Journal of Nuclear Medicine, published on November 9, 2017 as doi:10.2967/jnumed.117.198853

Reproducibility of MR-based Attenuation Maps in PET/MRI and the Impact on PET Quantification in Lung Cancer

Anders Olin[1], Claes N. Ladefoged[1], Natasha H. Langer[1], Sune H. Keller[1], Johan Löfgren[1], Adam E. Hansen[1], Andreas Kjær[1], Seppo W. Langer[2], Barbara M. Fischer[1], and Flemming L. Andersen[1]

[1]Department of Clinical Physiology, Nuclear Medicine & PET, Rigshospitalet, University of Copenhagen, Copenhagen, Denmark

[2]Department of Oncology, Rigshospitalet, University of Copenhagen, Copenhagen, Denmark

Running Title: Reproducibility of thoracic MR-AC maps

Corresponding author: Flemming Littrup Andersen, Ph.D, Department of Clinical Physiology, Nuclear Medicine and PET, Rigshospitalet, University of Copenhagen, Blegdamsvej 9, 2100 Copenhagen, Denmark, Telephone: +45 35458143, Fax: +45 35453898, Email: flemming.andersen@regionh.dk

First author: Anders Olin, MSc, Department of Clinical Physiology, Nuclear Medicine and PET, Rigshospitalet, University of Copenhagen, Blegdamsvej 9, 2100 Copenhagen, Denmark, Email: anders.olin@regionh.dk

Key words: PET/MR, MR-based attenuation correction, reproducibility, non-small cell lung cancer

Total words: 4879

Abstract words: 329

ABSTRACT

AIM: Quantitative positron emission tomography/magnetic resonance imaging (PET/MRI) is dependent upon reliable and reproducible MR-based attenuation correction (MR-AC). In this study we evaluated the quality of current vendor-provided thoracic MR-AC maps, and further investigated the reproducibility of these in terms of their impact on fluorodeoxyglucose (FDG)-PET quantification in patients with non-small cell lung cancer (NSCLC). **MATERIAL AND METHODS:** Eleven patients with inoperable NSCLC underwent two to five thoracic PET/MRI scan-rescan examinations within 22 days. FDG-PET data were acquired along with two Dixon MR-AC maps for each examination. Two PET images (PET_A and PET_B) were reconstructed using identical PET emission data but with MR-AC from these intra-subject repeated attenuation maps. A total of 90 MR-AC maps were evaluated visually for quality and occurrence of categorized artifacts by two PET/MRI experienced physicians. Each tumor was outlined by a volume of interest (VOI; 40% isocontour of maximum) on PET_A, which was then projected onto the corresponding PET_B. Mean standardized uptake value (SUV_{mean}) and maximum (SUV_{max}) were assessed from the PET images. Within-examination coefficients of variation and Bland-Altman analyses were conducted for the assessment of SUV variations between PET_A and PET_B. **RESULTS:** Image artifacts were observed in 86% of the MR-AC maps and 30% of the MR-AC maps were subjectively expected to affect the tumor SUV. SUV_{mean} and SUV_{max} resulted in coefficients of variation of 5.6% and 6.6%, respectively, and scan-rescan SUV variations were within ±20% in 95% of the cases. Substantial SUV variations were mainly seen for scan-rescan examinations affected by respiratory motion. **CONCLUSION:** Artifacts occur frequently in standard thoracic MR-AC maps affecting the reproducibility of PET/MRI. This, in combination with other well-known sources of error associated

with PET/MRI examinations, leads to inconsistent SUV measurements in serial studies, which may affect the reliability of therapy response assessment. A thorough visual inspection of the thoracic MR-AC map and Dixon images from which it is derived remains crucial for the detection of MR-AC artifacts that may influence the reliability of SUV.

INTRODUCTION

PET imaging with various tracers, including fluorine-18 FDG, is increasingly used for evaluation of biologic response to cancer treatment (*1*). Reliable and repeatable quantitative PET images are therefore important for follow-up examinations. The reproducibility of SUV in PET-only and combined PET and computed tomography (PET/CT) examinations has been well studied (*2*–*11*). A single study by Rasmussen *et al.* also analyzed the reproducibility of FDG-PET/MRI scans (*12*). Reproducibility of PET imaging is affected by many sources of variations arising from emission and attenuation data. However, none of the mentioned studies have tried to isolate and quantify the effect of the different sources of variation, which include errors of the images used for attenuation correction of PET. Attenuation correction is an essential prerequisite for quantification of PET images and the reproducibility of attenuation maps is therefore expected to be strongly linked to the overall PET reproducibility.

Whereas a PET/CT system utilizes the CT transmission scan to generate an attenuation map, PET/MRI systems lack such information. For the Siemens mMR scanner, the current vendor-provided MR-based attenuation correction (MR-AC) maps are derived from a segmentation of Dixon images (*13*) into four classes (soft tissue, fat, lungs, and air) each with a predetermined linear attenuation coefficient value (*14*). It is well-known that this method for MR-AC introduces a systematic and considerable bias, especially due to the lacking bone (*15*) and the allocation of a single predetermined attenuation coefficients to specific tissues. Nevertheless, the Dixon-based method is the reality in vendor-provided whole-body clinical routine today, and it is therefore also of interest to investigate the reproducibility of the Dixon-based MR-AC maps. This reproducibility has been questioned because

the underlying Dixon MR images are sensitive to artifacts (*16*–*18*). Truncation artifacts due to the limited field-of-view, water/fat inversion artifacts and patient movements have all shown to affect the measured SUVs of lesions (*19*–*22*). Thoracic MR-AC maps are especially challenging to derive due to respiratory motion, which may occur despite clear breathing instructions given to the patient. Paulus *et al.* exemplified that incorrect lung classification in an MR-AC map resulted in 53% increase of SUV_{mean} relative to CT-based attenuation correction for a lesion in the vicinity of the lung (*23*). Furthermore, the fast signal decay induced by air-tissue interfaces may cause small lesions or fine reticulations to be missed (*24*,*25*).

In the present study we evaluate the quality of current vendor-provided thoracic MR-AC maps and investigate how the reproducibility of these maps affects PET quantification in patients with NSCLC.

MATERIALS AND METHODS

Patients

This prospective clinical study included eleven patients (four females and seven males) with inoperable stage IIIB/IV NSCLC but who where all in good physical condition, i.e. performance status 0 and 1 (*26*).

The study was approved by the departmental science committees at Rigshospitalet, by the Regional Ethics Committee, Copenhagen, Denmark, approval number H-3-2013-090, and by the Danish Data Protection Agency.

