# Quantitative Reality: A Physical Ground Truth to Assess Heterogeneity Beyond Visual Limits in Molecular Imaging

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

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

***Correspondence :**

Olivier Caselles, PhD

Medical Physics Manager, Department of Nuclear Medicine

IUTC - Institut universitaire du cancer de Toulouse, Université De Toulouse,

1 Av. Irène Joliot-Curie, 31100 Toulouse

Email: Caselles.Olivier@iuct-oncopole.fr

ORCID : 0000-0002-5404-5066

Quentin Maronnier: **The work for this manuscript was conducted while affiliated with Oncopole Claudius Regaud, Toulouse, France**

**Conflict of Interest : NONE**

**Financial Statement : NONE**

**Data Availability :** "*Research data are stored in an institutional repository and will be shared upon request to the corresponding author*"

**Word Count**: 6097 (4295 Main Article – Introduction, Methods, Results, Discussion, Appendix)

Sai Kiran Kumar Nalla : 0000-0002-8044-6330
Quentin Maronnier : 0000-0002-2266-9122
John A Kennedy**:** 0000-0002-2043-4397

**Keywords:** Spatial heterogeneity, 3D Printing, Multi Uptake, PET Imaging, Radiomics

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

**Abstract:**

**Background:** Accurate quantification of intratumoral heterogeneity is essential for validating radiomics features and AI models in oncology. However, standard physical phantoms typically utilize homogeneous compartments, lacking the spatially complex uptake patterns required to rigorously benchmark these metrics against a known reality. To address this, we introduce porous 3D-printed inserts for simulating controlled spatial heterogeneity.

**Methods:** Modular grid inserts (17, 22, 28 mm targets) were developed with contrast of 4:1. Nine heterogeneity models, simulating progressions from uniform uptake to multifocal to necrosis, were imaged on a 32cm Omni Legend PET/CT system. Two different tracers: [$^{18}$F]FDG ([$^{18}$F]-2-fluoro-2-deoxy-d-glucose) and [$^{68}$Ga] (68-Gallium) were used on different days. Quantitative performance was assessed using conventional Standardized Uptake Values (SUV) and Target-to-Background Ratios (TBR), alongside the Gini Index (GI) to quantify spatial heterogeneity and radiotracer distribution inequality.

**Results:** The phantom demonstrated high consistency, with reference solution variability remaining under 2.2% across layers. In 28 mm inserts, experimental $TBR_{max}$ values fell within 10% of the theoretical 4:1 design ratio. GI successfully quantified structural complexity, ranging from 0.05 in homogeneous regions to 0.30 in necrotic targets. Cross-tracer analysis demonstrated robust reproducibility with a mean bias of <10%. While Partial Volume Effects reduced visual distinctness in 17 mm geometries, quantitative metrics successfully preserved the intended pathological trends.

**Conclusion**: This study establishes a foundational ground truth for heterogeneous uptake in PET. By providing defined physical standards validated by quantitative metrics, it enables the robust validation of segmentation algorithms and radiomics features even when spatial resolution limits visual assessment.

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

## 1. Introduction:

Radiotracer uptake is a defining characteristic of lesions in nuclear medicine playing a crucial role in quantification, feature extraction, and characterization [1–3]. In clinical practice, however, lesions rarely demonstrate uniform uptake; instead, they often display heterogeneous patterns such as necrotic cores, peripheral gradients, or multifocal regions [3–5]. Quantitative metrics derived from such patterns, including radiomic features, form the basis for advanced Artificial Intelligence (AI) models used in prognostication and personalized medicine. Crucially, in some tumours underlying heterogeneity can be present as low-contrast or visually indistinct patterns, leading to the risk of underestimating the true biological complexity. Hence, accurate quantification of this heterogeneity is essential for treatment response evaluation and multi-centre harmonization [6–8]. Physical phantoms that are widely used for performance testing and standardization of imaging [9,10] typically represent homogeneous activity distributions. Despite clinical importance, there remains a lack of practical tools for studying spatial heterogeneous uptake patterns under controlled and repeatable conditions.

In this context, Three dimensional (3D) printing offers a flexible platform for phantom development. Precise control of geometry and material composition were previously achieved for nuclear medicine and radiation therapy based studies [11–13]. Existing 3D printed phantom designs range from organ models derived from patient segmentations [14–16] to emulated lesions with controllable uniform uptake characteristics [15,17–21]. From the perspective of spatial heterogeneity, compartment-based phantoms have been created based on regions with different activity levels [18,22], but they require labour-intensive preparation, careful manual filling, and risk leakage or cross-contamination. Approaches that incorporate radioactivity directly into the printing resin have been explored as an alternative but only for uniform targets [23,24]. Among existing strategies, permeable grid-based designs are particularly promising [15,17,19]. These grids alternate between regions smaller than the scanner's spatial resolution, producing an averaged effect on the reconstructed image. This principle allows for fine control over tracer distribution using single injection without the need for complex compartmentalization or multiple manual filling steps. Such inserts have been proven to be conceptually effective with strong transferability and reproducibility across facilities [25,26]. However similar to direct radioactive printing, they were tested primarily to simulate uniform targets or organ-level phantoms. There is therefore a clear need for a simple, reproducible, and scalable approach to generate heterogeneous activity distributions while maintaining quantitative interpretability and potential for cross-site standardization.

In this work, we present a grid-based 3D-printed phantom inserts with controlled spatial heterogeneous uptake characteristics while preserving protocol simplicity and quantitative robustness. The proposed design incorporates homogeneous regions, "no-target" zones (background or non-avid tissue), necrotic areas (near zero uptake) and multifocal sub-regions representing varied activity. To assess consistency, a single phantom was imaged using two different

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

radiotracers on the same PET/CT (Positron Emission Tomography / Computed Tomography) system. The dual-tracer framework is geared to evaluate design robustness rather than quantitative cross-tracer performance analysis. With the help of conventional PET metrics, we further examine whether the design yields quantifiable heterogeneity, even in scenarios where such variation is not visually apparent.

## 2. Methods:

### 2.1 Base Phantom Design:

The basic design of phantom inserts is adapted from [19] in which all experiments were conducted within a standard Jaszczak cylinder phantom (total fill volume ≈ 5.5 L for this study). Each insert is based on a 40 mm porous cube in which spherical voids (17, 22, and 28 mm) were subtracted using Autodesk Fusion 360 (Autodesk Inc, San Francisco, CA, USA). These penetrable voids act as targets **(Fig.1a)**, allowing to be filled with the tracer solution and reach intended activity concentration (AC) following **Eq. 1, 3**. The **solidity** *S* (or infill percentage) of the background mesh **(Fig.1a)** represents the fraction of solid material and therefore controls tracer dilution. For a given background solidity ($S_{bkg}$), the effective tracer activity concentration within the mesh is given by **Eq.2.** A 3 × 3 array of the inserts as illustrated in **Fig.2** were placed concentrically within the cylinder phantom leading to 3 distinct regions as shown in **Fig 1**.

- Reference (Tracer uptake in the water outside of the inserts filled inside the phantom, used as reference to benchmark target uptake)

$$\mathrm{AC_{Ref}}\ (\mathrm{kBq/mL}) = \frac{\text{Injected Activity (kBq)}}{\text{Injected phantom volume (mL)}} \quad (1)$$

- Background mesh (Representing dimmer tissue emulation)

$$\mathrm{AC_{Bkg}}\ (\mathrm{kBq/mL}) = \frac{\left(100 - \mathrm{S_{bkg}}(\%)\right) * \text{ Injected Activity (kBq)}}{\text{Injected volume (mL)}} \quad (2)$$

- Targets (Uptake in the water within the voids - Representing hot lesion emulation)

$$\mathrm{AC_{Target}}\ (\mathrm{kBq/mL}) \approx \frac{\text{Injected Activity (kBq)}}{\text{Injected phantom volume (mL)}} \quad (3^*)$$

* '≈' is used since image derived value is determined by a range due to noise.

Target to background ratio (TBR) of these inserts with targets is defined using below **Eq.4.** Under ideal tracer accumulation within the void, $AC_{ref} \approx AC_{target}$ leading to $AC_{Bkg} \approx$ TBR x $AC_{ref}$ ~ TBR x $AC_{target}$. For this study, a 75% infill (solidity) was used for the background mesh. This choice was made to achieve a higher theoretical $TBR_{insert}$ ~ 4, pushing beyond the limits of previous grid-based studies [15,17,26], while ensuring sufficient range for creating heterogeneous patterns.

$$\mathrm{TBR_{Insert}} = \frac{100}{100 - \mathrm{S_{Bkg}}\ (\%)} \approx \frac{\mathrm{AC_{Target}} \text{ in kBq/mL}}{\mathrm{AC_{Bkg}} \text{ in kBq/mL}} \quad (4^*)$$

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

* '=' represents theoretical design value and '≈' represents image derived value.

