Measurement Repeatability of $^{18}$F-FDG-PET/CT versus $^{18}$F-FDG-PET/MRI in Solid Tumors of the Pelvis

Tyler J. Fraum, MD$^1$; Kathryn J. Fowler, MD$^1$; John P. Crandall$^1$; Richard A. Laforest, PhD$^1$; Amber Salter, PhD$^2$; Hongyu An, PhD$^1$; Michael A. Jacobs, PhD$^3$; Perry W. Grigsby, MD$^{4,5}$; Farrokh Dehdashti, MD$^{1,5}$; & Richard L. Wahl, MD$^{1,5}$

$^1$Mallinckrodt Institute of Radiology; Washington University School of Medicine; St. Louis, MO

$^2$Division of Biostatistics; Washington University School of Medicine; St. Louis, MO

$^3$Russell H. Morgan Department of Radiology and Radiological Science and Sidney Kimmel Comprehensive Cancer Center; Johns Hopkins School of Medicine; Baltimore, MD

$^4$Department of Radiation Oncology; Washington University School of Medicine; St. Louis, MO

$^5$Siteman Cancer Center; Washington University School of Medicine; St. Louis, MO

First/corresponding author: Tyler J. Fraum, MD (clinical fellow); E-mail: fraumt@wustl.edu; Telephone: (314) 362-1474; Mailing address: 510 S. Kingshighway Blvd, Campus Box 8131, St. Louis, MO 63110

Disclaimer: none

Financial disclosure: This project was supported by NIH grants U01CA140204, 5P30CA006973 (Imaging Response Assessment Team - IRAT), and 1R01CA190299. No potential conflicts of interest relevant to this article exist.

Word count: 4979

Running title: PET/MRI Repeatability of Pelvic Tumors

Immediate Open Access: Creative Commons Attribution 4.0 International License (CC BY) allows users to share and adapt with attribution, excluding materials credited to previous publications.

License: https://creativecommons.org/licenses/by/4.0/.

Details: http://jnm.snmjournals.org/site/misc/permission.xhtml.
ABSTRACT

Background: Knowledge of the within-subject variability of $^{18}$F-FDG-PET/MRI measurements is necessary for proper interpretation of quantitative PET or MRI metrics in the context of therapeutic efficacy assessments with integrated PET/MRI scanners. The goal of this study was to determine the test-retest repeatability of these metrics on PET/MRI, with comparison to similar metrics acquired by PET/CT. Methods: This prospective study enrolled patients with pathology-proven pelvic malignancies. Baseline imaging consisted of PET/CT immediately followed by PET/MRI, utilizing a single 370 MBq $^{18}$F-FDG dose. Repeat imaging was performed within 7 days using an identical imaging protocol, with no oncologic therapy between sessions. PET imaging on both scanners consisted of a list-mode acquisition at a single pelvic station. The MRI consisted of two-point Dixon imaging for attenuation correction, standard sequences for anatomic correlation, and diffusion-weighted imaging (DWI). PET data were statically reconstructed utilizing various frame durations and minimizing uptake time differences between sessions. Standard uptake value (SUV) metrics were extracted for both PET/CT and PET/MRI in each imaging session. Apparent diffusion coefficient (ADC) metrics were extracted for both PET/MRI sessions.

Results: The study cohort consisted of 14 subjects (13 female, 1 male) with various pelvic cancers (11 cervical, 2 rectal, 1 endometrial). The repeatability of the maximum SUV (SUV$_{\text{max}}$), as reflected by the within-subject coefficient of variation (wCV), was higher for PET/CT (8.5-12.8%) than PET/MRI (6.6-8.7%) across all PET reconstructions, though with no significant repeatability differences (all $p$ values $\geq 0.08$) between modalities. For lean body mass-adjusted peak SUV (SUL$_{\text{peak}}$), the wCVs were similar for PET/CT (9.9-11.5%) and PET/MRI (9.2-11.3%) across all PET reconstructions, again with no significant repeatability differences (all $p$ values $\geq 0.14$) between modalities. For PET/MRI, the wCV for median ADC (ADC$_{\text{median}}$) of 3.9% was lower than the wCVs for SUV$_{\text{max}}$ (6.6-8.7%) and SUL$_{\text{peak}}$ (9.2-11.3%), though without significant repeatability differences (all $p$ values $\geq 0.23$). Conclusion: For solid tumors of the pelvis, the repeatability of the evaluated SUV and ADC metrics on $^{18}$F-FDG-PET/MRI is both acceptably high and similar to previously published values for $^{18}$F-FDG-PET/CT and MRI, supporting the utilization of $^{18}$F-FDG-PET/MRI for quantitative oncologic treatment response assessments.

Keywords: PET/MRI, cervical cancer, repeatability, quantitative imaging, diffusion-weighted imaging, apparent diffusion coefficient
INTRODUCTION

Semi-quantitative assessments of $\text{^{18}F}$-fluoro-2-deoxy-D-glucose ($\text{^{18}F}$-FDG) uptake on positron emission tomography (PET) / computed tomography (CT) with metrics such as the standardized uptake value (SUV) are valuable tools for oncologic response assessment (1). Knowledge of measurement variability is essential to the interpretation of longitudinal changes in such metrics. Prior PET/CT studies have shown that repeated measurements of $\text{^{18}F}$-FDG uptake are highly repeatable (2). However, $\text{^{18}F}$-FDG-PET/CT has important limitations. For example, the low soft tissue contrast of CT impedes primary tumor staging, whereas high background $\text{^{18}F}$-FDG uptake in some organs may reduce the conspicuity of metastases (3). Consequently, many patients also undergo magnetic resonance imaging (MRI) to improve staging accuracy.

Simultaneous PET/MRI systems can provide whole-body staging and treatment response assessment in a single examination. These hybrid systems have necessitated the development of MRI-based methods for attenuation correction (4). MRI-based attenuation correction maps are often affected by artifacts that can vary between imaging sessions, potentially reducing the repeatability of PET quantitation (5). In contrast, the CT-based attenuation correction approach employed by PET/CT entails the direct measurement of photon attenuation by tissues, a method that is uncommonly affected by serious artifacts. The repeatability of PET metrics for PET/MRI relative to PET/CT is largely unknown.

Furthermore, PET/MRI permits the assessment of cellular density with diffusion-weighted imaging (DWI). In studies of MRI alone, quantitative metrics such as the apparent diffusion coefficient (ADC) have been used to track response to treatment or predict clinical outcomes (6,7). The repeatability of ADC metrics for integrated PET/MRI systems is unknown and might be different than for MRI only systems. For example, prior studies have suggested that the addition of PET hardware to MRI systems can worsen DWI artifacts related to eddy currents, creating the potential for greater variability in ADC values between imaging sessions (8).

Thus, the primary aim of this study was to determine the test-retest repeatability of several commonly used PET/MRI-based quantitative imaging metrics in patients with solid malignancies of the pelvis. The specific metrics of interest were maximum SUV (SUV$_{\text{max}}$), peak lean body mass-adjusted SUV (SUL$_{\text{peak}}$), and median ADC (ADC$_{\text{median}}$). The repeatability of other PET and MRI metrics of potential
clinical interest was also assessed in an exploratory analysis, the results of which are presented separately. A secondary aim was to evaluate the impact of various PET reconstruction techniques and frame durations on the repeatability of the above-described PET metrics.

MATERIALS AND METHODS

Subjects

This prospective study (ClinicalTrials.gov Identifier: NCT02717572) was approved by our Institutional Review Board. Inclusion/exclusion criteria are listed in Table 1. From June 2016 through May 2017, 17 subjects were enrolled. All subjects provided written informed consent. Two patients were excluded because of failure to complete the second imaging session, and one was excluded because of lack of 18F-FDG uptake presumably due to prior therapy.

Study Protocol

Subjects underwent two imaging sessions separated by 1-7 days without any intervening oncologic treatments (Fig. 1). Immediately prior to the PET/MRI, patients received 1 mg of intravenous glucagon to minimize artifacts related to bowel motility.

PET/CT

All subjects were imaged on a Siemens Biograph 40 TruePoint/TrueView PET/CT scanner (Siemens AG; Erlangen, Germany). List-mode PET data were collected for 15 min for a single pelvic station (tumor centered craniocaudally within 21.6 cm z-axis field-of-view), starting approximately 60 min after the intravenous injection of 370 MBq 18F-FDG dose. A low-dose CT was acquired for anatomic correlation and attenuation as follows: 50 mAs (effective; CareDose™ tube current modulation), 120 kVp, 0.8 pitch, 0.5 sec rotation time.