The characteristics of the studied patients with NSCLC are given as mean ± standard deviation: age 62±6 yr; weight 70±11 kg.

FDG PET/MRI

Each patient underwent two to five thoracic PET/MRI scan-rescan examinations within 22 days of the first examination (before and during chemotherapy), resulting in a total of 45 examinations included in this study. All scans were performed on the same Siemens Biograph mMR (Siemens Healthcare, Erlangen, Germany) with software version Syngo MR B20P at Rigshospitalet.

For each examination, the patients were instructed to fast for minimum six hours before undergoing a PET/MRI scan. PET scans were started approximately 60 min after injection of FDG (2 MBq/kg). Data from two bed positions were acquired covering thorax and upper abdomen to fully cover the lungs. The PET emission data acquisition time was 8 minutes per bed-position. Simultaneously, a standard two-point Dixon MRI sequence was acquired for derivation of the MR-AC map, using repetition time of 3.60 ms, echo times of 1.23 and 2.46 ms, and a flip angle of 10 degrees. Prior to this, the patients were instructed to hold their breath at end expiration in order to minimize respiration artifacts. The Dixon images (in-phase, opposed-phase, water, and fat) were generated on 192×126×128 matrices with a voxel size of 2.6×2.6×3.1 mm^3 for each bed-position and then composed into one image volume for the derivation of the MR-AC map following vendor-specific processing. Without repositioning the patient, a second PET/MRI rescan was performed as described above. However, the PET data from this rescan was not used in the subsequent analyzes.

The rescan MR-AC map was co-registered to the first MR-AC map by a 6-parameter rigid alignment procedure (minctracc, McConnel Imaging Center, Montreal, Canada) with cross-correlation as the objective function and using nearest-neighbor interpolation. The first MR-AC map is denoted as $\mu\text{-map}_A$ and the co-registered rescan MR-AC map as $\mu\text{-map}_B$.

Using the same raw PET emission data, two different PET images, PET_A and PET_B, were reconstructed offline using E7Tools (Siemens Medical Solutions, Knoxville, USA) with μ-map_A and μ-map_B for MR-AC, respectively. The reconstructions were performed using 3D OP-OSEM (Ordinary Poisson-Ordered Subset Expectation Maximization with 3 iterations, 21 Subsets, 4 mm Gaussian post filtering) on 344×344x224 matrices with a voxel size of 2.1×2.1×2.0 mm^3. MLAA (maximum likelihood reconstruction of attenuation and activity) was not applied.

Image Analyses

The general quality of MR-AC maps was assessed by a visual evaluation performed by PET/MRI-experienced physicians. The impact of MR-AC map reproducibility on PET quantification was analyzed by examining tumor SUV variability between PET_A and PET_B.

Visual Evaluation of MR-AC Maps

Two PET/MRI-experienced physicians visually evaluated all 90 MR-AC maps for occurrences of categorized artifacts, first separately and then in consensus. We report only the findings of the consensus artifact assessment. Each MR-AC map was inspected with only standard MRI sequences and PET for anatomical and metabolic references, respectively. The types of artifacts were based on those presented in studies by Keller *et al.* (*19*) and Brendle *et al.* (*22*), which include: Metal artifacts, trachea artifacts (the air-filled trachea segmented as soft tissue or fat), lung border artifacts, respiration artifacts, body contour artifacts, and erroneous tissue classification. Truncation artifacts were present in all MR-AC maps and were not included in this analysis.

Each MR-AC map was further subjectively categorized into two categories, *critical* or *non-critical*, based on whether any present artifacts were expected to have an impact on a tumor's SUV. Especially, the size and location of the artifacts relative to the tumor were taken into account.

Quantitative Evaluation of PET Images

One of the two PET/MRI-experienced physician outlined every tumor of each patient on PET_A by a VOI using the image analysis tool Mirada XD (Mirada Medical, Oxford, UK). The VOIs were defined as an automatically segmented 40% isocontour of SUV_{max}. Minor manual adjustments of the VOIs were done, e.g. excluding adjacent physiological uptake, if considered appropriate by the physician. The VOIs outlined on PET_A were projected onto to the corresponding PET_B without modifying the contours. Next, SUV_{mean} and SUV_{max} (body weight corrected) were assessed. The relative difference between the measured SUVs in PET_A versus PET_B was calculated as $\Delta\%SUV_{max}$ and $\Delta\%SUV_{max}$, always keeping PET_A as the reference.

A 10 cm diameter sphere was placed in $\mu\text{-map}_A$ and $\mu\text{-map}_B$ with the center located at the center-of-mass of each tumor. From these local spheres, the volumes of the different tissue compartments were assessed and the absolute differences between these volumes in $\mu\text{-map}_A$ and in $\mu\text{-map}_B$ were calculated for lung tissue (ΔV_{Lung}), fat (ΔV_{Fat}) and soft tissue (ΔV_{ST}).

Statistical Analyzes

Statistical analyzes were performed in R (R Development Core Team). First, both SUV_{mean} and SUV_{max} were transformed by the natural logarithm due to skewed distributions of their paired scan-rescan differences. To assess the components of variation for SUV_{mean} and SUV_{max}, a random-effects

one-way analysis of variance was conducted for the $\log_e$-transformed data with the specific examination as a random effect. From this we obtain the within-examination variation ($s_w{}^2$) and the between-examination variation ($s_b{}^2$). Next, the correlation among observations within the same examination, known as the intra-class correlation (ICC), and within-examination coefficient of variation (CV) was calculated according to (*27*) as:

$$\mathrm{ICC} = \frac{s_b{}^2}{s_b{}^2 + s_w{}^2} \quad (1)$$

$$\mathrm{CV} = \sqrt{\exp(s_w{}^2) - 1} \quad (2)$$

The scan-rescan SUV agreement was further assessed by the 95% limits of agreement (LOA) from the Bland-Altman analysis for repeated observations (*28*,*29*). The analysis was performed on $\log_e$-transformed SUV_{mean} and SUV_{max}, and the lower and upper LOA were back-transformed (by exponential e^x) into the original scale to represent ratios and expressed as percentage differences. The back-transformed lower and upper LOA were visualized graphically by Bland–Altman plots with the SUV percentage differences versus the SUV average for both SUV_{mean} and SUV_{max}.

Finally, linear regression analysis was used to assess the correlation between the difference variables: $\Delta\%SUV_{mean}$, $\Delta\%SUV_{max}$, ΔV_{Lung}, ΔV_{Fat}, and $\Delta V_{Soft\ tissue}$.

