Facilitating Multi-Center Consistency in PET/CT Assessment: Transferability and Harmonization of 3D-Printed Phantom Inserts

Sai Kiran Kumar Nalla[1], Quentin Maronnier[3], Tala Palchan-Hazan[4], Olivier Caselles[2, 3*], John A. Kennedy[4, 5]

1. IUTC - Institut universitaire du cancer de Toulouse, Université de Toulouse, Toulouse, France
2. Tachyons SAS, Toulouse, France
3. Department of Medical Physics, Oncopole Claudius Regaud, IUCT Oncopole, Toulouse, France
4. Department of Nuclear Medicine, Rambam Health Care Campus, P.O.B. 9602, 3109601 Haifa, Israel
5. Ruth and Bruce Rappaport Faculty of Medicine, Technion Israel Institute of Technology, Haifa, Israel

***Correspondance :** Olivier Caselles (**caselles.olivier@iuct-oncopole.fr**)

ORCID IDS :

Sai Kiran Kumar Nalla : 0000-0002-8044-6330

Quentin Maronnier : 0000-0002-2266-9122

John A Kennedy**:** 0000-0002-2043-4397

Tala Palchan-Hazan: 0009-0004-2632-1168

Olivier Caselles: 0000-0002-5404-5066

NOTE: This preprint reports new research that is not certified by peer and should not be used to guide clinical practice.

Abstract:

Objective: Three-dimensional (3D) printing has been used to build cost effective & realistic phantoms in medical imaging but their adaptability across facilities remains under-explored. Verification of the applicability of standardization protocols and quantitative metrics of positron emission tomography (PET) to printed phantoms has been sparse. This study aims to evaluate the harmonization and transferability of one such 3D printed phantom design across two facilities.

Approach: Cubic phantom inserts (40 mm) featuring spherical targets (0, 10, 15, and 20 mm), of a previously proven design, were fabricated at two different facilities using fused deposition modeling (FDM) printers and positioned within a cylinder of a Jaszczak phantom, filled with radiotracer and soap/water solution. At both facilities, acquisitions (Day 0,8) were performed on 5-ring Discovery MI and 32 cm OMNI Legend systems. Quantitative metrics, including CT Hounsfield units (HU), PET recovery coefficients (RC) and PET target-to-background ratios (TBR) were statistically analysed (rank sum tests) to assess the influence of facility, scanner model, and acquisition day.

Main Results: CT images revealed minor differences across facilities and scanners, with notable HU improvement (in emulating water) upon prolonged cube immersion, from about −12 HU to about −6 HU. PET image quality was visually good with respect to phantom feature representation. Average TBR_{Max} (mean±std. dev.) were 1.56 ± 0.03 (expected 2.0), 2.35 ± 0.05 (expected 2.5), and 3.10 ± 0.05 (expected 3.33). RC_{Max} and RC_{Mean} ranged from 0.69 – 1.11 and 0.56 - 0.89. No statistical differences were found for TBR_{Max}, RC_{Mean}, or RC_{Max} ($p \geq 0.07$ for Mann Whitney; $p \geq 0.57$ for Friedman) and Bland-Altman mean bias < 10%, though Wilcoxon post-hoc tests identified a few outliers.

Significance: Our study demonstrates that a simple digitally shared design file allows for easily reproducible 3D-printed phantoms of varying radiotracer contrasts, with consistency in PET quantitative metrics

Keywords: Quantitative PET/CT, Rapid prototyping, Radionuclide imaging, Standardization, Quality Control, Additive manufacturing

1. Background:

Quantitative Measurements of molecular level biological processes are key tools for diagnosis of diverse diseases in clinical oncology **[1,2]**. Positron emission tomography (PET) in conjunction with computed tomography (CT) is a widely used modality due its ability to provide both structural and functional information at high quantitative accuracy **[3]**. Consistency and reliability of PET/CT systems is ensured by rigorous quality control (QC) standards provided by the National Electrical Manufacturers Association (NEMA) **[4]**, the European Federation of Organisations for Medical Physics (EFOMP) **[5]**, and the European Association of Nuclear Medicine Research 4 Life (EANM EARL) **[6]** by means of test phantoms. Such standards are used to facilitate harmonization and transferability across facilities. Harmonization is generally focused on standardising acquisition protocols, reconstruction algorithms and quantitative analysis methods ensuring minimal inter-scanner variability **[7–9]**. Whereas, transferability ensures development of robust phantoms and data reproducibility across facilities **[7,8,10].** Recent advancements in PET/CT systems have enabled more precise imaging especially in terms of spatial resolution **[11,12]**. Additionally, the rate of adoption of artificial intelligence (AI) into clinical workflows is rapidly increasing resulting in faster and accurate diagnostic interpretation **[13,14]**. Some other significant advances include the growing contribution of novel radiotracers and greater application of quantitative PET/CT for personalized medicine and theranostics **[15,16]**. While existing standards provide a foundation, this fast paced innovation calls for new phantom designs and capabilities to push the limits of available testing techniques. Furthermore, phantoms designed to simulate complex anatomical structures and varying levels of tracer uptake can help in building radiomics and artificial intelligence datasets. Three dimensional (3D) printing is emerging as a powerful tool to create customizable and anthropomorphic phantoms **[17]**. Furthermore, it allows precise control over geometry and material composition which are essential for transferability. Various 3D printing based studies have confirmed that both lesions and tissue uptakes can be modelled for scanner assessment. One such methodology is the usage of permeable grids which can be filled with radioactive tracer **[18,19]**. The internal structure of the grid based on the porosity level can modulate the activity concentration (AC) thus allowing for the creation of varying radiotracer contrast within a single phantom with single tracer injection **[20]**. Multiple studies in both PET and SPECT imaging confirmed the compatibility of such inserts with existing simulation methods **[18]**. Also, 3D printing has been used to create simulated lesions with heterogeneous uptake, aiding in the development of radiomic metrics to bring them one step closer to clinical implementation **[21]**. Additionally, anthropomorphic phantoms mimicking patient geometries that can aid in testing AI algorithms were also produced using this technology **[22,23]**. A key challenge though for adoption of 3D printing is the lack of multi-facility studies to illustrate the ability to rapidly and cost-effectively distribute 3D printed phantoms. More comprehensive studies especially from molecular imaging perspective are needed to establish harmonization protocols and data transferability. Material standardization is paramount,

requiring reproducible results across different printers especially the characterization of attenuation properties and radiotracer activity distribution. Geometric accuracy is equally critical, necessitating strict quality control during the printing process to avoid dimensional variations that can impact quantitative measurements. A bigger challenge though for widespread adoption of 3D-printing is lack of universally accepted metrics. The choice and variability in metrics can have significant impact on the interpretation of results especially in multi-center studies **[24]**. The adaptation of existing standardized uptake value (SUV) thresholds to 3D printed grids can be impractical due to contrast bias provided by the grids to the expected value **[20]**, thus other metrics are to be explored. For harmonization between sites, one can rely on each site to accurately prepare the test phantoms to the prescribed contrasts, or one can ship an active phantom from site to site **[23]**. The former method introduces variability by requiring several activity and/or volume measurements, whereas the latter requires long-lived radioisotopes and special shipping. Printed phantoms lend themselves to having the "same" phantom shipped digitally **[23]**, with pre-determined contrast features. In this study we perform a dual-facility investigation with phantom design specifications provided by **[20]**. The main goal was to evaluate both harmonization and transferability concurrently. Under the same design, phantom preparation and acquisition protocols we attempt to quantify the variability existing in PET/CT measurements between different days, scanners, and 3D printers. Moreover, we evaluate the reproducibility of measured quantitative parameters derived from standard image analysis such as standardized uptake value, recovery coefficients and target-to-background ratios to establish their utility for 3D printing based studies.