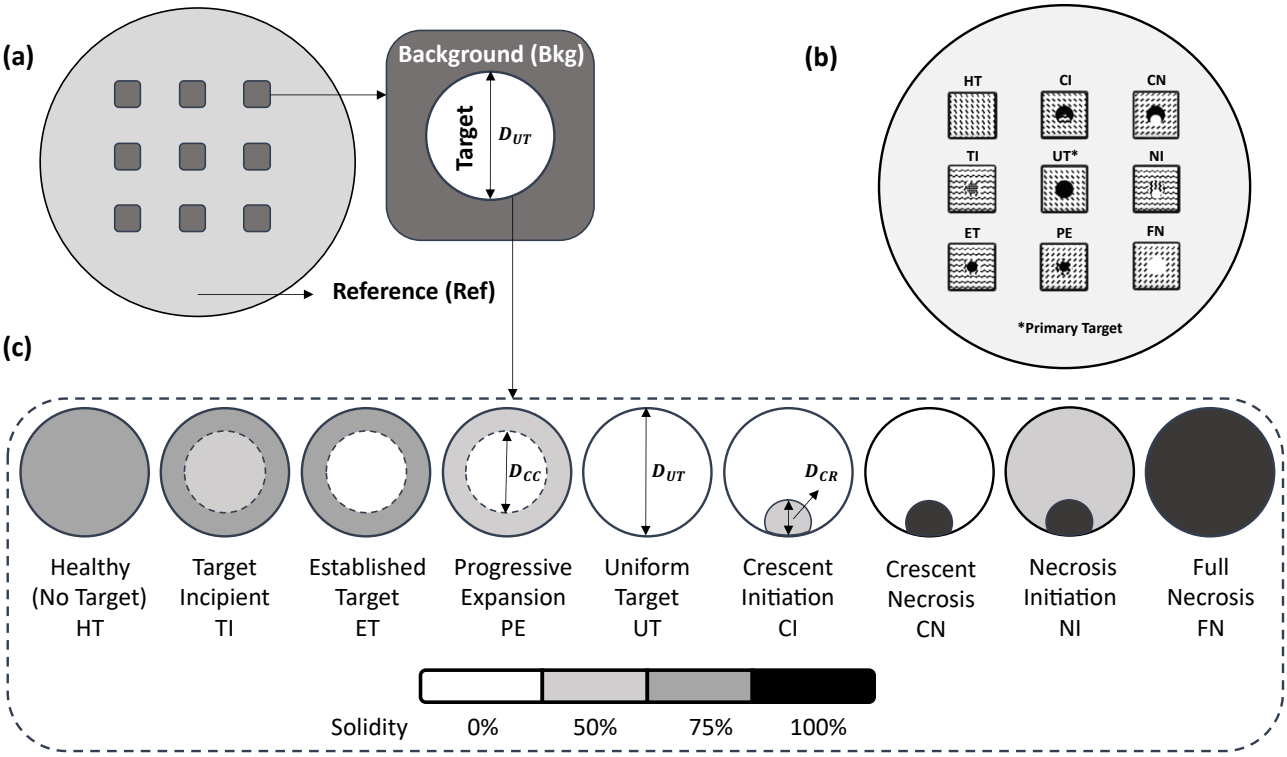


**Fig.1 (a)** Different regions of the custom phantom – Reference refers to the radiotracer + water solution; **(b)** 3 x 3 Layout used for each target size layer; **(c)** 9 targets types designed to simulate lesions with multi-uptake patterns, Solidity used to create dilution for different regions are represented using a colour bar (3 base controllable design parameters - $D_{UT}$ : Diameter of uniform target, $D_{CC}$ : Diameter of central core, $D_{CR}$ : Diameter of crescent region. Design configurations are also detailed in appendix (**Table A1**).

**2.2 Heterogeneous Target Design:**

Spatial heterogeneity simulation is achieved through introduction of a secondary sphere of different dimensions (**Appendix**) within these voids by another Boolean operation thus creating two nested regions (shown in **Fig. 1c**):

- The **primary region (region A)**, representing the highest tracer uptake
- The **secondary region (region B)**, representing reduced or necrotic activity

By varying the infill of these two regions (using either S: 0%, 50% or 100% representing TBR ~ 1, 2, Infinity) different uptake patterns can be generated to emulate various stages of lesion evolution - from healthy tissue to full necrosis or multifocal uptake. Nine distinct models were created for each base insert configuration (i.e., each uniform void size), resulting in a total of 27 inserts. The spatial arrangement of for each layer within the phantom is shown in **Fig.1b**. It is to be noted that the TBR for this new heterogeneous target now changes to a voxel based spatial function defined using **Eq.5**.

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

The $TBR_{insert}$ defined earlier is only used as a reference design parameter for cube identification and phantom positioning. Since background mesh and healthy insert share same infill percentage, the latter is always used in this study for quantitative calculations to have uniform assessment framework in case of delineation difficulties.

$$\text{TBR}\,(x,y,z) = \frac{AC_{target}\,(x,y,z)\text{ in kBq/mL}}{AC_{background}\,(x,y,z)\text{ in kBq/mL}} \approx \frac{AC_{target}\,(x,y,z)\text{ in kBq/mL}}{AC_{healthy}\text{ in kBq/mL}} \quad (5)$$

All designs were exported as .OBJ files and printed using a Prusa XL 3D printer using parameters detailed in [19] in acrylonitrile butadiene styrene (ABS). Printing success was verified via CT imaging (1.25 mm slice thickness).

**2.3 PET/CT Imaging Parameters & Data Analysis:**

The phantom was imaged on a Discovery OMNI Legend 32cm PET/CT system (*GE Healthcare, Chicago, IL, USA*) using two different tracers, [$^{18}$F]FDG ([$^{18}$F]-2-fluoro-2-deoxy-d-glucose) and [$^{68}$Ga] (68-Gallium), to evaluate design robustness and visualization consistency. The phantom was filled with soap & water solution followed FDG injection (approximately 6 hours post filling) for first acquisition. The phantom was then left undisturbed overnight for sufficient decay of the tracer. The following day Ga injection was carried out using same protocol for the second acquisition. The scan protocol was adjusted based on the tracer when necessary as shown in **Table 1** but following the recommendations from [26]. The attenuation CT (CTAC) parameters were maintained same as [26]. The acquired PET data was extracted using a semi-automatic workflow. Two distinct quantitative extractions were performed. First, for overall heterogeneity analysis, a standardized spherical Volume of Interest (VOI) with a diameter (+10% tolerance) than the insert's nominal design size (17, 22, 28mm), was automatically placed at the centroid of each insert. Second, for quantitative assessment of the primary region, an iso-surface segmentation analysis was performed on the most active regions within the same reference VOIs. Segmentation statistics were extracted as a function of the iso-surface threshold value. The results from the 75% iso-surface threshold (Corresponding to highest S: 75%) are utilised for main analysis.

**Table. 1 Acquisition & Reconstruction Parameters**

| Acquisition Parameter | $^{18}$F-FDG | $^{68}$Ga |
|---|---|---|
| **Bed Positions** | 1 | |
| **Matrix Dimensions** | 384 x 384 | |
| **Field of View** | 600 mm | |
| **Slice Thickness** | 2.07mm | |
| **Pixel Dimensions** | 1.56mm x 1.56mm | |
| **No. of Slices** | 153 | |
| **Voxel Volume** | 5.05mm$^3$ | |
| **Injected Activity** | 68.38 MBq | 93.04 MBq |
| **Injection to Scan relaxation time** | 42min | 25min |
| **Scan Duration** | 3 x 300s (50MBq)<br>3 x 600s (30MBq) | 3 x 600s (50MBq)<br>3 x 600s (30MBq) |
| **Reconstruction Algorithm** | QCHD (BPL*) | |
| **Reconstruction β parameter** | 650 | 750 |
| **Bayesian penalized-likelihood reconstruction with regularization parameter beta* | | |

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

Five different quantitative PET metrics were calculated from the extracted data, namely

- Standardized uptake value (SUV) for different regions from VOI data is given by

$$\mathrm{SUV}_{(\mathrm{mean}\,/\,\mathrm{max})} = \frac{\text{Activity Concentration in } \frac{\mathrm{Bq}}{\mathrm{mL}} \text{ (mean/max)}}{\text{(Injected activity in Bq)/(weight of phantom in g)}} \quad (7)$$

- Gini index (GI) was utilised as heterogeneity measure within the target volume, where a coefficient of 0 represents perfect homogeneity and 1 represents maximal heterogeneity **[18]**.

$$GI = \frac{2\sum_{i=1}^{n} i x_i}{n \sum_{i=1}^{n} x_i} - \frac{n+1}{n} \quad (8)$$

  Where n is total No. of voxels in the VOI, $x_i$ is voxel intensity of i-th voxel sorted in ascending order & i is the rank of the voxel in the sorted list

- Target to background ratio - $\mathrm{TBR}_{\mathrm{quant}}$ for segmentation (i.e. Hottest region) given by

$$\mathrm{TBR}_{\mathrm{quant}} = \frac{\mathrm{SUV}_{\mathrm{Segmentation}}}{\mathrm{SUV}_{\mathrm{HT}}} \approx \int \mathrm{TBR}(x, y, z)\, \mathrm{dV}_{\mathrm{seg}} \quad (9)$$

Statistical tests were performed to evaluate discrepancies. First, a Wilcoxon test was used to compare the extracted VOI data between each target type within each diameter group with a significance level of 0.05. The goal of Wilcoxon test is to quantify the combinations fail in similarity significance test (the higher the better). Second, a Bland-Altman analysis was conducted to assess the accuracy and potential bias. All data extraction and statistical analyses were conducted using MATLAB (*MathWorks INC, MA, USA*) and JASP (*JASP Team, Netherlands*).

