RESEARCH ARTICLE

Test–Retest Repeatability of Patlak Slopes versus Standardized Uptake Values for Hypermetabolic Lesions and Normal Organs in an Oncologic PET/CT Population

Semra Ince[1] · Richard Laforest[1] · Saeed Ashrafinia[2] · Anne M. Smith[2] · Richard L. Wahl[1,3] · Tyler J. Fraum[1]

Abstract

Purpose We aimed to determine the test–retest repeatability of quantitative metrics based on the Patlak slope (PS) versus the standardized uptake value (SUV) among lesions and normal organs on oncologic [^{18}F]FDG-PET/CT.

Procedures This prospective, single-center study enrolled adults undergoing standard-of-care oncologic [^{18}F]FDG-PET/CTs. Early (35–50 min post-injection) and late (75–90 min post-injection) SUV and PS images were reconstructed from dynamic whole-body PET data. Repeat imaging occurred within 7 days. Relevant quantitative metrics were extracted from lesions and normal organs. Repeatability was assessed via mean test–retest percent changes [T-RT %Δ], within-subject coefficients of variation (wCVs), and intra-class correlation coefficients (ICCs).

Results Nine subjects (mean age, 61.7 ± 6.2 years; 6 females) completed the test–retest protocol. Four subjects collectively had 17 [^{18}F]FDG-avid lesions. Lesion wCVs were higher (i.e., worse repeatability) for PS-early-max (16.2%) and PS-early-peak (15.6%) than for SUV-early-max (8.9%) and SUV-early-peak (8.1%), with similar early metric ICCs (0.95–0.98). Lesion wCVs were similar for PS-late-max (8.5%) and PS-late-peak (6.4%) relative to SUV-late-max (9.7%) and SUV-late-peak (7.2%), with similar late metric ICCs (0.93–0.98). There was a significant bias toward higher retest SUV and PS values in the lesion analysis (T-RT %Δ [95% CI]: SUV-late-max, 10.0% [2.6%, 17.0%]; PS-late-max, 20.4% [14.3%, 26.4%]) but not in the normal organ analysis.

Conclusions Among [^{18}F]FDG-avid lesions, the repeatability of PS-based metrics is similar to equivalent SUV-based metrics at late post-injection time points, indicating that PS-based metrics may be suitable for tracking response to oncologic therapies. However, further validation is required in light of our study's limitations, including small sample size and bias toward higher retest values for some metrics.

Keywords Patlak · Metabolic rate · PET · FDG · Repeatability · Test-retest

Introduction

In oncologic positron emission tomography (PET), the standardized uptake value (SUV) is commonly used for various interpretive tasks, including differentiation of benign from malignant lesions and assessment of treatment response. That said, the SUV, which reflects tracer concentrations at a single time point, has inherent limitations [1]. As many tumors accumulate tracer progressively, longer uptake times often result in higher SUVs, complicating longitudinal assessments of tumor activity [2]. Furthermore, the SUV reflects both bound tracer within cells and free tracer in the blood pool, which can obscure tracer-avid lesions in highly vascular organs (e.g., liver) [3].

The Patlak methodology, which offers a practical graphical approach to kinetic modeling, addresses these limitations by separating the signal derived from bound versus free tracer [4]. The Patlak slope (PS), a surrogate of K_i derived from compartmental kinetic modeling, reflects the rate of irreversible tracer trapping within a tissue. The PS has the potential to reduce the dependence of tracer-avidity

✉ Tyler J. Fraum
fraumt@wustl.edu

1   Department of Radiology, Washington University School of Medicine, 510 S. Kingshighway Blvd, Campus Box 8131, St. Louis, MO 63110, USA

2   Siemens Medical Solutions USA, Inc., 810 Innovation Drive, Knoxville, TN 37932, USA

3   Department of Radiation Oncology, Washington University School of Medicine, St. Louis, MO, USA

assessments on uptake times and to improve the conspicuity of mildly tracer-avid lesions, particularly within organs with relatively high extracellular tracer concentrations [5]. However, PS values, which are derived from dynamic PET data and require an estimation of the blood input function (IF), may be more prone to motion/noise artifacts and computational errors than SUVs [6].

Knowledge of the intrinsic variability of quantitative imaging biomarkers is essential to the accurate interpretation of changes in these metrics during therapy. Prior 2-deoxy-2-[^{18}F]fluoro-D-glucose ([^{18}F]FDG) PET/CT studies have shown that the maximum SUV exhibits relatively high repeatability, with within-subject coefficients of variation (wCVs) of 10–12% for solid tumors [7]. In contrast, the repeatability of PS values on oncologic PET/CT examinations has not been thoroughly studied. This information is critical for the clinical implementation of PS as a marker of tumor response, especially given the greater complexity inherent to the kinetic modeling required by the Patlak method. Furthermore, as PET interpretation guidelines often require quantitative assessments of internal reference structures, such as the liver, the repeatability of PS values relative to SUVs among normal organs is also of interest [8].

Thus, the aim of our study was to determine the test–retest repeatability of PS-based versus SUV-based quantitative imaging metrics for tracer-avid lesions and normal organs on oncologic [^{18}F]FDG-PET/CT.

Materials and Methods

Study Design

This prospective, IRB-approved, HIPAA-compliant study (NCT04283552) enrolled subjects scheduled to undergo standard-of-care (SOC) oncologic [^{18}F]FDG-PET/CT for various indications. All imaging occurred at a single tertiary care center from June 2020 through April 2023. Inclusion criteria were as follows: ≥ 18 years of age; ability to provide written informed consent; and self-reported ability to tolerate ~ 90 min of near-motionless supine positioning. A subset of these patients undergoing [^{18}F]FDG-PET/CT were eligible to enroll in a test–retest substudy. Study imaging was performed prior to and following the SOC PET/CT, utilizing the same [^{18}F]FDG dose.

