Feasibility of state of the art PET/CT systems performance harmonisation

Andres Kaalep 1, 2, 3, Terez Sera 2, 3, Sjoerd Rijnsdorp 4, Maqsood Yaqub 5, Anne Talsma 6, Martin A. Lodge 7, Ronald Boellaard 3, 5, 8

Abstract

Purpose The objective of this study was to explore the feasibility of harmonising performance for PET/CT systems equipped with time-of-flight (ToF) and resolution modelling/point spread function (PSF) technologies. A second aim was producing a working prototype of new harmonising criteria with higher contrast recoveries than current EARL standards using various SUV metrics.

Methods Four PET/CT systems with both ToF and PSF capabilities from three major vendors were used to acquire and reconstruct images of the NEMA NU2–2007 body phantom filled conforming EANM EARL guidelines. A total of 15 reconstruction parameter sets of varying pixel size, post filtering and reconstruction type, with three different acquisition durations were used to compare the quantitative performance of the systems. A target range for recovery curves was established such that it would accommodate the highest matching recoveries from all investigated systems. These updated criteria were validated on 18 additional scanners from 16 sites in order to demonstrate the scanners’ ability to meet the new target range.

Results Each of the four systems was found to be capable of producing harmonising reconstructions with similar recovery curves. The five reconstruction parameter sets producing harmonising results significantly increased SUVmean (25%) and SUVmax (26%) contrast recoveries compared with current EARL specifications. Additional prospective validation performed on 18 scanners from 16 EARL accredited sites demonstrated the feasibility of updated harmonising specifications. SUVpeak was found to significantly reduce the variability in quantitative results while producing lower recoveries in smaller (≤17 mm diameter) sphere sizes.

Conclusions Harmonising PET/CT systems with ToF and PSF technologies from different vendors was found to be feasible. The harmonisation of such systems would require an update to the current multicentre accreditation program EARL in order to accommodate higher recoveries. SUVpeak should be further investigated as a noise resistant alternative quantitative metric to SUVmax.

Keywords Performance · Harmonisation · PET/CT · Quantification · EARL accreditation
Introduction

$^{18}$F–fluorodeoxyglucose ($^{18}$F–FDG) positron emission tomography (PET) and computed tomography (CT) hybrid imaging (PET/CT) is an important functional imaging tool being widely used for diagnosis, staging and therapy response evaluation in, e.g., oncology [1–20]. Combined anatomical and functional information can be obtained in one session using hybrid PET/CT. In clinical practice, visual inspection of PET/CT images might be sufficient for the purposes of staging or restaging [7, 21], however PET is a quantitative technique [22–26] and can provide more accurate and less observer-dependent metrics for diagnosis, therapy assessment and response monitoring using quantitative data in addition to visual interpretation [27]. In recent oncological clinical trials quantitative PET/CT data are also used for patient selection, stratification and therapy response monitoring. However, variability, reproducibility and accuracy of quantitative PET/CT imaging [28–34] have to be considered. Scientific societies such as the European Association of Nuclear Medicine (EANM), American College of Radiology (ACR), American Association of Physicists in Medicine (AAPM), Radiological Society of North America (RSNA) and Society of Nuclear Medicine and Molecular Imaging (SNMMI) are closely collaborating to promote standardisation of practices in order to reduce variability of quantification in multicentre clinical trials. Initiatives such as QIBA-UPICT, SNMMI-CTN and EANM-EARL are providing quality control programs to assure quantitative comparability [35–40].

High utilisation of PET/CT in oncology can be attributed to the availability of $^{18}$F–FDG [5, 41]. Dynamic PET scans and pharmacokinetic modelling to evaluate the rate of glucose metabolism of tumours is an excellent method for quantification [27] but the technical impediments such as the limited scanner field of view and increased scan acquisition time make it unfeasible for routine use [42]. In clinical practice, a simplified uptake metric such as the standard uptake value (SUV) [43, 44] is therefore most commonly used. While SUV analysis is relatively easy to apply, it suffers from multiple technical, physical and biological factors that can significantly affect quantification [27]. The required level of harmonisation depends on the intended use of the PET study. When the same PET/CT system is used for therapy assessment and based on relative changes in SUV before and after therapy, a high reproducibility rather than absolute accuracy might be most important. It has been shown that in this case, when the scanner performance remains unchanged over time, consistent application of a certain methodology could be sufficient [34, 45]. However, patients are often scanned on different PET/CT systems, either because the scanner had been replaced by a new one, or in different institutions, which makes accurate cross-calibration of systems a crucial requirement. Absolute quantitative measures (e.g., residual uptake of $^{18}$F–FDG after therapy session) are also being used for differentiation between malignant and benign lesions, determining prognosis and response monitoring [27]. This again requires high reproducibility and comparability of the quantitative data, especially in multicentre settings.

One of the challenges in PET/CT systems performance harmonisation is the variability caused by different PET/CT technologies available in the field. Multicentre standards should not be based on the less performing systems; they need to fit with the highest, yet common denominator in systems’ performance. Additionally, in case of optimization of PET/CT systems performance for lesion detection, a single centre quantification does not necessarily coincide with a multicentre one. A particular challenge for recent PET/CT systems resulted from the introduction of time-of-flight (ToF) and resolution modelling (point spread function (PSF)) capabilities. The latter increased tumour detectability but also caused higher variability across centres, since some have and others lack these technologies. Currently a large number of the EARL accredited PET/CT systems [46] do not have PSF image reconstruction capabilities. However, it is expected that over the next couple of years the majority of the PET/CT systems will be equipped with these new reconstruction techniques.

The aim of this paper is to explore the feasibility of harmonising performance of PET/CT systems equipped with the latest PET technologies such as TOF and PSF, which were recently commercially released.

Materials and methods

PET/CT system selection

Four PET/CT systems equipped with both ToF and PSF capabilities from three major vendors (General Electric (GE), Siemens and Philips) were selected for this study. Systems included were the Siemens Biograph mCT (Siemens system 1), the Siemens Biograph mCT Flow (Siemens system 2), the GE Discovery 710 (GE system) and the Philips Ingenity TF 128 (Philips system). The equipment was calibrated in accordance with the corresponding manufacturer’s instructions. In addition, all systems were participating and accredited in the EANM/EARL $^{18}$F–FDG PET/CT accreditation program. Detailed specifications for the systems can be found in supplemental Table 1 and references [47–51].

Phantom experiments

The phantoms and filling procedures used complied with the EANM/EARL guidelines for Image Quality QC measurements which need to be performed annually as part of the EANM/EARL accreditation program [35]. The NEMA NU2–2007 body phantom was used, which is a plastic cylinder in the form of a fillable torso cavity, to act as a background compartment.
has a 5 cm diameter cylindrical lung insert in the centre and six
fillable spheres with internal diameters of 10, 13, 17, 22, 28 and
37 mm, positioned coaxially around the lung insert. The lung
insert is filled with polystyrene beads in order to mimic lung
tissue. The phantom background compartment and the spheri-
cal inserts were filled with 18F–FDG solutions aimed at activity
centrations of 2 kBq/mL and 20 kBq/mL, respectively,
at the start of the measurements, resulting in a sphere to back-
ground activity concentration ratio of 10:1.

Acquisition and reconstruction parameters

In accordance with current EANM/EARL guidelines for 18F–
FDG Image Quality QC phantom imaging [35], a low dose CT
acquisition, followed by an emission scan consisting of two bed
positions with an acquisition time of 5 min per bed position is to
be acquired for the “image quality” dataset to assess contrast
recovery performance. In this study, acquisition time of 5 min
per bed position was selected as the reference for high count
statistics. In order to investigate the effect of reduced count
statistics on contrast recovery, data acquired with shorter acquisi-
tion times, respectively 2 and 1 min per bed position, were
collected. The GE and Philips systems had list mode data ac-
quisition capability available, which meant that only the 5 min/
bed position emission scans were acquired and reconstructions
with shorter acquisition times were generated retrospectively
from the list mode data. On the Siemens systems included in
this study, multiple shorter emission scans were acquired with the
phantom left in an unchanged position. In order to facilitate
the Siemens Flow system’s (Siemens system 2) possibility of
performing scanning with continuous table movement, instead
of a specific bed position scanning duration, table feed speeds
of 0.5 mm/s, 1 mm/s and 2 mm/s were selected, resulting in
similar acquisition times as with the other scanners.

Reconstructions were performed using the software avail-
able on each of the PET/CT systems. TOF, PSF, normalisa-
tion, randoms, scatter and attenuation corrections were applied
and the reconstruction parameters were selected to increase
overall contrast recovery, meanwhile aiming at achieving compar-
able recovery values across systems (for each sphere). In addi-
tion, we also considered achieving comparable recovery
values between the spheres to minimise severe partial
volume effects as well as large Gibbs overshoots. Clinically
used and vendor recommended reconstruction parameters
were applied and varied. Three iterations with 21 subsets were
used for Siemens 1 (Biograph mCT) and two iterations with
21 subsets for Siemens 2 (mCT Flow) reconstruction. For GE
- B, D, F and G (Discovery 710) - two iterations with 24
subsets and the VPFXS reconstruction method were used,
while for GE - A, C and E - the QCFX reconstruction method,
with an unknown number of iterations and subsets, was used.
For the Philips systems the iterations/subsets were 3/33 but
these could not be selected prior to scanning, with no values
retrieved from the DICOM header of the images; so the
BLOB OS TF reconstruction method was used. Different
Gaussian filters and pixel sizes within clinically relevant
ranges were also investigated in order to study their effects
on contrast recovery. Additionally, for the GE system, a pro-
prietary reconstruction method, the “Q.Clear”, which uses a
Bayesian penalised-likelihood reconstruction algorithm, was
investigated using different penalization factors (β) and its
effect on quantitative image quality was evaluated. Due to
differences among vendors and models, the available recon-
struction parameters and their ranges were limited based on
availability and/or user selectability. In total, 15 reconstruction
parameter sets (reconstruction modes) were used to assess and
compare the quantitative performance of the investigated sys-
tems. Each reconstruction mode was applied on three different
scans, acquired with long (~4 min/bed for the Siemens Flow
system; ~5 min/bed for all other systems), with medium
(~2 min/bed) and short (~1 min/bed) frame durations. A sum-
mary of the acquisition and reconstruction settings of the 15
reconstruction modes is presented in Table 1.

