

ORIGINAL ARTICLE

Variability of dynamic 18F-FDG-PET data in breast cancer xenografts

KATHINKA E. PITMAN¹, ESPEN RUSTEN^{1,2}, ALEXANDR KRISTIAN³ & EIRIK MALINEN^{1,2}

¹Department of Physics, University of Oslo, Oslo, Norway, ²Department of Medical Physics, Oslo University Hospital, Oslo, Norway and ³Department of Tumour Biology, Oslo University Hospital, Oslo, Norway

ABSTRACT

Background. A murine breast cancer xenograft model was employed to evaluate inter- and intra-variability of various parameters derived from dynamic positron emission tomography with [18F]-fluorodeoxyglucose as tracer (FDG-PET). **Material and methods.** Seventeen female athymic nude foxn1/nu mice with bilaterally implanted triple-negative basal-like ductal carcinoma (MAS98.12) breast cancer xenografts underwent a dynamic PET scan over an hour after injection of approximately 10 MBq FDG. Inter-animal data were obtained from the entire animal cohort, while intra-animal data were from four mice receiving an additional scan after one or two days. Standardised uptake values (SUV_{max} , SUV_{mean} and SUV_{median}) were estimated for all tumours at different time points. Tumour uptake was analysed with a kinetic two-compartment model for estimation of pharmacokinetic parameters. The coefficient of variation (CV) was calculated for all PET-derived metrics.

Results. The CVs for SUV_{mean} and SUV_{median} were typically 10–20% for the tumours, depending on the time post-injection and group (intra vs. inter). The CV for SUV_{max} was mostly higher. The variability in the pharmacokinetic parameters ranged from 23 to almost 150%.

Conclusions. SUV_{mean} and SUV_{median} show less variability than SUV_{max} . The pharmacokinetic tumour metrics again display much greater variability than the SUV-based metrics. However, it is generally not known which of these metrics that best represents cancer aggressiveness and their use may still depend on the research questions addressed.

Positron emission tomography (PET) using 2-deoxy-2-[18F]-fluoro-D-glucose (FDG) as tracer may serve several purposes in oncology. First, it provides a means for detection of malignant lesions within the body, and may aid disease staging and subsequent treatment selection [1]. Second, FDG-PET images allow for tumour delineation in radiotherapy [2]. Third, FDG-PET may be used to monitor and evaluate treatment efficacy during and after therapy [3]. Thus, FDG-PET may serve as a ‘compass for safe navigation’ in the era of personalised medicine [4]. However, when using quantitative measures in the evaluation of treatment efficacy pre- and post-therapy, the measured change reflects a combination of the actual (biological) change and the inherent variability of the measurement. Therefore, a careful evaluation of FDG-PET as a test monitor is needed for both clinical and pre-clinical applications.

The tissue accumulation of the glucose analogue FDG reflects a combination of perfusion, transporter receptor (GLUT1–4) density and hexokinase activity [5]. As many solid tumours employ glycolysis as their main energy source (Warburg effect), they are distinguishable from the background in conventional PET scans performed typically one hour after FDG injection. FDG uptake can be quantified in terms of the standardised uptake value (SUV), which is the amount of tracer accumulated in a region of interest relative to the injected specific activity per body weight [6]. However, the SUV represents both the free (non-metabolised) and bound (metabolised) FDG, where only the latter reflects the Warburg effect. Furthermore, tissue accumulation is a continuous process, and conventional PET only provides a snapshot of this accumulation. Using dynamic PET, tissue accumulation can be monitored in both

time and space, allowing for not only estimation of SUVs at specific time points post-injection, but also full pharmacokinetic modelling [7]. The derived model parameters have been shown to reflect the underlying tissue processes leading to FDG uptake.

PET may be used for experimental testing of new therapies and assessment of treatment response [7–9]. We have used dynamic FDG-PET and pharmacokinetic analysis for assessing response to irradiation in prostate cancer xenografts [10], to imatinib therapy of gastrointestinal stromal xenografts [11] and to anti-angiogenic therapy of breast cancer xenografts [12]. In conjunction with such studies, it is timely to address the reproducibility of the parameters derived from dynamic PET used to assess therapy response. Of particular interest in this respect are the pharmacokinetic parameters, which derive from a statistical model analysis requiring an arterial input function [7]. In the current work, we investigate the reproducibility of parameters derived from dynamic PET of breast cancer xenografts and discuss the implications for experimental response monitoring.

Material and methods

Xenograft tumour model

Seventeen female athymic nude Foxn1^{nu} mice with triple-negative basal-like ductal carcinoma (MAS98.12) breast cancer xenografts [13] form the basis for this study. All experimental protocols were approved by the National Animal Research Authority and conducted in accordance with the guidelines of the Federation of European Laboratory Animal Science Association (FELASA).

Xenograft fragments of approximately 1 mm³ obtained from previous passage were implanted bilaterally in the mammary fat pads of 7–8-week-old mice (weight 23–25 g). Before implantation of xenografts or PET imaging, animals were anaesthetised by subcutaneous injection at dosage 75 µl/10 g body weight with a combination of 3.8 mg/ml xylazine (Narcoxy Vet, Roche, Basel, Switzerland), 2.4 mg/ml tiletamine and 2.4 mg/ml zolazepam (Zoletil Vet, Virbac Laboratories, Carros, France), and 0.1 mg/ml butorphanol (Torbugesic, Fort Dodge Laboratories, IA, USA), diluted 1:5 in sterile water.

