Repeatability of quantitative $^{18}$F-fluoromethylcholine PET/CT studies in prostate cancer

Daniela E. Oprea-Lager$^1$, Gem Kramer$^1$, Peter M. van de Ven$^2$, Alfons J. M. van den Eertwegh$^3$, Reindert J. A. van Moorselaar$^4$, Patrick Schober$^5$, Otto S. Hoekstra$^1$, Adriaan A. Lammertsma$^1$ and Ronald Boellaard$^1$

Departments of $^1$Radiology & Nuclear Medicine, VU University Medical Center, Amsterdam, the Netherlands, $^2$Epidemiology & Biostatistics, VU University Medical Center, Amsterdam, the Netherlands, $^3$Medical Oncology, VU University Medical Center, Amsterdam, the Netherlands, $^4$Urology, VU University Medical Center, Amsterdam, the Netherlands, $^5$Anesthesiology, VU University Medical Center, Amsterdam, the Netherlands

Corresponding author:
Daniela E. Oprea-Lager

Address: Department of Radiology & Nuclear Medicine, VU University Medical Center, PO Box 7057, 1007 MB Amsterdam, the Netherlands.

Telephone number: +31(0)204444366; Fax number: +31(0)204443090

E-mail: d.oprea-lager@vumc.nl

Word count: 5191

Short running foot line: Repeatability $^{18}$F-fluoromethylcholine
ABSTRACT

Repeatable quantification is essential when using $^{18}$F-fluoromethylcholine ($^{18}$F-FCH) PET/CT for monitoring treatment response in prostate cancer (PC). It has been shown that standardized uptake value (SUV) normalized to the area under the curve (AUC) of the blood activity concentration ($\text{SUV}_{\text{AUC}}$) provides better correlation with full kinetic analysis than standard SUV. However, precision of $\text{SUV}_{\text{AUC}}$ is not known yet. The purpose of this study was to assess repeatability of various semi-quantitative $^{18}$F-FCH parameters in PC. **Methods:** Twelve patients (64±8 years) with metastasized PC underwent 2 sets of $^{18}$F-FCH PET/CT scans, on consecutive days. Each set consisted of a 30 minutes dynamic PET/CT scan of the chest, after intravenous administration of 200 MBq $^{18}$F-FCH, followed by a whole body PET/CT at 40 minutes. Dynamic scan was used to derive AUC of the blood activity concentration. Lesion uptake was derived from the whole body scan using various types of volumes of interest: maximum, peak and mean. Each of these parameters was normalized to injected activity/weight, blood AUC and blood concentration itself at 40 minutes, resulting in several SUV, $\text{SUV}_{\text{AUC}}$ and $\text{SUV}_{\text{TBR}}$ values. Test-retest repeatability of these metrics, metabolic tumor volume (MTV) and total lesion choline uptake (TLCU), respectively, were studied. The level of agreement between test-retest data and reliability was assessed using Bland-Altman plots, repeatability coefficients (RC) and intraclass correlation coefficients. **Results:** A total of 67 choline avid metastases were identified, 44 bone and 23 lymph node lesions. In case of SUV$_{\text{max}}$, RC for SUV, $\text{SUV}_{\text{AUC}}$ and $\text{SUV}_{\text{TBR}}$ were 26% (ICC=0.95), 31% (ICC=0.95), and 46% (ICC=0.89), respectively. Similar values were obtained for SUV$_{\text{peak}}$ and SUV$_{\text{mean}}$. Repeatability of $\text{SUV}_{\text{AUC}}$ was comparable with that of SUV, for maximum, peak and mean values. Tissue type and tumor localization did not affect repeatability. MTV $<$4.2 cm$^3$ had larger variability than larger volumes (RC 45% versus 29%, $p=0.048$). Repeatability did not differ between lesions with SUV$_{\text{peak}}$ above or below the median value of 8.3 (RC 19% versus 28%, $p=0.264$). **Conclusion:** The repeatability of $\text{SUV}_{\text{AUC}}$ was comparable to that of standard SUV. RC of various semi-quantitative $^{18}$F-FCH parameters (SUV, MTV, TLCU) were $\sim$35%. Larger differences are likely to represent treatment effects.

