

ORIGINAL RESEARCH

Open Access



# Visual and quantitative evaluation of [ $^{18}\text{F}$ ]FES and [ $^{18}\text{F}$ ]FDHT PET in patients with metastatic breast cancer: an interobserver variability study

Lemonitsa H. Mammatas<sup>1</sup>, Clasina M. Venema<sup>2</sup>, Carolina P. Schröder<sup>2</sup>, Henrica C. W. de Vet<sup>3</sup>, Michel van Kruchten<sup>2</sup>, Andor W. J. M. Glaudemans<sup>4</sup>, Maqsood M. Yaqub<sup>5</sup>, Henk M. W. Verheul<sup>1</sup>, Epie Boven<sup>1</sup>, Bert van der Vegt<sup>6</sup>, Erik F. J. de Vries<sup>2</sup>, Elisabeth G. E. de Vries<sup>2</sup>, Otto S. Hoekstra<sup>5</sup>, Geke A. P. Hospers<sup>2</sup> and C. Willemien Menke-van der Houven van Oordt<sup>1\*</sup>

## Abstract

**Purpose:** Correct identification of tumour receptor status is important for treatment decisions in breast cancer. [ $^{18}\text{F}$ ]FES PET and [ $^{18}\text{F}$ ]FDHT PET allow non-invasive assessment of the oestrogen (ER) and androgen receptor (AR) status of individual lesions within a patient. Despite standardised analysis techniques, interobserver variability can significantly affect the interpretation of PET results and thus clinical applicability. The purpose of this study was to determine visual and quantitative interobserver variability of [ $^{18}\text{F}$ ]FES PET and [ $^{18}\text{F}$ ]FDHT PET interpretation in patients with metastatic breast cancer.

**Methods:** In this prospective, two-centre study, patients with ER-positive metastatic breast cancer underwent both [ $^{18}\text{F}$ ]FES and [ $^{18}\text{F}$ ]FDHT PET/CT. In total, 120 lesions were identified in 10 patients with either conventional imaging (bone scan or lesions > 1 cm on high-resolution CT,  $n = 69$ ) or only with [ $^{18}\text{F}$ ]FES and [ $^{18}\text{F}$ ]FDHT PET ( $n = 51$ ). All lesions were scored visually and quantitatively by two independent observers. A visually PET-positive lesion was defined as uptake above background. For quantification, we used standardised uptake values (SUV):  $\text{SUV}_{\text{max}}$ ,  $\text{SUV}_{\text{peak}}$  and  $\text{SUV}_{\text{mean}}$ .

**Results:** Visual analysis showed an absolute positive and negative interobserver agreement for [ $^{18}\text{F}$ ]FES PET of 84% and 83%, respectively ( $\text{kappa} = 0.67$ , 95% CI 0.48–0.87), and 49% and 74% for [ $^{18}\text{F}$ ]FDHT PET, respectively ( $\text{kappa} = 0.23$ , 95% CI – 0.04–0.49). Intraclass correlation coefficients (ICC) for quantification of  $\text{SUV}_{\text{max}}$ ,  $\text{SUV}_{\text{peak}}$  and  $\text{SUV}_{\text{mean}}$  were 0.98 (95% CI 0.96–0.98), 0.97 (95% CI 0.96–0.98) and 0.89 (95% CI 0.83–0.92) for [ $^{18}\text{F}$ ]FES, and 0.78 (95% CI 0.66–0.85), 0.76 (95% CI 0.63–0.84) and 0.75 (95% CI 0.62–0.84) for [ $^{18}\text{F}$ ]FDHT, respectively.

(Continued on next page)

\* Correspondence: [c.menke@amsterdamumc.nl](mailto:c.menke@amsterdamumc.nl)

Lemonitsa H. Mammatas and Clasina M. Venema are co-leading authors.  
<sup>1</sup>Department of Medical Oncology, Cancer Center Amsterdam, Amsterdam UMC, VUmc University Medical Center Amsterdam, de Boelelaan 1117, 1081, HV, Amsterdam, The Netherlands  
Full list of author information is available at the end of the article

(Continued from previous page)

**Conclusion:** Visual and quantitative evaluation of [ $^{18}\text{F}$ ]FES PET showed high interobserver agreement. These results support the use of [ $^{18}\text{F}$ ]FES PET in clinical practice. In contrast, visual agreement for [ $^{18}\text{F}$ ]FDHT PET was relatively low due to low tumour-background ratios, but quantitative agreement was good. This underscores the relevance of quantitative analysis of [ $^{18}\text{F}$ ]FDHT PET in breast cancer.

**Trial registration:** [ClinicalTrials.gov](https://clinicaltrials.gov), NCT01988324. Registered 20 November 2013, <https://clinicaltrials.gov/ct2/show/NCT01988324?term=FDHT+PET&draw=1&rank=2>.

**Keywords:** FES PET, FDHT PET, Breast cancer, Oestrogen receptor, Androgen receptor, Interobserver variability

## Introduction

Breast cancer is the most common malignancy in women in the Western world. The majority of breast tumours express the oestrogen receptor (ER), which is the main indicator of potential response to anti-oestrogen therapies [1, 2]. Therefore, it is mandatory to determine ER expression in breast cancer. Recently, the androgen receptor (AR) emerged as a possible target for breast cancer therapy. The AR is present in 70–80% of patients with breast cancer, and AR antagonists are under investigation in clinical trials [3–6].

A tumour biopsy is the gold standard to determine receptor expression. However, this is an invasive procedure, is not always feasible in case of inaccessible tumour sites, and is subject to sampling errors [7]. The  $16\alpha$ -[ $^{18}\text{F}$ ]fluoro- $17\beta$ -oestradiol ([ $^{18}\text{F}$ ]FES) and  $16\beta$ -[ $^{18}\text{F}$ ]fluoro- $5\alpha$ -dihydrotestosterone ([ $^{18}\text{F}$ ]FDHT) PET/CT have been developed to non-invasively visualise, respectively, the ER and AR status in the tumour lesions within a patient. Previously, it has been shown that [ $^{18}\text{F}$ ]FES and [ $^{18}\text{F}$ ]FDHT uptake correlate well with ER and AR expression levels in representative breast cancer biopsies [8–10]. As a diagnostic tool, [ $^{18}\text{F}$ ]FES PET leads to better diagnostic understanding in 88% and to a change of therapy in 48% of the patients presenting with a clinical dilemma [11]. To predict treatment effects, [ $^{18}\text{F}$ ]FES PET can be used to assess residual ER availability during treatment with, e.g. fulvestrant, a selective ER downregulator. Inadequate reduction of the [ $^{18}\text{F}$ ]FES PET signal (< 75%) by fulvestrant treatment was associated with early progression [12]. Similarly, in patients with prostate cancer, [ $^{18}\text{F}$ ]FDHT PET was used to determine the optimal dose of the AR blocker enzalutamide in a phase 1 trial [13]. Lastly, patients with ER-positive breast cancer and high [ $^{18}\text{F}$ ]FDG uptake showed a worse progression free survival if [ $^{18}\text{F}$ ]FES uptake was low in comparison to high [ $^{18}\text{F}$ ]FES uptake (3 versus 8 months, respectively) [14].

For all these potential applications, reliable, observer-independent identification and quantification of [ $^{18}\text{F}$ ]FES and [ $^{18}\text{F}$ ]FDHT uptake in tumour lesions is essential for translation to daily clinical practice. Up till now, there are no data on the interobserver variability of [ $^{18}\text{F}$ ]FES and [ $^{18}\text{F}$ ]FDHT PET in breast cancer. Therefore, the

primary objective of this study was to examine interobserver variability in visual and quantitative assessment of [ $^{18}\text{F}$ ]FES and [ $^{18}\text{F}$ ]FDHT PET. Secondary objectives included the effect of tumour to background ratio (TBR), tracer accumulation, tumour size and the use of different SUV parameters ( $\text{SUV}_{\text{max}}$ ,  $\text{SUV}_{\text{peak}}$  or  $\text{SUV}_{\text{mean}}$ ) on interobserver agreement. Also, the added value of quantitative assessment in comparison to visual assessment was examined, and the number of lesions detected on [ $^{18}\text{F}$ ]FES and [ $^{18}\text{F}$ ]FDHT was compared with those detected on conventional imaging methods (contrast enhanced CT scan and bone scan).

## Materials and methods

### Patient population

This prospective two-centre interobserver variability study was part of a study investigating the correlation between [ $^{18}\text{F}$ ]FES and [ $^{18}\text{F}$ ]FDHT uptake and ER and AR expression in simultaneously biopsied metastases, of which the results have been published elsewhere [8]. Patients were recruited from September 2014 to August 2015 at the CCA-VUmc University Medical Center Amsterdam and the University Medical Center Groningen in the Netherlands.

Eligibility criteria included metastatic breast cancer and an ER-positive primary tumour,  $\geq 1$  extrahepatic tumour lesion, ECOG performance status of  $\leq 2$  and a postmenopausal status or use of LHRH-agonists. Patients were excluded if they had used ER or AR binding drugs during the 6 weeks before study entry, because these ligands compete with tracer binding.

All patients had to give written informed consent before study participation. The study was conducted in compliance with the ethical principles originating in or derived from the Declaration of Helsinki and in compliance with all International Conference on Harmonization Good Clinical Practice guidelines. The local medical ethics committee approved the study (NCT01988324).