PET/MRI

Immediately after PET/CT, all subjects were imaged on a Siemens Biograph mMR PET/MRI scanner, software version VA40 (Siemens AG; Erlangen, Germany). List-mode PET data were collected
for 30 min (longer than for PET/CT due to duration of the MRI component) for a single pelvic station (tumor centered craniocaudally within 25.6 cm z-axis field-of-view). To minimize patient exclusion due to artifacts, DWI was performed twice during each session using the acquisition parameters in Supplemental Table 1. ADC maps were generated by the VA40 console software.

PET Reconstructions

For each session, PET reconstructions derived from 1 min, 3 min, and 5 min of list-mode data were performed for both PET/CT and PET/MRI. For each patient, variable intervals were skipped at the beginning of each list-mode acquisition to achieve identical effective uptake times for PET/CT between sessions 1 and 2 and for PET/MRI between sessions 1 and 2 (Fig. 1). The reconstruction intervals were also selected to minimize uptake time differences between PET/CT and PET/MRI. For each frame duration, static images were generated using an ordered-subset expectation maximization reconstruction (hereafter called OSEM) and an OSEM with point-spread function reconstruction (hereafter called PSF), the latter of which was utilized to improve spatial resolution. Image reconstruction parameters are shown in Supplemental Table 2.

Image Analysis

MIM version 6.7 (MIM Software; Cleveland, OH) was used for image analysis. For each PET session, the lesion of interest was manually delineated, based on the perceived boundary between tumor and background, to a generate a whole tumor contour that was propagated to all 6 reconstructions (1 min, 3 min, 5 min; OSEM and PSF).

For each whole tumor contour, multiple PET metrics were extracted. The peak was defined as the highest mean value achievable for a 1 cm³ sphere placed within the lesion contour. Notably, MIM software employs the James formula for calculating the lean body mass values needed for SUL computation. For the exploratory analysis, metrics were also extracted from a 40% isocontour, which contains all voxels with an SUV ≥ 40% of the SUV_max in the whole tumor contour, to determine the effects of semi-automated lesion segmentation on repeatability. The 40% isocontour was selected based on its correlation with metabolic tumor volume in prior PET/MRI studies of cervical cancer (9).
For each DWI acquisition, the lesion of interest was manually delineated on the corresponding ADC map, based on the perceived boundary between tumor and background diffusion properties. For each ADC contour, multiple metrics were extracted. For some acquisitions, the images were unusable due to severe artifacts related to incomplete fat suppression. For one patient, all four DWI acquisitions were inadequate. For another patient, both DWI acquisitions in the second imaging session were inadequate. Consequently, only 12 patients could be included in the ADC repeatability analysis.

**Statistical Analysis**

Subject characteristics were summarized descriptively using means ± standard deviations for continuous metrics. The Wilcoxon signed-rank test was used to assess for differences in metric values between sessions 1 and 2.

The repeatability analysis was performed according to the methods of Bland and Altman (10). Percent differences (%Δ) in measurements between sessions were used instead of absolute differences. For each metric, the standard deviation (SD) of the distribution of %Δ values from all subjects was calculated. The within-subject coefficient of variation (wCV) was defined as $wCV = SD / \sqrt{2}$. The repeatability coefficient (RC) was defined as $RC = 1.96 \cdot SD$. The %Δ between measurements is expected to fall within one RC of the mean %Δ (in either direction) in approximately 95% of cases.

Normality of the %Δ distributions was assessed via visual inspection of Q-Q plots. Natural log transformation was attempted for metrics with non-normal distributions, as repeatability statistics derived from natural log-transformed data are directly interpretable as non-transformed relative difference repeatability statistics (11). These transformations did not successfully normalize the distributions of the non-normal metrics (data not shown). However, as stated by Bland and Altman, deviations from normality are generally not problematic in the setting of repeatability analyses, in contrast to other areas of statistics (11). Consequently, RCs were generated from the non-normal metrics without transformation, as denoted in the appropriate tables.

The Wilcoxon signed-rank test (paired data) or Mann-Whitney U test (unpaired data) was used to evaluate for significant differences in repeatability. Provided that there are no systematic differences in measurements between sessions, the mean %Δ will be approximately 0% even if there are large %Δ
values, provided that those %Δ values are randomly distributed in the positive and negative directions. In contrast, the mean |%Δ| (i.e., the mean of the absolute values of the %Δ values) will be 0% only with perfect repeatability, with larger values indicating greater magnitudes of %Δ between sessions (irrespective of whether these changes are increases or decreases). On the basis of clinical interest/relevance, a limited subset of metrics was selected for pair-wise comparison via the mean |%Δ|, as the total number of possible pairs was prohibitively large.

Because of the large number of statistical tests, the methods of Benjamini and Hochberg were used to achieve a false discovery rate of 5% (12). This correction was performed separately for the primary and exploratory aims, with statistical significance defined as \( p \leq 0.01 \) and \( p \leq 0.007 \), respectively. All statistical analysis was performed in R version 3.4.

RESULTS

Subject Characteristics

The final study cohort consisted of 14 subjects with pelvic tumors, with characteristics summarized in Table 2. All patients but one were imaged at the time of initial cancer diagnosis. This previously treated patient had a pelvic recurrence of rectal adenocarcinoma. Mean serum glucose levels for sessions 1 and 2 were 96 mg/dl and 97 mg/dl, respectively.

PET Metrics on PET/CT versus PET/MRI

There were systematic differences in 18F-FDG uptake times for PET/CT versus PET/MRI (Fig. 1). The mean end-point for the PET/CT reconstructions was 75.1 min (range, 75-77 min) post-injection. The mean start-point for the PET/MRI reconstructions was 95.8 min (range, 88-105 min) post-injection.

Supplemental Table 3 shows mean values of SUV\(_{\text{max}}\) and SUL\(_{\text{peak}}\) on PET/CT versus PET/MRI for OSEM and PSF reconstructions (see Supplemental Table 4 for exploratory PET metrics). On average, the values of all PET metrics except metabolic tumor volume were numerically lower on PET/MRI than PET/CT, regardless of frame duration. These differences were statistically significant for 7 of 12 metrics (58.3%; 5 from OSEM, 2 from PSF) in the primary analysis and 12 of 72 metrics (76.5%; 11 from OSEM, 2 from PSF) in the exploratory analysis.
ADC Metrics for PET/MRI Session 1 versus Session 2

Supplemental Table 5 shows the mean values of ADC$_{\text{median}}$ (see Supplemental Table 6 for exploratory ADC metrics) on PET/MRI for both imaging sessions. Mean values were nearly identical between sessions for all ADC metrics. Notably, there was no significant difference in diffusional tumor volume between sessions ($p = 0.64$).

Repeatability of PET and ADC Metrics

Supplemental Table 7 shows repeatability results for SUV$_{\text{max}}$ and SUL$_{\text{peak}}$ on PET/CT and PET/MRI for the OSEM and PSF reconstructions (see Supplemental Table 8 for exploratory PET metrics). Relative differences between sessions, as reflected by the mean $\%\Delta$ (range: -9.8%, 6.0%), were small for all metrics. For a given metric, wCVs were generally similar across different reconstruction intervals (i.e., 1 min, 3 min, 5 min) and algorithms (i.e., OSEM, PSF). Repeatability results from the 3 min PSF reconstructions are shown in the form of Bland-Altman plots for SUV$_{\text{max}}$ (Fig. 2) and SUL$_{\text{peak}}$ (Fig. 3). The PET image analysis for a subject with excellent SUL$_{\text{peak}}$ repeatability is also shown (Fig. 4).

Supplemental Table 9 shows repeatability results for ADC$_{\text{median}}$ on PET/MRI (see Supplemental Table 10 for exploratory ADC metrics). Relative differences between sessions, as reflected by the mean $\%\Delta$ (range: -2.7%, 0.5%), were small for all metrics. Interestingly, ADC metric wCVs were generally lower than PET metric wCVs. Repeatability results for ADC$_{\text{median}}$ are shown in the form of a Bland-Altman plot (Fig. 5). The ADC map analysis for a subject with excellent ADC$_{\text{median}}$ repeatability is also shown (Fig. 6).

Supplemental Table 11 shows the results of pairwise repeatability comparisons of SUV$_{\text{max}}$, SUL$_{\text{peak}}$, and ADC$_{\text{median}}$ (see Supplemental Table 12 for pairwise repeatability comparisons of exploratory PET and ADC metrics). Notably, for many of the comparisons of PET/CT to PET/MRI, the mean $|\%\Delta|$ was numerically lower on PET/MRI though without statistical significance. There were also no significant differences in mean $|\%\Delta|$ for the PET versus ADC metric pairs evaluated.