2. Methods:

2.1. 3D Printed Inserts Fabrication, Phantom Preparation and Image Acquisition Protocols:

Phantom design specifications were followed based on a recent study by **[20]**. Gyroid cubic inserts (40mm) designed in Autodesk Fusion 360 (*Autodesk Inc, San Francisco, CA, USA*) with variations in solidity S (50, 60, 70%) and pore distance D_P (i.e. unit cell distance: 5, 7, 10 mm) were used. Spherical targets of T (10, 15, 20 mm) were incorporated to assess quantitation and detectability. The total set contained 36 cubes of which 9 were homogeneous (No Target) and 27 with targets. Detailed design parameters, including rationales behind S, D_P, and T configurations, as well as the insert support structure design, are discussed in **[20]**. The primary objective of this study was to evaluate the transferability of 3D-printed inserts across different facilities using a single standardized design file. The inserts designed were exported in .STL file format which acted as the common shared design file. Both facilities used their own FDM printers: a Prusa MK3S+ Printer (Prusa Research, Prague, Czech Republic) for Facility –1 and a Stratasys F170 (Stratasys, Rehovot, Israel) for Facility – 2, with acrylonitrile butadiene styrene (ABS) as the material. For the phantom preparation (**Figure 1)**, specifications from same reference were used **[20]** with an adjustment to the injection protocol. The effective activity (**Table 1**) at scan time was maintained as the base criteria for standardisation between facilities. At both facilities, image acquisitions were

performed on a 5-ring Discovery MI followed by a 32 cm OMNI Legend (*GE Healthcare, Chicago, IL, USA*). Acquisition parameters were maintained across facilities for each scanner. A helical CT scan using three different slice thickness reconstructions were acquired for each scanner – 2.5mm CT (clinical), 1.25mm CT (high resolution) and a 3.5mm CT for the PET attenuation correction. The clinical CT was the main reconstruction presented in this study. All clinical CT's were acquired at 120 keV with a 512 × 512 matrix and a 50 cm transverse field of view (pixel spacing 0.98mm) for both scanners. PET acquisition parameters are presented in **Table 1**. All PET acquisitions were carried out across 5 consecutive frames. To equalize for different count rates between scanners, the scan duration for each OMNI acquisitions was scaled based on the effective decay and scan duration used for DMI. The exact scan durations are laid out in **Table 1**.

2.2 Data Extraction and Analysis:

Visual comparison of CT and PET Images was performed for qualitative differences. From Volume Viewer - PETVCAR® (*GE Healthcare, Chicago, IL, USA*), DICOM Images were exported for quantitative data extraction. Hounsfield Units (HU) for CT were extracted manually in 3D Slicer based on VOI workflow used in **[25]**. Quantitative PET data extraction was carried out in MATLAB (*MathWorks INC, MA, USA*) using a semi-automatic program developed in house in different regions of the image **Figure 1c,d,e** (Extended methods in supplementary material). Standardized uptake value (SUV), recovery coefficient (RC) & target-background ratio (TBR) were calculated from this data

$$\mathrm{SUV} = \frac{\text{Activity Concentration Inside ROI in } \frac{\mathrm{Bq}}{\mathrm{ml}}}{(\text{Injected Activity in Bq})/(\text{Weight of Phantom in g})} \quad (1)$$

$$\mathrm{TBR}_{(\max)} = \frac{\text{SUV Inside Target ROI (Max)}}{\text{SUV in the Background Mesh (Mean)}} \quad (2)$$

$$\mathrm{RC}_{(\mathrm{mean/max})} = \frac{\text{SUV Inside the Target ROI (Mean/Max)}}{\text{Theoretical SUV (Mean/Max)}} \quad (3)$$

Weight of the phantom utilized was 4.5 kg. Data extraction was carried out on each frame individually and then averaged over 5 frames for each acquisition set. This study investigated the influence of three key parameters on quantitative results: facility (printer), scanner, and acquisition day. We utilized a hierarchical grouping method to determine the effect of each parameter. For each set, first normality test was performed based on a Shapiro Wilk test followed by non-parametric Mann Whitney test for statistical differences in data distribution **[25]**. Additionally, Friedman tests **[20]** were also carried out to assess pairwise discrepancies within the datasets. Cross facility validation also followed all the aforementioned tests with the addition of a post hoc Wilcoxon sign rank test with Bonferroni adjustment to identify outliers **[26]**. All the statistical

tests were performed in JASP (*JASP Team, Netherlands*) and R Studio (*PBC, Boston, MA*) with significance level of 0.05 for the *p values* and ± 10% for the bias and variability.

3. Results:

3.1 CT HU Interpretation:

Visually minimal differences in CT across facilities and scanners can be seen in **Figure 2a**. The improvement in CT image quality based on prolonged immersion is also noticeable for all images. Day 8 results of Facility-1 for D_P: 5mm are better than Facility -2 with minimal target visibility. Notably the HU_{Min} value within the VOI, revealed the presence of small air bubbles (approximately 2 voxels in size) with very low HU values (-800 to -925) for cubes: 70-5-10 and 70-5-20 on Day 0 (Facility -2), which dissolved over immersion period by Day 8 (**Supplementary Information - Figure S2**). Both facilities demonstrated similar initial HU values on Day 0, with a slight increase observed at Facility 1 (-11.89 to -10.68) and a more pronounced increase at Facility 2 (-11.89 to -6.39) by Day 8. Statistical tests yielded p values < 0.001 for Shapiro Wilk establishing non-normality also as seen in **Figure 2b** and **2c**. Mann-Whitney central tendency tests resulted in statistically significant differences in HU_{Mean} based on facility (*p value – 0.002*), acquisition day (*p value < 0.001*) and no differences based on scanner (*p value – 0.87*).

3.2 PET Quantitation:

For the PET, all slices were able to capture the contrasts and targets visually as illustrated in **Figure 3a**. The target to background ratios and recovery coefficients are for all days, scanners can be found in **Figure 3b,c (Supplementary Information – Table S1)**. The coefficient of variability for noise observed in the image were all less than 5%. SUV variability along the phantom axis **Figure 4a** between layers **[20]** was well within threshold of 10% with maximum variability of 4.1% (Facility - 1) and 8.7% (Facility - 2).

3.3 Parametric statistical tests:

Statistical tests on parameters (day, facility and scanner) resulted in non – normality and no significance was observed for the 3 PET metrics (TBR_{max}, RC_{mean} and RC_{max}) based on any grouping parameter. Mann Whitney p-values were ≥ 0.07 for most cubes except for RC_{mean} of homogeneous cubes across scanners (~ 0.04). Additionally, parametric friedman pairwise analysis resulted p values of ≥ 0.72 (TBR_{Max}), ≥ 0.76 (RC_{Mean}) and ≥ 0.63 (RC_{Max}) for target cubes and ≥ 0.57 (TBR_{Max}), ≥ 0.07 (RC_{Mean}) and ≥ 0.16 (RC_{Max}) for homogeneous cubes. Data distribution of the grouped metrics are represented in **Figure 4b,c**. Both centres observed no statistical differences with all intra facility *p – values > 0.05* and intraclass correlation (ICC) values > 90%. Cross facility statistical results are summarised in **Table 2**. All the metrics have $p > 0.05$ thus no statistical differences for central tendency distribution Bland Altman mean bias fall within the 10% threshold but

the SD band width varies for each pair. Post hoc Wilcoxon revealed some outliers with $p < 0.05$ (**Table 3**) without alpha rate Bonferroni adjustment.

