Evaluation of a Method Combining Physical Experiment and Data Insertion For the Assessment of PET-CT Systems

Quentin Maronnier (✉ maronnier.quentin@iuct-oncopole.fr)

IUCT Oncopole: Institut Universitaire du Cancer Toulouse Oncopole <https://orcid.org/0000-0002-2266-9122>

Frédéric Courbon

Institut Claudius Regaud: Institut Universitaire du Cancer Toulouse Oncopole

Olivier Caselles

Institut Claudius Regaud: Institut Universitaire du Cancer Toulouse Oncopole

Original research

Keywords: Positron-emission tomography, Performance, Methods, Phantom, Experiment, Simulation.

Posted Date: October 18th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-960402/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Evaluation of a method combining physical experiment and data insertion for the assessment of PET-CT systems

Quentin Maronnier¹, Frédéric Courbon¹, Olivier Caselles¹

1. Medical Physics Department, Institut Claudius Regaud, Toulouse (France)

Corresponding author: maronnier.quentin@iuct-oncopole.fr

Quentin Maronnier's ORCID: 0000-0002-2266-9122

Abstract

Background: To evaluate and compare Positron Emission Tomography (PET) devices among them, tests are performed on phantoms that generally consist in simple geometrical objects, fillable with radiotracers. On one hand, those tests bring a control over the experiment through the operator preparation but on the other hand, they are limited in terms of reproducibility, repeatability and are time-consuming, in particular, if several replications are required. To overcome these restrictions, we designed a method combining physical experiment and data insertion that aims to avoid experimental repetitions while testing multiple configurations for the performance evaluation of PET scanners.

Methods: Based on the National Electrical Manufacturers Association Image Quality standard, four experiments, with different spheres-to-background ratios: 2:1, 4:1, 6:1 and 8:1, were performed. An additional acquisition was done with a radioactive background and no activity within the spheres. It was created as a baseline to artificially simulate the radioactive spheres and reproduce initial experiments. Standard sphere set was replaced by smaller target sizes (4, 5, 6, 8, 10 and 13 mm) to match current detectability performance of PET scanners. Images were reconstructed following standard guidelines, i.e. using OSEM algorithm, and an additional BPL reconstruction was performed. We visually compared experimental and simulated images. We measured the activity concentration values into the spheres to calculate the mean and maximum recovery coefficient (RC_{mean} and RC_{max}) which we used in a quantitative analysis.

Results: No significant visual discrepancies were identified between experimental and simulated series. Mann-Whitney U tests comparing simulated and experimental distributions

showed no statistical differences for both RC_{mean} (P value = 0.611) and RC_{max} (P value = 0.720). Spearman tests revealed high correlation for RC_{mean} ($\rho = 0.974$, P value < 0.001) and RC_{max} ($\rho = 0.974$, P value < 0.001) between both datasets. According to Bland-Altman plots, we highlighted slight shifts in RC_{mean} and RC_{max} of respectively $2.1 \pm 16.9 \%$ and $3.3 \pm 22.3 \%$.

Conclusions: The method produced realistic results compared to experimental data. Known synthesized information fused with original data allows full exploration of the system's capabilities while avoiding the limitations associated with repeated experiments.

Keywords: Positron-emission tomography - Performance - Methods - Phantom - Experiment – Simulation.

Background

Since the first Positron Emission Tomography (PET) scanners were introduced, the molecular imaging modality has been technologically redesigned and enhanced to reach higher sensitivity, spatial and timing resolutions. This has been achieved through the development of its hardware components (scintillator crystal, photodetector, electronic) and software solutions (reconstruction and image analysis) [1]. As a result, overall performances of PET devices have been greatly increased, but with various technological and commercial strategies among manufacturers leading to different characteristics. Therefore, the assessment of their performances is essential to classify the systems, which can be achieved through tests based on standards that were defined by scientific experts from national and international authorities such as the National Electrical Manufacturers Association (NEMA) and the International Electrotechnical Commission (IEC) [2, 3]. These procedures involve experiments performed on phantoms that generally consist of relatively simple geometrical objects, fillable with radioactive aqueous solutions. Phantom tests are useful for controlling experimental parameters but, on the other hand, are limited by the operator's contribution in terms of reproducibility and repeatability [4]. In addition, some parameters do not vary during the acquisition, such as the activity ratio between two compartments. Hence, to test n configurations, the experiment will have to be repeated n times. Moreover, these phantoms are often no longer adapted neither to PET performances nor to current clinical challenges such as the detectability of sub-centimeter lesions [5, 6]. An auditable way to study PET system performances would be to simulate data to avoid the introduction of biases and experimental repetitions. Indeed, simulation has several advantages among experiments performed on physical phantoms, such as a better control on

