



**HAL**  
open science

## **EORTC PET response criteria are more influenced by reconstruction inconsistencies than PERCIST but both benefit from the EARL harmonization program**

Charline Lasnon, Elske Quak, Pierre-Yves Le Roux, Philippe Robin, Michael S Hofman, David Bourhis, Jason Callahan, David S Binns, Cédric Desmots, Pierre-Yves Salaun, et al.

### ► To cite this version:

Charline Lasnon, Elske Quak, Pierre-Yves Le Roux, Philippe Robin, Michael S Hofman, et al.. EORTC PET response criteria are more influenced by reconstruction inconsistencies than PERCIST but both benefit from the EARL harmonization program. *EJNMMI Physics*, 2017, 4 (1), pp.17. 10.1186/s40658-017-0185-4 . inserm-01856103

**HAL Id: inserm-01856103**

**<https://inserm.hal.science/inserm-01856103>**

Submitted on 9 Aug 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ORIGINAL RESEARCH

Open Access



# EORTC PET response criteria are more influenced by reconstruction inconsistencies than PERCIST but both benefit from the EARL harmonization program

Charline Lasnon<sup>1,2</sup>, Elske Quak<sup>1</sup>, Pierre-Yves Le Roux<sup>3</sup>, Philippe Robin<sup>3</sup>, Michael S. Hofman<sup>4</sup>, David Bourhis<sup>3</sup>, Jason Callahan<sup>4</sup>, David S. Binns<sup>4</sup>, Cédric Desmonts<sup>5</sup>, Pierre-Yves Salaun<sup>3</sup>, Rodney J. Hicks<sup>4,6†</sup> and Nicolas Aide<sup>2,5,7,8\*†</sup> 

\* Correspondence: aide-n@chu-caen.fr

†Equal contributors

<sup>2</sup>INSERM U1086 ANTICIPE, BioTICLA, Caen University, Caen, France

<sup>5</sup>Nuclear Medicine Department, University Hospital, Caen, France  
Full list of author information is available at the end of the article

## Abstract

**Background:** This study evaluates the consistency of PET evaluation response criteria in solid tumours (PERCIST) and European Organisation for Research and Treatment of Cancer (EORTC) classification across different reconstruction algorithms and whether aligning standardized uptake values (SUVs) to the European Association of Nuclear Medicine acquisition (EANM)/EARL standards provides more consistent response classification.

**Materials and methods:** Baseline ( $P_{PET1}$ ) and response assessment ( $P_{PET2}$ ) scans in 61 patients with non-small cell lung cancer were acquired in protocols compliant with the EANM guidelines and were reconstructed with point-spread function (PSF) or PSF + time-of-flight (TOF) reconstruction for optimal tumour detection and with a standardized ordered subset expectation maximization (OSEM) reconstruction known to fulfil EANM harmonizing standards. Patients were recruited in three centres. Following reconstruction, EQ.PET, a proprietary software solution was applied to the PSF  $\pm$  TOF data (PSF  $\pm$  TOF.EQ) to harmonize SUVs to the EANM standards. The impact of differing reconstructions on PERCIST and EORTC classification was evaluated using standardized uptake values corrected for lean body mass (SUL).

**Results:** Using  $OSEM_{PET1}/OSEM_{PET2}$  (standard scenario), responders displayed a reduction of  $-57.5\% \pm 23.4$  and  $-63.9\% \pm 22.4$  for  $SUL_{max}$  and  $SUL_{peak}$ , respectively, while progressing tumours had an increase of  $+63.4\% \pm 26.5$  and  $+60.7\% \pm 19.6$  for  $SUL_{max}$  and  $SUL_{peak}$  respectively. The use of PSF  $\pm$  TOF reconstruction impacted the classification of tumour response. For example, taking the  $OSEM_{PET1}/PSF \pm TOF_{PET2}$  scenario reduced the apparent reduction in SUL in responding tumours ( $-39.7\% \pm 31.3$  and  $-55.5\% \pm 26.3$  for  $SUL_{max}$  and  $SUL_{peak}$ , respectively) but increased the apparent increase in SUL in progressing tumours ( $+130.0\% \pm 50.7$  and  $+91.1\% \pm 39.6$  for  $SUL_{max}$  and  $SUL_{peak}$ , respectively).

Consequently, variation in reconstruction methodology (PSF  $\pm$  TOF $_{PET1}/OSEM_{PET2}$  or  $OSEM_{PET1}/PSF \pm TOF_{PET2}$ ) led, respectively, to 11/61 (18.0%) and 10/61 (16.4%)

(Continued on next page)

(Continued from previous page)

PERCIST classification discordances and to 17/61 (28.9%) and 19/61 (31.1%) EORTC classification discordances. An agreement was better for these scenarios with application of the propriety filter, with kappa values of 1.00 and 0.95 compared to 0.75 and 0.77 for PERCIST and kappa values of 0.93 and 0.95 compared to 0.61 and 0.55 for EORTC, respectively.

**Conclusion:** PERCIST classification is less sensitive to reconstruction algorithm-dependent variability than EORTC classification but harmonizing SULs within the EARL program is equally effective with either.

**Keywords:** PET,  $^{18}\text{F}$ -FDG, Therapy response, PERCIST, EORTC, Harmonization

## Background

$^{18}\text{F}$ -FDG PET is increasingly being used for response evaluation in cancer patients, in clinical routine or in clinical trials [1–6]. Two main schemas based on the degree of standardized uptake value (SUV) change following treatment are currently used: the European Organisation for Research and Treatment of Cancer (EORTC) criteria [7] and PET evaluation response criteria in solid tumours (PERCIST) [8]. However, many sources of error in SUV measurement exist [9–11]. In particular, technological improvements can lead to significant device-dependent and reconstruction-dependent variations in quantitative values [12–14]. This could lead to classification errors by exceeding thresholds used for discriminating between responding and non-responding tumours unless acquisition and processing of pre- and post-treatment scans are acquired on the same scanner and processed identically.

The European Association Research Ltd (EARL) accreditation program [15] is an SUV harmonization strategy aiming at minimizing the variability in SUV measurements by harmonizing patient preparation and scan acquisition and processing [16]. While many sources of error in SUV measurements are overcome by complying with the EANM guidelines for PET tumour imaging [17–19], reconstruction-dependent variations require either the use of an additional filtering step [20] or the generation of two sets of images: one to provide optimal diagnostic quality and another to meet quantitative harmonization standards [21]. Previous research from the collaborators in this study have shown that  $\text{SUV}_{\text{max}}$  is more sensitive to reconstruction inconsistency than  $\text{SUV}_{\text{peak}}$  [20] and that reconstruction inconsistencies may affect PERCIST classification [22]. Consequently, one could expect a more significant impact of these inconsistencies on EORTC classification, which is based on  $\text{SUV}_{\text{max}}$  variation, than on PERCIST, which is based on  $\text{SUV}_{\text{peak}}$ .