### Imaging protocols

[ $^{18}\text{F}$ ]FES and [ $^{18}\text{F}$ ]FDHT were produced as described previously [15, 16]. On separate days,  $\leq 14$  days apart, 200 MBq ( $\pm 10\%$ ) of each tracer was injected. After 60

min ( $\pm 5$  min), a low-dose CT was performed during tidal breathing for attenuation correction, followed by a whole-body PET scan (skull vertex to mid-thigh, 2 min per bed position). PET/CT scans were made using a Philips Gemini TF-64 PET/CT (Amsterdam) or Siemens 64 slice mCT PET/CT (Groningen). Acquisition and reconstruction protocols used on both scanners were according to the recommendations of the European Association of Nuclear Medicine (EARL) [17].

In addition, a high-resolution, contrast-enhanced CT chest-abdomen and bone scan was performed within 6 weeks of the PET scans for comparison.

### Image analyses

Contrast enhanced CT scans were examined by experienced radiologists and bone scans by experienced nuclear medicine physicians, respectively, masked for the [ $^{18}\text{F}$ ]FES and [ $^{18}\text{F}$ ]FDHT PET results. Two independent observers from each centre (LM and CV), trained and supervised by two experienced nuclear medicine physicians, performed visual and quantitative analyses. The observers had knowledge of conventional imaging results (contrast enhanced CT and bone scans).

A visually PET-positive lesion was defined as focal uptake above local background incompatible with physiological uptake. Liver metastases were excluded from all analyses in this study because of high physiological [ $^{18}\text{F}$ ]FES and [ $^{18}\text{F}$ ]FDHT uptake in healthy liver tissue, making reliable identification of metastases difficult. In addition, if visual interpretation of uptake in a (potential) lesion was impossible, e.g. due to overlap with adjacent organs with high physiological tracer, the readers independently reported it as 'not evaluable' in the visual ratings, and these were excluded from further analyses. For each patient, the observers made a list that consisted of all lesions already detected on conventional imaging, followed by additional lesions discovered on [ $^{18}\text{F}$ ]FES or [ $^{18}\text{F}$ ]FDHT PET. An anatomical description of all the lesions was reported in order to match the results. In case a lesion was not reported by one of the two observers, it was scored as not visible for that observer. All visually PET-positive lesions were quantified, as well as PET-negative lesions that were identified on conventional imaging (i.e. lesions on bone scintigraphy and/or high resolution CT  $> 1$  cm).

Each observer manually drew volumes of interest (VOI) on the tumour contours, using PET images for PET-positive lesions and low-dose CT images for PET-negative lesions (lesions only seen on bone scan or high-resolution CT were visually matched on the low-dose CT). Lesions were separately analysed based on visibility on either PET or conventional imaging alone to investigate the influence of visibility on imaging techniques on interobserver agreement.

For every VOI, the standardised uptake values (SUV), i.e. the tracer uptake within a VOI normalised to the injected dose and body weight, were calculated using the software programs accurate (in-house build using IDL, observer 1) and syngo.via version VB10B, Siemens (observer 2). Both programs yielded identical results on test images. Three types of SUV were compared in this study:  $\text{SUV}_{\text{max}}$  (voxel with highest SUV within the VOI),  $\text{SUV}_{\text{peak}}$  (average SUV of a  $1 \text{ cm}^3$  sphere containing the hottest voxels of the VOI) and  $\text{SUV}_{\text{mean}}$  with isocontour 50% of  $\text{SUV}_{\text{max}}$  (average SUV of all voxels with uptake  $\geq 50\%$  of  $\text{SUV}_{\text{max}}$ ).

Based on previous studies, an  $\text{SUV}_{\text{max}}$  [ $^{18}\text{F}$ ]FES cut-off  $\geq 1.5$  was used to define ER-positivity (corresponding with a IHC cut-off of  $\geq 1\%$ ) and an  $\text{SUV}_{\text{max}}$  [ $^{18}\text{F}$ ]FDHT cut-off  $\geq 1.9$  for AR positivity (corresponding with a IHC cut-off of  $\geq 10\%$ ) [8, 9].

For [ $^{18}\text{F}$ ]FES and [ $^{18}\text{F}$ ]FDHT, the  $\text{SUV}_{\text{max}}$  tumour-background ratio (TBR) was defined as the ratio of the  $\text{SUV}_{\text{max}}$  of a tumour lesion and the  $\text{SUV}_{\text{mean}}$  of healthy background tissue. To determine the  $\text{SUV}_{\text{mean}}$  of healthy background tissue, a VOI was drawn on reference tissue in the unaffected contralateral site whenever available or in the unaffected surrounding tissue of the same origin [18].

### Statistical analyses

For visual assessments, agreement was calculated with absolute and relative measures of interobserver agreement. Absolute agreement is the probability that if one observer would score a lesion as visible (positive agreement) or not visible (negative agreement) on the PET scan, the other observer would do the same. It is calculated by the following formulas: positive agreement =  $2 \times \text{lesions visible to both observers} / (2 \times \text{lesions visible to both observers} + \text{lesions only visible to observer 1} + \text{lesions only visible to observer 2})$  and negative agreement =  $2 \times \text{lesions not visible to both observers} / (2 \times \text{lesions not visible to both observers} + \text{lesions only not visible to observer 1} + \text{lesions only not visible to observer 2})$  [19]. In order to compare results with previous studies, also reliability (relative agreement) was calculated according to Cohen's kappa, and the results were interpreted as follows: kappa 0.01–0.20 as slight, 0.21–0.40 as fair, 0.41–0.60 as moderate, 0.61–0.80 as substantial and 0.81–1.00 as almost perfect interobserver agreement [20]. To account for potential within-person correlation in visual assessments, a chi-square test was performed to examine whether the percentage visual agreement differed per patient.

For quantitative assessments, parameters are presented as mean  $\pm$  SD, and reliability was calculated with intraclass correlation coefficients (ICC) using a two-way random effect model with absolute agreement. For the

interpretation of the ICCs, the following guideline was used:  $\geq 0.90$  as excellent,  $\geq 0.75$  as good,  $\geq 0.50$  as moderate and  $< 0.50$  as poor [21].

Absolute agreement on quantitative assessments were analysed with Bland-Altman plots (differences between observers showed a normal distribution). For each lesion, it graphically shows the average SUV of observers 1 and 2 on the  $x$ -axes and on the  $y$ -axes the difference between observers for each lesion, expressed as percentage of the average SUV value. Percentage differences were used instead of absolute differences to achieve independence of magnitude of differences from magnitude of SUV values, and it facilitates comparisons between the SUV parameters  $SUV_{max}$ ,  $SUV_{mean}$  and  $SUV_{peak}$ , which may show large differences in absolute values.

To investigate the effect of TBRs on interobserver variability, differences between TBRs of [ $^{18}F$ ]FES and [ $^{18}F$ ]FDHT PETs were tested with Wilcoxon matched pairs signed rank tests. In addition, correlations between tracer uptake or tumour size and percentage interobserver differences were determined using the Spearman correlation coefficient ( $r$ ). Finally, linear regression was performed to find the linear function between  $SUV_{max}$ ,  $SUV_{peak}$  and  $SUV_{mean}$  for [ $^{18}F$ ]FES and [ $^{18}F$ ]FDHT PET, and Cochran's Q and McNemar tests were used to analyse differences between visibility and quantitative uptake above or below cut-off for  $SUV_{max}$ ,  $SUV_{peak}$  and

$SUV_{mean}$ .  $P$  value  $< 0.05$  was considered significant. Statistical analyses were generated using the SPSS software (version 22; IBM, SPSS statistics).

## Results

### Patient characteristics

A total of 120 lesions were identified in 10 patients using the different imaging modalities (Table 1). Most lesions were skeletal (66%), followed by lymph node (25%) and visceral metastases (9%). The median number of lesions per patient was 9 (range 2–32).

