ORIGINAL RESEARCH

Open Access

Repeatability of ^{18}F-FDG uptake in metastatic bone lesions of breast cancer patients and implications for accrual to clinical trials

Mark Muzi[1*] , Lanell M. Peterson[1], Jennifer M. Specht[1], Daniel S. Hippe[1], Alena Novakova-Jiresova[2], Jean H. Lee[1], Brenda F. Kurland[1], David A. Mankoff[3], Nancy Obuchowski[4], Hannah M. Linden[1] and Paul E. Kinahan[1]

Abstract

Background Standard measures of response such as Response Evaluation Criteria in Solid Tumors are ineffective for bone lesions, often making breast cancer patients that have bone-dominant metastases ineligible for clinical trials with potentially helpful therapies. In this study we prospectively evaluated the test-retest uptake variability of 2-deoxy-2-[18F]fluoro-D-glucose (^{18}F-FDG) in a cohort of breast cancer patients with bone-dominant metastases to determine response criteria. The thresholds for 95% specificity of change versus no-change were then applied to a second cohort of breast cancer patients with bone-dominant metastases.

Methods For this study, nine patients with 38 bone lesions were imaged with ^{18}F-FDG in the same calibrated scanner twice within 14 days. Tumor uptake was quantified by the most commonly used PET parameter, the maximum tumor voxel normalized by dose and body weight (SUVmax) and also by the mean of a 1-cc maximal uptake volume normalized by dose and lean-body-mass (SULpeak). The asymmetric repeatability coefficients with confidence intervals for SUVmax and SULpeak were used to determine the limits of ^{18}F-FDG uptake variability. A second cohort of 28 breast cancer patients with bone-dominant metastases that had 146 metastatic bone lesions was imaged with ^{18}F-FDG before and after standard-of-care therapy for response assessment.

Results The mean relative difference of SUVmax and SULpeak in 38 bone tumors of the first cohort were 4.3% and 6.7%. The upper and lower asymmetric limits of the repeatability coefficient were 19.4% and − 16.3% for SUVmax, and 21.2% and − 17.5% for SULpeak. ^{18}F-FDG repeatability coefficient confidence intervals resulted in the following patient stratification using SULpeak for the second patient cohort: 11-progressive disease, 5-stable disease, 7-partial response, and 1-complete response with three inevaluable patients. The asymmetric repeatability coefficients response criteria for SULpeak changed the status of 3 patients compared to the standard Positron Emission Tomography Response Criteria in Solid Tumors of ± 30% SULpeak.

Conclusion In evaluating bone tumor response for breast cancer patients with bone-dominant metastases using ^{18}F-FDG SUVmax, the repeatability coefficients from test-retest studies show that reductions of more than 17% and increases of more than 20% are unlikely to be due to measurement variability. Serial ^{18}F-FDG imaging in clinical trials

*Correspondence:
Mark Muzi
muzi@uw.edu

Full list of author information is available at the end of the article

investigating bone lesions in these patients, such as the ECOG-ACRIN EA1183 trial, benefit from confidence limits that allow interpretation of response.

Keywords Breast cancer, Bone metastases, ^{18}F-FDG PET, Test-retest, Repeatability

Introduction

Breast cancer is the most common malignancy and second leading cause of cancer death in women [1], and bone is the most common site of metastasis in breast cancer [2–5]. The appearance and behavior of bone metastases can be detected on a wide variety of clinical imaging studies (e.g. x-ray computed tomography, bone scan, magnetic resonance imaging, (2-deoxy-2-^{18}F-fluoro-D-glucose) ^{18}F-FDG using positron emission tomography with computed tomography attenuation mapping (PET/CT [6]). that are performed for different indications.

Imaging-based response criteria are often used to determine the efficacy of new therapeutic agents in cancer treatment trials. The most commonly used set of criteria in clinical trials is the Response Evaluation Criteria in Solid Tumors version 1.1 (RECIST) [7], which focuses predominantly on the physical dimensions of solid tumors from CT scans, similar to other size-based criteria such as those from the World Health Organization (WHO) [8]. However, CT does not evaluate the bone or bone marrow, but only the osteoblastic reaction in healing bone [9]. For this reason, RECIST criteria specify that bone lesions without soft-tissue components are non-measurable, non-target lesions. As a result, patients with bone dominant disease are often excluded from clinical trials due to a lack of RECIST measurable disease [10–12].

There is active interest in using measures of ^{18}F-FDG uptake with PET/CT imaging as a biomarker to assess early response to therapy for multiple types of cancer [13, 14]. For breast cancer, the AVATAXHER trial [15] and recently, the 2019 results of the TBCRC026 trial along with at least 11 other studies [6, 16–26], support using PET imaging as an effective method of measuring early breast cancer response in vivo.

An early effort to define PET-based response criteria for clinical trials was led by the European Organization for Research and Treatment of Cancer (EORTC) in 1999 [27]. The EORTC response criteria were expanded and modified by Wahl and colleagues in 2009 for the Positron Emission Tomography Response Criteria in Solid Tumors, or PERCIST [28]. Multiple clinical studies have shown that response assessment by EORTC criteria and PERCIST leads to similar response classifications [29]. In addition, there are preliminary data that suggest that response assessment by PERCIST is better correlated with patient outcome and may be a better predictor for the effectiveness of new anti-cancer therapies

than RECIST [30]. However there have only been very limited reported evaluations of the use of PET imaging specifically for response assessment of osseous metastases from breast cancer [9, 31], and an extension of PERCIST to metastatic bone disease is not yet established [21]. Peterson et al. [32] evaluated a modified version of PERCIST inclusion criteria (mPERCIST) accounting for the lower standardized-uptake-values (SUVs) of osseous lesions compared to the soft-tissue lesions previously studied using PERCIST. This study found that changes in ^{18}F-FDG-PET uptake during therapy were predictive of time to skeletal-related events (tSRE) and time-to-progression (TTP).

To design effective response criteria, an understanding of the test-retest variability is needed. In this study we prospectively evaluated the test-retest variability of ^{18}F-FDG-PET uptake in a cohort of breast cancer patients with metastatic bone-dominant lesions (BD-MBC) using the mPERCIST inclusion criteria. The calculated thresholds for 95% specificity of change versus no-change from the test-retest data were then applied to a second cohort of BD-MBC patients who had ^{18}F-FDG-PET scans both pre-therapy and after start of therapy. The classifications of change status were compared to those using EORTC, PERCIST, and recently published thresholds for soft-tissue cancers from the QIBA Profile [33].

Materials and methods
Patient selection
Cohort-1
Repeatability was assessed in a cohort of nine stage IV BD-MBC patients with stable bone disease that underwent two ^{18}F-FDG PET/CT studies on the same scanner within a two-week duration or less with no interval change in therapy. Patient and scan characteristics for cohort-1 appear in Table 1.

Cohort-2
A second retrospective cohort of 28 BD-MBC patients with planned standard-of-care therapy (including endocrine therapy, chemotherapy, and biological therapies) were imaged with ^{18}F-FDG before and within 30 days following therapy. Aspects of this study have been presented elsewhere [32].

