there was considerably less agreement. With $OSEM_{PET1}$/$PSF_{allpass-PET2}$, there were

Fig. 4 Representative coronal slices for OSEM, $PSF_{allpass}$ and PSF_{EANM} reconstructions in a patient with a lung tumour in the left upper lobe, bilateral nodal involvement (**a**) and distant metastases (lung, bone and liver) (**b**). Images have been scaled on the same maximum value. Note the improvement in activity recovery visible in a small lung metastasis on the $PSF_{allpass}$ image (*arrow*)
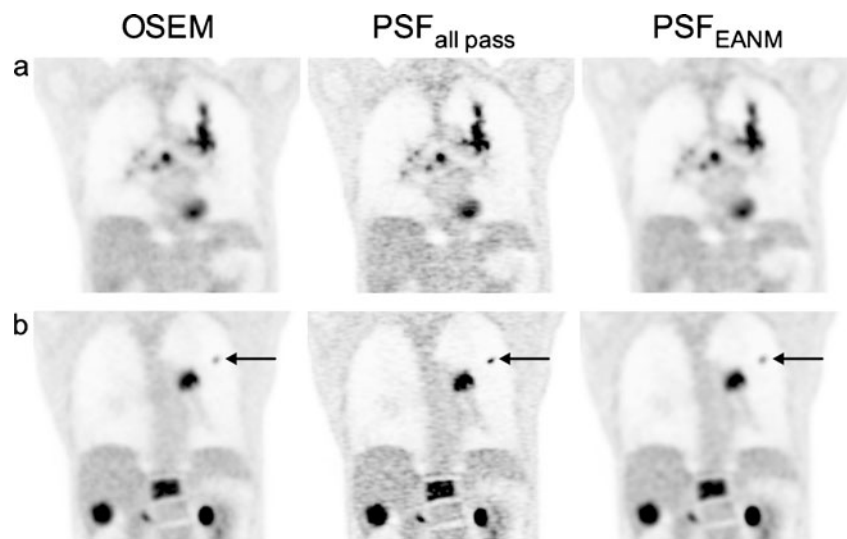
Table 2 Characteristics of patients who underwent post-therapy evaluation

Patient	Sex	Age	Number of lesions		Treatment	Delay between PET1 and PET2 (days)
			PET1	PET2		
1	F	79	7	3	Gefitinib	168
2	M	60	8	1	Cisplatin-pemetrexed	70
3	M	67	7	8	Cisplatin-vinorelbine	83
4	M	58	8	7	Carboplatin-pemetrexed	42
5	M	59	15	6	Cisplatin-pemetrexed-bevacizumab	56
6	M	80	1	2	Carboplatin-paclitaxel	54
7	M	65	11	8	Cisplatin-pemetrexed	52
8	M	53	2	2	Cisplatin-vinorelbine	105
9	M	54	6	6	Cisplatin-pemetrexed	49
10	M	65	13	5	Cisplatin-gemcitabine	118

F female, *M* male

overall 23 cases (27.4 %) of disagreement (4 tumours, 13 nodes and 6 metastatic lesions) in which $OSEM_{PET1}/PSF_{allpass-PET2}$ underestimated the therapeutic response when compared to $OSEM_{PET1}/OSEM_{PET2}$. With $PSF_{allpass-PET1}/OSEM_{PET2}$, there were 11 cases (13.1 %) of disagreement including 9 cases (6 nodes and 3 metastatic lesions) where a conclusion of partial response was reached, whereas $OSEM_{PET1}/OSEM_{PET2}$ diagnosed stable disease. The remaining two cases corresponded to tumours: one for which $PSF_{allpass-PET1}/OSEM_{PET2}$ reached a conclusion of stable disease, whereas $OSEM_{PET1}/OSEM_{PET2}$ identified partial response, and one for which $PSF_{allpass-PET1}/OSEM_{PET2}$ diagnosed progression, whereas $OSEM_{PET1}/OSEM_{PET2}$ identified stable disease.
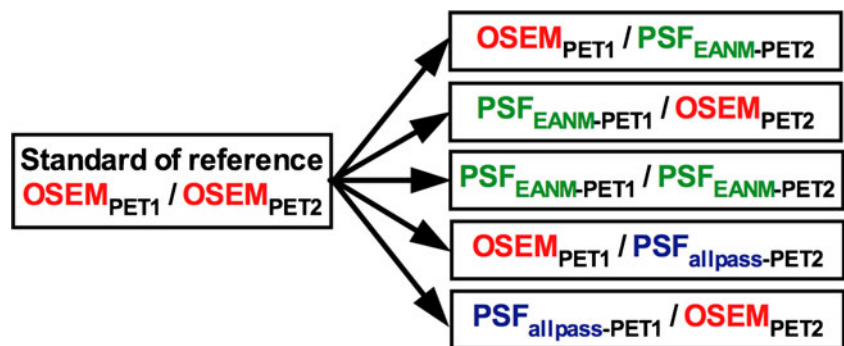
Discussion