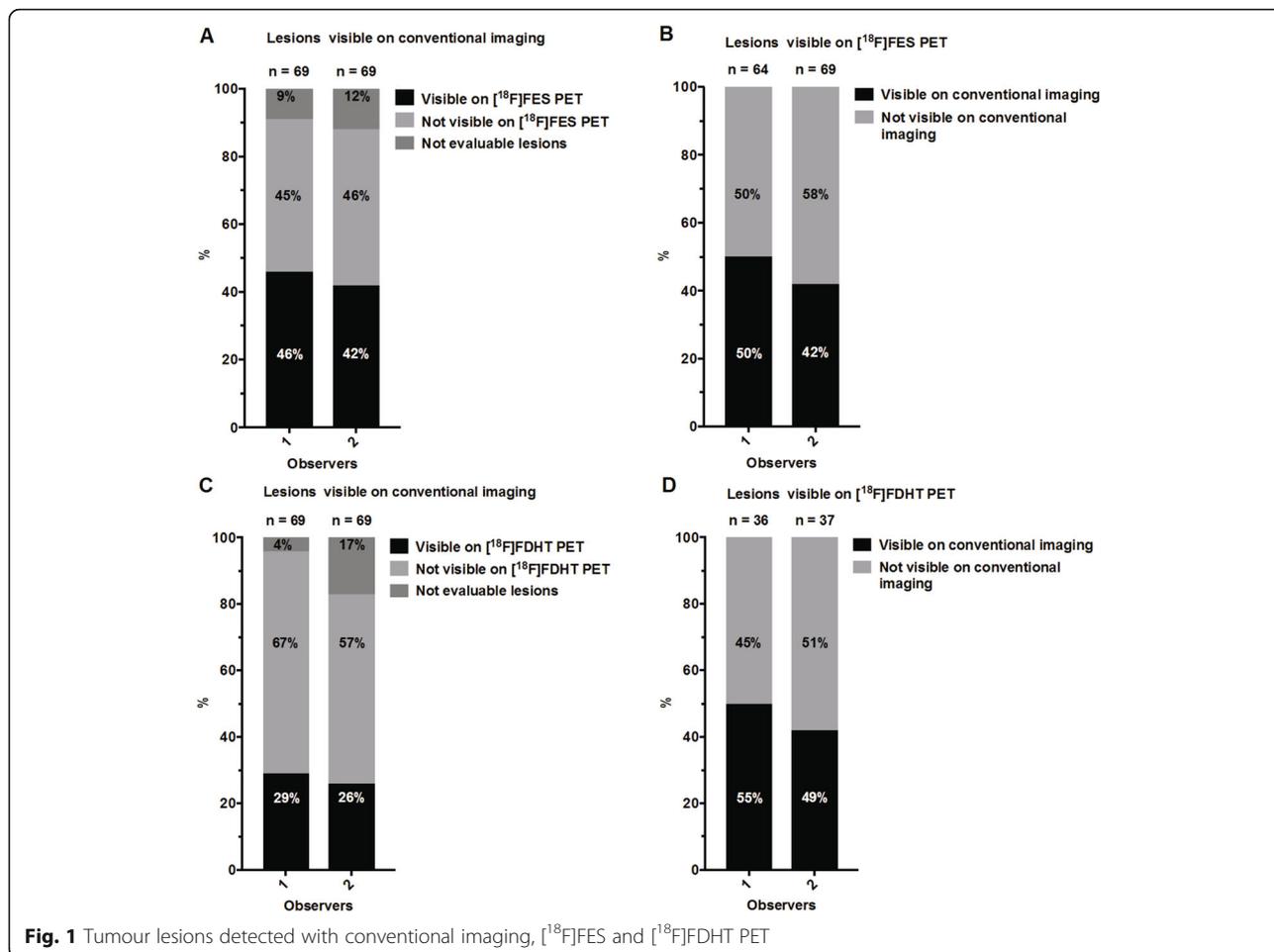
### Comparison of lesion detection on different imaging modalities

Of the 120 lesions in total (Table 1), most were identified on [ $^{18}F$ ]FES PET ( $n = 64$  [53%] and  $n = 69$  [58%] by observer 1 and 2, respectively), followed by high-resolution CT ( $n = 54$  [45%]), bone scintigraphy ( $n = 40$  [33%]) and [ $^{18}F$ ]FDHT PET ( $n = 36$  [30%] and  $n = 37$  [31%]). Fifty and 42% of the lesions identified on [ $^{18}F$ ]FES PET by observer 1 and 2, respectively, were also detected on high resolution CT or bone scintigraphy (Fig. 1). For [ $^{18}F$ ]FDHT PET, 55% and 49% of the identified lesions were seen with conventional imaging. Conversely, 46 and 42% of the lesions identified on conventional imaging were visible on [ $^{18}F$ ]FES PET by, respectively, observer 1 and 2, and 29% and 26% were seen on [ $^{18}F$ ]FDHT PET. In particular, more lymph node

**Table 1** Patient characteristics

Characteristic	Number ( $n = 10$ )	%
Age in years, mean (range)	67 (48–79)	
Biopsy of primary tumour		
ER+/AR+	10	100
Biopsy of metastases		
ER+/AR+	8	80
ER+/AR-	1	10
ER-/AR-	1	10
Previous treatment lines		
0–1	3	30
2–4	7	70
Visible lesions: total, median per patient (range)	120, 9 (2–32)	
Conventional imaging (CT, bone scan)	69 (54, 40)	58 (45, 33)
Visible on PET alone ([ $^{18}F$ ]FES or [ $^{18}F$ ]FDHT PET)	51 (33, 20)	42 (28, 16)
Total visible on [ $^{18}F$ ]FES PET (observer 1, 2)	64, 69	53, 58
Total visible on [ $^{18}F$ ]FDHT PET (observer 1, 2)	36, 37	30, 31
Location		
Bone (conventional imaging, [ $^{18}F$ ]FES, [ $^{18}F$ ]FDHT PET)	79 (55, 45, 37)	66 (80, 54, 64)
Lymph node (conventional imaging, [ $^{18}F$ ]FES; [ $^{18}F$ ]FDHT PET)	30 (8, 29, 16)	25 (12, 35, 28)
Visceral <sup>a</sup> (conventional imaging, [ $^{18}F$ ]FES, [ $^{18}F$ ]FDHT PET)	11 (6, 9, 5)	9 (9, 11, 9)

<sup>a</sup>Excluding liver



lesions were detected on  $[^{18}\text{F}]\text{FES}$  PET and  $[^{18}\text{F}]\text{FDHT}$  PET compared to conventional imaging: 97% and 53% versus 27% of all detected lymph node lesions, respectively.

**Visual analysis of  $[^{18}\text{F}]\text{FES}$  and  $[^{18}\text{F}]\text{FDHT}$  PET images**

Out of 120 lesions, a total of 87 and 74 on  $[^{18}\text{F}]\text{FES}$  and  $[^{18}\text{F}]\text{FDHT}$  PET, respectively, were analysed for visual interobserver agreement. The other lesions were excluded because one or both observers reported these as ‘not evaluable’ due to overlap with adjacent organs with high physiological tracer uptake.

For lesions *visible on conventional imaging*,  $[^{18}\text{F}]\text{FES}$  PET readings (Table 2) had substantial positive and negative agreement of 84% (95% CI 72–92%) and 83% (95% CI 70–91%), respectively (kappa = 0.67, 95% CI 0.48–0.87). By including lesions that were only visible on  $[^{18}\text{F}]\text{FES}$  PET, the positive agreement improved to 88% (95% CI 80–93%) for *all* lesions scored on  $[^{18}\text{F}]\text{FES}$  PET (negative agreement remained the same).  $[^{18}\text{F}]\text{FDHT}$  PET showed lower positive agreement of 49% (95% CI 32–65%) for lesions *visible on*

*conventional imaging*, while negative agreement was 74% (95% CI 62–83%) (kappa = 0.23, 95% CI – 0.04–0.49). Positive agreement for *all* lesions scored on  $[^{18}\text{F}]\text{FDHT}$  PET was 58% (95% CI 43–71%). By looking at lesions *only visible on PET* and not on conventional imaging, the positive agreement rate was the highest: 91% (95% CI 81–96%) for  $[^{18}\text{F}]\text{FES}$  PET and 80% (95% CI 55–93%) for  $[^{18}\text{F}]\text{FDHT}$  PET. Visual interobserver agreement was not significantly different between the 10 different patients in this study:  $P = 0.159$  for  $[^{18}\text{F}]\text{FES}$  PET and  $P = 0.387$  for  $[^{18}\text{F}]\text{FDHT}$  PET.

An important aspect in the identification of tumour lesions is how well tracer uptake can be distinguished from background uptake in normal reference tissue. The TBR of  $[^{18}\text{F}]\text{FDHT}$  was significantly lower than that of  $[^{18}\text{F}]\text{FES}$  (Fig. 2). In bone lesions, the mean TBR of  $[^{18}\text{F}]\text{FDHT}$  was 2.0 ( $\pm$  SD 0.6) versus 3.3 ( $\pm$  SD 2.2) for  $[^{18}\text{F}]\text{FES}$  ( $P = 0.003$ ). In addition, in lymph node lesions, the mean  $[^{18}\text{F}]\text{FDHT}$  TBR was 4.6 ( $\pm$  SD 1.9) compared to 10.7 ( $\pm$  SD 8.4) for  $[^{18}\text{F}]\text{FES}$  ( $P < 0.0001$ ).

**Table 2** Visual interobserver agreement for lesions visible (A, C) and not visible on conventional imaging (B, D) on [<sup>18</sup>F]FES and [<sup>18</sup>F]FDHT PET, respectively

<b>A Visual interobserver agreement on [<sup>18</sup>F]FES PET for lesions visible on conventional imaging</b>				
	Observer 1			
Observer 2	Visible	Not visible	Not evaluable <sup>a</sup>	Total
Visible	24	3	2	29
Not visible	6	22	4	32
Not evaluable <sup>a</sup>	2	6	0	8
Total	32	31	6	69

<b>B Visual interobserver agreement on [<sup>18</sup>F]FES PET for lesions not visible on conventional imaging</b>				
	Observer 1			
Observer 2	Visible	Not visible	Not evaluable <sup>a</sup>	Total
Visible	26	3	11	40
Not visible	2	1 <sup>b</sup>	3	6
Not evaluable <sup>a</sup>	4	1	0	5
Total	32	5	14	51

<b>C Visual interobserver agreement on [<sup>18</sup>F]FDHT PET for lesions visible on conventional imaging</b>				
	Observer 1			
Observer 2	Visible	Not visible	Not evaluable <sup>a</sup>	Total
Visible	9	8	1	18
Not visible	11	27	1	39
Not evaluable <sup>a</sup>	0	11	1	12
Total	20	46	3	69

<b>D Visual interobserver agreement on [<sup>18</sup>F]FDHT PET for lesions not visible on conventional imaging</b>				
	Observer 1			
Observer 2	Visible	Not visible	Not evaluable <sup>a</sup>	Total
Visible	6	2	11	19
Not visible	1	10 <sup>c</sup>	6	17
Not evaluable <sup>a</sup>	9	4	2	15
Total	16	16	19	51

<sup>a</sup>Not evaluable lesions due to overlap with adjacent organs with high physiological tracer uptake

<sup>b</sup>Lesions identified on [<sup>18</sup>F]FDHT PET

<sup>c</sup>Lesions identified on [<sup>18</sup>F]FES PET

### Quantitative analyses of [<sup>18</sup>F]FES and [<sup>18</sup>F]FDHT PET images

Out of 120 lesions, a total of 94 and 95 were quantified by both observers on [<sup>18</sup>F]FES and [<sup>18</sup>F]FDHT PET, respectively. The other lesions were not quantified by one or both of the observers as a result of overlap with adjacent organs with high physiological tracer uptake, unless there was a clear anatomical substrate on other imaging modalities allowing for reliable VOI definition.