4. Discussion:

We evaluated the standardization and transferability of 3D-printed inserts based on the design specifications outlined by **[20]**. This study was carried out across two different facilities, where the shared elements prior to image acquisition were solely the design files (.STL) and the phantom preparation protocol. The quantitative metrics presented a comprehensive assessment of cross facility consistency and potential variability in both CT and PET imaging employing two different scanners (DMI and OMNI). Although transaxial CT slices (**Figure 2a**) showed no significant visual changes, with this regard day 8 for Facility-1 performed better over Facility-2. The change in HU_{Mean} values by day 8 from an identical starting point on Day 0 underscores significance of the facility, most likely due to the printer used. This significance is further quantified by *p-values* of 0.002 and 0.0001 observed between the different facilities and acquisition days. The statistical difference across acquisition days is expected and is inherent to the study design. While **[20]** suggested the phantom's immersion history as a key criterion for CT attenuation, our observations highlight its amplified importance in relation to the printer used. Even while using the same printer and settings, the quality of the 3D-prints, and consequently their attenuation properties can vary. Additionally, with wear of the printed objects over time, the immersion time needed also changes. This HU change normally stabilizes over a period **[20]** as seen in **Figure 2c** where the range of HU is similar for both facilities on day 8. Although minimal, the presence of small air bubbles within 2 targets on Day 0 for Facility-2 is undesirable. The importance of the chosen immersion strategy is emphasized again, as these bubbles effectively dissipate by Day 8. The lack of statistical significance in HU_{Mean} based on scanner is an encouraging finding. Within the framework of this study, both scanners provide comparable results thus reducing a potential source of variability when combining data acquired on different models. Additionally the range of HU observed in both facilities are much closer to water than previous 3D printing studies. Finally, for cross facility harmonization if equivalent HU_{Mean} is desired, the immersion period needs to be tailored based on the printer or filament manufacturer. Conversely, if the objective is primarily PET quantitation, the effective attenuation correction for 3D-printed grids from CT is sufficiently accurate across materials **[18,20]** as all the design considerations (D_P: 5, 7, 10mm) fall well below PET spatial resolution.

PET visual contrasts and target visibility are in agreement with the base phantom design **[20]** and indicated the fundamental PET functionality at both facilities. The noise levels within the image are also below the 5% threshold used in standard practice **[5,6]**. A key characteristic of this phantom design is that it uses background dilution i.e. the targets and background solution always have a $SUV_{theoretical} \sim 1$ which makes the uptake assessment easier. SUV variability along the phantom axis based on variability suggested uniform mixing for both facilities. This approach was adopted because the

acceptable threshold (10%) **[5]** needed to be consistently achieved regardless of layer location as this is crucial for accurate target assessment without bias. We adapted the PET preparation protocol to suggestions made by **[20]** by removing and adding back certain volume of the solution during injection and mixing. This change was implemented as we fixed the post injection decay time (45-55 minutes), whereas the original protocol didn't have such criteria. This reduced SUV variations in the background and increased reproducibility for the same injected activity. Additionally, adjusting the scan duration for the activity at scan time also played a crucial role in this regard and can help mitigating injection errors.

We observe that the of mean values of practically achieved TBR_{Max} by 3D printed grids (**Table 2**) are slightly lower than the true values, although they remain within the range defined by the standard deviation. Notably, standard deviation increases with increase in contrast because the percentage of activity inside the mesh of the inserts reduces. Such bias was also previously reported in grid studies **[18–20]**. However, a positive finding is that this bias is consistently captured across facilities and scanners (**Figure 4c and Table S1**). Hence, the reproducibility of the bias suggests the potential for a global correction/calibration factor(s) determined based on grid design parameters. Furthermore, we did not report the TBR_{Mean} for the same reason as it would be unreliable metric. Agreement in TBR_{Max} values suggested by statistical *p-values* and mean percentage of variability (< 10%) for different contrasts indicate the independence on sensitivity of these values with respect to acquisition day, facility and scanner. This is a critical to achieve quantitative PET harmonization and assessment of recovery coefficients. Building on this, consistency observed in RC_{Mean} and RC_{Max} suggest that not only the physical setup but also the quantitation can be readily reproduced across facilities by simply sharing the design file and adhering to a common preparation protocol. One outlier observed was comparison of RC_{Mean} for reference cubes across scanners which showed significance in Mann Whitney ($p \sim 0.04$) but no significance in Friedman ($p \sim 0.07$). Additionally, the pairwise tests for some intra facility data warrant some attention to few outliers. More specifically, pairwise results of Facility-2 showed significance especially pertaining to a few Day 0 metrics which can be understood in conjunction with presence of air bubbles within 2 targets. Moving to cross facility tests, global tests on the metrics both Mann Whitney and Friedman showed no statistical significance. Also the Bland-Altman mean bias fell within 10% threshold for all inter facility tests suggesting strong evidence for overall harmonization. TBR_{Max} was consistent even for pairwise Wilcoxon tests establishing the transferability of the phantom setup. In contrast, for recovery coefficients only 2 Wilcoxon tests resulted in acceptable *p values.* We filtered the metrics with a 10% absolute variability between pairs to quantify the outliers. In overall percentage, this was equivalent to 12% (26 outliers out of the 216 comparisons) of the RC_{Mean} data and 10% (21 outliers out of the 216) of the RC_{Max} data. The pair-wise *p -values* excluding outliers were all greater than 0.05.

Though we address the data extraction limitation of **[20]**, the fundamental size of the ROI is still an issue. The strategy of using physical dimensions to create the ROI is not an ideal because of printing inaccuracies. This becomes even more challenging if heterogeneity is added to the targets. Hence an automatic segmentation method could be better suited for data extraction when more complexity is added. While we acknowledge the influence of target size on metrics as shown in **Figure 3b,c**, a detailed statistical analysis in this regard was not performed. A key limitation for comparison across scanners were the count rates of DMI and OMNI were not the same since they were performed sequentially on same setup. This was due to the availability of scanners for experiments and also a different day for each device would make the CT results incomparable. Another limitation of this study is that consistency in the protocol with other radiotracers was not addressed.

The findings of this study emphasize the easy transferability of 3D printed phantoms across facilities with just a single standard design file. Consistency of PET metrics across facilities and scanner has been established while signifying differences in CT results. Utilisation of generic low-cost desk FDM printers further enhances the cost-effectiveness of this phantom design with average production cost less than 100 euros. With the capacity of 36 cubes per acquisition, this opens up the possibility of a wide spectrum of experiments with a single tracer injection. Moving forward, with further studies focusing on wide range of grid design parameters and printers the noise bias in grids can be globally calibrated. Furthermore, exploring post-print processing techniques can reduce and standardize the immersion time needed for CT attenuation stability if necessary. Alternative printing materials, such as PLA, could also lead to improved tissue-equivalent performance on CT **[18]**; however, the immersion vs HU trade off needs an evaluation. Additionally, a dedicated target detectability study, encompassing a broader range of target sizes (e.g., 5mm to 35mm), is needed to benchmark the performance of these 3D-printed phantoms against established standards. Also a multi-vendor assessment with blind evaluation is required further to find the bottlenecks in material standardisation. Finally, relevance of the metrics used in this study should also be investigated when heterogeneity is introduced within the target shape or activity distribution.