Imaging Protocol

The study imaging protocol is summarized in Suppl. Figure 1 (see ESM). Each imaging session was repeated (i.e., retest session) utilizing identical or near-identical preparation and acquisition parameters within seven days of initial imaging (i.e., test session). Subjects were instructed to fast

for ≥ 4 h and had blood glucose (BG) levels ≤ 200 mg/dL at [^{18}F]FDG injection. The administered [^{18}F]FDG dose was weight-based, as follows: < 54 kg, 370 MBq; 54–113 kg, 555 MBq; > 113 kg, 740 MBq. Imaging occurred on a Biograph Vision 600 PET/CT scanner (Siemens Healthineers; Knoxville, USA) with FDA-approved, commercially available software for reconstruction of multiparametric PET images (FlowMotion Multiparametric PET Suite; Siemens Healthineers; Knoxville, USA). Subjects were placed on the scanner table in the supine position, and care was taken to maximize patient comfort due to the prolonged acquisition period. The [^{18}F]FDG was administered intravenously as a bolus upon initiating a 6-min dynamic PET acquisition centered on the mediastinal blood pool. Subsequent 'whole-body' (WB) PET passes extended from the skull base to proximal thighs for most subjects, with modification of the craniocaudal extent to include other structures as indicated by the SOC protocol. Each WB pass was performed in list mode, utilizing continuous bed motion. Ten WB passes of various lengths (five × 2 min, five × 5 min) were performed before the SOC PET. Subjects were instructed to leave the scanner table to empty their bladders immediately before SOC imaging. Three additional WB passes (5 min each) were performed after the SOC PET. Low-dose CT scans were acquired immediately before each PET portion utilizing CARE Dose4D (reference: 111mAs), CARE kV (reference: 120 kV), and ADMIRE (strength: 2).

Input Function Analysis

Volumes of interest (VOIs) were placed in the descending thoracic aorta on the 6-min dynamic thoracic acquisition and subsequent WB passes via fully automated scanner tools [9]. Time-activity curves were generated to estimate the input function (IF) for Patlak modeling. The area under the time-activity curve (AUTAC) was extracted for comparison with the injected [^{18}F]FDG dose to screen for IF errors. Inter-session percent changes (%Δ) in the AUTAC and [^{18}F]FDG dose were calculated. Substantial differences (i.e., [%Δ AUTAC] – [%Δ [^{18}F]FDG dose] ≥ 15% or ≤ -15%) prompted further investigation. Identifiable technical errors were corrected prior to image reconstruction. Furthermore, all WB PET passes were reviewed to confirm the absence of reconstruction errors related to patient motion.

PET Image Reconstruction

Using data from three 5 min WB passes at early (35–50 min post-injection; pre-SOC) and late (75–90 min; post-SOC) post-injection time points, SUV and PS images were reconstructed according to manufacturer-recommended parameters: SUV (PSF + TOF, 4 iterations, 5 subsets, 440 × 440 matrix, all-pass filter); Patlak slope (Patlak, 8 iterations, 5

subsets, 220×220 matrix, 2 mm Gaussian filter). The multiparametric software requires at least three WB passes for Patlak analysis. The three latest pre-SOC WB passes were selected to ensure adequate time for steady state conditions to be established. SUVs (g/ml) were adjusted by actual body weight. In generating PS images, the scanner software automatically scaled the PS by the provided blood glucose level, resulting in units of mg/min/ml. This PS is equivalent to the [18F]FDG metabolic rate (MR_{FDG}). To facilitate comparison with other studies, the PS units were scaled by 100 to mg/min/100 ml.

Quantitative Analysis

[18F]FDG-avid lesions interpreted as sites of probable malignancy on the SOC PET/CT study were analyzed by one author (S.I., 10 years of post-training experience in PET imaging). For cases with numerous lesions, the largest and/or most [18F]FDG-avid were selected (maximum of 5 per subject). Each lesion was manually segmented in MIM version 7.1.5 (MIM Software; Cleveland, USA), using the co-registered CT for guidance, on the PS-early, SUV-early, PS-late, and SUV-late images to generate a VOI for each reconstruction. Maximum (max) and peak values were extracted for each lesion [8]. Additionally, for each PET reconstruction, spherical VOIs of various sizes were placed in the liver, lung, blood pool, bone marrow, skeletal muscle, grey matter, and spleen (Suppl. Table 1; see ESM) to extract mean values. The blood glucose (BG) levels at the time of [18F]FDG injection were also analyzed to assess for potential biases in PS related to large BG differences between test and retest sessions This procedure was repeated for the same lesions and normal anatomic structures on all four reconstructions derived from the retest session. Care was taken to ensure visually similar lesion segmentations and normal anatomic structure VOI placements for both the test and retest sessions. Notably, as the lesion analysis focused on max and peak values (rather than mean), slight differences in lesion segmentation between test and retest sessions should not affect the extracted values.

Statistical Analysis

All statistical analysis was conducted by one author (T.J.F.; 5 years of post-training experience), with statistician guidance, in Prism 9 (GraphPad Software; San Diego, USA) or Excel 2016 (Microsoft, Inc.; Redmond, WA). Subject and scan characteristics were summarized descriptively. For all lesions and each normal anatomic structure, repeatability was assessed for four reconstructions (SUV-early, PS-early, SUV-late, PS-late) according to Bland–Altman methods [10]. Test–retest percent differences (T-RT %Δ) were calculated (instead of absolute differences) to facilitate comparison of results across metrics of different magnitudes.

The T-RT %Δ was defined as the retest value minus the test value, divided by the mean of the test and retest values. The mean T-RT %Δ reflects the mean of T-RT %Δ values across all lesions (lesion analysis) or subjects (normal organ analysis). The standard deviation (SD) of the T-RT %Δ values was calculated for each metric. The wCV was defined as $SD/\sqrt{2}$. The 95% confidence interval (95% CI) of the mean T-RT %Δ was calculated based on the t statistic due to the small sample size.