In general, interobserver agreement was excellent for PET quantification (Fig. 3) of *all* lesions combined (i.e.

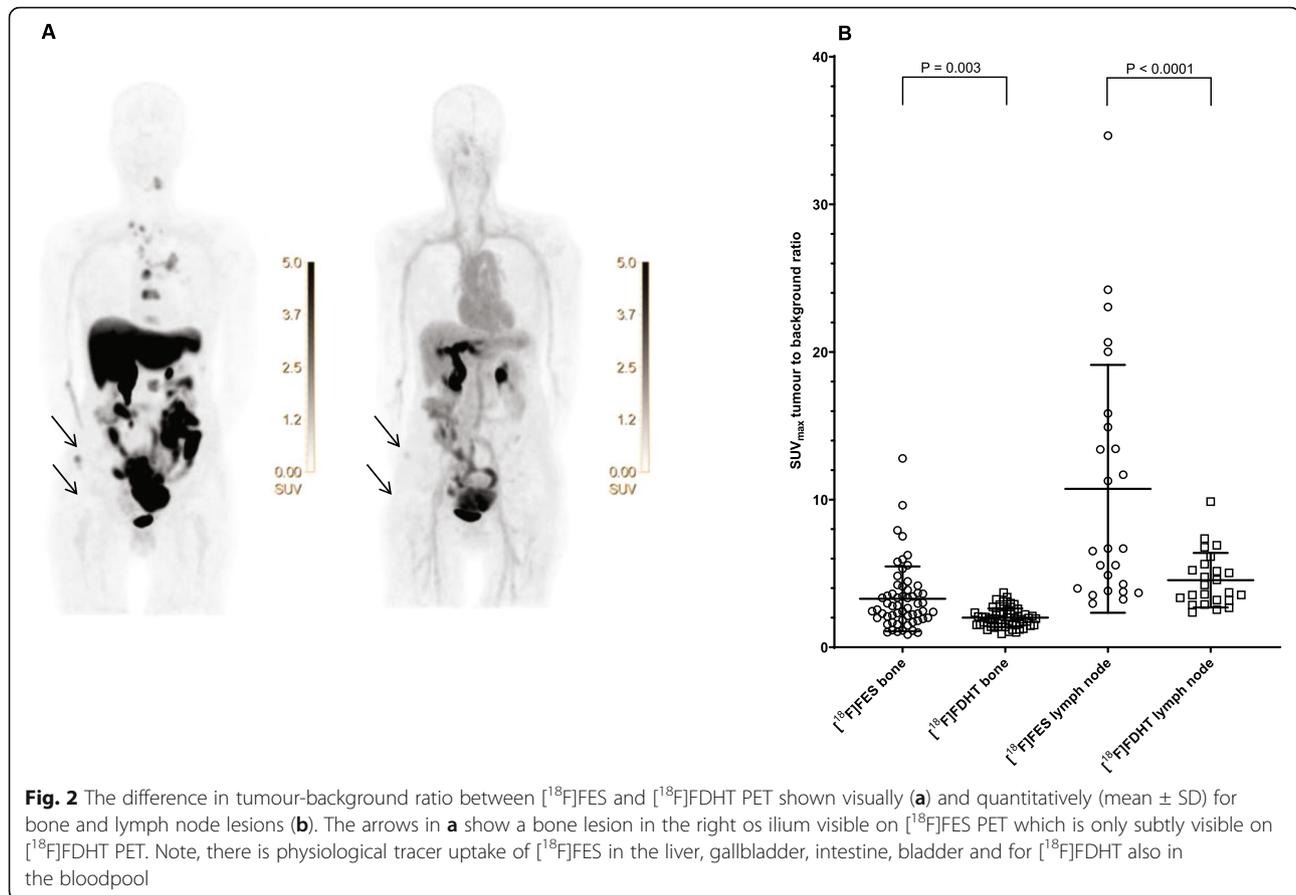
visible on PET or seen on conventional imaging). The ICCs for quantification of  $SUV_{max}$ ,  $SUV_{peak}$  and  $SUV_{mean}$  on [<sup>18</sup>F]FES PET were 0.98 (95% CI 0.96–0.98), 0.97 (95% CI 0.96–0.98) and 0.89 (95% CI 0.83–0.92). For [<sup>18</sup>F]FDHT PET, the ICCs were lower with 0.78 (95% CI 0.66–0.85), 0.76 (95% CI 0.63–0.84) and 0.75 (95% CI 0.62–0.84), respectively.

In addition, [<sup>18</sup>F]FES (Fig. 4) and [<sup>18</sup>F]FDHT PET (Fig. 5) quantification was analysed separately with Bland Altman plots for all lesions *visible on PET* or lesions *only visible on conventional imaging* (hence, PET-negative lesions). For [<sup>18</sup>F]FES PET, PET-positive lesions showed excellent quantitative interobserver agreement with mean differences < 2% and 95% limits of agreement ( $LOA_{95\%}$ ) being narrower for  $SUV_{max}$  ( $LOA_{95\%}$  – 31.3 to 34.3%) and  $SUV_{peak}$  ( $LOA_{95\%}$  – 31.1 to 28.4%), compared to  $SUV_{mean}$  ( $LOA_{95\%}$  – 46.5 to 44.3%). More differences were shown for PET-negative lesions with mean interobserver differences < 14% and larger  $LOA_{95\%}$  (within  $\pm$  75%), but note that absolute differences between observers were generally low due to a low SUV. Similarly, for [<sup>18</sup>F]FDHT PET, interobserver agreement was better for PET-positive (mean interobserver differences < 7%,  $LOA_{95\%}$  within  $\pm$  45%) compared to PET-negative lesions (mean interobserver differences < 12%,  $LOA_{95\%}$  within  $\pm$  76%).  $SUV_{max}$  and  $SUV_{peak}$  showed a better interobserver agreement in comparison to  $SUV_{mean}$  for the quantification of lesions visible on [<sup>18</sup>F]FES PET, while on [<sup>18</sup>F]FDHT PET the different SUV parameters were comparable.

Higher levels of tracer accumulation in PET positive lesions were not associated with improved interobserver agreement (for [<sup>18</sup>F]FES PET: Spearman  $r$  = 0.04, 0.26 and 0.14 for  $SUV_{max}$ ,  $SUV_{peak}$  and  $SUV_{mean}$ , respectively and for [<sup>18</sup>F]FDHT PET: Spearman  $r$  = 0.00,  $r$  = 0.03 and  $r$  = – 0.17, respectively). In addition, there was no correlation between tumour size and interobserver agreement (for [<sup>18</sup>F]FES PET: Spearman  $r$  = 0.10,  $r$  = 0.08 and  $r$  = 0.06, for  $SUV_{max}$ ,  $SUV_{peak}$  and  $SUV_{mean}$ , respectively and for [<sup>18</sup>F]FDHT PET: Spearman  $r$  = – 0.07,  $r$  = – 0.16 and  $r$  = – 0.42, respectively).

### The added value of quantitative assessment in comparison to visual assessment

Based on previous studies, [<sup>18</sup>F]FES and [<sup>18</sup>F]FDHT  $SUV_{max}$  cut-off levels of 1.5 and 1.9, respectively, have been identified. There are however limited data on quantitative thresholds and corresponding cut-off values for  $SUV_{peak}$  and  $SUV_{mean}$ . Based on linear regression of all lesions quantified in this study, an  $SUV_{max}$  cut-off of 1.5 on [<sup>18</sup>F]FES PET corresponded with an  $SUV_{peak}$  of 1.2 and an  $SUV_{mean}$  of 1.1 (Supplementary figure S1), and for [<sup>18</sup>F]FDHT PET, an  $SUV_{max}$  cut-off of 1.9 corresponded with an  $SUV_{peak}$  of 1.6 and an  $SUV_{mean}$  of 1.3.

