Published in final edited form as:

Mol Imaging Biol. 2013 February ; 15(1): 87–96. doi:10.1007/s11307-012-0564-0.

Reproducibility of Static and Dynamic ^{18}F -FDG, ^{18}F -FLT, and ^{18}F -FMISO MicroPET Studies in a Murine Model of HER2+ Breast Cancer

Jennifer G. Whisenant^{1,6}, Todd E. Peterson^{1,2,4}, Jacob U. Fluckiger³, Mohammed Noor Tantawy^{1,2}, Gregory D. Ayers⁵, and Thomas E. Yankeelov^{1,2,6}

¹Institute of Imaging Science, Vanderbilt University Medical Center, 1161 21st Ave. South, MCN AA-1105, Nashville, TN 37232-2675, USA

²Department of Radiology and Radiological Sciences, Vanderbilt University, Nashville, TN 37232-2675, USA

³Department of Radiology, Northwestern University, 676 N Saint Clair, Chicago, IL 60611, USA

⁴Department of Physics and Astronomy, Vanderbilt University, Nashville, TN 37232-2675, USA

⁵Department of Biostatistics, Vanderbilt University, Nashville, TN 37232-2675, USA

⁶Department of Cancer Biology, Vanderbilt University, Nashville, TN 37232-2675, USA

⁷Program in Chemical and Physical Biology, Vanderbilt University, Nashville, TN 37232-2675, USA

Abstract

Purpose—The objective of this study is to determine the reproducibility of static 2-deoxy-2- ^{18}F fluoro-D-glucose (^{18}F -FDG), 3'-deoxy-3'- ^{18}F fluorothymidine (^{18}F -FLT), and ^{18}F -fluoromisonidazole (^{18}F -FMISO) microPET measurements, as well as kinetic parameters returned from analyses of dynamic ^{18}F -FLT and ^{18}F -FMISO data.

Procedures—HER2+ xenografts were established in nude mice. Dynamic data were acquired for 60 min, followed by a repeat injection and second scan 6 h later. Reproducibility was assessed for the percent-injected dose per gram (%ID/g) for each radiotracer, and with kinetic parameters (K_1 – k_4 , K_i) for ^{18}F -FLT and ^{18}F -FMISO.

Results—The value needed to reflect a change in tumor physiology is given by the 95 % confidence interval (CI), which is ± 14 , ± 5 , and ± 6 % for ^{18}F -FDG ($n=12$), ^{18}F -FLT ($n=11$), and ^{18}F -FMISO ($n=11$) %ID/g, respectively. $V_d (=K_1/k_2)$, k_3 , and K_{FLT} are the most reproducible ^{18}F -FLT ($n=9$) kinetic parameters, with 95 % CIs of ± 18 , ± 10 , and ± 18 %, respectively. V_d and K_{FMISO} are the most reproducible ^{18}F -FMISO kinetic parameters ($n=7$) with 95 % CIs of ± 16 and ± 14 %, respectively.

Conclusions—Percent-injected dose per gram measurements are reproducible and appropriate for detecting treatment-induced changes. Kinetic parameters have larger threshold values, but are potentially sufficiently reproducible to detect treatment response.

Keywords

Reproducibility; ^{18}F -FDG; ^{18}F -FLT; ^{18}F -FMISO; Kinetic modeling

Introduction

Current radiographic analysis for monitoring anticancer treatments is based on the response evaluation criteria in solid tumors (RECIST), which measures response using changes in a unidimensional measurement [1]. With the development of novel therapeutics that target specific molecular pathways of breast cancer and the microenvironment [2], RECIST may not be the most sensitive or accurate approach [3]. Three techniques that have been investigated as surrogate biomarkers of tumor response are 2-deoxy-2- ^{18}F fluoro-D-glucose (^{18}F -FDG), 3'-deoxy-3'- ^{18}F fluorothymidine (^{18}F -FLT), and ^{18}F -fluoromisonidazole (^{18}F -FMISO) positron emission tomography (PET). ^{18}F -FDG provides an assessment of glucose metabolism as its accumulation is regulated by glucose transporters and hexokinase activity [3]. As ^{18}F -FLT accumulation is regulated by the cell-cycle-dependent thymidine salvage pathway and activity of thymidine kinase 1, it reflects DNA synthesis and thus cell proliferation [4]. ^{18}F -FMISO is a radiolabeled nitroimidazole that accumulates in regions of low oxygen content, and thus provides a measurement of hypoxia [4]. These molecular imaging techniques may provide more insight than RECIST into the assessment of disease response, as high glycolytic index, increased proliferative activity, and regions with elevated hypoxia are hallmarks of malignant tumors [5]. Indeed, PERCIST, or PET response evaluation criteria in solid tumors, was developed as a potentially more sensitive method to assess response to treatments that are cytostatic rather than cytocidal [1]. PERCIST assesses response using the change from baseline in the standardized uptake value normalized to a reference tissue from ^{18}F -FDG PET scans. With the increasing use of these techniques in pre-clinical studies to measure response to targeted therapies [6, 7], investigating reproducibility is important to understand whether changes in imaging metrics during therapeutic interventions reflect changes in underlying biology rather than measurement error [8–10]. The objectives of this study were to characterize static reproducibility of ^{18}F -FDG, ^{18}F -FLT, and ^{18}F -FMISO, as well as dynamic ^{18}F -FLT and ^{18}F -FMISO microPET studies in a murine model of HER2+ breast cancer.