Due to an apparent bias toward higher values on the retest study for multiple lesion metrics (i.e., mean T-RT %Δ substantially > 0.0%), the intraclass correlation coefficient (ICC) of each metric was also calculated for lesions and normal organs (ICC: range, 0–1; 1 = perfect repeatability), as an alternative means of assessing repeatability [11]. Note that the ICC is a metric of agreement rather than simple correlation. The apparent bias in the lesion repeatability analysis was also investigated by correlating subjects' clinical features (e.g., treatment history, disease trends on imaging performed before and after study imaging) with the T-RT %Δ values of their lesions and normal organs (specifically, liver and grey matter). The liver is a commonly used reference organ for clinical PET interpretation, though it violates the assumptions of Patlak analysis due to physiologic dephosphorylation and efflux of [18F]FDG [12]. In contrast, [18F]FDG dephosphorylation has minimal effects on metabolic rate measurements in grey matter during the time points assessed in our study [13], making the grey matter an ideal reference organ for tracking inter-scan PS stability.
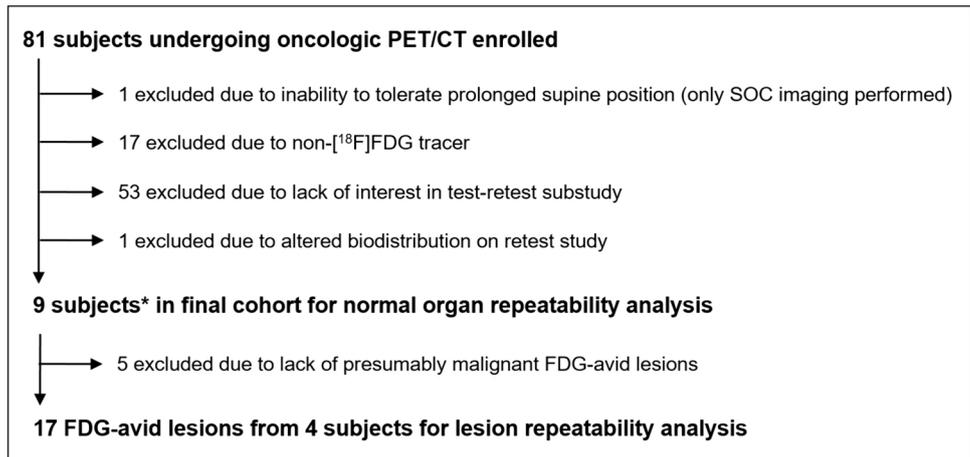
Results

Study Cohort

Among 81 study subjects, 10 opted to participate in the test–retest substudy; 9/10 were eligible for the normal organ repeatability analysis, and 4/10 (with 17 lesions) were eligible for the lesion repeatability analysis (Fig. 1). The normal organ cohort was 66.7% female (6/9), with a mean age of 61.7 years. Additional patient and scan characteristics are summarized in Table 1.

AUTAC Analysis

Results of the AUTAC analysis are summarized in Suppl. Table 2 (see ESM). A substantial difference between the %Δ AUTAC and the %Δ [18F]FDG dose existed for Subject 2 only (-44.1%). After correcting an error in the scanner's automated placement of the aortic VOI (Suppl. Figures 2 and 3; see ESM), the difference between the %Δ AUTAC and the %Δ [18F]FDG dose for Subject 2 fell within the expected range (7.5%). The adjusted Subject 2 IF was utilized for Patlak reconstruction. There were no large differences in blood

Fig. 1 Study flowchart. Abbreviations: SOC = standard-of-care. * 1 of 9 subjects did not complete post-SOC dynamic whole-body passes, so this subject was included in the normal organ repeatability analysis for the early time point only.

81 subjects undergoing oncologic PET/CT enrolled

→ 1 excluded due to inability to tolerate prolonged supine position (only SOC imaging performed)

→ 17 excluded due to non-[18F]FDG tracer

→ 53 excluded due to lack of interest in test-retest substudy

→ 1 excluded due to altered biodistribution on retest study

9 subjects* in final cohort for normal organ repeatability analysis

→ 5 excluded due to lack of presumably malignant FDG-avid lesions

17 FDG-avid lesions from 4 subjects for lesion repeatability analysis

Table 1 Subject and scan characteristics

Characteristic	Value
Age (mean ± standard deviation)	61.7 ± 6.2
Sex – n (%)	
Male	3 (33.3%)
Female	6 (66.7%)
Non-binary	0 (0.0%)
Cancer type – n (%)	
Thyroid	2 (22.3%)
Lung	1 (11.1%)
Breast	1 (11.1%)
Colorectal	1 (11.1%)
Lymphoma	1 (11.1%)
Cervical	1 (11.1%)
Multiple myeloma	1 (11.1%)
Atypical lymphocyte proliferation	1 (11.1%)
Indication for PET – n (%)	
Diagnosis of suspected malignancy	1 (11.1%)
Initial staging of confirmed malignancy	1 (11.1%)
Restaging during/after treatment	7 (77.8%)
Detection of suspected recurrence	0 (0.0%)
Surveillance	0 (0.0%)
Number of lesions analyzed per patient – n (%)	
0	4 (44.5%)
1	0 (0.0%)
2	2 (22.2%)
3	0 (0.0%)
4	0 (0.0%)
5	3 (33.3%)

Table 2 Repeatability of SUV and PS metrics among lesions

Metric	Mean T-RT %Δ [95% CI]	wCV	ICC
SUV-early-max	3.3% [-3.2%, 9.8%]	8.9%	0.96
SUV-early-peak	2.9% [-3.0%, 8.8%]	8.1%	0.98
PS-early-max	19.1% [7.3%, 30.9%]*	16.2%	0.95
PS-early-peak	17.8% [6.4%, 29.2%]*	15.6%	0.98
SUV-late-max	10.0% [2.6%, 17.0%]*	9.7%	0.93
SUV-late-peak	3.6% [-1.6%, 8.8%]	7.2%	0.98
PS-late-max	20.4% [14.3%, 26.4%]*	8.5%	0.94
PS-late-peak	16.2% [11.6%, 20.8%]*	6.4%	0.96

Abbreviations: *CI* confidence interval; *ICC* intraclass correlation coefficient; *PS* Patlak slope; *SUV* standardized uptake value; *T-RT* test–retest; *wCV* within-subject coefficient of variation

*Mean T-RT %Δ is significantly different from 0% at 95% confidence level

glucose levels between sessions for any subjects, including the four subjects in the lesion repeatability analysis (subjects 1–4; %Δ BG range: 11.2% to -13.4%).