**3. Results:**

CT air performed to confirm layer positioning and print success is presented in **Fig 2**.

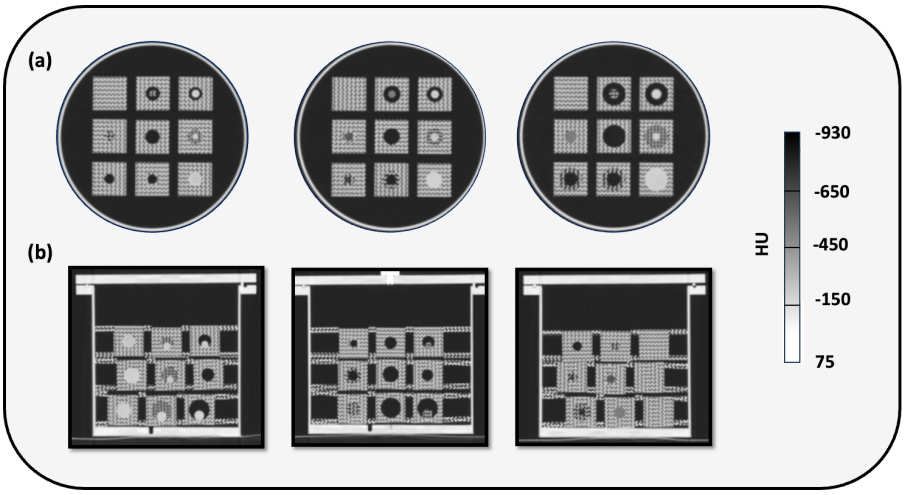


**Fig.2 (a)** Central Axial Slices for CT Air for 17mm (Left), 22mm (Middle), 28mm (Right); **(b)** Coronal view illustrating the crescent shapes of different infills and sizes

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

### 3.1 Visual and Quantitative Uptake Assessment:

The 3D-printed inserts successfully generated heterogeneous uptake patterns on the reconstructed PET data for both tracers (**Fig 3 & 4**). The nine designed uptake patterns, including uniform targets (UT), necrotic cores (e.g., NI, FN), and crescent-shaped regions (CI, CN) were most clearly represented in the largest (28 mm) inserts for both tracers. The smaller 22 mm inserts showed some intended patterns, though they were less distinct, while for the 17 mm inserts, these heterogeneous patterns (such as crescents) were visually indiscernible due to significant partial volume effects. The visual consistency between the $^{18}$F-FDG and $^{68}$Ga acquisitions was reasonably high. Mean Hounsfield ($HU_{mean}$) values within the VOIs on the CTAC varied between -3.2±6.4 to -38±20.4 for Fluorine and -3.1±4.6 to -31±13.4 for Gallium ensuring no presence of major air bubbles at scan resolution affecting the quantitation, falling within range prescribed by **[26]**. Uniform mixing of radiotracer was also confirmed based on layer wise variability as presented in **Table 2**.

Quantitative analysis of $SUV_{mean}$ and $SUV_{max}$ for each insert type demonstrated adherence to the design principles (**Fig 3b-e**). The profiles followed the expected pathological progression: uptake was minimal in the 'Healthy' (HT) insert, increased through the 'Incipient' (TI) , 'Established' (ET) and 'Expansion' (PE) stages, and peaked for the 'Uniform Target' (UT) and 'Crescent Initiation' (CI) inserts. As the simulated necrosis progressed ('Crescent Necrosis' (CN), 'Necrosis Initiation' (NI), and 'Full Necrosis' (FN)), the SUV progressively decreased, with FN showing near-zero uptake (for mean values). Partial volume and PET noise effects were also evident in the quantitative data. The peak SUV was visibly lower for the 17 mm inserts compared to the 22 mm and 28 mm inserts for both tracers and the mean necrosis SUV values were always higher than 0 but lower than HT. All mean bias values for cross activity comparison for both tracers led to < 10% bias within limits of agreement (LoA) as shown in **Fig 5a**. Kernel Density Estimation (KDE) plots of voxel-level SUV values were created for the 28 mm inserts (**Fig 4**) clearly captured the designed uptake patterns for both tracers.

**Table. 2 $SUV_{mean}$ for Reference Solution across layers**

| Layer No | $^{18}$F-FDG | | Maximum Variability in (%) | $^{68}$Ga | | Maximum Variability in (%) |
|---|---|---|---|---|---|---|
| | 50 MBq | 30 MBq | | 50 MBq | 30 MBq | |
| 1 (17mm) | 0.94±0.002 | 0.92±0.003 | 2.12% for 50MBq<br>1% for 30MBq | 0.91±0.003 | 0.91±0.001 | 1% for 50MBq<br>2.15% for 30MBq |
| 2 (22 mm) | 0.92±0.002 | 0.92±0.003 | | 0.92±0.003 | 0.92±0.002 | |
| 3 (28 mm) | 0.93±0.002 | 0.91±0.003 | | 0.92±0.003 | 0.93±0.003 | |

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

### 3.2 PET Quantitative Metrics:

Based on the phantom's morphological design progression, the Gini index is expected to follow a characteristic non-linear trend: starting low for homogeneous targets (Healthy/Background), rising for targets with focal uptake inequalities (Incipient/Established), dipping for the spatially uniform target (Uniform Target), and rising again for targets exhibiting high-contrast asymmetry or necrosis (Crescent/Necrosis). The capability of the phantom to generate this quantifiable heterogeneity is illustrated in **Fig. 5b,c**, which plots the Gini Index as a function of insert design and VOI size. The GI values ranged between 0.05 – 0.3 with more uniform / low contrast heterogeneity inserts closer to 0.05.