DISCUSSION

We have demonstrated the repeatability of various quantitative PET and MRI metrics on simultaneous PET/MRI. Our results point to the robustness of PET/MRI for utilization in clinical trials with
quantitative end-points, incorporation into treatment response assessment algorithms, and (eventually) prediction of tumor biology and clinical outcomes. Our focus on pelvic tumors reflects the expected value of PET/MRI for such malignancies based on prior studies of cervical (13–15) and rectal cancer (16). Furthermore, although other studies have assessed the repeatability of ADC metrics on MRI platforms, our study is the first (to our knowledge) to address DWI repeatability on simultaneous PET/MRI. The repeatability of DWI on PET/MRI is especially important to establish given the potential clinical value of PET/MRI-based biomarkers incorporating both SUV and ADC data (17).

With respect to the magnitude of PET metrics, values were generally lower on PET/MRI than PET/CT. These differences are likely not related to the systematically longer uptake times for PET/MRI, as delayed PET imaging generally results in higher SUVs for malignant lesions (18), but may instead be due to PET photon attenuation by the MRI body phased-array coils, which are not captured by PET/MRI attenuation correction algorithms. Furthermore, most cancers in our study arose in the cervix or rectum, both of which are surrounded by bony structures. The Siemens mMR, which utilizes a Dixon-based segmentation approach with 4 tissue classes (soft tissue, fat, lung, air) (4), may have under-corrected for the attenuation effects of cortical bone relative to PET/CT, resulting in lower SUVs. Newer approaches utilizing ultrashort echo times have been successful in delineating cortical bone for attenuation correction of PET data in PET/MRI studies (19). The PSF reconstructions seemed to reduce quantitative differences between PET/MRI and PET/CT, suggesting that PSF might be best for such scanner comparisons.

With respect to the repeatability of PET metrics, our results are consistent with published results for 18F-FDG-PET/CT. To facilitate comparisons, repeatability coefficients (RCs) reported by other authors were converted as needed into wCVs. For example, a meta-analysis of 12 studies of various malignancies found a mean wCV of 11.0% for SUVmax (20). In our study, SUVmax wCVs ranged from 8.5-12.8% for PET/CT and 6.6-8.7% for PET/MRI, with similar results for the other PET metrics. Two of the studies included in the meta-analysis focused on pelvic malignancies (though neither included PET/MRI), with an SUVmax wCV of 10.7% for colorectal cancer (21) and 6.3% for ovarian cancer (22).

To our knowledge, there are only two other studies that have assessed the repeatability of PET metrics on PET/MRI. One study of head/neck cancers found SUVmax wCVs of 7.6% and 6.4% for PET/CT and PET/MRI, respectively, with similar wCVs for SUVpeak and SUVmean (23). In contrast, our PET metric
wCVs were generally higher, potentially reflecting differences in biology between head/neck and pelvic cancers or differences in the surrounding anatomy, as various physiologic processes specific to the pelvis (e.g., bladder filling, bowel peristalsis) might introduce greater variability between imaging sessions. As in our study, the authors found no statistically significant differences in repeatability between PET/CT and PET/MRI. The second repeatability study enrolled 33 patients with various malignancies, including seven colorectal cancers but no cervical cancers (24). In a single session (i.e., one ¹⁸F-FDG dose), subjects underwent either one PET/CT followed by two PET/MRIs or two PET/MRIs followed by one PET/CT. Despite study design differences, their PET metric wCVs for PET/MRI were generally in the 7.9-11.2% range, on par with ours.

For the ADC metrics, our results compare favorably with those published in a recent meta-analysis of ADC repeatability for extracranial soft-tissue tumors (25). This study found a mean wCV across 12 studies of 4.1% for ADCmedian, slightly higher than our ADCmedian wCV of 3.5%. We observed similarly low wCVs for the exploratory ADC metrics, suggesting that these metrics are also quantitatively robust. Notably, the ADCtrough, unlike the other ADC metrics evaluated, can be determined quickly and in a relatively user-independent fashion, suggesting that it may be preferable to the other ADC measures from a workflow perspective. DWI is appealing for treatment response assessment, as ADC metrics (unlike PET metrics) do not require any radiation exposure, pre-examination fasting, or control for variation in uptake times between imaging sessions. However, unlike PET, DWI can be impeded by susceptibility artifacts; by chemical shift artifacts from fat suppression failure; or by intra-lesional fibrosis, which can develop in the tumor interstitium during treatment, resulting in low signal on ADC maps that can mimic or obscure residual/recurrent disease (26).

Our study had some limitations. First, the comparison of repeatability on PET/CT versus PET/MRI is confounded by the systematically longer uptake times for PET/MRI relative to PET/CT. Although the scan order could have been randomized on a per-patient basis, this approach was not adopted due to concerns about the potential for wide ranges of uptake times within each scan type. Next, the applicability of our results to PET/MRI for other solid tumors is uncertain, as most tumors in our study were cervical cancers. It is conceivable that distinct tumor types might exhibit differences in the intrinsic variability of PET and/or ADC metrics. Likewise, the impact of physiologic motion on repeatability was not tested in this
study, as we focused on the pelvis only. Additionally, given that study subjects were not on any oncologic therapies between imaging sessions, it is possible (though unlikely) that there were substantial interval changes in tumor metabolism or cellularity related to true tumor growth. The number of subjects in our study was relatively small, reducing statistical power for the detection of true (though likely small) differences in repeatability between metrics. Finally, the degree to which our results, which were derived from a single model of integrated PET/MRI scanner, are generalizable to other PET/MRI scanner models is unclear, though there is no reason to suspect substantial differences.

CONCLUSION

In conclusion, we have demonstrated that the test/retest measurement repeatability of various PET metrics on 18F-FDG PET/MRI is both acceptably high and similar to values for 18F-FDG-PET/CT, across all PET reconstructions. These findings support the utilization of quantitative 18F-FDG PET/MRI within the framework of treatment response assessment for cervical cancer and other solid tumors of the pelvis.

ACKNOWLEDGEMENTS

The authors would like to thank Trustin Saam, MD, of the Department of Radiology at the Saint Louis University School of Medicine for his assistance with portions of the data collection.
REFERENCES

1. Fletcher JW, Djulbegovic B, Soares HP, et al. Recommendations on the use of 18F-FDG PET in oncology. *J Nucl Med*. 2008;49:480-508.

2. Ben-Haim S, Ell P. 18F-FDG PET and PET/CT in the evaluation of cancer treatment response. *J Nucl Med*. 2008;50:88-99.

3. 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.

4. Lodge MA. Repeatability of SUV in oncologic 18F-FDG PET. *J Nucl Med*. 2017;58:523-532.

5. 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.

6. Fraum TJ, Fowler KJ, McConathy J. PET/MRI: emerging clinical applications in oncology. *Acad Radiol*. 2016;23:220-236.

7. Hofmann M, Pichler B, Schölkopf B, Beyer T. Towards quantitative PET/MRI: a review of MR-based attenuation correction techniques. *Eur J Nucl Med Mol Imaging*. 2009;36:93-104.

8. Nougaret S, Reinhold C, Alsharif SS, et al. Endometrial Cancer: Combined MR volumetry and diffusion-weighted imaging for assessment of myometrial and lymphovascular invasion and tumor grade. *Radiology*. 2015;276:797-808.

9. Gollub MJ, Hotker AM, Woo KM, Mazaheri Y, Gonen M. Quantitating whole lesion tumor biology in rectal cancer MRI: taking a lesson from FDG-PET tumor metrics. *Abdom Radiol (NY)*. 43(7):1575-1582.

10. Olsen JR, Esthappan J, Dewees T, et al. Tumor volume and subvolume concordance between FDG-PET/CT and diffusion-weighted MRI for squamous cell carcinoma of the cervix. *J Magn Reson Imaging*. 2013;37:431-434.

11. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986;327:307-310.

12. Bland JM, Altman DG. Measuring agreement in method comparison studies. *Stat Methods Med Res*. 1999;8:161-179.

13. Hochberg Y, Benjamini Y. More powerful procedures for multiple significance testing. *Stat Med*. 
14. Sarabhai T, Schaarschmidt BM, Wetter A, et al. Comparison of 18F-FDG PET/MRI and MRI for pre-therapeutic tumor staging of patients with primary cancer of the uterine cervix. *Eur J Nucl Med Mol Imaging*. 2017;45:67-76.