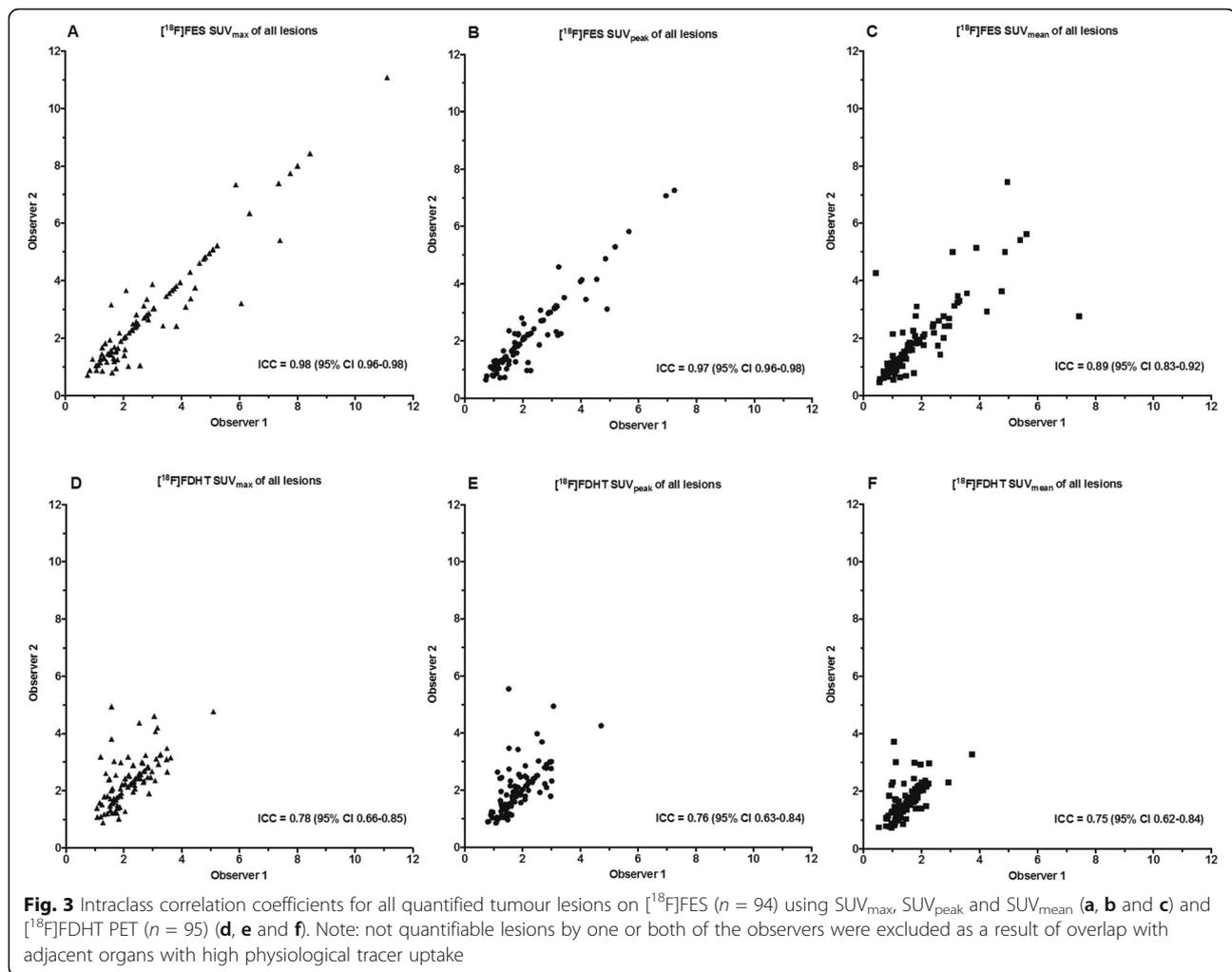


For diagnostic purposes, it is important to identify all receptor positive tumour lesions. Therefore, we compared visual and quantitative tracer uptake above/below cut-off levels (Table 3). In 3% and 1% of the lesions scored visually positive on  $[^{18}\text{F}]\text{FES}$  PET by observer 1 and 2 respectively,  $\text{SUV}_{\text{max}}$  was below the threshold of 1.5. For  $[^{18}\text{F}]\text{FDHT}$  PET, 14% of the visually positive lesions scored by observer 1 as well as observer 2 had an  $\text{SUV}_{\text{max}}$  below the threshold of 1.9. There were no structural differences between observer 1 and 2. The discrepancies were mostly seen in lesions located in tissue with low background uptake such as skin and lung metastases (Supplementary table S1). Conversely, in 44% and 39% of the lesions scored visually negative on  $[^{18}\text{F}]\text{FES}$  PET by observer 1 and 2, respectively,  $\text{SUV}_{\text{max}}$  was  $\geq 1.5$ . Similarly, 31% and 52% of the visually negative lesions had an  $\text{SUV}_{\text{max}} \geq 1.9$  on  $[^{18}\text{F}]\text{FDHT}$  PET, respectively. However, in most cases (60%), we observed overlap with organs having high physiological tracer accumulation such as the liver and bowel, followed by lesions that were determined to be visually positive at second glance (32%). After correction for these effects,  $\leq 4\%$  of the visually negative lesions had a  $\text{SUV}_{\text{max}}$  above cut-off for both tracers.

Comparing the impact of the different SUV parameters on discrepancies between visual and quantitative assessments showed no significant differences with the only exception that  $\text{SUV}_{\text{mean}}$  showed less visually negative lesions above cut-off on  $[^{18}\text{F}]\text{FES}$  PET than  $\text{SUV}_{\text{max}}$  or  $\text{SUV}_{\text{peak}}$  for observer 1 ( $P = 0.008$  and  $P = 0.001$ , respectively), but not for observer 2 ( $P = 0.125$  and  $P = 0.063$ , respectively).

## Discussion

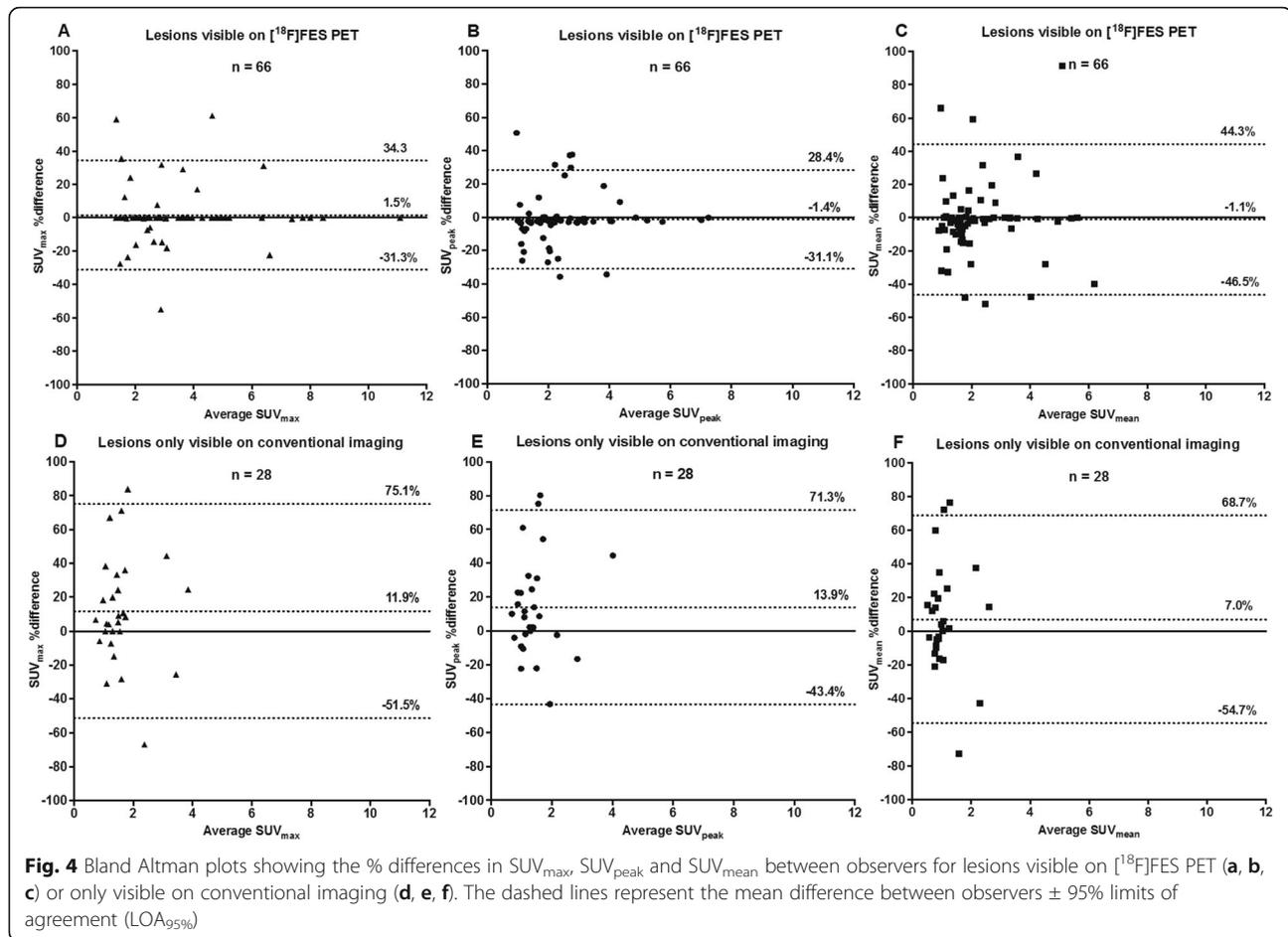
Interobserver variability is an important step in the clinical application of diagnostic tools. Here, we showed that both visual and quantitative evaluation were highly reproducible between independent observers evaluating  $[^{18}\text{F}]\text{FES}$  PET at separate centres using different scanners and software. Visual positive and negative absolute agreement was  $> 80\%$ , with a kappa of 0.67. Also, the interobserver reliability of quantitative metrics was excellent for  $\text{SUV}_{\text{max}}$  and  $\text{SUV}_{\text{peak}}$  (ICC of 0.98 and 0.97, respectively) and good for  $\text{SUV}_{\text{mean}}$  (ICC of 0.89). In comparison, staging patients with breast cancer showed similar results for bone scintigraphy (kappa 0.62–0.78) and  $[^{18}\text{F}]\text{FDG}$  PET (kappa 0.65 and an ICC of 0.93 for the quantification of  $[^{18}\text{F}]\text{FDG}$  uptake) [22–26].



