



Repeatability of quantitative ^{18}F -FLT uptake measurements in solid tumors: an individual patient data multi-center meta-analysis

G. M. Kramer¹ · Y. Liu² · A. J. de Langen³ · E. P. Jansma⁴ · I. Trigonis⁵ · M.-C. Asselin⁵ · A. Jackson⁵ · L. Kenny⁶ · E. O. Aboagye⁶ · O. S. Hoekstra¹ · R. Boellaard¹ · on behalf of the QuIC-ConCePT consortium

Received: 15 August 2017 / Accepted: 26 December 2017
© The Author(s) 2018. This article is an open access publication

Abstract

Introduction 3'-deoxy-3'-[^{18}F]fluorothymidine (^{18}F -FLT) positron emission tomography (PET) provides a non-invasive method to assess cellular proliferation and response to antitumor therapy. Quantitative ^{18}F -FLT uptake metrics are being used for evaluation of proliferative response in investigational setting, however multi-center repeatability needs to be established. The aim of this study was to determine the repeatability of ^{18}F -FLT tumor uptake metrics by re-analyzing individual patient data from previously published reports using the same tumor segmentation method and repeatability metrics across cohorts.

Methods A systematic search in PubMed, EMBASE.com and the Cochrane Library from inception-October 2016 yielded five ^{18}F -FLT repeatability cohorts in solid tumors. ^{18}F -FLT avid lesions were delineated using a 50% isocontour adapted for local background on test and retest scans. SUV_{max} , SUV_{mean} , SUV_{peak} , proliferative volume and total lesion uptake (TLU) were calculated. Repeatability was assessed using the repeatability coefficient ($\text{RC} = 1.96 \times \text{SD}$ of test-retest differences), linear regression analysis, and the intra-class correlation coefficient (ICC). The impact of different lesion selection criteria was also evaluated.

Results Images from four cohorts containing 30 patients with 52 lesions were obtained and analyzed (ten in breast cancer, nine in head and neck squamous cell carcinoma, and 33 in non-small cell lung cancer patients). A good correlation was found between test-retest data for all ^{18}F -FLT uptake metrics ($R^2 \geq 0.93$; $\text{ICC} \geq 0.96$). Best repeatability was found for SUV_{peak} ($\text{RC}: 23.1\%$), without significant differences in RC between different SUV metrics. Repeatability of proliferative volume ($\text{RC}: 36.0\%$) and TLU ($\text{RC}: 36.4\%$) was worse than SUV. Lesion selection methods based on $\text{SUV}_{\text{max}} \geq 4.0$ improved the repeatability of volumetric metrics ($\text{RC}: 26\text{--}28\%$), but did not affect the repeatability of SUV metrics.

Conclusions In multi-center studies, differences $\geq 25\%$ in ^{18}F -FLT SUV metrics likely represent a true change in tumor uptake. Larger differences are required for FLT metrics comprising volume estimates when no lesion selection criteria are applied.

Keywords Flt · Repeatability · PET · Oncology

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00259-017-3923-x>) contains supplementary material, which is available to authorized users.

✉ G. M. Kramer
ge.kramer@vumc.nl

¹ Department of Radiology and Nuclear Medicine, VU University Medical Center, PO Box 7057, 1007 MB Amsterdam, Netherlands

² European Organisation for Research and Treatment for Cancer (EORTC), Headquarters, Brussels, Belgium

³ Department of Pulmonology, VU University Medical Center, PO Box 7057, 1007 MB Amsterdam, Netherlands

⁴ Medical Library, VU University Medical Center, PO Box 7057, 1007 MB Amsterdam, The Netherlands

⁵ Division of Informatics, Imaging and Data Sciences Institute of Population Health, Wolfson Molecular Imaging Centre, Manchester Academic Health Sciences Centre, The University of Manchester, Manchester, UK

⁶ Department of Surgery and Cancer, Imperial College London, Hammersmith Campus, London, UK

Introduction

Despite the recent progress made in cancer diagnosis and treatment, cancer remains the number one cause of death in the Western world [1]. Although treatment can be very effective, most regimens fail for a substantial number of patients. Early response evaluation enables the treating physician to differentiate responders from non-responders, to stop the treatment in the non-responder cohort timely and reliably. This potentially helps to limit side effects of anticancer therapies and avoid treatment delay of subsequent lines, thereby reducing patient burden and healthcare costs.

Several imaging modalities can be used to non-invasively assess response to treatment. Most modalities only evaluate morphological features, yet slow changes in tumor morphology or even pseudoprogression, as can be seen in case of immunotherapy, impair the use of morphological features in early response assessment [2, 3]. However, morphological changes are often preceded by changes in tumor metabolism [4]. These early functional changes can be assessed using molecular imaging techniques such as PET, which may allow for more accurate early response evaluation.

There are several different radiotracers available to assess a variety of metabolic processes. One of these tracers is 3'-deoxy-3'-[^{18}F]fluorothymidine (^{18}F -FLT) and provides a method to evaluate cellular proliferation. Proliferation is a central hallmark of tumor growth and previous studies have validated ^{18}F -FLT against the immunohistochemistry proliferation marker Ki67 in pathological specimens for several tumor types [5–7]. Unfortunately, ^{18}F -FLT PET did not improve tumor detection or staging compared to 2-deoxy-2-[^{18}F]fluoro-D-glucose (^{18}F -FDG) due to lower sensitivity [8]. As proliferation is more cancer-specific compared to glycolysis, ^{18}F -FLT PET has potential as an imaging biomarker for response assessment.

Cytotoxic and cytostatic therapies aim, respectively, to kill tumor cells (mainly highly proliferating cells) and diminish tumor growth, both leading to a decrease in cellular proliferation. After initiation of any antitumor treatment, this change in proliferation can be evaluated using ^{18}F -FLT PET/CT. Several studies have been performed investigating ^{18}F -FLT PET/CT as quantitative imaging biomarker of response [9], nevertheless most did not take variability into account.

For ^{18}F -FDG, the repeatability of quantitative uptake measures has been widely investigated [10–13] and integrated into the response assessment criteria PERCIST [2]. Up to now, repeatability of quantitative ^{18}F -FLT PET/CT has only been studied in a few small single-center cohorts (≤ 10 patients) [14–17]. Moreover, there was variability in uptake intervals, tumor delineation methods, and image analyses. The aim of this study was therefore to perform an individual patient data meta-analyses by re-analyzing all available ^{18}F -FLT

repeatability data from previously published studies and to determine the repeatability of several quantitative ^{18}F -FLT tumor uptake metrics using similar uptake intervals, the same tumor segmentation method, and the same repeatability metrics as would be done in a prospective multi-center study.