Dynamic FDG-PET imaging

PET imaging commenced once tumours reached a diameter of 8–10 mm in the longest axis, approximately five weeks after implantation. Following overnight fasting, animals were anaesthetised and catheterised in the tail vein, and groups of three mice were placed with tumours centred in the gantry of a Siemens microPET Focus 120 (Erlangen, Germany)

pre-clinical scanner. Ten min transmission scans using a ⁶⁸Ge point source were conducted for scatter and attenuation correction. Sixty min PET acquisitions were then started in list mode prior to intravenous (i.v.) injection of approximately 10 MBq of FDG (GE Healthcare AS, Oslo, Norway) diluted in 0.9% NaCl. During scanning, animals were covered with an insulating material and kept warm by a heating lamp. Body temperature was maintained at about 35°C as measured by rectal probe in a representative mouse. After completed imaging, anaesthetised animals were sacrificed by cervical dislocation.

Images were reconstructed using OSEM3D-MAP (two OSEM iterations, 18 MAP iterations, $\beta = 0.5$, matrix size = $128 \times 128 \times 95$) [14] with resulting voxel size of $0.87 \times 0.87 \times 0.80$ mm³. Reconstructed time frames ranged from 10 s for early time points to 10 min for late time points.

All 17 animals each received one scan. Additionally, four animals underwent a repeat scan one or two days after the first and form the basis for the intra-reproducibility study. Mice were untreated and no major tumour progression or physiological changes were assumed to occur during the experiment.

Image analysis

Images were analysed with bespoke software in IDL 8.3 (ITT Visual Information Solutions, Boulder, CO, USA). Tumours were manually delineated in final frame images.

For normalisation of dPET-data, full-body VOIs were segmented and the total injected activity determined for all animals. This method was chosen as the extracted activity was thought to more accurately represent the actual injected activity than the nominal injected activity of 10 MBq, as all three mice were consecutively injected with a single syringe of FDG, making measurement of residual activity uninformative for individual injected activities. We did indeed find less variability in the data using this approach (data not shown). Pseudo-SUVs (hereafter SUVs) were calculated as the tissue activity concentration in MBq/ml divided by the integrated injected activity (MBq) normalised to full-body volume (ml). Maximum, mean and median SUVs over all contained voxels were determined for all tumours, and three corresponding SUV time activity curves (TACs) were generated for each individual VOI.

Individual arterial input functions (AIFs) were obtained by contouring of the maximal voxels of the aorta in the initial frame of injection for each animal (9–13 voxels). For a given animal, the time at which the AIF reached a maximum was determined and used to shift corresponding tumour TACs along the time axis to a common origin. Animals with AIFs which peaked in the first frame were excluded from

the analysis. As one set of animals was scanned for 45 min only, all TACs were interpolated to a uniform time axis with a temporal resolution of 10 s and range 0–45 min.

A total of 38 tumours could be identified in the image material using the final frame of the dynamic acquisition. Two animals were excluded from the analysis due to inadequate AIFs, and one borderline occult tumour was removed. This gave a data set of 34 tumours. Intra-animal reproducibility was evaluated over test-retest scans for a set of seven.

The injected activity estimated from full-body segmentations ranged from 6.5 to 14.6 MBq (mean, 10.8 MBq).

Kinetic modelling

Tumour TACs (max, mean and median) were fitted to a two-tissue compartment model using individual AIFs, as described by Røe et al. [10,15]. In brief, the model assumes that tracer taken up from the blood pool may be present in cells in a free (non-metabolised)

or a bound (metabolised) state or compartment. The four rate constants K_1 , k_2 , k_3 and k_4 describe the exchange of FDG between the two compartments; K_1 gives a measure of capillary permeability and perfusion rate, k_2 describes the efflux of FDG to the blood stream, k_3 gives a measure of metabolic phosphorylation of available FDG, whilst k_4 represents dephosphorylation of metabolised tracer. In our analysis, FDG is assumed to be irreversibly metabolised and k_4 was set to zero.

Variability was examined for K_1 , k_2 and k_3 , as well as two additional kinetic parameters; v_B and MR_{FDG} . v_B estimates the fraction of vascular volume in tissues, whilst MR_{FDG} gives a measure of the metabolic rate of FDG, given by $\text{MR}_{\text{FDG}} = (K_1 k_3) / (k_2 + k_3)$.

Patlak analysis was performed using $\text{SUV}_{\text{median}}$ AIFs for all tumour TACs by linear regression on data in the equilibrium region from 7.5 min to the end of acquisition. The slope of the regression line, K , gives another measure of metabolic rate of FDG whilst the intercept V_0 reflects volume of distribution.