Ethics and Consent
Patients in both cohorts were recruited from the Seattle Cancer Care Alliance or the University of Washington Medical center (Seattle, WA), and signed informed

Table 1 Cohort-1 test-retest patient characteristics

		Bone	Breast			^{18}F-FDG Scan Differences§			
		Lesions	Cancer	ER/PR/Her2†	Treatment	ΔDays	ΔUT	Δ%Dose	Δ[Glc]
Case	Age	(n)	Type*	Status	Therapy‡	(days)	(min)	(%)	(mg/dL)
01	51	9	IDC	+/+/+	Horm/HDT	14	3.0	-13%	5.0
02	55	2	IDC	+/+/=	Chemo	8	1.0	0%	-14.3
03	60	9	IDC	-/-/+	Chemo/HDT	8	0.0	-7%	1.0
05	38	1	IDC	+/+/=	Not Started	2	0.0	2%	-14.0
06	62	7	Mixed	+/+/-	Chemo	7	0.0	3%	-3.7
07	52	3	IDC	+/+/-	Horm	14	0.0	-3%	-6.7
18	67	3	DCIS	+/-/-	Not Started	12	6.0	-6%	19.0
24	32	4	Unk	+/-/+	Horm/HDT	8	3.0	1%	7.0
26	57	1	IDC	+/+/-	Horm	7	-4.0	-16%	4.0
Mean	51	5	-	-	-	9	1.0	-4%	-0.3
SD	11	3	-	-	-	4	2.8	7%	10.7

*Type of cancer is invasive ductal carcinoma (IDC), mixed lobular and ductal carcinoma (Mixed), ductal carcinoma in situ (DCIS) or unknown (Unk); †ER/PR/Her2 pathology status is positive (+), negative (-) or equivocal (=); ‡Treatment is endocrine-based or hormonal (Horm), chemotherapeutic (Chemo) or HER2-directed therapy (HDT). §^{18}F-FDG scan differences are ΔDays is the days between scans, ΔUT is the difference in minutes of uptake times (time between injection and scanning) between scans, Δ%Dose is the percent difference in dose between scans and Δ[Glc] is the change in blood glucose concentration between scans

consent prior to enrollment. All methods were performed in accordance with the ethical standards as laid down in the Declaration of Helsinki and its later amendments or comparable ethical standards, as approved by our local IRB (Institutional Review Board), Human Subjects and Radiation Safety committees.

PET/CT scanners and calibration

There were three PET scanners used in the study. Cohort-1 patients were all imaged on one of two General Electric (GE) Discovery STE PET/CT scanners [34], with identical reconstruction parameters, where each test-retest study was acquired on the same scanner. In addition to the recommended PET scanner calibration [33], the two scanners were cross-calibrated and quantitative performance was monitored with NIST-traceable reference sources to ensure similar quantitative accuracy [35, 36].

Most cohort-2 patients (15) were imaged on the same PET/CT scanner in serial studies. However, due to the addition of the GE Discovery STE PET/CT scanners at our center, thirteen cohort-2 patients were initially imaged on a GE Advance PET scanner [37] and underwent the second scan on a Discovery scanner. We have shown that our calibration and cross-calibration procedures and identical acquisition and reconstruction protocols provide test–retest accuracy comparable to a well-calibrated single scanner [38].

^{18}F-FDG-PET imaging protocol

The imaging protocol was performed according to clinical standards, consistent with the QIBA ^{18}F-FDG-PET/CT Profile [33]. Patients fasted for a minimum of 6 h before administration of ^{18}F-FDG. Medications that affect bone marrow uptake of the tracer (G-CSF, Epogen, or Procrit) were withheld for 2–3 weeks prior to scanning. The ^{18}F-FDG dose, obtained from Cardinal Health, ranged from 260 to 407 MBq (median 350MBq). Images were acquired with a target of 60 min after injection of ^{18}F-FDG (actual range 50–70 min) using multiple fields-of-view to image from the level of the eye orbits to mid-thigh.

Image analysis

Images from the Advance PET scanner were reconstructed using 2D filtered back projection reconstruction (4.29×4.29×4.25 mm voxel resolution), while images from the Discovery PET/CT scanners used iterative 3D reconstruction (4.29×4.29×3.27 mm voxel resolution). All reconstructions had corrections for dead time, random events, scatter, sensitivity, decay, branching ration, and attenuation. PET images were read by two qualified and experienced nuclear medicine physicians.

Quantitative uptake values (kBq/cc) for each lesion were extracted using the PMOD image analysis software (PMOD Technologies V4.1, Zurich, CH). SUVpeak volumes-of-interest (VOIs) were constructed as a cubic volume of approximately 1.5 cc centered on the maximum voxel (SUVmax) of each bone lesion. The average SUV of the VOI was the SUVpeak value. Both SUVmax and SUVpeak were normalized to lean-body-mass producing SULmax and SULpeak.

Statistical methods

Repeatability of SUV/SULs in metastatic bone lesions in cohort-1 patients was assessed using the procedures described by Velasquez et al. for gastrointestinal cancers [39] and Weber et al. for non–small cell lung cancer [40]. Both studies used ^{18}F-FDG-PET multicenter test-retest exams, as in the current study. A description of

the calculated metrics is summarized in Supplementary materials Table S1. Variability was assessed by calculating the difference of paired measurements, and the difference of the logs of the measurements using SUV to reflect SUVmax or SULpeak:

$$d_i = \mathrm{SUV}_{i,2} - \mathrm{SUV}_{i,1}$$

$$\Delta_i = ln\left(\mathrm{SUV}_{i,2}\right) - ln\left(\mathrm{SUV}_{i,1}\right) = ln\left(\frac{\mathrm{SUV}_{i,2}}{\mathrm{SUV}_{i,1}}\right)$$

The difference of the log of the measurements, Δ_i, can be useful where d_i does not follow a normal distribution or where the relative differences are found to be proportional to the mean [41]. The SUV measurements SUV_{i1} and SUV_{i2} are for lesion i at the time of the baseline and the follow-up scans, and are calculated using SUVmax and SULpeak, which are the most common clinical ^{18}F-FDG-PET biomarkers. The variability of the parameters d_i and the log-transformed values, Δ_i, were assessed using Bland-Altman plots. The consistency of d_i and Δ_i with a normal distribution were assessed with quantile–quantile plots and Kolmogorov–Smirnov tests.

The log-transformed data were used to calculate the mean percent difference in uptake between scans ($\%\bar{\Delta}$), within-subject coefficient of variation (wCV_Δ), the repeatability coefficient (RC), and asymmetric RC limits (-RC and +RC) as described in Supplementary materials, Table S1. The 95% confidence interval (CI) for $\%\bar{\Delta}$ (an estimate of bias between scans) did not include 0. However, this was hypothesized to be a sampling effect and to be conservative, the repeatability metrics were also calculated without subtracting the sample mean. This will include any bias into the estimate of variability and increases the associated metrics: the within-subject coefficient of variation with bias included ($wCV_{\Delta 0}$), the repeatability coefficient with bias included (RC_0) and the asymmetric repeatability coefficients with bias included (-RC_0 and +RC_0). Details of the calculations are provided in the Supplementary materials.