the input parameters (activity, volume, location) and the ability to explore multiple configurations without the need of the scanner access and the materials required for the experiments (phantoms, radiotracer). However, simulation often relies on an estimation of the physics processes involved in data acquisition and reconstruction [7-9] and is computational-demanding [7]. To overcome the limitations of both physical and simulated experiments, we designed a hybrid method combining both modalities while taking advantage of their respective benefits. The method is based on an experimental acquisition carried out on the physical PET system to which a data insertion process is applied to generate different configurations. The objective of data insertion is to embed artificial information with known characteristics such as location, volume, shape and activity into pre-acquired data using accurate system modeling in order to link realism and ground truth [10, 11]. It brings the possibility to generate a countless number of different datasets from a single sample and grant a control on the input given to the acquisition system, i.e. the inserted information [9-11]. Hence, the method allows to estimate the effects of different configurations on the reconstructed images while keeping a standard of truth. In this work, we described and evaluated the method applied to the in-depth performance evaluation of a PET system. We compared qualitatively and quantitatively results obtained from our method to equivalent experimental data.

Methods

PET-CT system

All the experiments were performed on the Discovery MI 5-ring Positron Emission Tomography - Computed Tomography (PET-CT) digital system (General Electric Healthcare, Chicago, IL, USA). This PET-CT device is part of the most technologically advanced product line, combining time-of-flight (TOF), low resolution and high sensitivity that improve overall image quality by reducing the noise in the reconstructed images and enhance lesion detection [12]. There are several reconstruction algorithms available to produce images from the acquired raw data such as the common ordered subset expectation maximization (OSEM). A most recent, Bayesian penalized likelihood (BPL) algorithm, gives access to a regularization parameter β that allows to reduce image noise through each iteration [13, 14]. Results from the NEMA NU2-2012 standard performance tests for this configuration of the device have been published [15]. The evaluation of its performances and its benchmark among other models is important but is not sufficiently discriminating compared to the opportunities that could offer this high performance PET-CT. We performed several experiments based on the image quality (IQ)

performance standard in which we adapted the set of fillable spheres for challenging smaller sizes.

Phantoms experiments

Based on the NEMA IQ NU-2 2018 procedure [2], we used the NEMA IEC body phantom with the lung insert. We replaced the standard fillable spheres by another set with smaller internal diameters of respectively 4, 5, 6, 8, 10 and 13 mm (Data Spectrum Corporation, Durham, NC, USA), which central section in the phantom is showed in Figure 1.

Figure 1: Central section of the spheres from CT images.

As required by the standard, a scatter phantom was placed outside the scanner field of view. The filling of the phantoms, their positioning and their acquisitions on the examination bed were carried out according to the standard recommendation. Five distinct experiments were performed corresponding to five concentrations of ^{18}F leading to five SBR of approximatively 2:1, 4:1, 6:1, 8:1 and finally 0:1 (water only within the spheres) as a baseline for data insertion. The radioactive concentration of the background was nearly the same for all the experiments to finally get rather the same global activity within the body phantom. Details of the filling level in the different compartments of the scanned phantom are available in Table 1.

Table 1: Activity concentration (kBq/mL) within the background and spheres of the body phantom for each series.

	Activity concentration (AC) (kBq/mL)		
	Background	Spheres	Resulting SBR
Experiment SBR 2:1	5.32	11.03	2.07
Experiment SBR 4:1	5.44	21.38	3.93
Experiment SBR 6:1	5.50	33.19	6.03
Experiment SBR 8:1	5.32	42.42	7.97
Experiment SBR 0:1	5.52	0.00	0.00

We chose to conduct this study with OSEM and BPL algorithms because the former is part of the standard procedure and the latter is used in the clinical routine of our institution. It allows to compare the impact of the reconstruction algorithm using our method. Images were obtained using acquisition and reconstruction parameters detailed below in Table 2.

Table 2: Acquisition and reconstruction parameters

Acquisitions durations (s):	323 / 334 / 346
Matrix size:	384 x 384
FOV (mm):	400
Slice Thickness (mm):	2.8
Voxel dimensions (mm):	1.042 x 1.042 x 2.8
Voxel volume (mm ³):	3.04
Algorithm 1	
Reconstruction algorithm:	OSEM
TOF:	Yes
Iterations:	4
Subsets:	34
Filter (FWHM):	2
Z-filter:	None
Corrections:	Attenuation, scatter, randoms
Point spread function (PSF) modeling:	No
Algorithm 2	
Reconstruction algorithm:	BPL
TOF:	Yes
Beta (β):	20
Corrections:	Attenuation, scatter, randoms
Point spread function (PSF) modeling:	Yes

1 Data insertion method

2 We aimed to simulate radioactivity upon spheres filled with water (SBR 0:1) using the
3 background as an activity reference in order to reproduce the SBR of other experiments (2.07:1,
4 3.93:1, 6.03:1 and 7.97:1). First, we model the spheres according to the internal volumes of the
5 real set using a coded function with MATLAB (The MathWorks Inc., Natick, MA, USA). Then,
6 we determine the spatial coordinates of the center of each physical sphere from the CT images
7 in order to precisely define the center of each artificial sphere. Lastly, we calculate the activity
8 concentration (AC) (Bq/mL) to be inserted inside them. To achieve that, we draw 12 Volumes-
9 Of-Interest (VOI) distributed in the phantom background. We extract two specific tags from
10 the DICOM images header: RescaleSlope and RescaleIntercept, which are used to convert
11 image intensities into activity concentrations. These tags have varying values depending on the
12 slice. Equation (1) represents this step for each image numbered S (Slice number).