Lesion Repeatability Analysis

Results of the lesion repeatability analysis are summarized in Table 2; source data for this analysis are provided in Suppl. Tables 3 and 4 (see ESM). The wCV was higher for PS-early-max (16.2%) than PS-late max (8.5%) and higher for PS-early-peak (15.6%) than PS-late-peak (6.4%), whereas the wCVs were similar for SUV metrics at the early and late time points. The mean T-RT %Δ was > 0.0% for all metrics (i.e., bias toward higher retest values), though generally of greater magnitude for PS metrics (e.g., PS-late-max: 20.4%) than the corresponding SUV metrics (SUV-late-max: 10.0%). The mean T-RT %Δ was significantly > 0.0% for all PS-based metrics, such as PS-late-max (95% CI: 14.3%, 26.4%), as well as for SUV-late-max (95% CI: 2.6%, 17.0%). This bias is also captured in the Bland–Altman plots for representative metrics (Fig. 2), with a preponderance of data points above the 0.0% T-RT %Δ line (though less pronounced for SUV-early-max). Scatterplots of lesion test–retest values for select metrics (Fig. 3) similarly show a bias toward higher retest values. However, the data points generally remain near or slightly above the unity line, resulting in ICCs near 1.0. In general, the ICCs were similar for SUV (0.93–0.98) and PS (0.94–0.98) metrics across the early and late time points.

Fig. 2 Bland–Altman lesion repeatability analysis for select metrics. Bland–Altman plots are shown for PS-early-max (**A**), PS-late-max (**B**), (**C**) SUV-early-max, and (**D**) SUV-late-max. Horizontal orange lines represent the mean T-RT %Δ. Horizontal dashed and solid black lines represent ± 1 standard deviation and ± 1.96 standard deviations, respectively, relative to the mean of the T-RT %Δ distribution. Abbreviations: PS = Patlak slope; SUV = standardized uptake value; T-RT = test–retest; wCV = within-subject coefficient of variation.
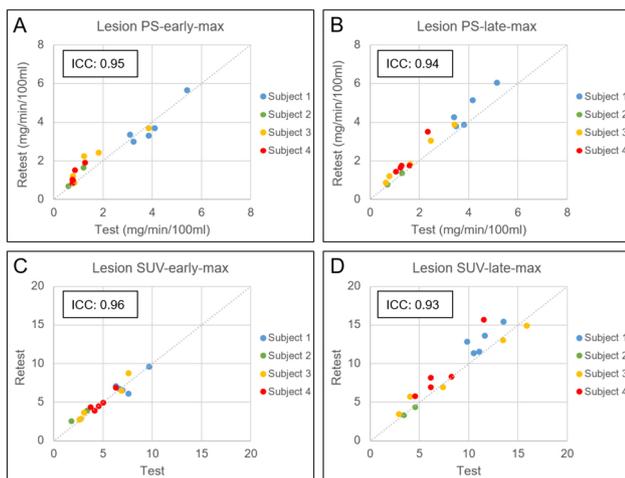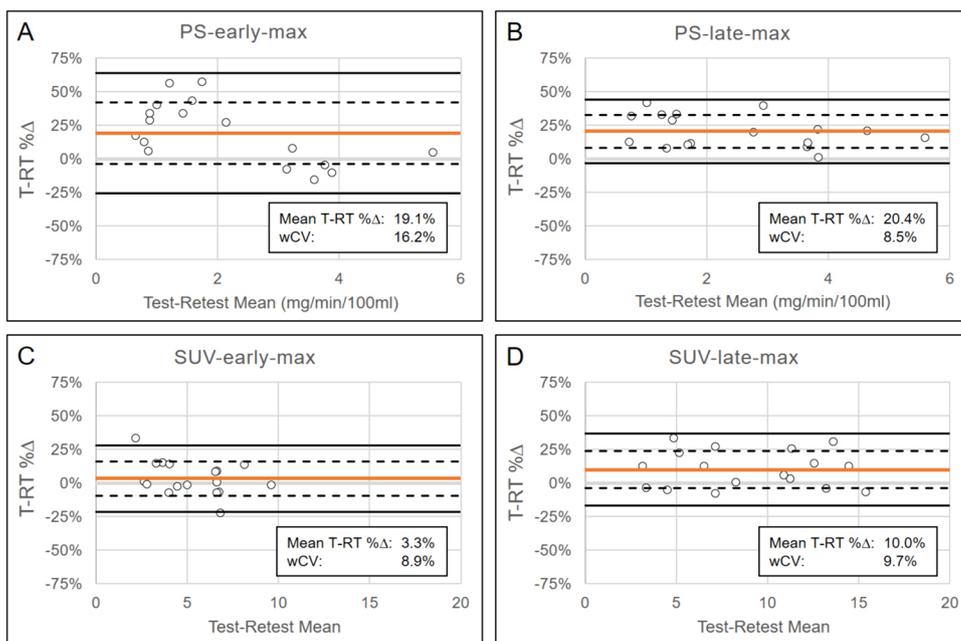
Fig. 3 Scatterplots from lesion repeatability analysis for select metrics. Lesion retest versus test values are plotted for PS-early-max (**A**), PS-late-max (**B**), (**C**) SUV-early-max, and (**D**) SUV-late-max. Individual subjects are color-coded. Dashed grey unity line (x = y) represents perfect repeatability. Dots above the unity line indicate higher values on retest. Abbreviations: ICC = intraclass correlation coefficient.