Metrics were calculated using the lesion as the unit of analysis. To account for non-independence of multiple lesions from the same patient, 95% CIs for the repeatability coefficients were calculated using the leave-one-patient-out jackknife method [42]. This involves estimating the standard error of the repeatability metric by recalculating the metric after one patient at a time (all lesions from that are excluded at each step) as this assumes the patients are independent but the lesions within patient are not. The Supplementary materials describes the approach in more detail.

PERCIST Quality Control
We applied the PERCIST recommendations for quality control by measuring the mean SUL of a 3 cm spherical VOI in a normal region of the right lobe of the liver to check that the difference between the scans is less than 20% and less than a SULmean value of 0.3 for both cohort-1 and cohort-2 patients.

Inclusion criteria
The PERCIST criteria for including lesions in evaluations of response to therapy is $\mathrm{SULpeak} \geq 1.5 \bullet r_L + 2 \bullet s_L$, where r_L is the mean SUL value of the normal liver region described above and s_L is the sample standard deviation of the VOI. As we have previously noted [32, 43], bone lesions appear to have lower average SULpeak values and lower coefficient of variation than soft-tissue lesions previously studied using PERCIST. In addition, it has been shown that the standard deviation of a VOI from a single image is not related to the true noise, i.e. the noise measured from multiple images of the same object [44]. For these reasons we proposed a modified PERCIST (mPERCIST) lesion inclusion criteria for bone lesions defined by liver $\mathrm{SULpeak} \geq 1.5 \bullet r_L$.

Cohort-2 patient data was used to assess the impact of PERCIST and mPERCIST thresholds for inclusion in studies, as well as the use of cohort-1 bone lesion±RC for the determination of response to therapy. The PERCIST approach uses the concept of a 'target' lesion to determine response, where only the percentage difference in SULpeak between the tumor with the highest value in study 1 and the tumor with the highest value in study 2 (i.e. not necessarily the same tumor) is used as the classifier for response. The criteria from EORTC and QIBA were also included where appropriate.

Results
Cohort-1 characteristics
Nine female breast cancer patients were enrolled in cohort-1 with an average age of 51 years (median 55, range 32–62) with metastatic bone disease. Patients had a mixture of sclerotic, lytic, or mixed-type lesions. Most of the patients were postmenopausal (7/9, 78%) with invasive ductal carcinoma (6/9, 67%). Most patients had ER positive disease (8/9, 89%), while some were HER2 negative (4/9, 44%). Seven patients were on therapy before enrolling in the study, and two had no therapy prior to the repeatability scans. For the patients that were on treatment, there were no changes to treatment between the two scans. The injected dose of ^{18}F-FDG ranged from 305 to 396 MBq for both test and retest scans (mean 368 MBq±20 MBq). The median time between scans was 8 days (range 2–14). Average glucose level for the first scan was 94 mg/dL (range 88–104) and for the second scan was 92 mg/dL (range 89–96). The uptake time from

tracer injection to the onset of imaging averaged 61 min (range 58–70 min for all scans), while the difference in uptake times between scan1 and scan2 per patient ranged from 0 to 6 min. Cohort-1 patient and scan characteristics appear in Table 1.

Repeatability of bone lesion ^{18}F-FDG uptake values

Individual SUVmax and SULpeak test-retest measurements for 38 lesions from 9 patients in cohort-1 are provided in Table S2 of the Supplementary materials. The median number of lesions per patient was 5 (range: 1 to 9 lesions). An example test-retest ^{18}F-FDG image set from cohort-1 is shown in Fig. 1, and illustrates the consistency of SUV measures between the scans. Also shown is a cohort-2 example of response to therapy as assessed by SUV.

For quantitative analysis, Bland-Altman plots of individual lesion differences for SUVmax and SULpeak are shown in Fig. 2. The corresponding Bland-Altman plots for within-patient averages of lesions are shown in Supplementary materials Figure S1.

The tests of normality of the differences, using both quantile-quantile plots (Supplemental materials Figure S2) and Kolmogorov–Smirnov tests ($p=0.88$), showed that all the results were consistent with a normal distribution. The Bland-Altman plots above indicated a potential increase in variance of the SUV differences as a function of the average value. This dependence was not apparent in the difference of the natural logarithms of the SUV values.

The derived repeatability metrics for metastatic bone lesions in breast cancer patients using log-transformed SUVmax and SULpeak measurements, which are normally distributed, are provided in Table 2. The

repeatability metrics for other extracted PET parameters, SUVpeak and SULmax, are presented in Supplementary materials Table S3.

PERCIST Quality Control: Cohort-1

For cohort-1, the average liver SULmean was 1.6 (range 1.2 to 2.0) in the first scan and 1.6 (range 1.3 to 1.8) in the second scan. The average difference between scans for ^{18}F-FDG SULmean in liver was −0.02 (range −0.21 to 0.15). The differences in patient liver SUL values between scans were well under the threshold of <0.3 SULmean suggested by PERCIST guidelines.

Cohort-2 characteristics

Patient and scanning characteristics of the 28 patients in cohort-2 are presented in Supplemental materials Table S4. After baseline ^{18}F-FDG imaging, patients received different therapies before post therapy PET imaging and were followed clinically thereafter. There were 146 metastatic bone tumors identified by a combination of ^{18}F-FDG-PET and CT imaging [32].

PERCIST Quality Control: Cohort-2

For cohort-2, 3 of the 28 patients did not meet the PERCIST quality control requirement (i.e. the difference between scans of the SULmean for a liver ROI was more than 0.3 and greater than 20%), and one patient had uninterpretable liver results from the second scan.

Assessment of inclusion criteria

The PERCIST threshold allowed assessment of 23 patients, and the mPERCIST threshold allowed assessment of 26 patients of the 28 patients in the cohort (Supplementary materials Table S5). We note that of the 3
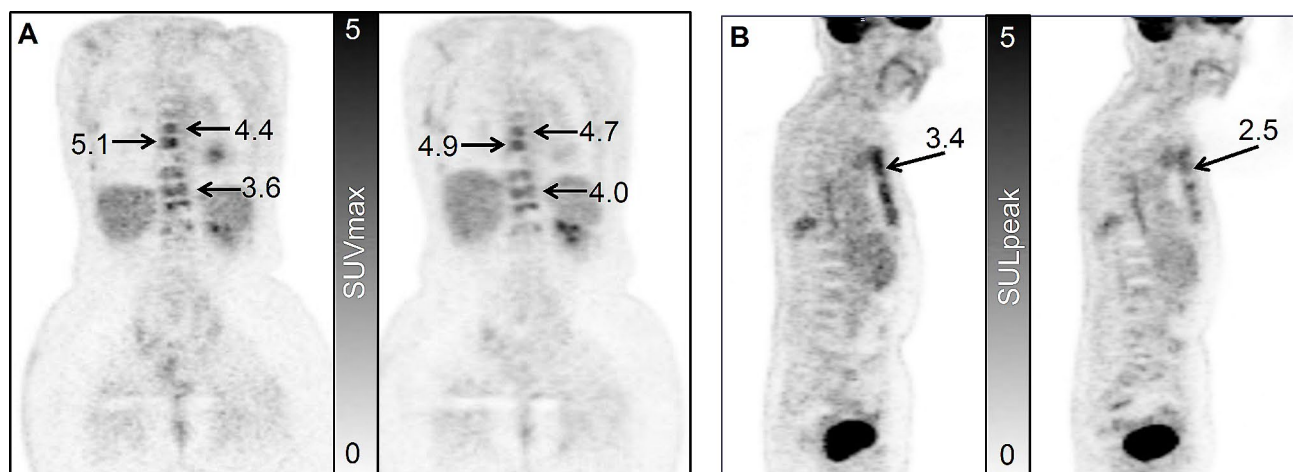