Data analysis

Data reconstructed on the PET/CT were exported to a PC for
further analysis using the EARL semi-automatic tool [35] de-
dsigned for quantitative analysis of images of the NEMA NU2–
2007 body phantom, filled conforming to EANM/EARL
guidelines for 18F–FDG Image Quality QC phantom imaging.
The software tool requires phantom images in DICOM format
and filling data as input, and extracts SUV recovery for the
spheres, a calibration factor for the background compartment
and standard deviation and coefficients of variation from uni-
form images of the background. The SUV recovery coefficient
(RC) is defined as the ratio between measured and expected
activity concentration in each spherical insert. RC values were
calculated based on 50% background corrected isocontour VOI
(RC_SUVmean), maximum voxel value included in VOI
(RC_SUVmax) and spherical VOI with a diameter of 12 mm, po-
positioned so to yield the highest uptake (RC_SUVpeak) [35, 39, 52].

Prior to further analysis, all data were corrected for system
 calibration bias in order to be able to compare the various recon-
struction modes’ impact on RCs and not to be effected by inter-
scanner calibration errors. For this purpose, to all RCs a correc-
tion factor, defined as the ratio between measured and expected
activity concentration in each spherical insert. RC values were
calculated based on 50% background corrected isocontour VOI
(RC_SUVmean), maximum voxel value included in VOI
(RC_SUVmax) and spherical VOI with a diameter of 12 mm, po-
positioned so to yield the highest uptake (RC_SUVpeak) [35, 39, 52].

Prior to further analysis, all data were corrected for system
calibration bias in order to be able to compare the various recon-
struction modes’ impact on RCs and not to be effected by inter-
scanner calibration errors. For this purpose, to all RCs a correc-
tion factor, defined as the ratio between measured and expected
activity concentration in each spherical insert. RC values were
calculated based on 50% background corrected isocontour VOI
(RC_SUVmean), maximum voxel value included in VOI
(RC_SUVmax) and spherical VOI with a diameter of 12 mm, po-
positioned so to yield the highest uptake (RC_SUVpeak) [35, 39, 52].

Prior to further analysis, all data were corrected for system
calibration bias in order to be able to compare the various recon-
struction modes’ impact on RCs and not to be effected by inter-
scanner calibration errors. For this purpose, to all RCs a correc-
tion factor, defined as the ratio between measured and expected
activity concentration in each spherical insert. RC values were
calculated based on 50% background corrected isocontour VOI
(RC_SUVmean), maximum voxel value included in VOI
(RC_SUVmax) and spherical VOI with a diameter of 12 mm, po-
positioned so to yield the highest uptake (RC_SUVpeak) [35, 39, 52].

Prior to further analysis, all data were corrected for system
calibration bias in order to be able to compare the various recon-
struction modes’ impact on RCs and not to be effected by inter-
scanner calibration errors. For this purpose, to all RCs a correc-
tion factor, defined as the ratio between measured and expected
activity concentration in each spherical insert. RC values were
calculated based on 50% background corrected isocontour VOI
(RC_SUVmean), maximum voxel value included in VOI
(RC_SUVmax) and spherical VOI with a diameter of 12 mm, po-
positioned so to yield the highest uptake (RC_SUVpeak) [35, 39, 52].

Prior to further analysis, all data were corrected for system
calibration bias in order to be able to compare the various recon-
struction modes’ impact on RCs and not to be effected by inter-
scanner calibration errors. For this purpose, to all RCs a correc-
tion factor, defined as the ratio between measured and expected
activity concentration in each spherical insert. RC values were
calculated based on 50% background corrected isocontour VOI
(RC_SUVmean), maximum voxel value included in VOI
(RC_SUVmax) and spherical VOI with a diameter of 12 mm, po-
positioned so to yield the highest uptake (RC_SUVpeak) [35, 39, 52].

Prior to further analysis, all data were corrected for system
calibration bias in order to be able to compare the various recon-
struction modes’ impact on RCs and not to be effected by inter-
scanner calibration errors. For this purpose, to all RCs a correc-
tion factor, defined as the ratio between measured and expected
activity concentration in each spherical insert. RC values were
calculated based on 50% background corrected isocontour VOI
(RC_SUVmean), maximum voxel value included in VOI
(RC_SUVmax) and spherical VOI with a diameter of 12 mm, po-
positioned so to yield the highest uptake (RC_SUVpeak) [35, 39, 52].

Prior to further analysis, all data were corrected for system
calibration bias in order to be able to compare the various recon-
struction modes’ impact on RCs and not to be effected by inter-
scanner calibration errors. For this purpose, to all RCs a correc-
tion factor, defined as the ratio between measured and expected
activity concentration in each spherical insert. RC values were
calculated based on 50% background corrected isocontour VOI
(RC_SUVmean), maximum voxel value included in VOI
(RC_SUVmax) and spherical VOI with a diameter of 12 mm, po-
positioned so to yield the highest uptake (RC_SUVpeak) [35, 39, 52].

Prior to further analysis, all data were corrected for system
calibration bias in order to be able to compare the various recon-
struction modes’ impact on RCs and not to be effected by inter-
scanner calibration errors. For this purpose, to all RCs a correc-
tion factor, defined as the ratio between measured and expected
activity concentration in each spherical insert. RC values were
calculated based on 50% background corrected isocontour VOI
(RC_SUVmean), maximum voxel value included in VOI
(RC_SUVmax) and spherical VOI with a diameter of 12 mm, po-
positioned so to yield the highest uptake (RC_SUVpeak) [35, 39, 52].
Selection of harmonising reconstruction modes

The primary objective of this study was to find reconstruction modes providing high, yet uniform contrast recoveries within the spheres of the NEMA NU2–2007 body phantom, which could be matched across all generations of PET/CT systems currently used in clinical practice – which would result in quantitative harmonisation of PET/CT systems.

RC\textsubscript{SUV\text{mean}}, RC\textsubscript{SUV\text{max}} and RC\textsubscript{SUV\text{peak}} curves for all reconstructed phantom images were plotted against sphere diameters (Fig. 1) and characterised using visual and quantitative analysis, for which the applied metrics are summarised in Table 2. Reconstruction modes with higher RCs than current EARL specifications, as well as tightly grouped and stable RC\textsubscript{SUV\text{mean}} and RC\textsubscript{SUV\text{max}} curves, were sought for harmonisation purposes.

The harmonising reconstruction modes were selected by simultaneously analysing quantitative characteristics of the reconstruction modes along with visual appearance of the RC curves. The following considerations were kept in mind while determining feasible reconstruction modes – (1) the proposed harmonising specifications should provide an increase over the current EARL compliant RC values, (2) the bandwidth of RCs should be similar to the current Earl specification limits and (3) the harmonising RC curves should not demonstrate major overshoots (upward bias) due to Gibbs artefacts. While the harmonising reconstruction modes were selected based on the abovementioned considerations, quantitative cut-off criteria were retrospectively determined and stated in Table 9 based on the bandwidth and characteristics of harmonising reconstruction modes. Performances of the candidate reconstruction modes were compared with the initial group of reconstructions as well as current EARL accreditation specifications.

Mean contrast recovery (MCR)

Mean contrast recovery (MCR) was calculated in order to evaluate overall contrast recovery potential of a reconstruction mode while Coefficient of Variation of the MCR parameter (CoV\textsubscript{MCR}) was used to characterise agreement among various reconstruction modes’ RC curves. Increased coinciding MCR and reduced CoV\textsubscript{MCR} values were preferred.

Contrast recovery variability (CRV)

Contrast Recovery Variability (CRV\text{medium} and CRV\text{short}) parameters were used to evaluate a reconstruction mode’s ability to produce consistent results in case of reduced count statistics. In order to achieve it, RCs of short and medium time frame acquisitions were compared to the long acquisition’s corresponding spheres’ RCs and relative differences calculated. Lower values were deemed preferable as being indicative of reconstruction mode’s stability and reduced variability in noisy environments.

Noise

Image noise was quantitatively evaluated by measuring the Coefficient of Variation (% SD/\text{Mean}x100) in the uniform background compartment (CoV\textsubscript{BG}) for each reconstruction mode and acquisition time frame. CoV\textsubscript{BG} cut-off limit of 15%, based on the existing EARL guideline and UPICT [35, 37, 40], was implemented to determine suitable reconstruction
modes for harmonisation. Reconstruction modes providing lower noise images were deemed preferable.

Curvature and absolute error

Curvature and absolute error parameters were used to evaluate RC variability and absolute accuracy of RC measurements due to changes in sphere/lesion size. Reduced values were preferable, but similar magnitude across systems/reconstructions was given priority.

Visual analysis

Visual analysis of the RC curves was used to identify reconstruction modes that exhibited abnormal behaviour or localised variations, such as exaggerated Gibbs artefacts, that were not identified by the previously described quantitative parameters.

The reconstruction modes, which were considered for harmonisation based on SUV\textsubscript{mean} and SUV\textsubscript{max} performance, were also used to develop provisional specifications for SUV\textsubscript{peak}.

Validation of reconstruction modes for harmonisation

In order to prospectively evaluate the reproducibility and inter-scanner variability of the proposed reconstruction modes for harmonisation, 16 EARL accredited facilities, equipped with current generation PET/CT systems, participated in the study and provided the requested reconstructions from independent phantom acquisitions applying acquisition and reconstruction parameters (supplemental Table 2) identical or similar to the reconstructions proposed for harmonisation purposes. Data received from the centres was analysed in the same way as the reconstructions in the pilot study.
Results

New specifications proposed for harmonisation

Analysis of the initial 15 reconstruction modes resulted in five reconstruction modes, which produced the highest uniform contrast recoveries and were feasible for all of the investigated systems considering SUVmean and SUVmax (Philips - B, GE – E, GE - F, Siemens 1 – D and Siemens 2 – A), to be considered for harmonisation. In order to accommodate unavoidable inter-scanner variability and reproducibility errors due to equipment calibration and user inaccuracy, all of the RC ranges were expanded to be proportional (i.e., using the same bandwidth of performance, but taking into account increased contrast recovery) to current EARL specifications for sphere recoveries. Bandwidths for proposed and current EARL specifications as well as the RC curves derived from the five reconstruction modes are presented in Fig. 2. For the provisional SUVpeak specifications, average sphere recoveries of the five reconstruction modes and a bandwidth of ±2 standard deviations was used. Additionally, recovery coefficients are plotted as a function of background noise for each sphere and per SUVmetric.