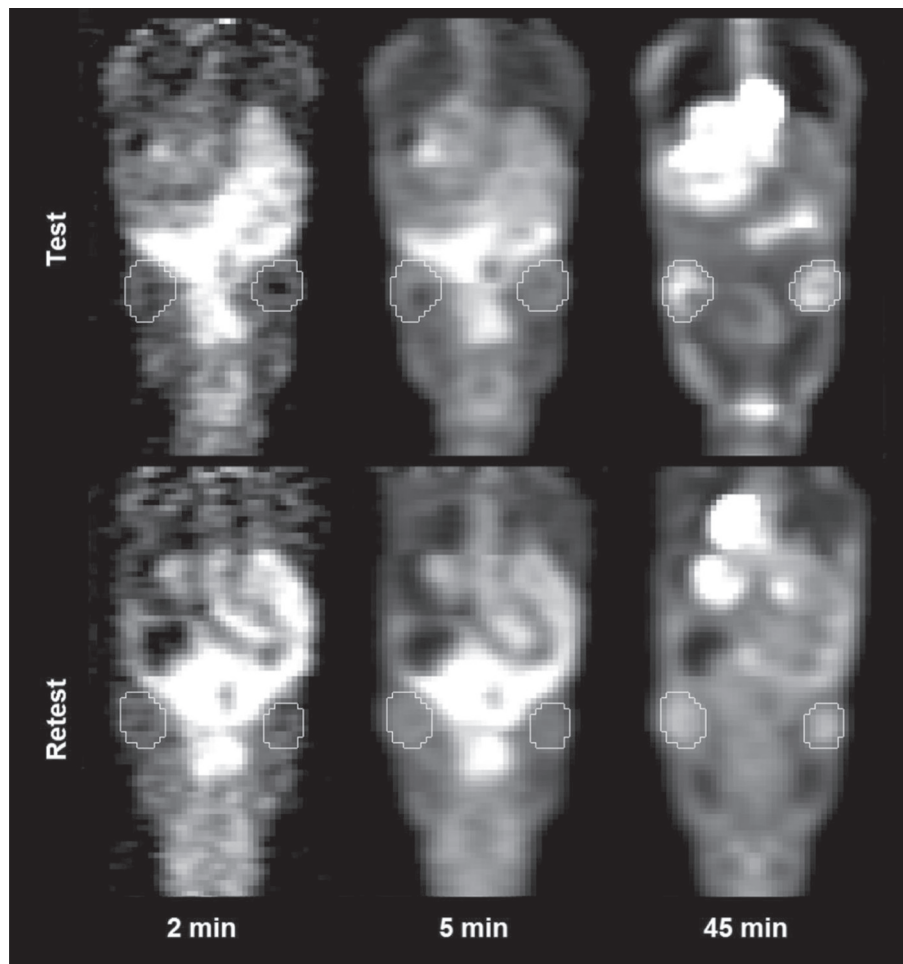


Figure 1. Coronal test (upper) and retest (lower) dynamic FDG-PET images for an individual representative animal implanted with MAS98.12 breast cancer xenografts at acquisition times 2 min, 5 min and 45 min. Tumours segmented in final frame images of test and retest scans are shown throughout.

Statistics

The variability of SUVs and pharmacokinetic parameters was examined in terms of the coefficient of variation (CV) expressed as a percentage, given by the standard deviation (SD) divided by the mean. The CV for intra-animal repeatability was calculated

as $\sqrt{\sum \frac{d^2}{2n}}$ divided by the grand mean of test and retest values, where d is the difference between test and retest values, and n is the number of mice in the intra group. Goodness of fit for the tracer kinetic curves resulting from kinetic modelling was evaluated with the squared Pearson's correlation coefficient (r^2) between the original and the fitted data.

Results

Figure 1 shows coronal slices of dPET images obtained for an individual representative animal on

Day 1 and 2 at three acquisition times; 2 min, 5 min and 45 min. The tumours, as delineated in the final PET frame, are also shown. As seen, the tumour uptake patterns appear quite similar, both temporally and spatially. At 2 min p.i., little uptake is present in either image, while at 45 min p.i. an active tumour rim is apparent in both tumours at Day 1 and 2.

Figure 2 shows TACs for all the tumours, extracted for the maximum, mean and median uptake in the tissue. An initially rapid uptake is succeeded by a slower accumulation of FDG after about 5–10 minutes. All tumours displayed overall a relatively modest FDG uptake, with a population-based mean SUV_{max} of 1.9 at 45 min p.i. (Table I). Furthermore, the corresponding SUV_{mean} and SUV_{median} in the tumours were similar to each other but almost 60% lower than SUV_{max} . The uptake curves in Figure 2 reflect rather similar uptake patterns from animal to animal, although SUV_{max} for the tumours shows greater variability at late time points.