The aim of this study was to evaluate the impact of SUV reconstruction dependency on PERCIST and EORTC classification and the ability of the EARL program to minimize variability in response assessment. To assess this, we reconstructed the same PET raw data with an OSEM algorithm known to meet EANM requirements and also with PSF with or without TOF reconstruction ( $\text{PSF} \pm \text{TOF}$ ). Post-reconstruction filtering was then applied to the  $\text{PSF} \pm \text{TOF}$  reconstruction with EQ.PET (Siemens Medical Solutions), a proprietary software solution allowing visualization of optimized images while simultaneously obtaining harmonized SUV values [20, 23].

## Methods

### Patients

Sixty-one patients with non-small cell lung cancer (NSCLC) who were scanned for monitoring efficacy of chemotherapy, molecularly targeted therapies or radiotherapy were included. The cohort was comprised of 51 patients prospectively included in a multicentre study involving three PET centres and 10 patients included in a single-centre prospective study. Informed consent was waived for this type of study by the local ethics committee (Ref A12-D24-VOL13, *Comité de protection des personnes Nord-Ouest III*) since the scans were performed for clinical indications, and the study procedures were performed independently without influencing clinical reporting.

Patient's sex ratio (male/female) was 2.4:1; mean  $\pm$  SD age was  $62.7 \pm 9.4$  years. The interval between the pre- and post-treatment PET scans was  $103 \pm 53$  days. Fifty-eight (95.1%) patients underwent chemotherapy, 1 (1.6%) patient had radiotherapy and 2 (3.3%) patients were administered targeted therapies (TKI and immunotherapy).

### PET systems

Data from the following three PET systems were used for this study: a Biograph 6 TrueV with PSF reconstruction, a mCT with PSF + TOE, and a Biograph 64 TrueV with PSF reconstruction (Siemens Medical Solutions). Both the Biograph systems were equipped with an extended axial field-of-view.

### Patient preparation, PET acquisition and reconstruction parameters

All patients were requested to fast for 6 h prior to the  $^{18}\text{F}$ -FDG injection. Patient height, weight and blood glucose levels were recorded. Patients were injected intravenously with  $^{18}\text{F}$ -FDG, followed by a 60 min rest in a warm room.

A daily calibration of each PET system was performed with a  $^{68}\text{Ge}$  source according to the manufacturer's protocol. A quarterly cross-calibration of each PET system was performed according to the EANM guidelines, as described elsewhere [17, 18], and clocks from workstations were synchronized weekly.

Patients were scanned from the skull vertex or base to the mid-thighs. All raw PET data were reconstructed with the local PSF  $\pm$  TOF settings for optimal lesion detection and an OSEM-3D reconstruction algorithm fulfilling the EANM guidelines regarding recovery coefficients (Table 1). Scatter and attenuation corrections were applied on all PET acquisitions.

### EQ.PET methodology

For each PET system, the EQ.PET filter was calculated on the phantom data of each PSF  $\pm$  TOF reconstruction as described in details elsewhere [21]. Briefly, the recovery coefficients (RCs; defined as the ratio between the measured and true activity concentration for each sphere) of a National Electrical Manufacturers Association NU2 phantom scanned as per EANM guidelines were aligned to the EANM reference RCs by applying a Gaussian filter.

### PERCIST and EORTC evaluation

All PET exams were analyzed on Syngo.via software equipped with EQ.PET (Siemens Medical Solutions). For interpretation purposes, both the reconstruction for optimal



lesion detection (PSF  $\pm$  TOF) and the OSEM reconstruction were displayed on the screen together with the EQ.PET-filtered harmonized SUV results for the tumour region(s) of interest. The EQ.PET-filtered images were not displayed on the screen.

For PERCIST criteria [8], the measurable target lesion is the single most intense tumour site on pre- and post-treatment scans, which means that the target lesion is not necessarily the same pre- and post-treatment. As per EORTC PET response criteria, the volumes of interest (VOI) should involve the same tumour lesion on pre- and post-treatment scan.

In practice, the target lesion on baseline scan was chosen as the most intense lesion and located by scaling the 3D MIP view both on the OSEM and PSF  $\pm$  TOF reconstructions. VOIs were drawn on one reconstruction and automatically propagated to the second set of reconstruction (propagation from OSEM to PSF  $\pm$  TOF and vice versa). Within these volumes of interest, lean body mass SUV<sub>peak</sub> (SUL<sub>peak</sub>) and SUL<sub>max</sub> were measured.

The same VOI methodology was used on the post-treatment scan, where the target lesion was chosen as the most intense lesion for PERCIST, while the same target lesion for baseline and post-treatment scans was used for EORTC classification.

Based on the SUL<sub>peak</sub> and SUL<sub>max</sub> variation between the pre- and post-treatment scans, patients were classified according to PERCIST and EORTC as follows:

- *Complete metabolic response* (CMR): complete resolution of <sup>18</sup>F-FDG uptake in the tumour volume, with tumour SUL lower than liver SUL and background blood pool, and disappearance of all lesions if multiple.
- *Partial metabolic response* (PMR): at least 30% (PERCIST) or 25% (EORTC) reduction in tumour uptake.
- *Stable metabolic disease* (SMD): less than 30% (PERCIST) or 25% (EORTC) increase, or less than 30 or 25% (EORTC) decrease in tumour <sup>18</sup>F-FDG SUL<sub>peak</sub> and no new lesions.
- *Progressive metabolic disease* (PMD): greater than 30% (PERCIST) or 25% (EORTC) increase in <sup>18</sup>F-FDG tumour SUL<sub>peak</sub> within the tumour or appearance of new lesions.

### Statistical analysis

Quantitative data from clinical PET/CT examinations are presented as mean (standard deviation  $\pm$  SD). The relationship between PSF  $\pm$  TOF, PSF  $\pm$  TOF.EQ and OSEM quantitative values were assessed with Bland-Altman plots. Levels of agreement between the different types of reconstruction were evaluated using the kappa statistic. The use of OSEM reconstruction for both pre- and post-therapeutic PET examinations (OSEM<sub>PET1</sub>/OSEM<sub>PET2</sub>) was used as the “current standard” to classify the therapeutic response of each lesion and compared to other scenarios. Kappa values were reported using the benchmarks of Landis and Koch [24].