Table 2 Description of quantitative metrics used

| Metric            | Description of metric                                                                 |
|-------------------|---------------------------------------------------------------------------------------|
| SUVmean           | Ratio of image derived average radioactivity concentration within a region of interest and the whole body concentration of the injected radioactivity |
| SUVmax            | Ratio of image derived maximum (single pixel) radioactivity concentration within a region of interest and the whole body concentration of the injected radioactivity |
| SUVpeak           | Ratio of image derived average radioactivity concentration within a 12 mm diameter spherical volume within the region of interest, positioned to yield the highest uptake, and the whole body concentration of the injected radioactivity |
| RC                | Recovery Coefficient - the ratio between image derived and expected activity concentration |
| MCR*              | Mean Contrast Recovery - mean RC of all spheres in corresponding reconstruction mode’s long duration acquisition. Parameter is indicative of reconstruction mode’s overall contrast recovery potential. |
| CoVMCR            | Coefficient of Variation (SD/mean*100, %) of a group of MCR values. Parameter is indicative of RC curves’ alignment within a group. |
| CRVmedium*        | Contrast Recovery Variability - Mean deviation of medium duration acquisition spheres’ RCs from the corresponding values of long duration acquisition. |
| CRVshort*         | Contrast Recovery Variability - Mean deviation of short duration acquisition spheres’ RCs from the corresponding values of long duration acquisition. |
| CoVBG*            | Coefficient of Variation (SD/mean *100, %) of measured activity concentration within the uniform background compartment of the phantom. Parameter is indicative of the noise present in the images. |
| Curvature         | Long acquisition duration root-mean-square deviation of spheres’ RC values from RC value of the largest (37 mm) sphere. Parameter characterises the deviation of smaller spheres’ RC values which usually cause the RC-object size relation to assume a curved shape. |
| Absolute error    | Long acquisition duration root-mean-square deviation of spheres’ RC values from unity. The parameter characterises the reconstruction mode’s ability to report accurate activity concentration values. |
| Curvature (excl. 10 mm sphere) | Same as “curvature” but excluding the smallest (10 mm) sphere. |
| Absolute error (excl. 10 mm sphere) | Same as “absolute error” but excluding the smallest (10 mm) sphere. |

*Quantitative metrics that were retrospectively used to determine harmonising cut-off criteria.
Axial slices of the phantom data from the five harmonising reconstructions are shown in supplemental Fig. 7.

Mean contrast recovery (MCR)

SUVmean and SUVmax RC curves vary substantially among different systems and reconstruction modes as seen in Fig. 1 and Tables 3 and 4. The reconstruction mode showing the lowest recoveries (Siemens 1 – E) produced a SUVmean MCR value of 0.714 and SUVmax MCR of 0.948 while for the highest recovery reconstruction mode (Siemens 1 – A), the corresponding values were 1.09 and 1.56 – a difference of more than 50%. SUVpeak MCR values were found to be between 0.754 and 0.929. CoV_{MCR} values for the 15 reconstruction modes were 12.4% and 15.4% for SUVmean and SUVmax, respectively, while for SUVpeak, CoV_{MCR} was 6.0%.

Contrast recovery variability (CRV)

The initial 15 reconstruction modes demonstrated a variable sensitivity as a function of count statistics. The expected increase in variability with decrease in count statistics was observed in all reconstruction modes by comparing CRV_{medium} and CRV_{short} values (Tables 3, 4 and 5). The CRV_{medium} results for SUVmean, SUVmax and SUVpeak ranged from 2.4% to 8.4%, 2.7% to 17.8% and 1.6% to 4.5%, respectively. The CRV_{short} results for SUVmean, SUVmax and SUVpeak

For the five reconstruction modes proposed for harmonisation, the range of MCR values were 0.770–0.816 and 1.01–1.09 for SUVmean and SUVmax, respectively. The harmonising reconstruction modes produced SUVpeak MCR values in the range of 0.784–0.823. CoV_{MCR} values for SUVmean, SUVmax and SUVpeak were 2.2%, 2.9% and 2.2%, respectively.
ranged from 2.3% to 14.5%, 4.9% to 20.4% and 2.7% to 6.3%, respectively.

For the five reconstruction modes proposed for harmonisation, the CRV_{\text{medium}} results for SUV_{\text{mean}},

| Reconstruction mode | MCR | CRV_{\text{medium}} | CRV_{\text{short}} | Curvature | Absolute error | Curvature (excl. 10 mm sphere) | Absolute error (excl. 10 mm sphere) |
|---------------------|-----|----------------------|---------------------|-----------|----------------|----------------------------------|-------------------------------------|
| GE - A              | 0.956 | 5.0%                 | 6.8%                | 0.031     | 0.053          | 0.202                            | 0.040                               |
| GE - B              | 0.903 | 8.4%                 | 8.7%                | 0.139     | 0.147          | 0.022                            | 0.050                               |
| GE - C              | 0.887 | 6.7%                 | 6.4%                | 0.109     | 0.140          | 0.025                            | 0.077                               |
| GE - D              | 0.859 | 6.3%                 | 6.6%                | 0.168     | 0.188          | 0.053                            | 0.092                               |
| GE - E              | 0.806 | 5.2%                 | 6.2%                | 0.218     | 0.253          | 0.075                            | 0.134                               |
| GE - F              | 0.770 | 5.3%                 | 5.3%                | 0.228     | 0.277          | 0.120                            | 0.183                               |
| GE - G              | 0.725 | 3.8%                 | 4.9%                | 0.253     | 0.321          | 0.147                            | 0.228                               |
| Philips - A         | 0.845 | 3.3%                 | 4.2%                | 0.149     | 0.192          | 0.088                            | 0.134                               |
| Philips - B         | 0.800 | 2.7%                 | 2.3%                | 0.236     | 0.271          | 0.124                            | 0.165                               |
| Siemens 1 - A       | 1.086 | 6.0%                 | 14.5%               | 0.097     | 0.117          | 0.108                            | 0.125                               |
| Siemens 1 - B       | 1.038 | 3.8%                 | 12.5%               | 0.072     | 0.076          | 0.071                            | 0.081                               |
| Siemens 1 - C       | 0.952 | 3.3%                 | 8.3%                | 0.111     | 0.101          | 0.048                            | 0.043                               |
| Siemens 1 - D       | 0.816 | 2.9%                 | 5.1%                | 0.197     | 0.222          | 0.097                            | 0.138                               |
| Siemens 1 - E       | 0.714 | 2.4%                 | 4.0%                | 0.269     | 0.329          | 0.166                            | 0.238                               |
| Siemens 2 - A       | 0.804 | 3.0%                 | 4.4%                | 0.203     | 0.238          | 0.100                            | 0.150                               |
| Min                 | 0.714 | 2.4%                 | 2.3%                | 0.031     | 0.053          | 0.022                            | 0.040                               |
| Max                 | 1.086 | 8.4%                 | 14.5%               | 0.269     | 0.329          | 0.166                            | 0.238                               |
| Average             | 0.864 | 4.5%                 | 6.7%                | 0.165     | 0.195          | 0.085                            | 0.125                               |
| COV_{\text{MCR}}   |      |                      |                     |           |                |                                  | 12.4%                               |

Table 3 Analysis results of 15 initial reconstruction modes using a SUV_{\text{mean}} quantitative metric. Values found to be outside of acceptable range during retrospective quantitative analysis, are coloured red

| Reconstruction mode | MCR | CRV_{\text{medium}} | CRV_{\text{short}} | Curvature | Absolute error | Curvature (excl. 10 mm sphere) | Absolute error (excl. 10 mm sphere) |
|---------------------|-----|----------------------|---------------------|-----------|----------------|----------------------------------|-------------------------------------|
| GE - A              | 1.245 | 17.8%                | 20.4%               | 0.081     | 0.255          | 0.089                            | 0.265                               |
| GE - B              | 1.201 | 11.9%                | 19.7%               | 0.160     | 0.236          | 0.052                            | 0.257                               |
| GE - C              | 1.142 | 12.9%                | 15.1%               | 0.076     | 0.157          | 0.036                            | 0.172                               |
| GE - D              | 1.139 | 10.6%                | 15.8%               | 0.181     | 0.194          | 0.047                            | 0.200                               |
| GE - E              | 1.036 | 7.2%                 | 7.7%                | 0.212     | 0.178          | 0.041                            | 0.119                               |
| GE - F              | 1.013 | 8.0%                 | 9.2%                | 0.235     | 0.170          | 0.085                            | 0.099                               |
| GE - G              | 0.951 | 5.5%                 | 6.6%                | 0.274     | 0.203          | 0.129                            | 0.094                               |
| Philips - A         | 1.146 | 7.2%                 | 15.0%               | 0.176     | 0.204          | 0.103                            | 0.218                               |
| Philips - B         | 1.061 | 3.7%                 | 5.2%                | 0.267     | 0.232          | 0.150                            | 0.197                               |
| Siemens 1 - A       | 1.555 | 10.1%                | 20.3%               | 0.126     | 0.566          | 0.139                            | 0.574                               |
| Siemens 1 - B       | 1.477 | 8.0%                 | 19.1%               | 0.116     | 0.487          | 0.112                            | 0.505                               |
| Siemens 1 - C       | 1.325 | 5.4%                 | 12.5%               | 0.148     | 0.346          | 0.104                            | 0.375                               |
| Siemens 1 - D       | 1.094 | 3.9%                 | 7.9%                | 0.218     | 0.179          | 0.080                            | 0.165                               |
| Siemens 1 - E       | 0.948 | 2.7%                 | 4.9%                | 0.290     | 0.199          | 0.145                            | 0.084                               |
| Siemens 2 - A       | 1.045 | 3.7%                 | 5.4%                | 0.246     | 0.184          | 0.104                            | 0.138                               |
| Min                 | 0.948 | 2.7%                 | 4.9%                | 0.076     | 0.157          | 0.036                            | 0.084                               |
| Max                 | 1.555 | 17.8%                | 20.4%               | 0.290     | 0.566          | 0.150                            | 0.574                               |
| Average             | 1.159 | 7.9%                 | 12.3%               | 0.187     | 0.253          | 0.094                            | 0.231                               |
| COV_{\text{MCR}}   |      |                      |                     |           |                |                                  | 15.4%                               |

Table 4 Analysis results of 15 initial reconstruction modes using a SUV_{\text{max}} quantitative metric. Values found to be outside of acceptable range during retrospective quantitative analysis, are coloured red
SUVmax and SUVpeak ranged from 2.7% to 5.3%, 3.7% to 8.0% and 2.8% to 3.0%, respectively. The CRV short results for SUVmean, SUVmax and SUVpeak ranged from 2.3% to 6.2%, 5.2% to 9.2% and 2.9% to 5.8%, respectively (Tables 6, 7 and 8).