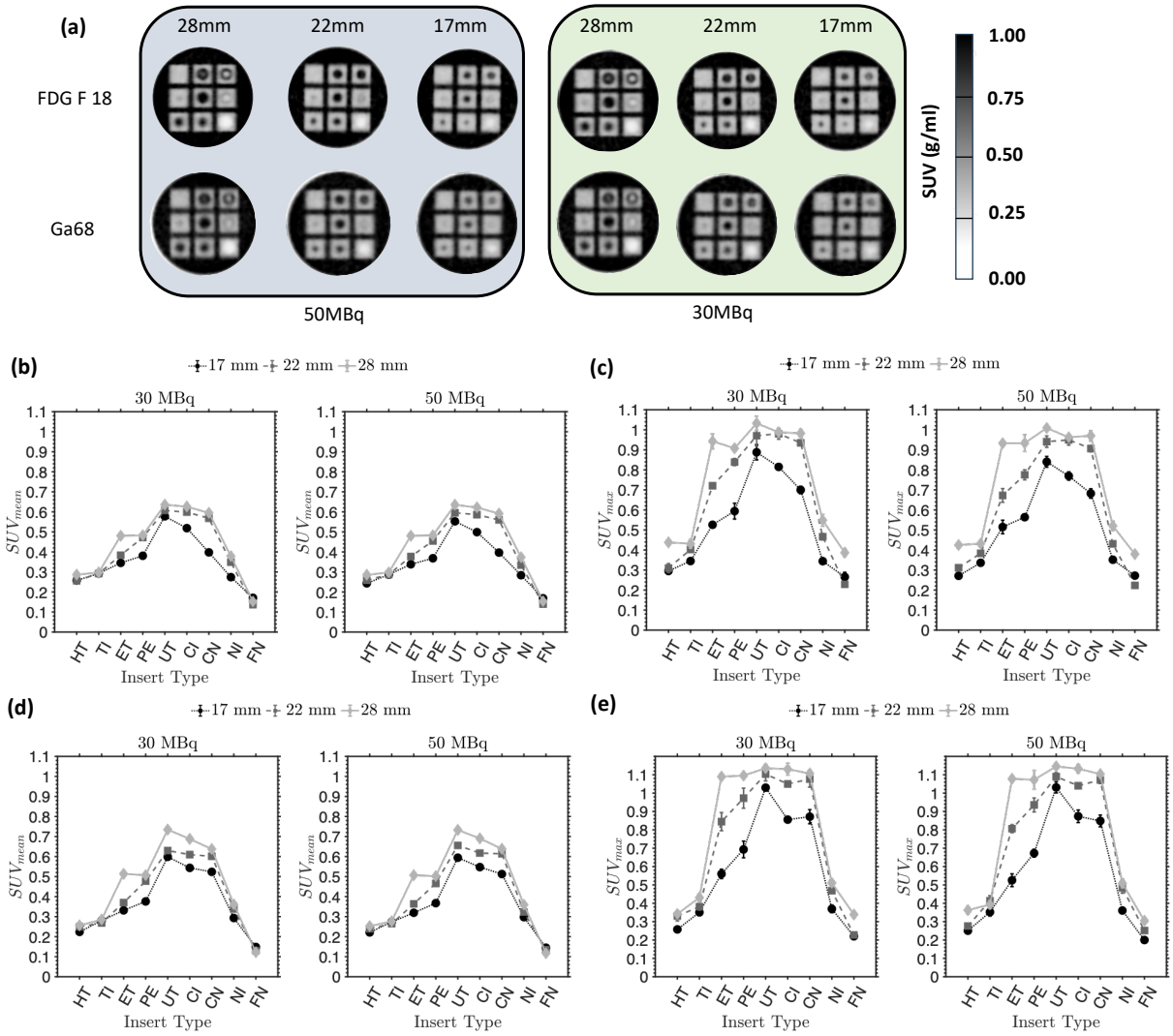


**Fig.3 (a)** PET axial slices of 3 layers for different activity concertation and tracers; Mean and Max SUV values as function of insert design type for **(b-c)** $^{18}$F-FDG and **(d-e)** $^{68}$Ga

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

To assess the design's effectiveness, the target-to-background ratio ($TBR_{quant}$) was evaluated and compared to the theoretical $TBR_{insert}$ of ~ 4. This analysis was performed on a 75% iso-value segmentation of the primary (hottest) target region, as defined in the methods. **Fig 6** plots the $TBR_{quant}$ (mean and max) for the five inserts containing a distinct, hot primary region (ET, PE, UT, CI, CN). The other four inserts (HT, TI, NI, FN) were excluded from this specific analysis as they either lack a hot target (HT), have a very low-contrast target (TI), or contain necrosis (FN, NI), making them challenging to delineate reliably with a 75% threshold. The $TBR_{max}$ values for the 28 mm inserts approached the theoretical TBR of 4. A qualitative comparison of the tracers showed that $^{68}$Ga (**Fig 6a, 6b**) consistently yielded a slightly lower TBR than $^{18}$F FDG (**Fig 6c, 6d**) for corresponding inserts and sizes. Furthermore, an analysis of $TBR_{max}$ values relative to the 10% range of the theoretical $TBR_{insert}$ (3.6 to 4.4) showed strong performance for the 28 mm inserts with all segmentations for both tracers fell within this range. This agreement degraded with size, as expected. On average the 28mm values fell with 10%, 22mm in 15-20% and 17mm in 30% bounds of $TBR_{insert}$.

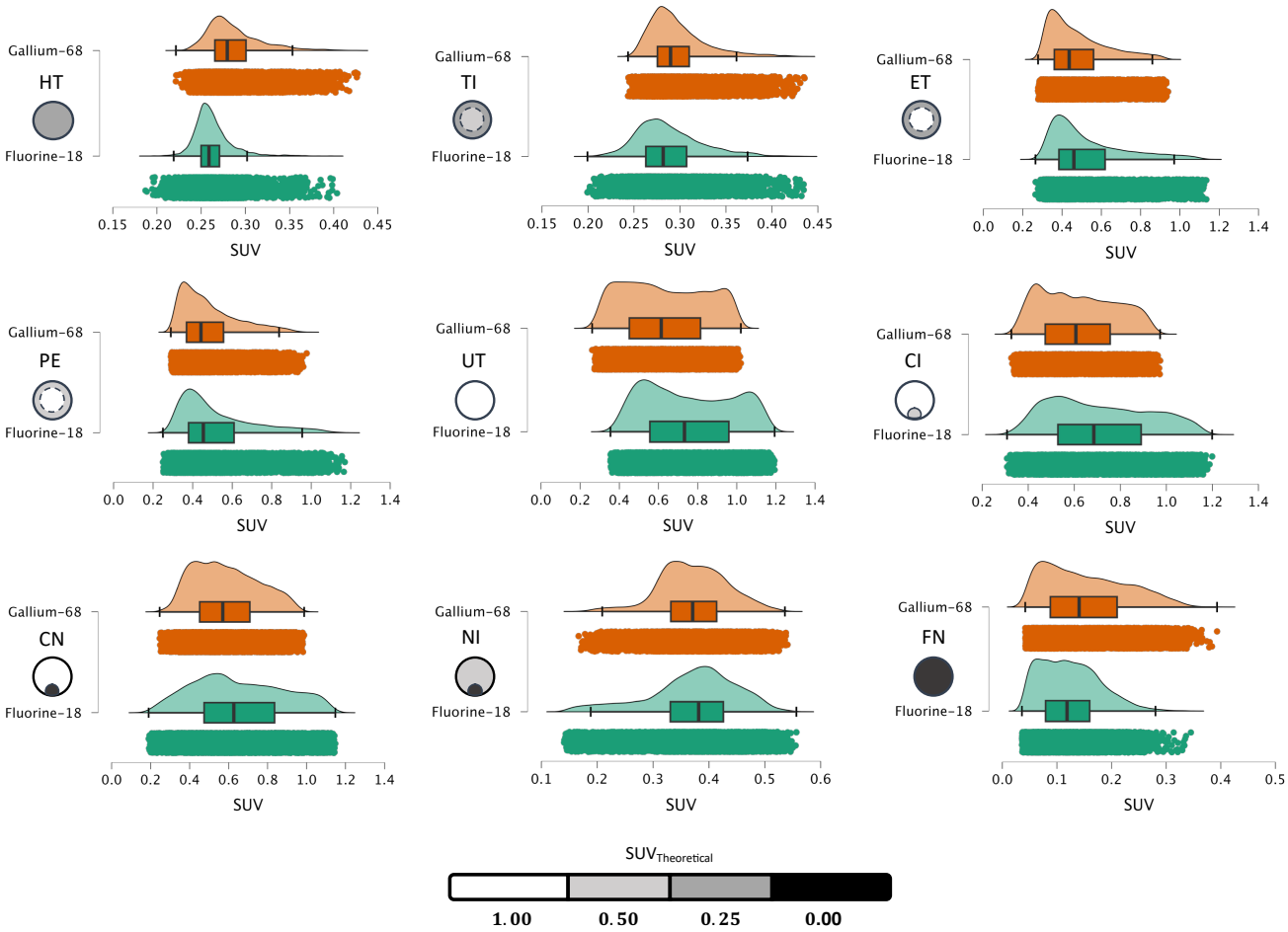


**Fig.4** Kernel Density Estimation (KDE) histogram plots from VOI data for the 28 mm inserts from the 50 MBq acquisition. Each plot shows the distribution of voxel-level SUV values for both $^{18}$F-FDG (Orange), $^{68}$Ga (Green)

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

**Discussion:**

This study successfully developed and validated a grid-inspired 3D-printed phantom inserts for simulating controlled spatial heterogeneous uptake patterns in PET imaging. The primary goal was to create a robust and reproducible tool that moves beyond simple hot-sphere phantoms with preparation simplicity. Hence, geometric simplicity (cubic outer shell, nested spherical voids) based on previous deign was chosen to ensure maximal quantitative interpretability and reproducibility as a prerequisite for future standardisation efforts. A crucial first step in validating the phantom's quantitative integrity was to confirm the success of printing and uniformity of tracer mixing in the reference solution. The former was confirmed through 1.25mm CT air as shown in **Fig 2**. The analysis of the reference $SUV_{mean}$ across the three phantom layers (**Table 2**) showed homogenous mixing for both tracers at both activity levels with a maximum variability of just 2.15%. Contrast to multi-compartment phantoms, proposed design functions with conventional acceptable standards of 10% **[10]** with a single tracer injection into the phantom body following findings of **[26].**