Methods

Search strategy and selection process

To identify all relevant publications, a systemic search was performed in PubMed, EMBASE.com and the Cochrane Library (via Wiley) from inception to October 20, 2016 (last elicitation). A combination of the search terms comprising 'FLT-PET' and 'neoplasms' was used. This included MeSH terms and controlled terms from Emtree for PubMed and EMBASE.com, respectively, as well as free-text terms. We only used free-text terms in the Cochrane Library (see supplemental data). All potentially relevant titles and abstracts were screened for eligibility. Full-text articles were checked for eligibility criteria where necessary. References of eligible publications were checked for relevant publications. We have also checked ClinicalTrials.gov and The European Union Clinical Trials Register for ongoing and unpublished studies.

Studies were included if they met the following criteria:

- The study investigated the repeatability of ^{18}F -FLT PET or PET/CT in oncological patients;
- Scans were performed on two separate days using the same scanner; and
- Patients were not treated in between both scans.

Studies were excluded if they met the following criteria:

- Animal or in vitro studies;
- Focused on tumors of the central nervous system (to avoid differences in pharmacokinetics due to the blood-brain barrier);
- Not available in full text or not written in English; and
- Reviews, editorials, letters, legal cases, interviews, case reports, and comments.

Data analysis

Sites from all identified cohorts were contacted, and permission was requested to re-analyze the original ^{18}F -FLT PET repeatability scans. All datasets consisted of 60- or 95-min dynamic test and retest ^{18}F -FLT PET scans. Where permission was granted, original ^{18}F -FLT scans of all individual patients were supplied in DICOM or Analyze format. Prior to re-analysis, all scans were

checked for technical issues and artifacts. If any technical issues or artifacts were present, data were cross-checked with the original research teams. After checking of the scan data, static standard uptake value (SUV) images were generated from the dynamic images: 40–65 or 45–60 min post-injection, depending on the original frame definition. A 5-mm Gaussian filter was applied to the non-smoothed reconstructed images to match the spatial resolution between existing datasets and with previously published data. New volumes of interest (VOI) were defined by segmenting tumors using a 50% isocontour of the SUV_{peak} (1.2 cm in diameter sphere positioned to maximize its mean value), adapted for local background (in-house developed software) [12, 18]. For each VOI, SUV_{max} , SUV_{mean} , SUV_{peak} , proliferative volume (50% threshold of SUV_{peak} corrected for local background) and total lesion uptake (TLU, product of SUV_{mean} and proliferative volume) was determined. These quantitative ^{18}F -FLT uptake metrics were checked for outliers and discrepancies with the original data, however no important issues were identified. In addition, tumor-to-blood ratios (TBR) were calculated by normalizing tumor SUVs to the bloodpool SUV_{mean} of a large vascular structure (2×2 voxel VOI in five consecutive planes) [19]. ^{18}F -FLT uptake in the tumor was normalized to the SUV_{mean} of the carotid artery in HNC data and to the ascending aorta for all other lesions. All SUVs were calculated by normalizing the radioactivity concentrations by the injected ^{18}F -FLT dose and body weight and were corrected for physical decay.

Statistical analysis

Repeatability of the quantitative uptake and volume metrics was determined by calculating the mean and standard deviation (SD) of the percentage differences between the two baseline scans:

$$\%Difference = \frac{Scan\ 2 - Scan\ 1}{(Scan\ 1 + Scan\ 2)/2} \times 100 \quad (1)$$

Normality of the data was assessed using histogram analyses and quantile-quantile plots (data not shown). The repeatability coefficients (RC) were calculated as $1.96 \times SD$ of the percentage differences. A paired t test was performed to test for significant differences in mean uptake between both baseline scans. To assess the significance of differences in RC, the Levene's test was used. Moreover, the intra-class correlation coefficient (ICC) using a two-way mixed model, model II regression analysis [20] and Bland–Altman plots were used to evaluate correlations and biases between the test-and-retest scans. The effect of various lesion selection strategies on

repeatability was evaluated: lesions ≥ 4.2 ml (diameter ≥ 20 mm) [18], $SUV_{max} \geq 4.0$ [10, 11], hottest lesion per scan (highest SUV_{max}) or primary lesions only. In addition, the uptake values of individual lesions were averaged per patient to obtain the averaged uptake and assess repeatability on a patient level. All statistical analyses were performed using SPSS 22.0 (SPSS, Chicago, IL, USA).

Results

Search results

The literature search generated 1728 results: 630 in PubMed, 1076 in [EMBASE.com](#) and 22 in the Cochrane Library. No ongoing or unpublished trials were identified. After removing duplicates, 1172 references remained (Fig. 1). Out of 1172, four articles (five patient cohorts) were considered eligible [14–17]. We obtained permission to re-analyze the original ^{18}F -FLT repeatability data from four of these cohorts, comprising data of 30 patients and 52 individual lesions (ten in breast cancer [14], nine in head and neck squamous cell carcinoma [15], and 33 in non-small cell lung cancer patients from two cohorts [15, 16]; Fig. 2). All patients were included in this individual patient data meta-analysis and no scans had to be excluded. An overview of the cohorts can be found in Table 1.

Repeatability

SUV metrics were lower in the lung cancer dataset from Trigonis et al. [16] compared to the other three datasets (average SUV_{mean} : 2.4 vs. 3.5, respectively; $p < 0.05$). In addition, the SUV_{max} and SUV_{peak} values in the breast cancer dataset from Kenny et al. [14] were higher compared to those from de Langen et al. [15]. Proliferative volumes and TLU were significantly smaller in the HNC group and the NSCLC lesions in the dataset from Trigonis et al. [16] were also significantly smaller than in the de Langen et al. dataset [15]. Despite overall proliferative volumes of the retest scan being significantly larger than the test scans (MATV: 14.5 vs. 15.6 ml, $p = 0.02$), no differences were found between the SUV metrics from test-and-retest scans (Table 2). When assessed per site, a small but significant difference in proliferative volume and TLU was only found in the dataset from Trigonis et al. (mean difference -2.3 ml and -4.2 ml respectively, $p < 0.01$) [16].

Correlations between test-and-retest scans were strong for all uptake metrics per lesion as well as averaged per patient ($R^2 \geq 0.93$ and $ICC \geq 0.96$, Fig. 3). Moreover, no systematic bias was present between both scans as revealed by the correlation plots (slope, 0.98–1.04, Fig. 3) and the Bland–Altman