Materials and Methods

Animal and Tumor Xenograft Model

Trastuzumab-resistant breast cancer cells (HR6) (gift from Dr. Carlos Arteaga at Vanderbilt University) were harvested from BT474 xenografts that initially responded but recurred in the presence of maintained trastuzumab as described previously [11]. (This cell line was selected as it is part of an ongoing investigation of trastuzumab resistance in breast tumors.) Female athymic nude mice ($n=13$, 4–6 weeks, Harlan, Indianapolis, IN) were implanted with 0.72 mg, 60-day release, 17β -estradiol pellets (Innovative Research of America, Sarasota, FL). Twenty-four hours later, approximately 10^7 HR6 cells were injected subcutaneously into the right flank. Mice were anesthetized with 2 % isoflurane in pure oxygen for both procedures. A 26-gauge jugular catheter was surgically implanted for radiotracer delivery. The mean tumor volume at start of imaging was 223 mm^3 [range, 103–312 mm^3], and the average mouse weight during imaging was 22.0 g [range, 18–25 g]. All animal procedures were approved by our Institution's Animal Care and Use Committee.

Radiotracer Synthesis

PETNET synthesized ^{18}F -FDG with a specific activity greater than 37 TBq/mmol. ^{18}F -FLT and ^{18}F -FMISO were prepared as a service by the radiochemistry core with average specific activities of 128.76 and 66.6 TBq/mmol, respectively, following standard protocols [12, 13].

Imaging Sessions

Test–retest scans were performed on 12 mice for each radiotracer. All mice except one were imaged with all three radiotracers. (A minimum of 48 h occurred between test–retest scans using different radiotracers.) Mice were anesthetized using 2 % isoflurane in pure oxygen, and animal body temperature was maintained by a heated water pad and heat lamp. List-mode data were collected for 60 min using a microPET Focus 220 system (Concorde Microsystems Inc., Knoxville, TN [14]) before, during, and after injection of 11.0 ± 2.0 MBq of ^{18}F -FDG, ^{18}F -FLT, or ^{18}F -FMISO diluted in 100 μl of sterile saline, and reconstructed into a 64-frame dynamic sequence (5×12 s, 59×60 s). A second scan with a repeat injection was performed approximately 6 h after the first injection. Six hours was chosen such that the radioactivity of the first scan would decay at least three half-lives; it was assumed that minimal tumor changes occurred during this period [8, 9]. A 5-min static scan was acquired just prior to the second injection to correct for residual activity, as a 2-fold decrease in coefficient of variation between repeated scans was previously observed using this method [9].

All microPET images were reconstructed using the 3D ordered-subsets expectation maximization (OSEM3D) (nine subsets) followed by a maximum *a posteriori* (MAP), 18 iterations, beta value 0.038. As the amount of attenuation in microPET studies of mice is small compared to human studies and significant differences between animals are not expected [14], attenuation and scatter correction were not performed.

Animals were fasted 4 h [8] and warmed 1 h [15] prior to ^{18}F -FDG injection. After the first scan, animals recovered from anesthesia, were given access to food for 1 h, then fasted another 4 h prior to the second scan [8]. It was not necessary to fast prior to ^{18}F -FLT imaging as significant differences in radiotracer accumulation between a fed or fasted state have not been observed in mice [9]; we assumed a condition held true for ^{18}F -FMISO, thus animals were not fasted for those studies either.

Image Analysis

Regions of interest (ROIs) were drawn around the entire tumor volume using MATLAB (The MathWorks, Natick, MA, USA). The mean ROI activity from the last 10 min of each dynamic scan was normalized by the injected dose to calculate percent-injected dose per gram (%ID/g). Average activity from tumor voxels was calculated at each dynamic frame to construct time-activity curves (TACs). The ROI activities and TACs from scan 2 were adjusted by subtracting the activity from the 5-min static scan to eliminate residual activity.

Compartmental modeling requires the time rate of change of radiotracer activity concentration in the blood plasma. As the gold-standard for making this measurement (blood sampling [16]) is not feasible for repeat studies in mice due to their small blood volume, we used the method developed by Kim et al. to obtain image-derived input functions (IDIFs) [17]. This method generates TACs from the left ventricle and corrects for partial-volume effects using a recovery coefficient (RC). In our implementation, we obtained TACs from cylindrical ROIs across three slices. A RC of 0.72 was determined from phantom studies acquired under similar acquisition conditions as described previously [18]. An identical procedure was followed for ^{18}F -FMISO. To the best of our knowledge, validation of an

IDIF for ^{18}F -FMISO microPET studies in mice has not been published, a point we return to in the “Discussion” section.

