EJNMMI Research

ORIGINAL RESEARCH

Open Access

Test-retest repeatability of quantitative organ and tissue uptake using 20-minute dynamic multiparametric whole-body [^{18}F]FDG PET/CT in patients with type 2 diabetes

Jonathan M. Baier[1,2], Kristian L. Funck[2,3], Anna Dons-Jensen[2], Ole L. Munk[1,5], Lars P. Tolbod[1,5], Esben Laugesen[2,4], Per L. Poulsen[1,2], Lars C. Gormsen[1,5] and André H. Dias[5*]

Abstract

Background Recently developed dynamic whole-body PET/CT (D-WB PET/CT) protocols allow for measurements of potentially more precise metabolic parameters than the commonly used semiquantitative SUV. Most notable is the metabolic rate of FDG uptake (MR_{FDG}), which reflects quantitative glucose uptake into tissues and organs. However, data on the reproducibility of MR_{FDG} measurements are scarce, particularly in patients with perturbed glucose homeostasis such as type 2 diabetes. We therefore aimed to evaluate the test-retest repeatability of both MR_{FDG} and SUV in these patients.

Results Fifteen participants (mean age 71 ± 7 years; 2 females) with type 2 diabetes underwent a short 20-minute [^{18}F]FDG D-WB PET/CT after 6 h fasting on two consecutive days. Both SUV and MR_{FDG} images were reconstructed from D-WB PET/CT data obtained 60–80 min post-injection of [^{18}F]FDG. MR_{FDG} and SUV data were measured in organs and tissues, and repeatability was assessed with Bland-Altman analysis, intraclass correlation coefficients (ICC), repeatability coefficients (RPC) and coefficients of variation (wCV). There was high repeatability of both SUV_{mean} and $MR_{FDG–mean}$ in all measured organs (ICC range: 0.65–0.95 for SUV_{mean} and 0.66–0.94 for $MR_{FDG–mean}$). SUV_{mean} generally demonstrated higher reliability (ICC) and lower variability (%RPC and %wCV) when compared to $MR_{FDG–mean}$. However, MR_{FDG} test-retest variation was < 19% in most analysed tissues, demonstrating that MR_{FDG} may be used as a precise marker of treatment response.

Conclusion This study demonstrates that MR_{FDG} calculated from D-WB PET/CT exhibit high repeatability, comparable to SUVs across most organs in patients with type 2 diabetes.

Keywords Test-retest, PET/CT, Dynamic whole-body PET, Parametric imaging, FDG, Diabetes

*Correspondence:
André H. Dias
andre.dias@auh.rm.dk
[1]Dept. of Clinical Medicine, Aarhus University, Aarhus, Denmark
[2]Steno Diabetes Center Aarhus, Aarhus University Hospital, Aarhus, Denmark
[3]Dept. of Endocrinology and Internal Medicine, Aarhus University Hospital, Aarhus, Denmark
[4]Diagnostic Centre, Silkeborg Regional Hospital, Silkeborg, Denmark
[5]Department of Nuclear Medicine & PET-Centre, Aarhus University Hospital, Palle Juul-Jensens Boulevard 165, Aarhus N DK-8200, Denmark

Introduction

Positron emission tomography (PET) with ^{18}F-fluorodeoxyglucose ([^{18}F]FDG) has become a key tool in the diagnostic and therapeutic follow-up of both malignant and inflammatory diseases. In daily clinical practice, the analysis of PET images is mostly based on a visual qualitative evaluation. This is often supplemented by a semiquantitative analysis using the standardized uptake value (SUV), a measure of relative tracer radioactivity in a region of interest, normalized to the total injected activity and body weight.

The SUV, however, has several inherent shortcomings [1]. Pointedly, the SUV reflects not only the metabolically phosphorylated [^{18}F]FDG, but also includes the unbound circulating tracer. This is particularly important in the evaluation of highly perfused areas such as the vessels, the kidneys, and the liver. Furthermore, as SUV measurement is extremely dependent on the exact timing of the scan after tracer injection, it may be seriously erroneous if tracer is extravasated, and is also impacted by scanner calibration settings [2]. The European Association of Nuclear Medicine Research Ltd. (EARL) accreditation program aims to harmonize and standardise scan and reconstruction protocols, with the purpose of reducing variability both within and across departments [3, 4]. However, even with such measures in place, the evaluation of SUV remains prone to biases, which have limited the widespread use of semiquantitative values both in a clinical and research setting.

Recent technological advances have made dynamic whole-body PET/CT (D-WB PET/CT) easier to implement in clinical routine [5]. D-WB PET/CT imaging produces not only the classical SUV images but also two other reconstructions using the Patlak method, which account for the time course of the arterial input function (AIF) [6, 7]: MR_{FDG}, representing the metabolic rate of [^{18}F]FDG into the tissue; and DV_{FDG} representing the distribution volume of free [^{18}F]FDG in the reversible compartment and the fractional blood volume [8]. Thus, the Patlak parameters should be more specific and robust to the variability of the AIF.

Several studies have assessed the clinical feasibility of D-WB PET/CT [9, 10], improved the methodology thus allowing for an overall reduction of scan time [11, 12, 13], and provided a database of normal reference values for various organs and tissues [14]. However, literature evaluating the reproducibility and repeatability of multiparametric [^{18}F]FDG D-WB PET/CT remains sparse. A single study of 90 min dynamic scanning in nine patients by Ince et al. [15] found that the repeatability of Patlak and SUV values at late post-injection time points was similar in FDG avid tumour lesions and selected organs within seven days. However, in diabetic patients, widespread insulin resistance may severely hamper glucose uptake in e.g. skeletal muscle and the heart, increase hepatic gluconeogenesis, increase circulating glucose levels and consequently cause excretion of glucose by the kidneys. This disrupted glucose homeostasis is theoretically expected to impact the reproducibility of semiquantitative PET measurements.

With the present study we aimed to assess the repeatability of MR_{FDG} compared to SUVs in normal tissues and organs using [^{18}F]FDG PET/CT in a cohort of 15 patients with type 2 diabetes scanned on two consecutive days with a short 20-min D-WB protocol. This approach allows for decay of the tracer activity from the first scan and minimizes changes in the metabolic milieu that could be a consequence of altered eating habits, exercise, or drug administration.

Methods

Study design

This test-retest study was part of a randomized controlled trial (EudraCT-no.: 2021-003525-30) aiming to evaluate the effect of the anti-inflammatory drug colchicine on vessel wall inflammation in patients with type 2 diabetes measured by [^{18}F]FDG PET/CT. Of the 30 participants that underwent [^{18}F]FDG D-WB PET/CT at baseline, 15 participants opted to participate in the present reproducibility study. In these patients, retest-scans were carried out on the day following the baseline scan (24.0 ± 0.8 h mean interval between studies). Both scans were performed before the trial's pharmacological intervention. Data were collected between November 2022 and May 2023 at a single trial site at Department of Nuclear Medicine & PET-Centre, Aarhus University Hospital, Aarhus, Denmark. The trial was approved by the local ethics committee (Committees on Health Research Ethics of the Central Region of Denmark, approval no. 1-10-72-345-21) and complied with the Helsinki declaration. All participants gave written informed consent prior to enrolment.