Conclusion:

Based on a dual facility study, transferability of 3D printed targets for PET/CT assessment of varying radiotracer contrasts solely based on the transfer of a standardised design file has been comprehensively established. We further demonstrated the compatibility of standard PET metrics with 3D printed phantoms and also laid out some modifications to statistical guidelines by a hierarchical approach for such phantoms. The key takeaway being that the PET results are invariant of differences observed in CT i.e. the attenuation correction works effectively resolving differences.

References:

1. Sokolenko AP, Imyanitov EN. Molecular Diagnostics in Clinical Oncology. Front Mol Biosci [Internet]. 2018 [cited 2025 Mar 31];5. Available from: https://www.frontiersin.org/journals/molecular-biosciences/articles/10.3389/fmolb.2018.00076/full

2. Das S, Dey MK, Devireddy R, Gartia MR. Biomarkers in Cancer Detection, Diagnosis, and Prognosis. Sensors. 2023;24:37.

3. Trotter J, Pantel AR, Teo B-KK, Escorcia FE, Li T, Pryma DA, et al. Positron Emission Tomography (PET)/Computed Tomography (CT) Imaging in Radiation Therapy Treatment Planning: A Review of PET Imaging Tracers and Methods to Incorporate PET/CT. Adv Radiat Oncol [Internet]. 2023 [cited 2025 Mar 31];8. Available from: https://www.advancesradonc.org/article/S2452-1094(23)00041-6/fulltext

4. Houda Baghous NE, Kafhali ME, Tahmasbi M, Chakir EM, Kessioui AE, Elkhatib A, et al. Evaluating long-term performance and quality control of the uMI 550 positron emission tomography- computed tomography (PET-CT) system: A comprehensive scientific analysis. Radiography. 2025;31:102920.

5. Matheoud R, Boellaard R, Pike L, Ptacek J, Reynés-Llompart G, Soret M, et al. EFOMP's protocol quality controls in PET/CT and PET/MR. Phys Med. 2023;105:102506.

6. Aide N, Lasnon C, Veit-Haibach P, Sera T, Sattler B, Boellaard R. EANM/EARL harmonization strategies in PET quantification: from daily practice to multicentre oncological studies. Eur J Nucl Med Mol Imaging. 2017;44:17–31.

7. Akamatsu G, Tsutsui Y, Daisaki H, Mitsumoto K, Baba S, Sasaki M. A review of harmonization strategies for quantitative PET. Ann Nucl Med. 2023;37:71–88.

8. Fahey F, Christian P, Zukotynski K, Sexton-Stallone B, Kiss C, Clarke B, et al. Use of a Qualification Phantom for PET Brain Imaging in a Multicenter Consortium: A Collaboration Between the Pediatric Brain Tumor Consortium and the SNMMI Clinical Trials Network. J Nucl Med. 2019;60:677–82.

9. Monsef A, Ay MR, Sheikhzadeh P, Geramifar P, Rahmim A, Ghafarian P. Harmonization based on quantitative analysis of standardized uptake value variations across PET/CT scanners: a multicenter phantom study. Nucl Med Commun. 2022;43:1004.

10. Multicenter CT phantoms public dataset for radiomics reproducibility tests - Kalendralis - 2019 - Medical Physics - Wiley Online Library [Internet]. [cited 2025 Mar 31]. Available from: https://aapm.onlinelibrary.wiley.com/doi/10.1002/mp.13385

11. PET/CT imaging in cancer: Current applications and future directions - Farwell - 2014 - Cancer - Wiley Online Library [Internet]. [cited 2025 Mar 31]. Available from: https://acsjournals.onlinelibrary.wiley.com/doi/10.1002/cncr.28860

12. Hussain D, Abbas N, Khan J. Recent Breakthroughs in PET-CT Multimodality Imaging: Innovations and Clinical Impact. Bioengineering. 2024;11:1213.

13. Wang Y-R (Joyce), Baratto L, Elizabeth Hawk K, Theruvath AJ, Pribnow A, Thakor AS, et al. Artificial Intelligence enables whole body Positron Emission Tomography Scans with minimal radiation exposure. Eur J Nucl Med Mol Imaging. 2021;48:2771–81.

14. Qi C, Wang S, Yu H, Zhang Y, Hu P, Tan H, et al. An artificial intelligence-driven image quality assessment system for whole-body [18F]FDG PET/CT. Eur J Nucl Med Mol Imaging. 2023;50:1318–28.

15. Dhoundiyal S, Srivastava S, Kumar S, Singh G, Ashique S, Pal R, et al. Radiopharmaceuticals: navigating the frontier of precision medicine and therapeutic innovation. Eur J Med Res. 2024;29:26.

16. Królicki L, Kunikowska J. Theranostics – present and future. Bio-Algorithms Med-Syst. 2021;17:213–20.

17. Filippou V, Tsoumpas C. Recent advances on the development of phantoms using 3D printing for imaging with CT, MRI, PET, SPECT, and ultrasound. Med Phys. 2018;45:e740–60.

18. Jessen L, Gustafsson J, Ljungberg M, Curkic-Kapidzic S, Imsirovic M, Sjögreen-Gleisner K. 3D printed non-uniform anthropomorphic phantoms for quantitative SPECT. EJNMMI Phys. 2024;11:8.

19. Theisen A-L, Lassmann M, Tran-Gia J. Toward a Patient-Specific Traceable Quantification of SPECT/CT-Based Radiopharmaceutical Distributions. J Nucl Med. 2022;63:1108.

20. Nalla SKK, Maronnier Q, Hazan TP, Kennedy J, Caselles O. Enhancing PET/CT Assessment with Porous 3D Printed Grids: A Pilot Study [Internet]. medRxiv; 2025 [cited 2025 Apr 2]. p. 2025.04.01.25325022. Available from: https://www.medrxiv.org/content/10.1101/2025.04.01.25325022v3

21. Gallivanone F, Interlenghi M, D'Ambrosio D, Trifirò G, Castiglioni I. Parameters Influencing PET Imaging Features: A Phantom Study with Irregular and Heterogeneous Synthetic Lesions. Contrast Media Mol Imaging. 2018;2018:5324517.

22. Gear JI, Cummings C, Craig AJ, Divoli A, Long CDC, Tapner M, et al. Abdo-Man: a 3D-printed anthropomorphic phantom for validating quantitative SIRT. EJNMMI Phys. 2016;3:17.

23. Jessen L, Curkic Kapidzic S, Gustafsson J, Larsson E, Ljungberg M, Sandström M, et al. A validation protocol for 177Lu-SPECT image quantification as a basis for multi-centre kidney dosimetry. Phys Med Biol. 2025;70:115001.

24. Fahey FH, Kinahan PE, Doot RK, Kocak M, Thurston H, Poussaint TY. Variability in PET quantitation within a multicenter consortium. Med Phys. 2010;37:3660–6.

25. Maronnier Q, Courbon F, Caselles O. Evaluation of a method based on synthetic data inserted into raw data prior to reconstruction for the assessment of PET scanners. EJNMMI Phys. 2022;9:68.

26. Graef J, Senger C, Wetz C, Baur ADJ, Kluge AK, Lukas M, et al. Shortened Tracer Uptake Time in GA-68-DOTATOC-PET of Meningiomas Does Not Impair Diagnostic Accuracy and PET Volume Definition. Diagnostics. 2020;10:1084.