Due to the relatively low spatial resolution, coupled with myocardial spillover and partial-volume effects, the IDIF from user-defined ROIs cannot be used for ^{18}F -FDG. [16]. Several groups have developed mathematical models to derive the input function from ^{18}F -FDG image data [16, 19] or use hybrid TACs from other organs [18]; however, these methods require at least one blood sample. Blood samples were not collected in this study, thus we were unable to assess ^{18}F -FDG dynamic reproducibility.

Kinetic Modeling

We employed a three-compartment model with rate constants K_1 – k_4 , that for ^{18}F -FLT represent the rate of transport between plasma and tissue, rate of outflow from the tissue to plasma, thymidine kinase 1 phosphorylation rate, and the dephosphorylation rate, respectively. Similar definitions apply for ^{18}F -FMISO kinetic parameters except k_3 and k_4 that are the rates of tracer retention and transport between the trapped and reversible compartments, respectively. As sensitivity assessments performed by Muzi et al. showed high covariance between K_1 and k_2 in tumors [20], we combined those parameters and performed model fitting with K_1 , $V_d (=K_1/k_2)$, the initial volume of distribution), k_3 , and k_4 . Data were analyzed using four parameters assuming reversible phosphorylation ($k_4 \neq 0$) and three parameters assuming irreversible phosphorylation ($k_4 = 0$). Modeling also included the blood volume fraction (V_b). Model fitting used a nonlinear least-squares method with the governing differential equations, parameter ranges, and initial guesses described previously (also listed in Table 1) [20]. We also calculated the net influx constant (K_i):

$$K_i = \frac{K_1 \cdot k_3}{\frac{K_1}{V_d} + k_3}. \quad (1)$$

The Akaike Information Criterion (AIC) was calculated to determine the most parsimonious kinetic model for ^{18}F -FLT and ^{18}F -FMISO [21]. For the least-squares case, the AIC with a bias correction for small samples is:

$$\text{AIC} = n \cdot \ln\left(\frac{\text{RSS}}{n}\right) + 2k + \frac{2k(k+1)}{n-k-1}, \quad (2)$$

where RSS is the residual sum of squares, n number dynamic frames, and k number of model parameters. Average AIC was calculated for each model, and reproducibility was assessed for the model with the lowest AIC.

Reproducibility Statistics

Statistical methods for evaluating reproducibility follow Bland and Altman [22] as implemented by Galbraith et al. [23]. For each mouse, the difference between parameter measurements (d) was calculated. The distribution of d was tested for normality with the Shapiro–Wilk test, and Kendall’s tau test was used to determine if d correlated with the mean. A Wilcoxon signed-rank test was performed to test the null hypothesis of no bias between repeated measurements. A significance value of $P < 0.05$ was used.

The following statistical calculations were performed:

1. Mean squared difference (dsd) is computed as the standard deviation of d . The dsd is then used to calculate the 95 % confidence interval (CI), which provides a threshold of measurement error in a group of n mice:

$$CI = \pm \frac{t_{\text{stat}} \cdot \text{dsd}}{\sqrt{n}}, \quad (3)$$

where t_{stat} is the appropriate t statistic corresponding to the sample size. Any change in the population greater than this value would reflect changes in the underlying biology.

2. Within-subject standard deviation (wSD) is:

$$\text{wSD} = \frac{\text{dsd}}{\sqrt{2}}. \quad (4)$$

3. Repeatability coefficient (r) is:

$$r = 2.77 \cdot \text{wSD}, \quad (5)$$

and indicates that the difference between two measurements will be less than this figure for 95 % of observation pairs.

If the mean parameter difference is correlated with the mean (Kendall's tau, $P < 0.05$), the data would need to be transformed to a $\log_{10}$ scale and Eqs. (3) and (5) would need to be modified accordingly [23].

To compare our results with previous reproducibility studies [8, 9], we calculated the coefficient of variation (CV) by dividing the standard deviation of the two measurements by the mean. Statistical analyses were performed using Microsoft Excel (Redmond, WA) and statistical toolbox in MATLAB.

Results

Static Reproducibility

Twelve data sets were analyzed for ^{18}F -FDG; one animal was removed from each of the ^{18}F -FLT and ^{18}F -FMISO data sets due to a large variation in injected activity between scans. Examples of repeated %ID/g parametric images are shown in Fig. 1. No parameter had a mean difference significantly different from normal except ^{18}F -FLT ($P = 0.011$). However, the conclusion that a population does or does not follow a Gaussian distribution may be incorrect for small sample sizes due to insufficient power of the normality test [24], thus reproducibility statistics were still performed for ^{18}F -FLT %ID/g. No parameter had a dependence of d on the mean, thus a logarithmic transformation was not required.