Trial participants

Patients with type 2 diabetes and $HbA1_C \geq 48$ mmol/mol (6,5%) and either age ≥ 50 years and a history of cardiovascular disease or age ≥ 60 years and at least 1 risk factor for CVD (e.g. albuminuria), were eligible for inclusion. Participants were excluded if they had an estimated glomerular filtration rate (eGFR) < 50 mL/min/1.73 m^2, active non-cutaneous cancer, liver disease, severe heart failure, inflammatory bowel disease or cytopenia (anaemia, leukopenia, or thrombocytopenia). An exhaustive list of inclusion and exclusion criteria can be found in the Supplemental Table 1.

Imaging protocol

All participants were scanned using a fully automated short multiparametric 20-minute [^{18}F]FDG D-WB PET protocol 60–80 min post-injection performed on a Siemens Biograph Vision 600 PET/CT scanner (Siemens Healthineers, Knoxville, TN, USA). Participants fasted for at least 8 h prior to imaging and abstained from caffeine and nicotine products on the study days. Blood glucose levels and blood pressure were measured in all participants. A standardized tracer infusion was administered using an Intego PET Infusion System (MEDRAD, Inc., Warrendale, PA, USA) through a peripheral venous catheter in the participant's arm.

A low-dose WB CT (25 Ref mAs, 120 kV, Care Dose4D, Care kV, Admire level 3) was performed, followed by a 20-minute multiparametric PET acquisition protocol, consisting of 4 × 5 min WB passes, started 60 min after tracer administration. The same protocol was repeated on the following day for retest evaluation.

Input function analysis

To obtain multiparametric images from this short 20-min D-WB PET acquisition protocol, a population-based input function (PBIF) was used, scaled to the tail of the image-derived input function (IDIF) following the methodology described in the study by Dias et al. [12].

In brief, an automated delineation tool, developed by Siemens Healthineers, known as automated learning and parsing of human anatomy (ALPHA) [16] was employed to generate a 1.6 mm^3 cylinder VOI in the proximal descending aorta in order to extract an IDIF. The correct position of the VOI was verified by visual inspection. The PBIF was then individually applied by scaling it to the participant's IDIF as measured on the four late passes. Thus, this scaled PBIF (sPBIF) accounts for variations in injected dose, the patient's body composition, and tracer distribution.

PET image reconstruction

In this study, 4 PET reconstructions were performed: (1) The standard-of-care SUV image using PET data from 60 to 70 min (two passes): TrueX + TOF, 4 iterations, 5 subsets, 440 × 440 matrix, 2-mm Gaussian filter, and relative scatter correction; (2) An SUV image using PET data from 60 to 80 min (four passes): TrueX + TOF, 6 iterations, 5 subsets, 440 × 440 matrix, no filtering, and relative WB scatter correction; (3) and (4) Direct Patlak with non-negativity constraints [6] was performed to reconstruct the two parametric images (MR_{FDG} and DV_{FDG}) using data from 60 to 80 min (four passes) using the previously described PBIF: TrueX + TOF, 8 iterations, 5 subsets, 30 nested loops, 440 × 440 matrix, 2-mm Gaussian filter, and relative scatter correction.

Image analysis

After image acquisition, the deep-learning (DL) software TotalSegmentator [17] was applied to the low-dose WB CT, automatically delineating structure VOIs. For this study, we selected VOIs of the aortic wall, myocardium of the left ventricle, right kidney, liver, lungs, pancreas, spleen, and vertebra (L3). Furthermore, geometric VOIs were placed in the liver (sphere, 30 mm radius), spleen (sphere, 20 mm radius), and lung (sphere, 20 mm radius), since these areas are more prone to movement which can introduce biases to the quantitative measurements when using VOIs automatically delineated from the CT information. All VOIs were individually checked by J.M.B and A.H.D. and manually adjusted to the respective areas if necessary using PMOD® 4.0 (PMOD Technologies Ltd., Zürich, Switzerland).

Quantitative values of SUV_{mean}, $MR_{FDG-mean}$ and $DV_{FDG-mean}$ were extracted respectively from the $SUV_{60-80\ min}$, $SUV_{60-70\ min,}$ and the Patlak PET reconstructions. MR_{FDG} is calculated as Ki × plasma glucose and can be converted to the metabolic rate of glucose (MR_{Glu}) using $MR_{Glu} = MR_{FDG}/LC$, where LC is the lumped constant. All images and tables assume LC = 1, as LC is tissue-dependent and not established for all tissues. To provide a more comprehensive analysis and account for body composition variations, additional calculations were performed, including SUV_{glu} (plasma glucose-corrected SUV), SUL (lean body mass-corrected SUV), and SUL_{glu} (plasma glucose-corrected lean body mass-corrected SUV). These metrics while not commonly used in a clinical setting can potentially reduce intersubject variability and provide more accurate assessments of tracer kinetics. Specifically, SUV_{glu} was calculated by multiplying SUV by the plasma glucose concentration, while SUL and SUL_{glu} were obtained by normalizing SUV and SUV_{glu} to lean body mass, respectively.

Statistical analysis

Normally distributed data are presented as mean ± standard deviation or with 95% confidence intervals, data with a non-normal distribution are presented as median values with interquartile ranges. Data normality was evaluated with histograms and QQ-plots.

The repeatability of SUVs, MR_{FDG} and DV_{FDG} was assessed with a Bland-Altman Analysis similar to Jochumsen et al. [18]. Non-normal distributed data were log-transformed, and the within-subject coefficient of variability, intraclass correlation coefficients (ICCs) and repeatability coefficients (RPC) were calculated according to the methodology described by Lodge et al. [19]. In brief, the RPC was calculated as $1.96 \times \sqrt{2} \times wCV$, where wCV represents the within-subject coefficient of variation for log-transformed data. Bland–Altman plots of MR_{FDG} are presented in the original scale, whereas limits

of agreement are back-transformed using the methodology of Euser et al. [20]. The percentage RPC (%RPC) was expressed for easier understanding of the measurements relative to the size of the measured values. A lower %RPC suggests better repeatability, indicating that the measurement error is small relative to the size of the measurements [21].

Examples of sample size calculations were performed using a two-sided significance test of no difference for paired log-normally distributed data, with a significance level of 5% and a power of 95%. These calculations were based on the standard deviation of the difference between the logarithm of the test and the logarithm of the retest.

Study data were collected and managed using REDCap (Vanderbilt University Medical Center, Nashville, TN, USA) [22, 23]. Data analysis was conducted using Stata version 18.1 (StataCorp LLC, College Station, TX, USA).

Results

Participant characteristics

Fifteen participants were included in this test-retest study. Mean age of the participants was 71 years, and the median diabetes duration was 13 years. Overall, the participants demonstrated effective management of traditional risk factors, including glucose levels, blood pressure, and LDL cholesterol. Detailed clinical and biochemical characteristics of the participants are presented in Table 1.

Repeatability analyses

Values of $SUV_{mean,}$ $MR_{FDG-mean}$ and $DV_{FDG-mean}$ were obtained for the automatic delineations of the aorta wall, myocardium, kidney, liver, lung, pancreas, spleen, bone, as well as for geometric VOI delineations in the liver, spleen and lung. These results, displayed in Table 2, are similar to those available in previous literature [14].