Graphs and analyses were carried out using Prism GraphPad and the Vassar University website for statistical computation (<http://vassarstats.net>).

## Results

### Ability of the EQ.PET methodology to harmonize SUL assessments

The mean percentage difference (% difference) between PSF  $\pm$  TOF and OSEM reconstructions were 37.19% (95%CI 9.99–64.40) and 19.94% (95%CI 3.12–36.80) for SUL<sub>max</sub>

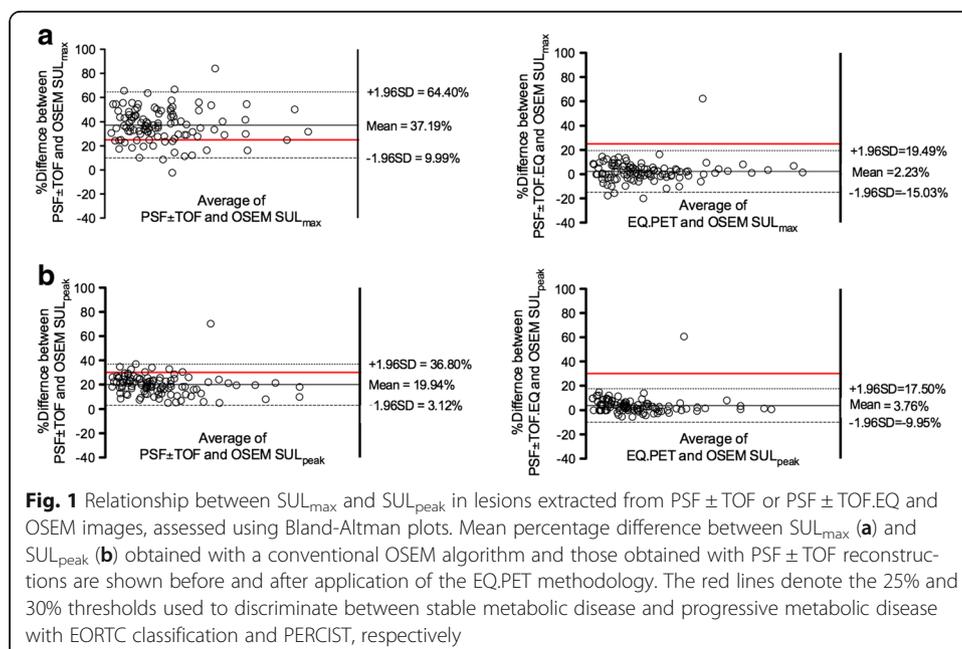
and  $SUL_{peak}$ , respectively. After application of the EQ.PET filter, this was reduced to 2.23% (95%CI -15.03-19.49) and 3.76% (95%CI -9.95-17.50) for  $SUL_{max}$  and  $SUL_{peak}$ , respectively (Fig. 1). Noticeably, in both cases, confidence intervals were slightly narrower for  $SUL_{peak}$  values.

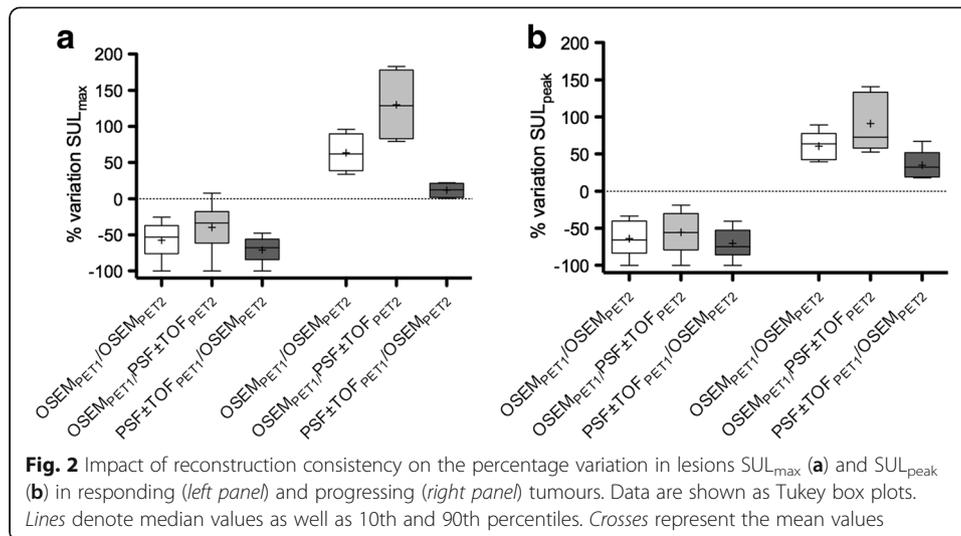
**Impact of reconstruction-dependent variation on SUL changes between baseline and post-treatment scans**

The same target lesion for baseline and post-treatment scans was used for EORTC classification except for two patients. The first patient displayed a large tumoural and nodal complex for which the EQ.PET software was unable to differentiate nodes from a tumour on post-treatment scan. The second patient had a complete disappearance of the initial target lesion in a patient with multiple tumour lesions, requiring to use the hottest remaining lesion on post-treatment scan.

The variations in  $SUL_{max}$  and  $SUL_{peak}$  between the pre- and post-treatment scans are shown in Fig. 2. For the  $OSEM_{PET1}/OSEM_{PET2}$  scenario, which was taken as the reference standard, the change in  $SUL_{max}$  was  $-57.5\% \pm 23.4$  and  $+63.4\% \pm 26.5$  in the groups of tumours showing a decrease and an increase in  $^{18}F$ -FDG uptake, respectively. For  $SUL_{peak}$ , it was  $-63.9\% \pm 22.4$  and  $+60.7\% \pm 19.6$ , respectively.

The use of PSF reconstruction impacted SULs, depending whether this reconstruction was used for the pre- or post-treatment scans. For example,  $OSEM_{PET1}/PSF \pm TOF_{PET2}$  scenario reduced the apparent reduction in SUL in responding tumours ( $-39.7\% \pm 31.3$  and  $-55.5\% \pm 26.3$  for  $SUL_{max}$  and  $SUL_{peak}$ , respectively) but increased the apparent increase in SUL in progressing tumours ( $+130.0\% \pm 50.7$  and  $+91.1\% \pm 39.6$  for  $SUL_{max}$  and  $SUL_{peak}$ , respectively) as compared to the  $OSEM_{PET1}/OSEM_{PET2}$  scenario described above. Accordingly, inconsistent reconstructions induced discordant response classifications amongst the different scenarios, as described in the section below.