Reproducibility statistics for %ID/g are listed in Table 2. Bland–Altman plots for each radiotracer are displayed in Fig. 2, with each panel plotting the mean %ID/g against the mean difference between repeated scans. The mean difference (solid line) is also shown with 95 % CIs (dotted line) that represent the required change to surpass the expected measurement variability for a group of mice, which is ± 0.62 (14 %), ± 0.23 (5 %), and ± 0.24 (6 %) for ^{18}F -FDG, ^{18}F -FLT, or ^{18}F -FMISO, respectively. Repeatability ranges (dashed line) are ± 1.9 , ± 0.66 , or ± 0.69 for ^{18}F -FDG, ^{18}F -FLT, or ^{18}F -FMISO, respectively. The CVs for %ID/g (mean $\pm$ SD) are 11.4 % $\pm$ 10.6 %, 4.0 % $\pm$ 3.8 %, and 6.0 % $\pm$ 3.8 % for ^{18}F -FDG, ^{18}F -FLT, and ^{18}F -FMISO, respectively.

The percentage values for the 95 % CIs will change with the mean. We note that the absolute values are appropriate when determining measurement error [23]; however, the

percentage values are easier to interpret in a longitudinal study when parameter changes are normalized to baseline as is commonly done. The reader is encouraged to refer to Table 2 for the absolute values; however, the 95 % CIs will be quoted as a percentage when defining the threshold of change in a mouse cohort.

¹⁸F-FLT Dynamic Reproducibility

Two additional data sets from the static reproducibility analysis were excluded due to abnormal TACs for the liver and kidney ($n=1$) and the left ventricle not easily located on the images ($n=1$), resulting in nine usable data sets. The AIC (mean $\pm$ SD) for the four-parameter and three-parameter models were 739 ± 44 and 758 ± 51 , respectively. Reproducibility was assessed using the four-parameter model, as it had the lowest AIC. Figure 3 shows an example of a repeated dynamic ¹⁸F-FLT data set, where the TACs from the left ventricle and tumor are shown in panels a and b, respectively. The four-parameter model fits are also shown, and the kinetic parameter values for each scan were: K_1 , 0.031 and 0.038; V_d , 0.629 and 0.578; k_3 , 0.027 and 0.029; k_4 , 0.091 and 0.061; K_{FLT} , 0.011 and 0.012.

¹⁸F-FLT dynamic reproducibility statistics are listed in Table 3. No parameter was significantly different from normal, and no systematic bias was detected among repeats for any parameter. Additionally, the variability associated with each parameter was independent of the mean, thus no logarithmic transformation was required. Bland–Altman plots for each kinetic parameter are displayed in Fig. 4; similar to Fig. 2, the mean difference is graphed with 95 % CIs and repeatability ranges. V_d , k_3 , and K_{FLT} were the most reproducible parameters, with 95 % CIs of ± 18 , ± 10 , ± 18 %, respectively. Results for K_1 and k_4 were more variable, with 95 % CIs of ± 33 and ± 43 %, respectively. The repeatability coefficients for K_1 , V_d , k_3 , k_4 , and K_{FLT} were ± 0.035 , ± 0.231 , ± 0.007 , ± 0.07 , and ± 0.005 , respectively. To compare dynamic reproducibility results with %ID/g, the CVs (mean $\pm$ SD) for K_1 , V_d , k_3 , k_4 , and K_{FLT} were 25 ± 16 %, 14 ± 9 %, 6 ± 5 %, 28 ± 27 %, and 14 ± 9 %, respectively.

¹⁸F-FMISO Dynamic Reproducibility

Four additional data sets from the static reproducibility analysis were excluded due to abnormal liver and kidney TACs ($n=4$), resulting in seven useable data sets. The AIC (mean $\pm$ SD) for the four-parameter and three-parameter models were 756 ± 42 and 747 ± 39 , respectively. Reproducibility was assessed using the three-parameter model, as it had the lowest AIC. Figure 5 shows an example of repeated dynamic ¹⁸F-FMISO data, where the TACs from the left ventricle and tumor are shown in panels a and b, respectively. The three-parameter model fits are also shown, and the kinetic parameter values for each scan were: K_1 , 0.028 and 0.035; V_d , 0.432 and 0.35; k_3 , 0.014 and 0.02; K_{FMISO} , 0.005 and 0.006.

¹⁸F-FMISO dynamic reproducibility statistics are listed in Table 3. No kinetic parameter was significantly different from normal, and no systematic bias was detected among repeats for any parameter. Additionally, no parameter had a significant dependence of d on the mean, thus no logarithmic transformation was required. Bland–Altman plots for each ¹⁸F-FMISO kinetic parameter are displayed in Fig. 6; again, the mean difference is graphed with 95 % CIs and repeatability ranges. V_d and K_{FMISO} were the most reproducible with the lowest 95 % CIs of ± 16 and ± 14 %, respectively. Results for K_1 and k_3 were more variable, with 95 % CIs of ± 25 and ± 42 %, respectively. The repeatability coefficients for K_1 , V_d , k_3 , and K_{FLT} were ± 0.022 , ± 0.16 , ± 0.017 , and ± 0.003 , respectively. To compare dynamic reproducibility results with %ID/g, the CVs (mean $\pm$ SD) for K_1 , V_d , k_3 , and K_{FMISO} were 22 ± 13 %, 11 ± 7 %, 24 ± 13 %, and 12 ± 13 %, respectively.