Table 1. PET Acquisition and reconstruction parameters used by both facilities

Parameter	DMI	OMNI
No.of Bed Positions	1	1
Matrix Size	256 × 256	
Slice Thickness	2.8mm	2.07mm
No. of Slices	89	153
Scan Time	5 × 300 s	5 × 420 s
Reconstruction Diameter	600 cm	
Pixel Dimensions	[2.34 2.34] mm	
Voxel Volume	15.43 mm^3	11.33 mm^3
Reconstruction Algorithm	QCFX (time-of-flight BPL*)	QCHD (BPL*)
Beta	500	350
Injected Activity in MBq	Facility – 1: 51.0 (*Day 0*), 68.3 (*Day 8*) Facility – 2: 51.1 (*Day 0*), 49.9 (*Day 8*)	
Effective Activity at first scan in MBq	Facility – 1: 37.1 (*Day 0*), 37.1 (*Day 8*) Facility – 2: 37.7 (*Day 0*), 37.5 (*Day 8*)	

*Bayesian penalized-likelihood reconstruction with regularization parameter beta.

Table 2. Cross Facility Statistical Tests Summary for the PET Metrics of Targets

Comparison Pair	Noise (%)	TBR_{Max}			RC_{Max}			RC_{Mean}		
		Mann Whitney	**Wilcoxon Bonferroni (True, Adjusted)**	**Bland Altman**	**Mann Whitney**	**Wilcoxon Bonferroni (True, Adjusted)**	**Bland Altman**	**Mann Whitney**	**Wilcoxon Bonferroni (True, Adjusted)**	**Bland Altman**
DMI -1 vs DMI - 2	4.36 vs 3.60	1.00	0.16, 0.64	2.4 (21, -16)	0.86	0.81, 1	0.6 (11, -10)	0.55	0.05, 0.2	-1.5 (16, -13)
OMNI -1 vs OMNI - 2	4.25 vs 3.56	0.79	0.09, 0.36	-1.4 (11, -14)	0.47	0.03, 0.12	-1.9 (10, -13)	0.72	0.02, 0.08	-1.4 (13, -16)

DMI 1, OMNI 1 correspond to combined data of Facility -1 and DMI 2, OMNI 2 correspond to combined data of Facility – 2

Mann Whitney (N_{Sample} = 108) and Wilcoxon Sign Rank with Bonferroni Correction (N_{Sample} = 108 i.e. 54 pairs) represents the *p values*

Bland Altman represents % deviation (Mean Bias, ± 1.96 SD)

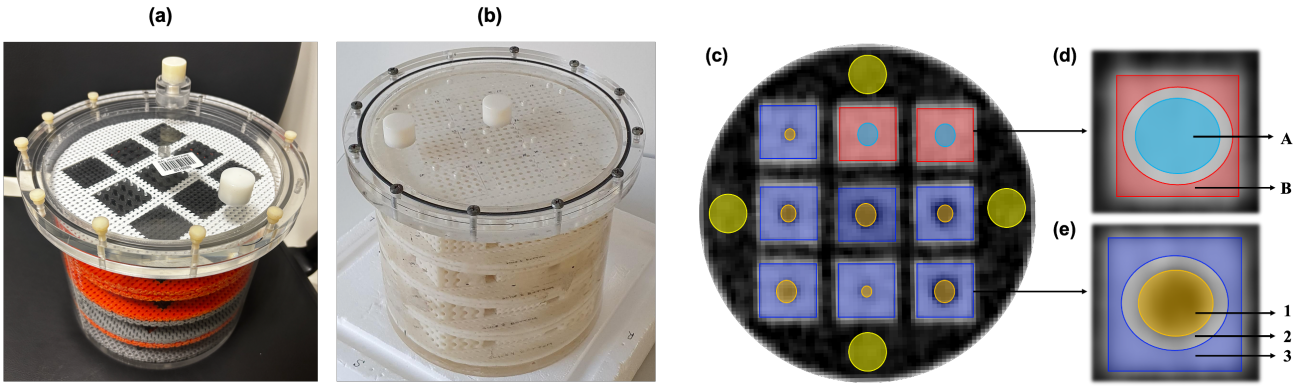

Figure 1. Jaszczak cylinder phantom with 36 cube inserts assembled in 4 layers prior to filling with a soap and distilled water solution – **(a)** Facility 1 , **(b)** Facility 2 – exact cube layout for each layer can be found in **[20]; (c)** ROI placement for each layer on transverse PET slice – reference solution (yellow), target (orange), homogeneous core (blue), background mesh homogeneous (red) and background mesh target cube (violet); **(d)** Exploded view for homogenous cubes – central core (region A), corner mesh (region B); **(e)** Exploded view for target cubes – target (region – 1), no measurement zone (region 2), background mesh (region 3).

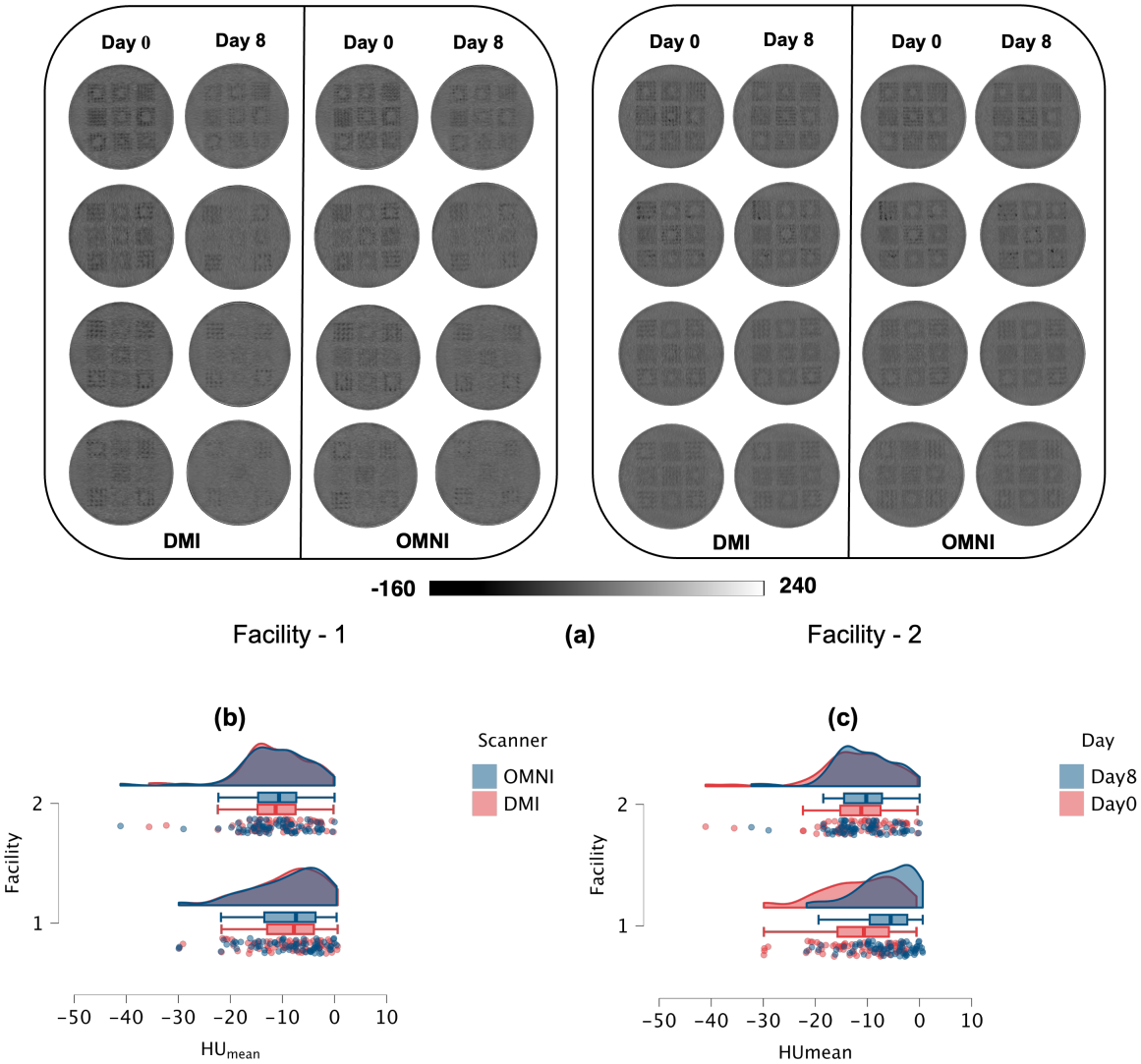

Figure 2. CT analysis. (a) Black areas are air. Transverse slice of CT for both scanners and facilities viewed on abdominal windowing (W: -40, L: 400; abdominal windowing); raincloud HU_{Mean} data distribution plots for both facilities separated as function of (b) Scanner; (c) Acquisition day