#### Impact of reconstruction-dependent variation of SUL on PERCIST and EORTC evaluation

By using OSEM for the pre- and post-treatment scans, PET classified 7 patients as CMR, 18 as PMR, 14 as SMD and 22 as PMD according to EORTC classification (Fig. 3) and 7 patients as CMR, 14 as PMR, 17 as SMD and 23 as PMD according to PERCIST (Fig. 4). According to EORTC evaluation, CMR occurred in five patients with a decrease in  $SUL_{max}$  to a level below the liver and blood pool background and in two patients to complete disappearance of the target lesions. PMD occurred in four patients with an increase in tumour  $SUL_{max}$  greater than 25% and in 18 patients with new lesions on the post-treatment scan. According to PERCIST classification, CMR occurred in five patients with a decrease in  $SUL_{peak}$  to a level below the liver and blood pool background and in two patients to complete disappearance of the target lesions. PMD occurred in five patients with an increase in tumour  $SUL_{peak}$  greater than 30% and in 18 patients with new lesions on the post-treatment scan.

The agreement level between EORTC and PERCIST therapeutic evaluations was almost perfect with a kappa value equal of 0.84 (0.73–0.95). Eight discordances (13%) occurred: one patient classified as CMR with EORTC and PMR with PERCIST, one patient classified as PMR with EORTC and CMR with PERCIST, four patients classified as PMR with EORTC and SMD with PERCIST and one patient classified as SMD with EORTC and PD with PERCIST.

Agreement levels between the  $OSEM_{PET1}/OSEM_{PET2}$  scenario and other scenarios involving reconstruction inconsistency were found to be almost perfect with narrow confidence intervals for the scenarios using EQ.PET-filtered data either pre- or post-treatment and the reconstruction-consistent scenario for both EORTC and PERCIST classifications (Table 2). For EORTC and PERCIST evaluations, agreement levels were moderate to substantial for the scenario  $OSEM_{PET1}/PSF \pm TOF_{PET2}$  and  $PSF \pm TOF_{PET1}/OSEM_{PET2}$ , with wide confidence intervals. Noticeably, kappa values were lower for EORTC classification than for PERCIST, especially for the  $OSEM_{PET1}/PSF \pm TOF_{PET2}$  scenario (0.55 quoted as moderate vs 0.77 quoted as substantial).

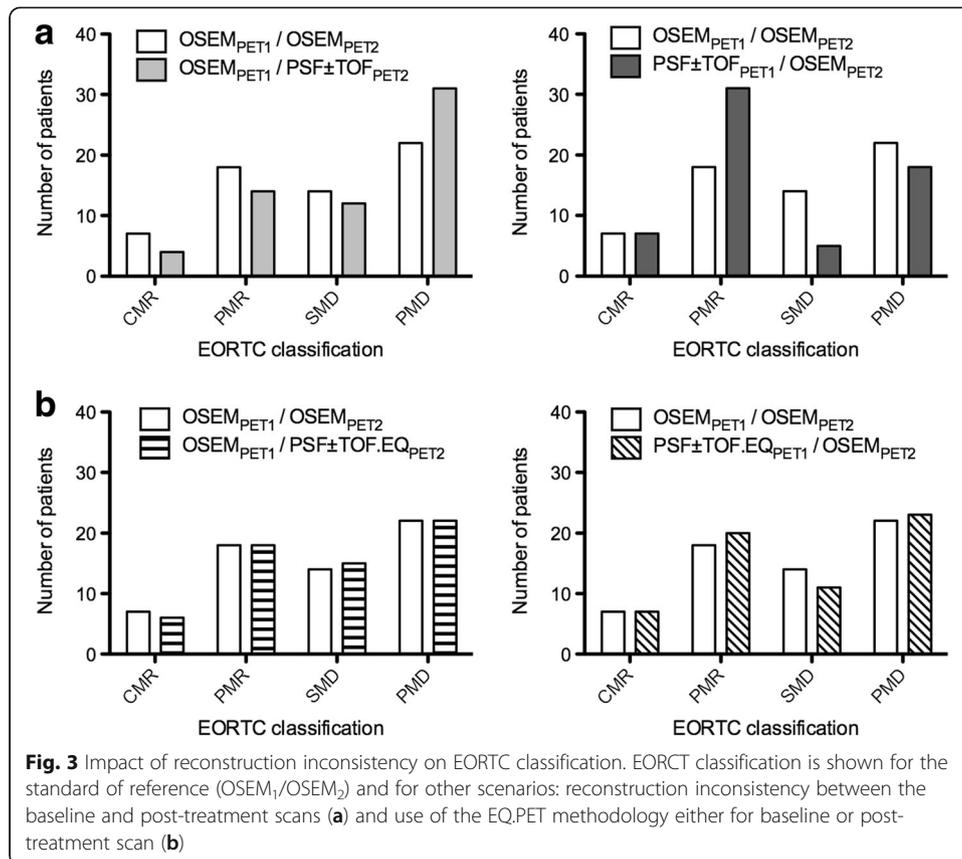
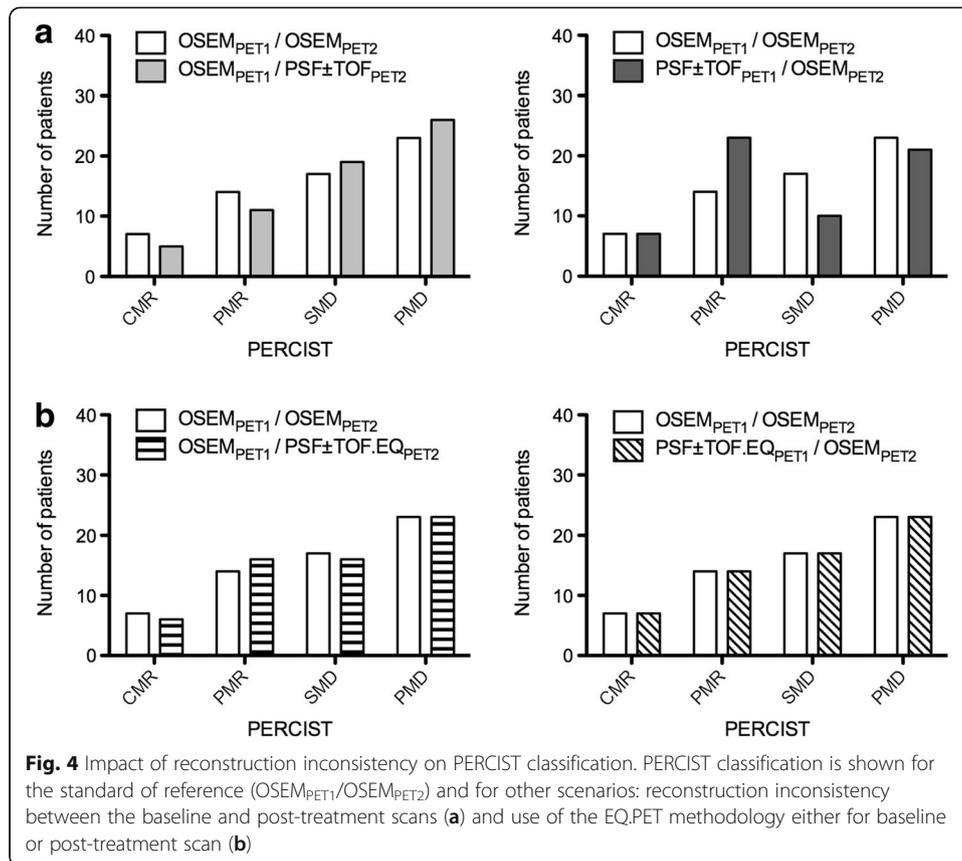