Noise

The CoVBG values are summarised in supplemental Fig. 8. The average CoVBG of all reconstruction modes with a long time frame was 12.6%. For medium and short acquisition times, the corresponding values were 19.7% and 27.0%, respectively. The selected reconstruction modes for harmonisation purposes produced average CoVBG values of 9.4%, 14.0% and 18.4% for long, medium and short acquisition time frames, respectively.

Curvature and absolute error

Curvatures for the initial 15 reconstruction modes were in the ranges of 0.031–0.269, 0.076–0.290 and 0.305–0.413 for long, medium and short acquisition times, respectively.

Table 5  Analysis results of 15 initial reconstruction modes using SUVpeak quantitative metric

| Reconstruction mode | MCR   | CRV<sub>medium</sub> | CRV<sub>short</sub> | Curvature | Absolute error | Curvature (excl. 10 mm sphere) | Absolute error (excl. 10 mm sphere) |
|---------------------|-------|----------------------|---------------------|-----------|----------------|---------------------------------|----------------------------------|
| GE - A              | 0.848 | 3.9%                 | 3.7%                | 0.334     | 0.287          | 0.187                           | 0.153                            |
| GE - B              | 0.833 | 3.4%                 | 5.7%                | 0.381     | 0.310          | 0.237                           | 0.179                            |
| GE - C              | 0.840 | 2.3%                 | 3.6%                | 0.359     | 0.302          | 0.211                           | 0.166                            |
| GE - D              | 0.823 | 3.9%                 | 6.3%                | 0.389     | 0.320          | 0.248                           | 0.191                            |
| GE - E              | 0.821 | 2.9%                 | 4.1%                | 0.400     | 0.339          | 0.250                           | 0.203                            |
| GE - F              | 0.784 | 3.3%                 | 5.8%                | 0.404     | 0.346          | 0.272                           | 0.223                            |
| GE - G              | 0.757 | 3.1%                 | 5.9%                | 0.413     | 0.367          | 0.287                           | 0.248                            |
| Philips - A         | 0.874 | 3.2%                 | 3.4%                | 0.328     | 0.281          | 0.192                           | 0.161                            |
| Philips - B         | 0.796 | 2.8%                 | 2.9%                | 0.383     | 0.341          | 0.263                           | 0.229                            |
| Siemens 1 - A       | 0.901 | 4.5%                 | 6.3%                | 0.305     | 0.232          | 0.148                           | 0.090                            |
| Siemens 1 - B       | 0.929 | 1.6%                 | 4.2%                | 0.325     | 0.240          | 0.154                           | 0.103                            |
| Siemens 1 - C       | 0.872 | 3.3%                 | 5.0%                | 0.308     | 0.251          | 0.151                           | 0.107                            |
| Siemens 1 - D       | 0.823 | 3.0%                 | 4.5%                | 0.350     | 0.291          | 0.204                           | 0.155                            |
| Siemens 1 - E       | 0.754 | 3.9%                 | 2.7%                | 0.382     | 0.346          | 0.255                           | 0.226                            |
| Siemens 2 - A       | 0.789 | 2.9%                 | 4.9%                | 0.355     | 0.323          | 0.240                           | 0.214                            |
| Min                 | 0.754 | 1.6%                 | 2.7%                | 0.305     | 0.232          | 0.148                           | 0.090                            |
| Max                 | 0.929 | 4.5%                 | 6.3%                | 0.413     | 0.367          | 0.287                           | 0.248                            |
| Average             | 0.830 | 3.2%                 | 4.6%                | 0.361     | 0.305          | 0.220                           | 0.177                            |
| COVMCR              | 6.0%  |                      |                     |           |                |                                 |                                  |

Table 6  Results of the analysis of five reconstruction modes considered for harmonisation using the SUVmean quantitative metric

| Reconstruction mode | MCR   | CRV<sub>medium</sub> | CRV<sub>short</sub> | Curvature | Absolute error | Curvature (excl. 10 mm sphere) | Absolute error (excl. 10 mm sphere) |
|---------------------|-------|----------------------|---------------------|-----------|----------------|---------------------------------|----------------------------------|
| GE - E              | 0.806 | 5.2%                 | 6.2%                | 0.218     | 0.253          | 0.075                           | 0.134                            |
| GE - F              | 0.770 | 5.3%                 | 5.3%                | 0.228     | 0.277          | 0.120                           | 0.183                            |
| Philips - B         | 0.800 | 2.7%                 | 2.3%                | 0.236     | 0.271          | 0.124                           | 0.165                            |
| Siemens 1 - D       | 0.816 | 2.9%                 | 5.1%                | 0.197     | 0.222          | 0.097                           | 0.138                            |
| Siemens 2 - A       | 0.804 | 3.0%                 | 4.4%                | 0.203     | 0.238          | 0.100                           | 0.150                            |
| Min                 | 0.770 | 2.7%                 | 2.3%                | 0.197     | 0.222          | 0.075                           | 0.134                            |
| Max                 | 0.816 | 5.3%                 | 6.2%                | 0.236     | 0.277          | 0.124                           | 0.183                            |
| Average             | 0.799 | 3.8%                 | 4.6%                | 0.216     | 0.252          | 0.103                           | 0.154                            |
| COVMCR              | 2.2%  |                      |                     |           |                |                                 |                                  |
| EARL min            | 0.570 | N/A                  | N/A                 | 0.282     | 0.466          | 0.198                           | 0.393                            |
| EARL max            | 0.710 | N/A                  | N/A                 | 0.277     | 0.342          | 0.176                           | 0.251                            |
| EARL Average        | 0.640 | N/A                  | N/A                 | 0.279     | 0.403          | 0.187                           | 0.321                            |
SUV\textsubscript{mean}, SUV\textsubscript{max} and SUV\textsubscript{peak}, respectively. For the five reconstruction modes suggested for harmonisation, the SUV\textsubscript{mean}, SUV\textsubscript{max} and SUV\textsubscript{peak} curvatures were in the ranges of $0.197 - 0.236$, $0.212 - 0.267$ and $0.350 - 0.404$, respectively.

Absolute errors for the initial 15 reconstruction modes were in the ranges of $0.053 - 0.329$, $0.157 - 0.566$ and $0.232 - 0.367$ for SUV\textsubscript{mean}, SUV\textsubscript{max} and SUV\textsubscript{peak}, respectively. For the five reconstruction modes selected for harmonisation, the SUV\textsubscript{mean}, SUV\textsubscript{max} and SUV\textsubscript{peak} curvatures ranged between $0.222 - 0.277$, $0.170 - 0.232$ and $0.291 - 0.346$, respectively.

### Visual analysis

Significant variations in investigated RC curves’ shapes and positions of Siemens 1 - A, B, C, GE - A, B, C, D and Philips - A reconstruction modes were noticed when compared with other systems or acquisition times and considered unsuitable for harmonisation. Based on the bandwidth and characteristics of harmonising reconstruction modes, quantitative cut-off criteria were determined and are stated in Table 9.

#### Additional reconstructions

Sixteen EARL accredited sites participated in the prospective evaluation of the newly proposed specifications for harmonisation and performed reconstructions according to instructions provided. Data received included 23 distinctive reconstructions from three GE Discovery 710 systems, two Philips Ingenuity systems, three Siemens mCT systems, three Siemens mCT Flow systems, one GE Discovery IQ system, two GE Discovery MI systems and one Philips Vereos system. RC curves derived from the 18 systems along with proposed new harmonising specifications can be seen in Fig. 3. For SUV\textsubscript{mean}, 16 out of 138 analysed spheres produced RC values outside of the suggested

| Reconstruction mode | MCR  | CRV\textsubscript{medium} | CRV\textsubscript{short} | Curvature | Absolute error | Curvature (excl. 10 mm sphere) | Absolute error (excl. 10 mm sphere) |
|---------------------|------|---------------------------|---------------------------|-----------|----------------|----------------------------------|-------------------------------------|
| GE - E              | 1.036| 7.2%                      | 7.7%                      | 0.212     | 0.178          | 0.041                            | 0.119                               |
| GE - F              | 1.013| 8.0%                      | 9.2%                      | 0.235     | 0.170          | 0.085                            | 0.099                               |
| Philips - B         | 1.061| 3.7%                      | 5.2%                      | 0.267     | 0.232          | 0.150                            | 0.197                               |
| Siemens 1 - D       | 1.094| 3.9%                      | 7.9%                      | 0.218     | 0.179          | 0.080                            | 0.165                               |
| Siemens 2 - A       | 1.045| 3.7%                      | 5.4%                      | 0.246     | 0.184          | 0.104                            | 0.138                               |
| Min                 | 1.013| 3.7%                      | 5.2%                      | 0.212     | 0.170          | 0.041                            | 0.099                               |
| Max                 | 1.094| 8.0%                      | 9.2%                      | 0.267     | 0.232          | 0.150                            | 0.197                               |
| Average             | 1.050| 5.3%                      | 7.1%                      | 0.236     | 0.189          | 0.092                            | 0.144                               |
| COVMCR              | 2.9% |                           |                           |           |                |                                  |                                     |
| EARL min            | 0.730| N/A                       | N/A                       | 0.347     | 0.355          | 0.220                            | 0.237                               |
| EARL max            | 0.970| N/A                       | N/A                       | 0.339     | 0.236          | 0.176                            | 0.121                               |
| EARL Average        | 0.850| N/A                       | N/A                       | 0.342     | 0.277          | 0.198                            | 0.142                               |