$[^{18}\text{F}]\text{FDHT}$  PET also showed good interobserver reliability for quantitative assessments with ICCs  $\geq 0.75$ . These values are slightly lower than those of  $[^{18}\text{F}]\text{FES}$  PET, and this was probably due to the lower lesional  $[^{18}\text{F}]\text{FDHT}$  uptake, because quantitative agreement according to Bland Altman analyses were comparable for both tracers. The TBR of  $[^{18}\text{F}]\text{FDHT}$  was considerably lower compared to  $[^{18}\text{F}]\text{FES}$ . This probably explains the higher variability in visual interpretation ( $\kappa = 0.23$ ), mainly caused by a low visual positive agreement (49%) in lesions already identified by conventional imaging modalities, while positive agreement in lesions not identified by conventional imaging was much higher (80%), as well as negative visual agreement between observers (74%). An important impeding factor was the significantly lower TBR of  $[^{18}\text{F}]\text{FDHT}$  in bone and lymph node lesions compared to  $[^{18}\text{F}]\text{FES}$  PET. The TBR of  $[^{18}\text{F}]\text{FDHT}$  in the current study (2.0 for bone and 4.6 for lymph nodes) was also lower than in prostate cancer metastases (3.3 for bone and 5.7 for soft tissue metastases) with an  $\text{SUV}_{\text{max}}$  three times higher in prostate

cancer (7.1–9.1 versus 2.0 in the present breast cancer study) [27, 28]. This suggests that higher AR expression likely results in better interobserver reliability.

Our study had some limitations. There were only a limited number of patients included in this study. However, receptor expression between lesions within a single patient can be heterogeneous [29], which was confirmed in the present study resulting in the coverage of a large range of data in 120 lesions [8]. In addition, we showed there was no within-patient correlation in visual assessments. A second limitation is a substantial number of ‘not evaluable’ lesions, due to overlap with adjacent organs with high physiological background. The decision for evaluability was left to each observer individually, which may have contributed to the low agreement ( $\leq 6\%$ ) on these ‘not evaluable lesions’. For future studies, we recommend that all lesions with physiological background overlap from the liver, gallbladder, intestine, bladder and for  $[^{18}\text{F}]\text{FDHT}$  also from bloodpool are regarded as not evaluable. A third limitation is the lack of robust  $[^{18}\text{F}]\text{FES}$  and  $[^{18}\text{F}]\text{FDHT}$  thresholds for test

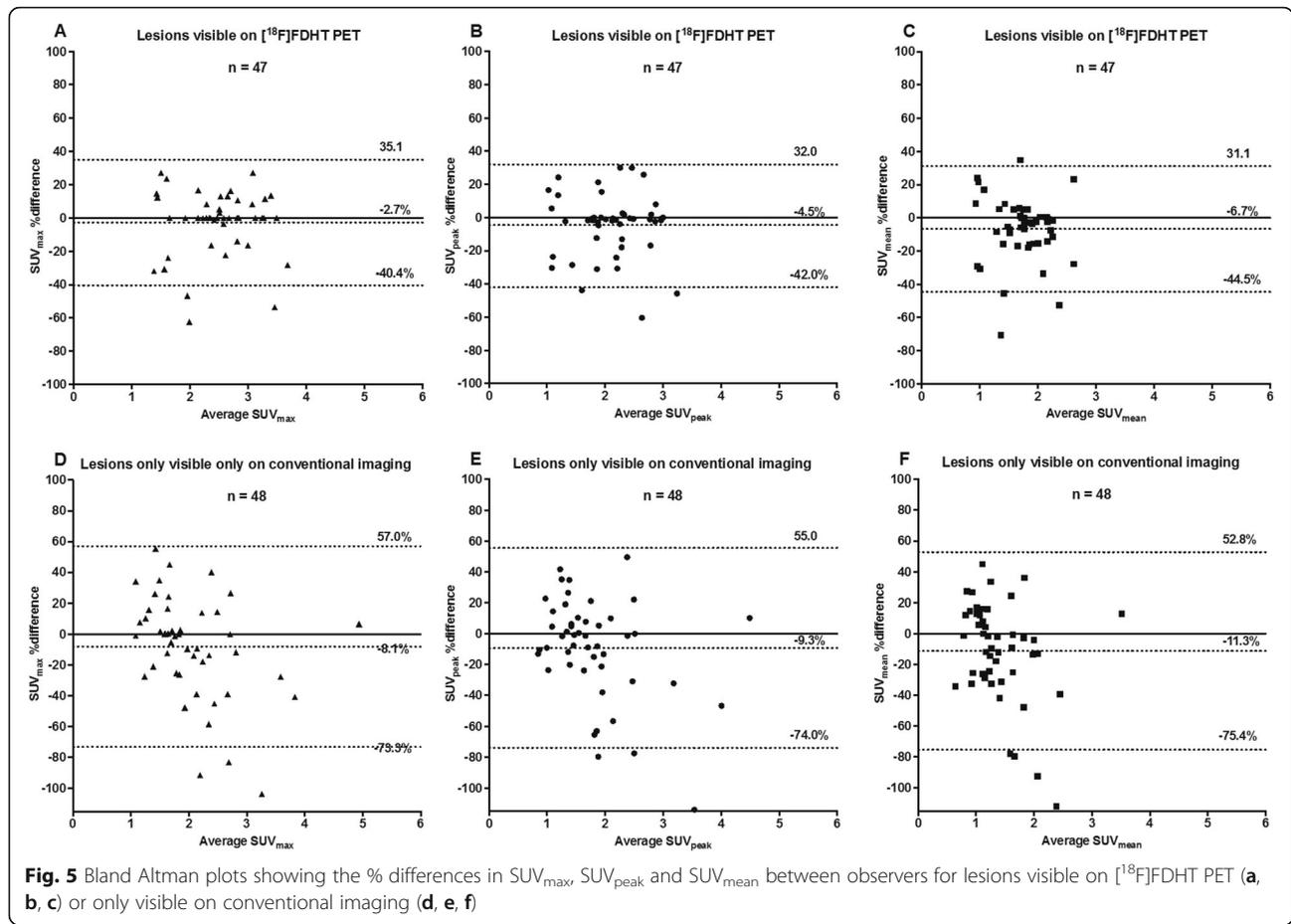


positivity. We used an  $\text{SUV}_{\text{max}}$  cut-off of 1.5 for  $[^{18}\text{F}]\text{FES}$  and 1.9 for  $[^{18}\text{F}]\text{FDHT}$  PET based on previous data corresponding with ER and AR positivity in biopsies and so far showing the best predictive value for response to endocrine therapy [8, 9, 30, 31]. Some studies suggested an  $\text{SUV}_{\text{max}}$  cut-off of 2.0 for  $[^{18}\text{F}]\text{FES}$  PET, taking into account the background  $[^{18}\text{F}]\text{FES}$  uptake in normal tissues which can exceed the cut-off of 1.5 [29–31]. Tissue specific cut-off values may indeed be more appropriate as there are responders to endocrine therapy with a tumour  $\text{SUV}_{\text{max}} < 2.0$ . In the current study, up to 20% of the visually positive lesions had an  $\text{SUV}_{\text{max}} < 2.0$ , while  $< 3\%$  had an  $\text{SUV}_{\text{max}} < 1.5$  (Supplementary table S2).

For diagnostic purposes, simple visual assessment of  $[^{18}\text{F}]\text{FES}$  uptake may suffice to determine the receptor status of a tumour lesion (agreement was high between visual assessment and the applied  $\text{SUV}_{\text{max}}$  cut-off value of 1.5 for ER-positivity). True discrepancies between visibility and corresponding uptake above or below cut-off were low ( $< 4\%$ ), making quantification of visually negative lesions not only cumbersome, but also unnecessary. Also, quantification of lesions without visual  $[^{18}\text{F}]\text{FES}$  uptake leads to higher interobserver variability due to

differences in VOI definition. However, quantification remains a helpful tool for nuclear medicine physicians in ‘equivocal  $[^{18}\text{F}]\text{FES}$  lesions’. In addition, quantification is useful to measure receptor availability over time for the evaluation of treatment effects. In contrast, quantification of  $[^{18}\text{F}]\text{FDHT}$  uptake is still required in future breast cancer studies, as we have shown relatively low visual agreement.