**Keywords:** repeatability, positron emission tomography (PET), prostate cancer, choline, standardized uptake value (SUV)
INTRODUCTION

Prostate cancer (PC) is the second most common cancer in men worldwide and was the third most diagnosed malignancy in Europe in 2012, with 92,000 deaths (1,2). This androgen-dependent neoplasm is characterized by good initial response to (anti)hormonal therapy and an unpredictable latent castration-resistant status (3). At the beginning of this decennium, molecular profiling studies have improved knowledge about the heterogeneous biological behavior of PC. It was found that even in the presence of a castrate range (<1.7 nmol/L) of serum testosterone in castration-resistant PC patients, a proportion of these tumors remains dependent on androgen-receptor signaling for growth (4). Five potential mechanisms of development of castration-resistant PC were described, based on ligand and androgen-receptor dependence (5).

Nowadays, several therapeutic options against castration-resistant PC prevail, including cytotoxic (docetaxel, cabazitaxel), hormonal (abiraterone, enzalutamide), immunotherapeutic (sipuleucel-T) and bone-targeting (Ra-223 dichloride) agents (6–13). However, despite this variety of new agents with demonstrated improvement in life expectancy, proper sequencing (e.g., modality, timing) in individual patients with metastatic PC is unclear (14). Since therapeutic options vary greatly with stage and grade of the disease, specific patterns of metastatic spread (i.e., hematogeneous and/or lymphatic) and dominant phenotype, accurate diagnostic “instruments” for response evaluation are essential (15,16).

Non-invasive, hybrid positron emission tomography/computed tomography (PET/CT) is a valuable diagnostic tool by combining metabolic and anatomic information in vivo (17). Encouraging results have been reported on the usefulness of radiolabeled-choline PET/CT in PC (18,19). Apart from its main recognized application in restaging disease in case of biochemical relapse (20,21), \(^{18}\text{F}-\text{fluoromethylcholine}\) (\(^{18}\text{F}-\text{FCH}\)) might also qualify as a biomarker of response to therapy. Since conceptually, choline uptake represents viable tumor cells, tracer uptake changes over time might serve as an improved read-out of treatment efficacy.

In vitro experiments have shown promising results in the use of radiolabeled-choline to monitor anti-androgen treatment or chemotherapy (22,23). Recently, simplified quantitative methods for \(^{18}\text{F}-\text{FCH}\) have been developed and validated (24); standardized uptake value (SUV) normalized to the area under the curve (AUC) of the blood activity
concentration (SUV\textsubscript{AUC}) provides better correlation with full kinetic analysis than standard SUV (24). However, precision of SUV\textsubscript{AUC}, and also that of SUV itself, is as yet unknown.

The purpose of this study was to prospectively assess repeatability of semi-quantitative \textsuperscript{18}F-FCH PET/CT parameters in PC, also including metabolic tumor volume (MTV) and total lesion choline uptake (TLCU). Such knowledge is essential for proper interpretation of signal changes of \textsuperscript{18}F-FCH over time, thus improving personalized therapy strategies for PC patients.

**MATERIALS AND METHODS**

**Patients**

Twelve patients with histologically proven PC (n=4 castration-resistant PC), with lymphatic and/or hematogeneous metastases, were included prospectively. Inclusion criteria required at least 2 metastases (diameter $\geq$1.5cm) detected by conventional imaging performed $\leq$3 months prior to PET/CT and ability to remain supine for 60 minutes. Exclusion criteria were claustrophobia and coexistence of multiple malignancies. The study was approved by the Medical Ethics Review Committee of VU University Medical Center. Prior to inclusion, each patient signed a written informed consent, after receiving verbal and written explanation.