$$13 \quad AC \text{ (Bq/mL)} = \text{DICOM Intensity (S)} * \text{RescaleSlope (S)} + \text{RescaleIntercept (S)} \quad (1)$$

14 Next, in equation (2), we calculate an averaged value of the AC within the VOI taking into
15 account all nv voxels that compose it.

$$16 \quad \text{VOI Mean AC (Bq/mL)} = \frac{1}{nv} \sum_{v=1}^{nv} (\text{VOI AC (v) (Bq/mL)}) \quad (2)$$

Finally, we average across all VOIs (nm number of measurements) in (3) to obtain an average activity value (Bq/mL) defined as the baseline for generating insertions directly related to the initial experiment.

$$\text{Baseline AC (Bq/mL)} = \frac{1}{nm} \sum_{m=1}^{nm} (\text{VOI Mean AC (m) (Bq/mL)}) \quad (3)$$

In order to simulate the four synthetic sets of images reproducing the already acquired experiments, we multiply the mask by the exact experimental value of each SBR, i.e. respectively 2.07, 3.93, 6.03 and 7.97 (4).

$$\text{Simulated AC (Bq/mL)} = \text{Baseline AC (Bq/mL)} * \text{SBR} \quad (4)$$

Once these preliminary steps have been completed, we proceed to the generation of modified raw data using specific functions provided by the manufacturer. It relies on the forward projection through the scanner model of the mask image into sinogram. During this process, scanner and phantoms effects (geometric efficiency, detector efficiency variations, resolution and attenuation) are applied using raw data and CT images from the original acquisition [10, 11]. Original and simulated sinograms are summed to obtain new raw data used for the reconstruction of the simulated datasets. Images are reconstructed using a dedicated high performance workstation coupled with an associated reconstruction research toolbox for offline reconstruction (Duetto v02.13, General Electric Healthcare, Chicago, IL, USA). Those reconstructions are numerically equivalent to the reconstruction processing available on the PET console. The insertion process is illustrated in Figure 2. Reconstructed DICOM images are finally uploaded to the interpretation console database for analysis.

Figure 2: Workflow of the data insertion process.

Images and data analysis

A statistical data analysis was conducted in order to assess the similarity between the simulated and the experimental data considered as the reference. We first assessed the visual equivalence of the synthetic images compared to the experimental results. For each set of acquisitions, we had three sets of images, equivalent in terms of total counts, which we averaged to obtain a single frame. Visual inspection was performed on the central slice allowing to get all the spheres in the same plane. Afterwards, we also performed a quantitative analysis using the image interpretation software PETVCAR® on the AWS server client console (General Electric Healthcare, Chicago, IL, USA). To perform this analysis, we placed for each set of images 6

spherical VOIs centered on each sphere. The volume of each VOI aimed to reproduce the real internal volume of the corresponding spheres and was identical for experimental and simulated series. This study focused on the maximum and mean activity concentrations (kBq/mL) within the spheres. As required by the NEMA standard, each experimental and simulated series consisted of three successive acquisitions which were analyzed individually, following the same measurements conditions, and then averaged to improve the reproducibility of the results. Finally, taking into account all ratios studied and all sphere sizes, we obtained 48 measurements of mean and maximum activity concentrations (kBq/mL), AC_{mean} and AC_{max} , for both experimental and simulated series (6 spheres x 2 algorithms x 4 SBR). We then calculated the mean and maximum recovery coefficient, RC_{mean} and RC_{max} , dividing AC_{mean} and AC_{max} by the theoretical activity concentrations deducted from phantom preparation activity measurements (5).

$$RC_{mean/max} = \frac{AC_{mean/max} \text{ (kBq/mL)}}{AC_{theoretical} \text{ (kBq/mL)}} \quad (5)$$

As the background activity may vary from one experiment to another, even if the SBR is rather the same, we performed this calculation in order to normalize the different datasets and avoid discrepancies between reiterations (Table 1). Simulated and experimental averaged recovery coefficient (RC) are compared considering entire datasets and reconstruction algorithms datasets. First, we realized a non-parametric Mann-Whitney test to highlight any differences in RC_{mean} and RC_{max} distributions between the experimental and simulated data. Then, we performed a Spearman test [16] to estimate their crossed correlation. Moreover, we verified the normality of the distribution of differences between experimental and simulated quantitative data using D'Agostino-Pearson test [17], in order to further refine the statistical analysis creating Bland-Altman plots [18, 19]. These tests were carried out on all the experimental and simulated data and subsequently repeated for each algorithm. The criterion for the significant difference was $p < 0.05$ for the Mann-Whitney and D'Agostino-Pearson tests. A strong correlation was identified by $p < 0.05$ and $\rho > 0.8$.