^{18}F-FDG PET has an increasing role in oncology for staging, restaging and therapy monitoring of chemotherapy and molecularly targeted therapies and is being increasingly implemented in clinical trials, especially for the early assessment of antineoplastic treatments. This prospective study in NSCLC patients validates a strategy allowing the use of quantitative values within the framework of multicentre trials, which is based on the production of

protocol-specific images, in addition to images optimized for diagnostic purpose.

Standardized quantification of PET data in multicentre trials as described in the EANM guidelines allows for reliable and reproducible treatment response assessment. However, standardization remains a major challenge as new, more sensitive PET systems and reconstruction algorithms are continuously being developed and introduced into clinical practice [20, 23, 35]. In the present study, we validated a strategy in which the recently introduced PSF reconstruction algorithm can be used not only for visual but also for quantitative analysis of PET imaging, whilst adhering to the EANM guidelines. Our results demonstrate, by mimicking a situation in which a patient would undergo the pre- and post-therapy PET scans on different generation PET systems, that it is possible to minimize reconstruction-dependent variability. Hence, Bland-Altman analysis (Fig. 2) showed that after having applied an adequate filter (PSF_{EANM}) the upper limit of the confidence intervals was 12 %, a value well below the 25 and 30 % cut-off values recommended by EORTC [30] and PERCIST [36], respectively, to discriminate between responders and non-responders when using ^{18}F-FDG PET for therapy monitoring. Importantly, we confirmed this finding in a subset of ten patients who underwent two PET examinations for response assessment (Table 2). In these

Fig. 5 Flow chart for the evaluation of the level of agreement when using PSF_{EANM} or $PSF_{allpass}$ reconstructions for response monitoring (EORTC criteria) either pre- or post-treatment as compared to the exclusive use of OSEM reconstruction (current standard)

Table 3 Impact of PSF_{EANM} on response evaluation

	Agreements (95 % CI)				
	$OSEM_{PET1}/OSEM_{PET2}$ vs $OSEM_{PET1}/PSF_{EANM-PET2}$	$OSEM_{PET1}/OSEM_{PET2}$ vs $PSF_{EANM-PET1}/OSEM_{PET2}$	$OSEM_{PET1}/OSEM_{PET2}$ vs $PSF_{EANM-PET1}/PSF_{EANM-PET2}$	$OSEM_{PET1}/OSEM_{PET2}$ vs $OSEM_{PET1}/PSF_{allpass-PET2}$	$OSEM_{PET1}/OSEM_{PET2}$ vs $PSF_{allpass-PET1}/OSEM_{PET2}$
All lesions	0.95 (0.91–1.00)	0.99 (0.96–1.00)	0.98 (0.94–1.00)	0.75 (0.67–0.83)	0.86 (0.78–0.94)
Tumours	1.00 (1.00–1.00)	1.00 (1.00–1.00)	1.00 (1.00–1.00)	0.59 (0.28–0.90)	0.76 (0.52–0.99)
Nodes	0.94 (0.86–1.00)	0.97 (0.91–1.00)	0.94 (0.86–1.00)	0.69 (0.59–0.78)	0.80 (0.67–0.92)
Metastasis	0.95 (0.88–1.00)	1.00 (1.00–1.00)	1.00 (1.00–1.00)	0.83 (0.70–0.96)	0.92 (0.83–1.00)

Linear-weighted kappa values for all lesions as well as for different lesion types, when using PSF_{EANM} or $PSF_{allpass}$ reconstructions for response monitoring (EORTC criteria) either pre- or post-treatment as compared to the exclusive use of OSEM reconstruction (current standard)

CI confidence interval

patients, an excellent agreement was found (kappa values 0.95 and 0.99) in the post-treatment classification of 84 lesions according to EORTC criteria when comparing PSF_{EANM} either pre- or post-therapy to OSEM as the current standard, and no major discordance occurred. However, when the $PSF_{allpass}$ data were used either pre- or post-therapy compared to OSEM, we saw considerably less agreement. Due to system updates on existing PET systems or the purchase of a new PET machine, $OSEM_{PET1}/PSF_{allpass-PET2}$ is the situation most likely to occur. In this situation, our data showed discordance in 27.4 % of lesions.