Fig. 1 (**A**) Cohort-1 study. Coronal ^{18}F-FDG-PET images from the same patient imaged 7 days apart with SUVmax values indicated. (**B**) Cohort-2 study. Sagittal ^{18}F-FDG-PET images of a 90-year old female with bone-dominant MBC. Left: Pre-therapy baseline scan showing the SULpeak of the index lesion. Right: Post therapy 4mo scan, that shows a decrease of 25%, which met our LRC threshold of -18%, but not the PERCIST threshold of -30%. The SUVmax of the index lesion decreased by 22%. The response was considered stable disease by the criteria developed in this report
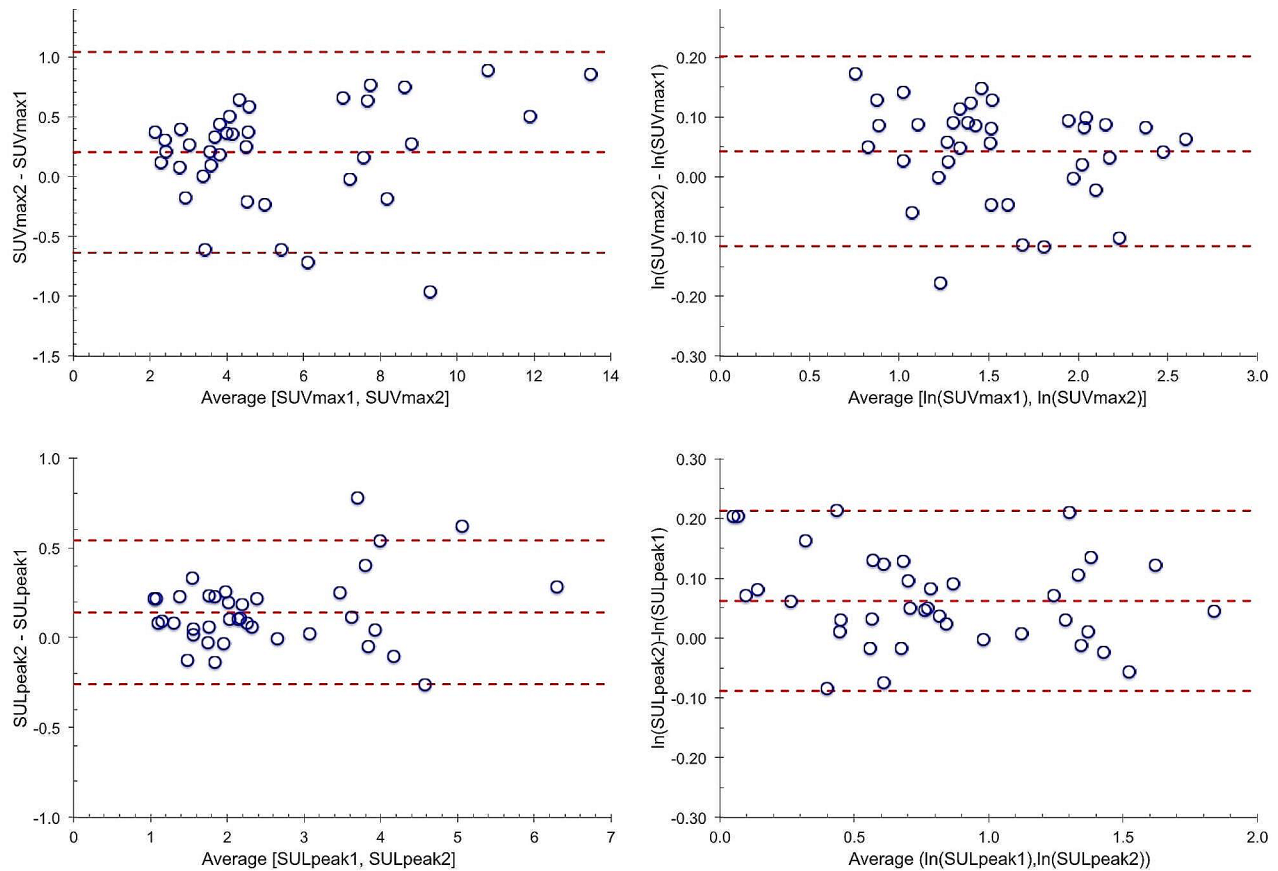
Fig. 2 Bland-Altman plots for all 38 lesions for the 9 patients. Top: SUVmax. Bottom: SULpeak. Left: Test-retest difference versus average value. Right: Differences of the natural logarithms. Dashed lines are the mean difference and the upper and lower limits of agreement

Table 2 SUV Repeatability metrics for all $n=38$ lesions

Parameter	SUVmax	SULpeak
SUVmin, SUVmax	2.0, 13.9	0.9, 6.4
$\bar{d} \pm$ SD	0.20±0.43	0.14±0.20
d_i Min, Max	-0.96, 0.88	-0.26, 0.77
*$\bar{\Delta}$ [-CI, +CI] (%)	4.4 [0.4, 8.3]	6.4 [1.1, 11.7]
-RC [-CI, +CI] (%)	-14.7 [-17.9, -11.3]	-14.0 [-18.6, -9.1]
+RC [-CI, +CI] (%)	17.2 [12.7, 21.8]	16.3 [10.0, 22.9]
$\boldsymbol{wCV}_{\Delta}$ [-CI, +CI] (%)	5.9 [4.4, 8.3]	5.6 [3.5, 7.7]
-RC$_0$ [-CI, +CI] (%)	-16.3 [-19.7, -12.8]	-17.5 [-25.3, -8.8]
+RC$_0$ [-CI, +CI] (%)	19.4 [14.6, 24.5]	21.2 [9.7, 33.9]
$\boldsymbol{wCV}_{\Delta 0}$ [-CI, +CI] (%)	6.6 [5.0, 8.2]	7.2 [3.4, 11.1]