Results

Visual comparison

As expected, due to the creation process of the simulated images, the visual aspect of the background in both experimental and inserted data was similar. Considering the spheres, the same objects could be visualized in both series. However, we could spot slight differences at

the PET performance limits for low contrast and small targets, such as the 8 mm sphere at SBR 2:1 and 5 mm at SBR 4:1. This observation was reported for both OSEM and BPL algorithms. An overview of the images is available in Figure 3 and 4, respectively for OSEM and BPL reconstruction.

Figure 3: Visual comparison of OSEM averaged reconstructed images: (a) Experimental series (from top to bottom respectively SBR 2:1, 4:1, 6:1 and 8:1). (b) Simulated series (from top to bottom respectively SBR 2:1, 4:1, 6:1 and 8:1).

Figure 4: Visual comparison of BPL averaged reconstructed images: (a) Experimental series (from top to bottom respectively SBR 2:1, 4:1, 6:1 and 8:1). (b) Simulated series (from top to bottom respectively SBR 2:1, 4:1, 6:1 and 8:1).

RC comparison

The results provided by the quantitative analysis for both experimental and inserted data are expressed as average and standard deviation (SD) for each SBR (column) and sphere size (row) in Tables 3 and 4.

Table 3: Experimental and simulated RC values from OSEM algorithm for each SBR and sphere size.

Table 4: Experimental and simulated RC values from BPL algorithm for each SBR and sphere size.

We highlights RC_{mean} relative errors inferior to 20 % for all configurations, and rather the same for RC_{max} except for 6 mm sphere at SBR 2:1 (OSEM), which is not visible on the image and giving a relative error of 39 %. Otherwise, neglecting this extreme value, the RC_{max} relative error was within 23 %. Considering the impact of the reconstruction algorithm, OSEM algorithm gave the largest relative error differences for both RC_{mean} and RC_{max} .

A graphical representation of these data is available in Figures 5 (OSEM) and 6 (BPL), which indicates similar trends for the variation of RC_{mean} and RC_{max} versus sphere size for both artificial and physical spheres whatever the SBR.

Figure 5: Experimental and simulated RC curves using OSEM algorithm as a function of sphere size: RC_{mean} and RC_{max} curves are arranged in two columns (left and right respectively) and SBR ordered in rows (from top to bottom: 2:1, 4:1, 6:1 and 8:1).

Figure 6 Experimental and simulated RC curves using BPL algorithm as a function of sphere size: RC_{mean} and RC_{max} curves are arranged in two columns (left and right respectively) and SBR ordered in rows (from top to bottom: 2:1, 4:1, 6:1 and 8:1)

The Mann-Whitney test showed no statistical differences between experimental and simulated RC values (both RC_{mean} and RC_{max}). We found the same result comparing the impact of the reconstruction algorithm on RC values, even if P value for the BPL algorithm was higher. Spearman tests revealed a strong correlation for all datasets ($\rho > 0.950$). From Bland-Altman plots, we determined the mean differences and agreement intervals (within 95 % of the differences). Considering all data, we obtained results for RC_{mean} and RC_{max} of 2.1 ± 16.9 % and 3.3 ± 22.3 % respectively. This analysis showed a small difference between OSEM and BPL for both RC_{mean} (respectively 4.4 ± 14.5 % vs -0.32 ± 18.1 %) and RC_{max} (5.9 ± 22.2 % vs 0.77 ± 21.6 %). All the results of the statistical analysis are presented in Table 5. Figure 7 and 8 show the correlation curves and Bland-Altman plots for RC_{mean} and RC_{max} respectively.

	Mann-Whitney U test (P value)	Spearman correlation test (ρ , P value)	Bland-Altman plot ($d \pm 1.96SD$ (%))
RC_{mean}	0.611	0.974, P value < 0.001	2.1 ± 16.9 %
OSEM RC_{mean}	0.599	0.979, P value < 0.001	4.4 ± 14.5 %
BPL RC_{mean}	0.650	0.960, P value < 0.001	-0.32 ± 18.1 %
RC_{max}	0.720	0.974, P value < 0.001	3.3 ± 22.3 %
OSEM RC_{max}	0.578	0.977, P value < 0.001	5.9 ± 22.2 %
BPL RC_{max}	0.853	0.967, P value < 0.001	0.77 ± 21.6 %

Table 5: Results from the statistical tests performed on RC values for overall and individual algorithm data.