Normal Organ Repeatability Analysis

Results of the normal organ repeatability analysis are summarized in Table 3, with descriptive statistics for each organ in Suppl. Table 5 (see ESM). In contrast to the lesion analysis, there was no bias toward higher retest values for the normal organs, as the mean T-RT %Δ values were fairly evenly distributed above and below 0.0% (all within ± 10%, except for PS-early-mean in the liver). None of the mean T-RT %Δ values were significantly different from 0.0%,

with the exception of SUV-mean-early (95% CI: -11.5%, -2.0%) in the liver. As in the lesion analysis, the normal organ wCVs were mostly lower at the late time point for PS metrics but similar between early and late time points for SUV metrics. For example, the wCV for grey matter was substantially higher for PS-early-mean (13.0%) than for PS-late-mean (6.4%), with a relatively small difference in wCVs between SUV-early-mean (12.4%) and SUV-late-mean (9.6%). In general, the wCVs were lower (i.e., better repeatability) for SUV metrics than for PS metrics at equivalent time points, with the exception of grey matter at the late time point. In terms of ICCs, repeatability was also generally better for SUV metrics than PS metrics, including the liver (Fig. 4) and grey matter (Fig. 5), though with similar values for the grey matter at the late time point. These scatterplots confirm a lack of bias toward higher retest values, with data points fairly evenly distributed above and below the unity lines for most metrics.

Analysis of Repeatability Results by Subject

Table 4 summarizes the oncologic history and post-study follow-up for the four subjects included in the lesion analysis, along with T-RT %Δ values for the analyzed lesions and several normal organs. Colored dots in Fig. 3 represent individual lesions for each subject, as numbered in Table 4. These subjects are also color-coded in Figs. 4 and 5. Fairly large positive T-RT %Δ values were seen for multiple lesions from subjects 1, 3, and 4 at the late time point for both SUV and PS (e.g., subject 1, lesion 4: SUV-late-max, 14.6%; PS-late-max, 20.5%; see Fig. 6). Despite this bias toward higher lesion retest values, the grey matter and liver did not show

Table 3 Repeatability of SUV and PS metrics for normal organs

| Metric	Mean T-RT %Δ [95% CI]	wCV	ICC
Liver			
SUV-early-mean	-6.7% [-11.5%, -2.0%]*	4.3%	0.74
PS-early-mean	-10.6% [-28.3%, 7.1%]	16.3%	0.37
SUV-late-mean	-4.3% [-8.5%, -0.05%]	3.6%	0.73
PS-late-mean	-0.9% [-11.4%, 9.6%]	8.9%	0.55
Lung			
SUV-early-mean	-2.0% [-9.1%, 5.1%]	6.6%	0.93
PS-early-mean	-4.2% [-32.0%, 23.6%]	25.6%	0.68
SUV-late-mean	1.3% [-7.6%, 10.2%]	7.6%	0.89
PS-late-mean	5.5% [-5.1%, 16.1%]	9.0%	0.76
Blood pool			
SUV-early-mean	-2.8% [-9.5%, 3.9%]	6.2%	0.73
PS-early-mean	9.4% [-8.8%, 27.6%]	16.8%	0.47
SUV-late-mean	-0.6% [-6.1%, 4.9%]	4.7%	0.86
PS-late-mean	1.0% [-11.9%, 13.9%]	11.0%	0.23
Bone marrow			
SUV-early-mean	9.0% [-0.1%, 18.1%]	8.4%	0.84
PS-early-mean	1.8% [-8.7%, 12.3%]	9.7%	0.95
SUV-late-mean	-1.8% [-10.7%, 7.1%]	7.6%	0.95
PS-late-mean	1.4% [-14.3%, 17.1%]	13.3%	0.84
Skeletal muscle			
SUV-early-mean	-4.7% [-16.1%, 6.7%]	10.6%	0.62
PS-early-mean	1.4% [-24.9%, 27.7%]	24.3%	0.53
SUV-late-mean	-3.3% [-19.6%, 13.1%]	13.9%	0.47
PS-late-mean	-0.5% [-18.5%, 17.5%]	15.3%	0.39
Grey matter			
SUV-early-mean	1.0% [-12.4%, 14.4%]	12.4%	0.75
PS-early-mean	-8.0% [-22.1%, 6.1%]	13.0%	0.39
SUV-late-mean	-3.6% [-14.9%, 7.8%]	9.6%	0.86
PS-late-mean	0.5% [-7.0%, 8.0%]	6.4%	0.75
Spleen			
SUV-early-mean	-2.3% [-12.3%, 7.7%]	9.3%	0.43
PS-early-mean	4.7% [-21.3%, 30.7%]	24.0%	0.14
SUV-late-mean	2.4% [-11.9%, 16.7%]	12.1%	0.41
PS-late-mean	6.0% [-12.3%, 24.3%]	15.5%	0.36

Abbreviations: *ICC* intraclass correlation coefficient; *PS* Patlak slope; *SUV* standardized uptake value; *T-RT* test–retest; *wCV* within-subject coefficient of variation

*Mean T-RT %Δ is significantly different from 0.0% at 95% confidence level

the same bias for these subjects (e.g., subject 1, grey matter: SUV-late-max, -0.2%; PS-late-max, 3.0%; see Fig. 6).