*The 95% confidence interval of the average difference of log-transformed data, $CI_{\bar{\Delta}}$, was found to not contain zero. This was hypothesized to be a sampling effect and RCs were recalculated without subtracting the sample mean, as described in the Supplementary materials. This approach led to a slightly more conservative (i.e. larger) estimates of the repeatability coefficients and coefficient of variation, denoted in Supplementary materials Table S2 as $\pm RC_0$ and wCV$_{\Delta 0}$.

additional patients included by the mPERCIST criteria, one did not meet the PERCIST quality control requirement for liver (Case 50 in Table S5). While the PERCIST approach uses only the change in the single target lesion(s) to determine response, we also evaluated the impact of the change in inclusion thresholds on all 146 metastatic lesions in the 28 patients and found that the PERCIST threshold allowed assessment of 76 of the bone tumors (52%). The mPERCIST threshold allowed assessment of 102 (70%) of the bone tumors, a substantial increase. These changes for the target lesion ^{18}F-FDG SUVmax are illustrated in Fig. 3 along with thresholds for partial response (PR) and progressive disease (PD) based on QIBA, PERCIST, EORTC and the $\pm RC_0$ threshold developed from cohort-1 test-retest study (-RC$_0$ = -16.3% for PR and +RC$_0$=19.4% for PD). The $\pm$RC values using SULpeak are -RC$_0$ = -17.5% for PR and +RC$_0$=21.5% indicating PD.
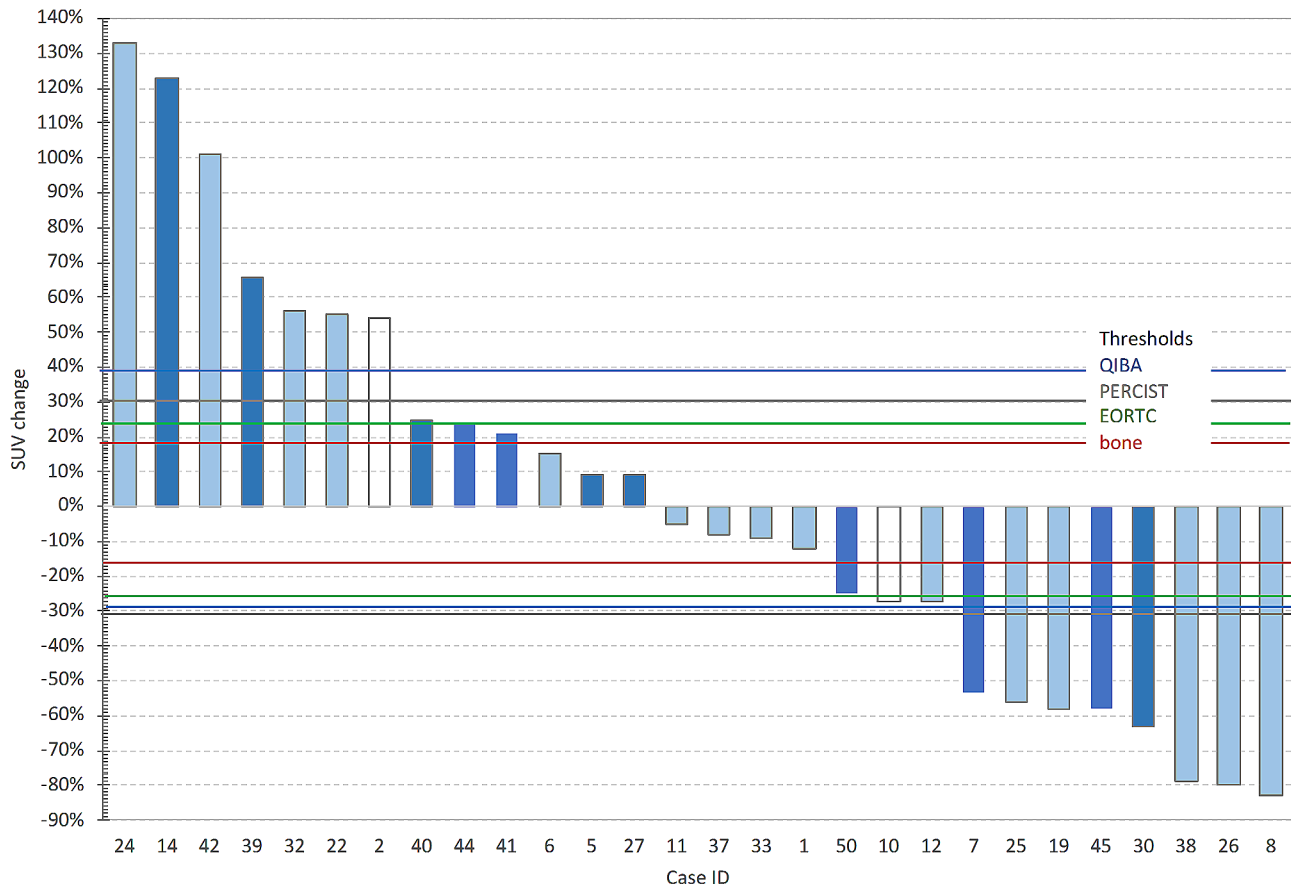
Fig. 3 Percentage change in SUVmax for cohort-2 patients ordered by magnitude of change. Dark bars are cases where new lesions appeared in the second PET scan and white bars indicate an issue with initially low SUVmax (1.0-1.8) values. The horizontal lines are the thresholds for classifying a change as determined by the PERCIST, QIBA, EORTC and this study especially for bone metastases

Assessment of Response thresholds

The response criteria developed from cohort-1 test-retest studies of ^{18}F-FDG SULpeak values in bone lesions ($\pm RC_0$) changed the response status of 4/28 patients compared to standard PERCIST response criteria. The changes were evenly divided between shifts from stable disease (SD) to progressive disease (PD) or to partial response (PR) when shifting from the PERCIST thresholds of ±30% to the bone metastasis±RC threshold of change (-17.5%, +21.2%) for ^{18}F-FDG SULpeak values. In some cases new lesions appear, which is considered an overriding determination of progressive disease, regardless of the change in SUL, PERCIST/mPERCIST threshold or PERCIST inclusion criteria.

Discussion

Our primary finding, albeit based on a study of 9 patients with a total of 38 metastatic bone lesions, was that the test-retest variability of ^{18}F-FDG uptake in bone is lower than has been previously published for soft-tissue tumors [39, 40, 45–47] or mixed tumors typical of breast cancer recurrence [36]. As summarized in the QIBA Profile summary paper [33], the within-subject coefficient of variation ranged from 10 to 12% in the above cited publications. In our study we estimated a within-subject coefficient of variation (wCV$_\Delta$) for SUVmax of 6.6% (95% CI: 5.0–8.2%) and for SULpeak of 7.2% (95% CI: 3.4–11.1%). There are two implications from this reduction in variability: First that inclusion criteria can be relaxed compared to the EORTC, PERCIST, and QIBA proposals. Second, that the thresholds for determining response can also be reduced. These comparisons are described in Table 3.