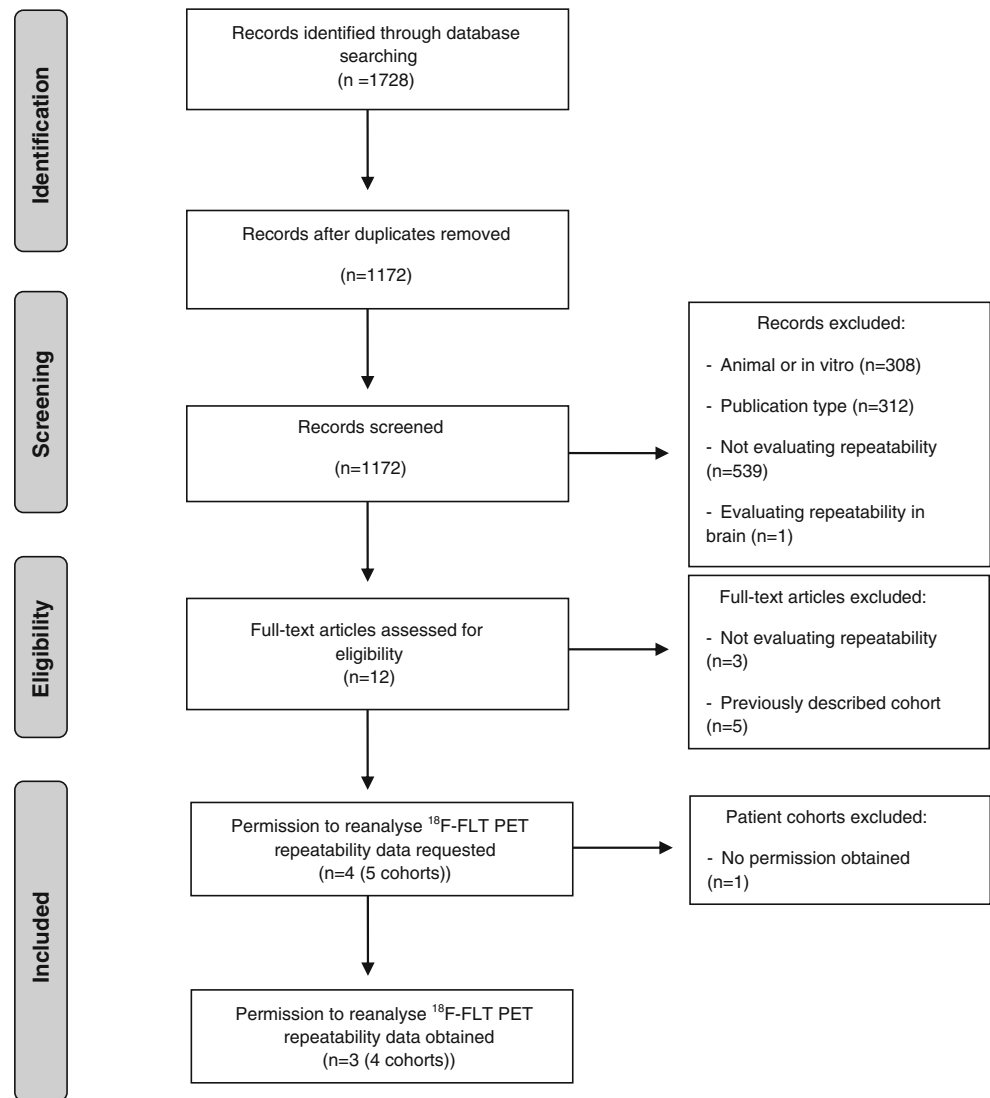
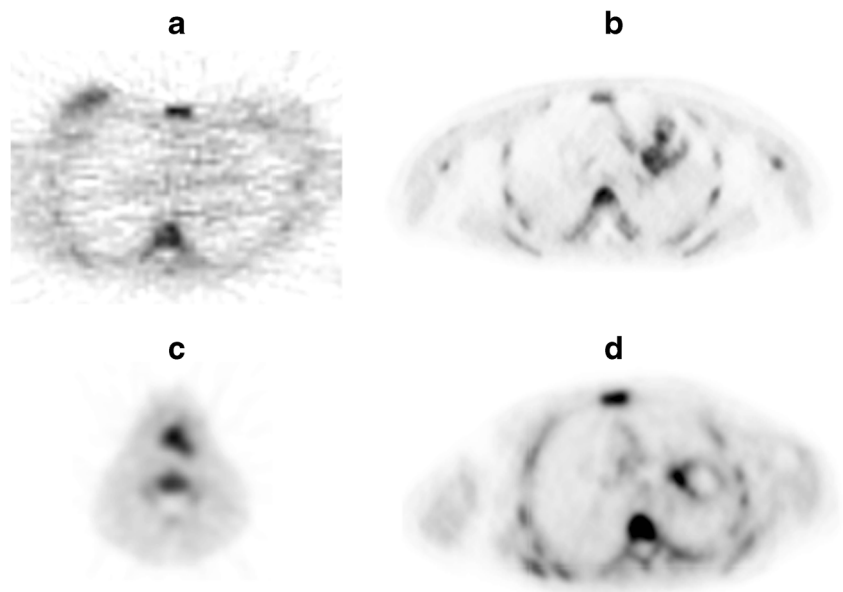
Fig. 1 Flowchart of the search-and-selection procedure of studies**Fig. 2** ^{18}F -FLT PET scan of all four cohorts. **a** Kenny et al. (breast); **b** Trigonis et al. (NSCLC); **c**, **d** de Langen et al., HNC and NSCLC, respectively

Table 1 Cohort and patient characteristics; median (range)

Characteristic	Kenny	de Langen	HNC	Trigonis
Cancer type	BC	NSCLC	HNC	NSCLC
Patients	8	9	6	7
Tumors	10	15	9	18
Scanner - Manufacturer - Model	Siemens ECAT/962 HR+	Siemens ECAT EXACT HR+	Siemens ECAT EXACT HR+	Siemens Biograph 6 Truepoint TrueV
Reconstruction	FBP	OSEM	OSEM	OSEM
Iteration	–	2	2	4
Subsets	–	16	16	21
Voxel size (mm ³)	2.62 × 2.62 × 2.42	2.57 × 2.57 × 2.43	3.43 × 3.34 × 2.43	2.67 × 2.67 × 2.00
Static reconstruction - Scan interval (min) - Frames:	45–65 Averaged	45–60 Averaged	45–60 Averaged	45–60 Summed
Time between scans (days)	4.5 (2–9)	2 (1–6)	1 (1)	4 (2–6)
Weight (kg) - Test - Retest	61.6 (53–106) 61.9 (51.3–107)	71 (61–83) 71 (61–86.5)	77 (65–85) 77 (65–85)	81.5 (66–96.8) 81.4 (65.6–97)
Injected dose (MBq) - Test - Retest	369 (246–380) 312 (153–379)	385 (253–389) 365 (341–397)	375 (334–390) 376 (354–405)	289 (254–332) 328 (283–361)

*No significant differences were present between the test-and-retest scans for any of the studies (Wilcoxon signed-rank test)

BC breast cancer, NSCLC non-small cell lung cancer, HNC head and neck cancer

plots (Fig. 3). Overall, the best repeatability of quantitative ¹⁸F–FLT PET/CT was obtained using SUV_{peak} (RCs 23.1%, Table 3). No differences in RCs were found between the individual SUV metrics.

Variability of proliferative volume and TLU (RCs 36.0 and 36.4%, respectively) were significantly worse than for SUV metrics, with an average increase in RC of $9.6 \pm 6.6\%$ ($p \leq 0.02$) (Fig. 4). When the datasets were evaluated individually, variability of SUV_{peak} and SUV_{mean} within the de Langen et al. [15] cohorts was significantly smaller compared to those of the breast cancer dataset, the only one reconstructed with FBP ($p < 0.02$) [14]. In general, the largest variability was seen in the latter dataset. When comparing only the OSEM reconstructed datasets, RCs for SUV_{max}, SUV_{peak}, and SUV_{mean} changed to 25, 20, and 17% respectively, but RCs of proliferative volumes and TLU remained close to 35%. An overview of the absolute repeatability coefficients for each quantitative uptake metric can be found in supplemental Tables 4 and 5.