Figure 7: Correlation curve (left) and Bland-Altman plot (right) of RC_{mean} data: Different configurations are ordered in rows (from top to bottom: whole RC_{mean} , OSEM RC_{mean} and BPL RC_{mean})

Figure 8: Correlation curves (left) and Bland-Altman plots (right) of RC_{max} data: Different configurations are ordered in rows (from top to bottom: whole RC_{max} , OSEM RC_{max} and BPL RC_{max})

Discussion

In nuclear medicine, physical phantom experiments are the gold standard for evaluating system performances in order to forecast outcomes in clinical practice. However, replication of these tests is limited in terms of reproducibility and repeatability, due to the introduction of uncertainties during preparation and acquisition. Hence, similar experiments repeated at different times show variations, which can impact their results [4]. An example of such variations can be observed in the measurements of activity concentrations (kBq/mL) inside the compartments of the phantom related to the experimental series in Table 1. Alternatively, some studies focused on the generation of artificial data for the assessment of imaging systems. A

widely used method is Monte-Carlo simulation [7], but it requires to model the system and the objects investigated and this option is computationally demanding. Other studies have focused on developing tools to shorten data generation times and adaptable to every PET-CT system [8, 9]. A drawback related to the generic aspect of this last solution is the need of a calibration step required to normalize the data, which reflects empirical status of the estimation made on the acquisition and reconstruction processes to mimic the system response.

In order to overcome these limitations, we designed and evaluated a method combining physical experiment and data insertion that aims to test multiple configurations for the performance evaluation of PET-CT devices, while avoiding experimental reiterations. We used the method to reproduce from one test different experimental scenarios we had conducted. We then focused our analysis on the realism of the synthesized objects through a visual and quantitative comparison with acquired experimental data.

Based on real images and using accurate model of the scanner, the method generates images whose visual rendering and visualization of objects is similar to experimental images. Discrepancies observed in the visual comparison occur for target presenting size and contrast challenging in terms of detection for the device. It is the limit of the method which shows slight deviation from the physical images when the limits in performance of the system are reached. Nevertheless, given this specificity and weakness of these differences, we considered them negligible for our study and assessed the equivalence between the two data sets. From the quantitative analyses, we were able to verify that the experimental and artificial data were comparable, correlated, with differences normally distributed. Both algorithms showed no statistical differences and close results in terms of correlations and limits of agreement. Although, we can highlight higher mean difference for OSEM RC_{mean} and RC_{max} which is consistent with the noise reduction linked to the regularization algorithm. Indeed, BPL applies activity-dependent smoothing and suppresses image noise in low-activity regions [11], in our case, the cold spheres from the original images, which represent the insertions location. Hence, it results in a deviation for the mean difference between BPL (less than 1 %) compared to OSEM (within 6 %) considering RC_{mean} and RC_{max} , which can thus be explained by the study setting and the choice to have no activity inside the spheres for the original acquisition.

In this study, we demonstrated the reliability of the method applied under specific experimental conditions through the insertion of artificial spheres and their comparison with equivalent experimental data defined as reference. We showed that we can simulate close visual and quantitative results compared to real data even under challenging situations such as small and

low contrast targets. Based on computer programs developed by the PET-CT manufacturer, the method uses the same processes available on the physical systems. In comparison to some simulation studies [7-9], the processes rely on a real acquisition performed on the system. Hence, files generated during the initial acquisition and reconstruction are used to generate new datasets inserting virtual information to get finally a realistic render of the exam. These synthetic datasets are useful for qualitative and quantitative assessment of system performance as they combine real backgrounds with inserted objects of known size, activity and location. From a single experiment, it allows to generate as many configurations as needed without requiring access to the scanner, which may be limited in terms of device and radiotracer availability. In addition, it can be applied directly to clinical data in order to evaluate impacts of acquisition and reconstruction parameter on patient examination [10, 11]. We intend to employ this method to support the physical and clinical evaluation phase of a new PET-CT device as part of the collaborative research partnership with the manufacturer.

One of underlying assumptions in the manufacturer functions implied in the method is that inserted information has a negligible impact on the scatter and random coincidences in the resulting sinogram (simulated + original). Hence, the modified data include only the original random and scattered coincidences. In this study, given the sizes and activities present in the spheres, we assume that their impacts are negligible.

Conclusion

We developed and validated a method allowing the generation of virtual realistic objects in phantom images. The insertions can be fully controlled and provide opportunities to evaluate medical imaging functions and image processing techniques.

In the context of a collaborative research partnership, this study is a first step in the method development for the performance evaluation of the next generation of PET systems. It will then be extended to more complex phantom models and clinical data.

LIST OF ABBREVIATIONS:

- PET:** Positron Emission Tomography
- CT:** Computed Tomography
- NEMA:** National Electrical Manufacturers Association
- IEC:** International Electrotechnical Commission
- OSEM:** Ordered Subset Expectation Maximization
- BPL:** Bayesian Penalized Likelihood
- SBR:** Sphere-to-Background Ratio
- TOF:** Time-Of-Flight
- FOV:** Field Of View
- IQ:** Image Quality
- ^{18}F :** Fluorine-18
- PSF:** Point Spread Function
- VOI:** Volume-Of-Interest
- AC:** Activity Concentration
- AC_{mean}:** Mean Activity Concentration
- AC_{max}:** Maximum Activity Concentration
- RC:** Recovery Coefficient
- RC_{mean}:** Mean Recovery Coefficient
- RC_{max}:** Maximum Recovery Coefficient
- SD:** Standard Deviation

Declarations

Ethics approval and consent to participate: Not applicable.