Personal and demographic data [age, body weight, Gleason score, prostate-specific antigen (ng/ml) at the time of PET/CT, and information on previous therapy] were collected, as well as metastatic lesion characteristics [location (intrathoracically/intra-abdominal/ pelvic), number and type (bone/lymph nodes)]. Values are presented as mean±SD.

**Data Acquisition**

All patients underwent \textsuperscript{18}F-FCH [synthesis details see (24)] PET/CT scans on two consecutive days. The minimal interval of time between last treatment and first PET/CT scan was 19 days. Patient preparation was similar to that required for \textsuperscript{18}F-2-fluoro-2-deoxy-D-glucose (\textsuperscript{18}F-FDG) (25). Patients were scanned using a Gemini TF-64 PET/CT scanner (Philips Medical Systems, Cleveland, Ohio, USA).

Each patient received a low dose CT (30 mAs, 120 kV), followed by a 30 minutes dynamic PET scan of the chest [for details see (24)], centered over a large blood pool structure (e.g., ascending aorta), to obtain an image derived
input function. At the start of the dynamic $^{18}$F-FCH scan a bolus injection of 205±9 and 206±7 MBq $^{18}$F-FCH (days 1 and 2, respectively) was administered intravenously using an automated injector (Medrad, Pittsburgh, USA), which was flushed with 40mL of saline (5mL at 0.8mL·s$^{-1}$ followed by 35mL at 2mL·s$^{-1}$). Dynamic PET data were normalized and corrected for decay, scatter, random coincidences and photon attenuation, and reconstructed into 25 frames (1x10, 8x5, 5x20, 5x60, 3x150, 3x300 s) with a matrix size of 144x144x45 voxels (4x4x4 mm$^3$) using a 3-dimensional row action maximum likelihood reconstruction algorithm (3D-RAMLA).

After a standard (5 min) micturition break to warrant a proper visual assessment of the pelvic region, a whole body (WB) $^{18}$F-FCH scan (mid-thigh to skull base) was performed 40 min post injection. Following this PET acquisition (10 bed positions; each of 2 min), a second low dose CT (50 mAs, 120 kV) was acquired for anatomical correlation and attenuation correction. WB data were corrected for dead time, decay, scatter and randoms, and reconstructed into 34 frames (1x10, 8x5, 4x10, 3x20, 5x30, 5x60, 4x150, 4x300 s) with a matrix size of 144x144 and voxels of 4x4x4 mm$^3$, using time of flight iterative reconstruction (BLOB-OS-TF). The transaxial spatial resolution was ~5 mm full width at half maximum in the center of the field of view, similarly to that of the dynamic scan.

**Data Analysis**

Reconstructed images were transferred to off-line workstations for further analysis. Data were analyzed on a volume of interest (VOI) basis (26). The dynamic scan was used to derive the AUC of the blood activity concentration, by defining a cylindrical VOI with 1.5 cm diameter extending over 5 consecutive axial planes within the lumen of the ascending aorta. Next, lesion uptake (defined as $^{18}$F-FCH accumulation exceeding local background and incompatible with physiological tracer biodistribution) was derived from the WB scan.

The term MTV was used to indicate tumor volumes that were derived directly from the PET studies, quantified as the VOI size (26). VOIs were defined by a semiautomatic delineation tool, applying a background adapted 50% of maximum isocontour (i.e., contour value equals 50% of maximum + background). For each VOI, maximum, peak and mean uptakes were calculated. In addition, each of these parameters was normalized to injected activity/body weight (SUV), AUC of blood activity concentration (SUV$_{AUC}$), as derived from the dynamic scan, and blood concentration at 40 minutes post injection (SUV$_{TBR}$), as derived from the WB scan. For SUVmean associated parameters, also total lesion choline uptake (defined as SUVmean x MTV) was calculated, resulting in TLCU, TLCU$_{AUC}$ and TLCU$_{TBR}$, respectively. Test-retest repeatability of all metrics was calculated using both standard
repeatability coefficients (RC) (defined as $1.96 \times \text{SD of the difference between test-retest}$) and relative RC (test-retest difference in percentage).