Discussion

Reproducibility for individuals or groups can be described by several statistical metrics. The 95 % CIs of the mean difference in Tables 2 and 3 dictate thresholds by which greater changes would reflect tumor physiology rather than measurement variability in groups of similar size to our study. For individuals, the repeatability coefficient represents the required threshold of change. Both the 95 % CI and repeatability coefficient are useful in, for example, a treatment response study where the objective is to quantify changes in parameters after a specific therapy. For example, ^{18}F -FDG %ID/g needs to either increase or decrease by 1.9 for individuals or 14 % in a group analysis to reflect a treatment-induced change, whereas ^{18}F -FLT can be used to detect a change in %ID/g of ± 0.66 in individual tumors or ± 5 % in a group. These measurement uncertainties are not unreasonable, and much higher changes due to therapy have been reported in pre-clinical studies of cancer [6, 7].

Our results are comparable to previously published ^{18}F -FDG and ^{18}F -FLT static reproducibility *via* the CV. Average CV observed by Dandekar et al. [8] and Tseng et al. [9] for whole tumor ROIs were 15.4 and 14 % for ^{18}F -FDG and ^{18}F -FLT, respectively. We observed CVs that were 4 and 10 % less than the previous studies for ^{18}F -FDG and ^{18}F -FLT, respectively. This difference in measurement uncertainty could be attributed to differences in animal preparation and PET data reconstruction. We used jugular vein catheters that are implanted into the animal once, while the previous studies used tail vein catheters which can be quite difficult to place. A difficult catheter placement could cause activity accumulation in the tail due to a less than optimal tracer injection. Additionally, the two previous studies used an OSEM image-reconstruction algorithm. Disselhorst et al. investigated the effect of different reconstruction algorithms on PET image quality, and observed that OSEM3D followed by MAP increased overall image quality by increasing the recovery coefficient, and lowering the spillover ratio and percent standard deviation of the noise. [25]. The improved image quality obtained by using an OSEM3D-MAP algorithm would decrease measurement error; however, the slight differences in animal handling procedures and the uncontrollable variable that is animal stress between scans affects ^{18}F -FDG accumulation and would increase measurement error [15]. This may explain why we observed a much lower CV for ^{18}F -FLT and a similar CV to the previous study for ^{18}F -FDG.

Measurements of tracer uptake normalized to the injected dose are straightforward to acquire; however, changes in tracer clearance from the blood or tumor metabolism are not considered [26]. Performing measurements that use the full dynamic curve of tracer retention/metabolism can overcome some of these issues. However, dynamic acquisitions and compartmental modeling can be complex and can introduce additional sources of noise in the measurement, which in turn, lowers reproducibility as has been observed previously [26] and also in our study.

One potential parameter of interest in a tumor response study with dynamic ^{18}F -FLT is K_{FLT} , which was the second most reproducible dynamic parameter after k_3 . The 95 % CI for K_{FLT} is ± 18 %, suggesting that a change greater than ± 18 % would reflect true treatment differences in a mouse cohort ($n=9$). A larger difference in K_{FLT} due to treatment has been observed in a murine model of prostate cancer [27]. The least reproducible parameter was k_4 , which agrees with the large error previously reported [10], especially when using a 60-min dynamic acquisition [26]. The greater K_1 variability (than K_{FLT}) may be because it is more sensitive to changes in the model input function and/or tracer injection; a similar phenomenon has been reported for ^{18}F -FMISO data [28] as well as dynamic contrast-

enhanced MRI [23]. Using a power injector with a constant injection rate instead of a manual injection may lower variability in this parameter.

The only published ^{18}F -FLT dynamic reproducibility study in mice reported measurement error as the mean $\pm$ standard deviation of the percent difference [10]. Choi et al. concluded that a parameter was reproducible if the standard deviation of the percent difference was $\leq 20\%$. The kinetic parameters that fit this criterion from their study were K_{FLT} (17.5 %) for A431 tumors, and K_1 (19.7 %), V_d (17.4 %), and k_3 (8.6 %) for LLC tumors [10]. The parameters that fit this criterion from our study were V_d (13 %), k_3 (6.4 %), and K_{FLT} (13 %).

The use of static ^{18}F -FMISO measurements to assess treatment response is increasing; however, to the best of our knowledge, a three-compartment analysis of tracer accumulation within a treatment response protocol has not been performed. Thus, we can only hypothesize that the observed measurement errors for ^{18}F -FMISO are reasonable and the kinetic parameters can be used to reliably assess tumor response. The most reproducible parameter is K_{FMISO} , which might be of interest in a treatment response study as this parameter provides both a measure of tracer perfusion as well as tracer retention and/or trapping. For K_{FMISO} , a change greater than $\pm 14\%$ would reflect changes due to treatment in a mouse cohort ($n=7$) instead of measurement variability.