### Table 8 Results of the analysis of five reconstruction modes considered for harmonisation using the SUV\textsubscript{peak} quantitative metric

| Reconstruction mode | MCR  | CRV\textsubscript{medium} | CRV\textsubscript{short} | Curvature | Absolute error | Curvature (excl. 10 mm sphere) | Absolute error (excl. 10 mm sphere) |
|---------------------|------|---------------------------|---------------------------|-----------|----------------|----------------------------------|-------------------------------------|
| GE - E              | 0.821| 2.9%                      | 4.1%                      | 0.400     | 0.339          | 0.250                            | 0.203                               |
| GE - F              | 0.784| 3.3%                      | 5.8%                      | 0.404     | 0.346          | 0.272                            | 0.223                               |
| Philips - B         | 0.796| 2.8%                      | 2.9%                      | 0.383     | 0.341          | 0.263                            | 0.229                               |
| Siemens 1 - D       | 0.823| 3.0%                      | 4.5%                      | 0.350     | 0.291          | 0.204                            | 0.155                               |
| Siemens 2 - A       | 0.789| 2.9%                      | 4.9%                      | 0.355     | 0.323          | 0.240                            | 0.214                               |
| Min                 | 0.784| 2.8%                      | 2.9%                      | 0.350     | 0.291          | 0.204                            | 0.155                               |
| Max                 | 0.823| 3.3%                      | 5.8%                      | 0.404     | 0.346          | 0.272                            | 0.229                               |
| Average             | 0.803| 3.0%                      | 4.4%                      | 0.378     | 0.328          | 0.246                            | 0.205                               |
| COVMCR              | 2.2% |                           |                           |           |                |                                  |                                     |
accreditation interval, while for SUVmax and SUVpeak, the number of outliers was 12. Quantitative results describing additional reconstructions can be found in Tables 10, 11 and 12. Specifications, based on the current findings, proposed for harmonisation along with current EARL specifications are presented in Table 13.

Discussion

The SUVmean and SUVmax RC curves of the initial 15 reconstruction modes vary significantly, even within one system. This reflects the high degree of variability that could be introduced into quantitative PET with variation in reconstruction settings. The selection of harmonising reconstruction modes, and the validation which followed on additional reconstructions, demonstrated that the variability can be reduced to acceptable limits.

The acquisition time of 5 min per bed position specified in the current EARL accreditation settings, while characterising system performance in high statistics scenarios, may not provide an accurate representation of the reconstruction mode’s performance in clinical settings. Therefore, the observation of reduced CRVmedium and CRVshort in reconstruction modes for harmonisation is important since the acquisition times when
Significant increase in both SUVmean and SUVmax MCR values was observed in the reconstruction modes proposed for harmonisation compared to the corresponding current EARL specifications. The trend is in agreement with results recently published by Sunderland et al. demonstrating that high-end PET/CT systems are having significantly increased SUVmax values in anthropomorphic phantom scans [53]. The metrics for all of the spheres demonstrated a noticeable increase; however, for the smaller spheres (≤ 17 mm) the effect was relatively stronger. This could be explained by the so-called Gibbs artefact which produces an overshoot of measured activity at the edges of the spheres, becoming more dominant at smaller sizes, also described by Lasnon et al. [54]. To some extent the effect can be considered beneficial, compensating for the inherently lower recoveries seen in the smaller spheres. It should, however, be noticed that with the use of resolution modelling (PSF) without any or with minimal post filtering applied, the overshoot could introduce significant positive SUV bias, in particular when using SUVmax. Methods like regularised (MAP) reconstruction with a regularising prior (such as Q.Clear implemented by GE) can also be used to suppress Gibbs artefacts and were therefore also considered in this study.

The increased SUVmean and SUVmax recoveries seen in the proposed reconstruction modes for harmonisation would significantly reduce the gap that exists today between standardised quantitative reconstruction protocols used in multicentre settings and the locally developed non-standard protocols for lesion detection and general visual assessment – both of which are used in parallel in many nuclear medicine departments. Close agreement between the two could lead to the adoption of a single reconstruction mode that would provide standardised SUV data while maintaining increased lesion detectability.

In the reconstruction modes identified as suitable candidates for harmonisation, a relatively higher increase was found in the

| PET/CT system | MCR | CRVmedium | CRVshort | Curvature | Absolute error | Curvature (excl. 10 mm sphere) | Absolute error (excl. 10 mm sphere) |
|---------------|-----|-----------|----------|-----------|----------------|-------------------------------|-----------------------------------|
| Ingenuity 1   | 0.820 | N/A       | N/A      | 0.213     | 0.249          | 0.106                         | 0.145                             |
| Ingenuity 2   | 0.694 | N/A       | N/A      | 0.276     | 0.365          | 0.164                         | 0.263                             |
| mCT Flow 1    | 0.691 | N/A       | N/A      | 0.303     | 0.368          | 0.196                         | 0.270                             |
| mCT Flow 2    | 0.711 | N/A       | N/A      | 0.298     | 0.339          | 0.190                         | 0.242                             |
| mCT Flow 3    | 0.816 | N/A       | N/A      | 0.193     | 0.231          | 0.079                         | 0.136                             |
| mCT 1         | 0.847 | N/A       | N/A      | 0.176     | 0.194          | 0.080                         | 0.112                             |
| mCT 2         | 0.786 | N/A       | N/A      | 0.194     | 0.250          | 0.115                         | 0.181                             |
| mCT 3         | 0.825 | N/A       | N/A      | 0.188     | 0.208          | 0.113                         | 0.142                             |
| mCT 4         | 0.765 | N/A       | N/A      | 0.174     | 0.262          | 0.091                         | 0.195                             |
| mCT 5         | 0.786 | N/A       | N/A      | 0.195     | 0.245          | 0.119                         | 0.179                             |
| mCT 6         | 0.811 | N/A       | N/A      | 0.136     | 0.207          | 0.078                         | 0.161                             |
| Discovery 710 1 | 0.847 | N/A       | N/A      | 0.153     | 0.182          | 0.079                         | 0.120                             |
| Discovery 710 2 | 0.793 | N/A       | N/A      | 0.217     | 0.254          | 0.129                         | 0.174                             |
| Discovery 710 1 Q.Clear 1 | 0.887 | N/A       | N/A      | 0.120     | 0.145          | 0.027                         | 0.074                             |
| Discovery 710 2 Q.Clear 2 | 0.817 | N/A       | N/A      | 0.211     | 0.236          | 0.110                         | 0.146                             |
| Discovery 710 3 Q.Clear 3 | 0.895 | N/A       | N/A      | 0.121     | 0.144          | 0.042                         | 0.073                             |
| GE Discovery MI 1 | 0.794 | N/A       | N/A      | 0.150     | 0.228          | 0.099                         | 0.182                             |
| GE Discovery MI 2 | 0.813 | N/A       | N/A      | 0.171     | 0.214          | 0.102                         | 0.155                             |
| GE Discovery MI 1 Q.Clear 1 | 0.857 | N/A       | N/A      | 0.081     | 0.151          | 0.055                         | 0.129                             |
| GE Discovery MI 2 Q.Clear 2 | 0.869 | N/A       | N/A      | 0.118     | 0.156          | 0.039                         | 0.096                             |
| GE Discovery IQ 1 | 0.817 | N/A       | N/A      | 0.219     | 0.244          | 0.077                         | 0.123                             |
| GE Discovery IQ 1 Q.Clear 1 | 0.818 | N/A       | N/A      | 0.221     | 0.246          | 0.069                         | 0.118                             |
| Vereos 1      | 0.757 | N/A       | N/A      | 0.191     | 0.277          | 0.087                         | 0.195                             |
| Min           | 0.691 | 0.081     | 0.144    | 0.027     | 0.073          |                               |                                   |
| Max           | 0.895 | 0.303     | 0.368    | 0.196     | 0.270          |                               |                                   |
| Average       | 0.805 | 0.188     | 0.235    | 0.098     | 0.157          |                               |                                   |
| COVMCR        | 6.6%  |           |          |           |                |                               |                                   |
recoveries of smaller spheres. This would lead to more “flat” RC curves, making subsequent quantitative analysis less dependent on lesion size. With the proposed reconstruction modes, the recoveries remained largely size-independent for ≥17 mm diameter lesions. Moreover, it is important to notice that a possible new harmonising standard for systems with PSF implies SUVmax recoveries to exceed 1.0. This suggests that if SUVmax remains the de facto field standard for PET/CT quantification, one should accept a positive bias of about 10 to 25% for larger homogeneous objects (≥17 mm diameter).

For both SUVmean and SUVmax the proposed reconstruction modes for harmonisation yielded promising results. The two largest spheres (28 mm diameter, 37 mm diameter) showed excellent agreement across all systems for both SUVmean and SUVmax. Even though there is not enough data for a reproducibility assessment, it can be predicted that a harmonising performance bandwidth is feasible for the next generation of PET/CT systems. The results from prospective validation using additional reconstructions will be further improved in the EARL accreditation process, where the centres will be guided to optimise their reconstruction settings in order to meet the new specifications.

As the harmonising RCs for SUVmean, SUVmax and SUVpeak all demonstrated a noticeable curve, the curvature and absolute error parameters exhibited increased or similar values with the initial reconstruction modes. Calculations excluding the smallest sphere demonstrated much better performance, which illustrated the high impact the smallest sphere has, that led to a significant decrease in the RCs range.