**Statistical Analysis**

Data were analysed using SPSS version 15.0. Test-retest repeatability was quantified using intraclass correlation coefficients (ICC; based on absolute agreement), RC and displayed graphically using Bland-Altman and/or box plots. ICCs were calculated for each SUV measure as the ratio of between-lesion variability and total variability (both between-lesion and within-lesion). Variance components for patient, lesion within patient and (repeated) measurements within lesions were estimated using mixed models. Total variability was calculated as the sum of these three variance components. Between-lesion variability was calculated as the sum of the variance components for patient and lesion-within patient. Confidence intervals for ICCs were determined using the delta-method. Reliability coefficients were based on the relative difference:

$$100\% \times \frac{\text{SUV}_{\text{test}} - \text{SUV}_{\text{retest}}}{0.5 \times (\text{SUV}_{\text{test}} + \text{SUV}_{\text{retest}})}$$

The method in Bland (27) was used to take into account correlation of measurements between different lesions within the same patient (intra-patient), when calculating RC. Differences in repeatability of the different SUV measures (max, peak, mean) and methods of normalization (SUV, SUV\text{\textsubscript{AUC}} and SUV\text{\textsubscript{TBR}}) were assessed by comparing the variances of the relative test-retest differences, using the Pitman-Morgan test (28) for correlated variances. Differences in repeatability of SUV measures related to location, metastatic lesion type and size were assessed by comparing the variances of the relative test-retest differences between the subgroups of lesions, using Levene’s test (29). Overall type I error, within each set of comparisons, was controlled at 5%, using a Bonferroni correction.

**RESULTS**

**Patients**

Patient characteristics (Table 1) were: age 64±8 years, weight 88±9 kg, Gleason score 7 (n=3) or more (n=9) and a median prostate-specific antigen at the time of PET/CT of 46 ng/ml (range 2-226). Patients had been treated previously by prostatectomy and lymph node dissection (n=3), (anti)hormone therapy (n=7), external beam
radiotherapy on the prostate (n=1), external beam radiotherapy in combination with (anti)hormone therapy (n=5), chemotherapy (n=4) or immunotherapy (n=1).

67 metastases were identified at PET, with a median of 6 per patient (range 3-8), 44 of which were bone and 23 lymph node metastases. Twelve metastases were localized above the diaphragm, the others were intra-abdominal and/or in the pelvic region (n=55). The malignant nature of the metastases was confirmed radiologically, based on progression of pre-existent lesions and/or new metastatic sites. VOI size (median/interquartile range) was 4.9/7.6 cm³ with a lesion SUVpeak (median/interquartile range) of 8.3/5.2. Repeatability data was analyzed using a volume threshold of 4.2 cm³, based on a repeatability study of metabolic tumor volume with $^{18}$F-FDG and $^{18}$F-fluorothymidine ($^{18}$F-FLT) in lung cancer (26). In that study, changes of $>$37% for $^{18}$F-FDG in lesions larger than 4.2 cm³ were found to represent a biologic effect. This volume threshold corresponds by approximation to a diameter of 2 cm (for spherical metastatic lesions), which equals about 4 times the spatial resolution of PET below which quantification, VOI definition and detectability are hampered by partial-volume effects.

Repeatability of semi-quantitative $^{18}$F-FCH parameters

Repeatability of each semi-quantitative $^{18}$F-FCH parameters (i.e., SUV, MTV and TLCU) was studied as a function of uptake (median SUVpeak 8.3), MTV (larger or smaller than 4.2 cm³), metastatic tissue type (bone/lymph node) and location (intrathoracically versus abdominal/pelvic), respectively. These test-retest aspects are discussed below.

Repeatability of SUV

In case of SUVmax, RC (for relative differences) were 26%, 31% and 46%, respectively. Similar values were observed for SUVpeak and SUVmean. ICCs for SUV and SUV_{AUC} were all approximately 0.95, whereas ICCs for SUV_{TBR} were 0.89 (Table 2).