Results

In this study 11 patients were included. Each patient underwent several test-retest scans over a time span of a few days resulting in 90 MR-AC maps for visual evaluation. In total, 19 separate tumors where found in the 11 patients resulting in 80 tumors for reproducibility analysis.

Image artifacts were observed in 86% (77/90) of the MR-AC maps. The trachea artifact was most frequent and seen in 79% of the MR-AC maps, followed by erroneous tissue classification (34%), body contour artifacts (19%), metal artifacts (13%), respiration artifacts (11%), and lung border artifacts (10%).

Critical artifacts, subjectively expected to affect SUV, were present in 30% (27/90) of the MR-AC maps. Fig. 1 reports an overview of the MR-AC maps categorized as *critical* and *non-critical*, including a color-coding specifying the type of artifact causing a *critical* categorization. The information in Fig. 1 can be related to Fig. 2, where $\Delta\%SUV_{mean}$ and $\Delta\%SUV_{max}$ can be seen for each tumor of each examination of each patient. In relation to reproducibility it is of interest that there are nine examinations where the MR-AC map pair, $\mu\text{-map}_A$ and $\mu\text{-map}_B$, are categorized differently i.e. one *critical* and the other *non-critical*. The $\Delta\%SUV_{max}$ exceeds ±20% for four tumors (originating from four different repeated MR-AC map pairs in three different patients).

We consider all four cases where $\Delta\%SUV_{max}$ exceeds ±20% in detail:

Case 1: In the first examination of patient 9 (Fig. 3), $\mu\text{-map}_A$ was categorized as *non-critical* and $\mu\text{-map}_B$ was categorized as *critical* (lung border artifact). Obvious differences in what is classified as lung tissue can be seen in the paired MR-AC maps.

Case 2: In the fourth examination of patient 9 (Fig. 4), $\mu\text{-map}_A$ was categorized as *critical* (lung border artifact and erroneous tissue classification) and $\mu\text{-map}_B$ was categorized as *non-critical*. Obvious differences in the paired MR-AC maps can be seen and respiratory motion is apparent from the corresponding Dixon water images. Part of the lungs (red arrows) and trachea (blue arrows) are misclassified as soft tissue in $\mu\text{-map}_A$.

Case 3: In the first examination of patient 7 (Fig. 5), $\mu\text{-map}_A$ was categorized as *critical* (erroneous tissue classification) and $\mu\text{-map}_B$ was categorized as *non-critical*. Pronounced respiratory motion artifacts can be seen in the Dixon water image of $\mu\text{-map}_A$, which result in tissue misclassification in $\mu\text{-map}_A$.

Case 4: In the second examination of patient 3 (Fig. 6), both MR-AC maps were categorized as *non-critical*. Nevertheless, one of the three tumors had a large percentage difference with a $\Delta\%SUV_{max}$ of 27%. Through a joint inspection of the MR-AC maps and the corresponding Dixon water image, several different respiration states in the Dixon water image of $\mu\text{-map}_B$ become apparent. This was undetected by the physicians, as the evaluation of each MR-AC map was limited to MR-AC maps not including the Dixon water images.

The Bland-Altman plots with the percentage differences in SUV_{mean} and SUV_{max} are shown in Fig. 7. The back-transformed bias between the observations and the back-transformed lower and upper LOA are also shown in Fig. 7. Table 1 summarizes the results and depicts the lower and upper LOA, ICC, and CV for SUV_{mean} and SUV_{max}. The 95% limits of agreement are approximately ±20% for both SUV_{mean} and SUV_{max}.

The correlation matrix for the variables $\Delta\%SUV_{mean}$, $\Delta\%SUV_{max}$, ΔV_{Lung}, ΔV_{Fat}, and $\Delta V_{Soft\ tissue}$ is shown in Table 2. The relative SUV differences are seen to correlate with both ΔV_{Lung} and $\Delta V_{Soft\ tissue}$. Furthermore, ΔV_{Lung} and $\Delta V_{Soft\ tissue}$ have a strong negative correlation.

Discussion

To the best of our knowledge, this study is the first to examine how the reproducibility of thoracic MR-AC maps affects PET quantification by comparing PET image pairs (PET_A and PET_B) reconstructed with the same emission data but different intra-subject repeated MR-AC maps.

In general, we have found that the relative SUV variations were independent of the absolute values (Fig. 7) and both SUV_{mean} and SUV_{max} had a CV of approximately 6% and a scan-rescan SUV variation within ±20% in 95% of the cases using intra-subject repeated MR-AC map pairs. If we assume SUV_{max} to be similar to lean body mass corrected peak SUV (SUL_{peak}), then the observed LOA are within the proposed PERCIST criteria for a stable disease (change within ±30%) (*9*). However, these criteria are based on studies where SUV variations were assessed from two different PET acquisitions, thus the variation comes from several sources including differences in emission and attenuation data as well as biological variations. The only source of variation in our study comes from differences in the MR-AC maps alone, which adds to other well-known sources of variation in measuring SUV. Thus, the SUV variations observed in this study must be considered alarmingly high. Note that the distribution is skewed. This bias might be related to a systematic difference in the paired MR-AC maps, which for example could be caused by the patients being more relaxed during the later scan. For comparison, a meta-analysis by de Langen *et al.* reported LOA for day-to-day SUV variations for FDG PET-only and PET/CT of ±20% for SUV_{mean} and ±25% for SUV_{max}. Rasmussen *et al.* reported LOA of ±20% and CV of ~6% for day-to-day SUV variations in FDG PET/CT and PET/MRI in patients with head and neck squamous cell carcinoma (*12*).

Substantial SUV_{max} differences (exceeding ±20%) were observed in four different examinations. In three of the four cases, the paired MR-AC maps were categorized differently, which agrees with the expectation that large differences between MR-AC maps will result in large differences in SUV measures. This emphasizes that physicians must examine the MR-AC maps and be aware that SUV measures can be unreliable if artifacts and/or patient motion are present. On the other hand, of the nine examinations where the paired MR-AC maps were categorized differently, five of them had SUV_{max} differences of less than 10% and one had a difference of 18%. Furthermore, in the one case where $\Delta\%SUV_{max}$ exceeded 20% but no critical artifacts were detected, different states of the breathing cycle could be observed when comparing the paired MR-AC maps side-by-side or when inspecting the Dixon water image of $\mu\text{-map}_B$ (Fig. 5). This suggests that artifacts and respiratory motion affecting quantitative PET are difficult to identify when inspecting the corresponding MR-AC map only. Therefore, a joint inspection of the MR-AC map and the underlying Dixon images for artifacts and thus patient compliance with breath-hold instruction is recommended. Alternatively, acquisition of more than one MR-AC map per patient could be a safety precaution in case the first acquisition has obvious artifacts.