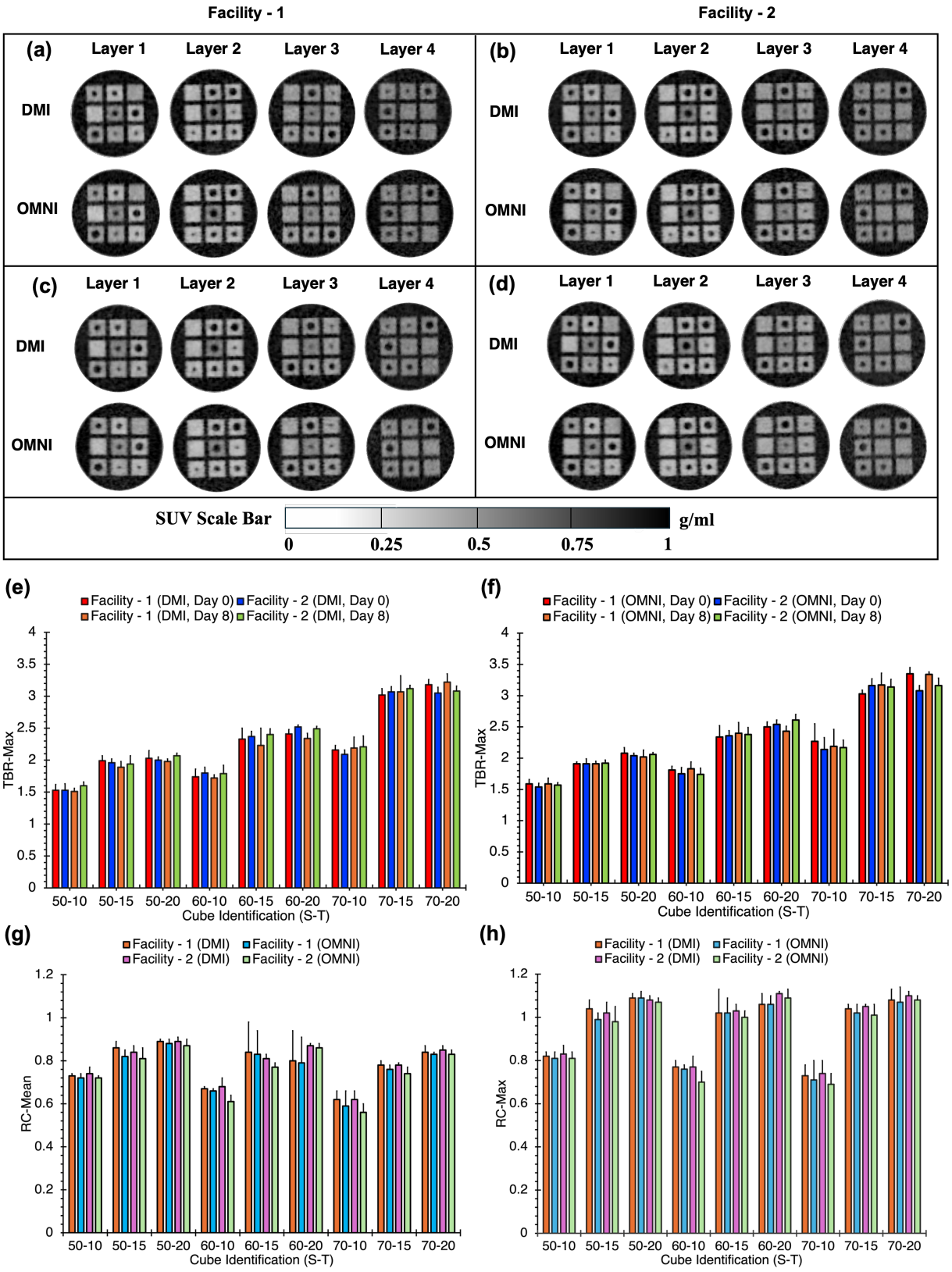

Figure 3. Central Axial Slice of PET for both scanners and facilities: **(a) & (b)** Day 0 and **(c) & (d)** Day 8 *(Note: For Layer -3, the 20mm (3 × 1) and 15mm (3 X 3) cubes of* D_P*:10 are misplaced between facilities);* TBR_{max} comparison between facilities (averaged w.r.t unit cell) as a function of S-T for **(e)** DMI and **(f)** OMNI; Recovery coefficient comparison across scanners and facilities for **(g)** RC_{mean} & **(h)** RC_{max}