15. Sarabhai T, Tschischka A, Stebner V, et al. Simultaneous multiparametric PET/MRI for the assessment of therapeutic response to chemotherapy or concurrent chemoradiotherapy of cervical cancer patients: preliminary results. *Clin Imaging*. 2018;49:163-168.

16. Sawicki LM, Kirchner J, Grueneisen J, et al. Comparison of 18F-FDG PET/MRI and MRI alone for whole-body staging and potential impact on therapeutic management of women with suspected recurrent pelvic cancer: a follow-up study. *Eur J Nucl Med Mol Imaging*. 2017;45:622-629.

17. Paspulati RM, Partovi S, Herrmann KA, Krishnamurthi S, Delaney CP, Nguyen NC. Comparison of hybrid FDG PET/MRI compared with PET/CT in colorectal cancer staging and restaging: a pilot study. *Abdom Imaging*. 2015;40:1415-1425.

18. Surov A, Meyer HJ, Schob S, et al. Parameters of simultaneous 18F-FDG-PET/MRI predict tumor stage and several histopathological features in uterine cervical cancer. *Oncotarget*. 2017;8:28285-28296.

19. Schillaci O. Use of dual-point fluorodeoxyglucose imaging to enhance sensitivity and specificity. *Semin Nucl Med*. 2012;42:267-80.

20. Berker Y, Franke J, Salomon a., et al. MRI-Based attenuation correction for hybrid PET/MRI systems: a 4-class tissue segmentation technique using a combined ultrashort-echo-time/Dixon MRI sequence. *J Nucl Med*. 2012;53:796-804.

21. Van Velden FHP, Nissen IA, Jongsma F, et al. Test-retest variability of various quantitative measures to characterize tracer uptake and/or tracer uptake heterogeneity in metastasized liver for patients with colorectal carcinoma. *Mol Imaging Biol*. 2014;16:13-18.

22. Rockall AG, Avril N, Lam R, et al. Repeatability of quantitative FDG-PET/CT and contrast-enhanced CT in recurrent ovarian carcinoma: test-retest measurements for tumor FDG uptake, diameter, and volume. *Clin Cancer Res*. 2014;20:2751-2760.

23. Rasmussen JH, Fischer BM, Aznar MC, et al. Reproducibility of 18F-FDG PET uptake
measurements in head and neck squamous cell carcinoma on both PET/CT and PET/MR. *Br J Radiol.* 2015;88:20140655.

24. Groshar D, Bernstine H, Goldberg N, et al. Reproducibility and repeatability of same-day two sequential FDG PET/MR and PET/CT. *Cancer Imaging.* 2017;17:11.

25. Winfield JM, Tunariu N, Rata M, et al. Extracranial soft-tissue tumors: repeatability of apparent diffusion coefficient estimates from diffusion-weighted MR imaging. *Radiology.* 2017;284:88-99.

26. Nougaret S, Tirumani SH, Addley H, Pandey H, Sala E, Reinhold C. Pearls and pitfalls in MRI of gynecologic malignancy with diffusion-weighted technique. *Am J Roentgenol.* 2013;200:261-276.
Figure 1. Schematic of study imaging protocol. All patients received 370 MBq \(^{18}\text{F}\)-FDG at 0 min. PET/CT imaging began at 60 min, lasting for 15 min. Immediately after PET/CT, the PET/MRI began, lasting for 30 min. DWI was performed at the beginning (DWI 1) and at the end (DWI 2) of the PET/MRI session. Occasionally, due to various patient-related factors, session 2 imaging began at a different time than session 1 imaging (note horizontal offset in the bars representing PET/CT sessions 1 & 2 and PET/MRI sessions 1 & 2). To control for these differences, static PET images were reconstructed using overlapping intervals of 1 min, 3 min, and 5 min in the PET data (brackets), thereby achieving identical effective uptake times. To minimize differences in uptake times between PET/MRI and PET/CT, these reconstruction intervals were selected from the latest overlapping portion for the PET/CT and the earliest overlapping portion for the PET/MRI.
Figure 2. Bland-Altman plots for \( \text{SUV}_{\text{max}} \). For the \( \text{SUV}_{\text{max}} \) from the point-spread function (PSF) 3 min reconstructions on PET/CT (A) and PET/MRI (B), Bland-Altman plots are shown. For each subject (small black circles), the percent change (\( \%\Delta \)) between measurements (y-axis) is plotted against the mean of the two measurements. The horizontal red line (\( \mu \)) indicates the mean \( \%\Delta \) across all subjects; the dotted black horizontal lines indicate one standard deviation (SD) from the mean \( \%\Delta \); and the solid horizontal lines indicate 1.96 SDs from the mean \( \%\Delta \), constituting the 95% limits of repeatability (the range of \( \%\Delta \) within which 95% of observations are expected to fall). The distance along the y-axis between \( \mu \) and either 1.96 SD line is RC. Comparing the plots in A and B (identical y-axis ranges), the 95% limits of repeatability (LOR) appear substantially more narrow for PET/MRI; correspondingly, the repeatability coefficient (RC) and within-subject coefficient of variation (wCV) were larger for PET/CT than PET/MRI.
Figure 3. Bland-Altman plots for SUL_{peak}. For the SUL_{peak} from point-spread function (PSF) 3 min reconstructions on PET/CT (A) and PET/MRI (B), Bland-Altman plots are shown. Comparing the plots in A and B (identical y-axis ranges), the 95% limits of repeatability (LOR) appear slightly more narrow for PET/MRI; correspondingly, the repeatability coefficient (RC) and within-subject coefficient of variation (wCV) were slightly larger for PET/CT than PET/MRI.
Figure 4. SUL\textsubscript{peak} repeatability from PET image analysis. A 65-year-old woman with biopsy-proven squamous cell carcinoma of the cervix was imaged with 24 hours between session 1 (A) and session 2 (B). Transaxial PET images with manual whole tumor contours (pink), T2-weighted MR images, and fused PET/MR images are shown from top to bottom for each imaging session. Note the inter-session contour similarities. The SUL\textsubscript{peak} (small pink sphere) was located in the same tumor region for both sessions, with similar values.
Figure 5. Bland-Altman plots for $AD_{C_{median}}$. For the $AD_{C_{median}}$ on PET/MRI, Bland-Altman plots are shown. The 95% limits of repeatability (LOR) were substantially narrower than for SUV$_{max}$ and SUL$_{peak}$, with numerically lower repeatability coefficient (RC) and within-subject coefficient of variation (wCV) values.
Figure 6. ADC$_{median}$ repeatability from ADC map analysis. A 33-year-old woman with biopsy-proven squamous cell carcinoma of the cervix was imaged with 48 hours between session 1 (A) and session 2 (B). Transaxial ADC maps with manual contours (pink), T2-weighted MR images, and fused PET/MR images are shown from top to bottom for each imaging session. The ADC$_{median}$ values for sessions 1 and 2 were similar.
### Table 1. Inclusion and exclusion criteria.

| Inclusion                                                                                       |
|------------------------------------------------------------------------------------------------|
| • Histologically confirmed malignant solid tumor of the pelvis (primary or metastatic; newly diagnosed or recurrent) |
| • Maximum tumor diameter ≥ 2.0 cm                                                              |
| • Patient age ≥ 18 years                                                                       |
| • Ability to provide informed consent                                                           |
| • Ability to tolerate 60 min of supine imaging                                                 |

| Exclusion                                                                                       |
|------------------------------------------------------------------------------------------------|
| • Oncologic therapy within 30 days prior to enrollment (to minimize treatment-related changes in tumor behavior between sessions) |
| • Uncontrolled intercurrent illness (e.g., active infections)                                    |
| • Insulin-dependent diabetes mellitus                                                           |
| • Prostheses incompatible with 3 Tesla magnetic fields                                           |
| • Pregnancy or nursing                                                                         |
Table 2. Subject characteristics.

| Characteristic                        | Value       |
|---------------------------------------|-------------|
| Age (years)                           | 48.1 ± 10.5*|
| Height (m)                            | 1.7 ± 0.1*  |
| Weight (kg)                           | 82.5 ± 21.0*|
| Body-mass index (kg/m²)               | 29.7 ± 6.4* |
| Gender                                |             |
| Male                                  | 1 (7)†      |
| Female                                | 13 (93)†    |
| Race/ethnicity                        |             |
| Caucasian (non-hispanic)             | 12 (86)†    |
| Caucasian (hispanic)                 | 1 (7)†      |
| African-American                      | 1 (7)†      |
| Histologic diagnosis                  |             |
| Cervical squamous cell carcinoma      | 11 (79)†    |
| Colorectal adenocarcinoma             | 2 (14)†     |
| Endometrial adenocarcinoma            | 1 (7)†      |
| Treatment status                      |             |
| Initial diagnosis                     | 13 (93)†    |
| Recurrent disease                     | 1 (7)†      |