The majority of the large SUV differences were a consequence of respiratory motion, which led to a misclassification of lung tissue as soft tissue and vice versa in the MR-AC maps (Table 2). Thus, we propose that strategies to obtain highly reproducible MR-AC maps should focus on a reduction of movement during data acquisition. Furthermore, optimization of the segmentation algorithm deriving the MR-AC maps should be pursued.

In this study, 19 different tumors from 11 patients were investigated. Each patient underwent several test-retest scans over a time span of a few days, giving 80 tumors for analysis but with multiple repetitions of the same tumors. Therefore, multiple lesions within the same patient may have an undesirable correlated reproducibility; as may the same tumor from different days. We accounted for this using the Bland-Altman analysis for repeated observations (*28*,*29*). Nevertheless, it would be desirable to have a larger study population to generalize observations.

The method of projecting the VOIs from one PET image to the other instead of a recalculation of the isocontour may also affect our results. The projection method is more conservative and was applied because we wanted to quantify the effect of MR-AC reproducibility in a specific and fixed tumor volume, whereas recalculating the isocontour would change these quantitative values while giving insight to the morphological changes of the VOI such as the volume.

Furthermore, it should be mentioned that for studies where repeated PET scans are performed, the allocation of predetermined attenuation coefficients to specific tissues not only introduces a bias but is also a potential source of variation as the attenuation in lung can vary dramatically depending on disease and respiration pattern. However, in the presented study we used the same raw PET data but applied different intra-subject repeated MR-AC maps, which were acquired without repositioning the patient. Thus the only variation in the resulting PET images will be that arising from the differences in the MR-AC maps and not a consequence of the assumption of a fixed single attenuation value.

Finally, while new methods for deriving improved MR-AC map including bone representatives is under intense research (*23*,*30*), a reproducibility study of such methods should also be conducted.

Conclusion

In the present study we have found that artifacts occur frequently in standard thoracic MR-AC maps affecting the reproducibility of PET/MRI by resulting in large SUV variations. This, in combination with other well-known sources of error associated with PET/MRI examinations, leads to inconsistent SUV measurements in serial studies, which may affect the reliability of therapy response assessment. Therefore, a thorough visual inspection of thoracic MR-AC maps and the corresponding unsegmented Dixon images remain crucial for the detection of artifacts and errors influencing the reproducibility of SUV in test-retest settings. Based on our study, especially a reduction of respiratory movement during data acquisition is important for improving MR-AC map reproducibility, possibly by shortening the acquisition time or by gated data acquisition.

Disclosure

All authors declare that they have no conflict of interest.

Acknowledgements

We thank the John and Birthe Meyer Foundation who donated the PET/MRI system.

References

1. Pinker K, Riedl C, Weber WA. Evaluating tumor response with FDG PET: updates on PERCIST, comparison with EORTC criteria and clues to future developments. *Eur J Nucl Med Mol Imaging*. 2017;44:1-12.
2. Krak NC, van der Hoeven JJM, Hoekstra OS, Twisk JWR, van der Wall E, Lammertsma AA. Measuring [18F]FDG uptake in breast cancer during chemotherapy: comparison of analytical methods. *Eur J Nucl Med Mol Imaging*. 2003;30:674-681.
3. Minn H, Zasadny KR, Quint LE, Wahl RL. Lung cancer: reproducibility of quantitative measurements for evaluating 2-[F-18]-fluoro-2-deoxy-D-glucose uptake at PET. *Radiology*. 1995;196:167-173.
4. Nahmias C, Wahl LM. Reproducibility of standardized uptake value measurements determined by 18F-FDG PET in malignant tumors. *J Nucl Med*. 2008;49:1804-1808.
5. Nakamoto Y, Zasadny KR, Minn H, Wahl RL. Reproducibility of common semi-quantitative parameters for evaluating lung cancer glucose metabolism with positron emission tomography using 2-deoxy-2-[18F]fluoro-D-glucose. *Mol Imaging Biol*. 2002;4:171-178.
6. Hoekstra CJ, Hoekstra OS, Stroobants SG, et al. Methods to monitor response to chemotherapy in non-small cell lung cancer with 18F-FDG PET. *J Nucl Med*. 2002;43:1304-1309.
7. Weber WA, Ziegler SI, Thödtmann R, Hanauske AR, Schwaiger M. Reproducibility of metabolic measurements in malignant tumors using FDG PET. *J Nucl Med*. 1999;40:1771-1777.
8. Velasquez LM, Boellaard R, Kollia G, et al. Repeatability of 18F-FDG PET in a multicenter phase I study of patients with advanced gastrointestinal malignancies. *J Nucl Med*. 2009;50:1646-1654.

9. Wahl RL, Jacene H, Kasamon Y, Lodge MA. From RECIST to PERCIST: evolving considerations for PET response criteria in solid tumors. *J Nucl Med*. 2009;50:122S-150S.

10. Weber WA, Gatsonis CA, Mozley PD, et al. Repeatability of 18F-FDG PET/CT in advanced non-small cell lung cancer: prospective assessment in 2 multicenter trials. *J Nucl Med*. 2015;56:1137-1143.

11. de Langen AJ, Vincent A, Velasquez LM, et al. Repeatability of 18F-FDG uptake measurements in tumors: a metaanalysis. *J Nucl Med*. 2012;53:701-708.

12. Rasmussen JH, Fischer BM, Aznar MC, et al. Reproducibility of FDG PET uptake measurements in head and neck squamous cell carcinoma on both PET/CT and PET/MR. *Br J Radiol*. 2015;88:20140655.

13. Dixon WT. Simple proton spectroscopic imaging. *Radiology*. 1984;153:189-194.

14. Martinez-Möller A, Souvatzoglou M, Delso G, et al. Tissue classification as a potential approach for attenuation correction in whole-body PET/MRI: evaluation with PET/CT data. *J Nucl Med*. 2009;50:520-526.

15. Samarin A, Burger C, Wollenweber SD, et al. PET/MR imaging of bone lesions - Implications for PET quantification from imperfect attenuation correction. *Eur J Nucl Med Mol Imaging*. 2012;39:1154-1160.