Nine pairwise comparisons for different methods of SUV normalization were performed to estimate variances of relative test-retest differences. After correcting for multiple comparisons (Bonferroni corrected significant difference at 5% level $p<0.0056$ and at 1% level $p<0.0011$, respectively), only SUV_{TBR} parameters were found to have
consistently larger variances ($p<0.001$) (supplemental Table 1). The relative differences (in percentage) between test-retest data and the mean values for the different types of SUVs (max, mean and peak) and their normalizations (SUV, $SUV_{AUC}$ and $SUV_{TBR}$) are presented in Figure 1. The repeatability of $SUV_{AUC}$ for max, peak and mean values were comparable with those of the corresponding SUV measures (supplemental Table 2).

Repeatability did not differ between lesions with a $SUV_{peak}$ above or below the median value of 8.3 (RC 19% versus 28%, $p=0.264$) (Fig. 2, and supplemental Fig.1). These values were comparable with those of $SUV_{peakAUC}$ [RC 23% ($SUV_{peak}>8.3$) versus 31% ($SUV_{peak}<8.3$), $p=0.136$]. Moreover, repeatability of $SUV_{peak}$ and $SUV_{mean}$ was independent of MTV (Fig. 3) as well as of tissue type and tumor location (supplemental Fig. 2 and supplemental Table 3).

**Repeatability of MTV**

For MTV, test-retest difference was 0.03±1.63 and the relative test-retest difference 36% (Fig. 4). Repeatability of MTV was independent of $SUV_{peak}$ [RC 34.25% ($SUV_{peak}>8.3$) versus 36.43% ($SUV_{peak}<8.3$), $p=0.933$]. MTV <4.2 cm$^3$ had larger variability than larger volumes (RC 45% versus 29%, $p=0.048$) (supplemental Fig. 1). Repeatability of MTV did not differ between bone/lymph nodes (RC 34% versus 36.4%, $p=0.684$) and location [RC 36.7% (intrathoracically) versus RC 34.4% (intra-abdominal/pelvic), $p=0.820$].

**Repeatability of TLCU**

Repeatability of the TLCU and TLCU$_{AUC}$ were comparable (RC 33% versus 31%, $p=0.954$), while TLCU$_{TBR}$ showed larger variance of 51% ($p<0.001$) (supplemental Fig. 3). No significant difference was found between lesions with a total choline uptake below or above the median value of 30.9 [RC 40.9% (<30.9) versus 23.1% (>30.9), $p=0.093$].

Repeatability of TLCU was independent of uptake (RC 31% ($SUV_{peak}<8.3$) versus 34% ($SUV_{peak}>8.3$), $p=0.139$) and MTV [RC 42.1% (<4.2 cm$^3$) versus 25.3% (>4.2 cm$^3$), $p=0.037$], respectively. Repeatability of TLCU was also independent of tissue type [RC 30.8% (bone) versus 35.7% (lymph nodes), $p=0.241$] and location [RC 35.5% (intrathoracically) versus RC 32.4% (intra-abdominal/pelvic), $p=0.778$].
An overview of all RC of the semi-quantitative $^{18}$F-FCH parameters (SUV, MTV and TLCU), as a function of uptake, MTV and metastatic tissue type/location, is presented in supplemental Tables 4 and 5. Individual patient RC for all semi-quantitative parameters are provided in supplemental figures 4, 5 and 6. Comparable RC across the majority of subjects were observed. Patient 1 had the poorest RC regarding SUV (~35% on all 9 combinations of SUV types and normalizations). Patients 2, 4 and 7 had the poorest RC regarding MTV (all ~51%) and TLCU (55%, 51% and 67%, respectively).