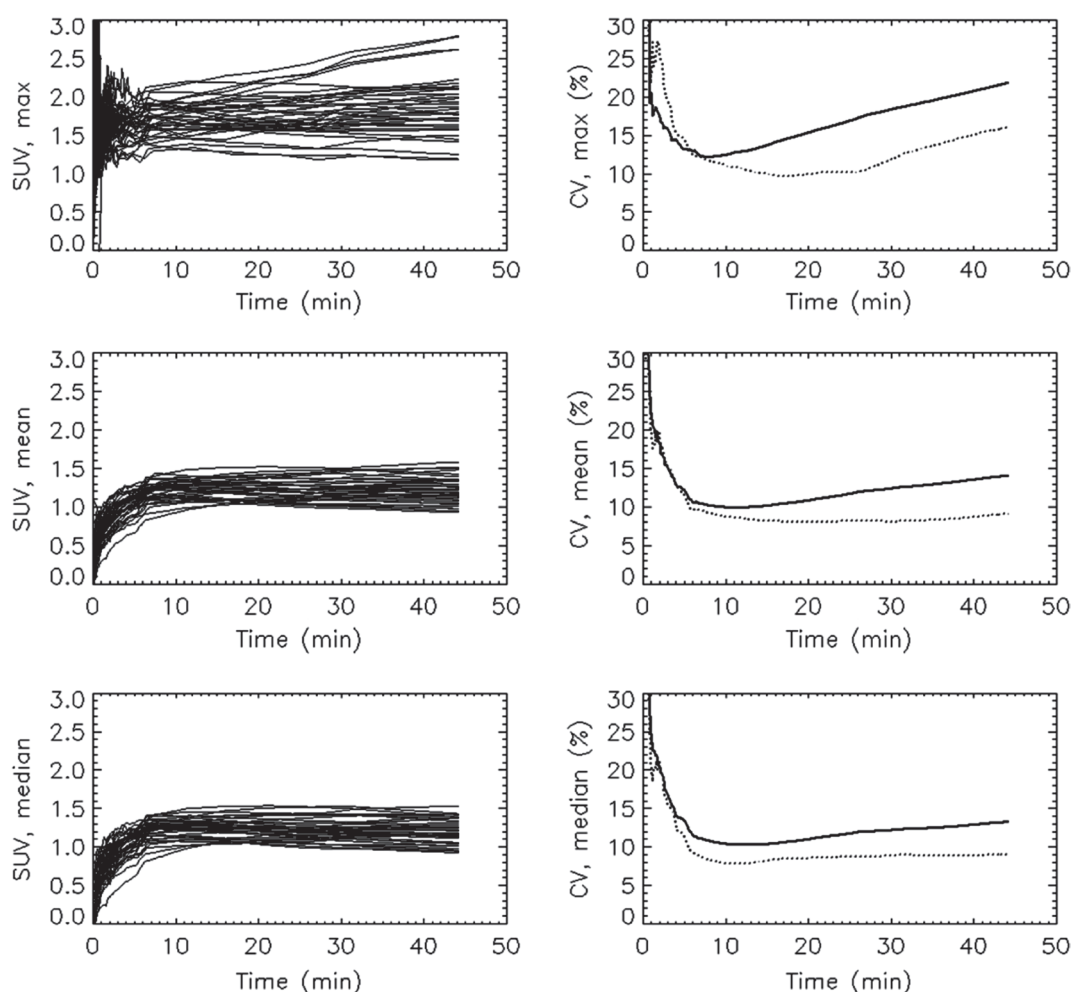


Figure 2. The left column shows individual max (top), mean (middle) and median (bottom) TACs for all tumours, demonstrating inter-animal variability of dynamic FDG-uptake in 34 tumours bilaterally implanted in 19 animals (four animals had only one tumour). In the right column, inter- (solid) and intra-animal (dotted) variability expressed in terms of CVs of tracer uptake as a function of time after injection for max, mean and median TACs are shown ($n_{inter} = 34$, $n_{intra} = 7$).

Table I. Overview of PET-data from tumours with average values, standard deviations (SD) and inter and intra coefficients of variation (CV). Pharmacokinetic parameters (non-SUVs) are given as $\times 10^2$.

Tumour Metric	Average $\pm$ SD, Inter (n = 34)	Average $\pm$ SD, Intra (n = 7)	CV, Inter (%)	CV, Intra (%)
SUV (2 min)				
max	1.7 $\pm$ 0.3	1.7 $\pm$ 0.4	17	26
mean	0.81 $\pm$ 0.15	0.80 $\pm$ 0.16	18	20
median	0.80 $\pm$ 0.16	0.80 $\pm$ 0.17	20	21
SUV (5 min)				
max	1.6 $\pm$ 0.2	1.7 $\pm$ 0.2	13	14
mean	1.0 $\pm$ 0.1	1.1 $\pm$ 0.1	12	11
median	1.1 $\pm$ 0.1	1.1 $\pm$ 0.1	13	11
SUV (45 min)				
max	1.9 $\pm$ 0.4	2.2 $\pm$ 0.4	22	16
Mean	1.2 $\pm$ 0.2	1.3 $\pm$ 0.1	14	9
median	1.2 $\pm$ 0.2	1.3 $\pm$ 0.1	13	9
K ₁				
max	20 $\pm$ 8	19 $\pm$ 6	40	33
mean	10 $\pm$ 2	10 $\pm$ 2	24	25
median	10 $\pm$ 3	10 $\pm$ 3	26	27
k ₂				
max	40 $\pm$ 17	37 $\pm$ 17	43	46
mean	20 $\pm$ 5	20 $\pm$ 6	25	28
median	19 $\pm$ 5	19 $\pm$ 6	29	31
k ₃				
max	3.5 $\pm$ 1.9	4.0 $\pm$ 2.1	56	53
mean	2.1 $\pm$ 1.2	2.1 $\pm$ 1.1	59	50
median	1.8 $\pm$ 1.0	1.8 $\pm$ 0.9	55	48
MR _{FDG}				
max	1.5 $\pm$ 0.7	1.8 $\pm$ 0.6	46	36
mean	0.87 $\pm$ 0.37	0.94 $\pm$ 0.34	43	36
median	0.80 $\pm$ 0.35	0.85 $\pm$ 0.31	43	36
v _B				
max	11 $\pm$ 7	12 $\pm$ 6	68	53
mean	1.8 $\pm$ 2.2	1.5 $\pm$ 1.6	121	111
median	1.4 $\pm$ 2.0	1.1 $\pm$ 1.5	143	137
K				
max	0.24 $\pm$ 0.12	0.28 $\pm$ 0.10	49	36
mean	0.15 $\pm$ 0.06	0.16 $\pm$ 0.05	43	31
median	0.14 $\pm$ 0.06	0.15 $\pm$ 0.05	43	31
V ₀				
max	58 $\pm$ 13	59 $\pm$ 14	22	24
mean	45 $\pm$ 9	44 $\pm$ 9	20	21
median	46 $\pm$ 9	47 $\pm$ 8	19	18