SUV_{mean} generally demonstrated higher reliability (ICC) and lower variability (%RPC and %wCV) compared to $MR_{FDG-mean}$ and $DV_{FDG-mean}$ across most organs. These results, summarized in Table 3, align with our post hoc power calculations for each organ (Supplemental Table 2), where $MR_{FDG-mean}$ required larger sample sizes than SUV_{mean} to achieve sufficient statistical power for detecting change. Reliability and variability results for $SUV(60–70)_{mean}$ and Patlak intercept values are presented in Supplemental Table 3, reliability and variability results for SUV_{glc}, SUL, and SUL_{glc}, along with the respective formulas for each calculation are presented in Supplemental Table 4.

Of all the analysed organs, the myocardium exhibited the highest reliability for both SUV_{mean} (ICC = 0.95) and $MR_{FDG-mean}$ (ICC = 0.94), albeit with considerable variability (%RPC = 40% and 65%, respectively). The lungs equally showed strong reliability (SUV_{mean} ICC = 0.85, $MR_{FDG-mean}$ ICC = 0.84) with moderate variability (%RPC = 21% and 38%, respectively). Similarly, the pancreas demonstrated good reliability for SUV_{mean} and $MR_{FDG-mean}$ but with notably higher %RPC for $MR_{FDG-mean}$ (49%).

In contrast, organs such as the spleen and kidney displayed only moderate reliability. For the spleen, SUV_{mean} exhibited an ICC of 0.73 with a %RPC of 27%, while $MR_{FDG-mean}$ showed reduced reliability (ICC = 0.67) and slightly increased variability (%RPC = 36%). The kidney showed comparable reliability for both metrics (ICC ~ 0.66–0.68), with high %RPC for $MR_{FDG-mean}$ (67%). Curiously, the liver exhibited only moderate reliability for SUV_{mean} (ICC = 0.65, % RPC = 15%) and improved reliability for $MR_{FDG-mean}$ (ICC = 0.83, %RPC = 31%).

Interestingly, the use of geometric delineation showed mixed results in the 3 chosen organs. For the liver, there was similar reliability of SUV_{mean} values and $MR_{FDG-mean}$

Table 1 Clinical and biochemical characteristics of study patients

Age– years		**71 ± 7**
Sex– n (%)	***Male***	13 (87%)
	Female	2 (13%)
Body mass index		32.0 ± 3.2
Diabetes duration– years		13 [4–17]
24 h blood pressure	***Systolic*** (mmHg)	123 ± 11
	Diastolic (mmHg)	75 ± 6
History of cardiovascular disease– N (%)		11 (73)
Biochemical characteristics	***HbA1*$_c$** (mmol/mol)	55 ± 7
	LDL cholesterol (mmol/L)	1.5 [1.2–1.9]
	B-Glc Scan Day 1 (mmol/L)	7.6 [5.8–11.8]
	B-Glc Scan Day 2 (mmol/L)	7.8 [6.0-12.9]
Medication use– N. (%)	***Metformin***	13 (87)
	SGLT2 inhibitor	7 (47)
	GLP-1 receptor agonist	7 (47)
	Insulin	2 (13)

Table 2 Quantitative study results

Organ/metric		$SUV\ (60–80)_{mean}$ (g/mL)	$SUV\ (60–70)_{mean}$ (g/mL)	$MR_{FDG–mean}$ 100 x (µmol/g/min)	$DV_{FDG–mean}$ (%)
Aortic wall	D1	1.88 [1.45–2.26]	1.9 [1.44–2.23]	2.44 [1.91–4.01]	52.53 [36.77–63.55]
	D2	1.84 [1.49–2.4]	1.86 [1.46–2.13]	2.58 [1.48–3.27]	50.82 [42.19–59.85]
Myocardium	D1	1.25 [0.86–6.88]	1.25 [0.87–6.48]	1.93 [0.88–22.62]	33.57 [16.84–68.39]
	D2	1.26 [0.85–8.33]	1.29 [0.89–8.16]	1.74 [1.28–21.74]	26.92 [19.01-131.75]
Kidney (right)	D1	3.29 [2.4–4.07]	3.34 [2.38–4.1]	4.7 [2.87–10.85]	79.51 [38.76-133.06]
	D2	3.24 [2.53–4.45]	3.25 [2.47–4.46]	5.66 [2.09–7.72]	72.42 [54.03-133.72]
Liver	D1	2.27 [1.77–2.59]	2.36 [1.77–2.64]	3.14 [2.24–6.05]	62.4 [51.8-80.65]
	D2	2.27 [2.12–2.71]	2.32 [2.14–2.56]	3.35 [2.44–4.8]	56 [43.32–74.51]
Lungs	D1	0.51 [0.31–0.72]	0.52 [0.32–0.71]	0.93 [0.57–1.94]	10.01 [6.12–15.18]
	D2	0.53 [0.37–0.8]	0.54 [0.38–0.72]	0.89 [0.47–1.52]	10.18 [6.65–15.87]
Pancreas	D1	1.26 [0.86–1.72]	1.26 [0.89–1.65]	2.65 [1.54–5.66]	22.33 [12.66–31.93]
	D2	1.34 [0.87–1.99]	1.32 [0.86–1.79]	2.76 [1.25–4.3]	22.56 [11.91–32.52]
Spleen	D1	1.8 [1.29–2.13]	1.79 [1.29–2.18]	3.34 [2.1–4.92]	39.29 [26.78–67.63]
	D2	1.91 [0.89–2.16]	1.92 [0.88–2.05]	3.35 [2.6–4.24]	37.34 [13.31–51.95]
Vertebra (L3)	D1	1.55 [1.1–1.73]	1.45 [1.06–1.64]	3.92 [2.02–7.98]	13.24 [7.15–32.16]
	D2	1.62 [1.06–2.05]	1.5 [1.06-2]	4.09 [1.15–6.25]	14.73 [6.99–27.73]
Liver (Geometric)	D1	2.25 [1.77–2.66]	2.33 [1.77–2.72]	3.4 [2.05–6.47]	60.99 [44.96–81.2]
	D2	2.32 [2.06–2.67]	2.31 [2.1–2.65]	3.55 [2.23–5.4]	57.51 [42.67–73.94]
Spleen (Geometric)	D1	1.85 [1.28–2.26]	1.86 [1.29–2.3]	3.15 [2.03–6.4]	42.06 [26.62–58.05]
	D2	1.95 [0.61–2.11]	1.92 [0.62–2.11]	3.26 [1.57–4.23]	38.37 [11.02–54.68]
Lung (Geometric)	D1	0.38 [0.19–0.58]	0.37 [0.19–0.57]	0.66 [0.37–1.44]	7.64 [3.23–11.52]
	D2	0.37 [0.25–0.66]	0.38 [0.25–0.6]	0.68 [0.31–1.12]	7.74 [3.79–13.4]

Values are median [min-max]