16. Bailey DL, Barthel H, Beuthin-Baumann B, et al. Combined PET/MR: Where are we now? Summary report of the second international workshop on PET/MR imaging April 8-12, 2013, Tubingen, Germany. *Mol Imaging Biol*. 2014;16:295-310.

17. Rausch I, Rust P, Difranco MD, et al. Reproducibility of MRI Dixon-based attenuation correction

in combined PET/MR with applications for lean body mass estimation. *J Nucl Med*. 2016;57:1096-1102.

18. Beyer T, Lassen ML, Boellaard R, et al. Investigating the state-of-the-art in whole-body MR-based attenuation correction: an intra-individual, inter-system, inventory study on three clinical PET/MR systems. *MAGMA*. 2016;29:75-87.
19. Keller SH, Holm S, Hansen AE, et al. Image artifacts from MR-based attenuation correction in clinical, whole-body PET/MRI. *MAGMA*. 2013;26:173-181.
20. Schramm G, Langner J, Hofheinz F, et al. Influence and compensation of truncation artifacts in MR-based attenuation correction in PET/MR. *IEEE Trans Med Imaging*. 2013;32:2056-2063.
21. Ladefoged CN, Hansen AE, Keller SH, et al. Impact of incorrect tissue classification in Dixon-based MR-AC: fat-water tissue inversion. *EJNMMI Phys*. 2014;1:101.
22. Brendle C, Schmidt H, Oergel A, et al. Segmentation-based attenuation correction in positron emission tomography/magnetic resonance. *Invest Radiol*. 2015;50:339-346.
23. Paulus DH, Quick HH, Geppert C, et al. Whole-body PET/MR imaging: quantitative evaluation of a novel model-based MR attenuation correction method including bone. *J Nucl Med*. 2015;56:1061-1066.
24. Biederer J, Beer M, Hirsch W, et al. MRI of the lung (2/3). Why... when ... how? *Insights Imaging*. 2012;3:355-371.
25. Puderbach M, Hintze C, Ley S, Eichinger M, Kauczor HU, Biederer J. MR imaging of the chest: a practical approach at 1.5 T. *Eur J Radiol*. 2007;64:345-355.
26. Oken MM, Creech RH, Tormey DC, et al. Toxicity and response criteria of the Eastern

Cooperative Oncology Group. *Am J Clin Oncol*. 1982;5:649-656.

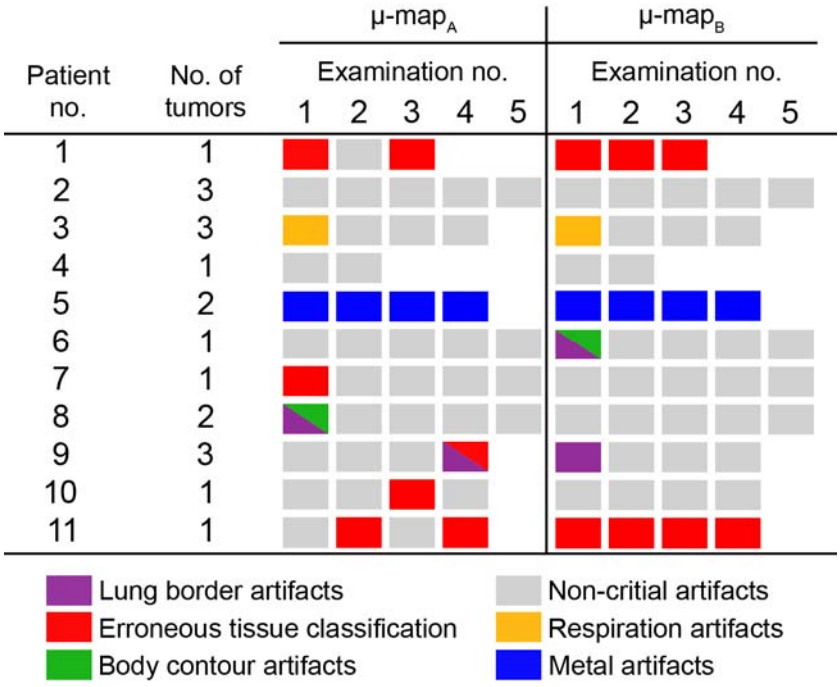
27. Raunig DL, McShane LM, Pennello G, et al. Quantitative imaging biomarkers: a review of statistical methods for technical performance assessment. *Stat Methods Med Res*. 2015;24:27-67.

28. Bland JM, Altman DG. Measuring agreement in method comparison studies. *Stat Methods Med Res*. 1999;8:135-160.

29. Bland JM, Altman DG. Agreement between methods of measurement with multiple observations per individual. *J Biopharm Stat*. 2007;17:571-582.

30. Ladefoged CN, Law I, Anazodo U, et al. A multi-centre evaluation of eleven clinically feasible brain PET/MRI attenuation correction techniques using a large cohort of patients. *Neuroimage*. 2016;147:346-359.

FIGURE 1: Overview of the categorization of all 90 MR-AC maps including a color-coding of the type of artifact causing the categorization. Empty fields indicate that no examination was performed, while gray squares indicate MR-AC maps categorized as *non-critical*. All other colored squares indicate MR-AC maps categorized as *critical.*