Assessment of repeatability on a patient level improved repeatability in general (Table 4). Improvement of repeatability weighted for lesions number was $< 2\%$ compared to unweighted averaging per patient. For the SUV metrics, a decrease in RC was largest in the de Langen dataset [15]. Only SUV_{mean} showed a slight increase in variability, which was caused by one lesion with a 53% difference (4 SDs) between both scans from the breast cancer dataset. If excluded, repeatability of SUV_{mean} improved to 19%, while other SUV metrics remained unaffected. RCs of proliferative volume and TLU also decreased to $< 30\%$, with the exception of the breast dataset [14].

Lesion selection

Assessing repeatability including only lesions with SUV_{max} ≥ 4.0 decreased variability of volumetric metrics but did not influence RCs of SUV metrics (RCs 26–28%, Fig. 5). The former is mainly caused by a large decrease of RCs in the Trigonis dataset (-20%). If only lesions larger than 4.2 ml were included in the analysis, no significant change in variability of SUV, proliferative volume or TLU was seen (RCs 22–25% and 34–36%, respectively). Similar results were observed when only the hottest or primary lesions were assessed. Combining the two selection criteria SUV_{max} ≥ 4.0 and lesions ≥ 4.2 ml did not further improve results. No significant change in repeatability of SUV metrics was seen when analyzing cohorts individually. In addition, applying lesions selection criteria to the per-patient analysis did not decrease variability of SUV and volumetric ¹⁸F–FLT uptake measures.

Normalization to blood uptake

Overall, repeatability deteriorated significantly when TBR was used (RCs +49–52%; $p < 0.02$). The effect on the HNC dataset using the carotid artery was not different compared to the lung cancer datasets using the larger ascending aorta. In particular, repeatability of the breast dataset worsened by calculating the TBR, showing an increase of $> 50\%$ for all metrics. This is likely explained by the variability of the bloodpool SUV being significantly larger in the FBP reconstructed dataset compared to the OSEM reconstructed datasets (SD: 34 vs. 13%). When this cohort was excluded, RCs of

Table 2 Mean ^{18}F -FLT uptake values of different uptake metrics overall and per cohort

Quantitative tracer uptake measures	Overall		Kenny		de Langen		Trigonis				
	Test (mean \pm SD)	Retest (mean \pm SD)	Test (mean \pm SD)	Retest (mean \pm SD)	Test (mean \pm SD)	Retest (mean \pm SD)	Test (mean \pm SD)	Retest (mean \pm SD)			
			BC		NSCLC		HNC				
			Overall		NSCLC		HNC				
SUV _{max}	5.0 \pm 2.4	4.9 \pm 2.4	7.2 \pm 3.4	6.7 \pm 3.5	4.9 \pm 1.7	4.8 \pm 1.7	5.2 \pm 1.5	4.9 \pm 1.6	4.5 \pm 2.0	3.8 \pm 1.5	4.0 \pm 1.8
SUV _{peak}	4.0 \pm 2.1	3.9 \pm 2.0	5.7 \pm 3.0	5.3 \pm 3.1	4.0 \pm 1.6	4.0 \pm 1.6	4.3 \pm 1.5	4.1 \pm 1.5	3.7 \pm 1.8	2.9 \pm 1.2	2.9 \pm 1.3
SUV _{mean}	3.1 \pm 1.5	3.0 \pm 1.5	4.1 \pm 2.3	3.8 \pm 2.4	3.3 \pm 1.2	3.2 \pm 1.1	3.4 \pm 1.0	3.3 \pm 1.0	3.1 \pm 1.4	2.3 \pm 0.9	2.4 \pm 1.0
TLLU	49 \pm 66	51 \pm 66	86 \pm 40	88 \pm 55	56 \pm 83	56 \pm 81	77 \pm 100	76 \pm 97	23 \pm 18	25 \pm 32	30 \pm 35
Volume	14 \pm 16	16 \pm 17	25 \pm 25	26 \pm 27	15 \pm 17	15 \pm 16	19 \pm 20	20 \pm 19	7.0 \pm 4.7	10 \pm 8.9	12 \pm 12

BC breast cancer, NSCLC non-small cell lung cancer, HNC head and neck cancer, SUV standardized uptake value, TLLU total lesion uptake

TBR metrics were no longer significantly different from the SUV metrics.

Discussion

This individual patient data meta-analysis combined available data from four different ^{18}F -FLT PET test-retest cohorts acquired in three different cancer types at three different centers. Of the quantitative ^{18}F -FLT uptake measures commonly used in oncological setting, SUV metrics showed better repeatability overall than the volumetric metrics. Unfortunately, we did not obtain permission from one study to re-analyze their data [17]. However, individual SUV_{max}, SUV_{peak}, and SUV_{mean} values were reported in this article. If these numbers are included in the analysis, RCs of the SUV metrics improve by approximately 2%, yet do not influence the results significantly.

If we compare our results to those published in the original reports, similar variability was found for SUV_{max} [15, 16]. Repeatability of SUV_{mean} improved when threshold based segmentation was applied for the Trigonis et al. [16] cohort (RC: 29.8 vs. 21.1). In contrast, variability of SUV_{mean} increased in the FBP dataset compared to manual delineation (RC: 20.6 vs. 41.9) [14]. This is also seen when other segmentation algorithms are used for lesion delineation in this FBP reconstructed dataset and raises the issue of appropriateness of semi-automatic segmentation in FBP reconstructed images [21]. Unfortunately, the raw data of this dataset were not available, so no reconstruction using OSEM could be performed.

The repeatability of ^{18}F -FLT SUV metrics from this study is better than the 30% threshold suggested by PET response criteria in solid tumors (PERCIST) for ^{18}F -FDG PET. The repeatability is similar to that found in a recent prospective multi-center study ($n = 10$ patients, one lesion per patient; five institutions) on ^{18}F -FLT in gliomas (RCs 19–23%) [22]. In addition, our results are in line with multiple other single-center repeatability studies for several different tracers [12, 23, 24]. In general, multi-institutional studies yield higher variability (RCs 28–47%) [10, 11, 13]. The lower variability found in this study might be partly explained by the fact that data were acquired in strictly controlled single-center setting. Moreover, no differences in uptake time between the test and retest scans were present because static images were generated from dynamic scans. This removed the variability in uptake time on SUV that is typically encountered when acquiring static images. However, a previous study has shown that ^{18}F -FLT tumor uptake reached equilibrium at 30 min post injection in NSCLC [19].