As mentioned previously, validation of an IDIF for ^{18}F -FMISO microPET studies in mice has not been published. However, an IDIF from ROIs within the carotid artery has been used in a human study to assess ^{18}F -FMISO pharmacokinetics in cancer [29]. Using a large blood pool such as the left ventricle or carotid artery for an IDIF is reasonable as these tissues are not normally hypoxic and therefore retention of ^{18}F -FMISO should be minimal.

Study Limitations

We noted in the “Results” section that ^{18}F -FLT %ID/g test–retest measurements failed normality testing; however, interpreting results from normality tests with small samples can be difficult. Indeed, a large P value does not guarantee that the data follow a Gaussian distribution with small sample sizes [24]. Therefore, we proceeded to compute the reproducibility statistics knowing that the statistical methods depend strongly on the assumption of normally distributed data. The two previous static reproducibility studies [8, 9] did not perform the statistical methods outlined in this work, instead concluded that the threshold of required change in an individual mouse would be twice that of the CV. Following their analysis with our data, the threshold of change for ^{18}F -FLT %ID/g is $\pm 8\%$ (twice the CV) in an individual mouse. Our reproducibility results are more conservative, and the repeatability coefficient as a percentage of the mean is $\pm 14\%$.

The anesthesia used in this study was 2 % isoflurane in pure oxygen. A previous study has shown that the use of pure oxygen lowers accumulation of ^{18}F -FMISO, thereby potentially leading to systematic errors in the measurement of tumor hypoxia [30]. However, the measurement error associated with the actual imaging protocol should not change depending on the anesthesia method.

Changes in physiological variables (e.g., blood glucose, cardiac output, and tracer metabolic rates of tumors), scanner electronics efficiency, or radiotracer-specific activities might increase longitudinal variability. The ability to generalize these same day test–retest reproducibility results to the variability of serial studies is important. To minimize these sources of error, it is imperative to develop and strictly follow a well-designed imaging protocol. Furthermore, implementing quantitative methods to normalize imaging data could

prove beneficial if, for example, changes in blood glucose and scanner electronic drift become significant sources of variability during longitudinal studies.

Conclusion

Imaging biomarkers that can detect early changes in specific molecular characteristics of tumors would greatly benefit both pre-clinical and clinical cancer studies. Before an imaging biomarker can be employed to assess treatment response, reproducibility must be assessed. In this study, we assessed static reproducibility for ^{18}F -FDG, ^{18}F -FLT, and ^{18}F -FMISO, as well as dynamic ^{18}F -FLT and ^{18}F -FMISO microPET data. Our test–retest analyses of %ID/g measurements were very reproducible (95 % CI $\pm$ 14 %). Although, the kinetic analysis of ^{18}F -FLT and ^{18}F -FMISO resulted in higher measurement error, the variability associated with the net influx constants (K_{FLT} and K_{FMISO}) is reasonable (95 % CI $\pm$ 19 %). Our study indicates that static and dynamic measurements of these radiotracers can be used in serial studies to measure changes in tracer metabolism following anticancer therapies in mouse models of HER2+ human breast cancer.

Acknowledgments

We thank the National Institutes of Health for funding through R01 CA138599, P50 CA98131, P30 CA68485, U24 CA126588, P50 CA12832, and 1 S10 RR17858. We thank Clare Osborne for imaging assistance; Jarrod True, Dr. Zoe Yu, M.D., and Carlo Malabanon (Vanderbilt University MMPC) for animal care assistance. We also thank Cammie Rinehart for cell culture and animal assistance.

References

1. 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. [PubMed: 19403881]
2. Guarneri V, Barbieri E, Dieci MV, Piacentini F, Conte P. Anti-HER2 neoadjuvant and adjuvant therapies in HER2 positive breast cancer. *Cancer Treat Rev.* 2010; 36(Suppl 3):S62–6. [PubMed: 21129613]
3. Harry VN, Semple SI, Parkin DE, Gilbert FJ. Use of new imaging techniques to predict tumour response to therapy. *Lancet Oncol.* 2010; 11:92–102. [PubMed: 20129132]
4. Dunphy MP, Lewis JS. Radiopharmaceuticals in preclinical and clinical development for monitoring of therapy with PET. *J Nucl Med.* 2009; 50(Suppl 1):106S–21S. [PubMed: 19380404]
5. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell.* 2011; 144:646–74. [PubMed: 21376230]
6. Brepoels L, Stroobants S, Verhoef G, et al. (18)F-FDG and (18)F-FLT uptake early after cyclophosphamide and mTOR inhibition in an experimental lymphoma model. *J Nucl Med.* 2009; 50:1102–9. [PubMed: 19525456]
7. Oehler C, O'Donoghue JA, Russell J, et al. 18F-fluoromisonidazole PET imaging as a biomarker for the response to 5,6-dimethylxanthenone-4-acetic acid in colorectal xenograft tumors. *J Nucl Med.* 2011; 52:437–44. [PubMed: 21321262]
8. Dandekar M, Tseng JR, Gambhir SS. Reproducibility of 18F-FDG microPET studies in mouse tumor xenografts. *J Nucl Med.* 2007; 48:602–7. [PubMed: 17401098]
9. Tseng JR, Dandekar M, Subbarayan M, et al. Reproducibility of 3'-deoxy-3'-(18)F-fluorothymidine microPET studies in tumor xenografts in mice. *J Nucl Med.* 2005; 46:1851–7. [PubMed: 16269599]
10. Choi SJ, Kim SY, Kim SJ, et al. Reproducibility of the kinetic analysis of 3'-deoxy-3'-(18)F-fluorothymidine positron emission tomography in mouse tumor models. *Nucl Med Biol.* 2009; 36:711–9. [PubMed: 19720283]
11. Ritter CA, Perez-Torres M, Rinehart C, et al. Human breast cancer cells selected for resistance to trastuzumab *in vivo* overexpress epidermal growth factor receptor and ErbB ligands and remain