The utility of the SUVpeak was investigated as being a possible metric for standardised quantification. A recent prospective repeatability study by Kramer et al. [55] demonstrated the robustness of using the SUVpeak in non–small cell lung cancer patients. As previously shown by Makris et al. [56], and presented in supplemental Figs. 4–6, SUVpeak is significantly less sensitive to changes in reconstruction parameters and acquisition durations than SUVmean or SUVmax. The difference is mostly prominent in the initial group of 15

| PET/CT system                  | MCR   | $\text{CRV}_{\text{medium}}$ | $\text{CRV}_{\text{short}}$ | Curvature | Absolute error | Curvature (excl. 10 mm sphere) | Absolute error (excl. 10 mm sphere) |
|--------------------------------|-------|------------------------------|-----------------------------|-----------|---------------|-------------------------------|-------------------------------------|
| Ingenuity 1                    | 1.094 | N/A                          | N/A                         | 0.278     | 0.264         | 0.143                         | 0.228                               |
| Ingenuity 2                    | 0.917 | N/A                          | N/A                         | 0.334     | 0.288         | 0.188                         | 0.167                               |
| mCT Flow 1                     | 0.911 | N/A                          | N/A                         | 0.347     | 0.270         | 0.207                         | 0.159                               |
| mCT Flow 2                     | 0.943 | N/A                          | N/A                         | 0.350     | 0.234         | 0.187                         | 0.109                               |
| mCT Flow 3                     | 1.071 | N/A                          | N/A                         | 0.237     | 0.211         | 0.110                         | 0.179                               |
| mCT 1                          | 1.118 | N/A                          | N/A                         | 0.185     | 0.179         | 0.057                         | 0.179                               |
| mCT 2                          | 1.038 | N/A                          | N/A                         | 0.173     | 0.140         | 0.065                         | 0.108                               |
| mCT 3                          | 1.098 | N/A                          | N/A                         | 0.168     | 0.148         | 0.082                         | 0.151                               |
| mCT 4                          | 1.019 | N/A                          | N/A                         | 0.160     | 0.130         | 0.041                         | 0.082                               |
| mCT 5                          | 1.033 | N/A                          | N/A                         | 0.176     | 0.127         | 0.067                         | 0.092                               |
| mCT 6                          | 1.067 | N/A                          | N/A                         | 0.113     | 0.107         | 0.033                         | 0.105                               |
| Discovery 710 1                | 1.139 | N/A                          | N/A                         | 0.151     | 0.176         | 0.051                         | 0.188                               |
| Discovery 710 2                | 1.045 | N/A                          | N/A                         | 0.213     | 0.168         | 0.086                         | 0.130                               |
| Discovery 710 1 Q.Clear 1      | 1.172 | N/A                          | N/A                         | 0.085     | 0.189         | 0.054                         | 0.207                               |
| Discovery 710 2 Q.Clear 2      | 1.049 | N/A                          | N/A                         | 0.204     | 0.172         | 0.064                         | 0.131                               |
| Discovery 710 3 Q.Clear 3      | 1.154 | N/A                          | N/A                         | 0.114     | 0.184         | 0.042                         | 0.200                               |
| GE Discovery MI 1              | 1.055 | N/A                          | N/A                         | 0.105     | 0.100         | 0.032                         | 0.095                               |
| GE Discovery MI 2              | 1.066 | N/A                          | N/A                         | 0.179     | 0.142         | 0.065                         | 0.125                               |
| GE Discovery MI 1 Q.Clear 1    | 1.119 | N/A                          | N/A                         | 0.040     | 0.123         | 0.017                         | 0.108                               |
| GE Discovery MI 2 Q.Clear 2    | 1.124 | N/A                          | N/A                         | 0.107     | 0.157         | 0.039                         | 0.168                               |
| GE Discovery IQ 1              | 1.102 | N/A                          | N/A                         | 0.255     | 0.240         | 0.047                         | 0.201                               |
| GE Discovery IQ 1 Q.Clear 1    | 1.083 | N/A                          | N/A                         | 0.234     | 0.219         | 0.052                         | 0.177                               |
| Vereos 1                       | 1.029 | N/A                          | N/A                         | 0.230     | 0.176         | 0.074                         | 0.115                               |
| Min                            | 0.911 | N/A                          | N/A                         | 0.040     | 0.100         | 0.017                         | 0.082                               |
| Max                            | 1.172 | N/A                          | N/A                         | 0.350     | 0.288         | 0.207                         | 0.228                               |
| Average                        | 1.063 | 0.193                        | 0.180                       | 0.078     | 0.148         |                               |                                     |

$\text{COV}_{\text{MCR}}$ 6.3%
relatively loosely selected reconstruction modes, while within the five reconstructions for harmonisation and 23 additional ones, the difference became less apparent. On the other hand, the benefits of SUVpeak were offset by its consistently low recoveries for spheres with ≤17 mm diameter and therefore low MCR, which is comparable to that of SUVmean but significantly (20–40%) lower than that of SUVmax. This is due to peak VOI size approaching or even exceeding the size of the sphere, therefore missing some of the active volume. If this issue could be addressed by, for example, reducing the SUVpeak VOI size, SUVpeak may become an effective alternative to SUVmax, especially if quantitative comparison among reconstructions of unknown origin or non-harmonised PET/CT systems is desired. Harmonisation among systems

| PET/CT system | MCR | CRVmedium | CRVshort | Curvature | Absolute error | Curvature (excl. 10 mm sphere) | Absolute error (excl. 10 mm sphere) |
|---------------|-----|-----------|----------|-----------|----------------|---------------------------------|------------------------------------|
| Ingenuity 1   | 0.789 | N/A | N/A | 0.376 | 0.341 | 0.246 | 0.218 |
| Ingenuity 2   | 0.736 | N/A | N/A | 0.405 | 0.383 | 0.284 | 0.267 |
| mCT Flow 1    | 0.737 | N/A | N/A | 0.439 | 0.390 | 0.324 | 0.280 |
| mCT Flow 2    | 0.750 | N/A | N/A | 0.476 | 0.379 | 0.353 | 0.263 |
| mCT Flow 3    | 0.797 | N/A | N/A | 0.393 | 0.328 | 0.274 | 0.217 |
| mCT 1         | 0.858 | N/A | N/A | 0.348 | 0.282 | 0.214 | 0.162 |
| mCT 2         | 0.812 | N/A | N/A | 0.347 | 0.302 | 0.225 | 0.188 |
| mCT 3         | 0.847 | N/A | N/A | 0.365 | 0.281 | 0.242 | 0.169 |
| mCT 4         | 0.781 | N/A | N/A | 0.326 | 0.313 | 0.198 | 0.192 |
| mCT 5         | 0.803 | N/A | N/A | 0.355 | 0.304 | 0.243 | 0.199 |
| mCT 6         | 0.827 | N/A | N/A | 0.297 | 0.269 | 0.184 | 0.163 |
| Discovery 710 1 | 0.829 | N/A | N/A | 0.357 | 0.301 | 0.234 | 0.188 |
| Discovery 710 2 | 0.794 | N/A | N/A | 0.398 | 0.342 | 0.274 | 0.227 |
| Discovery 710 1 Q.Clear 1 | 0.867 | N/A | N/A | 0.372 | 0.294 | 0.231 | 0.171 |
| Discovery 710 2 Q.Clear 2 | 0.824 | N/A | N/A | 0.413 | 0.344 | 0.276 | 0.221 |
| Discovery 710 3 Q.Clear 3 | 0.884 | N/A | N/A | 0.370 | 0.298 | 0.212 | 0.166 |
| GE Discovery MI 1 | 0.797 | N/A | N/A | 0.351 | 0.313 | 0.233 | 0.202 |
| GE Discovery MI 2 | 0.819 | N/A | N/A | 0.375 | 0.308 | 0.237 | 0.180 |
| GE Discovery MI 1 Q.Clear 1 | 0.838 | N/A | N/A | 0.328 | 0.285 | 0.200 | 0.166 |
| GE Discovery MI 2 Q.Clear 2 | 0.859 | N/A | N/A | 0.356 | 0.294 | 0.202 | 0.157 |
| GE Discovery IQ 1 | 0.814 | N/A | N/A | 0.407 | 0.342 | 0.263 | 0.210 |
| GE Discovery IQ 1 Q.Clear 1 | 0.831 | N/A | N/A | 0.412 | 0.336 | 0.258 | 0.199 |
| Vereos 1      | 0.803 | N/A | N/A | 0.381 | 0.320 | 0.251 | 0.199 |
| Min           | 0.736 | N/A | N/A | 0.297 | 0.269 | 0.184 | 0.157 |
| Max           | 0.884 | N/A | N/A | 0.476 | 0.390 | 0.353 | 0.280 |
| Average       | 0.813 | N/A | N/A | 0.376 | 0.320 | 0.246 | 0.200 |
| COVMCR        | 4.7%             | | | | | | |

Table 13  SUVmean, SUVmax and SUVpeak specifications proposed for harmonisation along with current EARL specifications

| Sphere diameter (mm) | Current EARL RC bandwidth | Proposed RC bandwidth |
|----------------------|---------------------------|-----------------------|
|                      | SUVmean  | SUVmax  | SUVpeak | SUVmean  | SUVmax  | SUVpeak |
| 37                   | 0.76–0.89 | 0.95–1.16 | N/A     | 0.85–1.00 | 1.05–1.29 | 0.99–1.07 |
| 28                   | 0.72–0.85 | 0.91–1.13 | N/A     | 0.82–0.97 | 1.01–1.26 | 0.95–1.07 |
| 22                   | 0.63–0.78 | 0.83–1.09 | N/A     | 0.80–0.99 | 1.01–1.32 | 0.90–1.09 |
| 17                   | 0.57–0.73 | 0.73–1.01 | N/A     | 0.76–0.97 | 1.00–1.38 | 0.75–0.99 |
| 13                   | 0.44–0.60 | 0.59–0.85 | N/A     | 0.63–0.86 | 0.85–1.22 | 0.45–0.69 |
| 10                   | 0.27–0.43 | 0.34–0.57 | N/A     | 0.39–0.61 | 0.52–0.88 | 0.27–0.41 |
remains necessary in order to enable reliable use of SUV\text{max}. Further studies are needed in order to explore the optimal peak VOI diameter maintaining noise cancelling effects, while producing higher, yet harmonised recoveries.

An alternative to the described methodology of achieving harmonised recoveries, such as suggested in this paper, could be to gradually increase the post smoothing on high recovery PET data until harmonised RC-s are obtained (supplemental Figs. 9–11). Such a method is available on some systems and previously validated by Lasnon et al. [54]. Potentially a post-smoothing feature on a workstation could be used for this purpose. This could result in higher recoveries and may reduce noise and Gibbs artefacts to acceptable levels for multicentre harmonisation. However, when offline post-smoothing needs to be applied to a dataset in order for it to achieve quantitative harmonisation, the filter information for the specific system always needs to accompany the PET data and extra care be taken that the filter be actually applied and clearly reported every time when required.

**Limitations and future directions**

Quantification of PET images is affected by uncertainties derived from reconstruction settings as well as global system (cross-) calibration. In this study the experimental data were corrected for global calibration errors, but in clinical practice both effects should be taken into consideration. Therefore, an accurate system calibration remains of utmost importance for all PET/CT systems used for quantification in order to keep the uncertainties as low as possible.