Several other studies also found poorer repeatability of volumetric metrics compared to SUV metrics (RCs >

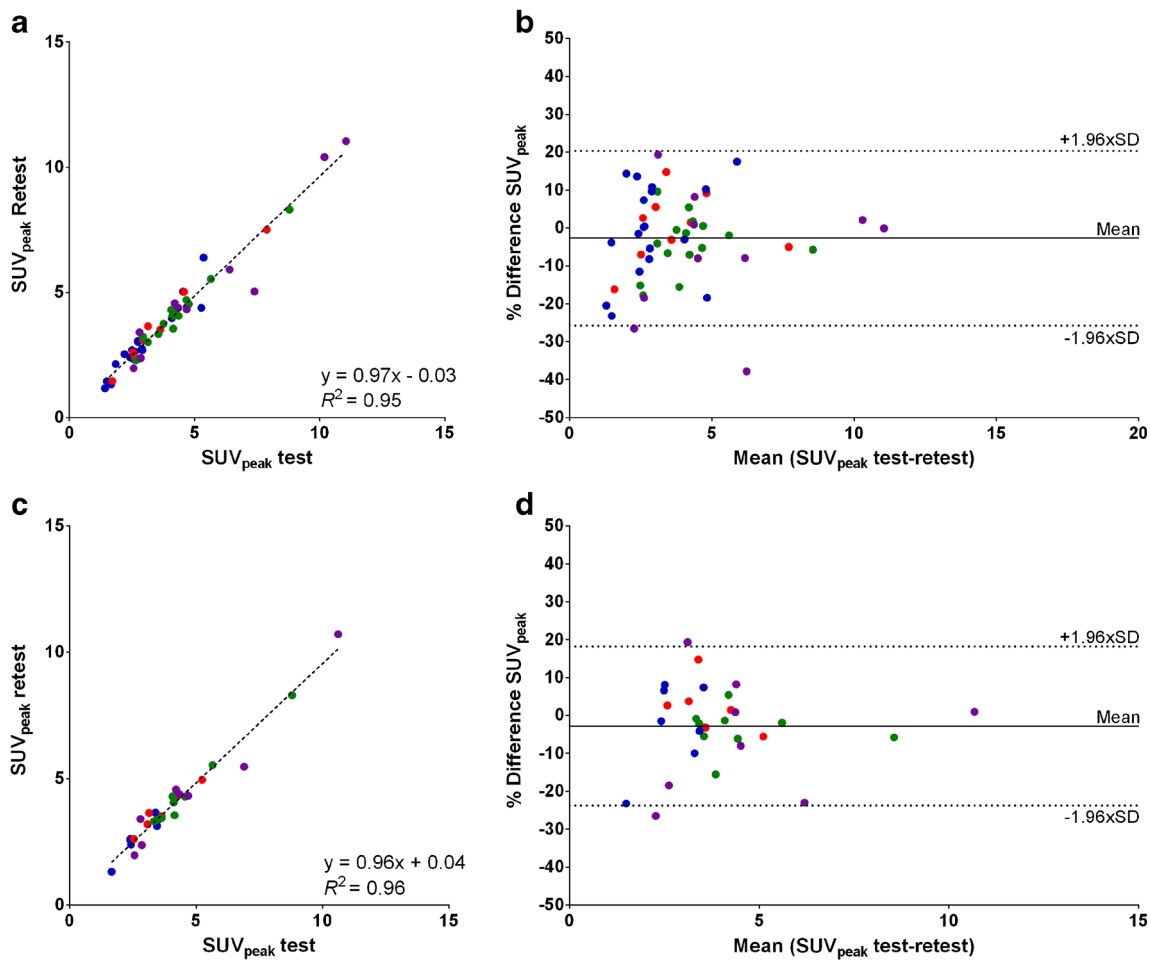


Fig. 3 Test-and-retest SUV_{peak} plotted reciprocally per lesion (a) and per patient (c) with corresponding Bland–Altman plots (b and d, respectively). Similar patterns were seen for other SUV metrics. (● Trigonis; ● de Langen [HNC]; ● de Langen [NSCLC]; ● Kenny)

30%) [12, 18]. In our study, VOIs were defined using semi-automatic segmentation to minimize user dependency. In two out of three original reports, manual delineation was used, potentially contributing to the observed differences [14, 16]. It was expected that repeatability of volumetric metrics would be slightly worse in the FBP dataset

due to higher noise levels and streak artifacts. In contrast to our expectation, PET/CT data showed a higher variability of proliferative volume and TLU compared to PET only data. Moreover, variability of proliferative volume was larger in our study compared to the original report for the PET/CT data (RCs 43.7 vs. 30.6%) [16]. This

Table 3 Mean relative differences and RCs on lesion level for several uptake metrics

Quantitative tracer uptake measures	Overall		Kenny		de Langen						Trigonis	
	Mean difference (%)	RC (%)	Mean difference (%)	RC (%)	Overall		NSCLC		HNC		NSCLC	
					Mean difference (%)	RC (%)	Mean difference (%)	RC (%)	Mean difference (%)	RC (%)	Mean difference (%)	RC (%)
SUV _{max}	-3.14	25.54	-9.05	25.86	-3.38	19.26	-5.60	19.80	0.32	16.91	0.47	31.13
SUV _{peak}	-2.72	23.06	-6.83	33.22	-2.56	16.42	-4.24	14.96	0.24	18.16	-0.65	24.29
SUV _{mean}	-3.32	25.16	-12.62	41.89	-1.40	14.42	-2.80	13.01	0.93	16.24	-0.72	21.12
TLU	3.70	36.38	-5.43	37.03	0.06	27.30	1.69	29.75	-2.65	23.32	12.09	41.88
Volume	5.43	35.95	-0.01	40.64	1.45	24.35	4.47	27.46	-3.59	14.47	12.84	43.68

BC breast cancer, NSCLC non-small cell lung cancer, HNC head and neck cancer, SUV standardized uptake value, TLU total lesion uptake