D1: First scan day; D2: Second scan day

Table 3 Reliability and variability results for SUV(60–80)mean and MRFDG–mean

Organ/metric		ICC	%RPC	%wCV
Aortic wall	SUV_{mean}	0.69	19	6.7
	$MR_{FDG–mean}$	0.66	38	13.3
Myocardium	SUV_{mean}	0.95	40	14.0
	$MR_{FDG–mean}$	0.94	65	23.1
Kidney (right)	SUV_{mean}	0.68	25	9.0
	$MR_{FDG–mean}$	0.66	67	23.4
Liver	SUV_{mean}	0.65	15	5.4
	$MR_{FDG–mean}$	0.83	31	11.3
Lungs	SUV_{mean}	0.85	21	8.1
	$MR_{FDG–mean}$	0.84	38	13.2
Pancreas	SUV_{mean}	0.88	14	6.6
	$MR_{FDG–mean}$	0.81	49	17.3
Spleen	SUV_{mean}	0.73	27	9.4
	$MR_{FDG–mean}$	0.67	36	12.9
Vertebra (L3)	SUV_{mean}	0.77	19	7.5
	$MR_{FDG–mean}$	0.78	53	18.5
Liver (Geometric)	SUV_{mean}	0.66	15	5.7
	$MR_{FDG–mean}$	0.76	43	16.7
Spleen (Geometric)	SUV_{mean}	0.61	44	15.7
	$MR_{FDG–mean}$	0.47	64	22.3
Lung (Geometric)	SUV_{mean}	0.87	24	9.2
	$MR_{FDG–mean}$	0.88	38	13.4

ICC: Interclass correlation coefficients; RPC: Repeatability coefficients; wCV: Within-subject coefficient of variation

values, whereas there was a notable increase in %RPC for $MR_{FDG-mean}$ values (43% vs. 31%), reflecting a reduced precision in the measured values. Conversely, for the spleen, geometric segmentation appeared to have reduced reliability, especially for $MR_{FDG-mean}$ (ICC = 0.47 vs. 0.67), as well as leading to a significant increase in %RPC for both SUV_{mean} and $MR_{FDG-mean}$. Finally, in the lung, reliability and precision were largely unaffected by the use of geometric delineation.

Correlation plots and Bland-Altman plots are presented in Figs. 1, 2, 3 and 4 and Supplemental Figs. 1–7. The Bland-Altman analyses showed that the test-retest results across all organs were evenly distributed, indicating no significant systematic bias toward either a negative or positive trend.

Adjusting SUV values for plasma glucose and/or lean body mass resulted in only minor changes to reliability while noticeably increasing variability. Compared to SUV, these corrections did not consistently enhance reliability and introduced greater variability, suggesting SUV remains a reliable option in daily clinical practice. Interestingly, SUL demonstrated superior reliability and the lowest variability across most organs, surpassing SUV_{glc} and SUL_{glc}, thus emerging as the most robust metric among the analysed corrections.

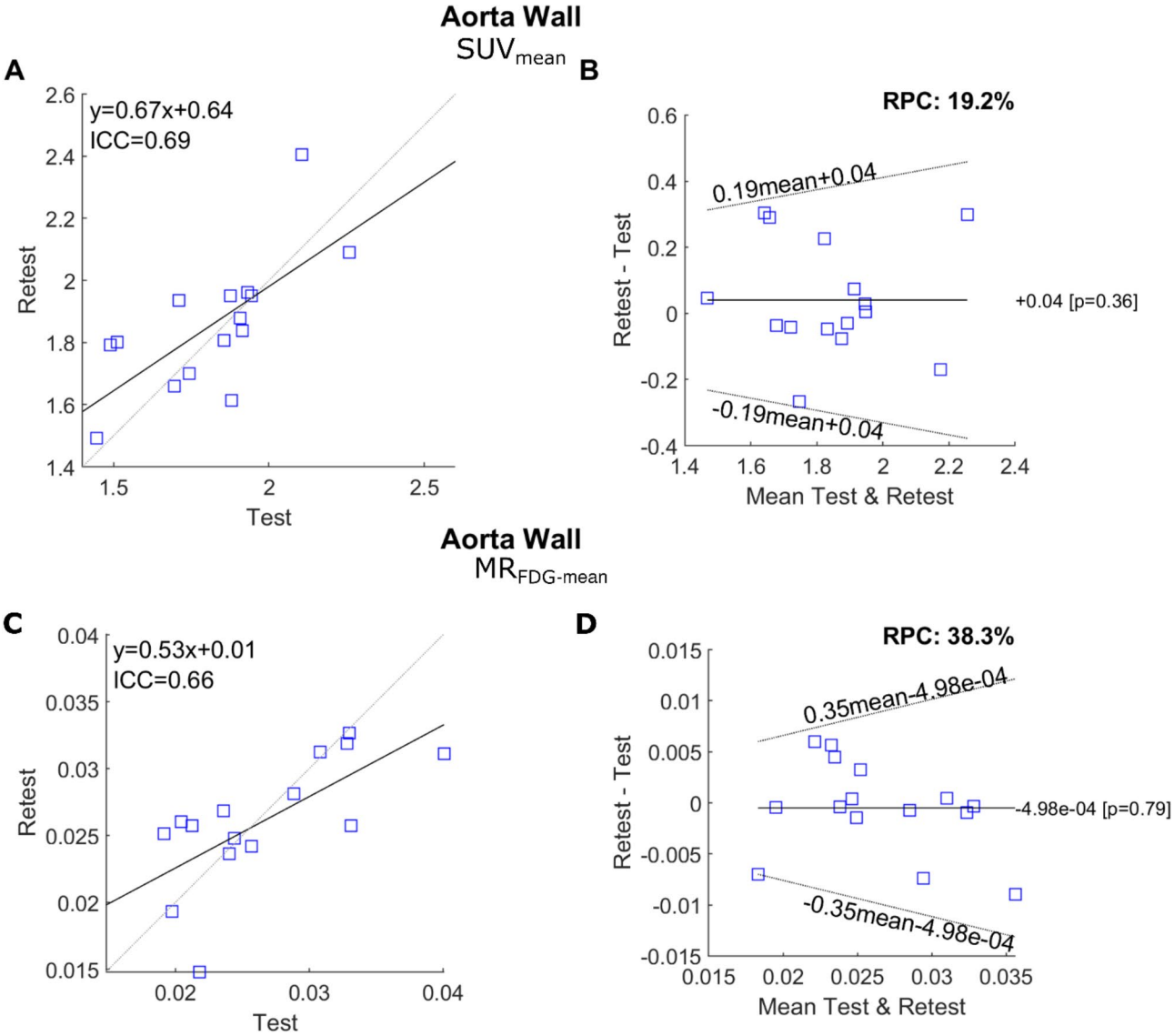

Fig. 1 Correlation plots (**A** and **C**) and Bland–Altman plots (**B** and **D**) demonstrating the repeatability of $SUV(60–80)_{mean}$ (**A** and **B**) and $MR_{FDG-mean}$ (**C** and **D**) in the aortic wall. In the correlation plots, the dashed lines indicate the lines of identity, whereas the solid lines represent the linear regression fit. In the Bland–Altman plots, the solid line marks the mean difference between the test and retest results, while the dashed lines represent the 95% upper and lower limits of agreement. The limits of agreement of log-transformed data are transformed back to original scale