Table 3 and Figs. 3 and 4 show the number of discordances in the EORTC and PERCIST classifications that occurred for the different scenarios tested. The EORTC classification displayed more discordances than what PERCIST did for all scenarios. For example, the scenario  $OSEM_{PET1}/PSF \pm TOF_{PET2}$  led to three patients being classified as PMR instead of CMR, seven as SMD instead of PMR, and nine as PMD instead of SMD with the EORTC classification whereas these same changes occurred, respectively, in two, five and three cases with the PERCIST classification. Figure 5 illustrates a patient classified as SMD according to the  $OSEM_{PET1}/OSEM_{PET2}$  standard of reference with EORTC classification and PERCIST, while  $PSF + TOF_{PET1}/OSEM_{PET2}$  led to PMR with both classifications and  $OSEM_{PET1}/PSF + TOF_{PET2}$  led to PD with EORTC classification.

Consistent reconstruction (i.e. the  $PSF \pm TOF_{PET1}/PSF \pm TOF_{PET2}$  and  $PSF \pm TOF.EQ_{PET1}/PSF \pm TOF.EQ_{PET2}$  scenarios) did not give a perfect agreement compared to the  $OSEM_{PET1}/OSEM_{PET2}$  standard of reference (Additional file 1: Figure S1). This was more pronounced for the EORTC classification in the  $PSF \pm TOF_{PET1}/PSF \pm TOF_{PET2}$  scenario where six discordances occurred (Table 3), leading to a kappa value of 0.86 (Table 2).

### Discussion

In the framework of therapy monitoring with PET, pre- and post-treatment scans should ideally involve identical scan acquisition and image processing. However, this is often impractical in busy PET centres, especially those running several scanners. This can also be challenged by a scanner upgrade during the conduct of a trial or when a



patient relocates. Previous studies aimed at validating the EARL harmonization strategy in the clinical setting have shown that  $SUV_{max}$  is more sensitive to reconstruction inconsistency than  $SUV_{peak}$  or their lean body mass equivalents,  $SUL_{max}$  and  $SUL_{peak}$ . Consequently, one could expect a more significant impact of reconstruction inconsistencies on EORTC classification than on PERCIST.

In the present study, we evaluated the impact of inconsistent reconstruction on both EORTC and PERCIST response classifications, demonstrating variation in up to 31% of cases for EORTC classification vs up to 18% for PERCIST classification. Further, we showed that applying the EARL harmonization strategy provided more consistent response classification with kappa values greater than 0.93 for all the scenarios involving harmonized SULs, compared to the  $OSEM_{PET1}/OSEM_{PET2}$  scenario used as a standard of

**Table 2** Agreement levels between the  $OSEM_1/OSEM_2$  scenario and other scenarios involving reconstruction inconsistency for EORTC and PERCIST therapeutic evaluations

	Kappa (95%CI)	
	EORTC	PERCIST
$OSEM_{PET1}/OSEM_{PET2}$ vs $OSEM_{PET1}/PSF \pm TOF_{PET2}$	0.55 (0.39–0.72)	0.77 (0.63–0.90)
$OSEM_{PET1}/OSEM_{PET2}$ vs $PSF \pm TOF_{PET1}/OSEM_{PET2}$	0.61 (0.45–0.77)	0.75 (0.62–0.88)
$OSEM_{PET1}/OSEM_{PET2}$ vs $OSEM_{PET1}/PSF \pm TOF.EQ_{PET2}$	0.95 (0.89–1.00)	0.95 (0.89–1.00)
$OSEM_{PET1}/OSEM_{PET2}$ vs $PSF \pm TOF.EQ_{PET1}/OSEM_{PET2}$	0.93 (0.86–1.00)	1.00 (1.00–1.00)
$OSEM_{PET1}/OSEM_{PET2}$ vs $PSF \pm TOF_{PET1}/PSF \pm TOF_{PET2}$	0.86 (0.75–0.97)	0.93 (0.85–1.00)
$OSEM_{PET1}/OSEM_{PET2}$ vs $PSF \pm TOF.EQ_{PET1}/PSF \pm TOF.EQ_{PET2}$	0.93 (0.85–1.00)	0.93 (0.85–1.00)

**Table 3** Number of discordances between the OSEM<sub>1</sub>/OSEM<sub>2</sub> scenario and other scenarios involving reconstruction inconsistency for EORTC and PERCIST therapeutic evaluations