Consent for publication: Not applicable.

Availability of data and material: The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Competing interests:

Author Q. Maronnier declares that he has no conflict of interest.

Author F. Courbon declares that he has no conflict of interest.

Author O. Caselles declares that he has no conflict of interest.

Funding: This study was supported by the provision of free IT equipment by General Electric Healthcare.

Authors' contributions: QM, OC wrote the manuscript. QM produced, analyzed and interpreted the data. OC and FC reviewed the manuscript. All authors read and approved the final manuscript.

Acknowledgments: We wish to express our gratitude to General Electric Healthcare, in particular, Michael Spohn, Kuan-Hao Su and Moshe Levy for their support and availability during this study.

References :

1. Jones T, Townsend D. History and future technical innovation in positron emission tomography. J Med Imaging (Bellingham) [Internet]. 2017 Jan [cited 2021 Mar 28];4(1). Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5374360/>
2. Performance Measurements of Positron Emission Tomographs (PET) [Internet]. NEMA. [cited 2021 Mar 28]. Available from: <https://www.nema.org/standards/view/Performance-Measurements-of-Positron-Emission-Tomographs>
3. Ilan E, Deller T, Kjellberg F, Peterson W, Lubberink M. Performance comparison of three commercially available PET systems : SIGNA PET/MR, Discovery IQ and Discovery MI. Journal of Nuclear Medicine [Internet]. 2017 [cited 2021 Oct 6];58(S1). Available from: <http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-333334>
4. Vallot D, De Ponti E, Morzenti S, Gramek A, Pieczonka A, Llompert GR, et al. Evaluation of PET quantitation accuracy among multiple discovery IQ PET/CT systems via NEMA image quality test. EJNMMI Phys [Internet]. 2020 May 12 [cited 2021 Mar 28];7. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7218035/>
5. van der Vos CS, Koopman D, Rijnsdorp S, Arends AJ, Boellaard R, van Dalen JA, et al. Quantification, improvement, and harmonization of small lesion detection with state-of-the-art PET. Eur J Nucl Med Mol Imaging. 2017 Aug;44(Suppl 1):4–16.
6. Miwa K, Wagatsuma K, Nemoto R, Masubuchi M, Kamitaka Y, Yamao T, et al. Detection of sub-centimeter lesions using digital TOF-PET/CT system combined with Bayesian penalized likelihood reconstruction algorithm. Ann Nucl Med. 2020 Oct 1;34(10):762–71.
7. Sarrut D, Bała M, Bardiès M, Bert J, Chauvin M, Chatzipapas K, et al. Advanced Monte Carlo simulations of emission tomography imaging systems with GATE. Phys Med Biol. 2021 May 21;66(10):10TR03.
8. Pfaehler E, De Jong JR, Dierckx RAJO, van Velden FHP, Boellaard R. SMART (SiMulAtion and ReconsTruction) PET: an efficient PET simulation-reconstruction tool. EJNMMI Physics. 2018 Sep 18;5(1):16.
9. Berthon B, Häggström I, Apte A, Beattie BJ, Kirov AS, Humm JL, et al. PETSTEP: Generation of synthetic PET lesions for fast evaluation of segmentation methods. Physica Medica. 2015 Dec 1;31(8):969–80.
10. Gabrani-Juma H, Bimani ZA, Zuckier LS, Klein R. Development and validation of the Lesion Synthesis Toolbox and the Perception Study Tool for quantifying observer limits of detection of lesions in positron emission tomography. JMI. 2020 Apr;7(2):022412.
11. Wangerin KA, Ahn S, Wollenweber S, Ross SG, Kinahan PE, Manjeshwar RM. Evaluation of lesion detectability in positron emission tomography when using a convergent penalized likelihood image reconstruction method. J Med Imaging (Bellingham). 2017 Jan;4(1):011002.
12. Conti M, Bendriem B. The new opportunities for high time resolution clinical TOF PET. Clin Transl Imaging. 2019 Apr 1;7(2):139–47.