* Entries are means ± standard deviations.
† Entries are numbers of cases (percentages).
Supplemental Table 1. DWI acquisition parameters.

| Parameter                        | Value                  |
|----------------------------------|------------------------|
| Time to repetition (TR)          | 5300 ms                |
| Time to echo (TE)                | 81 ms                  |
| Averages                         | 3                      |
| Field of view                    | 330 mm x 248 mm        |
| Voxel dimensions (in-plane)      | 2.6 mm x 2.6 mm        |
| Slice thickness                  | 5 mm                   |
| B-values                         | 50, 500, 1000 s/mm²    |
| Diffusion directions             | 3                      |
| Fat suppression                  | SPAIR                  |
| Parallel acquisition technique   | GRAPPA                 |
| Acceleration factor              | 2                      |

Abbreviations: GRAPPA – generalized autocalibrating partial parallel acquisition; SPAIR – spectral attenuated inversion recovery
Supplemental Table 2. PET reconstruction parameters.

| Parameter       | OSEM PET/CT | OSEM PET/MRI | PSF PET/CT | PSF PET/MRI |
|-----------------|-------------|--------------|------------|-------------|
| Dimensionality  | 2D          | 3D           | 3D         | 3D          |
| Iterations      | 4           | 3            | 2          | 3           |
| Subsets         | 8           | 21           | 21         | 21          |
| Matrix          | 168 x 168   | 172 x 172    | 168 x 168  | 172 x 172   |
| Zoom            | 1           | 1            | 1          | 1           |
| Filter          | 5 mm Gaussian | 7 mm Gaussian | 2 mm Gaussian | 3 mm Gaussian |

Note: For the OSEM reconstructions, parameters were selected to achieve identical recovery coefficients for PET/CT and PET/MRI. Similarly, for the PSF reconstructions, manufacturer-provided resolution recovery parameters were first applied, and then the filters were adjusted to achieve identical recovery coefficients for PET/CT and PET/MRI. Time-of-flight imaging was not available for either scanner used in this study.

Abbreviations: OSEM – ordered-subset expectation maximization; PSF – point-spread function
Supplemental Table 3. Mean values of SUV\text{max} and SUL\text{peak} from OSEM and PSF reconstructions for PET/CT and PET/MRI.

| Reconstruction | Metric  | 1 min | 3 min | 5 min |
|----------------|---------|-------|-------|-------|
|                |         | PET/CT | PET/MRI | p     | PET/CT | PET/MRI | p     | PET/CT | PET/MRI | p     |
| OSEM           | SUV\text{max} | 18.7 ± 5.6 | 16.3 ± 5.5 | 0.006* | 18.3 ± 5.4 | 15.7 ± 5.3 | 0.006* | 18.2 ± 5.6 | 16.4 ± 5.8 | 0.02  |
| PSF            | SUV\text{max} | 24.0 ± 8.0 | 23.6 ± 6.5 | 0.83   | 23.2 ± 7.6 | 22.8 ± 6.4 | 0.472  | 22.7 ± 7.7 | 22.0 ± 6.2 | 0.12  |
| OSEM           | SUL\text{peak} | 9.3 ± 2.5  | 8.2 ± 2.3  | 0.001* | 9.3 ± 2.5  | 8.1 ± 2.2  | 0.001* | 9.2 ± 2.5  | 8.4 ± 2.4  | 0.001* |
| PSF            | SUL\text{peak} | 10.5 ± 2.8 | 9.3 ± 2.2  | 0.001* | 10.3 ± 2.6 | 9.8 ± 2.5  | 0.036  | 10.3 ± 2.7 | 9.5 ± 2.4  | 0.003* |

Note: All data are based on whole tumor contours. Numbers provided are means ± standard deviations across all patients and reflect values obtained for both imaging sessions averaged on a per-patient basis. The p values are based on results of the Wilcoxon signed-rank test.

Abbreviations: OSEM – ordered-subset expectation maximization; PSF – point spread function; SUL – lean body mass-adjusted standardized uptake value

* Significant p value, based on results of the Wilcoxon signed-rank test.
Supplemental Table 4. Mean values of exploratory $^{18}$F-FDG-PET metrics from OSEM and PSF reconstructions for PET/CT and PET/MRI.

| Reconstruction | Contour | Metric | 1 min | 3 min | 5 min | 1 min | 3 min | 5 min | 1 min | 3 min | 5 min |
|----------------|--------|--------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| OSEM           | WT     | SUV$_{\text{mean}}$ | 7.8 ± 2.4 | 7.3 ± 2.2 | 0.05 | 7.9 ± 2.3 | 7.1 ± 2.1 | 0.03 | 7.9 ± 2.3 | 7.5 ± 2.3 | 0.08 |
| PSF            | WT     | SUV$_{\text{mean}}$ | 8.0 ± 2.3 | 7.3 ± 1.8 | 0.16 | 8.1 ± 2.3 | 8.0 ± 2.3 | 0.64 | 8.0 ± 2.3 | 7.8 ± 2.2 | 0.25 |
| OSEM           | WT     | SUV$_{\text{peak}}$ | 14.8 ± 4.7 | 13.1 ± 4.2 | 0.001* | 14.8 ± 4.7 | 12.9 ± 4.1 | 0.001* | 14.8 ± 4.7 | 13.5 ± 4.4 | 0.001* |
| PSF            | WT     | SUV$_{\text{peak}}$ | 16.4 ± 5.0 | 14.9 ± 4.2 | 0.01 | 16.1 ± 4.3 | 15.6 ± 4.8 | 0.16 | 16.4 ± 5.2 | 15.3 ± 4.6 | 0.004* |
| OSEM           | WT     | SUV$_{\text{TLG}}$ | 611 ± 769 | 575 ± 778 | 0.25 | 616 ± 771 | 558 ± 730 | 0.03 | 615 ± 770 | 596 ± 790 | 0.33 |
| PSF            | WT     | SUV$_{\text{TLG}}$ | 608 ± 759 | 538 ± 676 | 0.05 | 616 ± 757 | 615 ± 805 | 0.87 | 614 ± 755 | 595 ± 758 | 0.51 |
| OSEM           | WT     | SUV$_{\text{max}}$ | 11.7 ± 2.9 | 10.2 ± 3.1 | 0.001* | 11.5 ± 2.8 | 9.8 ± 3.0 | 0.006* | 11.4 ± 2.9 | 10.2 ± 3.3 | 0.02 |
| PSF            | WT     | SUV$_{\text{max}}$ | 15.3 ± 4.2 | 14.7 ± 3.3 | 0.27 | 14.5 ± 4.0 | 14.2 ± 3.4 | 0.27 | 14.2 ± 4.0 | 13.7 ± 3.2 | 0.18 |
| OSEM           | WT     | SUV$_{\text{mean}}$ | 4.9 ± 1.4 | 4.5 ± 1.2 | 0.03 | 5.0 ± 1.3 | 4.5 ± 1.1 | 0.006* | 5.0 ± 1.3 | 4.7 ± 1.3 | 0.04 |
| PSF            | WT     | SUV$_{\text{mean}}$ | 5.1 ± 1.3 | 4.5 ± 1.0 | 0.04 | 5.1 ± 1.3 | 5.0 ± 1.3 | 0.47 | 5.1 ± 1.3 | 4.9 ± 1.2 | 0.20 |
| OSEM           | WT     | SUV$_{\text{TLG}}$ | 386 ± 493 | 361 ± 497 | 0.10 | 389 ± 495 | 350 ± 466 | 0.01 | 388 ± 494 | 374 ± 506 | 0.36 |
| PSF            | WT     | SUV$_{\text{TLG}}$ | 391 ± 487 | 338 ± 431 | 0.01 | 388 ± 485 | 386 ± 515 | 0.73 | 387 ± 484 | 371 ± 483 | 0.20 |
| OSEM           | WT     | MTV (ml) | 67.8 ± 74.6 | 68.7 ± 77.2 | 0.33 | 67.8 ± 74.6 | 68.7 ± 77.2 | 0.33 | 67.8 ± 74.6 | 68.7 ± 77.2 | 0.33 |
| PSF            | WT     | MTV (ml) | 67.8 ± 74.6 | 68.7 ± 77.2 | 0.33 | 67.8 ± 74.6 | 68.7 ± 77.2 | 0.33 | 67.8 ± 74.6 | 68.7 ± 77.2 | 0.33 |
| OSEM           | 40%    | SUV$_{\text{mean}}$ | 11.0 ± 3.4 | 9.7 ± 3.1 | 0.002* | 10.9 ± 3.3 | 9.4 ± 3.0 | 0.001* | 10.8 ± 3.4 | 9.8 ± 3.2 | 0.02 |
| PSF            | 40%    | SUV$_{\text{mean}}$ | 13.4 ± 4.3 | 13.0 ± 3.6 | 0.25 | 13.1 ± 4.3 | 13.0 ± 3.9 | 0.60 | 12.9 ± 4.3 | 12.6 ± 3.8 | 0.22 |
| OSEM           | 40%    | SUV$_{\text{TLG}}$ | 436 ± 602 | 409 ± 546 | 0.78 | 463 ± 640 | 399 ± 495 | 0.18 | 468 ± 643 | 430 ± 537 | 0.10 |
| PSF            | 40%    | SUV$_{\text{TLG}}$ | 351 ± 504 | 266 ± 338 | 0.04 | 389 ± 546 | 387 ± 609 | 0.87 | 390 ± 564 | 390 ± 569 | 0.93 |
| OSEM           | 40%    | SUV$_{\text{mean}}$ | 6.9 ± 1.7 | 6.0 ± 1.7 | 0.002* | 6.9 ± 1.7 | 5.9 ± 1.6 | 0.001* | 6.8 ± 1.7 | 6.1 ± 1.7 | 0.004* |
| PSF            | 40%    | SUV$_{\text{mean}}$ | 8.6 ± 2.2 | 8.1 ± 1.8 | 0.04 | 8.2 ± 2.2 | 8.1 ± 2.0 | 0.36 | 8.1 ± 2.2 | 7.8 ± 2.0 | 0.12 |
| OSEM           | 40%    | MTV (ml) | 277 ± 387 | 257 ± 348 | 0.73 | 294 ± 412 | 251 ± 317 | 0.04 | 295 ± 413 | 271 ± 345 | 0.08 |
| PSF            | 40%    | MTV (ml) | 226 ± 323 | 168 ± 216 | 0.004* | 247 ± 350 | 247 ± 391 | 0.98 | 248 ± 361 | 245 ± 364 | 0.60 |
| OSEM           | 40%    | MTV (ml) | 36.1 ± 46.8 | 37.6 ± 43.6 | 0.08 | 39.4 ± 51.5 | 38.3 ± 42.9 | 0.08 | 39.9 ± 52.2 | 39.4 ± 44.2 | 0.08 |
| PSF            | 40%    | MTV (ml) | 25.7 ± 36.2 | 19.9 ± 24.8 | 0.01 | 29.6 ± 41.2 | 29.0 ± 44.2 | 0.98 | 29.9 ± 43.1 | 30.4 ± 44.0 | 0.83 |