**DISCUSSION**

In a previous study we investigated the $^{18}$F-FCH kinetics in metastatic PC (24), and demonstrated that SUV cannot be used to estimate $^{18}$F-FCH uptake. Whole-blood activity concentration SUV ($\text{SUV}_{\text{AUC}}$), based on two consecutive PET scans was proposed as a clinically feasible alternative. In the present study, we prospectively assessed the repeatability of quantitative $^{18}$F-FCH PET/CT parameters in patients with PC, and found that repeatability of $\text{SUV}_{\text{AUC}}$ is comparable to that of standard SUV and that $^{18}$F-FCH PET/CT uptake differences of 30% or more are likely to represent treatment effects.

Test-retest repeatability is essential for clinical implementation of any parameter of response assessment. Due to the heterogeneous biological behaviour of PC (4) and in the light of rapidly evolving treatment modalities (14), biological markers are needed that adequately monitor response to therapy. Standard treatment response evaluation criteria in solid tumors do not apply to metastatic PC, further complicating the issue of treatment response (30). A newly proposed system for measuring functional response with $^{18}$F-FDG PET/CT, PET Response Criteria In Solid Tumours (PERCIST), might also apply to radiolabeled choline (31). Nevertheless, before $^{18}$F-FCH PET/CT can be implemented as a biomarker for response evaluation in PC, the repeatability of the tracer should be known (32).

To the best of our knowledge, repeatability of $^{18}$F-FCH measurements in metastatic PC has not been assessed previously. Pegard et al. (33) addressed the reproducibility of observer interpretation (i.e., visual evaluation and classification of foci with increased uptake as being malignant or benign) of $^{18}$F-FCH PET/CT examinations in patients with biochemically recurrent PC. The authors described good concordance when evaluating bone metastases.
and abdominal/pelvic lymphatic recurrences in previously treated patients. A limited usefulness was found at the prostate level in untreated patients.

Observed repeatabilities were within the range seen with other commonly used radiotracers, such as \(^{18}\)F-FDG and \(^{18}\)F-FLT. When analyzing \(^{18}\)F-FDG uptake changes, the generally accepted PERCIST response classification system assumes a true biological change when SUVpeak changes exceed 30%, in combination with 0.8 unit change of the absolute SUVpeak value (31). In a meta-analysis on the repeatability of \(^{18}\)F-FDG uptake measurements in tumors, de Langen et al. (34) identified 8 eligible studies. SUVmean was found to have better repeatability than SUVmax. A minimal relative change of 20% in combination with 1.2 unit change of SUVmean was presumed to represent a biological change. Comparable results were reported by Rockall et al. (35) in a study on repeatability of quantitative \(^{18}\)F-FDG PET/CT in recurrent ovarian carcinoma. RC suggested that a decrease in \(^{18}\)F-FDG uptake (SUV) up to 20% from baseline and a decrease in tumor size up to 15% could be used to determine (early) tumor response. In a study addressing repeatability and reproducibility of \(^{18}\)F-FDG and \(^{18}\)F-FLT PET tumor volume measurements, Hatt et al. (36) found comparable percentage differences for these two tracers datasets. Differences larger than 30% were considered indicative for treatment response evaluation.

Frings et al. (26) analyzed the repeatability of MTV with \(^{18}\)F-FDG and \(^{18}\)F-FLT in lung cancer. Repeatability was better for larger tumors. Changes of >37% for \(^{18}\)F-FDG in lesions larger than 4.2 cm\(^3\) represented a biologic effect. We obtained comparable results when using this MTV threshold for analysing \(^{18}\)F-FCH test-retest data. A larger variability was found in small metabolic volumes (<4.2 cm\(^3\)), suggesting that a similar lower threshold for MTV in treatment response evaluation studies should be used. In case of TLCU, RC were not significantly different for lesions smaller or larger than 4.2 cm\(^3\) (after correcting for multiple comparisons; Bonferroni corrected significance level \(p<0.0056\)). However, the uncorrected \(p\) equalled 0.037, suggesting a trend towards poorer repeatability for smaller lesions.