13. Teoh EJ, McGowan DR, Macpherson RE, Bradley KM, Gleeson FV. Phantom and clinical evaluation of the Bayesian penalized likelihood reconstruction algorithm Q.Clear on an LYSO PET/CT system. *J Nucl Med*. 2015 Sep;56(9):1447–52.
14. Vallot D, Caselles O, Chaltiel L, Fernandez A, Gabiache E, Dierickx L, et al. A clinical evaluation of the impact of the Bayesian penalized likelihood reconstruction algorithm on PET FDG metrics. *Nucl Med Commun*. 2017 Nov;38(11):979–84.
15. Pan T, Einstein SA, Kappadath SC, Grogg KS, Gomez CL, Alessio AM, et al. Performance evaluation of the 5-Ring GE Discovery MI PET/CT system using the national electrical manufacturers association NU 2-2012 Standard. *Med Phys*. 2019 Jul;46(7):3025–33.
16. Mukaka M. A guide to appropriate use of Correlation coefficient in medical research. *Malawi Med J*. 2012 Sep;24(3):69–71.
17. DagoSPtest Plot - File Exchange - MATLAB Central [Internet]. [cited 2021 Aug 24]. Available from: <https://fr.mathworks.com/matlabcentral/fileexchange/3954-dagospstest>
18. Giavarina D. Understanding Bland Altman analysis. *Biochem Med (Zagreb)*. 2015 Jun 5;25(2):141–51.
19. Bland-Altman and Correlation Plot - File Exchange - MATLAB Central [Internet]. [cited 2021 Aug 24]. Available from: <https://fr.mathworks.com/matlabcentral/fileexchange/45049-bland-altman-and-correlation-plot>

1 *Table 3: Experimental and simulated RC values from OSEM algorithm for each SBR and sphere*

	OSEM							
	Experimental series				Simulated series			
	SBR 2:1	SBR 4:1	SBR 6:1	SBR 8:1	SBR 2:1	SBR 4:1	SBR 6:1	SBR 8:1
Sphere	MEAN RECOVERY COEFFICIENT: Average $\pm$ SD							
13 mm	0.67 $\pm$ 0.04	0.65 $\pm$ 0.02	0.64 $\pm$ 0.01	0.60 $\pm$ 0.01	0.71 $\pm$ 0.03	0.66 $\pm$ 0.02	0.64 $\pm$ 0.01	0.63 $\pm$ 0.01
10 mm	0.65 $\pm$ 0.03	0.58 $\pm$ 0.02	0.57 $\pm$ 0.02	0.56 $\pm$ 0.01	0.69 $\pm$ 0.04	0.63 $\pm$ 0.02	0.61 $\pm$ 0.02	0.60 $\pm$ 0.01
8 mm	0.56 $\pm$ 0.04	0.45 $\pm$ 0.03	0.47 $\pm$ 0.01	0.45 $\pm$ 0.01	0.58 $\pm$ 0.04	0.53 $\pm$ 0.03	0.51 $\pm$ 0.03	0.50 $\pm$ 0.02
6 mm	0.43 $\pm$ 0.06	0.46 $\pm$ 0.01	0.40 $\pm$ 0.01	0.36 $\pm$ 0.03	0.49 $\pm$ 0.05	0.43 $\pm$ 0.03	0.40 $\pm$ 0.02	0.38 $\pm$ 0.02
5 mm	0.42 $\pm$ 0.08	0.32 $\pm$ 0.06	0.30 $\pm$ 0.02	0.29 $\pm$ 0.05	0.45 $\pm$ 0.06	0.35 $\pm$ 0.04	0.31 $\pm$ 0.02	0.30 $\pm$ 0.02
4 mm	0.42 $\pm$ 0.14	0.29 $\pm$ 0.03	0.32 $\pm$ 0.05	0.24 $\pm$ 0.05	0.40 $\pm$ 0.06	0.31 $\pm$ 0.06	0.26 $\pm$ 0.05	0.23 $\pm$ 0.04
Sphere	MAXIMUM RECOVERY COEFFICIENT: Average $\pm$ SD							
13 mm	1.67 $\pm$ 0.29	1.22 $\pm$ 0.02	1.12 $\pm$ 0.02	0.99 $\pm$ 0.03	1.49 $\pm$ 0.08	1.26 $\pm$ 0.04	1.16 $\pm$ 0.04	1.11 $\pm$ 0.03
10 mm	1.46 $\pm$ 0.24	1.11 $\pm$ 0.06	1.05 $\pm$ 0.11	0.93 $\pm$ 0.06	1.49 $\pm$ 0.17	1.27 $\pm$ 0.13	1.16 $\pm$ 0.11	1.10 $\pm$ 0.09
8 mm	1.11 $\pm$ 0.08	0.83 $\pm$ 0.15	0.83 $\pm$ 0.03	0.87 $\pm$ 0.06	1.08 $\pm$ 0.23	0.92 $\pm$ 0.03	0.84 $\pm$ 0.04	0.75 $\pm$ 0.10
6 mm	0.70 $\pm$ 0.04	0.79 $\pm$ 0.13	0.70 $\pm$ 0.13	0.63 $\pm$ 0.03	0.98 $\pm$ 0.07	0.81 $\pm$ 0.03	0.74 $\pm$ 0.04	0.65 $\pm$ 0.11
5 mm	0.67 $\pm$ 0.23	0.48 $\pm$ 0.08	0.47 $\pm$ 0.05	0.42 $\pm$ 0.10	0.72 $\pm$ 0.06	0.55 $\pm$ 0.05	0.49 $\pm$ 0.06	0.44 $\pm$ 0.09
4 mm	0.55 $\pm$ 0.14	0.38 $\pm$ 0.07	0.41 $\pm$ 0.11	0.30 $\pm$ 0.08	0.62 $\pm$ 0.16	0.42 $\pm$ 0.08	0.34 $\pm$ 0.07	0.31 $\pm$ 0.05