From the individual SUV_{max}, SUV_{mean} and SUV_{median} data displayed in Figure 2, CVs were generated for the inter- and intra-animal groups. The tumours were treated as independent. We also performed the analysis treating tumours from the same animal as a single lesion, but this yielded virtually no change in the numbers ($< \pm 1\%$). The right column of Figure 2 shows CVs plotted against time p.i. For SUV_{mean} and SUV_{median} intra-animal variability is relatively stable 10–45 min p.i. at a level of about 9%. However, SUV_{max} displays increasing variability after 10 min. The inter-animal CV is mostly greater than the CV for the intra-animal group, especially at late time points. These findings are summarised in Table I for selected time points.

Shown in Figure 3A are the individual AIFs (SUV_{median}) obtained for all animals. Panel B shows the range of measured arterial activity over the first 2 min post-injection (p.i.). CVs were calculated for the inter and intra groups, and are shown plotted against time p.i. in Panels C and D of Figure 3. The inter-animal variability is high with CVs of approximately 35% immediately after injection. Test-retest variability is somewhat higher at 50%. After the initial peak, the intra- and inter-animal CVs drop to 10–25%, with intra-animal variability slightly lower than inter-animal variability.

The tumour TACs were fitted to the two-compartment model using individual AIFs, and estimates for pharmacokinetic parameters K₁, k₂, k₃,

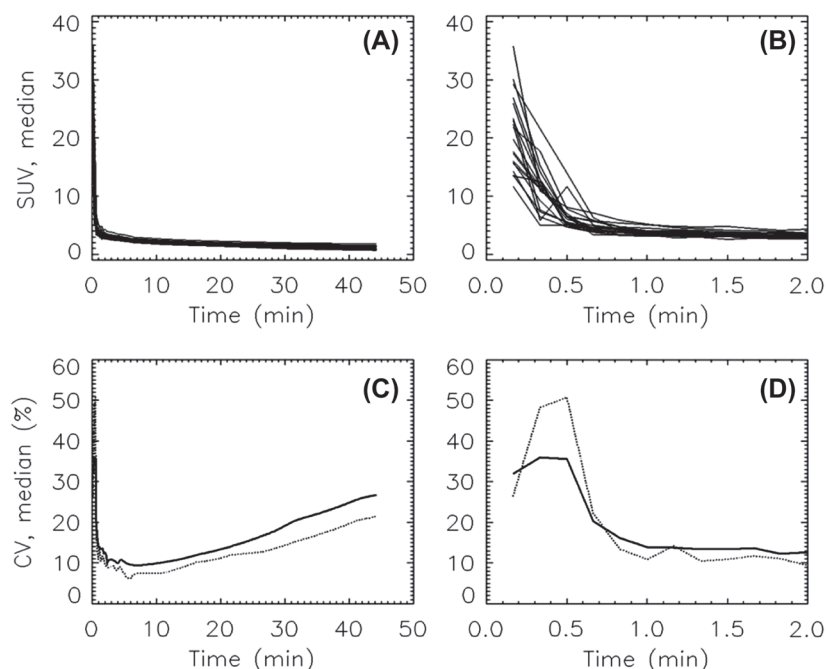


Figure 3. The top panels show median AIFs obtained for individual animals for all times p.i. (A) and for the first 2 min of the acquisition (B). In the bottom panels, the inter- (solid) and intra-animal (dotted) variability in median AIFs are shown as a function of time p.i., for the entire dynamic range (C) and over the first 2 min p.i. (D).

MR_{FDG} and v_B were obtained. A high agreement between experimental and fitted uptake values was found for SUV_{mean} and SUV_{median} ($r^2=0.97$), but was lower for SUV_{max} ($r^2=0.66$). Population-based mean values and inter- and intra-animal variability for kinetic parameters derived from SUV_{max} , SUV_{mean} and SUV_{median} are given in Table I. As seen, the relative variability for the pharmacokinetic parameters was much higher than for SUV-derived metrics. While CVs for SUV were mostly limited to the range 9–22%, CVs for the pharmacokinetic parameters ranged from 24% to as high as 143%.

Patlak plots (not shown) were constructed for all tumour TACs. Agreement between experimental data and values fitted in the linear region (7.5–45 min) was high ($r^2=0.98$). Estimates for the parameters K and V_0 were extracted and evaluated for inter- and intra-variability, see Table 1. The variability in K is very similar to that found for MR_{FDG} , with CV approximately 35% for the intra group and CV approximately 45% for the inter group. V_0 displayed somewhat lower variability with CV approximately 20% for all metrics.

Discussion

In this study, we evaluated the variability of SUV-derived metrics and of parameters from pharmacokinetic analysis of dynamic ^{18}F -FDG-PET of breast cancer xenografts in athymic mice. It was found that: 1) SUV_{max} shows greater variability compared to

SUV_{mean} and SUV_{median} ; 2) intra-animal variability was lower than inter-animal variability; 3) variability is dependent on the time post-injection; and 4) pharmacokinetic parameters show much greater variability compared with SUV-based metrics.