The role of  $[^{18}\text{F}]\text{FES}$  and  $[^{18}\text{F}]\text{FDHT}$  PET in addition to conventional imaging modalities needs to be defined further. It has to be taken into account that besides partial volume effects and constraints due to background tracer uptake limiting their detection, receptor expression can be heterogeneous and variable during the course of the disease [11, 32]. In addition, treatment may induce changes in receptor expression, but also eradicated tumour cells can leave a visible lesion on conventional imaging (e.g. sclerotic bone lesions), in absence of viable tumour cells. In the current study with heavily pretreated patients, 42–46% and 26–29% of the lesions identified by conventional imaging were detected on  $[^{18}\text{F}]\text{FES}$  and  $[^{18}\text{F}]\text{FDHT}$  PET, respectively. Vice versa, only approximately 50% of the lesions observed on



**Table 3** Discrepancies between visual and quantitative assessments (above/below cut-off values for receptor positivity) for  $[^{18}\text{F}]$ FES (A) and  $[^{18}\text{F}]$ FDHT PET (B)

A $[^{18}\text{F}]$ FES	Observer 1		Observer 2	
	Visible (n = 64)	Not visible (n = 36)	Visible (n = 69)	Not visible (n = 38)
$\text{SUV}_{\text{max}} \geq 1.5$	62 (97%)	16 (44%)	68 (99%)	15 (39%)
$\text{SUV}_{\text{max}} < 1.5$	2 (3%)	20 (56%)	1 (1%)	23 (61%)
$\text{SUV}_{\text{peak}} \geq 1.2$	54 (84%)	19 (53%)	67 (97%)	16 (42%)
$\text{SUV}_{\text{peak}} < 1.2$	10 (16%)	17 (47%)	2 (3%)	22 (58%)
$\text{SUV}_{\text{mean}} \geq 1.1$	57 (89%)	8 (22%)	67 (97%)	11 (29%)
$\text{SUV}_{\text{mean}} < 1.1$	7 (11%)	28 (78%)	2 (3%)	27 (71%)
B $[^{18}\text{F}]$ FDHT	Observer 1		Observer 2	
	Visible (n = 36)	Not visible (n = 62)	Visible (n = 37)	Not visible (n = 56)
$\text{SUV}_{\text{max}} \geq 1.9$	31 (86%)	19 (31%)	32 (86%)	29 (52%)
$\text{SUV}_{\text{max}} < 1.9$	5 (14%)	43 (69%)	5 (14%)	27 (48%)
$\text{SUV}_{\text{peak}} \geq 1.6$	30 (83%)	25 (40%)	33 (89%)	30 (54%)
$\text{SUV}_{\text{peak}} < 1.6$	6 (17%)	37 (60%)	4 (11%)	26 (46%)
$\text{SUV}_{\text{mean}} \geq 1.3$	31 (86%)	20 (32%)	33 (89%)	30 (54%)
$\text{SUV}_{\text{mean}} < 1.3$	5 (14%)	42 (68%)	4 (11%)	26 (46%)

Not evaluable lesions were excluded as reported in Table 2

[<sup>18</sup>F]FES PET and [<sup>18</sup>F]FDHT PET were identified by conventional imaging.

Therefore, a potential role for [<sup>18</sup>F]FES PET may be in staging of early ER-positive breast cancer as an addition to existing imaging techniques. Standard staging with [<sup>18</sup>F]FDG PET can miss low-intermediate grade ER-positive lesions due to their low metabolic activity [33]. We are currently investigating [<sup>18</sup>F]FES PET in staging patients with low grade, ER-positive locally advanced or recurrent breast cancer versus [<sup>18</sup>F]FDG PET (NCT03726931), and in metastatic breast cancer versus addition to conventional diagnostics (NCT01957332). The *non-invasive* visualisation of receptor status in metastatic lesions with PET offers a number potential clinical advantages. For example, in case conventional diagnostics cannot establish a final diagnosis of suspected metastatic breast cancer lesions (e.g. as a result of inaccessible biopsy sites or repeated biopsy sampling errors). Also, PET imaging may help to determine the hormone receptor status of *different tumour sites within a patient* and guide treatment decisions, for instance, to decide on the origin of a metastatic lesion in case of multiple primary tumours or to determine whether receptor conversion occurred in metastases from a single primary tumour [11]. If validated, this may help with multimodality treatment strategies for heterogeneous tumour sites of breast cancer, such as endocrine therapy for [<sup>18</sup>F]FES positive lesions combined with a local modality such as radiotherapy for concurrent [<sup>18</sup>F]FES negative lesions [34].

## Conclusion

In conclusion, our findings demonstrate that visual and quantitative evaluation of [<sup>18</sup>F]FES PET has a high inter-observer concordance and support the use in clinical practice. Although [<sup>18</sup>F]FDHT PET showed relatively low visual agreement, presumably a result of the low AR expression and consequently low TBR in patients with breast cancer, there was good quantitative agreement between observers, acceptable for further [<sup>18</sup>F]FDHT PET imaging studies in breast cancer.

## Abbreviations

ER: Oestrogen receptor; AR: Androgen receptor; [<sup>18</sup>F]FES: 16α-[<sup>18</sup>F]fluoro-17β-oestradiol; [<sup>18</sup>F]FDHT: 16β-[<sup>18</sup>F]fluoro-5α-dihydrotestosterone; VOI: Volume of interest; SUV: Standardised uptake value; TBR: Tumour-background ratio; ICC: Intraclass correlation coefficient; LOA<sub>95%</sub>: 95% limits of agreement

## Acknowledgements

We thank the patients who participated in this study and their families. In addition, we acknowledge the efforts of the clinical and imaging teams at the university medical centres participating in this study.

## Authors' contributions

CS, MvK, AG, HV, EB, ErdV, EldV, OH, GH and WM were responsible for the study concept and design. LM, CV, OH and AG carried out the data acquisition. MY, OH and AG performed the quality control of the data. HdV supervised the statistical analyses. BvdV performed the pathological analyses.

LM and CV prepared the manuscript. CS, MvK, AG, HV, EB, ErdV, EldV, OH, GH, WM, MY, HdV and BvdV edited and revised the manuscript. All authors read and approved of the final manuscript.

## Funding

This study was financially supported by the Centre for Translational Molecular Medicine (CTMM) project as part of the Mammary Carcinoma Molecular Imaging for Diagnosis and Therapeutics (MAMMOTH) project.

## Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

## Ethics approval and consent to participate

All procedures performed in studies involving human participants were in accordance with the ethical standards of the Institutional Review Board of University Medical Center, Groningen, and also approved by the Institutional Review Board of the Amsterdam UMC, location VUmc University Medical Center, Amsterdam, the Netherlands (File no. 2014.501-NL41954.042.13) and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. Informed consent was obtained from all individual participants included in the study.

## Consent for publication

Not applicable.

## Competing interests

The authors declare that they have no competing interests.

## Author details

<sup>1</sup>Department of Medical Oncology, Cancer Center Amsterdam, Amsterdam UMC, VUmc University Medical Center Amsterdam, de Boelelaan 1117, 1081, HV, Amsterdam, The Netherlands. <sup>2</sup>Department of Medical Oncology, University of Groningen, University Medical Center Groningen, Hanzeplein 1, 9713, GZ, Groningen, The Netherlands. <sup>3</sup>Department of Epidemiology and Biostatistics, Amsterdam Public Health Institute, Amsterdam UMC, VUmc University Medical Center Amsterdam, De Boelelaan 1105, 1081, HV, Groningen, The Netherlands. <sup>4</sup>Department of Nuclear Medicine and Molecular Imaging, University of Groningen, University Medical Center Groningen, Hanzeplein 1, 9713, GZ, Groningen, The Netherlands. <sup>5</sup>Department of Radiology and Nuclear Medicine, Cancer Center Amsterdam, Amsterdam UMC, VUmc University Medical Center Amsterdam, De Boelelaan 1117, 1081, HV, Amsterdam, The Netherlands. <sup>6</sup>Department of Pathology and Medical Biology, University of Groningen, University Medical Center Groningen, DHanzeplein 1, 9713, GZ, Groningen, The Netherlands.

Received: 19 November 2019 Accepted: 3 April 2020

Published online: 19 April 2020

## References

- Blamey RW, Hornmark-Stenstam B, Ball G, Blichert-Toft M, Cataliotti L, Fourquet A, et al. ONCOPOOL - a European database for 16,944 cases of breast cancer. *Eur J Cancer*. 2010;46:56–71. <https://doi.org/10.1016/j.ejca.2009.09.009>.
- Yamashita H, Yando Y, Nishio M, Zhang Z, Hamaguchi M, Mita K, et al. Immunohistochemical evaluation of hormone receptor status for predicting response to endocrine therapy in metastatic breast cancer. *Breast Cancer*. 2006;13:74–83.
- Collins LC, Cole KS, Marotti JD, Hu R, Schnitt SJ, Tamimi RM. Androgen receptor expression in breast cancer in relation to molecular phenotype: results from the Nurses' Health Study. *Mod Pathol*. 2011;24:924–31. <https://doi.org/10.1038/modpathol.2011.54>.
- Krop I, Colleoni M, Traina T, Holmes F, Estevez L, et al. Results from a randomized placebo-controlled phase 2 trial evaluating exemestane ± enzalutamide in patients with hormone receptor-positive breast cancer. Abstract GS4-07. San Antonio Breast Cancer Symposium. San Antonio, Texas; 2017.
- Gucalp A, Tolanev S, Isakoff SJ, Ingle JN, Liu MC, Carey LA, et al. Phase II trial of bicalutamide in patients with androgen receptor-positive, estrogen receptor-negative metastatic breast cancer. *Clin Cancer Res*. 2013;19:5505–12. <https://doi.org/10.1158/1078-0432.CCR-12-3327>.