2 *Table 4: Experimental and simulated RC values from BPL algorithm for each SBR and sphere*

	BPL							
	Experimental series				Simulated series			
	SBR 2:1	SBR 4:1	SBR 6:1	SBR 8:1	SBR 2:1	SBR 4:1	SBR 6:1	SBR 8:1
Sphere	MEAN RECOVERY COEFFICIENT: Average $\pm$ SD							
13 mm	0.70 $\pm$ 0.04	0.73 $\pm$ 0.01	0.73 $\pm$ 0.01	0.70 $\pm$ 0.01	0.75 $\pm$ 0.04	0.73 $\pm$ 0.02	0.74 $\pm$ 0.01	0.74 $\pm$ 0.01
10 mm	0.69 $\pm$ 0.03	0.69 $\pm$ 0.01	0.71 $\pm$ 0.01	0.70 $\pm$ 0.01	0.73 $\pm$ 0.05	0.72 $\pm$ 0.02	0.74 $\pm$ 0.02	0.73 $\pm$ 0.01
8 mm	0.59 $\pm$ 0.05	0.54 $\pm$ 0.05	0.61 $\pm$ 0.02	0.59 $\pm$ 0.00	0.59 $\pm$ 0.03	0.61 $\pm$ 0.02	0.65 $\pm$ 0.03	0.65 $\pm$ 0.01
6 mm	0.44 $\pm$ 0.06	0.49 $\pm$ 0.02	0.51 $\pm$ 0.03	0.47 $\pm$ 0.08	0.44 $\pm$ 0.03	0.46 $\pm$ 0.03	0.51 $\pm$ 0.06	0.52 $\pm$ 0.02
5 mm	0.41 $\pm$ 0.09	0.34 $\pm$ 0.05	0.33 $\pm$ 0.04	0.35 $\pm$ 0.06	0.35 $\pm$ 0.07	0.28 $\pm$ 0.05	0.29 $\pm$ 0.03	0.30 $\pm$ 0.04
4 mm	0.40 $\pm$ 0.07	0.27 $\pm$ 0.03	0.31 $\pm$ 0.04	0.28 $\pm$ 0.05	0.40 $\pm$ 0.07	0.30 $\pm$ 0.05	0.27 $\pm$ 0.05	0.25 $\pm$ 0.04
Sphere	MAXIMUM RECOVERY COEFFICIENT: Average $\pm$ SD							
13 mm	1.65 $\pm$ 0.30	1.65 $\pm$ 0.38	1.37 $\pm$ 0.03	1.39 $\pm$ 0.12	1.78 $\pm$ 0.13	1.62 $\pm$ 0.06	1.51 $\pm$ 0.05	1.43 $\pm$ 0.07
10 mm	1.57 $\pm$ 0.31	1.62 $\pm$ 0.20	1.46 $\pm$ 0.05	1.40 $\pm$ 0.09	1.63 $\pm$ 0.21	1.77 $\pm$ 0.18	1.78 $\pm$ 0.19	1.59 $\pm$ 0.13
8 mm	1.06 $\pm$ 0.23	1.06 $\pm$ 0.30	1.28 $\pm$ 0.10	1.55 $\pm$ 0.42	1.05 $\pm$ 0.17	1.18 $\pm$ 0.27	1.46 $\pm$ 0.18	1.47 $\pm$ 0.06
6 mm	0.64 $\pm$ 0.12	0.85 $\pm$ 0.06	1.06 $\pm$ 0.28	1.02 $\pm$ 0.18	0.75 $\pm$ 0.04	0.80 $\pm$ 0.06	1.03 $\pm$ 0.08	1.11 $\pm$ 0.11
5 mm	0.54 $\pm$ 0.15	0.50 $\pm$ 0.01	0.45 $\pm$ 0.08	0.51 $\pm$ 0.06	0.49 $\pm$ 0.09	0.40 $\pm$ 0.05	0.42 $\pm$ 0.01	0.46 $\pm$ 0.02
4 mm	0.49 $\pm$ 0.06	0.40 $\pm$ 0.04	0.40 $\pm$ 0.07	0.36 $\pm$ 0.08	0.47 $\pm$ 0.12	0.37 $\pm$ 0.06	0.33 $\pm$ 0.07	0.31 $\pm$ 0.05

1 *Figure 1*

2
3
4
5
6
7
8
9
10
11
12
13
14
15
16

1 *Figure 2*

2

3

4

5

6

1 *Figure 3*

2

3

1 *Figure 4*

1 *Figure 5*

2

3

4

5

6

1 *Figure 6*

2

3

4

5

6

1 *Figure 7*

2

3

4

5

6

1 *Figure 8*

2

3

4