- dependent on the ErbB receptor network. *Clin Cancer Res.* 2007; 13:4909–19. [PubMed: 17699871]
12. Choi SJ, Kim JS, Kim JH, et al. [18F]3'-deoxy-3'-fluorothymidine PET for the diagnosis and grading of brain tumors. *Eur J Nucl Med Mol Imaging.* 2005; 32:653–9. [PubMed: 15711980]
 13. Tang G, Wang M, Tang X, Gan M, Luo L. Fully automated one-pot synthesis of [18F]fluoromisonidazole. *Nucl Med Biol.* 2005; 32:553–8. [PubMed: 15982586]
 14. Tai YC, Ruangma A, Rowland D, et al. Performance evaluation of the microPET focus: a third-generation microPET scanner dedicated to animal imaging. *J Nucl Med.* 2005; 46:455–63. [PubMed: 15750159]
 15. Fueger BJ, Czernin J, Hildebrandt I, et al. Impact of animal handling on the results of 18F-FDG PET studies in mice. *J Nucl Med.* 2006; 47:999–1006. [PubMed: 16741310]
 16. Fang YH, Muzic RF Jr. Spillover and partial-volume correction for image-derived input functions for small-animal 18F-FDG PET studies. *J Nucl Med.* 2008; 49:606–14. [PubMed: 18344438]
 17. Kim SJ, Lee JS, Im KC, et al. Kinetic modeling of 3'-deoxy-3'-18F-fluorothymidine for quantitative cell proliferation imaging in subcutaneous tumor models in mice. *J Nucl Med.* 2008; 49:2057–66. [PubMed: 18997037]
 18. Tantawy MN, Peterson TE. Simplified [18F]FDG image-derived input function using the left ventricle, liver, and one venous blood sample. *Mol Imaging.* 2010; 9:76–86. [PubMed: 20236605]
 19. Ferl GZ, Zhang X, Wu HM, Kreissl MC, Huang SC. Estimation of the 18F-FDG input function in mice by use of dynamic small-animal PET and minimal blood sample data. *J Nucl Med.* 2007; 48:2037–45. [PubMed: 18006615]
 20. Muzi M, Mankoff DA, Grierson JR, et al. Kinetic modeling of 3'-deoxy-3'-fluorothymidine in somatic tumors: mathematical studies. *J Nucl Med.* 2005; 46:371–80. [PubMed: 15695799]
 21. Burnham KP, Anderson DR. Multimodal inference: understanding AIC and BIC in model selection. *Socio Meth Res.* 2004; 33:261–304.
 22. Bland JM, Altman DG. Measuring agreement in method comparison studies. *Stat Meth Med Res.* 1999; 8:135–60.
 23. Galbraith SM, Lodge MA, Taylor NJ, et al. Reproducibility of dynamic contrast-enhanced MRI in human muscle and tumours: comparison of quantitative and semi-quantitative analysis. *NMR Biomed.* 2002; 15:132–42. [PubMed: 11870909]
 24. Motulsky, H. *Analyzing data with graphpad prism.* GraphPad Software Inc; San Diego: 1999.
 25. Disselhorst JA, Brom M, Laverman P, et al. Image-quality assessment for several positron emitters using the NEMA NU 4-2008 standards in the Siemens Inveon small-animal PET scanner. *J Nucl Med.* 2010; 51:610–7. [PubMed: 20237025]
 26. Shields AF, Lawhorn-Crews JM, Briston DA, et al. Analysis and reproducibility of 3'-Deoxy-3'-[18F]fluorothymidine positron emission tomography imaging in patients with non-small cell lung cancer. *Clin Cancer Res.* 2008; 14:4463–8. [PubMed: 18628460]
 27. Lee SJ, Kang HY, Kim SY, et al. Early assessment of tumor response to JAC106, an anti-tubulin agent, by 3'-deoxy-3'-[(1)F] fluorothymidine in preclinical tumor models. *Eur J Nucl Med Mol Imaging.* 2011; 38:1436–48. [PubMed: 21484374]
 28. Wang W, Georgi JC, Nehmeh SA, et al. Evaluation of a compartmental model for estimating tumor hypoxia via FMISO dynamic PET imaging. *Phys Med Biol.* 2009; 54:3083–99. [PubMed: 19420418]
 29. Wang W, Lee NY, Georgi JC, et al. Pharmacokinetic analysis of hypoxia (18)F-fluoromisonidazole dynamic PET in head and neck cancer. *J Nucl Med.* 2010; 51:37–45. [PubMed: 20008982]
 30. Kersemans V, Hueting R, Cornelissen B, et al. Confounding factors arising from the use of anesthesia in preclinical imaging studies of 18F-FMISO as a diagnostic marker of hypoxia. *J Nucl Med: Abstract Book Supplement.* 2011; 52:72P–3P.