Note: Numbers provided are means ± standard deviations across all patients and reflect values obtained for both imaging sessions averaged on a per-patient basis.

SUV$_{\text{peak}}$ and SUV$_{\text{max}}$ were calculated only for the WT contour, as some of the 40% isocontours were not large enough for calculation of a peak value and because maximum values for the WT contour and the 40% isocontours are identical.

Abbreviations: 40% – 40% isocontour; MTV – metabolic tumor volume; OSEM – ordered-subset expectation maximization; PSF – point spread function; SUL – lean body mass-adjusted standardized uptake value; TLG – total lesion glycolysis; WT – whole tumor contour

* Significant $p$ value, based on results of the Wilcoxon signed-rank test.
Supplemental Table 5. Mean values of ADC\textsubscript{median} from PET/MRI.

| Metric     | Session 1       | Session 2       | p    |
|------------|-----------------|-----------------|------|
| ADC\textsubscript{median} (mm/s\textsuperscript{2}) | 904 ± 104       | 904 ± 113       | 0.76 |

Note: Numbers provided in the session 1 and session 2 columns are means ± standard deviations across all patients. For the majority of patients, two usable DWI acquisitions were available for each imaging session. For these patients, the two data points from a given session were averaged to obtain a single value per imaging session. The p values are based on results of the Wilcoxon signed-rank test.

Abbreviations: ADC – apparent diffusion coefficient
Supplemental Table 6. Mean values of exploratory ADC metrics from PET/MRI.

| Metric          | Session 1     | Session 2     | p   |
|-----------------|---------------|---------------|-----|
| ADCmean (mm/s²) | 942 ± 92      | 933 ± 101     | 0.64|
| ADC₂₀ (mm/s²)   | 708 ± 79      | 704 ± 75      | 0.81|
| ADCtrough (mm/s²)| 750 ± 106    | 753 ± 115     | 0.39|
| DTV (ml)        | 64.7 ± 80.5   | 62.7 ± 77.5   | 0.64|

Note: Numbers provided in the session 1 and session 2 columns are means ± standard deviations across all patients. For the majority of patients, two usable DWI acquisitions were available for each imaging session. For these patients, the two data points from a given session were averaged to obtain a single value per imaging session. The p values are based on results of the Wilcoxon signed-rank test.

Abbreviations: ADC – apparent diffusion coefficient; ADC₂₀ – mean value obtained from the intra-lesion voxels with the lowest 20% of ADC values; ADCtrough – lowest mean ADC value obtainable for a 1 cm³ sphere placed within the confines of the lesion (analogous to the inverse of peak for PET imaging); DTV – diffusional tumor volume.
Supplemental Table 7. Repeatability results for $SUV_{\text{max}}$ and $SUL_{\text{peak}}$ from OSEM and PSF reconstructions for PET/CT and PET/MRI.

| Reconstruction | Metric | 1 min | 3 min | 5 min |
|---------------|--------|-------|-------|-------|
|               | PET/CT | PET/MRI | PET/CT | PET/MRI | PET/CT | PET/MRI |
| OSEM          | $SUV_{\text{max}}$ | -2.2% 8.5% | -5.3% 7.9% | -1.7% 9.6% | -4.7% 6.7%* | -1.6% 9.7% | -5.2% 8.0% |
|               | $SUL_{\text{peak}}$ | -0.6% 9.9% | -4.6% 10.4% | 0.6% 10.5% | -4.7% 9.2%* | -0.2% 10.7% | -5.1% 9.7%* |
| PSF           | $SUV_{\text{max}}$ | 0.9% 12.8% | -2.3% 8.7% | 2.6% 11.7% | -1.8% 6.6% | -0.8% 8.8% | -3.7% 7.7% |
|               | $SUL_{\text{peak}}$ | 2.3% 11.2% | -3.7% 11.3% | 2.3% 11.5% | -6.0% 10.0%* | 0.5% 10.4% | -5.5% 9.5%* |

Note: All data are based on whole tumor contours.

Abbreviations: %Δ – percent change between sessions; OSEM – ordered-subset expectation maximization; PSF – point-spread function; wCV – within-subject coefficient of variation; SUL – lean body mass-adjusted standardized uptake value

* Distribution of values for mean %Δ was non-normal.
Supplemental Table 8. Repeatability results for exploratory 18F-FDG-PET parameters from OSEM and PSF reconstructions for PET/CT and PET/MRI.