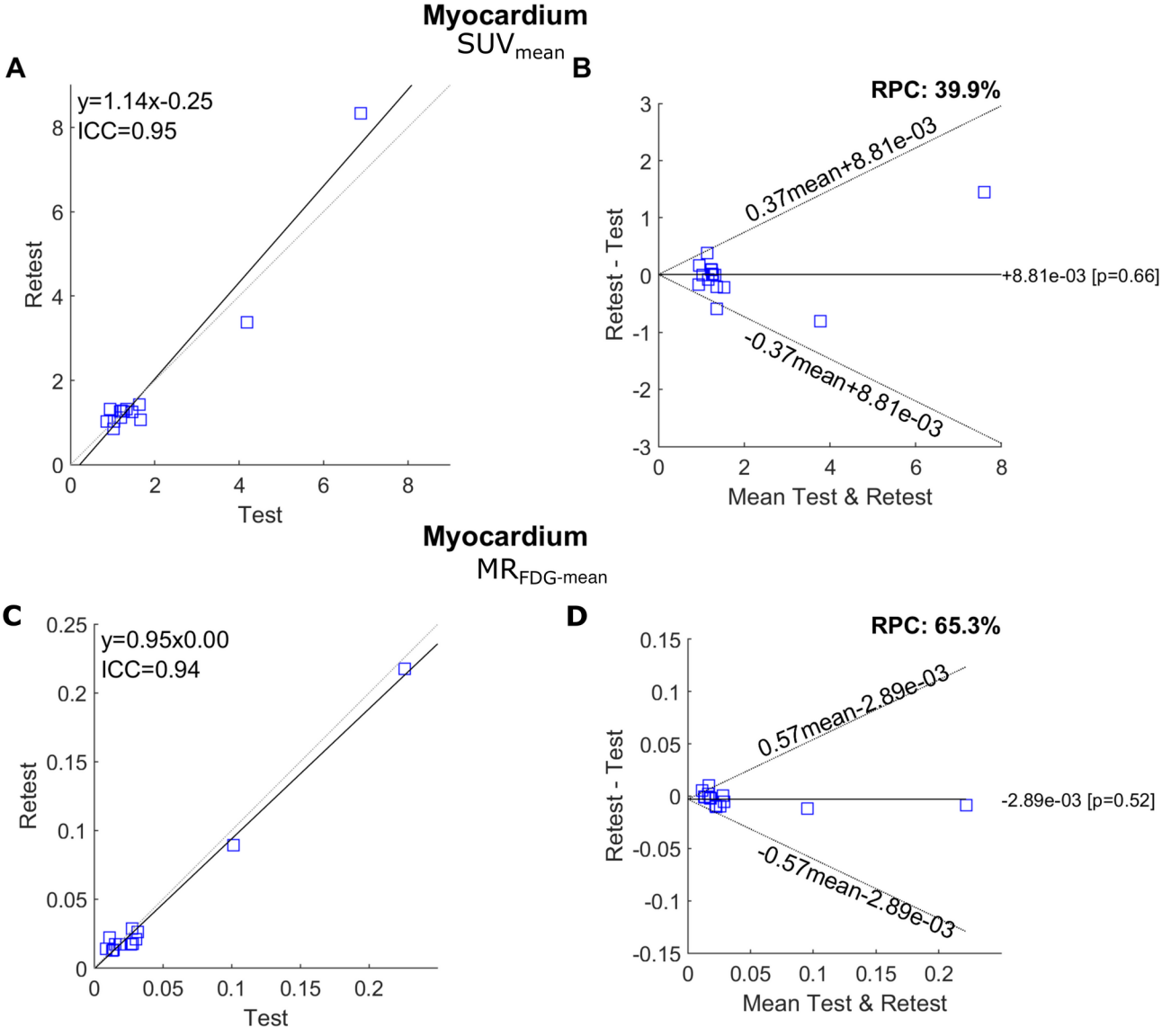

Fig. 2 Correlation plots (**A** and **C**) and Bland–Altman plots (**B** and **D**) demonstrating the repeatability of $SUV(60–80)_{mean}$ (**A** and **B**) and $MR_{FDG\text{-}mean}$ (**C** and **D**) in the left ventricular myocardium. In the correlation plots, the dashed lines indicate the lines of identity, whereas the solid lines represent the linear regression fit. In the Bland–Altman plots, the solid line marks the mean difference between the test and retest results, while the dashed lines represent the 95% upper and lower limits of agreement. The limits of agreement of log-transformed data are transformed back to original scale

Discussion

This study assessed the reliability and precision of SUV_{mean} and $MR_{FDG–mean}$ in a cohort of patients with type 2 diabetes, across various selected automatically delineated organs, as well as evaluating the impact of geometric subsegmentations.

Previous studies reporting on the repeatability in [^{18}F] FDG SUV measurements have mostly focused on the assessment of tumour lesions [24, 25, 26, 27, 28]. A meta-analysis by Langen et al. [29] suggested that a 25% variation in SUV_{mean} variation in studies performed within 30 days of each other might be used as a threshold for significant changes in lesions between studies, while the PERCIST criteria established a 30% variation of SUV_{peak} as significant [30]. Fewer studies have evaluated the reproducibility of quantitative values in organs, with most focusing on specific organs such as the liver, showing variabilities in the 6–11% range over studies spaced up to several months [31, 32]. A study by Malaih et al. showed variations of SUV_{mean}, in studies performed up to 5 days apart, to be as high as 65% in the myocardium, 23% in the skeletal muscle, and up to 31% in the lung [27]. The results of our study, across all evaluated organs, displayed much smaller variations of SUV_{mean} values (< 14% between studies).