The phantom experiments conducted were sensitive to measurement uncertainties of dose calibrators and human error during the phantom preparation phase. The uncertainties related to phantom filling procedure are not part of this study and may increase the bandwidth of achievable harmonisation.

All experiments on various PET/CT models were performed on appointed systems. The inter-system variability stemming from the individual differences among the systems of the same make and differences due to manufacturers allowed variability in well counter calibration factors, and may increase the bandwidth of achievable harmonisation even further, although the newly proposed harmonisation specification was set using the same bandwidth as current EARL, which was shown to be appropriate and feasible.

As the position of VOI-s used in the analysis and comparison of SUV\text{mean} data is based on PET images rather that CT data, it is to some extent affected by image noise and may induce a small additional uncertainty to the results. This, however, is reflective of the clinically used method of VOI positioning. When this strategy is followed, it is therefore important to also put a threshold on acceptable noise levels (in this paper background noise should be lower than 15\%). Yet, use of CT-based VOI definition could be of interest in order to mitigate the effects of noise on VOI definition and subsequently on the measurement of the recovery coefficients. Another alternative could be the use of SUV\text{peak} rather than SUV\text{max} as a starting point for VOI definition, as was applied in Frings et al. [57]. These strategies may be considered when developing future standards.

Current study investigated harmonisation of PET/CT systems using \textsuperscript{18}F tracer based FDG. The results cannot be directly transferred to system performance harmonisation involving other PET isotopes such as \textsuperscript{68}Ga or \textsuperscript{82}Rb which have a substantially longer positron range. System performance harmonisation with positron emitting isotopes other than \textsuperscript{18}F requires further investigation.

In this feasibility study we primarily made use of reconstruction methods and parameter settings that were predefined or could be easily set by the user on commercially released systems. Where the software permitted, we applied additional reconstructions to include at least PSF and TOF, and also tried other reconstruction parameter settings which were expected to yield higher recoveries than the current EARL specification. Yet, in this study we did not extensively explore a wide range of reconstruction settings as, e.g., iterations, subsets, matrix sizes, etc., since our aim was to investigate clinically available protocols which are accessible for the users. Moreover, the investigated reconstruction modes had similar, but still different, voxel sizes as well as the number of iterations/subsets between various systems which complicates direct comparison. In conclusion, the harmonisation investigated in this study should be considered as a first feasibility test aiming at improving the current EARL specifications. Of course, a higher level of harmonisation would also be possible by considering more parameters, but then the question will be the feasibility in clinical practice. Further work is also needed to more extensively explore the impact of PSF reconstructions, voxel size and number of iterations/subsets on the variability of quantitative metrics of clinical datasets. Some reports have already been published showing that the repeatability and ICC of SUV\text{max}, SUV\text{peak} and SUV\text{mean} are at an acceptable level [58].

To conclude, despite possible limitations, we have studied the feasibility of the harmonising state of the art PET/CT system performances, and the results suggest that an update of the EARL current specification is feasible and achievable in practice.

**Conclusions**

This study investigated the feasibility of harmonising performance for PET/CT systems equipped with the latest Time-of-Flight (ToF) and resolution modelling (PSF) technology. Also, new possible specifications with higher contrast recoveries were investigated using various metrics such as average, maximum and peak SUV. Harmonisation state of the art PET/CT systems with ToF and PSF technologies was found to be
feasible. The harmonisation of such systems would require an update to the current multicentre accreditation program of EARL in order to accommodate higher recoveries. SUVpeak could be used as an uptake metric being less sensitive to noise and variation in image quality resulting from different reconstruction settings. It could be considered as an alternative to SUVmax if lower recoveries are considered to be acceptable for lesions of 17 mm in diameter and smaller.

Acknowledgements The publication of this article was supported by funds of the European Association of Nuclear Medicine (EANM). We would like to thank the EARL sites that provided additional reconstructions of the image quality QC phantom experiments, namely: University Clinic for Nuclear Medicine and Endocrinology, Salzburg, Austria.
VU University Medical Centre, Amsterdam, The Netherlands.
GIE Humanitép, Lille, France.
University Hospital of Cologne, Cologne, Germany.
Ghent University Hospital, Ghent, Belgium.
Zentralklinik Bad Berka, Bad Berka, Germany.
University Medical Centre Utrecht, Utrecht, The Netherlands.
Centre Antoine Lacassagne, Nice, France.
VieCuri Medical Centre, Venlo, The Netherlands.
University Hospital Magdeburg, Magdeburg, Germany.
Cancer Institute Nantes-Atlantiques, St. Herblain, France.
Groene Hart Ziekenhuis, Gouda, The Netherlands.
University Hospital of Navarra, Pamplona, Spain.
University Hospital Antwerp, Edegem, Belgium.
Heinrich-Heine University Hospital, Düsseldorf, Germany.
University Hospital Brussels, Jette, Belgium.
Rispbjerg Hospital, Copenhagen, Denmark.
St. Vincent’s Hospital PET – CT Centre, Linz, Austria.
Postgraduate Institute of Medical Education and Research, Chandigarh, India.
King’s College Hospital NHS Foundation Trust, London, United Kingdom.
Uppsala University Hospital, Uppsala, Sweden.

Compliance with ethical standards
Andres Kaalep declares that he has no conflict of interest. Terez Sera has received travel grants and honoraria from EARL. Sjoerd Rijnisendorp declares that he has no conflict of interest. Maqsood Yaqub declares that he has no conflict of interest. Anne Talsma declares that he has no conflict of interest. Maqsood Yaqub declares that he has no conflict of interest. Maqsood Yaqub declares that he has no conflict of interest. "Erdi YE. The use of PET for radiotherapy. Curr Med Imaging Rev. 2007;3:3–16.
Fletcher JW. Djulbegovic B, Soares HP, Siegel B a, Lowe VJ, Lyman GH, et al. recommendations on the use of 18F-FDG PET in oncology. J Nucl Med. 2008;49:480–508.
de Geus-Oei L-F, van der Heijden HFM, Corstens FH, Oyen WJG. Predictive and prognostic value of FDG-PET in nonsmall-cell lung cancer. Cancer [internet]. Wiley subscription services, Inc. A Wiley Company. 2007;110:1654–64. https://doi.org/10.1002/cncr.22979.
Hoeckstra CJ, Stroobants SG, Smit EF, Vansteenkiste J, Van Tinteren H, Postmus PE, et al. Prognostic relevance of response evaluation using [18F]-2-fluoro-2-deoxy-D-glucose positron emission tomography in patients with locally advanced non-small-cell lung cancer. J Clin Oncol. 2005;23:8362–70.
Juweid ME, Stroobants S, Hoeckstra OS, Mottaghy FM, Dietlein M, Guermazi A, et al. Use of positron emission tomography for response assessment of lymphoma: consensus of the imaging subcommittee of international harmonization project in lymphoma. J Clin Oncol. 2007;25:571–8.
Vansteenkiste JF, Stroobants SG. The role of positron emission tomography with 18F-fluoro-2-deoxy-D-glucose in respiratory oncology. Eur Respir J. 2001;17:802–20.
Weber WA. Use of PET for monitoring cancer therapy and for predicting outcome. J. Nucl. Med. [Internet]. 2005;46:983–95. Available from: http://jnm.snmjournals.org/content/46/6/983.
abstract
Borst GR, Belderbos JS, Boellaard R, EFI C, De Jaeger K, Lammertsma AA, et al. Standardised FDG uptake: a prognostic factor for inoperable non-small cell lung cancer. Eur J Cancer. 2005;41:1533–41.
Shankar LK, Hoffman JM, Bacharach S, Graham MM, Karp J, Lammertsma AA, et al. Consensus recommendations for the use of 18F-FDG PET as an indicator of therapeutic response in patients in National Cancer Institute trials. J Nucl Med. 2006;47:1059–66.
Gupta T, Master Z, Kannan S, Agarwal JP, Gishob-Laskar S, Rangarajan V, et al. Diagnostic performance of post-treatment FDG PET or FDG PET/CT imaging in head and neck cancer: a systematic review and meta-analysis. Eur J Nucl Med Mol Imaging. 2011;38:2083.
Ung VC, Mazia DE, Vanderveen JA, Smith CA, Gulenchyn K, Lacchetti C, et al. 18fluorodeoxyglucose positron emission tomography in the diagnosis and staging of lung cancer: a systematic review. J Natl Cancer Inst. 2007;99:1753–67.
Lv Y-L, Yuan D-M, Wang K, Miao X-H, Qian Q, Wei S-Z, 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 International Association for the Study of Lung Cancer. 2011;6:1350–8.
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.
Hicks RJ, Kalff V, Macmanus MP, Ware RE, Mckenzie AF, Matthews JP, et al. The utility of F-FDG PET for suspected