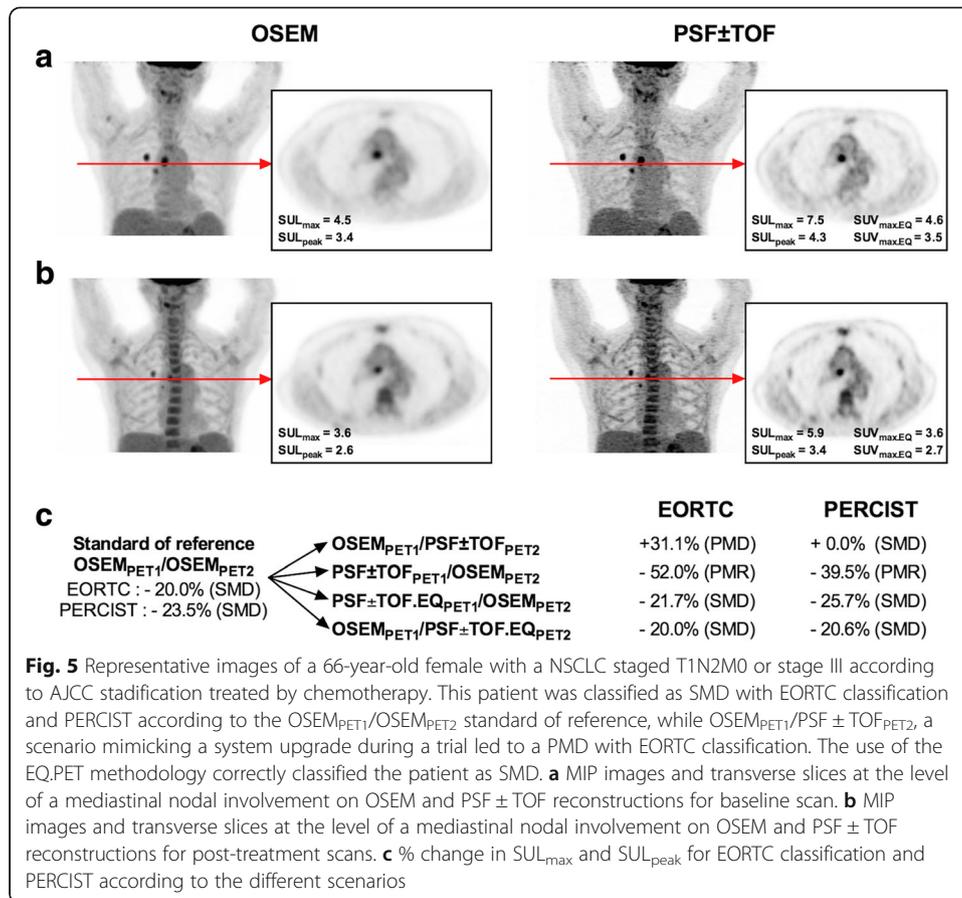
	Numbers of discordances <i>n</i> (%)	
	EORTC	PERCIST
OSEM <sub>PET1</sub> /OSEM <sub>PET2</sub> vs OSEM <sub>PET1</sub> / PSF ± TOF <sub>PET2</sub>	19 (31)	10 (16)
OSEM <sub>PET1</sub> /OSEM <sub>PET2</sub> vs PSF ± TOF <sub>PET1</sub> /OSEM <sub>PET2</sub>	17 (28)	11 (18)
OSEM <sub>PET1</sub> /OSEM <sub>PET2</sub> vs OSEM <sub>PET1</sub> / PSF ± TOF.EQ <sub>PET2</sub>	2 (3)	2 (3)
OSEM <sub>PET1</sub> /OSEM <sub>PET2</sub> vs PSF ± TOF.EQ <sub>PET1</sub> /OSEM <sub>PET2</sub>	3 (5)	0 (0)
OSEM <sub>PET1</sub> /OSEM <sub>PET2</sub> vs PSF ± TOF <sub>PET1</sub> / PSF ± TOF <sub>PET2</sub>	6 (10)	3 (5)
OSEM <sub>PET1</sub> /OSEM <sub>PET2</sub> vs PSF ± TOF.EQ <sub>PET1</sub> / PSF ± TOF.EQ <sub>PET2</sub>	3 (5)	3 (5)

reference. In line with its greater sensitivity to reconstruction inconsistencies, the EORTC classification benefited more from the EARL harmonization strategy, with kappa values increasing from 0.55 to 0.95 for the worst case scenario (OSEM<sub>PET1</sub>/PSF ± TOF<sub>PET2</sub>), compared with an improvement from 0.77 to 0.95 for PERCIST (Table 2).

This has practical advantages when there is variation of acquisition/reconstruction settings. This situation seems relatively common even in centres running the same PET system, as recently described by Sunderland and colleagues [25] in a survey involving 237 PET/CT systems in 170 international imaging centres with technology advancements spanning more than a decade, reporting that site-specific reconstruction parameters increased the quantitative variability of similar scanners, post-reconstruction smoothing filters being the most influential parameter. Harmonization has also practical advantages when the use of the same scanner for both scans is impractical, for instance in centres running two or more PET systems, as illustrated by the study by Skougaard et al. [26], in which 12 of 81 (14%) patients undergoing pre- and post-treatment PET in the same department were excluded for analysis because they were scanned on two different generation PET systems.

Taking, for example, the scenario of a system upgrade during a trial, the use of OSEM for the pre-treatment scan while using PSF ± TOF for the post-treatment scan led to discordant response assessments in 19/61 (31%) for EORTC classification and 10/61 (16%) for PERCIST (Table 3). Using a harmonization strategy (hereby aligning quantitative values to the EARL/EANM harmonizing standards with a proprietary filter, the EQ.PET methodology) either for the pre- or post-treatment scans gave almost perfect agreement levels in comparison with the OSEM<sub>PET1</sub>/OSEM<sub>PET2</sub> reference standard, with narrow confidence intervals. We observed only two discordances for the OSEM<sub>PET1</sub>/PSF ± TOF.EQ<sub>PET2</sub> vs OSEM<sub>PET1</sub>/OSEM<sub>PET2</sub> scenario for both the EORTC and PERCIST classifications and three discordances which occurred for the PSF ± TOF.EQ<sub>PET1</sub>/OSEM<sub>PET2</sub> vs OSEM<sub>PET1</sub>/OSEM<sub>PET2</sub> scenario for the EORTC classification. No discordance occurred for the PSF ± TOF.EQ<sub>PET1</sub>/OSEM<sub>PET2</sub> vs OSEM<sub>PET1</sub>/OSEM<sub>PET2</sub> scenario for PERCIST classification. The three discordances that occurred only with EORTC classification for the PSF ± TOF.EQ<sub>PET1</sub>/OSEM<sub>PET2</sub> were due to SUL<sub>max</sub> variations between the pre and post-treatment scans very close to the cut-off value of +25 or -25% with the standard scenario OSEM<sub>PET1</sub>/OSEM<sub>PET2</sub> resulting in changes from SMD to either PMR or PMD and vice versa for other scenarios.