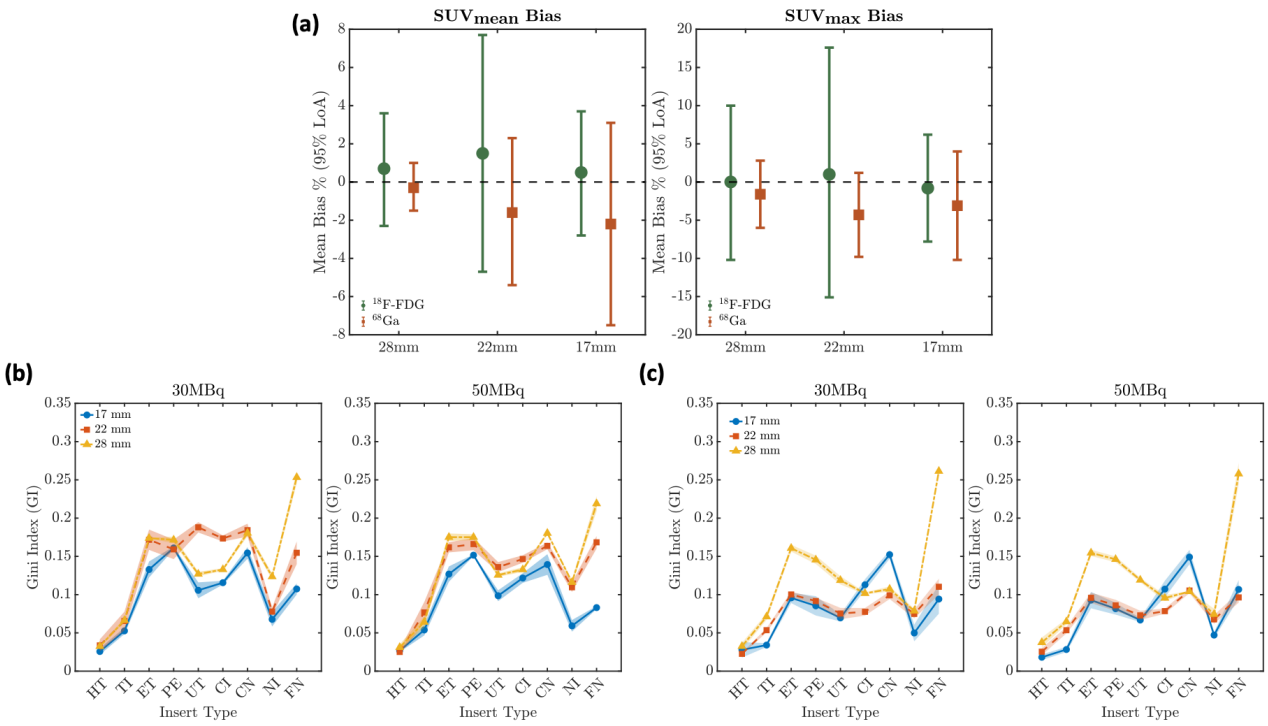


**Fig.5 (a)** Bland-Altman mean bias with 95% limits of agreement for 50 vs 30 MBq of extracted SUV values (Corresponding data in **Table A4**); Heterogeneity Gini index for **(b)** $^{18}$F-FDG & **(c)** $^{68}$Ga

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

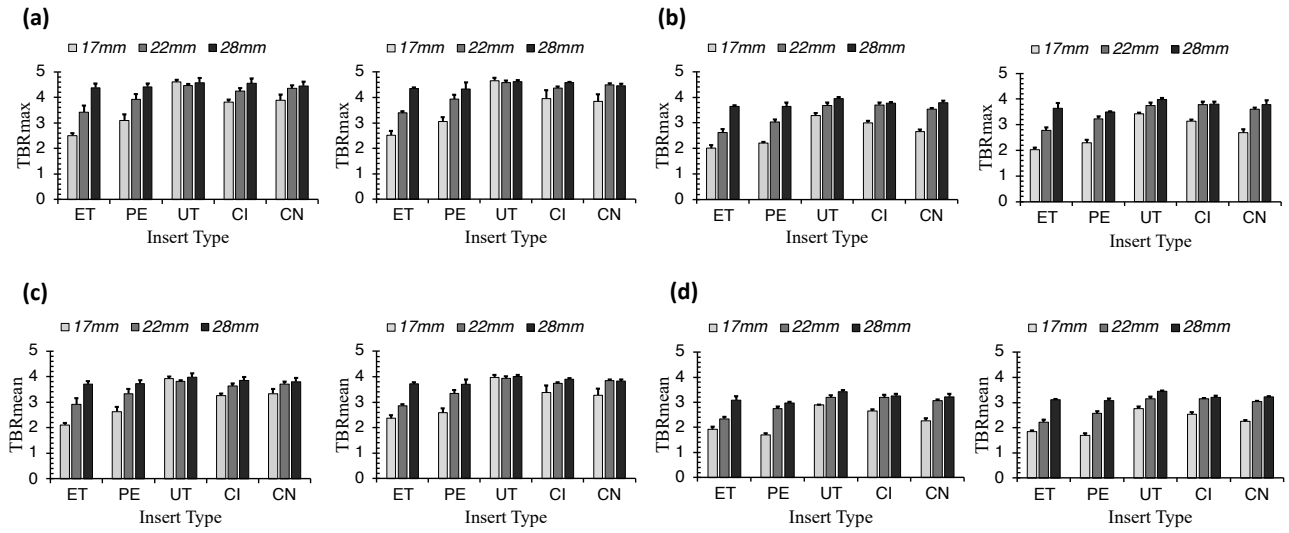


**Fig.6** Target-to-Background Ratio ($TBR_{quant}$) for the five hottest insert types, calculated from a 75% iso-value segmentation. Plots are shown for $TBR_{mean}$ and $TBR_{max}$ – (a) $^{18}$F-FDG; (b) $^{68}$Ga

The visual analysis of the PET images (**Fig 3a**) immediately confirmed the design's success and its utility for studying the limits of spatial resolution. The intended heterogeneous patterns, such as crescent shapes and necrotic cores, were clearly discernible in the 28 mm inserts. These fine structures became less distinct in the 22 mm inserts and were visually indiscernible in the 17 mm inserts, collapsing into a more uniform-appearing hot spot. This is not a failure of the phantom but rather a key feature: it provides a clear, physical demonstration of partial volume effects (PVE) and the threshold of visual detectability for complex heterogeneity in small lesions. However, a critical finding is that despite this visual blurring, the quantitative SUV profiles (**Fig 3b-e**) for the all inserts still demonstrated the distinct trends of the designed progression from healthy (HT) to necrotic (FN). This illustrates a crucial clinical challenge: small lesions can appear visually homogeneous while possessing significant underlying quantitative heterogeneity [27]. The phantom's ability to replicate this, is a significant advantage confirming that modulating the grid solidity is a highly effective method for controlling apparent tracer uptake. We acknowledge that clinical lesions often exhibit much higher contrast (typically TBR ~ 8-10) than the TBR ~ 4 used in this study. However, the primary goal was to build a reliable tool closer to the limits of detection and quantitative measurement, yet still practically achievable in terms of print resolution.

The voxel histograms in **Fig 4** reveal the distribution differences among the targets. While homogeneous targets (HT, UT) displayed symmetric peak (bell-shaped or flat peak), the heterogeneous inserts produced distinctly complex intensity profiles. Specifically, the multi-focal designs exhibited flattened, broad distributions indicating a

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

mix of signal intensities, whereas necrotic models featured high-uptake peaks with extended tails stretching into the low-activity range. This contrast confirms the phantom's ability to generate structural texture beyond simple averaging, with consistent voxel-level behaviour observed across both radiotracers. This suggests that the 3D-printed varying porosity successfully modulates the radiotracer distribution at a resolution detectable by the scanner. Furthermore, the pairwise statistical analysis (**Fig A1, A2**) validated that these simulated targets are not just visually different but are also quantitatively distinct except for 4 outliers.

We focused on validating the phantom's quantitative integrity using standard, reproducible metrics such as SUV, GI, and TBR. The consistency of SUV trends between $^{18}$F-FDG and $^{68}$Ga, and across different activity concentrations is a crucial finding. This qualitative trend performance (**Fig 3b vs 3d**) demonstrates that the phantom's relative contrast generation is robust, and independent of activity / tracer, making it a reliable tool for future standardization studies. Crucially, the mean SUV values for the necrotic regions were always higher than zero, reflecting the expected influence of printing resolution limitations and intrinsic PET noise/spill-in. The quantitative assessment of heterogeneity using the gini index revealed distinct trends correlating with the phantom's morphological design. As illustrated in **Fig. 5**, the GI profiles generally exhibited a characteristic trend: starting low for the homogeneous insert, rising for the focal targets, dipping for the spatially uniform, and rising again for the asymmetric and necrotic targets. This non-linear behaviour confirms the metric's sensitivity to structural complexity with the dip at the UT correctly distinguishing it from the surrounding heterogeneous grid structures. Full Necrosis frequently exhibited the highest Gini values (~0.25). This spike is attributable to two compounding factors. First, the necrotic core represents a region of near-zero counts, creating a high-noise environment. Second, the inclusion of the outer boundary captures the high-contrast gradient between the cold target and the warm background. This creates maximal statistical inequality driving the Gini index. Certain outliers were also observed, in the lower-activity 30 MBq acquisition, the GI for the UT in the 22 mm inserts disrupted the expected dip. Finally, it is important to note that the observed Gini coefficients ranged from 0.05 to 0.30, well below the theoretical maximum of 1.0. This limited dynamic range is an expected consequence of the study design. As the phantom was engineered with TBR ~ 4:1, the voxel intensity distribution is relatively tight leaded to tighter range for GI.