References
1. Avril NE, Weber WA. Monitoring response to treatment in patients utilizing PET. Radiol Clin N Am. 2005;43:189–204.
2. Bastiaanmet E, Groen B, Jager PL, Cobben DCP, van der Graaf WTA, Vialburg W, et al. The value of FDG-PET in the detection, grading and response to therapy of soft tissue and bone sarcomas; a systematic review and meta-analysis. Cancer Treat Rev. 2004;30:83–101.
3. Erdi YE. The use of PET for radiotherapy. Curr Med Imaging Rev. 2007;3:3–16.
4. Fletcher JW. Djulbegovic B, Soares HP, Siegel B a, Lowe VJ, Lyman GH, et al. recommendations on the use of 18F-FDG PET in oncology. J Nucl Med. 2008;49:480–508.
de Geus-Oei L-F, van der Heijden HFM, Corstens FH, Oyen WJG. Predictive and prognostic value of FDG-PET in nonsmall-cell lung cancer. Cancer [internet]. Wiley subscription services, Inc. A Wiley Company. 2007;110:1654–64. https://doi.org/10.1002/cncr.22979.
Hoeckstra CJ, Stroobants SG, Smit EF, Vansteenkiste J, Van Tinteren H, Postmus PE, et al. Prognostic relevance of response evaluation using [18F]-2-fluoro-2-deoxy-D-glucose positron emission tomography in patients with locally advanced non-small-cell lung cancer. J Clin Oncol. 2005;23:8362–70.
Juweid ME, Stroobants S, Hoeckstra OS, Mottaghy FM, Dietlein M, Guermazi A, et al. Use of positron emission tomography for response assessment of lymphoma: consensus of the imaging subcommittee of international harmonization project in lymphoma. J Clin Oncol. 2007;25:571–8.
Vansteenkiste JF, Stroobants SG. The role of positron emission tomography with 18F-fluoro-2-deoxy-D-glucose in respiratory oncology. Eur Respir J. 2001;17:802–20.
Weber WA. Use of PET for monitoring cancer therapy and for predicting outcome. J. Nucl. Med. [Internet]. 2005;46:983–95. Available from: http://jnm.snmjournals.org/content/46/6/983.
abstract
Borst GR, Belderbos JS, Boellaard R, EFI C, De Jaeger K, Lammertsma AA, et al. Standardised FDG uptake: a prognostic factor for inoperable non-small cell lung cancer. Eur J Cancer. 2005;41:1533–41.
Shankar LK, Hoffman JM, Bacharach S, Graham MM, Karp J, Lammertsma AA, et al. Consensus recommendations for the use of 18F-FDG PET as an indicator of therapeutic response in patients in National Cancer Institute trials. J Nucl Med. 2006;47:1059–66.
Gupta T, Master Z, Kannan S, Agarwal JP, Gishob-Laskar S, Rangarajan V, et al. Diagnostic performance of post-treatment FDG PET or FDG PET/CT imaging in head and neck cancer: a systematic review and meta-analysis. Eur J Nucl Med Mol Imaging. 2011;38:2083.
Ung VC, Mazia DE, Vanderveen JA, Smith CA, Gulenchyn K, Lacchetti C, et al. 18fluorodeoxyglucose positron emission tomography in the diagnosis and staging of lung cancer: a systematic review. J Natl Cancer Inst. 2007;99:1753–67.
Lv Y-L, Yuan D-M, Wang K, Miao X-H, Qian Q, Wei S-Z, 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 International Association for the Study of Lung Cancer. 2011;6:1350–8.
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.
Hicks RJ, Kalff V, Macmanus MP, Ware RE, Mckenzie AF, Matthews JP, et al. The utility of 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;42:10–21.

17. Dijkman BG, Schuurbers OCJ, Vriens D, Looijen-Salomon M, Bussink J, Timmer-Bonte JNH, 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 Berlin/Heidelberg: Springer-Verlag. 2010;37:207–47.

18. Hellwig D, Graeter TP, Ukkena D, Groeschel A, Ukena D, Groeschel A, et al. 18F-FDG PET/CT for mediastinal staging of lung cancer: which SUV threshold makes sense? J Nucl Med. 2007;48:1761–6.

19. Agarwal M, Brahmanday G, Bajaj SK, Ravikrishnan KP, Wong C-YO. Revisiting the prognostic value of preoperative (18)F-fluoro-2-deoxyglucose ([18F-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.

20. 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 Mol Imaging. 2012;39:27–38.

21. Cheson BD, Pfistner B, Juweid ME, Specht L, Wagner SN, et al. Revised response criteria for malignant lymphoma. J Clin Oncol. 2007;25:579–86.

22. Hicks RJ. Role of 18F-FDG PET in assessment of response in non-small cell lung cancer. J Nucl Med. 2009;50:31S–42.

23. Czernin J, Weber WA, Horschman HR. Molecular imaging in the development of cancer therapeutics. Annu Rev Med United States. 2006;57:99–118.

24. Frank R, Hargreaves R. Clinical biomarkers in drug discovery and development. Nat Rev Drug Discov England. 2003;2:566–72.

25. Weber WA, Petersen V, Schmidt B, Tyndale-Hines L, Link T, Biermann C, et al. Performance evaluation of the biograph mCT flow PET/CT system according to the NEMA NU2-2012 standard. EJNMMI Phys. 2015;2:26.

26. Rausch I, Cal-González J, Dapra D, Gallowitsch HJ, Lind P, Beyer T, et al. Performance evaluation of the biograph mct flow PET/CT system according to the NEMA NU2-2012 standard. EJNMMI Phys. 2015;2:26.

27. Boellaard R. Standards for PET image acquisition and quantitative data analysis. J Nucl Med. 2009;50(Suppl 1):1S–20S.

28. Boellaard R. Need for standardization of 18F-FDG PET/CT for treatment response assessments. J Nucl Med [Internet]. 2011;52(Suppl 2):93S–100S. Available from: http://jnm.snmjournals.org/content/52/Supplement_2/93S.full

29. Doot RK, Pierce LA, Byrd D, Eliston B, Allberg KC, Kinahan PE. Biases in multicenter longitudinal PET standardized uptake value measurements. Transl Oncol [Internet]. 2014;7:48–54. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3998681&tool=pmcrender&rendertype=abstract

30. Feuardent J, Soret M, De Dreuleille O, Foehrenbach H, Buvat I. Reliability of uptake estimates in FDG PET as a function of acquisition and processing protocols using the CPET. IEEE Trans Nucl Sci. 2005;52:1447–52.

31. Kurland BF, Gerstner ER, Mountz JM, Schwartz LH, Ryan CW, Graham MM, et al. Promise and pitfalls of quantitative imaging in oncology clinical trials. Magn Reson Imaging. 2012;30:1301–12.

32. Lammertsma AA, Hoekstra CJ, Giaccone G, Hoekstra OS. How should we analyse FDG PET studies for monitoring tumour response? Eur J Nucl Med Mol Imaging. 2006;33

33. Lammertsma AA, Hoekstra CJ, Giaccone G, Hoekstra OS. How should we analyse FDG PET studies for monitoring tumour response? Eur J Nucl Med Mol Imaging. 2006;33

34. 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.

35. Boellaard R, Delgado-Bolton R, Oyen WJG, 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. 2014;42:328–54.

36. Delbeke D, Coleman RE, Guiberteau MJ, Brown ML, Royal HD. Siegel BA, et al. procedure guideline for tumor imaging with 18F-FDG PET/CT. J Nucl Med. 2006;47:885–95.

37. Graham MM, Walth RL, Hoffman JM, Yap JT, Sunderland JJ, Boellaard R, et al. Summary of the UPICCT protocol for 18F-FDG PET/CT imaging in oncology clinical trials. J Nucl Med. 2015;56:955–61.

38. Scheuermann JS, Saffer JR, Karp JS, Levering AM, Siegel A. Qualification of PET scanners for use in multicenter cancer clinical trials: the American College of Radiology Imaging Network experience. J Nucl Med. 2010;50:1187–93.

39. Kinahan P, Wahl R, Shao L, Frank R, Perlm A. The QIBA profile for quantitative FDG-PET/CT oncology imaging. J Nucl Med. 2014;55:1520.

40. Boellaard R, Willemsen A, Tenders, Visser EP. EARL procedure for assessing PET/CT system specific patient FDG activity prepreparation for quantitative PET/CT studies. 2013, p. 1–3.

41. Weber WA. PET for response assessment in oncology: radiotherapy and chemotherapy. Br J Radiol. 2005;78:42–9.

42. Young H, Baum R, Cremerius U, Herholz K, Hoeckstra O, Lammersma AA, et al. Measurement of clinical and subclinical tumour response using [18F]-fluorodeoxyglucose and positron emission tomography: review and 1999 EORTC recommendations. Eur J Cancer. 1999;35:1773–82.

43. Huang A. Anatomy of SUV. Nucl Med Biol [Internet]. 2000;27:1063–1071. Available from: https://doi.org/10.1016/S0969-8051(00)00155-4.

44. Thie JA. Understanding the standardized uptake value, its methods, and implications for usage. J Nucl Med. 2004;45:1431–4.

45. Krak NC, Boellaard R, Hoekstra OS, Twisk JWR, Hoekstra CJ, Lammersma AA. Effects of ROI definition and reconstruction method on quantitative outcome and applicability in a response monitoring trial. J Nucl Med Mol Imaging. 2005;32:294–301.

46. Sera T. Experiences and Results of the PET/CT Accreditation Program: The 150 Sites Milestone. EANM Comm. Symp. 2016.

47. Rausch I, Cal-González J, Dapra D, Gallowitsch HJ, Lind P, Beyer T, et al. Performance evaluation of the biograph mct flow PET/CT system according to the NEMA NU2-2012 standard. EJNMMI Phys. 2015;2:26.

48. Karlberg AM, Sarther O, Ekenes L, Gao P, Quantitative comparison of PET performance-Siemens Biograph mCT and mMR. EJNMMI Phys. 2015;2:3.

49. Kolthammer J, Su K-H, Grover A, Narayanan M, Jordan DW, Muzic RF. Performance evaluation of the ingenuity TF PET/CT scanner with a focus on high count-rate conditions. Phys Med Biol. 2014;59:3843–59. Available from: http://iopscience.iop.org/article/10.1088/0031-9155/59/14/3843

50. GE Healthcare Discovery PET/CT 710 system manufacturer webpage [Internet]. Available from: http://www3.gehealthcare.com/en/products/categories/pet/ct/pet-ct_scanners/discovery_pet-ct_710

51. Philips USA website [Internet]. Available from: https://www.usa.philips.com/healthcare/product/HC882456/ingenuity-tf-pet-ct-scanner

52. 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.

53. 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.
54. Lasnon C, Desmonds 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.
55. Kramer GM, Frings V, Hoetjes N, Hoekstra OS, Smit EF, de Langen AJ, et al. Repeatability of quantitative whole-body 18F-FDG PET/CT uptake measures as function of uptake interval and lesion selection in non-small cell lung cancer patients. J Nucl Med [Internet]. 2016;57:1343–9. Available from: http://jnm.snmjournals.org/cgi/doi/10.2967/jnumed.115.170225
56. 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.
57. Frings V, van Velden FHP, Velasquez LM, Hayes W, van de Ven PM, Hoekstra OS, et al. Repeatability of metabolically active tumor volume measurements with FDG PET/CT in advanced gastrointestinal malignancies: a multicenter study. Radiology [Internet] Radiological Society of North America. 2014;273:539–48. https://doi.org/10.1148/radiol.14132807.
58. van Velden FHP, Kramer GM, Frings V, Nissen IA, Mulder ER, de Langen AJ, et al. Repeatability of Radiomic features in non-small-cell lung cancer [18F]FDG-PET/CT studies: impact of reconstruction and delineation. Mol. Imaging biol. [internet]. Mol Imaging Biol. 2016;18:788–95. https://doi.org/10.1007/s11307-016-0940-2.