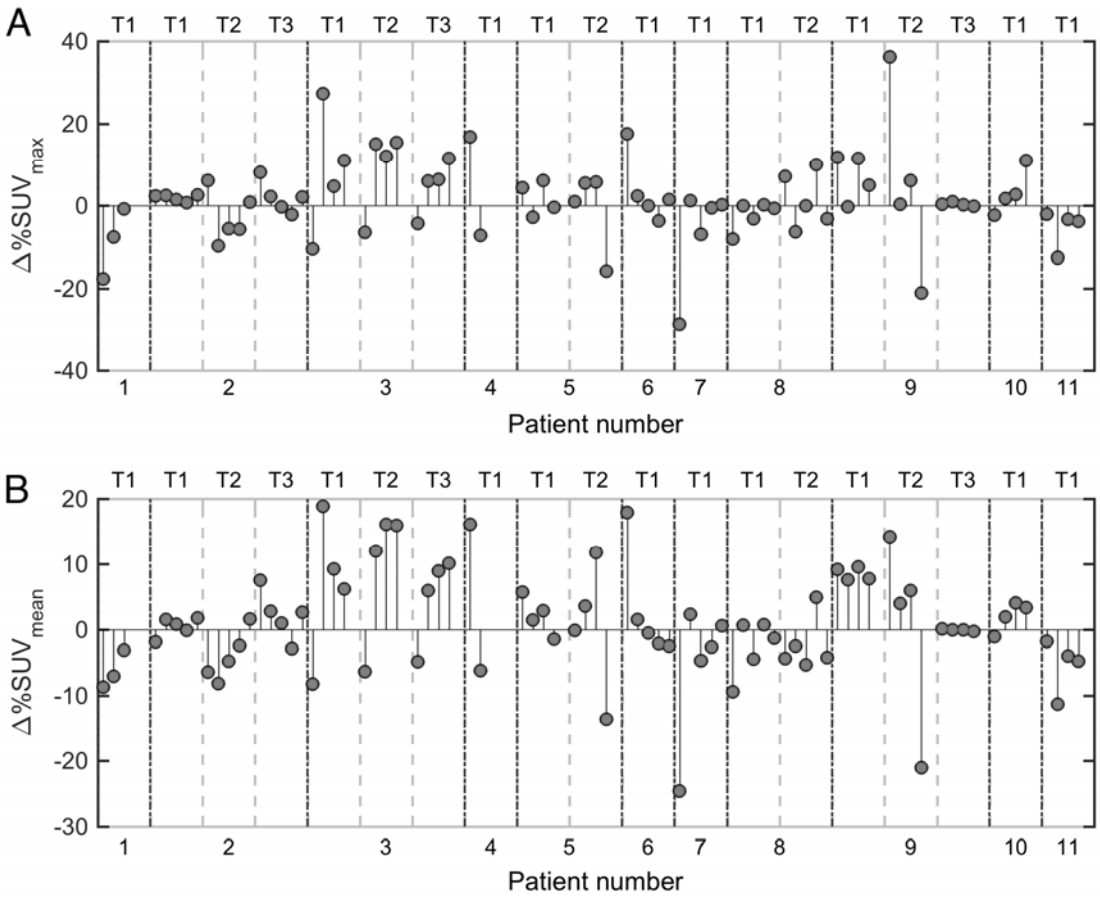


FIGURE 2: Overview of the $\Delta\%SUV_{max}$ (A) and $\Delta\%SUV_{mean}$ (B) for each tumor of each examination of each patient. The x-axis depicts the patient number (separated by the vertical black dashed lines). Each patient has up to three tumors (T1, T2, T3; separated by the vertical gray dashed lines). For each tumor the relative scan-rescan SUV difference is display on the y-axis for each examination (up to five examinations). This information can be compared to that of Fig. 1.

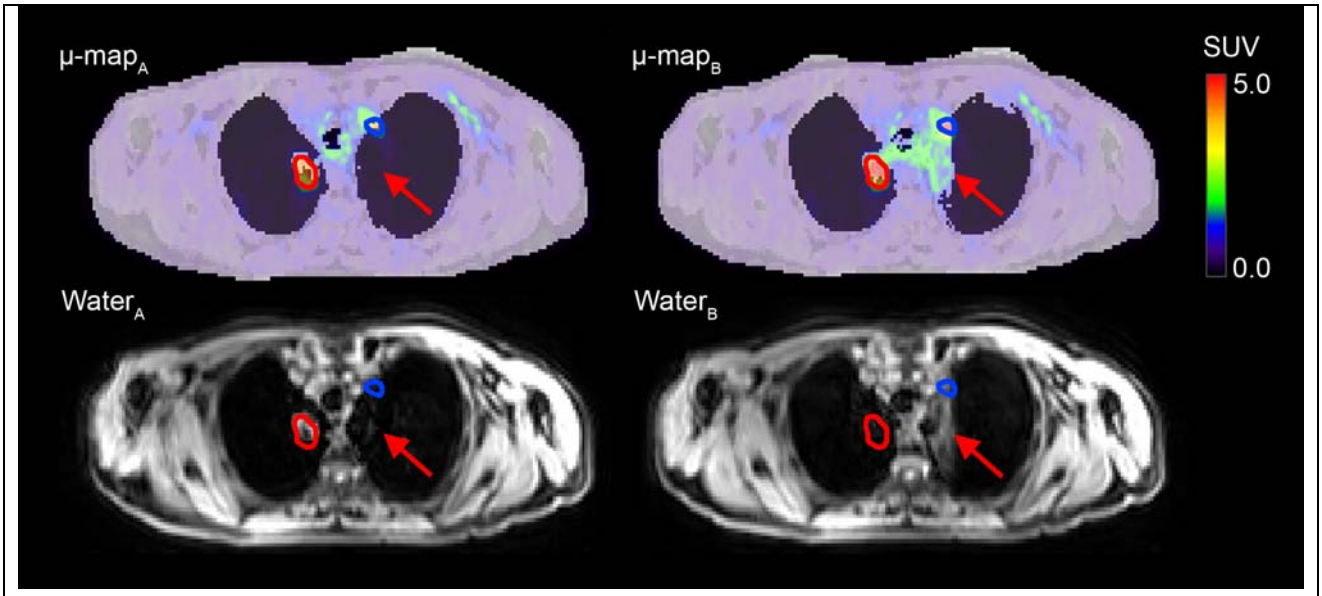


FIGURE 3: Patient 9, examination 1. Intra-subject repeated MR-AC maps (μ-map$_A$ and μ-map$_B$) fused with PET$_A$ and PET$_B$, along with their respective Dixon water images. Δ%SUV$_{max}$ of 12% (tumor delineated in red), 36% (tumor delineated in blue). Red arrows indicate noteworthy differences in the paired MR-AC maps and the Dixon water images.