Fig. 1. Coronal sections of HR6 xenografts (*T*) showing static reproducibility of ^{18}F -FDG (*a*, *b*), ^{18}F -FLT (*c*, *d*), and ^{18}F -FMISO (*e*, *f*). Please note how the HER2+ xenografts are clearly visible for accurate tumor segmentation.

Fig. 2. Bland–Altman plots displaying mean %ID/g plotted against mean parameter difference between repeat scans for ^{18}F -FDG (a), ^{18}F -FLT (b), and ^{18}F -FMISO (c). The mean difference (solid line) is shown with 95 % confidence intervals (dotted lines) for the population. Repeatability is also shown (dashed lines), which represents the threshold of required change in an individual mouse.

Fig. 3.

Representative example of ^{18}F -FLT dynamic reproducibility. Left ventricle time-activity curves for both scans (a) are shown over the entire 60-min acquisition, with the *inset* figure displaying the first 5 min. Tumor time-activity curves from each scan (b) are shown with curve fits from the most parsimonious kinetic model (four-parameter fit, determined by AIC analysis).

Fig. 4. Bland–Altman plots displaying mean of each ^{18}F -FLT kinetic parameter plotted against the mean difference between scans. The mean difference (*solid line*) is shown with 95 % confidence intervals (*dotted lines*) for the population. Repeatability is also shown (*dashed lines*) for each parameter, which represents the threshold of required change in an individual mouse.

Fig. 5. Representative example ^{18}F -FMISO dynamic reproducibility. Left ventricle time-activity curves for both scans (a) are shown over the entire 60-min acquisition, with the *inset* figure displaying the first 5 min. Tumor time-activity curves of each scan (b) are shown with curve fits from the most parsimonious kinetic model (three-parameter fit, determined by AIC analysis).

Fig. 6. Bland–Altman plots displaying mean of each ^{18}F -FMISO kinetic parameter plotted against the mean difference between scans. The mean difference (*solid line*) is shown with 95 % confidence intervals (*dotted lines*) for the population.

Table 1

Initial guesses and parameter ranges for kinetic modeling

Parameter (units)	Initial guess	Parameter range
K_1 (ml/min/g)	0.1	0.0–0.5
$V_d (K_1/k_2)$ (ml/g)	1	0.1–5.0
k_3 (min ⁻¹)	0.1	0.001–1.0
k_4 (min ⁻¹)	0.02	0.001–0.2
V_b (ml/g)	0.05	0.0–0.1

Table 2

Reproducibility analysis for %ID/g

Radiotracer	Mean	Mean difference	95 % CI for mean difference	wSD	Repeatability	CV (mean±SD)
¹⁸ F-FDG	4.3	-0.01	±0.62 (14 %)	0.69	1.9	11.4±10.6 %
¹⁸ F-FLT ^a	4.6	-0.09	±0.23 (5 %)	0.24	0.66	4.0±3.8 %
¹⁸ F-FMISO	3.8	-0.16	±0.24 (6 %)	0.25	0.69	6.0±3.8 %

CI confidence interval, wSD within-subject standard deviation, CV coefficient of Variation

^aNormality cannot be assumed

Table 3

Dynamic reproducibility for ^{18}F -FLT and ^{18}F -FMISO

Parameter	Mean	Mean difference	95 % CI for mean difference	wSD	Repeatability	CV (mean±SD)
^{18}F -FLT with four-parameter fit						
K_1	0.042	-0.012	±0.014 (33 %)	0.012	0.035	25±16 %
$V_d (K_1/k_2)$	0.501	0.013	±0.091 (18 %)	0.083	0.231	14±9 %
k_3	0.031	-0.001	±0.003 (10 %)	0.003	0.007	6±5 %
k_4	0.065	-0.006	±0.028 (43 %)	0.025	0.070	28±27 %
K_{FLT}	0.011	-0.001	±0.002 (18 %)	0.002	0.005	14±9 %
^{18}F -FMISO with three-parameter fit						
K_1	0.040	-0.007	±0.010 (25 %)	0.008	0.022	21±13 %
$V_d (K_1/k_2)$	0.473	0.042	±0.075 (16 %)	0.058	0.160	11±7 %
k_3	0.019	0.003	±0.008 (42 %)	0.006	0.017	22±13 %
K_{FMISO}	0.007	0.001	±0.001 (14 %)	0.001	0.003	11±12 %

CI confidence interval, wSD within-subject standard deviation, CV coefficient of variation