Discussion

In this study, we assessed the test–retest repeatability of PS versus SUV among tracer-avid lesions on oncologic [^{18}F]FDG-PET/CT and found that the wCVs of PS-based
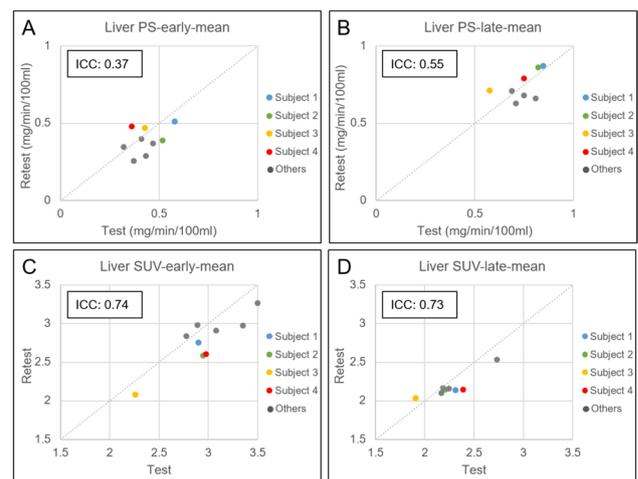
Fig. 4. Scatterplots from normal liver repeatability analysis for select metrics. Normal liver retest versus test values are plotted for PS-early-max (**A**), PS-late-max (**B**), SUV-early-max (**C**), and SUV-late-max (**D**). Individual subjects from the lesion analysis are color-coded. The other subjects correspond to grey dots. Dashed grey unity line (x = y) represents perfect repeatability. Dots above the unity line indicate higher values on retest. Abbreviations: ICC = intraclass correlation coefficient.
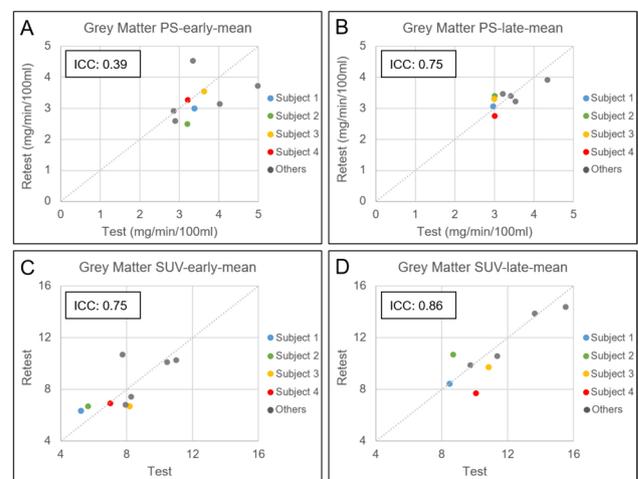
Fig. 5 Scatterplots from normal grey matter repeatability analysis for select metrics. Normal grey matter retest versus test values are plotted for PS-early-mean (**A**), PS-late-mean (**B**), SUV-early-mean (**C**), and SUV-late-mean (**D**). Individual subjects from the lesion analysis are color-coded. The other subjects correspond to grey dots. Dashed grey unity line (x = y) represents perfect repeatability. Dots above the unity line indicate higher values on retest. Abbreviations: ICC = intraclass correlation coefficient.

metrics were generally worse than those of equivalent SUV-based metrics at the early post-injection time point (e.g., SUV-early-max: 8.9%; PS-early-max: 16.2%) but similar at the late post-injection time point (e.g., SUV-late-max: 9.7%; PS-late-max: 8.5%). As lesion retest values

Table 4 Detailed clinical features of subjects in lesion analysis

Subject	Cancer / treatment	PET indication	Follow-up	T-RT interval	Lesion or *normal organ*	T-RT %Δ SUV-early[a]	T-RT %Δ PS-early[a]	T-RT %Δ SUV-late[a]	T-RT %Δ PS-late[a]
1	Metastatic well-differentiated, intermediate-grade NET; on TKI (clinical trial) for 18 months	Restaging during/after therapy	CT one week after PET/CT showed anatomically stable disease at all sites, relative to CT from 3 months before PET/CT	7 days	Lesion 1 (liver)	-2.0%	4.7%	12.2%	15.4%
					Lesion 2 (liver)	8.9%	-10.6%	25.1%	21.4%
					Lesion 3 (liver)	0.3%	7.7%	3.0%	8.7%
					Lesion 4 (liver)	-6.8%	-7.9%	14.6%	20.5%
					Lesion 5 (l.n.)	-22.5%	-15.6%	5.8%	0.8%
					Grey matter	18.9%	-13.2%	-0.2%	3.0%
					Liver	-5.3%	-12.8%	-8.5%	2.3%
2	Lung adenocarcinoma s/p CRT (4 mos prior); on consolidation ICI for 3 months	Restaging during/after therapy	Repeat PET/CT 6 months later showed stable disease	7 days	Lesion 1 (RML)	14.8%	33.6%	-5.3%	7.5%
					Lesion 2 (LUL)	33.0%	16.8%	-3.9%	12.4%
					Grey matter	16.5%	-24.6%	20.7%	12.5%
					Liver	-13.0%	-28.6%	-3.2%	4.8%
3	Unknown (mesenteric and inguinal lymphadenopathy); no prior treatment	Diagnosis of suspected malignancy	Excisional biopsies showed atypical lymphocytic infiltration without monoclonal population	6 days	Lesion 1 (l.n.)	13.1%	-5.0%	-4.2%	12.0%
					Lesion 2 (l.n.)	-7.5%	27.1%	-6.9%	19.8%
					Lesion 3 (l.n.)	14.4%	57.1%	-7.8%	11.5%
					Lesion 4 (l.n.)	0.8%	40.0%	33.0%	41.2%
					Lesion 5 (l.n.)	-1.1%	5.8%	12.3%	31.6%
					Grey matter	-20.5%	-3.1%	-11.7%	8.9%
					Liver	-8.8%	8.9%	5.6%	20.2%
4	Colorectal adenoca. with liver metastases; on TKI + ICI (clinical trial) for 2 months	Restaging during/after therapy	Liver MRI 1 month after PET/CT showed anatomic progression of liver lesions relative to MRI from 1 month before PET/CT	6 days	Lesion 1 (liver)	-1.6%	56.2%	0.1%	10.0%
					Lesion 2 (liver)	7.9%	42.8%	30.5%	39.5%
					Lesion 3 (liver)	14.1%	28.2%	22.0%	32.4%
					Lesion 4 (liver)	-2.7%	33.7%	12.4%	28.6%
					Lesion 5 (liver)	-7.2%	12.5%	26.8%	33.1%
					Grey matter	-1.6%	1.5%	-26.7%	-8.3%
					Liver	-13.3%	28.6%	-11.0%	5.2%

Abbreviations: *CRT* chemoradiotherapy; *ICI* immune checkpoint inhibitor; *l.n.* lymph node; *LUL* left upper lobe; *NET* neuroendocrine tumor; *PS* Patlak slope; *RML* right middle lobe; *s/p* status-post; *SUV* standardized uptake value; *TKI* tyrosine kinase inhibitor; *T-RT* test–retest