$TBR_{quant}$ analysis (**Fig 6**) served to validate the design's high-contrast performance against its theoretical $TBR_{insert}$ of ~ 4. The agreement for $TBR_{max}$ in the 28 mm inserts confirmed the grid design is well-calibrated. The predictable degradation of $TBR_{quant}$ (both mean and max) with decreasing insert size (22 mm and 17 mm) provided another clear, quantitative measure of PVE. Finally, the slight overestimation of $TBR_{max}$ by $^{18}$F-FDG (especially in the UT insert)

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

confirms a known bias in grid-based phantoms related to statistical noise [15,19], which the more-blurred $^{68}$Ga signal was incidentally less susceptible to. Despite its demonstrated utility, this phantom study has inherent limitations that define the necessity of future work. This work focused on spatial heterogeneity and conventional PET metrics (SUV, GI, TBR) to validate the foundational ground truth. We did not perform a detailed radiomics analysis within the scope of this study. Although patterns like the crescent and necrotic shapes (CI, CN) were introduced, these models do not fully replicate the highly irregular or spiculated morphologies observed in real patient tumours. It is important to note that the exclusion of the lower-contrast inserts (TI, NI) from the segmentation validation, was an imposed limitation of the fixed 75% iso-surface threshold. This specific threshold, optimized for validating the highest contrast primary regions, does not reflect a failure of the phantom design to simulate these challenging patterns. Instead, it underscores the need for advanced adaptive segmentation algorithms that can reliably delineate low-contrast or near-zero uptake regions, confirming the phantom's utility as a tool for developing and validating such methods. We acknowledge that full radiomic texture analysis is the next logical step. However, the primary aim of this manuscript is to validate the design feasibility and quantitative stability of the phantom. Establishing this ground truth is a prerequisite before evaluating complex radiomic feature stability. Crucially, the entire spatially heterogeneous structure for a given target size is produced as a single object from one design file, completely eliminating the multiple assembly steps or risk of leakage or cross-contamination associated with manually combining different regions of activity. Also, it is essential to note that this phantom is not intended for routine QC testing or system commissioning, which are adequately covered by current guidelines. Instead, it serves as a specialized additional tool for controlled heterogeneity simulation.

To conclude, the phantom provides a practical platform for radiomics studies where controlled, repeatable heterogeneity is essential. Known ground-truth patterns including cases where heterogeneity is quantitatively present but visually suppressed could be highly suited for evaluating feature robustness, sensitivity to partial-volume effects, segmentation uncertainty, and reconstruction biases. This tool could also support cross-site harmonization efforts by providing a standardized test object for feature stability benchmarking across centres, scanners, and protocols. It's simple preparation workflow further enables large-scale, multi-institutional reproducibility studies.

**Conclusion:**

This study successfully developed and validated 3D-printed phantom inserts designed to simulate controlled spatial heterogeneity in PET imaging. The phantom provides a foundational ground-truth reference validated using conventional PET metrics.

- The tool demonstrates the quantitative presence of heterogeneity even when visually suppressed by PVE.
- Its reproducible nature across tracer supports potential for cross-site standardization efforts.

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

**References:**


1. Lee JW, Lee SM. Radiomics in Oncological PET/CT: Clinical Applications. Nucl Med Mol Imaging. 2018;52:170–89. https://doi.org/10.1007/s13139-017-0500-y

2. Chen R, Ng YL, Yang X, Zhu Y, Li L, Zhao H, et al. Assessing dynamic metabolic heterogeneity in prostate cancer patients via total-body [68Ga]Ga-PSMA-11 PET/CT imaging: quantitative analysis of [68Ga]Ga-PSMA-11 uptake in pathological lesions and normal organs. Eur J Nucl Med Mol Imaging. 2024;51:896–906. https://doi.org/10.1007/s00259-023-06475-y

3. Dubash S, Inglese M, Mauri F, Kozlowski K, Trivedi P, Arshad M, et al. Spatial heterogeneity of radiolabeled choline positron emission tomography in tumors of patients with non-small cell lung cancer: first-in-patient evaluation of [$^{18}$F]fluoromethyl-(1,2-$^{2}$H$_4$)-choline. Theranostics. Ivyspring International Publisher; 2020;10:8677–90. https://doi.org/10.7150/thno.47298

4. Lovinfosse P, Hatt M, Visvikis D, Hustinx R. Heterogeneity analysis of 18F-FDG PET imaging in oncology: clinical indications and perspectives. Clin Transl Imaging. 2018;6:393–410. https://doi.org/10.1007/s40336-018-0299-2

5. Doumou G, Siddique M, Tsoumpas C, Goh V, Cook GJ. The precision of textural analysis in 18F-FDG-PET scans of oesophageal cancer. Eur Radiol. 2015;25:2805–12. https://doi.org/10.1007/s00330-015-3681-8

6. Du D, Shiri I, Yousefirizi F, Salmanpour MR, Lv J, Wu H, et al. Impact of harmonization and oversampling methods on radiomics analysis of multi-center imbalanced datasets: application to PET-based prediction of lung cancer subtypes. EJNMMI Phys. 2025;12:34. https://doi.org/10.1186/s40658-025-00750-7

7. Antunes JT, Ismail M, Hossain I, Wang Z, Prasanna P, Madabhushi A, et al. RADIomic Spatial TexturAl Descriptor (RADISTAT): Quantifying Spatial Organization of Imaging Heterogeneity Associated With Tumor Response to Treatment. IEEE J Biomed Health Inform. 2022;26:2627–36. https://doi.org/10.1109/JBHI.2022.3146778

8. Geady C, Bannon JJ, Reza S, Madanat-Harjuoja L, Reinke D, Schuetze S, et al. Measured intrapatient radiomic variability as a predictor of treatment response in multi-metastatic soft tissue sarcoma patients. Sci Rep. Nature Publishing Group; 2025;15:27838. https://doi.org/10.1038/s41598-025-12451-3

9. 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. https://doi.org/10.1186/s40658-022-00496-6

10. Koole M, Armstrong I, Krizsan AK, Stromvall A, Visvikis D, Sattler B, et al. EANM guidelines for PET-CT and PET-MR routine quality control. Z Für Med Phys. 2023;33:103–13. https://doi.org/10.1016/j.zemedi.2022.08.003

11. Tino R, Yeo A, Leary M, Brandt M, Kron T. A Systematic Review on 3D-Printed Imaging and Dosimetry Phantoms in Radiation Therapy. Technol Cancer Res Treat. SAGE Publications Inc; 2019;18:1533033819870208. https://doi.org/10.1177/1533033819870208

12. Qiu J, Hou K, Dyer BA, Chen J-C, Shi L, Sun Y, et al. Constructing Customized Multimodal Phantoms Through 3D Printing: A Preliminary Evaluation. Front Phys [Internet]. Frontiers; 2021 [cited 2025 Apr 9];9. https://doi.org/10.3389/fphy.2021.605630

13. Okkalidis N. 3D printing methods for radiological anthropomorphic phantoms. Phys Med Biol. IOP Publishing; 2022;67:15TR04. https://doi.org/10.1088/1361-6560/ac80e7

14. 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. https://doi.org/10.1186/s40658-016-0151-6

15. 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. https://doi.org/10.1186/s40658-024-00613-7


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

16. Gallivanone F, Interlenghi M, D'Ambrosio D, Fantinato D, Alberizzi L, Trifirò G, et al. An anthropomorphic phantom for advanced image processing of realistic 18F-FDG PET-CT oncological studies. 2016 IEEE Nucl Sci Symp Med Imaging Conf Room-Temp Semicond Detect Workshop NSSMICRTSD [Internet]. 2016 [cited 2025 Oct 24]. p. 1–7. https://doi.org/10.1109/NSSMIC.2016.8069418

17. 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. https://doi.org/10.2967/jnumed.121.262925

18. 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. https://doi.org/10.1155/2018/5324517

19. 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. https://doi.org/10.1101/2025.04.01.25325022