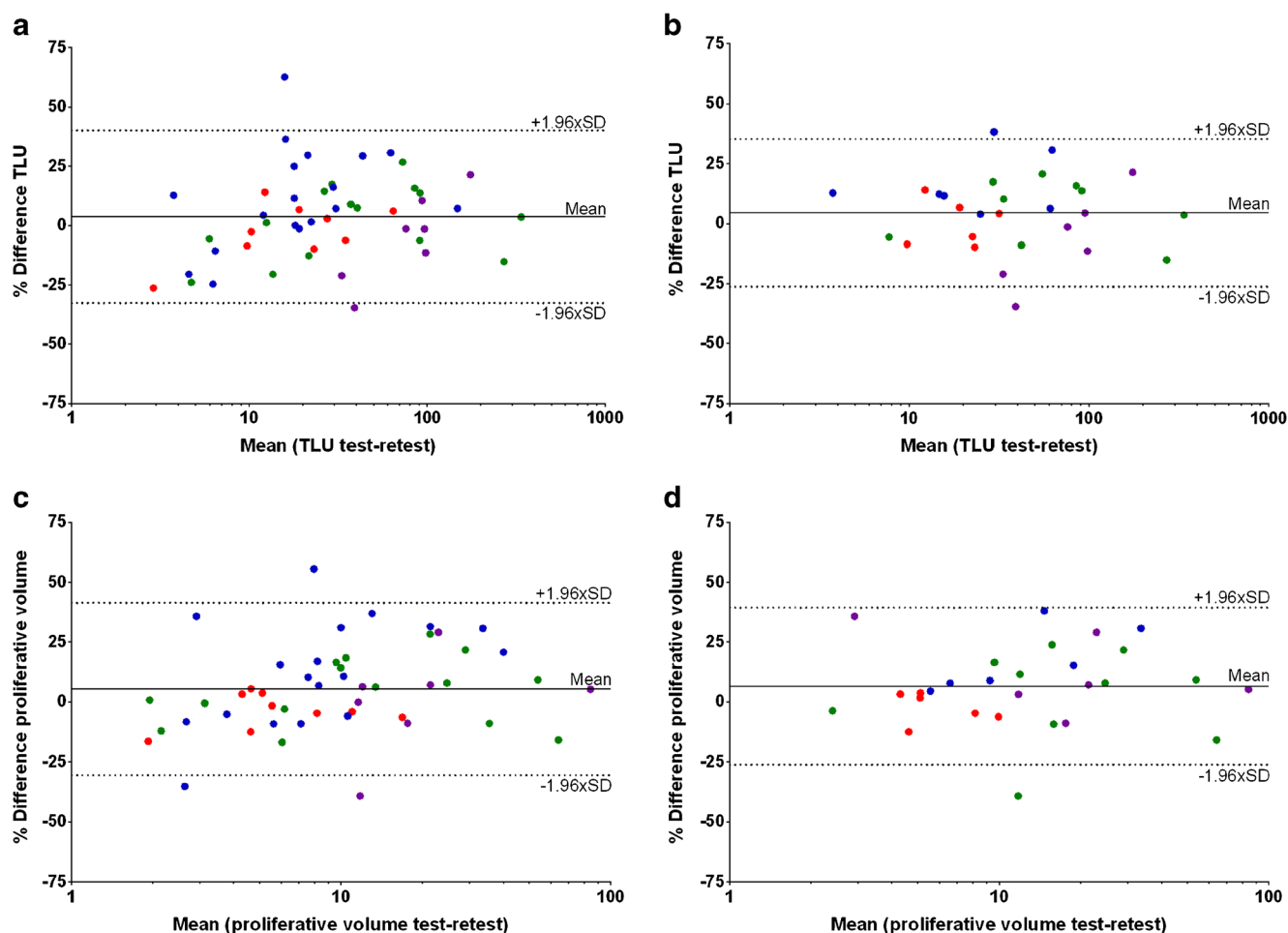


Fig. 4 Bland–Altman plots of total lesion uptake (TLU) and proliferative volume on lesion (a and c, respectively) and patient level (b and d, respectively). (● Trigonis; ● de Langen [HNC]; ● de Langen [NSCLC]; ● Kenny)

discrepancy was mainly caused by low ^{18}F -FLT uptake of lesions in the PET/CT dataset, resulting in low tumor-to-background ratios. As semi-automatic segmentation methods require adequate contrast between tumor and background radioactivity, accurate VOI definition can be

compromised. This is supported by the fact that results significantly improve when including only lesions with $\text{SUV}_{\text{max}} > 4.0$.

Two studies validating simplified quantitative metrics of ^{18}F -FLT uptake in NSCLC showed a stronger

Table 4 Mean relative differences and RCs on patient level for several uptake metrics

Quantitative tracer uptake measures	Overall		Kenny		de Langen					Trigonis		
	Mean difference (%)	RC (%)	Mean difference (%)	RC (%)	Overall		NSCLC		HNC		NSCLC	
					Mean difference (%)	RC (%)	Mean difference (%)	RC (%)	Mean difference (%)	RC (%)	Mean difference (%)	RC (%)
SUV_{max}	-3.54	20.63	-8.61	21.82	-1.82	14.92	-4.27	14.47	1.85	13.59	-1.43	28.31
SUV_{peak}	-2.76	21.00	-5.81	31.50	-1.32	13.21	-3.74	11.07	2.31	13.82	-2.37	22.34
SUV_{mean}	-3.99	26.44	-12.14	43.26	-0.18	10.93	-1.72	9.06	2.14	12.72	-2.83	20.79
TLU	4.51	30.85	-7.10	38.73	3.52	22.83	5.75	25.26	0.18	18.79	16.58	25.22
Volume	6.67	32.81	-0.52	44.24	3.23	23.60	6.96	26.95	-2.36	12.53	20.19	27.98

BC breast cancer, NSCLC non-small cell lung cancer, HNC head and neck cancer, SUV standardized uptake value, TLU total lesion uptake

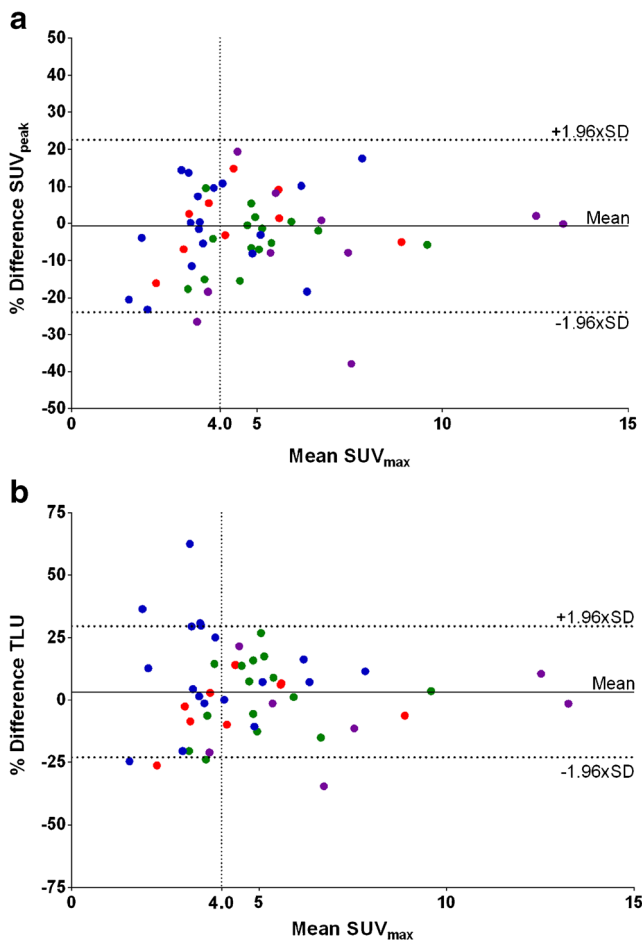


Fig. 5 Variability of SUV_{peak} (a) and proliferative volume (b) plotted against SUV_{max} . The dashed horizontal lines indicate the cut-off values used for the lesions selection strategies. (● Trigonis; ● de Langen [HNC]; ● de Langen [NSCLC]; ● Kenny)

correlation of TBR with the uptake constant K_i (estimated from kinetic analysis) compared to SUV [19, 25]. In our study, we found that normalizing SUV to blood pool radioactivity concentrations significantly increases variability for ^{18}F -FLT images reconstructed with FBP. Moreover, TBR has been shown to be highly time dependent for ^{18}F -FLT, limiting its use in response assessment, especially in busy clinical settings [19, 26].