[a]For both PS and SUV, T-RT %Δ is based on maximum values for lesions and mean values for normal organs
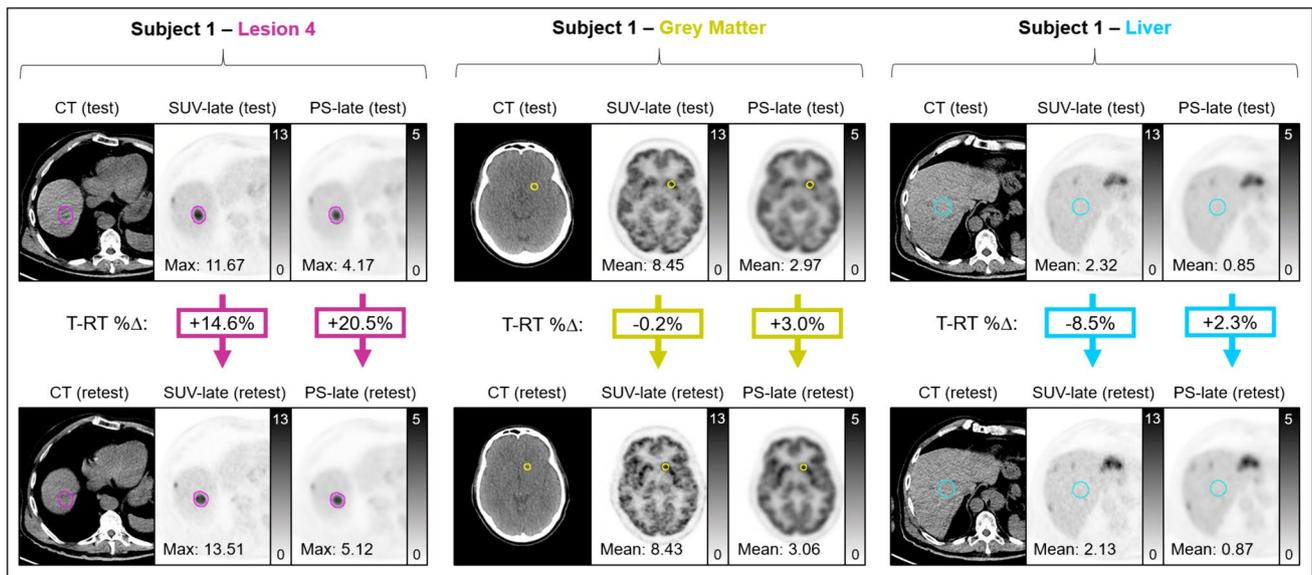
Fig. 6 Stability of normal organs despite higher lesion retest activity. Late test and retest SUV and Patlak slope (PS) images are shown for subject 1, highlighting a substantial test-to-retest increase in the SUV-max and PS-max for lesion 4. Other subject 1 lesions (not shown) also exhibited substantial test-to-retest increases. In contrast, the SUV-mean and PS-mean of the grey matter and liver were very similar for test and retest sessions. Abbreviations: T-RT = test–retest.

were significantly higher than test values for many metrics (likely a combination of true progression and stochastic variation in the setting of a small sample size), we alternatively utilized the ICC as a marker of repeatability and found high degrees of correlation (ICC: 0.93–0.98) between test and retest values for all PS and SUV metrics. Among normal organs, as for tracer-avid lesions, we found wCVs were generally lower (i.e., better repeatability) for SUV-based metrics than equivalent PS-based metrics and also lower for later time points than early time points (particularly for PS). For nearly all metrics across nearly all normal organs evaluated, the test and retest values were statistically similar at the early and late time points.

Robust repeatability metrics are required for PS-based metrics to be incorporated into clinical trials and treatment response algorithms. To our knowledge, no prior studies have assessed the test–retest repeatability of PS-based metrics on whole-body [18F]FDG-PET/CT. Several studies from the pre-PET/CT era compared the repeatabilities of SUV and PS for malignant lesions on [18F]FDG-PET [14, 15]. However, the analytic methodologies utilized in these studies preclude direct comparison to our results. For example, Weber et al. utilized mean lesion values rather than maximum/peak values, which were assessed in our study and are more commonly utilized clinically. Furthermore, in that study, the uptake times for utilized PET data were not identical for SUV and PS reconstructions. Importantly, our study showed that repeatability (as captured by the wCV) can change during the uptake period. That said, the prior study reported wCVs of 6.4% and

7.1% for SUV-mean and PS-mean, respectively. For comparison, we observed wCVs of 7.2% and 6.4% for SUV-late-peak and PS-late-peak, respectively. Our wCVs for maximum SUVs and maximum PS values were slightly higher, as maximum values are intrinsically noisier given their dependence on single voxels. Furthermore, our wCVs of 8.9% and 9.7% for SUV-early-max and SUV-late-max align with expected values, based on prior [18F]FDG-PET/CT studies [7].

Surprisingly, we observed a statistically significant bias toward higher values for lesions in the retest study than in the test study. This effect was apparent for all PS-based metrics (e.g., PS-late-max: mean T-RT %Δ, 20.4% [14.3%, 26.4%]), as well as for SUV-late-max (mean T-RT %Δ, 10.0% [2.6%, 17.0%]), suggesting that this finding was not simply related to an error in acquisition, reconstruction, or analysis of the PS images. This notion is further corroborated by the lack of bias toward higher retest values among normal organs, indicating that the observed bias was likely derived from true biological variation in the lesions assessed. Furthermore, the small %Δ BG values, particularly for the subjects with hypermetabolic lesions, confirm that the higher retest values were not simply a consequence of including BG values in the PS calculation (i.e., MR$_{FDG}$). On assessment of the clinical features for the four subjects included in the lesion analysis, subject 4 had enlargement of the study lesions between MRI examinations performed before and after study imaging (2-month interval). Thus, the substantial test-to-retest increases in lesion SUV and PS for this subject (particularly lesions 2–5), without

corresponding increases in liver and grey matter, likely represented true metabolic progression during the six-day test–retest period. In contrast, subject 1 had anatomic stability of the study lesions between CT examinations performed before and after study imaging (3-month interval). Therefore, the substantial test-to-retest increases in SUV and PS for this subject (particularly lesions 1, 2, 4), again without corresponding increases in liver and grey matter, were more likely due to stochastic temporal variations in the [^{18}F]FDG-avidity of the study lesions. Because the 17 lesions analyzed in our study were derived from only four subjects, the results of our analysis depend heavily on the particular behaviors of their individual lesions. Hence, if a larger percentage of enrolled subjects had had tracer-avid lesions available for analysis, the bias toward higher retest values might have been negated by directionally opposite stochastic variations in lesion activity.