SUVs obtained late in the acquisition phase (45 min p.i.) were found to have test-retest CVs of 9% for $SUV_{\text{mean/median}}$ and 16% for SUV_{max} . This level of variability is comparable with previous pre-clinical work on a murine melanoma xenograft model in which SUV_{mean} and SUV_{max} of test-retest scans acquired 6 h apart yielded CVs of around 15% and 18%, respectively [16]. Furthermore, the large difference in variability found between SUV_{max} and $SUV_{\text{mean/median}}$ is in line with previous clinical findings [17]. Nahmias and Wahl studied variability in SUV-based metrics (max and mean) for various cancers in patients [17]. For breast cancer lesions, the variability of test-retest SUV_{mean} ranged from 2% to 4% for the hottest lesions through 10% for intermediate lesions and up to almost 18% for the occult lesions. Considering the rather low uptake and small lesion sizes of the xenografts included in our study, the reproducibility is rather good as compared with clinical data. This is likely due to the superior resolution of animal scanners relative to clinical scanners, reducing partial volume effects. Furthermore, the intra- and inter-animal variability was rather similar in the current work. This is not unexpected, given the highly standardised conditions implemented in the animal model, anticipating a between subject

variability which is greater than, but close to the level of within subject variability.

It is not known which SUV-metric is most representative of tumour aggressiveness or best predicts prognosis. Several studies show that a higher FDG uptake as measured by SUV_{max} typically 1 hr p.i. is associated with a poorer prognosis [18]. Some studies have also compared the prognostic power of different metrics; Borst et al. found that SUV_{max} better predicted treatment response in inoperable NSCLC patients than SUV_{mean} [19]. Higgins et al., however, found better prognostic value for disease-free survival in SUV_{mean} than SUV_{max} for head and neck cancer patients [20]. However, tumour hot spots, which may influence prognosis, will not be detected by SUV_{mean} . Furthermore, none of the aforementioned metrics will reflect intra-tumour heterogeneity. The recently published PERCIST guidelines suggest the use of SUV_{max} or SUV_{peak} for evaluation of treatment response in solid cancers. In that work, a 30% relative decrease (and 0.8 units absolute decrease) in $SUV_{max/peak}$ is taken to indicate a significant treatment effect (partial response) [8]. For our data (late acquisition phase; 45 min p.i.), this level of confidence is reached with a relative change of 17% in $SUV_{mean/median}$ and 28% in SUV_{max} (95% relative test-retest variability). However, in the meta-analysis on test-retest variability in SUV-derived metrics, de Langen et al. [21] found less variability in SUV_{mean} than SUV_{max} , suggesting that treatment response may be quantified in terms of a 20% relative reduction combined with 1.2 units absolute reduction in SUV_{mean} .

We found that the tumour TACs overall follow similar patterns, with relatively low FDG-uptake which stabilises quite quickly, typically after 5–10 min. The variability of the SUV-based metrics, as measured by the CV, varies over time post-injection. The high variability in the initial phase is in part due to inherent variability and noise in the early uptake of individual tumours, but also a result of the CV being a mean normalised quantity. There is a slight increasing trend in variability with time between tumours, which is likely a reflection of the varying metabolic activity of the different lesions.

The SUV obtained at a given time point p.i. represents the amount of FDG accumulated in the tissue in question up to that point. It is as such a time dependent measure, an aspect which gives rise to variability, and standardisation of time point is thus of importance when comparing SUVs between studies. Through pharmacokinetic modelling, the observed FDG accumulation is explained in terms of vascular perfusion and permeability, GLUT receptor density and glycolytic activity of the tissue. The resulting rate constants are time-invariant, providing

physiological descriptors which could be readily comparable between studies. We have, however, found that these pharmacokinetic parameters are highly variable, with poorer reproducibility than SUV-based metrics.

The late SUV (45 min p.i.) is commonly used as a surrogate for metabolic activity. Yet, pharmacokinetic parameters such as MR_{FDG} and k_3 are expected to more closely relate to the true metabolism, and are as such attractive as prognostic markers. For instance, Okazumi et al. [22] found a significant decrease in k_3 of liver tumours after chemotherapy, and suggest the use of this parameter as a quantitative index for tumour viability. Also, Wahl et al. [23] found a significant correlation between outcome of therapy and early decrease in k_3 in breast cancer patients. However, Minn et al. [24] found substantial variability in pharmacokinetic rate constants in lung cancers, with levels of variability slightly lower than in our study, and advocate caution for their use in monitoring of individual patients. One might expect MR_{FDG} to be a more robust metabolic parameter than k_3 , as large relative noise in k_3 propagated from poor signal into the free compartment (low K_1) is partially compensated for in MR_{FDG} . Indeed, a substantially lower variability in MR_{FDG} compared to k_3 was found in the current work. However, the variability in MR_{FDG} is still roughly three times higher than SUV at 45 min p.i. Both the full pharmacokinetic analysis and the Patlak approach yielded roughly the same level of variability in the metabolic parameters MR_{FDG} and K . As the Patlak approach employs linear regression with only two parameters, whilst the full pharmacokinetic analysis uses non-linear regression with five parameters, it could be expected that Patlak analysis is more robust than pharmacokinetic modelling. This was however not reflected in our findings.