It is noteworthy that consistent reconstruction (i.e. the PSF ± TOF<sub>PET1</sub>/PSF ± TOF<sub>PET2</sub> and PSF ± TOF.EQ<sub>PET1</sub>/ PSF ± TOF.EQ<sub>PET2</sub> scenarios) did not give perfect agreement



**Fig. 5** Representative images of a 66-year-old female with a NSCLC staged T1N2M0 or stage III according to AJCC stadification treated by chemotherapy. This patient was classified as SMD with EORTC classification and PERCIST according to the OSEM<sub>PET1</sub>/OSEM<sub>PET2</sub> standard of reference, while OSEM<sub>PET1</sub>/PSF ± TOF<sub>PET2</sub>, a scenario mimicking a system upgrade during a trial led to a PMD with EORTC classification. The use of the EQ.PET methodology correctly classified the patient as SMD. **a** MIP images and transverse slices at the level of a mediastinal nodal involvement on OSEM and PSF ± TOF reconstructions for baseline scan. **b** MIP images and transverse slices at the level of a mediastinal nodal involvement on OSEM and PSF ± TOF reconstructions for post-treatment scans. **c** % change in SUL<sub>max</sub> and SUL<sub>peak</sub> for EORTC classification and PERCIST according to the different scenarios

compared to the OSEM<sub>PET1</sub>/OSEM<sub>PET2</sub> standard of reference. These discordances were due to PSF reconstruction increasing SUV metrics in the tumours while not impacting the background (blood pool and liver) [27, 28], leading to CMR being changed to PMR. Also, both the EORTC and PERCIST classifications were affected by %change in SUL close to +30%/+25% or -30%/-25% for the OSEM<sub>PET1</sub>/OSEM<sub>PET2</sub> scenario resulting in changes from SMD to either PMR or PMD and vice versa for other scenarios.

A limitation of this study is that we used EQ.PET, a software solution developed for and applied only to scanners and reconstruction algorithms of the company that developed this product. EQ.PET has not been validated for equipment from other manufacturers but has been shown to be as effective as the alternative approach of obtaining a second reconstruction dataset, as recommended by the EARL accreditation program for quantitation [29, 30]. The ability of this algorithm to correct for scans performed on different scanners and then processed with different reconstruction methods was not tested.

### Conclusions

PERCIST classification is less sensitive to reconstruction algorithm-dependent variability than EORTC classification. The EORTC and PERCIST classifications would benefit from harmonization strategies such as the EARL accreditation program in multicentre studies or in sites equipped with multiple PET systems.

## Additional file

**Additional file 1: Figure S1.** Impact of reconstruction consistency on EORTC classification and PERCIST. EORTC classification and PERCIST are shown for the standard of reference (OSEM<sub>1</sub>/OSEM<sub>2</sub>) and for other scenarios involving reconstruction consistency between the baseline and post-treatment scans using either PSF ± TOF (a) or the EQPET methodology (PSF ± TOF.EQ; b). (TIFF 1787 kb)

### Authors' contributions

Study design: CL, EQ, PYL, RH; Study coordination: NA, CL; Data gathering: CL, EQ, PYL, PR, MH, DB, JC, DS, CD, PYS, RL, NA; Data analysis: CL, EQ, PYL, PR, NA; Manuscript writing: CL, NA, EQ, PYL, MH, RH. All authors read and approved the final manuscript.

### Competing interests

The authors declare that they have no competing interests.

### Ethics approval and consent to participate

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent was waived for this type of study by the local ethics committee (Ref A12-D24-VOL13, Comité de protection des personnes Nord-Ouest III), since the PET scans were performed for clinical indications and the trial procedures were performed independent of usual clinical reporting.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

### Author details

<sup>1</sup>Nuclear Medicine Department, François Baclesse Cancer Centre, Caen, France. <sup>2</sup>INSERM U1086 ANTICIPE, BioTICLA, Caen University, Caen, France. <sup>3</sup>Nuclear Medicine Department and EA 3878 IFR 148, University Hospital, Brest, France. <sup>4</sup>Cancer Imaging, Peter Mac Callum Cancer Institute, Parkville, Australia. <sup>5</sup>Nuclear Medicine Department, University Hospital, Caen, France. <sup>6</sup>The Sir Peter MacCallum Department of Oncology, the University of Melbourne, Melbourne, Australia. <sup>7</sup>Normandy University, Caen, France. <sup>8</sup>Nuclear Medicine Department, Caen University Hospital, Avenue Côte de Nacre, 14000 Caen, France.

Received: 7 April 2017 Accepted: 19 May 2017

Published online: 30 May 2017

### References

- Bazan JG, Duan F, Snyder BS, Horng D, Graves EE, Siegel BA, et al. Metabolic tumor volume predicts overall survival and local control in patients with stage III non-small cell lung cancer treated in ACRIN 6668/RTOG 0235. *Eur J Nucl Med Mol Imaging*. 2017;44:17–24. doi:10.1007/s00259-016-3520-4.
- Ho KC, Fang YD, Chung HW, Liu YC, Chang JW, Hou MM, et al. TLG-S criteria are superior to both EORTC and PERCIST for predicting outcomes in patients with metastatic lung adenocarcinoma treated with erlotinib. *Eur J Nucl Med Mol Imaging*. 2016;43:2155–65. doi:10.1007/s00259-016-3433-2.
- Hyun OJ, Lubner BS, Leal JP, Wang H, Bolejack V, Schuetz SM, et al. Response to early treatment evaluated with 18F-FDG PET and PERCIST 1.0 predicts survival in patients with Ewing sarcoma family of tumors treated with a monoclonal antibody to the insulin-like growth factor 1 receptor. *J Nucl Med*. 2016;57:735–40.
- Michl M, Lehner S, Paprottka PM, Ilhan H, Bartenstein P, Heinemann V, et al. Use of PERCIST for prediction of progression-free and overall survival after radioembolization for liver metastases from pancreatic cancer. *J Nucl Med*. 2016;57:355–60. doi:10.2967/jnumed.115.165613.
- Pinker K, Riedl CC, Ong L, Jochelson M, Ulaner GA, McArthur H, et al. The impact that number of analyzed metastatic breast cancer lesions has on response assessment by 18F-FDG PET/CT using PERCIST. *J Nucl Med*. 2016; 57:1102–4. doi:10.2967/jnumed.115.166629.
- Shang J, Ling X, Zhang L, Tang Y, Xiao Z, Cheng Y, et al. Comparison of RECIST, EORTC criteria and PERCIST for evaluation of early response to chemotherapy in patients with non-small-cell lung cancer. *Eur J Nucl Med Mol Imaging*. 2016;43:1945–53. doi:10.1007/s00259-016-3420-7.
- 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.
- 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–50. doi:10.2967/jnumed.108.057307.
- Boellaard R. Standards for PET image acquisition and quantitative data analysis. *J Nucl Med*. 2009;50(Suppl 1):11S–20. doi:10.2967/jnumed.108.057182.
- Boellaard R. Methodological aspects of multicenter studies with quantitative PET. *Methods Mol Biol*. 2011;727:335–49. doi:10.1007/978-1-61779-062-1\_18.
- Boellaard R. Mutatis mutandis: harmonize the standard! *J Nucl Med*. 2012;53:1–3. doi:10.2967/jnumed.111.094763.