As noted above, a small bias in the mean test-retest relative difference was observed for log-transformed SUVmax and SULpeak, where corresponding 95% CIs did not include 0. However, this was thought to be due to sampling variability rather than a true bias between the two scans. To be conservative in the repeatability coefficient estimates, we recalculated the repeatability metrics without subtracting the sample mean, assuming the true bias was zero, which would in effect include the estimated bias as part of the variability and thus somewhat increasing the variability estimates. This increased the estimated

Table 3 , Comparison of EORTC and PERCIST response criteria, QIBA Profile Claims and current study

Category	EORTC criteria	PERCIST criteria	QIBA Profile Claim	This study (mPERCIST)
Measurable lesions	Region of high ^{18}F-FDG uptake representing viable tumor	SULpeak (hottest tumor) ≥ (1.5 x mean liver uptake + 2 SD of liver region)	SUVmax ≥ 4 and diameter ≥ 2 cm	SULpeak (hottest tumor) ≥ (1.5 x mean liver uptake)
No change condition (stable disease for EORTC and PERCIST)	Increase in SUV < 25% or decrease of < 15% after one cycle of treatment (and < 25% after more than one treatment cycle) and no change in extent > 20%	Change in SULpeak < ± 30%, If there are other changes that indicate response or progression.	Claim 2: Increase in SUVmax < 39% and decrease in SUVmax < 28%.	SULpeak < 21% or and decrease in SULpeak <-17% or an increase in SUVmax < 19%, and decrease in SUVmax< -16% assuming a mean difference between test and retest of zero.
Conditions for valid measurements	Scanners should provide reproducible data	Normal liver SUL must be within 20% and 0.3 SUL for baseline and follow-up studies	Acquisition and analysis must comply to the QIBA Profile checklists	Same as PERCIST.
Protocol	Specified	Refers to NIH-consensus (40) and Netherlands (41) protocols	Specified	Kurland 2019 and Peterson 2018

Note that this only contains an excerpt of the detailed EORTC and PERCIST response criteria. In addition, the EORTC and PERCIST response criteria are intended to provide information on disease status, while the QIBA Profile Claims are providing information about the statistical variability of SUVs under the assumption of no true biological change. Acronyms. EORTC: European Organization for Research and Treatment of Cancer. PERCIST: PET Response Criteria in Solid Tumors. QIBA: Quantitative Imaging Biomarkers Alliance. ^{18}F-FDG: ^{18}F-flourodexyglucose. SUV: standardized uptake value. SUVmax,: SUV calculated using the maximum value of a region placed over the image of an ^{18}F-FDG-avid lesion. SULpeak,: The mean SUV of a 1 cm diameter region centered over the maximum value of an ^{18}F-FDG-avid lesion with biodistribution normalization by lean body mass

within-subject coefficient of variation (wCV_{Δ}) from 5.9 to 6.6%. Justification for assuming a mean relative difference of zero includes; patients were scanned on the identical scanner for test and retest scans and had similar injected doses, blood glucose concentrations and uptake times. Additionally, the soft tissue tumors for these same patients in cohort-1 did not show a bias in test-retest

SUV metrics [36], which may be related to the small size but intense ^{18}F-FDG uptake in bone metastases.

We did not see a difference in reproducibility for metastatic bone lesion between types of primary breast cancer disease, such as lobular or ductal, however the number of lesions studied was limited and most patients had ductal disease.

Conclusions

Quantitative ^{18}F-FDG-PET SUV uptake values can be highly repeatable measures in breast cancer patients with bone metastases, when acquired in a well-calibrated PET scanner with careful attention to scanner calibration, acquisition protocols and image analysis. This small cohort indicates that repeat bone metastases SUV metrics can be measured with a within-patient COV ($wCV_{\Delta 0}$) of less than 8%. In evaluating response assessment in breast cancer patients with bone-dominant metastases, a percentage decrease in ^{18}F-FDG SUVmax of more than 17% (SULpeak < 18%) would indicate response, while an increases of more than 20% (SULpeak > 22%) would indicate disease progression, and unlikely to be due to measurement variability. Multicenter clinical trials, such as ECOG-ACRIN EA1183 (FEATURE) trial, assessing metastatic bone disease with ^{18}F-FDG PET/CT will directly benefit from (1) a relaxed bone PERCIST threshold for bone tumor assessment, and (2) confidence limits of bone tumor SULpeak or SUVmax that allow interpretation of response to therapy using ^{18}F-FDG uptake in bone lesions from breast cancer patients with bone-dominant metastatic disease.

Abbreviations

^{18}F-FDG	2-deoxy-2-[^{18}F]fluoro-D-glucose
95% CI	95% Confidence interval
AVATAXHER	A Study of Bevacizumab added to Trastuzumab plus Docetaxel in the Neoadjuvant Setting in Participants With Early Stage HER2-Positive Breast Cancer
BD-MBC	Bone-dominant metastatic breast cancer
ECOG-ACRIN	Eastern Cooperative Oncology Group-American Collage of Radiology Imaging Network
EORTC	European Organization for Research and Treatment of Cancer
ER	Estrogen Receptor
FEATURE ECOG-ACRIN EA1183 clinical trial	FDG PET to Assess Therapeutic Response in Patients with Bone-dominant Metastatic Breast Cancer
G-CSF	Granulocyte colony-stimulating factor
HER2	Human epidermal growth factor receptor 2
IRB	Institutional Review Board
mPERCIST	A modified version of PERCIST inclusion criteria for bone metastases

MBq	Mega-bequerel
mg/dL	Milligrams per deciliter, the clinical units of blood glucose concentration
MRI	Magnetic Resonance Imaging
NIST	National Institute of Standards and Technology
PERCIST	Positron Emission tomography Response Criteria In Solid Tumors
PET/CT	Positron Emission Tomography and Computed Tomography scanning
QIBA	Quantitative Imaging Biomarkers Alliance
RECIST	Response Evaluation Criteria In Solid Tumors
SUV	Standardized Uptake Value normalized by dose and body weight
SUVmax	Maximum voxel in a region in SUV units
SULpeak	Average activity in a ~ 1 cc volume normalized by dose and lean body mass
RC	Repeatability Coefficient
TBCRC026	Phase II Trial Correlating Standardized Uptake Value With Pathologic Complete Response to Pertuzumab and Trastuzumab in breast cancer
tSRE	Time to Skeletal Related Event
TTP	Time To Progression
VOI	Volume of Interest
wCV	Within patient coefficient of variation
WHO	World Health Organization

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13550-024-01093-7.

Supplementary Material 1

Declarations

Human ethics approval
All methods were performed in accordance with the ethical standards as laid down in the Declaration of Helsinki and its later amendments or comparable ethical standards, as approved by our local (University of Washington) IRB (Institutional Review Board), Human Subjects and Radiation Safety committees. (See Manuscript, page 6 paragraph 3 section on Ethics and Consent).

Consent to participate
Patients in both study cohorts were recruited from the Seattle Cancer Care Alliance or University of Washington Medical Center (Seattle, WA), and signed informed consent prior to enrollment. (See Manuscript, page 6 paragraph 3 section on Ethics and Consent).