The potential of dynamic PET for evaluating tumour vasculature [12,25] and the associated predictive significance has not been commonly addressed. PET parameters of relevance are in this case K_1 , V_0 and SUV at 2 and 5 min p.i., all reflecting the early tissue distribution phase. Partridge et al. used longitudinal FDG-PET/CT, and found a correlation between reduction in K_1 and pathologic response status for patients with breast cancer undergoing neoadjuvant chemotherapy [26]. For complete pathologic responders (pCR), a median reduction of 83% in K_1 was found, against a median reduction of 38% for patients not classified as pCR. Previously, we studied the effects of anti-angiogenic therapy in breast cancer xenografts, and could show that K_1 was significantly altered 24 h after therapy [12]. Thus, K_1 may play a role in response assessment following chemotherapy. Comparing K_1 with SUV at 2 min p.i.

and V_0 , the present work indicates that the variability in these metrics is comparable at around 20%. The only exception was K_1 obtained for maximum tumour uptake, for which the variability was roughly twice as high. Furthermore, there was a 25–30% reduction in all K_1 metrics and a 30–35% reduction in k_2 metrics between our test-retest data. As no major tumour progression is assumed to occur in the interim between scans, this is most likely due to the small population size sampled, and may be ascribed to randomness.

The use of an image-derived input function (IDIF) offers an elegant, non-invasive alternative to more laborious blood sampling procedures, particularly in small animal imaging where blood supply is limited. There are however drawbacks to the method, including sensitivity to partial volume effects, movement of the animal during the acquisition and limited temporal resolution. Imprecise representation of the AIF is potentially a great source of error for the estimated parameters. The AIFs obtained across animals in the current work were similar in shape, but, e.g. the SUV in the peak region ranged from 11 to 36. The AIF is thus expected to be a major contributor to the variability observed in pharmacokinetic parameters in this study. A pooled AIF (average of all individual AIFs) was therefore tested in the pharmacokinetic modelling for individual tumours, but this did not reduce variability (data not shown). It may well be that the temporal resolution of the initial frames of PET acquisition was slightly low, giving both smearing and erroneous sampling of the AIF peak. Smearing will lead to a reduction in the AIF peak and thus to overestimation of K_1 . Furthermore, erroneous sampling will for some miss the peak altogether, again leading to an overestimation of K_1 . These errors will be further magnified for estimation of parameters associated with the bound compartment (k_3 , MR_{FDG}), since these depends on input from the free compartment. One could improve accuracy by reconstructing shorter time frames for the first minute of acquisition, but this comes at a cost of decreased signal relative to noise. Our method of IDIF extraction should however be validated against blood sampling. De Geus-Oei *et al.* found a high correlation ($r = 0.98$) between IDIFs obtained from the ascending aorta in FDG-PET images and arterial blood samples for a cohort of 76 cancer patients [27]. The temporal sampling rate was 30 s and 15 s for PET and blood sampling, respectively. The plasma peak was underestimated in the IDIFs as compared with the blood samples, which was likely due to the relatively few initial frames acquired, giving rise to an increased averaging effect. Validation of IDIF by arterial sampling in mice is however more challenging because of the small volume which may

be withdrawn. Alf *et al.* suggested the use of an arteriovenous shunt for time-resolved validation [28]. They found a trend towards overestimation of the pharmacokinetic parameters $K_1 - k_3$ and MR_{FDG} and greater variability in these when using an IDIF as compared with the shunt input function. In particular, K_1 and k_2 could be estimated with higher precision using the shunt IF due to the improved temporal resolution in the interval immediately following injection. The authors do recognise the limitations of the shunt IF for high throughput and longitudinal studies and deem the IDIF a satisfactory alternative for such experiments.

To conclude, whether the level of variability in a potential predictive parameter is acceptable depends on the magnitude of the treatment effect. If there is a very small effect of the treatment relative to the variability of the parameter, increasing sample size is not a fruitful way to identify changes in the tumour. It is not certain how much variability is inherently present in the tumour physiology, and as such it may be preferable to use SUVs over pharmacokinetic parameters if the treatment effect is small. Even though SUVs may not provide any major insight into tumour physiology, such parameters display less variability than kinetic parameters.

Acknowledgements

Financial support received from MLS^{UiO} – Molecular Life Science, University of Oslo, and from the K.G. Jebsen Center for Breast Cancer Research, Oslo, Norway is gratefully acknowledged. The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

Declaration of interest: The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

References

- [1] Gambhir SS, Czernin J, Schwimmer J, Silverman DH, Coleman RE, Phelps ME. A tabulated summary of the FDG PET literature. *J Nucl Med* 2001;42(5 Suppl):1S–93S.
- [2] Troost EG, Schinagel DA, Bussink J, Oyen WJ, Kaanders JH. Clinical evidence on PET-CT for radiation therapy planning in head and neck tumours. *Radiother Oncol* 2010;96:328–34.
- [3] Weber WA. Use of PET for monitoring cancer therapy and for predicting outcome. *J Nucl Med* 2005;46:983–95.
- [4] de Geus-Oei LF, Vriens D, Arens AI, Hutchings M, Oyen WJ. FDG-PET/CT based response-adapted treatment. *Cancer Imaging* 2012;12:324–35.
- [5] Pauwels E, Ribeiro M, Stoot J, McCready V, Bourguignon M, Maziere B. FDG accumulation and tumor biology. *Nucl Med Biol* 1998;25:317–22.
- [6] Strauss LG, Conti PS. The applications of PET in clinical oncology. *J Nucl Med* 1991;32:623–48; discussion 49–50.