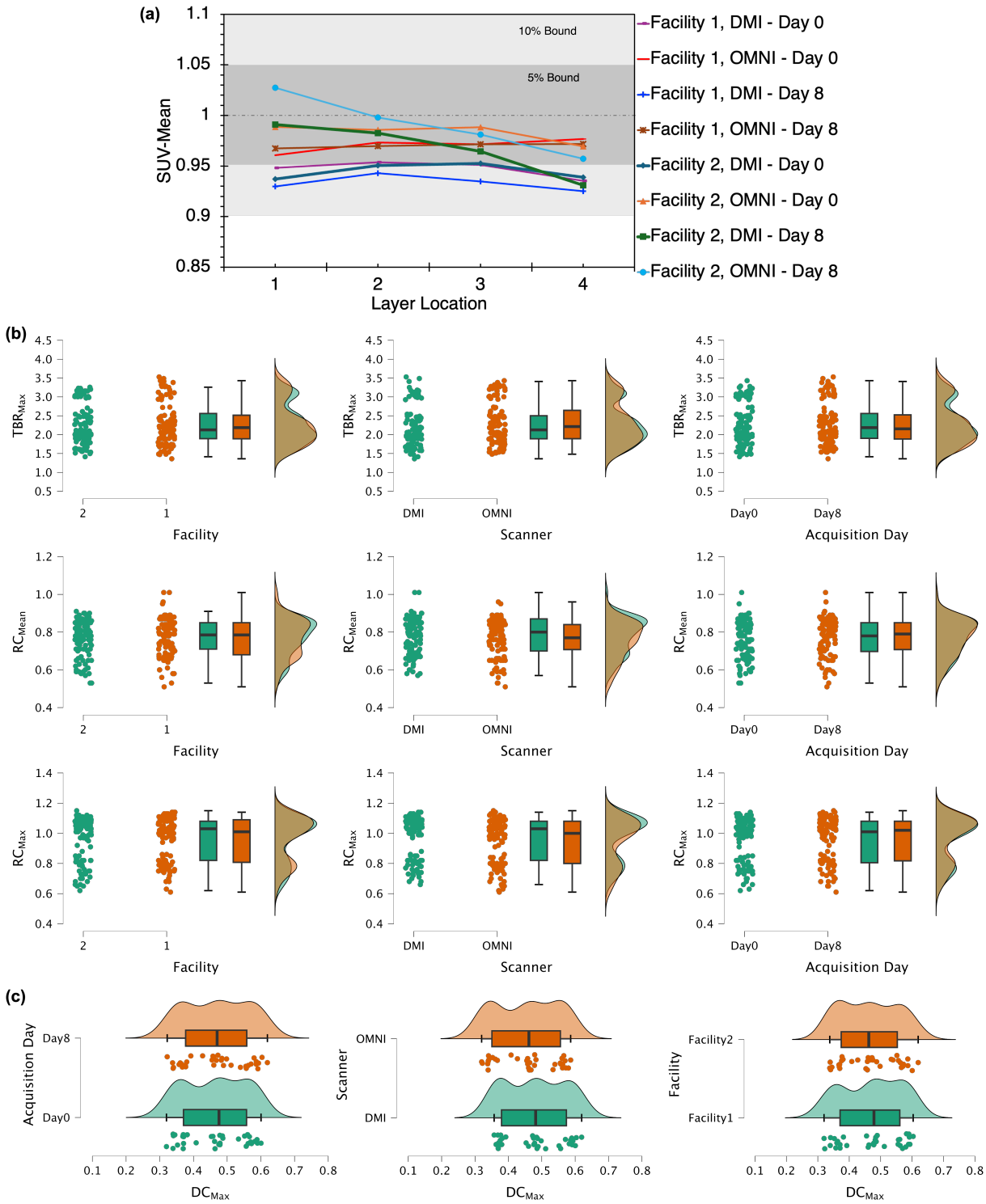

Figure 4. (a) SUV in reference solution across layers **(b)** Density Data *overlaid* for 3 quantitative PET metrics of Target Cubes – TBR_{Max}, RC_{Mean} and RC_{Max} (Rows) Data assorted together based on Facility, Scanner, and Acquisition Day (Columns); **(c)** Data Distribution Plots for homogeneous cubes –DC_{Max} , Data assorted together based on – Rows: Acquisition Day, Scanner, Facility. (*Expected clusters are centred on 0.3, 0.4, 0.5- A bias shift is observed in the acquired data*)

Supplementary Files:

Extended Methods:

Phantom preparation protocol - The cubes were placed in a cylinder from a Jaszczak phantom and filled with soap and water solution 4-6 hours prior to injection. Two studies corresponding to Day 0 and Day 8 were performed at each facility. Injection on each acquisition day was carried out 45-55 mins prior to scan. Tracer was administered by removing 10-50ml of background solution close to the phantom cap prior and was added back prior to scan.

Semiautomatic ROI data extraction - The best transverse slice for each layer was identified automatically for meticulous region of interest (ROI) placement. Exact ROI placement is detailed out in **Figure S1**. Target ROI size is based on converted nearest pixel size from corresponding theoretical dimensions. To mitigate spill-out artefacts and exclude the gradient region where partial volume effects are prominent, a second ROI dilated from target ROI was positioned, and a no measurement zone was established based on Boolean subtraction of the 2 ROI's. For non-target cubes, 20mm was the reference size for ROI. The extracted activity concentration (kBq/ml) were then converted to standardized uptake value (SUV) with phantom solution weight of 4.5 kg based on Equation 1, with the activities decay corrected to the acquisition start time.

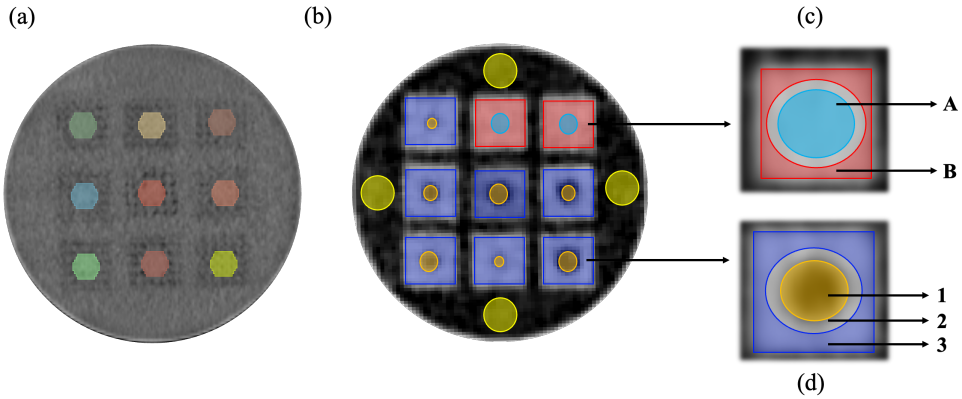

Figure S1. Phantom ROIs. CT: (a) 25 mm Spherical VOI Placement centred on axial plane for HU extraction; (b) ROI placement for each layer on transverse PET slice – background water (yellow), target (orange), homogeneous core (blue), background mesh non target (red) and background mesh target (violet); (c) Exploded view for non-target inserts – central core (region A), corner mesh (region B); (d) Exploded view for target cubes – target (region – 1), no measurement zone (region 2), background mesh (region 3).

$$\mathrm{SUV}_{(\mathrm{mean\ /\ max})} = \frac{\text{Activity Concentration Inside ROI in } \frac{\mathrm{Bq}}{\mathrm{ml}} \text{ (Mean/Max)}}{\text{(Injected Activity in Bq)/(Weight of Phantom in g)}} \quad (1)$$

SUV_{Mean} and SUV_{Max} were calculated for the entire data. Background SUV of the solution was averaged over 4 extracted values for each layer. Subsequently to quantify the contrast, target to background ratio (TBR_{Max} and TBR_{Mean}) were calculated using equation 2. Expected TBR values based on design parameters (S: 50, 60, 70) are 2, 2.5 and 3.33.

$$\mathrm{TBR}_{(\mathrm{mean/max})} = \frac{\text{SUV Inside Target ROI (Mean/Max)}}{\text{SUV in the Adjacent Background Mesh (Mean)}} \quad (2)$$

For target quantitation, as per standard practice, recovery coefficient (RC_{Max} and RC_{Mean}) were also calculated using Equation 3. Expected value of RC for targets is 1 and for homogeneous cubes is 0.5, 0.4, 0.3 for S: 50, 60, 70.

$$\mathrm{RC}_{(\mathrm{mean/max})} = \frac{\text{SUV Inside the Target ROI (Mean/Max)}}{\text{Theoretical SUV (Mean/Max)}} \quad (3)$$

Theoretical SUV is selected as actual measured value in reference solution rather than $SUV_{theoretical} \sim 1$, this allows relative assessment of each layer in case of mixing errors. Since all values fall in 10% of $SUV_{theoretical} \sim 1$, this should have minimal effect and improves reproducibility.