| Reconstruction | Contour | Metric     | PET/CT 1 min | PET/CT 3 min | PET/CT 5 min | PET/MRI 1 min | PET/MRI 3 min | PET/MRI 5 min |
|----------------|---------|------------|--------------|--------------|--------------|--------------|--------------|--------------|
| OSEM           | WT      | SUV\text{mean} | 2.1% 12.2%   | -4.2% 11.1%  | 3.0% 13.1%   | -4.1% 10.9%  | 2.2% 12.9%   | -4.6% 10.9%  |
|                |         | SUV\text{peak} | -0.5% 9.6%   | -5.8% 10.1%  | 0.7% 10.3%   | -4.5% 8.6%*  | 0.7% 10.7%   | -5.0% 9.4%*  |
|                | PSF     | SUV\text{mean} | 0.1% 13.7%   | -2.1% 14.4%  | 4.4% 14.5%   | -4.8% 11.9%  | 2.3% 13.4%   | -4.1% 11.6%  |
|                |         | SUV\text{peak} | -1.2% 11.2%  | -3.6% 11.0%  | -0.8% 14.5%  | -5.9% 9.8%*  | 0.5% 10.1%   | -5.4% 9.3%*  |
| OSEM           | WT      | SUVTLG    | -4.2% 13.8%  | -0.4% 10.5%  | -3.3% 13.6%  | -0.3% 10.2%  | -4.0% 13.9%  | -0.8% 10.4%  |
|                | PSF     | SUVTLG    | -6.1% 13.9%  | 1.8% 12.7%   | -1.9% 13.6%  | -0.9% 10.2%  | -4.0% 13.5%  | -0.3% 9.6%   |
| OSEM           | WT      | SULmax    | -2.3% 8.8%   | -5.4% 8.2%   | -1.8% 9.8%   | -4.8% 7.0%*  | -1.9% 10.0%  | -5.3% 8.4%*  |
|                | PSF     | SULmax    | 4.3% 13.0%   | -2.4% 8.9%   | 2.8% 11.8%   | -1.9% 6.9%   | -0.9% 9.1%   | -3.8% 7.9%   |
| OSEM           | WT      | SULmean   | 2.0% 12.5%   | -4.2% 11.4%  | 2.9% 13.3%   | -4.1% 11.1%  | 2.2% 13.2%   | -4.7% 11.2%  |
|                | PSF     | SULmean   | 3.5% 13.3%   | -2.1% 14.6%  | 4.3% 14.7%   | -4.8% 12.2%  | 2.2% 13.6%   | -4.1% 11.9%  |
| OSEM           | WT      | SULTLG   | -4.2% 13.9%  | -0.5% 10.8%  | -3.4% 13.7%  | -0.3% 10.6%  | -4.1% 14.0%  | -0.8% 10.7%  |
|                | PSF     | SULTLG   | -2.7% 14.0%  | 1.7% 12.9%   | -1.7% 13.6%  | -1.0% 10.5%  | -4.1% 13.6%  | 0.3% 9.3%    |
| OSEM           | WT      | MTV (ml)  | -6.2% 13.8%  | 3.8% 9.1%*   | -6.2% 13.8%  | 3.8% 9.1%*   | -6.2% 13.8%  | 3.8% 9.1%*   |
|                | PSF     | MTV (ml)  | -6.2% 13.8%  | 3.8% 9.1%*   | -6.2% 13.8%  | 3.8% 9.1%*   | -6.2% 13.8%  | 3.8% 9.1%*   |
| OSEM           | 40%     | SUV\text{mean} | -1.3% 8.2%   | -5.4% 7.9%   | -1.5% 9.6%   | -5.4% 7.6%*  | 0.0% 10.4%   | -5.2% 7.9%*  |
|                | PSF     | SUV\text{mean} | -0.7% 12.0%  | -3.2% 8.3%   | 1.8% 10.5%   | -4.4% 7.2%   | -0.3% 9.1%   | -5.4% 8.2%   |
| OSEM           | 40%     | SUVTLG   | -0.8% 16.3%  | 0.7% 15.0%   | 0.4% 13.7%   | 0.9% 14.1%   | -2.5% 15.2%  | 0.7% 13.5%   |
|                | PSF     | SUVTLG   | -9.8% 17.9%  | -1.5% 22.2%  | -0.7% 14.7%  | -2.9% 18.5%  | -4.5% 15.4%  | 0.5% 16.3%   |
| OSEM           | 40%     | SUL\text{mean} | -1.4% 8.5%   | -5.5% 8.2%   | -1.6% 10.0%  | -5.2% 7.4%*  | -2.3% 9.8%   | -5.3% 8.2%*  |
|                | PSF     | SUL\text{mean} | 2.7% 11.8%   | -3.3% 8.5%   | 1.7% 10.7%   | -4.2% 7.7%   | -0.4% 9.3%   | -5.5% 8.5%   |
| OSEM           | 40%     | SULTLG   | -0.9% 16.6%  | 0.6% 15.3%   | 0.4% 13.9%   | 0.8% 14.4%   | -0.6% 14.1%  | 0.6% 13.8%   |
|                | PSF     | SULTLG   | -6.4% 16.5%  | -1.5% 22.4%  | -0.8% 14.9%  | -5.1% 18.9%  | -4.6% 15.6%  | 0.4% 16.6%   |
| OSEM           | 40%     | MTV (ml)  | 0.5% 11.9%   | 6.0% 14.3%*  | 1.9% 8.3%   | 6.0% 11.7%*  | 1.7% 9.1%   | 5.9% 10.6%   |
|                | PSF     | MTV (ml)  | -9.2% 15.2%  | 1.5% 23.0%   | -2.6% 10.2%  | -1.2% 17.1%  | -4.2% 10.9%  | 5.8% 16.3%   |

Note: SUV\text{peak} and SUL\text{max} were calculated only for the WT contour, as some of the 40\% isocontours were not large enough for calculation of a peak value and because maximum values for the WT contour and the 40\% isocontour are identical.

Abbreviations: 40\% – 40\% isocontour; % Δ – percent change; 18F-FDG – 2-18F-fluoro-2-deoxy-D-glucose; MTV – metabolic tumor volume; OSEM – ordered-subset expectation maximization; PET – positron emission tomography; PSF – point-spread function; TLG – total lesion glycolysis; wCV – within-subject coefficient of variation; WT – whole tumor contour

*Distribution of values for mean %Δ was non-normal
Supplemental Table 9. Repeatability results for ADCmedian from PET/MRI.

|                | Mean | %Δ  | wCV |
|----------------|------|-----|-----|
| ADCmedian      | 1.6% | 2.4%|     |

Note: For patients with more than one usable set of ADC data for a given session (as two DWI acquisitions were performed per session), one set was selected at random from each session for this analysis.

Abbreviations: %Δ – percent change between sessions; ADC – apparent diffusion coefficient; wCV – within-subject coefficient of variation
Supplemental Table 10. Repeatability results for exploratory ADC metrics from PET/MRI.

| Metric      | Mean % Δ | wCV  |
|-------------|----------|------|
| ADC_{mean}  | -0.1%    | 3.4% |
| ADC_{median}| 0.5%     | 3.5% |
| ADC_{20}    | -0.3%    | 3.2% |
| ADC_{trough}| 0.5%     | 3.8% |
| DTV         | -2.7%    | 9.7% |

Note: For patients with more than one usable set of ADC data for a given session (as two DWI acquisitions were performed per session), one set was selected at random from each session for this analysis.

Abbreviations: ADC – apparent diffusion coefficient; ADC_{20} – mean value obtained from the intra-lesion voxels with the lowest 20% of ADC values; ADC_{trough} – lowest mean ADC value obtainable for a 1 cm³ sphere placed within the confines of the lesion (analogous to the inverse of peak for PET imaging); DTV – diffusional tumor volume; wCV – within-subject coefficient of variation.
Supplemental Table 11. Pairwise comparisons of repeatability for SUV\textsubscript{max}, SUL\textsubscript{peak}, and ADC\textsubscript{median}.

| Pair Description | Metric 1 | Mean | %Δ | SD | Metric 2 | Mean | %Δ | SD | p  |
|------------------|---------|------|----|----|---------|------|----|----|----|
| 1 SUV\textsubscript{max} OSEM 1 min PET/CT | 9.4% | 7.6% | SUV\textsubscript{max} OSEM 1 min PET/MRI | 8.0% | 9.2% | 0.64 |
| 2 SUL\textsubscript{peak} OSEM 1 min PET/CT | 11.0% | 8.2% | SUL\textsubscript{peak} OSEM 1 min PET/MRI | 10.2% | 11.2% | 0.40 |
| 3 SUV\textsubscript{max} OSEM 3 min PET/CT | 9.9% | 8.9% | SUV\textsubscript{max} OSEM 3 min PET/MRI | 6.5% | 8.3% | 0.22 |
| 4 SUL\textsubscript{peak} OSEM 3 min PET/CT | 10.8% | 9.9% | SUL\textsubscript{peak} OSEM 3 min PET/MRI | 8.5% | 10.7% | 0.14 |
| 5 SUV\textsubscript{max} OSEM 5 min PET/CT | 9.7% | 9.6% | SUV\textsubscript{max} OSEM 5 min PET/MRI | 7.7% | 9.7% | 0.73 |
| 6 SUL\textsubscript{peak} OSEM 5 min PET/CT | 10.4% | 10.6% | SUL\textsubscript{peak} OSEM 5 min PET/MRI | 8.2% | 11.9% | 0.16 |
| 7 SUV\textsubscript{max} PSF 1 min PET/CT | 15.9% | 7.6% | SUV\textsubscript{max} PSF 1 min PET/MRI | 9.7% | 7.5% | 0.08 |
| 8 SUL\textsubscript{peak} PSF 1 min PET/CT | 12.6% | 9.3% | SUL\textsubscript{peak} PSF 1 min PET/MRI | 12.1% | 10.4% | 0.93 |
| 9 SUV\textsubscript{max} PSF 3 min PET/CT | 12.5% | 10.5% | SUV\textsubscript{max} PSF 3 min PET/MRI | 7.1% | 6.0% | 0.12 |
| 10 SUL\textsubscript{peak} PSF 3 min PET/CT | 12.3% | 10.4% | SUL\textsubscript{peak} PSF 3 min PET/MRI | 9.0% | 12.3% | 0.36 |
| 11 SUV\textsubscript{max} PSF 5 min PET/CT | 8.9% | 8.5% | SUV\textsubscript{max} PSF 5 min PET/MRI | 8.4% | 7.6% | 0.87 |
| 12 SUL\textsubscript{peak} PSF 5 min PET/CT | 10.5% | 9.9% | SUL\textsubscript{peak} PSF 5 min PET/MRI | 8.8% | 11.4% | 0.25 |
| 13 SUV\textsubscript{max} PSF 3 min PET/MRI | 7.1% | 6.0% | ADC\textsubscript{median} | 3.9% | 2.7% | 0.23 |
| 14 SUL\textsubscript{peak} PSF 3 min PET/MRI | 9.0% | 12.3% | ADC\textsubscript{median} | 3.9% | 2.7% | 0.74 |
| 15 SUV\textsubscript{max} OSEM 3 min PET/MRI | 6.5% | 8.3% | ADC\textsubscript{median} | 3.9% | 2.7% | 0.90 |
| 16 SUL\textsubscript{peak} OSEM 3 min PET/MRI | 8.5% | 10.7% | ADC\textsubscript{median} | 3.9% | 2.7% | 0.41 |