It is suggested that assessment of response per patient rather than per lesion may improve correlation with patient outcome [27]. Similar to other studies, assessing repeatability per patient improved RCs by reducing the non-systematic differences between the test-and-retest scans. To our knowledge, only one study has been performed comparing response assessment per patient and per lesion [28]. Here, no significant differences in performance of the two methods were found. Yet, in this ^{18}F -FDG study, the same threshold of 30% to differentiate between stable disease and progressive disease or partial response was used for both methods [28]. We therefore propose that

future response assessment studies with ^{18}F -FLT PET/CT should also assess the response per patient, while taking the per-patient variability into account.

In the current study, we have used symmetric limits to assess repeatability of quantitative ^{18}F -FLT uptake metrics. Symmetrical RCs are commonly used in PET repeatability literature, however recent papers have discussed their applicability in daily clinical practice [10, 29]. In test-retest studies, often no golden standard is available and therefore relative differences are calculated using the average of the two measurements. This differs from response assessment in clinical setting where change is determined relative to a single baseline value and therefore asymmetrical RCs are suggested to be more suitable. If we calculate asymmetric RCs at lesion level, the overall upper (URC) and lower (LRC) limits of the RCs are: SUV_{max} (URC: 29.4%; LRC: -22.7%); SUV_{mean} (URC: 29.0%; LRC: -22.5%); SUV_{peak} (URC: 26.0%; LRC: -20.6%); TLU (URC: 44.6%; LRC: -30.9%); and volume (URC: 43.7%; LRC: -30.4%). These results show a slight shift in RCs of SUV metrics compared to the symmetric limits, however remain within 30%. On a patient level, asymmetrical RCs improved RCs of SUV: SUV_{max} (URC: 21.1%; LRC: -18.3%); SUV_{mean} (URC: 15.3%; LRC: -23.3%); SUV_{peak} (URC: 16.8%; LRC: -18.8%); TLU (URC: 34.1%; LRC: -27.9%); and volume (URC: 36.3%; LRC: -28.7%).

The use of different PET scanners and the heterogeneity in reconstruction methods between cohorts could have contributed to the variability in the uptake and volumetric metrics. However, despite these limitations, repeatability of ^{18}F -FLT was better compared to several other standardized multi-center studies that prospectively evaluated repeatability of ^{18}F -FDG. In contrast to other meta-analyses, we increased robustness by re-analyzing all scans and thus minimizing variability due to data analysis and allowing direct comparison of quantitative uptake metrics. To date, this individual patient data meta-analysis provides the largest test-retest ^{18}F -FLT PET cohort. These results should ideally be confirmed in a large prospective multi-center PET/CT study.

Conclusions

In this multi-center, individual patient data meta-analysis, we found that repeatability of ^{18}F -FLT tumor uptake is comparable to that of ^{18}F -FDG PET/CT. In multi-center studies, a 25% and 20% difference in individual ^{18}F -FLT SUV metrics likely represents a true change in tumor uptake at lesion and patient level, respectively. In case of volumetric measurements, higher thresholds are needed compared to SUV metrics, especially for lesions with $SUV_{max} < 4.0$ at baseline.

Funding The research leading to these results has received support from the Innovative Medicines Initiative Joint Undertaking (www.imi.europa.eu; grant agreement number 115151), whose resources are composed of a financial contribution from the European Union's Seventh Framework Programme (FP7/2007–2013) and an in-kind contribution from the companies of the European Federation of Pharmaceutical Industries and Associations. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Compliance with ethical standards

Conflict of interest There are no conflicts of interests.