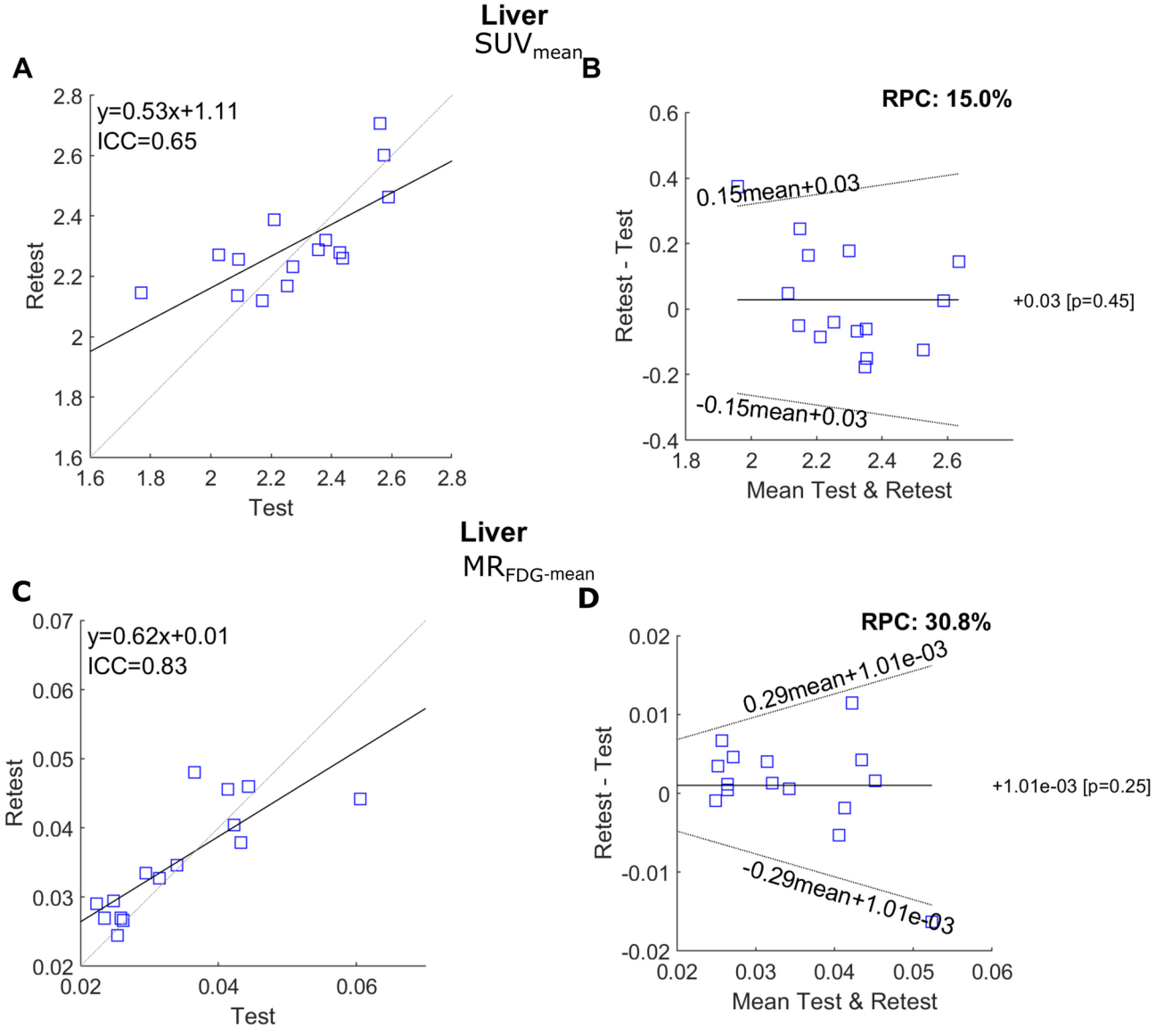

Fig. 3 Correlation plots (**A** and **C**) and Bland–Altman plots (**B** and **D**) demonstrating the repeatability of $SUV(60–80)_{mean}$ (**A** and **B**) and $MR_{FDG\text{-}mean}$ (**C** and **D**) in the liver. In the correlation plots, the dashed lines indicate the lines of identity, whereas the solid lines represent the linear regression fit. In the Bland–Altman plots, the solid line marks the mean difference between the test and retest results, while the dashed lines represent the 95% upper and lower limits of agreement. The limits of agreement of log-transformed data are transformed back to original scale

Furthermore, almost no studies have reported on the reproducibility of [^{18}F]FDG Patlak parametric values in organs. To the best of our knowledge, only a single previous study by Ince et al. [15] has reported on the test-retest repeatability of both MR_{FDG} and SUV measurements from [^{18}F]FDG PET/CT at early (35–50 min) and late (75–90) time points. In their study, reproducibility of these measurements was assessed in selected organs including the liver, lungs and spleen, as well as the skeletal muscle in nine patients with various types of cancer. As expected, the authors also reported a slightly worse performance of MR_{FDG} compared to SUV, with variations up to 15% in the skeletal muscle and spleen for MR_{FDG} from late PET dynamic data.

The results obtained in our study, based on data obtained from 60 to 80 min post tracer administration, revealed a similar pattern, indicating that SUV_{mean} generally outperforms $MR_{FDG–mean}$ in both reliability and precision across the analysed organs, with the myocardium, pancreas, and lungs showing the best repeatability results. However, for both SUV_{mean} and $MR_{FDG–mean}$ measurements, we did not observe statistically significant variations between study days in any of the assessed organs. Regarding the main aim of this study, it is particularly noteworthy that $MR_{FDG–mean}$ measurements

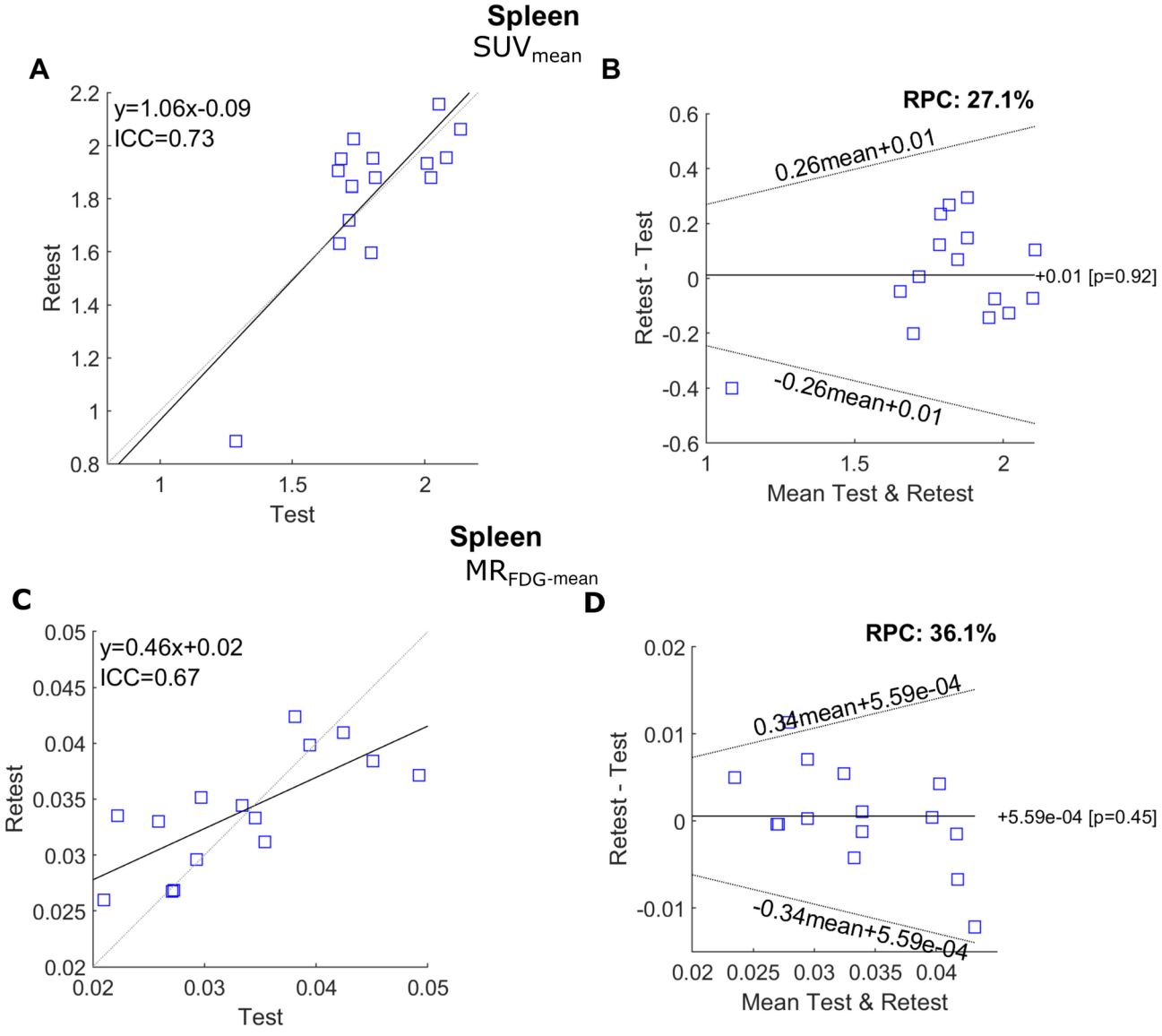

Fig. 4 Correlation plots (**A** and **C**) and Bland–Altman plots (**B** and **D**) demonstrating the repeatability of $SUV(60–80)_{mean}$ (**A** and **B**) and $MR_{FDG\text{-}mean}$ (**C** and **D**) in the spleen. In the correlation plots, the dashed lines indicate the lines of identity, whereas the solid lines represent the linear regression fit. In the Bland–Altman plots, the solid line marks the mean difference between the test and retest results, while the dashed lines represent the 95% upper and lower limits of agreement. The limits of agreement of log-transformed data are transformed back to original scale