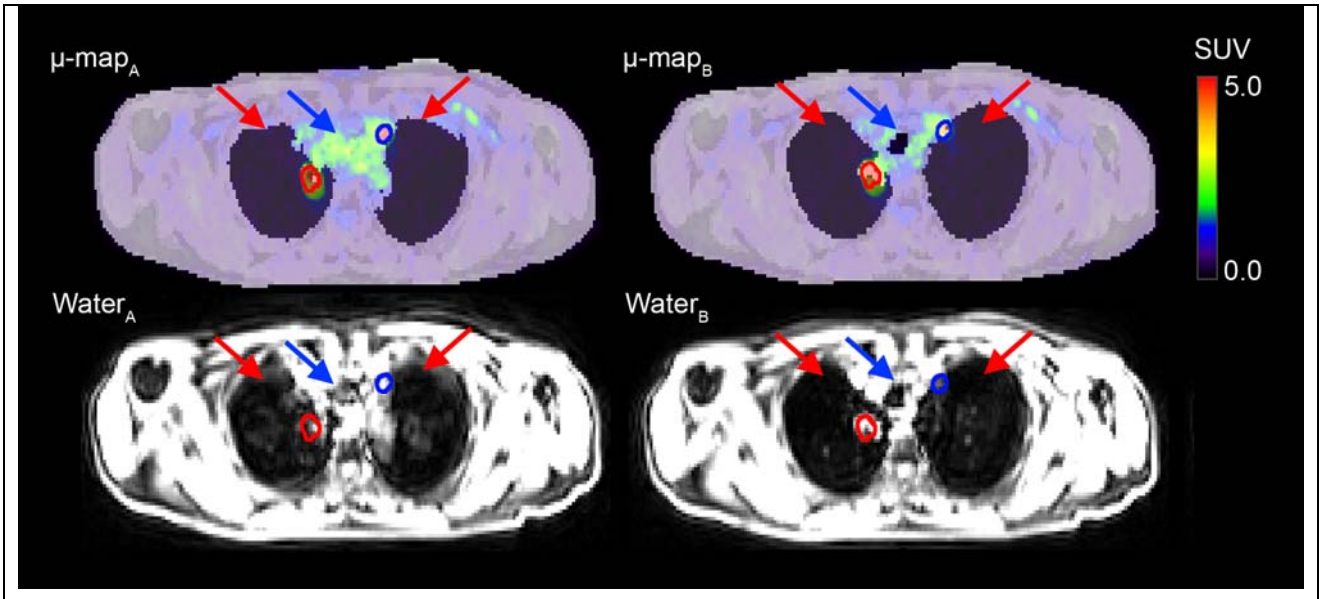


FIGURE 4: Patient 9, examination 4. Intra-subject repeated MR-AC maps (μ-map$_A$ and μ-map$_B$) fused with PET$_A$ and PET$_B$, along with their respective Dixon water images. Respiratory motion causes incorrect tissue classification in μ-map$_A$. The trachea is completely missed (blue arrows) and an incorrect classification of lung tissue as soft tissue can be seen (red arrows). As a result PET$_B$ is lower relative to PET$_A$ for the tumor delineated in blue (Δ%SUV$_{max}$ of -21%) as PET$_A$ is locally overcorrected for attenuation (SUV values too high). The tumor delineated in red is in μ-map$_A$ incorrectly classified as lung tissue, which causes undercorrection for attenuation in PET$_A$ and causes PET$_B$ to be slightly higher relative to PET$_A$ (Δ%SUV$_{max}$ of 5%).

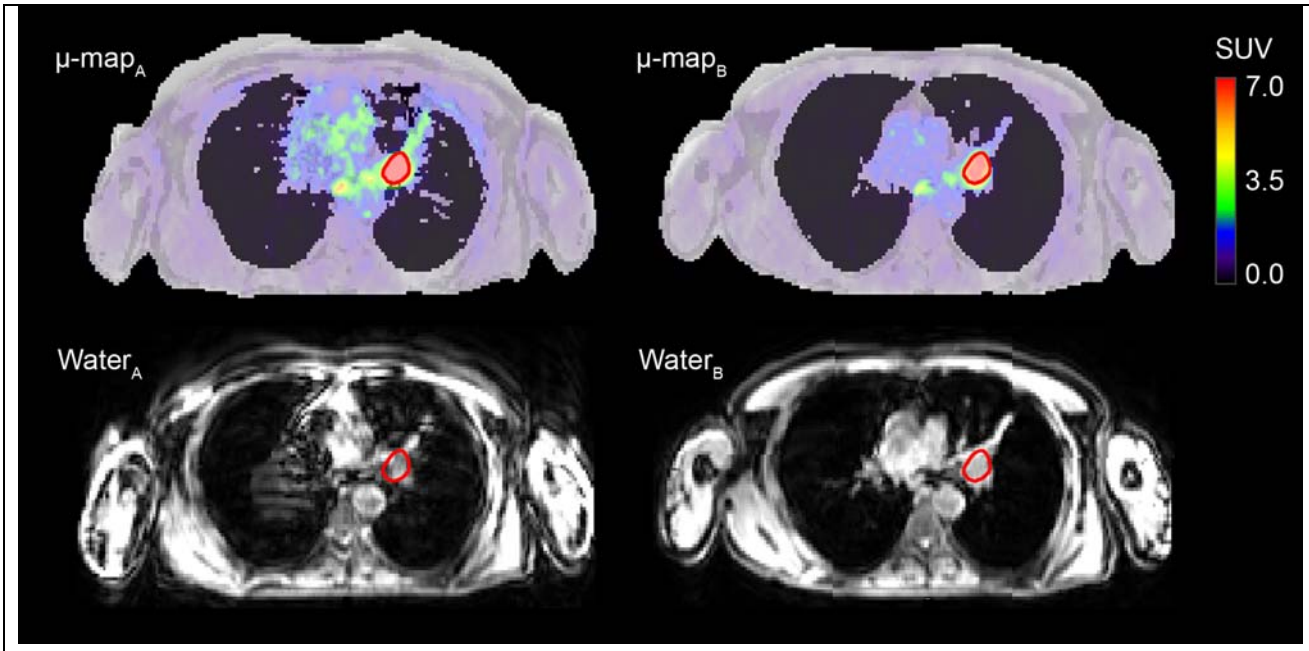


FIGURE 5: Patient 7, examination 1. Intra-subject repeated MR-AC maps (μ-map$_A$ and μ-map$_B$) fused with PET$_A$ and PET$_B$, along with their respective Dixon water images. A Δ%SUV$_{max}$ of -29% is found in the tumor delineated in red. Respiratory motion artifacts cause erroneous tissue classification in μ-map$_A$.