The proposed strategy can be useful in the case of patients undergoing pre- and post-treatment scans on different PET systems, for example in centres running two or more PET systems or updating their equipment during the course of a trial. Of course, it would be preferable to scan the patient repeatedly on the same machine, but in practice this is often not possible. Moreover, in the setting of multicentre trials there are two other situations in which standardization of PET quantitative values is required: when pooling SUV from different PET/CT systems for diagnostic purposes (i.e. to determine a specific diagnostic threshold value for a given disease) [8, 9] or as a prognostic tool (i.e. to search for the impact of tumour tracer uptake on disease-free and overall survival) [10, 11].

Regarding practical issues related to the proposed methodology, determination of the appropriate filter per PET system is required by performing the phantom studies and reconstructions with a Gaussian filter with increasing kernel as described in the "Materials and methods" section. Once the optimal filter meeting the EANM expected values is determined, the filtered PET data can be used for both local and multicentre quantitative PET analysis. This method can be readily applied on any PET scanner equipped with PSF; the purchase of additional software is not necessary. However, this method does not obviate the need to generate a second data set which is time consuming. Of course, the choice to use either an OSEM reconstruction or a filtered PSF algorithm for the standardized quantitative analysis remains a choice of local nuclear medicine physicians, physicists and researchers, just like the choice to systematically reconstruct non-attenuation-corrected images or only when clinically needed. Choosing PSF_{EANM} could be the preferred solution, as PSF reconstruction is meant to progressively replace former generation PET systems.

As pointed out by Boellaard [24], patients are frequently included in clinical trials after the first PET examination has been performed. This emphasizes the need to standardize the PET procedure from the very beginning of patient care. However, PET acquisition and reconstruction parameters are not the only source of variability that has to be taken into account. Other technical and biological factors also affect SUV measurements. These factors have been discussed extensively elsewhere [12, 24, 37]. In the present study, one technical factor, the reconstruction protocol, has been analysed. To minimize the influence of the other technical and biological factors affecting SUV measurements in this study, all PET examinations were performed according to the EANM guidelines. Of note, the injected activity per kilogram and the delay between injection and acquisition met the EANM requirements.

The potential impact of image noise on the accuracy of our method was evaluated in phantom studies by varying the acquisition time. Calculation of the RMSE values between PSF_{EANM} and EANM expected values showed that our strategy performed well when image noise was higher, the values being similar for the shortest and longest acquisition times. This was confirmed by clinical data showing no difference in PSF_{EANM}/OSEM ratios for the 2 min 40 s and 3 min 40 s per bed position acquisition times (Fig. 3e).

We found no confounding factors (lesion size and location, tumour heterogeneity, patient BMI) affecting the accuracy of our method. However, we noticed a trend towards higher PSF_{EANM}/OSEM ratios in overweight and obese patients for SUV_{max} (Fig. 3b). This may be due to the fact that noise in PET images is higher in obese patients and SUV_{max} is more affected by noise than SUV_{mean}. The

observed difference was minimal and did not affect the EORTC classification based on SUV_{max} (Table 3). The use of SUV_{peak}, which is defined as the mean value within an ROI centred on the area with the highest uptake, has been reported as a slightly more robust alternative for assessing the most metabolically active part of a tumour [19]. However, SUV_{peak} is highly sensitive to the ROI_{peak} definition (i.e. shape, size and location) [38], was shown to have similar repeatability as compared to SUV_{max} [39] and does not necessarily perform better than SUV_{max} for therapy assessment [40]. In the present study, a wide range of tumour intensities was studied and no systematic error was depicted by Bland-Altman analysis (i.e. the strategy performs equally for lesions with low ^{18}F-FDG avidity and for those with very intense ^{18}F-FDG uptake). This finding, taken together with the lack of confounding factors affecting our strategy, suggests that it could be applicable in other solid tumours.

Conclusion

The generation of protocol-specific images with NEMA NU-2 phantom-based filtering to meet EANM quantitative harmonizing standards, in addition to images optimized for diagnostic purposes, reduces reconstruction-dependent variation in SUVs. This can be of use in multicentre trials, when using SUV for therapy monitoring, or as a diagnostic or prognostic tool. As no confounding factors (lesion size and location, tumour heterogeneity, patient BMI, image noise) affecting the accuracy of our method were found, this strategy validated in NSCLC patients could be extrapolated to other solid tumours.