demonstrated consistent reliability, with a coefficient of variation below 19%. The only exceptions were the measurements of kidney activity (wCV = 23%), an organ which is obviously subject to marked activity variation due to its excretory function; and the myocardium (wCV = 23%), an organ known to show high variation in physiological glucose activity. Therefore, our study further reinforces the consistency of MR_{FDG} measurements in a test-retest setting, allowing a safe use of such measurements in both clinical and research conditions. We found that SUV measurements were less susceptible to day-to-day variation, likely due to the composite nature of the metric. The SUV image represents the total signal from 60 to 80 min, whereas the quantitative parametric images, MR_{FDG} and DV_{FDG}, decompose this signal into two components using kinetic modelling. Thus, the two Patlak-derived parametric images are inherently noisier than the SUV images. The quantitative parametric images offer greater physiological insight, but this comes at the expense of increased variability. Additionally, parametric images require the selection of a physiologically appropriate model and application of motion compensation to reduce artifacts, which are both subjects of ongoing research in the field.

Some limitations should be noted. Although larger than previous studies, we used a relatively small sample

size, which could limit the robustness of our findings. Also, contrary to most previous test-retest studies, we evaluated repeatability of [^{18}F]FDG D-WB PET/CT in normal organs within a non-oncologic population. While such data should, in theory, better reflect natural physiological functions unaffected by oncological drug interventions, it remains uncertain whether these insights can be effectively applied to oncological patients. Furthermore, the selected patients were not healthy controls but patients with type 2 diabetes selected to participate in an interventional drug trial. Therefore, a selection bias was present. However, since there were no pharmacological interventions between study dates, we consider the obtained data to reflect true physiological variation. The short time interval (~24 h) between the test and retest scans is a notable strength that could reflect the daily variation in organ physiology.

The Patlak model operates under the assumption of irreversible FDG uptake, where FDG is retained in the form of FDG-6-phosphate and should be used with D-WB data only after the blood pool and the reversible compartments (comprising free FDG in blood and tissue) have reached a steady state. It is important to note that these assumptions are not universally applicable to all tissues, which must be considered when evaluating multiparametric images. For instance, tissues like the liver and kidneys are probably more accurately represented by reversible kinetic models. Even though the calculated values might not all be physiologically meaningful, this bias should not interfere with the current test-retest analysis.

An automated delineation software of interest regions was chosen to simplify the data extraction process. However, since organ delineation is performed from CT data, the output VOIs can be misaligned with PET data, especially with multiparametric images, as previously reported [9]. This misregistration arises because PET and CT data are obtained at different time points, with respiratory movement particularly affecting regions placed in organs near the diaphragm, such as the liver, lungs, and spleen. While we attempted to manually correct significant errors, spillover from neighbouring organs was still observed in some organs, in particular with liver activity extending into the automatically delineated lung region. Curiously, the use of carefully placed geometric regions in the liver, spleen, and lung parenchyma did not improve measured reproducibility results, reinforcing the validity of the automatically extracted data.

It should also be noted that the study data was obtained by performing a short 20 min dynamic acquisition, using a recently published population-based input function (PBIF) [12]. The use of such a short protocol enhances the feasibility of performing more precise multiparametric quantitative evaluations in regular clinical practice, where traditional long dynamics studies are not viable. As noted in the previous publication, a slight correction to the standard PBIF may be advisable for diabetic patients. In our study, however, no such correction was applied. This should not impact the results of the test-retest evaluation, as any potential bias would be consistent across both scans.

Conclusions

In conclusion, this study demonstrates that both SUV and MR_{FDG} values derived from [^{18}F]FDG D-WB PET/CT are highly robust, showing excellent repeatability in studies conducted on consecutive days. This reliability enhances clinicians' confidence in interpreting these measurements for both clinical practice and research applications.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13550-025-01249-z.

Supplementary Material 1

Acknowledgements
Not applicable.

Author contributions
EL, PLP and LCG conceived the study; JMB and ADJ recruited patients for the study; OLM reconstructed the PET images; JMB and AHD analysed images and delineated VOIs; JMB, LPT and AHD performed statistics and quantitative data analysis; KLF, ADJ, OLM, LPT, EL, PLP, and LCG contributed to interpretation of the data; JMB wrote the first draft of the manuscript; AHD wrote the final version of the manuscript; JMB, KLF, ADJ, OLM, LPT, EL, PLP, LCG and AHD critically revised the manuscript and approved of the final version.

Funding
This work was supported by a research grant from the Danish Diabetes Academy to JMB (Novo Nordisk Foundation, NNF17SA0031406); as well as an unrestricted grant from The Novo Nordisk Foundation to LCG (NNF19OC0055100). The study funders were not involved in the study's design, data collection, analysis, or interpretation, nor in the writing of the report, and they did not impose any restrictions on the publication of the report. The participating researchers are paid by their institutions and have no affiliation with the study's financial sponsors.

Data availability
Within the restrictions applied by the EU GDPR, all data are available from the authors upon reasonable request.

Declarations

Ethics approval and consent to participate
Approval for the study was obtained from the ethics committee at Region Midtjylland (1-10-72-345-21). All methods were carried out in accordance with relevant guidelines and regulations. Informed consent was obtained from each patient before inclusion into the study.

Consent for publication
Publication consent was obtained from all study participants.

Competing interests
The authors declare that they have no competing interests.

Received: 31 January 2025 / Accepted: 25 April 2025
Published online: 12 May 2025