Two recently published papers explored the prognostic value of metabolic parameters when using \(^{18}\)F-FCH PET/CT in biochemical recurrent PC (37) or in castration-resistant PC (38). In a multivariate analysis, Colombié et al. (37) identified age<70 years, SUVmean ≥3 and a standardized metabolic activity ≥23, as independent prognostic factors for disease-free survival. In a prospective study, Kwee et al. (38) found that WB tumor indices based on quantifying net metabolically active tumor volume and total lesion activity \(^{18}\)F-FCH PET/CT were predictive of
overall survival. In our study we used comparable metrics, with an emphasis on repeatability of the use of $^{18}$F-FCH as a potential biomarker for response evaluation in PC.

RC for individual patients were comparable across the majority of subjects. However, patient 1 had the poorest RC regarding SUV, due to a small (1.5cm short axis diameter) at the right ventral edge of a vertebral body (Th10), close to the liver. We hypothesize that the relatively large test-retest difference (~35%) of this lesion was caused by incorrect scatter correction, due to high image-derived blood activity concentrations near physiologically avid $^{18}$F-FCH structures. This might lead to large quantification errors with image derived input function obtained from blood VOI in these areas (24). Three patients (2, 4 and 7) had larger RC for MTV and TLCU than the other subjects. This is likely explained by difficulty in lesion segmentation, such that errors in MTV are also propagated into poorer repeatability for TLCU. Besides, all these patients presented with small metastatic lesions (~1.5cm) with slightly increased $^{18}$F-FCH uptake.

With respect to the proposed 30% cut-off value for SUV measures, the test-retest differences pooled over all lesions or those with SUV >8.3 were ~20% (Fig. 2). However, for lesions with SUV <8.3, RC were ~30% (supplemental Table 4). Since in clinical practice, most patients will have a combination of metastatic lesions with SUV above or below 8.3, we have decided to adopt the more conservative value of 30%.

A possible limitation of our study is the limited number of subjects. The minimum required was calculated as a total of 12 patients (with minimally two measurements per patient). This sample size yields 80% power for testing the hypothesis: ICC ≤0.6 against the one-sided alternative (ICC >0.6) at a significance level of 5% when the true ICC is equal to 0.9 (39). Moreover, this sample size of 12 patients resulted in confidence intervals for the limits of agreement ranging to approximately 1.1 times the standard deviation of the difference scores, at either side of the estimated limit of agreement (27). Thus, present study provides a reasonable estimate of the expected RC.

CONCLUSION

In patients with metastatic PC, repeatability of SUV$_{AUC}$ was comparable to that of standard SUV and indicated that $^{18}$F-FCH PET/CT uptake differences of 30% or more are likely to represent treatment effects. Repeatability of MTV and TLCU, respectively, was ~35%. Observed repeatabilities are of the same order of magnitude as those seen for other commonly used radiotracers, such as $^{18}$F-FDG and $^{18}$F-FLT.
ACKNOWLEDGEMENTS

The authors would like to thank their colleagues at the Department of Radiology & Nuclear Medicine involved in tracer production and data acquisition. They are also grateful to the urologists from the VU University Medical Center (André N. Vis, MD, PhD) and Amstelland Hospital Amstelveen (Joop W. Noordzij MD; Sven Nadorp, MD) as well as the VU University Medical Center oncologists (Catharina Menke-van der Houven van Oordt, MD, PhD; Joyce van Dodewaard-de Jong, MD; Jens Voortman, MD, PhD) for their contribution to patient selection and inclusion.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.
REFERENCES


Figure 1. Relative differences between test-retest data and the mean values for the different types of SUVs (max, mean and peak) and their normalizations (SUV, SUV\textsubscript{AUC} and SUV\textsubscript{TBR}). Black dots represent outliers (more than 2 SD). Bonferroni-corrected significance levels $p < 0.0056$ (for significance level of 5%, denoted by *) and $p < 0.0011$ (for significance level of 1%, denoted by **)
Figure 2. Bland-Altman plots for relative differences between test-retest data and mean values for SUVpeak
Figure 3. Bland-Altman plots for the relative test-retest repeatability of SUV\textsubscript{peak} (A) and SUV\textsubscript{mean} (B), as a function of metabolic tumor volume (MTV)
**Figure 4.** Relative differences between test-retest for metabolic tumor volume (MTV) as a function of MTV
Table 1. Patient characteristics