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

Informed consent Informed consent was obtained from all individual participants included in the study.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer*. 2015;136(5):E359–86. <https://doi.org/10.1002/ijc.29210>.
2. Wahl RL, Jacene H, Kasamon Y, Lodge MA. From RECIST to PERCIST: evolving considerations for PET response criteria in solid tumors. *J Nucl Med : Off Publ, Soc Nucl Med*. 2009;50(Suppl 1):122S–50S. <https://doi.org/10.2967/jnumed.108.057307>.
3. Chiou VL, Burotto M. Pseudoprogression and immune-related response in solid tumors. *J Clin Oncol : Off J Am Soc Clin Oncol*. 2015;33(31):3541–3. <https://doi.org/10.1200/JCO.2015.61.6870>.
4. Bollineni VR, Collette S, Liu Y. Functional and molecular imaging in cancer drug development. *Chin Clin Oncol*. 2014;3(2):17. <https://doi.org/10.3978/j.issn.2304-3865.2014.05.05>.
5. Crippa F, Agresti R, Sandri M, Mariani G, Padovano B, Alessi A, et al. (1)(8)F-FLT PET/CT as an imaging tool for early prediction of pathological response in patients with locally advanced breast cancer treated with neoadjuvant chemotherapy: a pilot study. *Eur J Nucl Med Mol Imaging*. 2015;42(6):818–30. <https://doi.org/10.1007/s00259-015-2995-8>.
6. Buck AK, Halter G, Schirrmeister H, Kotzerke J, Wurziger I, Glatting G, et al. Imaging proliferation in lung tumors with PET: 18F-FLT versus 18F-FDG. *J Nucl Med : Off Publ, Soc Nucl Med*. 2003;44(9):1426–31.
7. Chalkidou A, Landau DB, Odell EW, Cornelius VR, O'Doherty MJ, Marsden PK. Correlation between Ki-67 immunohistochemistry and 18F-fluorothymidine uptake in patients with cancer: a systematic review and meta-analysis. *Eur J Cancer*. 2012;48(18):3499–513. <https://doi.org/10.1016/j.ejca.2012.05.001>.
8. Li XF, Dai D, Song XY, Liu JJ, Zhu YJ, Xu WG. Comparison of the diagnostic performance of 18F-fluorothymidine versus 18F-fluorodeoxyglucose positron emission tomography on pulmonary lesions: a meta analysis. *Molec Clin Oncol*. 2015;3(1):101–8. <https://doi.org/10.3892/mco.2014.440>.
9. Bollineni VR, Kramer GM, Jansma EP, Liu Y, Oyen WJ. A systematic review on [(18)F]FLT-PET uptake as a measure of treatment response in cancer patients. *Eur J Cancer*. 2016;55:81–97. <https://doi.org/10.1016/j.ejca.2015.11.018>.
10. Weber WA, Gatsonis CA, Mozley PD, Hanna LG, Shields AF, Aberle DR, et al. Repeatability of 18F-FDG PET/CT in advanced non-small cell lung cancer: prospective assessment in 2 multicenter trials. *J Nucl Med : Off Publ, Soc Nucl Med*. 2015;56(8):1137–43. <https://doi.org/10.2967/jnumed.114.147728>.
11. 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 metaanalysis. *J Nucl Med : Off Publ, Soc Nucl Med*. 2012;53(5):701–8. <https://doi.org/10.2967/jnumed.111.095299>.
12. Kramer GM, Frings V, Hoetjes N, Hoekstra OS, Smit EF, de Langen AJ, et al. Repeatability of quantitative whole-body 18F-FDG PET/CT uptake measures as function of uptake interval and lesion selection in non-small cell lung cancer patients. *J Nucl Med : Off Publ, Soc Nucl Med*. 2016;57(9):1343–9. <https://doi.org/10.2967/jnumed.115.170225>.
13. Velasquez LM, Boellaard R, Kollia G, Hayes W, Hoekstra OS, Lammertsma AA, et al. Repeatability of 18F-FDG PET in a multicenter phase I study of patients with advanced gastrointestinal malignancies. *J Nucl Med : Off Publ, Soc Nucl Med*. 2009;50(10):1646–54. <https://doi.org/10.2967/jnumed.109.063347>.
14. Kenny L, Coombes RC, Vigushin DM, Al-Nahhas A, Shousha S, Aboagye EO. Imaging early changes in proliferation at 1 week post chemotherapy: a pilot study in breast cancer patients with 3'-deoxy-3'-[18F]fluorothymidine positron emission tomography. *Eur J Nucl Med Mol Imaging*. 2007;34(9):1339–47. <https://doi.org/10.1007/s00259-007-0379-4>.
15. de Langen AJ, Klabbers B, Lubberink M, Boellaard R, Spreuwenberg MD, Slotman BJ, et al. Reproducibility of quantitative 18F-3'-deoxy-3'-fluorothymidine measurements using positron emission tomography. *Eur J Nucl Med Mol Imaging*. 2009;36(3):389–95. <https://doi.org/10.1007/s00259-008-0960-5>.
16. Trigonis I, Koh PK, Taylor B, Tamal M, Ryder D, Earl M, et al. Early reduction in tumour [18F]fluorothymidine (FLT) uptake in patients with non-small cell lung cancer (NSCLC) treated with radiotherapy alone. *Eur J Nucl Med Mol Imaging*. 2014;41(4):682–93. <https://doi.org/10.1007/s00259-013-2632-3>.
17. Shields AF, Lawhorn-Crews JM, Briston DA, Zalzal S, Gadgeel S, Douglas KA, et al. Analysis and reproducibility of 3'-deoxy-3'-[18F]fluorothymidine positron emission tomography imaging in patients with non-small cell lung cancer. *Clin Cancer Res : Off J Am Assoc Cancer Res*. 2008;14(14):4463–8. <https://doi.org/10.1158/1078-0432.CCR-07-5243>.
18. Frings V, de Langen AJ, Smit EF, van Velden FH, Hoekstra OS, van Tinteren H, et al. Repeatability of metabolically active volume measurements with 18F-FDG and 18F-FLT PET in non-small cell lung cancer. *J Nucl Med : Off Publ, Soc Nucl Med*. 2010;51(12):1870–7. <https://doi.org/10.2967/jnumed.110.077255>.
19. Frings V, Yaqub M, Hoyng LL, Golla SS, Windhorst AD, Schuit RC, et al. Assessment of simplified methods to measure 18F-FLT uptake changes in EGFR-mutated non-small cell lung cancer patients undergoing EGFR tyrosine kinase inhibitor treatment. *J Nucl Med : Off Publ, Soc Nucl Med*. 2014;55(9):1417–23. <https://doi.org/10.2967/jnumed.114.140913>.
20. Ludbrook J. Linear regression analysis for comparing two measurers or methods of measurement: but which regression? *Clin*

- Exp Pharmacol Physiol. 2010;37(7):692–9. <https://doi.org/10.1111/j.1440-1681.2010.05376.x>.
21. Hatt M, Cheze-Le Rest C, Aboagye EO, Kenny LM, Rosso L, Turkheimer FE, et al. Reproducibility of 18F-FDG and 3'-deoxy-3'-18F-fluorothymidine PET tumor volume measurements. *J Nucl Med : Off Publ, Soc Nucl Med*. 2010;51(9):1368–76. <https://doi.org/10.2967/jnumed.110.078501>.
 22. Lodge MA, Holdhoff M, Leal JP, Bag AK, Nabors LB, Mintz A, et al. Repeatability of 18F-FLT PET in a multicenter study of patients with high-grade glioma. *J Nucl Med : Off Publ, Soc Nucl Med*. 2017;58(3):393–8. <https://doi.org/10.2967/jnumed.116.178434>.
 23. Oprea-Lager DE, Kramer G, van de Ven PM, van den Eertwegh AJ, van Moorselaar RJ, Schober P, et al. Repeatability of quantitative 18F-fluoromethylcholine PET/CT studies in prostate cancer. *J Nucl Med : Off Publ, Soc Nucl Med*. 2016;57(5):721–7. <https://doi.org/10.2967/jnumed.115.167692>.
 24. Rockall AG, Avril N, Lam R, Iannone R, Mozley PD, Parkinson C, 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 : Off J Am Assoc Cancer Res*. 2014;20(10):2751–60. <https://doi.org/10.1158/1078-0432.CCR-13-2634>.
 25. Lubberink M, Direcks W, Emmering J, van Tinteren H, Hoekstra OS, van der Hoeven JJ, et al. Validity of simplified 3'-deoxy-3'-[18F]fluorothymidine uptake measures for monitoring response to chemotherapy in locally advanced breast cancer. *Molec Imaging Biol : MIB : Off Publ Acad Molec Imaging*. 2012;14(6):777–82. <https://doi.org/10.1007/s11307-012-0547-1>.
 26. Kumar V, Nath K, Berman CG, Kim J, Tanvetyanon T, Chiappori AA, et al. Variance of SUVs for FDG-PET/CT is greater in clinical practice than under ideal study settings. *Clin Nucl Med*. 2013;38(3):175–82. <https://doi.org/10.1097/RLU.0b013e318279ffdf>.
 27. Reck M, Heigener DF, Mok T, Soria JC, Rabe KF. Management of non-small-cell lung cancer: recent developments. *Lancet*. 2013;382(9893):709–19. [https://doi.org/10.1016/S0140-6736\(13\)61502-0](https://doi.org/10.1016/S0140-6736(13)61502-0).
 28. Pinker K, Riedl CC, Ong L, Jochelson M, Ulaner GA, McArthur H, et al. The impact that number of analyzed metastatic breast cancer lesions has on response assessment by 18F-FDG PET/CT using PERCIST. *J Nucl Med : Off Publ, Soc Nucl Med*. 2016;57(7):1102–4. <https://doi.org/10.2967/jnumed.115.166629>.
 29. Lodge MA. Repeatability of SUV in oncologic (18)F-FDG PET. *J Nucl Med : Off Publ, Soc Nucl Med*. 2017;58(4):523–32. <https://doi.org/10.2967/jnumed.116.186353>.