Extended Data:

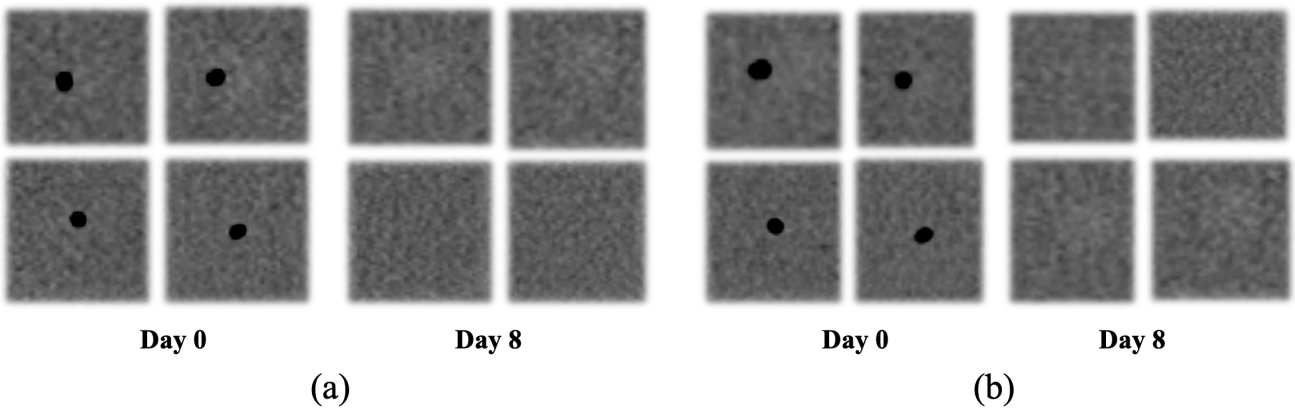

Figure S2. Snapshots of ROI with air bubbles on Day 0 and Day 8 for Facility 2

For DMI at Facility 1, HU value (mean ± Std.dev) was -11.65 ± 14 on Day 0 which changed to -10.8 ± 14 on Day 8. At Facility 2, the initial value was -11.89 ± 14, changing to -6.61 ± 14. OMNI also resulted in same range with the mean values at Facility 1 being -11.89 ± 14 (*Day 0*) and -10.68 ± 14 (*Day 8*) while at Facility 2 they were -11.61 ± 14 (*Day 0*) and -6.31 ± 14 (*Day 8*). For Facility 1, we observed the average noise in the image to be 4.19% for both scanners on Day 0 and 4.54 and 4.3% for DMI and OMNI respectively on Day 8. The average noise values for each scanner respectively at Facility 2 were 3.64 and 3.57% on Day 0, 3.45 and 3.56% on Day 8.

Table S1. PET Quantitative Metrics Calculated from Extracted Data

Insert Set (Solidity- Target Size)	TBR_{Max} Day 0 (Mean ± SD); Day 8 (Mean ± SD)			
	Facility - 1		Facility - 2	
	DMI	OMNI	DMI	OMNI
50-10	1.53 ± 0.09 1.51 ± 0.05	1.59 ± 0.07 1.59 ± 0.09	1.53 ± 0.10 1.60 ± 0.06	1.54 ± 0.06 1.57 ± 0.04
50-15	1.99 ± 0.08 1.89 ± 0.09	1.91 ± 0.03 1.91 ± 0.04	1.96 ± 0.06 1.94 ± 0.13	1.91 ± 0.08 1.92 ± 0.05
50-20	2.03 ± 0.12 1.98 ± 0.04	2.08 ± 0.09 2.02 ± 0.11	2.00 ± 0.05 2.07 ± 0.04	2.04 ± 0.04 2.06 ± 0.03
60-10	1.74 ± 0.12 1.72 ± 0.05	1.81 ± 0.06 1.83 ± 0.11	1.80 ± 0.09 1.79 ± 0.13	1.75 ± 0.10 1.74 ± 0.10
60-15	2.33 ± 0.17 2.23 ± 0.27	2.34 ± 0.18 2.40 ± 0.17	2.37 ± 0.08 2.40 ± 0.09	2.36 ± 0.08 2.38 ± 0.11
60-20	2.41 ± 0.07 2.34 ± 0.08	2.50 ± 0.08 2.43 ± 0.08	2.52 ± 0.03 2.49 ± 0.04	2.54 ± 0.07 2.61 ± 0.09
70-10	2.16 ± 0.07 2.19 ± 0.17	2.27 ± 0.28 2.19 ± 0.27	2.09 ± 0.07 2.21 ± 0.17	2.14 ± 0.19 2.17 ± 0.12
70-15	3.02 ± 0.10 3.07 ± 0.25	3.03 ± 0.06 3.17 ± 0.19	3.07 ± 0.08 3.12 ± 0.05	3.16 ± 0.11 3.14 ± 0.12
70-20	3.18 ± 0.08 3.22 ± 0.13	3.35 ± 0.10 3.34 ± 0.04	3.05 ± 0.09 3.08 ± 0.08	3.08 ± 0.08 3.16 ± 0.12
Insert Set (Solidity- Target Size)	**Averaged RC_{Max} (Mean ± SD); RC_{Mean} (Mean ± SD)**			
	Facility - 1		**Facility - 2**	
	DMI	**OMNI**	**DMI**	**OMNI**
50-10	0.82 ± 0.02 0.73 ± 0.01	0.81 ± 0.03 0.72 ± 0.02	0.83 ± 0.04 0.74 ± 0.03	0.81 ± 0.03 0.72 ± 0.01
50-15	1.04 ± 0.04 0.86 ± 0.03	0.99 ± 0.03 0.82 ± 0.03	1.02 ± 0.05 0.84 ± 0.03	0.98 ± 0.07 0.81 ± 0.05
50-20	1.09 ± 0.02 0.89 ± 0.01	1.09 ± 0.03 0.88 ± 0.02	1.08 ± 0.02 0.89 ± 0.02	1.07 ± 0.02 0.87 ± 0.03
60-10	0.77 ± 0.03 0.67 ± 0.01	0.76 ± 0.02 0.66 ± 0.01	0.77 ± 0.05 0.68 ± 0.04	0.70 ± 0.05 0.61 ± 0.03
60-15	1.02 ± 0.11 0.84 ± 0.14	1.02 ± 0.07 0.83 ± 0.11	1.03 ± 0.03 0.81 ± 0.02	1.00 ± 0.03 0.77 ± 0.02
60-20	1.06 ± 0.05 0.80 ± 0.14	1.06 ± 0.04 0.79 ± 0.12	1.11 ± 0.01 0.87 ± 0.01	1.09 ± 0.04 0.86 ± 0.02
70-10	0.73 ± 0.05 0.62 ± 0.04	0.71 ± 0.09 0.59 ± 0.07	0.74 ± 0.06 0.62 ± 0.04	0.69 ± 0.05 0.56 ± 0.04
70-15	1.04 ± 0.02 0.78 ± 0.02	1.02 ± 0.04 0.76 ± 0.02	1.05 ± 0.01 0.78 ± 0.01	1.01 ± 0.05 0.74 ± 0.03
70-20	1.08 ± 0.05 0.84 ± 0.03	1.07 ± 0.07 0.83 ± 0.01	1.10 ± 0.02 0.85 ± 0.02	1.08 ± 0.02 0.83 ± 0.02

Based on the values in **Table S1,** Cross facility TBR_{Max} (Mean ± SD) for 50% contrast cubes ($TBR_{Theoretical}$ ~ 2) were 1.82 ± 0.24 and 1.85 ± 0.23 for DMI. The values of same contrast for OMNI were observed to be 1.85 ± 0.21 and 1.84 ± 0.23. Inserts with a $TBR_{Theoretical}$ of 2.5 resulted in values (Facility 1, Facility 2) of 2.13 ± 0.31 and 2.23 ± 0.34 for DMI, 2.22 ± 0.32 and 2.23 ± 0.39 for OMNI. The 70% solidity cubes with $TBR_{Theoretcial}$ of 3.33 yielded DMI values of 2.8 ± 0.5 and 2.77 ± 0.48 while OMNI values were 2.89 ± 0.53 and 2.81 ± 0.5. RC_{Max} and RC_{Mean} were averaged for various insert sets with respect to target sizes for each facility and are presented in **Table S1**.