Consent to publish
The authors affirm that human research participants provided informed consent for publication of the images in Fig. 1A and B. The PET imaging technicians and PET physicists, Including Ms. Wanner, gave permission to acknowledge then in the Acknowledgements section.

Competing interests
We can confidently state that there is no conflict of interest—financial or otherwise—that may directly or indirectly influence the content of this manuscript.

Author details
[1]Department of Radiology, University of Washington Medical Center, 1959 NE Pacific Street, UW Box 356465, Seattle, Washington 98195, USA
[2]Thomayer University Hospital, Prague, Czech Republic
[3]University of Pennsylvania, Philadelphia, PA, USA
[4]Cleveland Clinic, Cleveland, OH, USA

References
1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. Cancer J Clin. 2019;69:7–34. https://doi.org/10.3322/caac.21551.
2. Coleman RE. Clinical features of metastatic bone disease and risk of skeletal morbidity. Clin Cancer Res. 2006;12:s6243–9. https://doi.org/10.1158/1078-0432.ccr-06-0931.
3. Coleman RE, Rubens RD. The clinical course of bone metastases from breast cancer. Br J Cancer. 1987;55:61–6. https://doi.org/10.1038/bjc.1987.13.
4. Disibio G, French SW. Metastatic patterns of cancers: results from a large autopsy study. Arch Pathol Lab Med. 2008;132:931–9. https://doi.org/10.1043/1543-2165(2008)132[931:mpocrf]2.0.co;2.
5. Lee YT. Breast carcinoma: pattern of metastasis at autopsy. J Surg Oncol. 1983;23:175–80. https://doi.org/10.1002/jso.2930230311.
6. Contractor KB, Kenny LM, Stebbing J, Rosso L, Ahmad R, Jacob J, et al. [18F]-3'Deoxy-3'-fluorothymidine positron emission tomography and breast cancer response to docetaxel. Clin Cancer Res. 2011;17:7664–72. https://doi.org/10.1158/1078-0432.ccr-11-0783.
7. Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). Eur J Cancer. 2009;45:228–47. https://doi.org/10.1016/j.ejca.2008.10.026.
8. World Health O. WHO handbook for reporting results of cancer treatment. Geneva: World Health Organization; 1979.
9. Riedl CC, Pinker K, Ulaner GA, Ong LT, Baltzer P, Jochelson MS, et al. Comparison of FDG-PET/CT and contrast-enhanced CT for monitoring therapy response in patients with metastatic breast cancer. Eur J Nucl Med Mol Imaging. 2017;44:1428–37. https://doi.org/10.1007/s00259-017-3703-7.

10. Cook GJR, Goh V. Molecular imaging of bone metastases and their response to Therapy. J Nucl Med. 2020;61:799–806. https://doi.org/10.2967/jnumed.119.234260.

11. Cook GJ, Azad GK, Goh V. Imaging bone metastases in breast Cancer: Staging and Response Assessment. J Nucl Med. 2016;57(Suppl 1):s27–33. https://doi.org/10.2967/jnumed.115.157867.

12. Costelloe CM, Chuang HH, Madewell JE, Ueno NT. Cancer Response Criteria and Bone metastases: RECIST 1.1, MDA and PERCIST. J Cancer. 2010;1:80–92. https://doi.org/10.7150/jca.1.80.

13. Gallamini A, Zwarthoed C, Borra A. Positron Emission Tomography (PET) in Oncology. Cancers. 2014;6:1821–89. https://doi.org/10.3390/cancers6041821.

14. Weber WA. Assessing tumor response to therapy. J Nucl Med. 2009;50(Suppl 1):S1–10. https://doi.org/10.2967/jnumed.108.057174.

15. Coudert B, Pierga JY, Mouret-Reynier MA, Kerrou K, Ferrero JM, Petit T, et al. Use of [(18)F]-FDG PET to predict response to neoadjuvant trastuzumab and docetaxel in patients with HER2-positive breast cancer, and addition of bevacizumab to neoadjuvant trastuzumab and docetaxel in [(18)F]-FDG PET-predicted non-responders (AVATAXHER): an open-label, randomised phase 2 trial. Lancet Oncol. 2014;15:1493–502. https://doi.org/10.1016/s1470-2045(14)70475-9.

16. Connolly RM, Leal JP, Goetz MP, Zhang Z, Zhou XC, Jacobs LK, et al. TBCRC 008: early change in 18F-FDG uptake on PET predicts response to preoperative systemic therapy in human epidermal growth factor receptor 2-negative primary operable breast cancer. J Nucl Med. 2015;56:31–7. https://doi.org/10.2967/jnumed.114.144741.

17. Connolly RM, Leal JP, Solnes L, Huang CY, Carpenter A, Gaffney K, et al. TBCRC026: phase II trial correlating standardized uptake value with pathologic complete response to Pertuzumab and Trastuzumab in breast Cancer. J Clin Oncol. 2019;37:714–22. https://doi.org/10.1200/jco.2018.78.7986.

18. Dunnwald LK, Doot RK, Specht JM, Gralow JR, Ellis GK, Livingston RB, et al. PET tumor metabolism in locally advanced breast cancer patients undergoing neoadjuvant chemotherapy: value of static versus kinetic measures of fluorodeoxyglucose uptake. Clin Cancer Res. 2011;17:2400–9. https://doi.org/10.1158/1078-0432.ccr-10-2649.

19. Gebhart G, Gámez C, Holmes E, Robles J, Garcia C, Cortés M, et al. 18F-FDG PET/CT for early prediction of response to neoadjuvant lapatinib, trastuzumab, and their combination in HER2-positive breast cancer: results from Neo-ALTTO. J Nucl Med. 2013;54:1862–8. https://doi.org/10.2967/jnumed.112.119271.

20. Groheux D, Cochet A, Humbert O, Alberini JL, Hindié E, Mankoff D. 18F-FDG PET/CT for staging and restaging of breast Cancer. J Nucl Med. 2016;57(Suppl 1):s17–26. https://doi.org/10.2967/jnumed.115.157859.

21. Groheux D, Biard L, Lehmann-Che J, Teixeira L, Bouhidel FA, Poirot B, et al. Tumor metabolism assessed by FDG-PET/CT and tumor proliferation assessed by genomic grade index to predict response to neoadjuvant chemotherapy in triple negative breast cancer. Eur J Nucl Med Mol Imaging. 2018;45:1279–88. https://doi.org/10.1007/s00259-018-3998-z.

22. Humbert O, Berriolo-Riedinger A, Riedinger JM, Coudert B, Arnould L, Cochet A, et al. Changes in 18F-FDG tumor metabolism after a first course of neoadjuvant chemotherapy in breast cancer: influence of tumor subtypes. Annals Oncology: Official J Eur Soc Med Oncol. 2012;23:2572–7. https://doi.org/10.1093/annonc/mds071.

23. 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;34:1339–47. https://doi.org/10.1007/s00259-007-0379-4.