References

1. Keyes JW. Jr. SUV: standard uptake or silly useless value? J Nucl Med. 1995;36:1836–9.
2. Adams MC, Turkington TG, Wilson JM, Wong TZ. A systematic review of the factors affecting accuracy of SUV measurements. AJR Am J Roentgenol. 2010;195. https://doi.org/10.2214/AJR.10.4923.
3. Boellaard R. Standards for PET image acquisition and quantitative data analysis. J Nucl Med. 2009;50(Suppl 1):s11–20. https://doi.org/10.2967/jnumed.108.057182.
4. Boellaard R, Delgado-Bolton R, Oyen WJ, Giammarile F, Tatsch K, Eschner W, et al. FDG PET/CT: EANM procedure guidelines for tumour imaging: version 2.0. Eur J Nucl Med Mol Imaging. 2015;42:328–54. https://doi.org/10.1007/s00259-014-2961-x.
5. Karakatsanis NA, Lodge MA, Tahari AK, Zhou Y, Wahl RL, Rahmim A. Dynamic whole-body PET parametric imaging: I. Concept, acquisition protocol optimization and clinical application. Phys Med Biol. 2013;58:7391–418. https://doi.org/10.1088/0031-9155/58/20/7391.
6. Patlak CS, Blasberg RG. Graphical evaluation of blood-to-brain transfer constants from multiple-time uptake data. Generalizations J Cereb Blood Flow Metabolism. 1985;5:584–90. https://doi.org/10.1038/jcbfm.1985.87.
7. Patlak CS, Blasberg RG, Fenstermacher JD. Graphical evaluation of blood-to-brain transfer constants from multiple-time uptake data. J Cereb Blood Flow Metabolism. 1983;3:1–7. https://doi.org/10.1038/jcbfm.1983.1.
8. Rahmim A, Lodge MA, Karakatsanis NA, Panin VY, Zhou Y, McMillan A, et al. Dynamic whole-body PET imaging: principles, potentials and applications. Eur J Nucl Med Mol Imaging. 2019;46:501–18. https://doi.org/10.1007/s00259-018-4153-6.
9. Dias AH, Pedersen MF, Danielsen H, Munk OL, Gormsen LC. Clinical feasibility and impact of fully automated multiparametric PET imaging using direct Patlak reconstruction: evaluation of 103 dynamic whole-body 18 F-FDG PET/CT scans. Eur J Nucl Med Mol Imaging. 2021;48. https://doi.org/10.1007/s00259-020-05007-2.
10. Fahrni G, Karakatsanis NA, Di Domenicantonio G, Garibotto V, Zaidi H. Does whole-body Patlak (18)F-FDG PET imaging improve lesion detectability in clinical oncology? Eur Radiol. 2019;29:4812–21. https://doi.org/10.1007/s00330-018-5966-1.
11. Naganawa M, Gallezot JD, Shah V, Mulnix T, Young C, Dias M, et al. Assessment of population-based input functions for Patlak imaging of whole body dynamic 18 F-FDG PET. EJNMMI Phys. 2020;7. https://doi.org/10.1186/s40658-020-00330-x.
12. Dias AH, Pigg D, Smith AM, Shah V, Gormsen LC, Munk OL. Clinical validation of a population-based input function for dynamic whole-body 18F-FDG multiparametric PET imaging using a standard injector. 34th annual Congress of the European association of nuclear medicine. Eur J Nucl Med Mol Imaging; 2021. p. (Suppl 1) 198–9.
13. Lan W, Sari H, Rominger A, Fougère C, Schmidt FP. Optimization and impact of sensitivity mode on abbreviated scan protocols with population-based input function for parametric imaging of [(18)F]-FDG for a long axial FOV PET scanner. Eur J Nucl Med Mol Imaging. 2024. https://doi.org/10.1007/s00259-024-06745-3.
14. Dias AH, Hansen AK, Munk OL, Gormsen LC. Normal values for 18 F-FDG uptake in organs and tissues measured by dynamic whole body multiparametric FDG PET in 126 patients. EJNMMI Res. 2022;12. https://doi.org/10.1186/s13550-022-00884-0.
15. Ince S, Laforest R, Ashrafinia S, Smith AM, Wahl RL, Fraum TJ. Test-Retest repeatability of Patlak slopes versus standardized uptake values for hypermetabolic lesions and normal organs in an oncologic PET/CT population. Mol Imaging Biol. 2024;26:284–93. https://doi.org/10.1007/s11307-024-01909-x.
16. Tao Y, Peng Z, Krishnan A, Zhou XS. Robust learning-based parsing and annotation of medical radiographs. IEEE Trans Med Imaging. 2011;30. https://doi.org/10.1109/TMI.2010.2077740.
17. Wasserthal J, Breit HC, Meyer MT, Pradella M, Hinck D, Sauter AW, et al. TotalSegmentator: robust segmentation of 104 anatomic structures in CT images. Radiol Artif Intell. 2023;5:e230024. https://doi.org/10.1148/ryai.230024.
18. Jochumsen MR, Bouchelouche K, Nielsen KB, Frøkiær J, Borre M, Sörensen J, et al. Repeatability of tumor blood flow quantification with (82)Rubidium PET/CT in prostate cancer - a test-retest study. EJNMMI Res. 2019;9:58. https://doi.org/10.1186/s13550-019-0529-2.
19. Lodge MA, Jacene HA, Pili R, Wahl RL. Reproducibility of tumor blood flow quantification with 15O-water PET. J Nucl Med. 2008;49:1620–7. https://doi.org/10.2967/jnumed.108.052076.
20. Euser AM, Dekker FW, le Cessie S. A practical approach to Bland-Altman plots and variation coefficients for log transformed variables. J Clin Epidemiol. 2008;61:978–82. https://doi.org/10.1016/j.jclinepi.2007.11.003.
21. Laird NM, Ware JH. Random-effects models for longitudinal data. Biometrics. 1982;38:963–74.
22. Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, et al. The REDCap consortium: Building an international community of software platform partners. J Biomed Inform. 2019;95. https://doi.org/10.1016/j.jbi.2019.103208.
23. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform. 2009;42. https://doi.org/10.1016/j.jbi.2008.08.010.
24. Shankar LK, Huang E, Litiere S, Hoekstra OS, Schwartz L, Collette S, et al. Meta-Analysis of the Test-Retest repeatability of [18F]-Fluorodeoxyglucose standardized uptake values: implications for assessment of tumor response. Clin Cancer Res. 2023;29:143–53. https://doi.org/10.1158/1078-0432.Ccr-21-3143.
25. Crandall JP, Fraum TJ, Lee M, Jiang L, Grigsby P, Wahl RL. Repeatability of (18) F-FDG PET radiomic features in cervical Cancer. J Nucl Med. 2021;62:707–15. https://doi.org/10.2967/jnumed.120.247999.
26. Fencl P, Belohlavek O, Harustiak T, Zemanova M. The analysis of factors affecting the threshold on repeated 18F-FDG-PET/CT investigations measured by the PERCIST protocol in patients with esophageal carcinoma. Nucl Med Commun. 2012;33:1188–94. https://doi.org/10.1097/MNM.0b013e3283573d0d.
27. Malaih AA, Dunn JT, Nygård L, Kovacs DG, Andersen FL, Barrington SF, et al. Test-retest repeatability and interobserver variation of healthy tissue metabolism using 18F-FDG PET/CT of the thorax among lung cancer patients. Nucl Med Commun. 2022;43:549–59. https://doi.org/10.1097/mnm.0000000000001537.
28. Muzi M, Peterson LM, Specht JM, Hippe DS, Novakova-Jiresova A, Lee JH, et al. Repeatability of 18F-FDG uptake in metastatic bone lesions of breast cancer patients and implications for accrual to clinical trials. Res Sq. 2024. https://doi.org/10.21203/rs.3.rs-3818932/v1.
29. de Langen AJ, Vincent A, Velasquez LM, van Tinteren H, Boellaard R, Shankar LK, et al. Repeatability of 18F-FDG uptake measurements in tumors: a meta-analysis. J Nucl Med. 2012;53:701–8. https://doi.org/10.2967/jnumed.111.095299.
30. 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(Suppl 1):s122–50. https://doi.org/10.2967/jnumed.108.057307.
31. Kanstrup IL, Klausen TL, Bojsen-Møller J, Magnusson P, Zerahn B. Variability and reproducibility of hepatic FDG uptake measured as SUV as well as tissue-to-blood background ratio using positron emission tomography in healthy humans. Clin Physiol Funct Imaging. 2009;29:108–13. https://doi.org/10.1111/j.1475-097X.2008.00846.x.
32. Paquet N, Albert A, Foidart J, Hustinx R. Within-patient variability of (18)F-FDG: standardized uptake values in normal tissues. J Nucl Med. 2004;45:784–8.

Publisher's note