References

1. Lv YL, Yuan DM, Wang K, Miao XH, Qian Q, Wei SZ, et al. Diagnostic performance of integrated positron emission tomography/computed tomography for mediastinal lymph node staging in non-small cell lung cancer: a bivariate systematic review and meta-analysis. J Thorac Oncol 2011;6:1350–8.
2. Eschmann SM, Friedel G, Paulsen F, Reimold M, Hehr T, Budach W, et al. 18F-FDG PET for assessment of therapy response and preoperative re-evaluation after neoadjuvant radio-chemotherapy in stage III non-small cell lung cancer. Eur J Nucl Med Mol Imaging 2007;34:463–71.
3. Hicks RJ, Kalff V, MacManus MP, Ware RE, McKenzie AF, Matthews JP, et al. The utility of (18)F-FDG PET for suspected recurrent non-small cell lung cancer after potentially curative therapy: impact on management and prognostic stratification. J Nucl Med 2001;42:1605–13.
4. Bengtsson T, Hicks RJ, Peterson A, Port RE. 18F-FDG PET as a surrogate biomarker in non-small cell lung cancer treated with erlotinib: newly identified lesions are more informative than standardized uptake value. J Nucl Med 2012;53:530–7.
5. Hicks RJ. Role of 18F-FDG PET in assessment of response in non-small cell lung cancer. J Nucl Med 2009;50 Suppl 1:31S–42S.
6. Huang W, Zhou T, Ma L, Sun H, Gong H, Wang J, et al. Standard uptake value and metabolic tumor volume of (18)F-FDG PET/CT predict short-term outcome early in the course of chemoradiotherapy in advanced non-small cell lung cancer. Eur J Nucl Med Mol Imaging 2011;38:1628–35.
7. Visser EP, Boerman OC, Oyen WJ. SUV: from silly useless value to smart uptake value. J Nucl Med 2010;51:173–5.
8. Dijkman BG, Schuurbiers OC, Vriens D, Looijen-Salamon M, Bussink J, Timmer-Bonte JN, et al. The role of (18)F-FDG PET in the differentiation between lung metastases and synchronous second primary lung tumours. Eur J Nucl Med Mol Imaging 2010;37:2037–47.
9. Hellwig D, Graeter TP, Ukena D, Groeschel A, Sybrecht GW, Schaefers HJ, et al. 18F-FDG PET for mediastinal staging of lung cancer: which SUV threshold makes sense? J Nucl Med 2007;48:1761–6.
10. Agarwal M, Brahmanday G, Bajaj SK, Ravikrishnan KP, Wong CY. Revisiting the prognostic value of preoperative (18)F-fluoro-2-deoxyglucose ((18)F-FDG) positron emission tomography (PET) in early-stage (I & II) non-small cell lung cancers (NSCLC). Eur J Nucl Med Mol Imaging 2010;37:691–8.
11. Liao S, Penney BC, Wroblewski K, Zhang H, Simon CA, Kampalath R, et al. Prognostic value of metabolic tumor burden on 18F-FDG PET in nonsurgical patients with non-small cell lung cancer. Eur J Nucl Med Mol Imaging 2012;39:27–38.
12. Boellaard R. Standards for PET image acquisition and quantitative data analysis. J Nucl Med 2009;50 Suppl 1:11S–20S.
13. Boellaard R. Mutatis mutandis: harmonize the standard! J Nucl Med 2012;53:1–3.
14. Buckler AJ, Boellaard R. Standardization of quantitative imaging: the time is right, and 18F-FDG PET/CT is a good place to start. J Nucl Med 2011;52:171–2.
15. Boellaard R, O'Doherty MJ, Weber WA, Mottaghy FM, Lonsdale MN, Stroobants SG, et al. FDG PET and PET/CT: EANM procedure guidelines for tumour PET imaging: version 1.0. Eur J Nucl Med Mol Imaging 2010;37:181–200.
16. Delbeke D, Coleman RE, Guiberteau MJ, Brown ML, Royal HD, Siegel BA, et al. Procedure guideline for tumor imaging with 18F-FDG PET/CT 1.0. J Nucl Med 2006;47:885–95.
17. Boellaard R, Oyen WJ, Hoekstra CJ, Hoekstra OS, Visser EP, Willemsen AT, et al. The Netherlands protocol for standardisation and quantification of FDG whole body PET studies in multi-centre trials. Eur J Nucl Med Mol Imaging 2008;35:2320–33.
18. Beyer T, Czernin J, Freudenberg LS. Variations in clinical PET/CT operations: results of an international survey of active PET/CT users. J Nucl Med 2011;52:303–10.
19. Lodge MA, Chaudhry MA, Wahl RL. Noise considerations for PET quantification using maximum and peak standardized uptake value. J Nucl Med 2012;53:1041–7.
20. Akamatsu G, Ishikawa K, Mitsumoto K, Taniguchi T, Ohya N, Baba S, et al. Improvement in PET/CT image quality with a combination of point-spread function and time-of-flight in relation to reconstruction parameters. J Nucl Med 2012;53:1716–22.
21. Alessio AM, Stearns CW, Tong S, Ross SG, Kohlmyer S, Ganin A, et al. Application and evaluation of a measured spatially variant system model for PET image reconstruction. IEEE Trans Med Imaging 2010;29:938–49.