24. Kostakoglu L, Duan F, Idowu MO, Jolles PR, Bear HD, Muzi M, et al. A phase II study of 3′-Deoxy-3′-18F-Fluorothymidine PET in the Assessment of early response of breast Cancer to Neoadjuvant Chemotherapy: results from ACRIN 6688. J Nucl Med. 2015;56:1681–9. https://doi.org/10.2967/jnumed.115.160663.

25. Lin NU, Guo H, Yap JT, Mayer IA, Falkson CI, Hobday TJ, et al. Phase II study of Lapatinib in Combination with Trastuzumab in patients with human epidermal growth factor receptor 2-Positive metastatic breast Cancer: clinical outcomes and predictive value of early [18F]Fluorodeoxyglucose Positron Emission Tomography Imaging (TBCRC 003). J Clin Oncol. 2015;33:2623–31. https://doi.org/10.1200/jco.2014.60.0353.

26. Connolly RM, Leal JP, Solnes L, Huang CY, Carpenter A, Gaffney K, et al. Updated results of TBCRC026: phase II trial correlating standardized uptake value with pathological complete response to Pertuzumab and Trastuzumab in breast Cancer. J Clin Oncol. 2021;39:2247–56. https://doi.org/10.1200/jco.21.00280.

27. Young H, Baum R, Cremerius U, Herholz K, Hoekstra O, Lammertsma AA, et al. Measurement of clinical and subclinical tumour response using [18F]-fluorodeoxyglucose and positron emission tomography: review and 1999 EORTC recommendations. European Organization for Research and Treatment of Cancer (EORTC) PET Study Group. Eur J Cancer. 1999;35:1773–82. https://doi.org/10.1016/s0959-8049(99)00229-4.

28. 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.

29. Pinker K, Riedl C, Weber WA. Evaluating tumor response with FDG PET: updates on PERCIST, comparison with EORTC criteria and clues to future developments. Eur J Nucl Med Mol Imaging. 2017;44:55–66. https://doi.org/10.1007/s00259-017-3687-3.

30. Pinker K. Advanced Imaging for Precision Medicine in breast Cancer: from morphology to function. Breast care (Basel Switzerland). 2017;12:208–10. https://doi.org/10.1159/000480397.

31. Tateishi U, Gamez C, Dawood S, Yeung HW, Cristofanilli M, Macapinlac HA. Bone metastases in patients with metastatic breast cancer: morphologic and metabolic monitoring of response to systemic therapy with integrated PET/CT. Radiology. 2008;247:189–96. https://doi.org/10.1148/radiol.2471070567.

32. Peterson LM, O'Sullivan J, Wu QV, Novakova-Jiresova A, Jenkins I, Lee JH, et al. Prospective study of serial (18)F-FDG PET and (18)F-Fluoride PET to predict time to skeletal-related events, Time to Progression, and Survival in patients with Bone-Dominant metastatic breast Cancer. J Nucl Med. 2018;59:1823–30. https://doi.org/10.2967/jnumed.118.211102.

33. Kinahan PE, Perlman ES, Sunderland JJ, Subramaniam R, Wollenweber SD, Turkington TG, et al. The QIBA Profile for FDG PET/CT as an imaging biomarker measuring response to Cancer Therapy. Radiology. 2020;294:647–57. https://doi.org/10.1148/radiol.2019191882.

34. Macdonald LR, Schmitz RE, Alessio AM, Wollenweber SD, Stearns CW, Ganin A, et al. Measured count-rate performance of the Discovery STE PET/CT scanner in 2D, 3D and partial collimation acquisition modes. Phys Med Biol. 2008;53:3723–38. https://doi.org/10.1088/0031-9155/53/14/002.

35. Byrd DW, Doot RK, Allberg KC, MacDonald LR, McDougald WA, Elston BF, et al. Evaluation of Cross-calibrated (68)Ge/(68)Ga Phantoms for assessing PET/CT Measurement Bias in Oncology Imaging for single- and Multicenter trials. Tomography. 2016;2:353–60. https://doi.org/10.18383/j.tom.2016.00205.

36. Kurland BF, Peterson LM, Shields AT, Lee JH, Byrd DW, Novakova-Jiresova A, et al. Test-retest reproducibility of (18)F-FDG PET/CT uptake in Cancer patients within a qualified and calibrated local network. J Nucl Med. 2019;60:608–14. https://doi.org/10.2967/jnumed.118.209544.

37. DeGrado TR, Turkington TG, Williams JJ, Stearns CW, Hoffman JM, Coleman RE. Performance characteristics of a whole-body PET scanner. J Nucl Med. 1994;35:1398–406.

38. Peterson LM, Kurland BF, Schubert EK, Link JM, Gadi VK, Specht JM, et al. A phase 2 study of 16α-[18F]-fluoro-17β-estradiol positron emission tomography (FES-PET) as a marker of hormone sensitivity in metastatic breast cancer (MBC). Mol Imaging Biology. 2014;16:431–40. https://doi.org/10.1007/s11307-013-0699-7.

39. 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. 2009;50:1646–54. https://doi.org/10.2967/jnumed.109.063347.

40. 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. 2015;56:1137–43. https://doi.org/10.2967/jnumed.114.147728.

41. Bland JM, Altman DG. Measurement error proportional to the mean. BMJ (clinical research ed). 1996;313:106. https://doi.org/10.1136/bmj.313.7049.106.

42. Lipsitz SR, Laird NM, Harrington DP. Using the jackknife to estimate the variance of regression estimators from repeated measures studies. Commun Stat - Theory Methods. 1990;19:821–45. https://doi.org/10.1080/03610929008830234.

43. Muzi M, Peterson L, Novakova A, Lee J, Kurland B, Specht J, et al. Repeatability of 18F-FDG uptake values in bone lesions from breast cancer patients with metastatic bone disease. J Nucl Med. 2018;59:1369.

44. Tong S, Alessio AM, Kinahan PE. Evaluation of Noise Properties in PSF-Based PET Image Reconstruction. IEEE Nuclear Science Symposium conference

record Nuclear Science Symposium. 2009;2009:3042-7. https://doi.org/10.1109/nssmic.2009.5401574.

45. 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. 2010;51:1368–76. https://doi.org/10.2967/jnumed.110.078501.

46. Krak NC, Boellaard R, Hoekstra OS, Twisk JW, Hoekstra CJ, Lammertsma AA. Effects of ROI definition and reconstruction method on quantitative outcome and applicability in a response monitoring trial. Eur J Nucl Med Mol Imaging. 2005;32:294–301. https://doi.org/10.1007/s00259-004-1566-1.

47. Nakamoto Y, Zasadny KR, Minn H, Wahl RL. Reproducibility of common semi-quantitative parameters for evaluating lung cancer glucose metabolism with positron emission tomography using 2-deoxy-2-[18F]fluoro-D-glucose. Mol Imaging Biology. 2002;4:171–8. https://doi.org/10.1016/s1536-1632(01)00004-x.

Publisher's Note