- [7] Castell F, Cook G. Quantitative techniques in 18FDG PET scanning in oncology. *Br J Cancer* 2008;98:1597–601.
- [8] 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):122S–50S.
- [9] Fuss M. Strategies of assessing and quantifying radiation treatment metabolic tumor response using F18 FDG positron emission tomography (PET). *Acta Oncol* 2010;49:948–55.
- [10] Røe K, Aleksandersen TB, Kristian A, Nilsen LB, Seierstad T, Qu H, et al. Preclinical dynamic 18F-FDG PET-tumor characterization and radiotherapy response assessment by kinetic compartment analysis. *Acta Oncol* 2010;49:914–21.
- [11] Revheim M-E, Kristian A, Malinen E, Bruland ØS, Berner J-M, Holm R, et al. Intermittent and continuous imatinib in a human GIST xenograft model carrying KIT exon 17 resistance mutation D816H. *Acta Oncol* 2013;52:776–82.
- [12] Kristian A, Revheim ME, Qu H, Mælandsmo GM, Engebråten O, Seierstad T, et al. Dynamic 18F-FDG-PET for monitoring treatment effect following anti-angiogenic therapy in triple-negative breast cancer xenografts. *Acta Oncol* 2013;52:1566–72.
- [13] Bergamaschi A, Hjortland GO, Triulzi T, Sørli T, Johnsen H, Ree AH, et al. Molecular profiling and characterization of luminal-like and basal-like in vivo breast cancer xenograft models. *Mol Oncol* 2009;3:469–82.
- [14] Qi J, Leahy RM. Resolution and noise properties of MAP reconstruction for fully 3-D PET. *Med Imaging IEEE Trans* 2000;19:493–506.
- [15] Rusten E, Rødal J, Revheim ME, Skretting A, Bruland ØS, Malinen E. Quantitative dynamic 18FDG-PET and tracer kinetic analysis of soft tissue sarcomas. *Acta Oncol* 2013;52:1160–7.
- [16] Dandekar M, Tseng JR, Gambhir SS. Reproducibility of 18F-FDG microPET studies in mouse tumor xenografts. *J Nucl Med* 2007;48:602–7.
- [17] Nahmias C, Wahl LM. Reproducibility of standardized uptake value measurements determined by 18F-FDG PET in malignant tumors. *J Nucl Med* 2008;49:1804–8.
- [18] Baba S, Isoda T, Maruoka Y, Kitamura Y, Sasaki M, Yoshida T, et al. Diagnostic and prognostic value of pretreatment SUV in 18F-FDG/PET in breast cancer: Comparison with apparent diffusion coefficient from diffusion-weighted MR imaging. *J Nucl Med* 2014;55:736–42.
- [19] Borst GR, Belderbos JSA, Boellaard R, Comans EFI, Jaeger KD, Lammertsma AA, et al. Standardised FDG uptake: A prognostic factor for inoperable non-small cell lung cancer. *Eur J Cancer* 2005;41:1533–41.
- [20] Higgins KA, Hoang JK, Roach MC, Chino J, Yoo DS, Turkington TG, et al. Analysis of pretreatment FDG-PET SUV parameters in head-and-neck cancer: Tumor SUVmean has superior prognostic value. *Int J Radiat Oncol Biol Phys* 2012;82:548–53.
- [21] de Langen AJ, Vincent A, Velasquez LM, van Tinteren H, Boellaard R, Shankar LK, et al. Repeatability of 18F-FDG uptake measurements in tumors: A metaanalysis. *J Nucl Med* 2012;53:701–8.
- [22] Okazumi S, Isono K, Enomoto K, Kikuchi T, Ozaki M, Yamamoto H, et al. Evaluation of liver tumors using fluorine-18-fluorodeoxyglucose PET: Characterization of tumor and assessment of effect of treatment. *J Nucl Med* 1992;33:333–9.
- [23] Wahl RL, Zasadny K, Helvie M, Hutchins G, Weber B, Cody R. Metabolic monitoring of breast cancer chemohormonotherapy using positron emission tomography: Initial evaluation. *J Clin Oncol* 1993;11:2101–11.
- [24] Minn H, Zasadny KR, Quint LE, Walh RL. Lung cancer: Quantitative measurements for evaluating 2-IF-1 8J-Fluoro-2deoxy-Dglucose uptake at PET. *Radiology* 1995;1:167–73.
- [25] Malinen E, Rødal J, Knudtsen IS, Sovik A, Skogmo HK. Spatiotemporal analysis of tumor uptake patterns in dynamic (18)FDG-PET and dynamic contrast enhanced CT. *Acta Oncol* 2011;50:873–82.
- [26] Partridge SC, Vanantwerp RK, Doot RK, Chai X, Kurland BF, Eby PR, et al. Association between serial dynamic contrast enhanced MRI and dynamic 18F-FDG PET measures in patients undergoing neoadjuvant chemotherapy for locally advanced breast cancer. *J Magn Reson Imaging* 2010;32:1124–31.
- [27] de Geus-Oei L-F, Visser EP, Krabbe PF, van Hoorn BA, Koenders EB, Willemsen AT, et al. Comparison of image-derived and arterial input functions for estimating the rate of glucose metabolism in therapy-monitoring 18F-FDG PET studies. *J Nucl Med* 2006;47:945–9.
- [28] Alf M, Martić-Kehl M, Schibli R, Krämer S. FDG kinetic modeling in small rodent brain PET: Optimization of data acquisition and analysis. *EJNMMI Res* 2013;3:1–14.