Our study has several limitations, including its single-center design and utilization of a single make/model of PET scanner. Our findings will need to be confirmed on different Patlak-enabled scanners at other institutions. As discussed in detail above, the greatest limitation of our study was the small number of patients with tracer-avid lesions, which likely contributed to the observed bias. A shorter test–retest interval may have mitigated this bias by reducing the potential for true biological changes in tumor activity. Thus, studies with larger numbers of subjects and more diverse pathologies are needed to confirm our lesion repeatability data. Furthermore, as only patients who reported an ability to tolerate prolonged supine positioning were enrolled (with an even smaller subset agreeing to the test–retest substudy), the subjects in our analysis were particularly well-suited to the demands of dynamic PET imaging (i.e., prolonged motionlessness). Consequently, the true repeatability of PS-based metrics may be worse in a general oncology population. However, PS-based metrics may prove more repeatable than SUV-based metrics in actual clinical practice, as scan-to-scan variations in uptake times often occur and would be expected to affect the repeatability of SUV-based metrics more than PS-based metrics. Finally, future studies might assess the repeatability of PS-based metrics when a population-based input function (PBIF) is employed, as the PBIF has proven to be a useful, relatively unbiased surrogate for the image-derived input function (IDIF), such as when the IDIF is unavailable or unreliable [16, 17].

Conclusions

In this study, we found that PS-based metrics of tumor activity exhibited similar repeatability to equivalent SUV-based metrics, provided imaging was performed at late post-injection time points. However, our study was limited by the small number of subjects with analyzable lesions, resulting in a significant bias toward higher retest values in the lesion analysis As such, the specific values of repeatability metrics derived from our data require validation in larger, more diverse patient cohorts. For example, the true value of the wCV for PS-max among [^{18}F]FDG-avid lesions might, on average, be higher than what we have reported. Consequently, we caution against utilizing our results in isolation to inform PS-based treatment response assessments. Despite this limitation, our study points to several issues that are important to consider for future more robust PS repeatability studies, as well as for utilization of Patlak images in routine clinical PET interpretation. More specifically, confirmation of accurate VOI placement is essential when image-based input functions are utilized. Time-activity curve analysis can also be helpful to screen for VOI placement errors or to identify issues related to suboptimal bolus injection (e.g., extravasation, venous stenosis), particularly when the injection site is not within the PET field-of-view. Finally, reviewing the whole-body passes dynamically to screen for patient motion is a critical element of quality control, as patient motion can introduce artifacts in images derived from Patlak analysis.

Declarations

References

1. Keyes JW (1995) SUV: Standard uptake or silly useless value? J Nucl Med 36:1836–1839
2. Beaulieu S, Kinahan P, Tseng J et al (2003) SUV varies with time after injection in 18F-FDG PET of breast cancer: Characterization and method to adjust for time differences. J Nucl Med 44:1044–1050
3. Lee JW, Kim SK, Lee SM, Moon SH, Kim TS (2011) Detection of hepatic metastases using dual-time-point FDG PET/CT scans in patients with colorectal cancer. Mol Imaging Biol 13:565–572
4. Rahmim A, Lodge MA, Karakatsanis NA et al (2019) Dynamic whole-body PET imaging: principles, potentials and applications. Eur J Nucl Med Mol Imaging 46:501–518
5. Zasadny KR, Wahl RL (1996) Enhanced FDG-PET tumor imaging with correlation-coefficient filtered influx-constant images. J Nucl Med 37:371–374
6. Verhaeghe J, Gravel P, Mio R et al (2010) Motion compensation for fully 4D PET reconstruction using PET superset data. Phys Med Biol 55:4063–4082
7. Kinahan PE, Perlman ES, Sunderland JJ et al (2020) The QIBA profile for FDG PET/CT as an imaging biomarker measuring response to cancer therapy. Radiology 294:647–657
8. Wahl RL, Jacene H, Kasamon Y, Lodge MA (2009) From RECIST to PERCIST: Evolving Considerations for PET response criteria in solid tumors. J Nucl Med 50(Suppl 1):122S-S150
9. Tao Y, Peng Z, Krishnan A, Zhou XS (2011) Robust learning-based parsing and annotation of medical radiographs. IEEE Trans Med Imaging 30:338–350
10. Martin Bland J, Altman DG (1986) Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 327:307–310
11. Koo TK, Li MY (2016) A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. J Chiropr Med 15:155–163
12. Chirindel A, Alluri KC, Tahari AK et al (2015) Liver standardized uptake value corrected for lean body mass at FDG PET/CT: Effect of FDG uptake time. Clin Nucl Med 40:e17–e22
13. Kuwabara H, Gjedde A (1991) Measurements of glucose phosphorylation with FDG and PET are not reduced by dephosphorylation of FDG-6-phosphate. J Nucl Med 32:692–698
14. Minn H, Zasadny KR, Quint LE, Wahl RL (1995) Lung cancer: Reproducibility of quantitative measurements for evaluating 2-[F-18]-fluoro-2-deoxy-D-glucose uptake at PET. Radiology 196:167–173
15. Weber WA, Ziegler SI, Thödtmann R, Hanauske AR, Schwaiger M (1999) Reproducibility of metabolic measurements in malignant tumors using FDG PET. J Nucl Med 40:1771–1777
16. Naganawa M, Gallezot JD, Shah V et al (2020) Assessment of population-based input functions for Patlak imaging of whole body dynamic 18F-FDG PET. EJNMMI Phys 7:67
17. Dias AH, Smith AM, Shah V, Pigg D, Gormsen LC, Munk OL (2022) Clinical validation of a population-based input function for 20-min dynamic whole-body 18F-FDG multiparametric PET imaging. EJNMMI Phys 9:60