Note: All data are based on whole tumor contours. Numbers in the mean | %Δ | column are means (across all subjects) of the absolute values of relative differences (i.e., percentage changes) between imaging sessions 1 and 2. The p values are based on results of the Wilcoxon signed-rank test for comparisons 1-12 (paired data) and the Mann-Whitney U test for comparisons 13-16 (unpaired data).

Abbreviations: ADC – apparent diffusion coefficient; OSEM – ordered-subset expectation maximization; PSF – point-spread function; SD – standard deviation
Table 12. Pairwise comparisons of repeatability for exploratory 18F-FDG-PET and ADC metrics.

| Pair | Metric 1 | Mean | %Δ | SD | Metric 2 | Mean | %Δ | SD | p |
|------|----------|------|----|----|----------|------|----|----|---|
| 1    | Volume PET/CT | 15.4% | 13.0% | | Volume PET/MRI | 8.5% | 10.1% | | 0.11 |
| 2    | SUVmean PSF 1 min PET/CT | 15.6% | 10.7% | | SUVmean PSF 1 min PET/MRI | 15.2% | 13.1% | | 0.73 |
| 3    | 40% SUVmean PSF 1 min PET/CT | 9.2% | 8.4% | | 40% SUVmean PSF 1 min PET/MRI | 9.9% | 6.5% | | 0.16 |
| 4    | SUVmean PSF 3 min PET/CT | 16.5% | 12.1% | | SUVmean PSF 3 min PET/MRI | 12.1% | 12.4% | | 0.18 |
| 5    | 40% SUVmean PSF 3 min PET/CT | 10.9% | 9.8% | | 40% SUVmean PSF 3 min PET/MRI | 7.6% | 7.9% | | 0.16 |
| 6    | SUVmean PSF 5 min PET/CT | 15.1% | 10.9% | | SUVmean PSF 5 min PET/MRI | 11.8% | 11.7% | | 0.22 |
| 7    | 40% SUVmean PSF 5 min PET/CT | 8.8% | 9.0% | | 40% SUVmean PSF 5 min PET/MRI | 9.0% | 8.9% | | 0.87 |
| 8    | SUVmean OSEM 1 min PET/CT | 13.8% | 10.0% | | SUVmean OSEM 1 min PET/MRI | 11.3% | 11.5% | | 0.36 |
| 9    | 40% SUVmean OSEM 1 min PET/CT | 9.0% | 7.0% | | 40% SUVmean OSEM 1 min PET/MRI | 7.8% | 9.4% | | 0.47 |
| 10   | SUVmean OSEM 3 min PET/CT | 14.8% | 10.8% | | SUVmean OSEM 3 min PET/MRI | 11.4% | 10.8% | | 0.25 |
| 11   | 40% SUVmean OSEM 3 min PET/CT | 9.4% | 9.6% | | 40% SUVmean OSEM 3 min PET/MRI | 7.5% | 9.3% | | 0.33 |
| 12   | SUVmean OSEM 5 min PET/CT | 14.4% | 10.7% | | SUVmean OSEM 5 min PET/MRI | 10.3% | 12.1% | | 0.12 |
| 13   | 40% SUVmean OSEM 5 min PET/CT | 10.3% | 10.2% | | 40% SUVmean OSEM 5 min PET/MRI | 7.3% | 9.8% | | 0.26 |
| 14   | ADCmean (inter-session) | 3.1% | 2.1% | | ADCmean (intra-session) | 2.7% | 1.9% | | 0.35 |
| 15   | ADCmean (inter-session) | 3.1% | 2.1% | | ADCmedian (inter-session) | 3.9% | 2.7% | | 0.90 |
| 16   | ADCmean (inter-session) | 3.1% | 2.1% | | ADC20 (inter-session) | 2.9% | 3.4% | | 0.22 |
| 17   | ADCmean (inter-session) | 3.1% | 2.1% | | ADCtrough (inter-session) | 3.9% | 3.6% | | 0.75 |
| 18   | SUVmax PSF 3 min PET/MRI | 7.1% | 6.0% | | ADCmean (intra-session) | 3.1% | 2.1% | | 0.17 |
| 19   | SUVmax PSF 3 min PET/MRI | 9.0% | 12.3% | | ADCmean (inter-session) | 3.1% | 2.1% | | 0.70 |
| 20   | SUVmax OSEM 3 min PET/MRI | 6.5% | 8.3% | | ADCmean (inter-session) | 3.1% | 2.1% | | 0.96 |
| 21   | SUVmax OSEM 3 min PET/MRI | 8.5% | 10.7% | | ADCmean (inter-session) | 3.1% | 2.1% | | 0.50 |
| 22   | SUVmean PSF 3 min PET/CT | 16.5% | 12.1% | | 40% SUVmean PSF 3 min PET/CT | 10.9% | 9.8% | | 0.02 |
| 23   | SUVmean PSF 3 min PET/MRI | 12.1% | 12.4% | | 40% SUVmean PSF 3 min PET/MRI | 7.6% | 7.9% | | 0.06 |

Note: Data are based on whole tumor contours except for metrics preceded by a ‘40%’ designation; these metrics are based on the 40% isocontour. Numbers provided in the mean | %Δ | column are means (across all subjects) of the absolute values of relative differences (i.e., percentage changes) between imaging sessions 1 and 2. The p values are based on results of the Wilcoxon signed-rank test for comparisons 1-17 & 22-23 (paired data) and the Mann-Whitney U test for comparisons 18-21 (unpaired data).

Abbreviations: 40% – 40% isocontour; ADC – apparent diffusion coefficient; OSEM – ordered-subset expectation maximization; PSF – point-spread function; WT – whole tumor
Measurement Repeatability of $^{18}$F-FDG-PET/CT versus $^{18}$F-FDG-PET/MRI in Solid Tumors of the Pelvis

Tyler J. Fraum, Kathryn J. Fowler, John P. Crandall, Richard A. Laforest, Amber Salter, Hongyu An, Michael A. Jacobs, Perry W. Grigsby, Farrokh Dehdashti and Richard L. Wahl

*J Nucl Med.*
Published online: February 7, 2019.
Doi: 10.2967/jnumed.118.218735

This article and updated information are available at:
[http://jnm.snmjournals.org/content/early/2019/01/31/jnumed.118.218735](http://jnm.snmjournals.org/content/early/2019/01/31/jnumed.118.218735)

Information about reproducing figures, tables, or other portions of this article can be found online at:
[http://jnm.snmjournals.org/site/misc/permission.xhtml](http://jnm.snmjournals.org/site/misc/permission.xhtml)

Information about subscriptions to JNM can be found at:
[http://jnm.snmjournals.org/site/subscriptions/online.xhtml](http://jnm.snmjournals.org/site/subscriptions/online.xhtml)

*JNM* ahead of print articles have been peer reviewed and accepted for publication in *JNM*. They have not been copyedited, nor have they appeared in a print or online issue of the journal. Once the accepted manuscripts appear in the *JNM* ahead of print area, they will be prepared for print and online publication, which includes copyediting, typesetting, proofreading, and author review. This process may lead to differences between the accepted version of the manuscript and the final, published version.