20. Pfaehler E, Beukinga RJ, de Jong JR, Slart RHJA, Slump CH, Dierckx RAJO, et al. Repeatability of 18F-FDG PET radiomic features: A phantom study to explore sensitivity to image reconstruction settings, noise, and delineation method. Med Phys. 2019;46:665–78. https://doi.org/10.1002/mp.13322

21. Valladares A, Beyer T, Rausch I. Physical imaging phantoms for simulation of tumor heterogeneity in PET, CT, and MRI: An overview of existing designs. Med Phys. 2020;47:2023–37. https://doi.org/10.1002/mp.14045

22. Pfaehler E, Sluis J van, Merema BBJ, Ooijen P van, Berendsen RCM, Velden FHP van, et al. Experimental Multicenter and Multivendor Evaluation of the Performance of PET Radiomic Features Using 3-Dimensionally Printed Phantom Inserts. J Nucl Med. Society of Nuclear Medicine; 2020;61:469–76. https://doi.org/10.2967/jnumed.119.229724

23. Zounek AJ, Joerg NM, Lindheimer F, Zatcepin A, Palumbo G, Oos R, et al. 3D printing of radioactive wall-less PET phantoms improves threshold-based target delineation and quantification. EJNMMI Phys. 2025;12:53. https://doi.org/10.1186/s40658-025-00768-x

24. Meier LP, Läppchen T, Hentschel M, Rominger A, Prenosil GA. 3-D Printing of Calibrated Germanium-68 PET Phantoms Without Inactive Walls. IEEE Trans Radiat Plasma Med Sci. 2023;7:775–84. https://doi.org/10.1109/TRPMS.2023.3275994

25. 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. IOP Publishing; 2025;70:115001. https://doi.org/10.1088/1361-6560/add708

26. Nalla SKK, Maronnier Q, Palchan-Hazan T, Caselles O, Kennedy J. Facilitating Multi-Center Consistency in PET/CT Assessment: Transferability and Harmonization of 3D-Printed Phantom Inserts [Internet]. medRxiv; 2025 [cited 2025 June 1]. p. 2025.05.30.25328639. https://doi.org/10.1101/2025.05.30.25328639

27. Lokre O, Perk TG, Weisman AJ, Govindan RM, Chen S, Chen M, et al. Quantitative evaluation of lesion response heterogeneity for superior prognostication of clinical outcome. Eur J Nucl Med Mol Imaging. Springer; 2024;51:3505–17. https://doi.org/10.1007/s00259-024-06764-0

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

## APPENDIX

Description for the nomenclature of these models shown in **main article (Fig 1.c)** is as below

- **Healthy (HT) :** Represents the baseline simulation of healthy tissue, characterized by a uniform 75% solidity
- **Target Incipient (TI) :** Simulates the earliest stage of lesion development, featuring a faint **primary region** with 50% infill. This initial region is typically smaller than the final intended target size – Multifocal Target
- **Established Target (ET) :** Simulates the first clearly detectable lesion. The *primary region*, maintaining the same size as in the IT stage, transitions to 0% infill, creating a distinct, hollow core - Multifocal Target
- **Progressive Expansion (PE) :** Simulates the active growth phase. A *secondary region* (outer ring) with 50% infill forms and expands around the hollow primary region established in the ET stage - Multifocal Target
- **Uniform Target (UT) :** Represents a single, homogeneous lesion that fills an entire intended target area (e.g., 17, 22, or 28 mm in diameter) – Uniform Target
- **Crescent Initiation (CI) :** Simulates a pattern forming within a Uniform Target. A new internal secondary region with 50% infill develops, causing the remaining original tissue to form a crescent-shaped primary region - Multifocal Target
- **Crescent Necrosis (CN) :** Simulates the progression of the crescent pattern. The internal secondary region becomes fully necrotic, transitioning to 100% solidity, while being enclosed by the same crescent-shaped primary region – Necrosis Target
- **Necrosis Initiation (NI) :** Simulates the beginning of total lesion collapse in a target structure. The secondary region becomes fully necrotic (100% infill), while necrosis begins to consume the primary crescent region represented by a transition to 50% infill - Necrosis Target
- **Full Necrosis (FN) :** Represents the final end-stage. Both the primary and secondary regions are consumed, resulting in a single, homogeneous necrotic lesion emulation with 100% infill across the entire intended target size - Necrosis Target

**Table. A1 Design Parameters Used for the Inserts**

| Healthy / Background | $D_{UL}$ in mm & Solidity % (Primary Region) | $D_{CC}$ in mm & Solidity % (Secondary Region) | $D_{CR}$ & Solidity % (Secondary Region) |
|---|---|---|---|
| 0mm, 75% (HT) | 17mm, 0% (UT)<br>17mm, 100% (FN)<br>17mm, 50% (PE & NI) | 11mm, 50% (TI) | 11mm, 50% (CI) |
| | | 11mm, 0% (ET) | 11mm, 100% (CN) |
| | 22 mm, 0% (UT)<br>22mm, 100% (FN)<br>22mm, 50% (PE & NI) | 12mm, 50% (TI) | 12mm, 50% (CI) |
| | | 12mm, 0% (ET) | 12mm, 100% (CN) |
| | 28 mm, 0% (UT)<br>28mm, 100% (FN)<br>28mm, 50% (PE & NI) | 17mm, 50% (TI) | 17mm, 50% (CI) |
| | | 17mm, 50% (ET) | 17mm, 50% (CN) |

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

Recovery coefficient for the segmented data is calculated as below

$$RC_{(mean/max)} = \frac{\text{SUV Inside the Target ROI (Mean/Max)}}{\text{Theoretical SUV (Mean/Max)}}$$

**Table. A2 $RC_{mean}$ calculated from segmented data for primary region**

| Layer No | $^{18}$F-FDG | | | $^{68}$Ga | | |
|---|---|---|---|---|---|---|
| | 17mm | 22mm | 28mm | 17mm | 22mm | 28mm |
| 50MBq | | | | | | |
| ET | 0.58±0.02 | 0.76±0.02 | 1.03±0.02 | 0.52±0.01 | 0.62±0.03 | 0.87±0.01 |
| PE | 0.64±0.02 | 0.89±0.03 | 1.02±0.04 | 0.47±0.03 | 0.72±0.03 | 0.86±0.03 |
| UT | 0.98±0.03 | 1.04±0.02 | 1.11±0.01 | 0.77±0.02 | 0.88±0.03 | 0.96±0.01 |
| CI | 0.83±0.03 | 0.99±0.01 | 1.08±0.02 | 0.71±0.02 | 0.88±0.01 | 0.90±0.01 |
| CN | 0.81±0.03 | 1.02±0.01 | 1.06±0.00 | 0.63±0.02 | 0.85±0.01 | 0.90±0.02 |
| 30MBq | | | | | | |
| ET | 0.53±0.02 | 0.81±0.05 | 1.03±0.01 | 0.53±0.04 | 0.65±0.01 | 0.86±0.03 |
| PE | 0.66±0.04 | 0.92±0.05 | 1.03±0.01 | 0.47±0.01 | 0.76±0.02 | 0.83±0.01 |
| UT | 0.98±0.01 | 1.05±0.03 | 1.10±0.02 | 0.81±0.02 | 0.89±0.04 | 0.96±0.02 |
| CI | 0.81±0.00 | 1.00±0.01 | 1.07±0.02 | 0.74±0.02 | 0.89±0.02 | 0.91±0.00 |
| CN | 0.83±0.03 | 1.02±0.03 | 1.05±0.01 | 0.63±0.01 | 0.85±0.01 | 0.90±0.02 |

**Table. A3 $RC_{max}$ calculated from segmented data for primary region**

| Layer No | $^{18}$F-FDG | | | $^{68}$Ga | | |
|---|---|---|---|---|---|---|
| | 17mm | 22mm | 28mm | 17mm | 22mm | 28mm |
| 50MBq | | | | | | |
| ET | 0.62±0.05 | 0.90±0.02 | 1.20±0.02 | 0.62±0.01 | 0.73±0.04 | 1.02±0.01 |
| PE | 0.75±0.02 | 1.03±0.04 | 1.20±0.05 | 0.56±0.04 | 0.85±0.03 | 1.02±0.05 |
| UT | 1.15±0.03 | 1.22±0.02 | 1.28±0.01 | 0.92±0.03 | 1.03±0.03 | 1.10±0.01 |
| CI | 0.98±0.04 | 1.16±0.01 | 1.26±0.03 | 0.84±0.02 | 1.04±0.02 | 1.05±0.01 |
| CN | 0.95±0.04 | 1.20±0.02 | 1.23±0.00 | 0.75±0.02 | 0.99±0.02 | 1.06±0.03 |
| 30MBq | | | | | | |
| ET | 0.62±0.03 | 0.94±0.05 | 1.21±0.02 | 0.64±0.04 | 0.77±0.02 | 1.01±0.04 |
| PE | 0.77±0.05 | 1.08±0.06 | 1.22±0.01 | 0.57±0.01 | 0.90±0.02 | 0.98±0.01 |
| UT | 1.15±0.01 | 1.23±0.04 | 1.27±0.02 | 0.95±0.04 | 1.04±0.05 | 1.11±0.04 |
| CI | 0.95±0.00 | 1.17±0.01 | 1.26±0.04 | 0.88±0.00 | 1.05±0.03 | 1.06±0.01 |
| CN | 0.97±0.04 | 1.20±0.05 | 1.23±0.02 | 0.75±0.02 | 1.00±0.01 | 1.06±0.02 |