6. Traina TA, Yardley DA, Schwartzberg LS, O'Shaughnessy J, Cortes J, Awada A, et al. Overall survival (OS) in patients (Pts) with diagnostic positive (Dx+) breast cancer: subgroup analysis from a phase 2 study of enzalutamide (ENZA), an androgen receptor (AR) inhibitor, in AR+ triple-negative breast cancer (TNBC) treated with 0-1 prior lines of therapy. *J Clin Oncol*. 2017;35:1089. [https://doi.org/10.1200/JCO.2017.35.15\\_suppl.1089](https://doi.org/10.1200/JCO.2017.35.15_suppl.1089).
7. Youk JH, Kim EK, Kim MJ, Lee JY, Oh KK. Missed breast cancers at US-guided core needle biopsy: how to reduce them. *Radiographics*. 2007;27:79–94. <https://doi.org/10.1148/rg.271065029>.
8. Venema CM, Mammatas LH, Schroder CP, van Kruchten M, Apollonio G, Glaudemans A, et al. Androgen and estrogen receptor imaging in metastatic breast cancer patients as a surrogate for tissue biopsies. *J Nucl Med*. 2017;58:1906–12. <https://doi.org/10.2967/jnumed.117.193649>.
9. van Kruchten M, de Vries EG, Brown M, de Vries EF, Glaudemans AW, Dierckx RA, et al. PET imaging of oestrogen receptors in patients with breast cancer. *Lancet Oncol*. 2013;14:e465–75. [https://doi.org/10.1016/S1470-2045\(13\)70292-4](https://doi.org/10.1016/S1470-2045(13)70292-4).
10. Chae SY, Ahn SH, Kim SB, Han S, Lee SH, Oh SJ, et al. Diagnostic accuracy and safety of 16alpha-[(18)F]fluoro-17beta-oestradiol PET-CT for the assessment of oestrogen receptor status in recurrent or metastatic lesions in patients with breast cancer: a prospective cohort study. *Lancet Oncol*. 2019;20:546–55. [https://doi.org/10.1016/S1470-2045\(18\)30936-7](https://doi.org/10.1016/S1470-2045(18)30936-7).
11. van Kruchten M, Glaudemans AW, de Vries EF, Beets-Tan RG, Schroder CP, Dierckx RA, et al. PET imaging of estrogen receptors as a diagnostic tool for breast cancer patients presenting with a clinical dilemma. *J Nucl Med*. 2012;53:182–90. <https://doi.org/10.2967/jnumed.111.092734>.
12. van Kruchten M, de Vries EG, Glaudemans AW, van Lanschot MC, van Faassen M, Kema IP, et al. Measuring residual estrogen receptor availability during fulvestrant therapy in patients with metastatic breast cancer. *Cancer Discov*. 2015;5:72–81. <https://doi.org/10.1158/2159-8290.CD-14-0697>.
13. Scher HI, Beer TM, Higano CS, Anand A, Taplin ME, Efstathiou E, et al. Antitumour activity of MDV3100 in castration-resistant prostate cancer: a phase 1-2 study. *Lancet*. 2010;375:1437–46. [https://doi.org/10.1016/S0140-6736\(10\)60172-9](https://doi.org/10.1016/S0140-6736(10)60172-9).
14. Kurland BF, Peterson LM, Lee JH, Schubert EK, Currin ER, Link JM, et al. Estrogen receptor binding (18F-FES PET) and glycolytic activity (18F-FDG PET) predict progression-free survival on endocrine therapy in patients with ER+ breast cancer. *Clin Cancer Res*. 2017;23:407–15. <https://doi.org/10.1158/1078-0432.CCR-16-0362>.
15. Liu A, Dence CS, Welch MJ, Katzenellenbogen JA. Fluorine-18-labeled androgens: radiochemical synthesis and tissue distribution studies on six fluorine-substituted androgens, potential imaging agents for prostatic cancer. *J Nucl Med*. 1992;33:724–34.
16. Römer J, Steinbach J, Kasch H. Studies on the synthesis of 16 $\alpha$ -[18F] fluoroestradiol. *Appl Rad Isotop*. 1996;47:395–9.
17. 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>.
18. Jansen BHE, Kramer GM, Cysouw MCF, Yaqub MM, de Keizer B, Lavalaye J, et al. Healthy tissue uptake of (68)Ga-prostate specific membrane antigen (PSMA), (18)F-DCFpYL, (18)F-fluoromethylcholine (FCH) and (18)F-dihydrotestosterone (FDHT). *J Nucl Med*. 2019. <https://doi.org/10.2967/jnumed.118.222505>.
19. de Vet HC, Mokkink LB, Tenwee CB, Hoekstra OS, Knol DL. Clinicians are right not to like Cohen's kappa. *BMJ*. 2013;346:f2125. <https://doi.org/10.1136/bmj.f2125>.
20. Landis JR, Koch GG. An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics*. 1977;33:63–74.
21. Portney LG, Watkins MP. Foundations of clinical research: applications to practice: Pearson/Prentice Hall Upper Saddle River, NJ; 2009.
22. Sawicki LM, Grueneisen J, Schaarschmidt BM, Buchbender C, Nagarajah J, Umutlu L, et al. Evaluation of 18 F-FDG PET/MRI, 18 F-FDG PET/CT, MRI, and CT in whole-body staging of recurrent breast cancer. *Eur J Radiol*. 2016;85:459–65.
23. van der Hoeven JJ, Hoekstra OS, Comans EF, Pijpers R, Boom RP, van Geldere D, et al. Determinants of diagnostic performance of [F-18] fluorodeoxyglucose positron emission tomography for axillary staging in breast cancer. *Ann Surg*. 2002;236:619.
24. Shackleton M, Yuen K, Little AF, Schlicht S, McLachlan SA. Reliability of X-rays and bone scans for the assessment of changes in skeletal metastases from breast cancer. *Intern Med J*. 2004;34:615–20. <https://doi.org/10.1111/j.1445-5994.2004.00637.x>.
25. Gutzeit A, Doert A, Froehlich JM, Eckhardt BP, Meili A, Scherr P, et al. Comparison of diffusion-weighted whole body MRI and skeletal scintigraphy for the detection of bone metastases in patients with prostate or breast carcinoma. *Skeletal Radiol*. 2010;39:333–43. <https://doi.org/10.1007/s00256-009-0789-4>.
26. Jacene HA, Lebourleux S, Baba S, Chatzifotiadis D, Goudarzi B, Teytelbaum O, et al. Assessment of interobserver reproducibility in quantitative 18F-FDG PET and CT measurements of tumor response to therapy. *J Nucl Med*. 2009;50:1760–9. <https://doi.org/10.2967/jnumed.109.063321>.
27. Fox JJ, Autran-Blanc E, Morris MJ, Gavane S, Nehme S, Van Nuffel A, et al. Practical approach for comparative analysis of multilesion molecular imaging using a semiautomated program for PET/CT. *J Nucl Med*. 2011;52:1727–32. <https://doi.org/10.2967/jnumed.111.089326>.
28. Vargas HA, Wassberg C, Fox JJ, Wibmer A, Goldman DA, Kuk D, et al. Bone metastases in castration-resistant prostate cancer: associations between morphologic CT patterns, glycolytic activity, and androgen receptor expression on PET and overall survival. *Radiology*. 2014;271:220–9. <https://doi.org/10.1148/radiol.13130625>.
29. Nienhuis HH, van Kruchten M, Elias SG, Glaudemans A, de Vries EFJ, Bongaerts AHH, et al. (18)F-fluoroestradiol tumor uptake is heterogeneous and influenced by site of metastasis in breast cancer patients. *J Nucl Med*. 2018;59:1212–8. <https://doi.org/10.2967/jnumed.117.198846>.
30. Dehdashti F, Mortimer JE, Trinkaus K, Naughton MJ, Ellis M, Katzenellenbogen JA, et al. PET-based estradiol challenge as a predictive biomarker of response to endocrine therapy in women with estrogen-receptor-positive breast cancer. *Breast Cancer Res Treat*. 2009;113:509–17. <https://doi.org/10.1007/s10549-008-9953-0>.
31. Mortimer JE, Dehdashti F, Siegel BA, Trinkaus K, Katzenellenbogen JA, Welch MJ. Metabolic flare: indicator of hormone responsiveness in advanced breast cancer. *J Clin Oncol*. 2001;19:2797–803. <https://doi.org/10.1200/JCO.2001.19.11.2797>.
32. Amir E, Miller N, Geddie W, Freedman O, Kassam F, Simmons C, et al. Prospective study evaluating the impact of tissue confirmation of metastatic disease in patients with breast cancer. *J Clin Oncol*. 2012;30:587–92. <https://doi.org/10.1200/JCO.2010.33.5232>.
33. Groheux D, Giacchetti S, Moretti JL, Porcher R, Espie M, Lehmann-Che J, et al. Correlation of high 18F-FDG uptake to clinical, pathological and biological prognostic factors in breast cancer. *Eur J Nucl Med Mol Imaging*. 2011;38:426–35. <https://doi.org/10.1007/s00259-010-1640-9>.
34. Mammatas LH, Verheul HM, Hendrikse NH, Yaqub M, Lammertsma AA, Menke-van der Houven van Oordt CW. Molecular imaging of targeted therapies with positron emission tomography: the visualization of personalized cancer care. *Cell Oncol (Dordr)*. 2015;38:49–64. <https://doi.org/10.1007/s13402-014-0194-4>.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:**

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)