<table>
<thead>
<tr>
<th>Patient</th>
<th>Age (yr)</th>
<th>Weight (kg)</th>
<th>Gleason</th>
<th>PSAtt (ng/ml)</th>
<th>HT</th>
<th>CT</th>
<th>EBRT</th>
<th>RALP</th>
<th>LND</th>
<th>Immun</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>65</td>
<td>78</td>
<td>9</td>
<td>93</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>55</td>
<td>89</td>
<td>8</td>
<td>18</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>74</td>
<td>90</td>
<td>7</td>
<td>43</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>68</td>
<td>79</td>
<td>10</td>
<td>8</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>66</td>
<td>83</td>
<td>8</td>
<td>54</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>60</td>
<td>85</td>
<td>9</td>
<td>137</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>54</td>
<td>98</td>
<td>7</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>72</td>
<td>84</td>
<td>7</td>
<td>54</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>54</td>
<td>85</td>
<td>9</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>54</td>
<td>108</td>
<td>9</td>
<td>226</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>71</td>
<td>81</td>
<td>8</td>
<td>39</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>70</td>
<td>101</td>
<td>8</td>
<td>49</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

PSAtt prostate-specific antigen at the time of performing the PET/CT scans; HT (anti)hormone therapy; CT chemotherapy; EBRT external beam radiotherapy; RALP robot assisted laparoscopic prostatectomy; LND lymph node dissection; Immun immunotherapy.
Table 2. Test-retest differences, intraclass correlation coefficients and repeatability coefficients, for different types of SUVs (max, mean and peak) and their normalizations (SUV, SUV\textsubscript{AUC} and SUV\textsubscript{TBR}).

<table>
<thead>
<tr>
<th></th>
<th>TRT diff (mean±SD)</th>
<th>ICC (95%CI)</th>
<th>RC*</th>
</tr>
</thead>
<tbody>
<tr>
<td>SUVmax</td>
<td>SUV</td>
<td>-0.38±1.45</td>
<td>0.95 (0.91-0.98)</td>
</tr>
<tr>
<td></td>
<td>SUV\textsubscript{AUC}</td>
<td>0±0.88</td>
<td>0.95 (0.92-0.99)</td>
</tr>
<tr>
<td></td>
<td>SUV\textsubscript{TBR}</td>
<td>1.0±3.9</td>
<td>0.89 (0.81-0.97)</td>
</tr>
<tr>
<td>SUVpeak</td>
<td>SUV</td>
<td>-0.26±0.97</td>
<td>0.96 (0.94-0.98)</td>
</tr>
<tr>
<td></td>
<td>SUV\textsubscript{AUC}</td>
<td>0.02±0.70</td>
<td>0.95 (0.93-0.98)</td>
</tr>
<tr>
<td></td>
<td>SUV\textsubscript{TBR}</td>
<td>0.76±3.02</td>
<td>0.89 (0.83-0.95)</td>
</tr>
<tr>
<td>SUVmean</td>
<td>SUV</td>
<td>-0.13±0.76</td>
<td>0.96 (0.94-0.99)</td>
</tr>
<tr>
<td></td>
<td>SUV\textsubscript{AUC}</td>
<td>0.05±0.52</td>
<td>0.95 (0.92-0.98)</td>
</tr>
<tr>
<td></td>
<td>SUV\textsubscript{TBR}</td>
<td>0.68±2.3</td>
<td>0.89 (0.81-0.96)</td>
</tr>
</tbody>
</table>

*TRT diff* test-retest differences; *ICC* intraclass coefficients; *RC* repeatability coefficients; *for relative differences according to Bland (2007) method.*