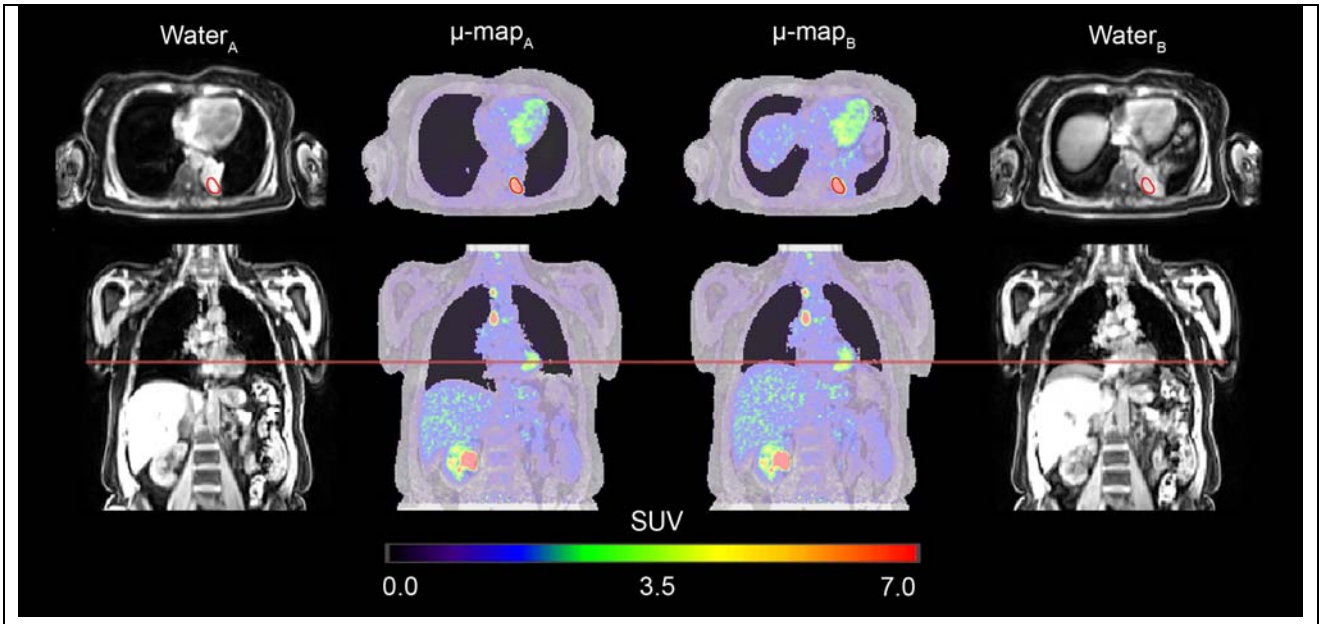


FIGURE 6: Patient 3, examination 2. Intra-subject repeated MR-AC maps (μ-map$_A$ and μ-map$_B$) fused with PET$_A$ and PET$_B$, along with their respective Dixon water images. Initially, no artifacts were found to have an impact on the tumor SUV based on the visual evaluation of the MR-AC maps alone. However, from the corresponding Dixon water image of μ-map$_B$ respiration artifacts are apparent. A comparison of μ-map$_A$ and μ-map$_B$ shows a partially elevated diaphragm in μ-map$_B$ (red line) causing a Δ%SUV$_{max}$ of 27%.

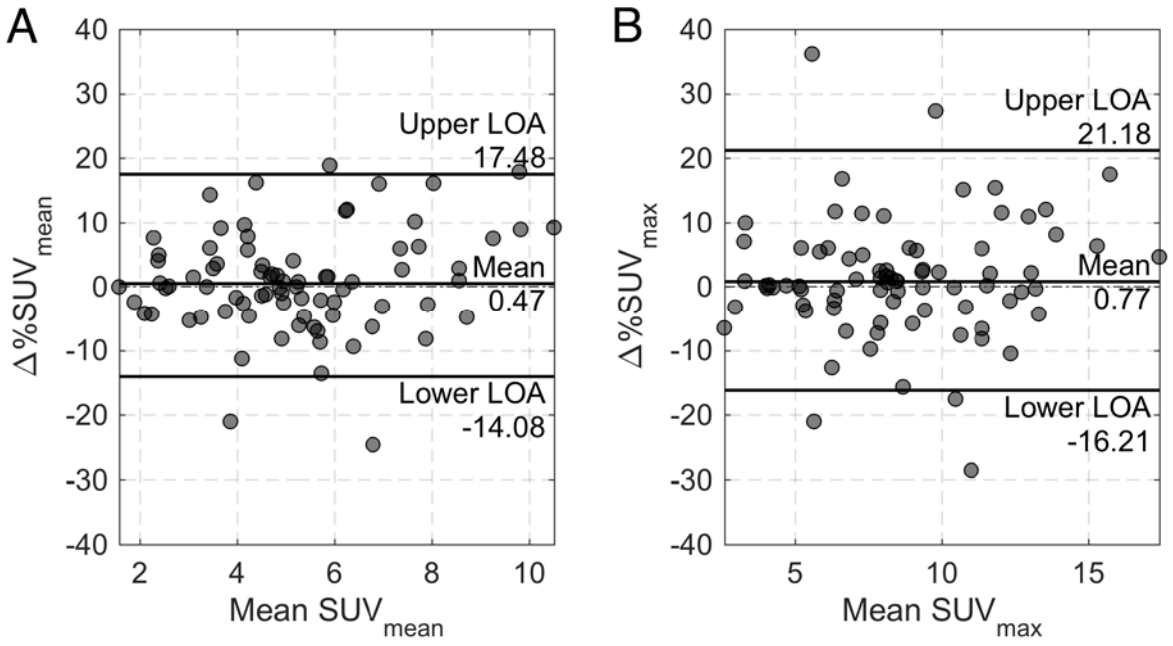


FIGURE 7: Bland-Altman plots showing the relative differences in SUV versus the average SUVs. The upper and lower LOA were calculated for the log-transformed SUV_{mean} (A) and log-transformed SUV_{max} (B) in which multiple observations for each patient were accounted for.

TABLE 1: The ICC and CV for both the SUV_{mean} and SUV_{max}. The lower and upper LOA are a result of the Bland-Altman analysis for repeated observations. All analyses were performed after a log-transformation and the CV and LOA are presented here after back-transformation.

Uptake measurement	ICC	CV (%)	95% LOA (%)	
			Lower	Upper
SUV_{mean}	0.982	5.6	-14.08	17.48
SUV_{max}	0.975	6.6	-16.21	21.18

TABLE 2: Correlation matrix for $\Delta\%SUV_{mean}$, $\Delta\%SUV_{max}$ and selected variables measured within a 10 cm sphere encapsulating the center-of-mass of the tumor. *Significant correlation with P<0.01.

	Variable				
Variable	$\Delta\%SUV_{mean}$	$\Delta\%SUV_{max}$	ΔV_{Lung}	ΔV_{Fat}	$\Delta V_{Soft\ tissue}$
$\Delta\%SUV_{mean}$	1	0.9*	-0.71*	0.03	0.70*
$\Delta\%SUV_{max}$	-	1	-0.72*	0.12	0.70*
ΔV_{Lung}	-	-	1	-0.17	-0.96*
ΔV_{Fat}	-	-	-	1	0.07
$\Delta V_{Soft\ tIssue}$	-	-	-	-	1