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

**Table. A4 Bland-Altman Mean Bian with Limits of Agreement in % for 50 vs 30 MBq**

| Mean Bias % (-1.96 Std.Dev , 1.96 Std.Dev ) | | |
|---|---|---|
| **Insert Size** | **$SUV_{mean}$** | **$SUV_{max}$** |
| **$^{18}$F-FDG** | | |
| **28mm** | 0.7 (-2.3, 3.6) | 0.01 (-10.2, 10.0) |
| **22mm** | 1.5 (-4.7, 7.7) | 1.0 (-15.1, 17.6) |
| **17mm** | 0.5 (-2.8, 3.7) | -0.8 (-7.8, 6.2) |
| **$^{68}$Ga** | | |
| **28mm** | -0.3 (-1.5, 1.0) | -1.6 (-6.0, 2.8) |
| **22mm** | -1.6 (-5.4, 2.3) | -4.3 (-9.8, 1.2) |
| **17mm** | -2.2 (-7.5, 3.1) | -3.1 (-10.2, 4.0) |

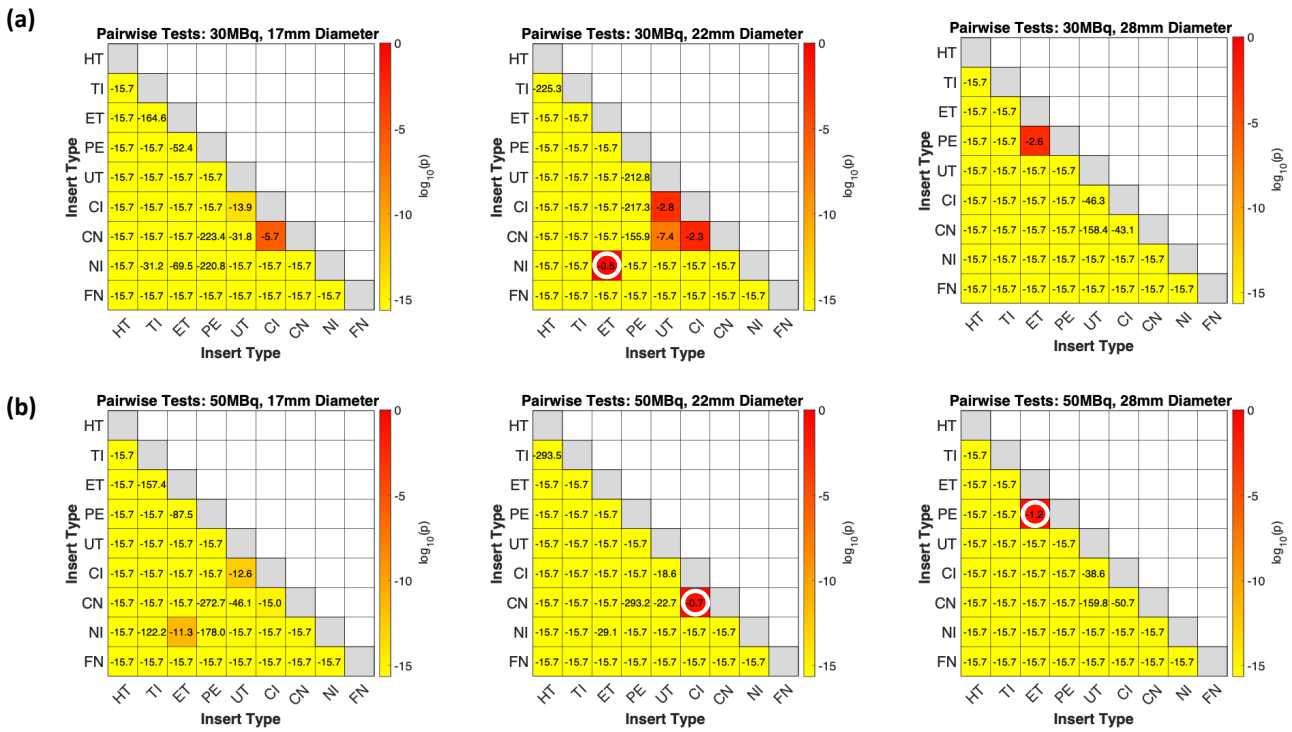


**Fig.A1** Wilcoxon Rank Sum test for the VOI histogram data of 28mm inserts for [$^{18}$F]FDG **(a)** 30Mbq**; (b)** 50MBq – p value plotted in log scale : **≤ -1.3** which is log(0.05) - (Values higher than -1.3 indicates failures i.e. the two targets are statistically similar to each other)

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

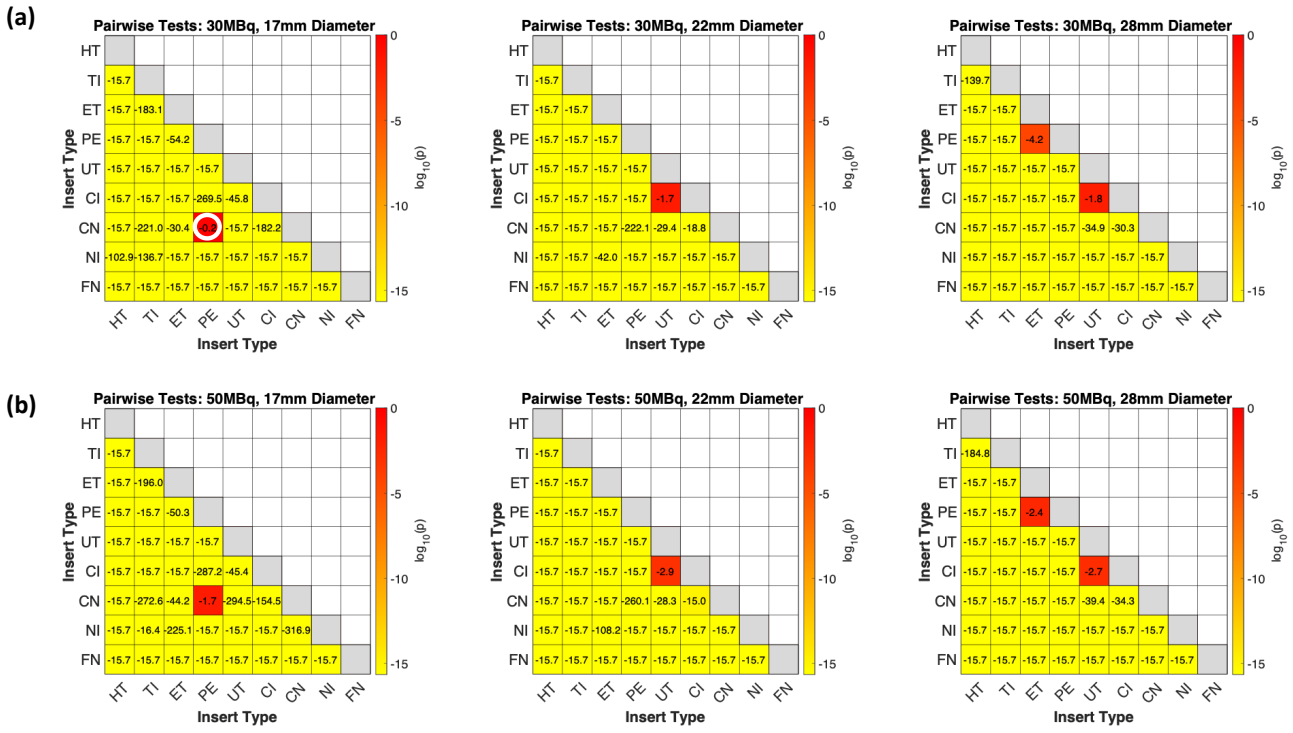


**Fig.A2** Wilcoxon Rank Sum test for the VOI histogram data of 28mm inserts for [$^{68}$Ga] **(a)** 30Mbq**; (b)** 50MBq – p value plotted in log scale : **≤ -1.3** which is log(0.05) - (Values higher than -1.3 indicates failures i.e. the two targets are similar to each other)

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