12. Bellevre D, Blanc Fournier C, Switsers O, Dugue AE, Levy C, Allouache D, et al. Staging the axilla in breast cancer patients with (1)(8)F-FDG PET: how small are the metastases that we can detect with new generation clinical PET systems? *Eur J Nucl Med Mol Imaging*. 2014;41:1103–12. doi:10.1007/s00259-014-2689-7.
13. Koopman D, Groot Koerkamp M, Jager PL, Arkies H, Knolles S, Slump CH, et al. Digital PET compliance to EARL accreditation specifications. *EJNMMI physics*. 2017;4:9. doi:10.1186/s40658-017-0176-5.
14. Teoh EJ, McGowan DR, Macpherson RE, Bradley KM, Gleeson FV. Phantom and clinical evaluation of the Bayesian penalized likelihood reconstruction algorithm Q.Clear on an LYSO PET/CT system. *J Nucl Med*. 2015. doi:10.2967/jnumed.115.159301.
15. European Association of Nuclear Medicine. EARL FDG-PET/CT accreditation. 2015. [http://earl.eanm.org/cms/website.php?id=en/projects/fdg\\_pet\\_ct\\_accreditation.htm](http://earl.eanm.org/cms/website.php?id=en/projects/fdg_pet_ct_accreditation.htm).
16. Makris NE, Huisman MC, Kinahan PE, Lammertsma AA, Boellaard R. Evaluation of strategies towards harmonization of FDG PET/CT studies in multicentre trials: comparison of scanner validation phantoms and data analysis procedures. *Eur J Nucl Med Mol Imaging*. 2013;40:1507–15. doi:10.1007/s00259-013-2465-0.
17. Boellaard R, Delgado-Bolton R, Oyen WJ, Giammarile F, Tatsch K, Eschner W, et al. FDG PET/CT: EANM procedure guidelines for tumour imaging: version 2.0. *Eur J Nucl Med Mol Imaging*. 2015;42:328–54. doi:10.1007/s00259-014-2961-x.
18. 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. doi:10.1007/s00259-009-1297-4.
19. 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.
20. Quak E, Le Roux PY, Hofman MS, Robin P, Bourhis D, Callahan J, et al. Harmonizing FDG PET quantification while maintaining optimal lesion detection: prospective multicentre validation in 517 oncology patients. *Eur J Nucl Med Mol Imaging*. 2015;42:2072–82. doi:10.1007/s00259-015-3128-0.
21. Lasnon C, Desmots C, Quak E, Gervais R, Do P, Dubos-Arvis C, et al. Harmonizing SUVs in multicentre trials when using different generation PET systems: prospective validation in non-small cell lung cancer patients. *Eur J Nucl Med Mol Imaging*. 2013;40:985–96. doi:10.1007/s00259-013-2391-1.
22. Lasnon C, Le Roux PY, Quak E, Robin P, Hofman MS, Bourhis D, et al. EORTC PET response criteria are more influenced by reconstruction inconsistencies than PERCIST, but both equally benefit from the EARL harmonization program. *J Nucl Med*. 2017. doi:10.2967/jnumed.115.171983.
23. Kelly MD, Declerck JM. SUVref: reducing reconstruction-dependent variation in PET SUV. *EJNMMI Res*. 2011;1:16. doi:10.1186/2191-219X-1-16.
24. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33:159–74.
25. Sunderland JJ, Christian PE. Quantitative PET/CT scanner performance characterization based upon the society of nuclear medicine and molecular imaging clinical trials network oncology clinical simulator phantom. *J Nucl Med*. 2015;56:145–52. doi:10.2967/jnumed.114.148056.
26. Skougard K, Nielsen D, Jensen BV, Hendel HW. Comparison of EORTC criteria and PERCIST for PET/CT response evaluation of patients with metastatic colorectal cancer treated with irinotecan and cetuximab. *J Nucl Med*. 2013; 54:1026–31. doi:10.2967/jnumed.112.111757.
27. Kuhnert G, Boellaard R, Sterzer S, Kahraman D, Scheffler M, Wolf J, et al. Impact of PET/CT image reconstruction methods and liver uptake normalization strategies on quantitative image analysis. *Eur J Nucl Med Mol Imaging*. 2015. doi:10.1007/s00259-015-3165-8.
28. Quak E, Hovhannisyan N, Lasnon C, Fruchart C, Vilque JP, Musafiri D, et al. The importance of harmonizing interim positron emission tomography in non-Hodgkin lymphoma: focus on the Deauville criteria. *Haematologica*. 2014; 99:e84–5. doi:10.3324/haematol.2014.104125.
29. Lasnon C, Salomon T, Desmots C, Do P, Oulhouir Y, Madelaine J, et al. Generating harmonized SUV within the EANM EARL accreditation program: software approach versus EARL-compliant reconstruction. *Ann Nucl Med*. 2016. doi:10.1007/s12149-016-1135-2.
30. Aide N, Lasnon C, Veit Haibach P, Sera T, Sattler B, Boellaard R. EANM/EARL harmonization strategies in PET quantification: from daily practice to multicentre oncological studies. *Eur J Nucl Med Mol Imaging*. 2017. doi:10.1007/s00259-017-3740-2.

Submit your manuscript to a SpringerOpen® journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

---

Submit your next manuscript at ► [springeropen.com](http://springeropen.com)

---