22. Westerterp M, Pruim J, Oyen W, Hoekstra O, Paans A, Visser E, et al. Quantification of FDG PET studies using standardised uptake values in multi-centre trials: effects of image reconstruction, resolution and ROI definition parameters. Eur J Nucl Med Mol Imaging 2007;34:392–404.

23. Lasnon C, Hicks RJ, Beauregard JM, Milner A, Paciencia M, Guizard AV, et al. Impact of point spread function reconstruction on thoracic lymph node staging with 18F-FDG PET/CT in non-small cell lung cancer. Clin Nucl Med 2012;37:971–6.

24. Boellaard R. Need for standardization of 18F-FDG PET/CT for treatment response assessments. J Nucl Med 2011;52 Suppl 2:93S–100S.

25. Kelly MD, Declerck JM. SUVref: reducing reconstruction-dependent variation in PET SUV. EJNMMI Res 2011;1:16.

26. Aide N, Desmonts C, Beauregard JM, Beyer T, Kinross K, Roselt P, et al. High throughput static and dynamic small animal imaging using clinical PET/CT: potential preclinical applications. Eur J Nucl Med Mol Imaging 2010;37:991–1001.

27. Kadrmas DJ, Casey ME, Conti M, Jakoby BW, Lois C, Townsend DW. Impact of time-of-flight on PET tumor detection. J Nucl Med 2009;50:1315–23.

28. Andersen FL, Klausen TL, Loft A, Beyer T, Holm S. Clinical evaluation of PET image reconstruction using a spatial resolution model. Eur J Radiol 2012.

29. Mortazavi-Jehanno N, Giraudet AL, Champion L, Lerebours F, Le Stanc E, Edeline V, et al. Assessment of response to endocrine therapy using FDG PET/CT in metastatic breast cancer: a pilot study. Eur J Nucl Med Mol Imaging 2012;39:450–60.

30. Young H, Baum R, Cremerius U, Herholz K, Hoekstra O, Lammertsma AA, et al. Measurement of clinical and subclinical tumour response using [18F]-fluorodeoxyglucose and positron emission tomography: review and 1999 EORTC recommendations. European Organization for Research and Treatment of Cancer (EORTC) PET Study Group. Eur J Cancer 1999;35:1773–82.

31. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;1:307–10.

32. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159–74.

33. Zeng GL. Gibbs artifact reduction by nonnegativity constraint. J Nucl Med Technol 2011;39:213–9.

34. Boussion N, Cheze Le Rest C, Hatt M, Visvikis D. Incorporation of wavelet-based denoising in iterative deconvolution for partial volume correction in whole-body PET imaging. Eur J Nucl Med Mol Imaging 2009;36:1064–75.

35. Levin Klausen T, Høgild Keller S, Vinter Olesen O, Aznar M, Andersen FL. Innovations in PET/CT. Q J Nucl Med Mol Imaging 2012;56:268–79.

36. Wahl RL, Jacene H, Kasamon Y, Lodge MA. From RECIST to PERCIST: evolving considerations for PET response criteria in solid tumors. J Nucl Med 2009;50 Suppl 1:122S–50S.

37. Vriens D, Visser EP, de Geus-Oei LF, Oyen WJ. Methodological considerations in quantification of oncological FDG PET studies. Eur J Nucl Med Mol Imaging 2010;37:1408–25.

38. Vanderhoek M, Perlman SB, Jeraj R. Impact of the definition of peak standardized uptake value on quantification of treatment response. J Nucl Med 2012;53:4–11.

39. Velasquez LM, Boellaard R, Kollia G, Hayes W, Hoekstra OS, Lammertsma AA, et al. Repeatability of 18F-FDG PET in a multicenter phase I study of patients with advanced gastrointestinal malignancies. J Nucl Med 2009;50:1646–54.

40. Benz MR, Evilevitch V, Allen-Auerbach MS, Eilber FC, Phelps ME, Czernin J, et al. Treatment monitoring by 18F-FDG PET/CT in patients with sarcomas: interobserver variability of quantitative parameters in treatment-induced changes in histopathologically responding and nonresponding tumors. J Nucl Med 2008;49:1038–46.