# Reproducibility and Repeatability of Semiquantitative 18F-Fluorodihydrotestosterone Uptake Metrics in Castration-Resistant Prostate Cancer Metastases: A Prospective Multicenter Study

Hebert Alberto Vargas*[1], Gem M. Kramer*[2], Andrew M. Scott[3,4], Andrew Weickhardt[5], Andreas A. Meier[1], Nicole Parada[6], Bradley J. Beattie[7], John L. Humm[7], Kevin D. Staton[1], Pat B. Zanzonico[1], Serge K. Lyashchenko[1], Jason S. Lewis[1,8], Maqsood Yaqub[2], Ramon E. Sosa[1], Alfons J. van den Eertwegh[9], Ian D. Davis[10], Uwe Ackermann[11], Kunthi Pathmaraj[11], Robert C. Schuit[2], Albert D. Windhorst[2], Sue Chua[12], Wolfgang A. Weber[1,8], Steven M. Larson[1,8], Howard I. Scher[6,13], Adriaan A. Lammertsma[2], Otto S. Hoekstra[2], and Michael J. Morris[6,13]

[1]Department of Radiology, Memorial Sloan Kettering Cancer Center, New York, New York; [2]Department of Radiology and Nuclear Medicine, VU University Medical Centre, Amsterdam, The Netherlands; [3]Department of Molecular Imaging and Therapy, The University of Melbourne, Heidelberg, Victoria, Australia; [4]Department of Medicine, The University of Melbourne, Olivia Newton-John Cancer Research Institute, and La Trobe University, Austin Hospital, Heidelberg, Victoria, Australia; [5]Department of Medical Oncology, Olivia Newton-John Cancer Research Institute, Austin Hospital, Melbourne, Victoria, Australia; [6]Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, New York; [7]Department of Medical Physics, Memorial Sloan Kettering Cancer Center, New York, New York; [8]Department of Radiology, Weill Cornell Medicine, New York, New York; [9]Department of Medical Oncology, VU University Medical Centre, Amsterdam, The Netherlands; [10]Monash University and Eastern Health, Eastern Health Clinical School, Box Hill, Australia; [11]Department of Nuclear Medicine and Centre for PET, Austin Health, Heidelberg, Victoria, Australia; [12]Department of Nuclear Medicine, Royal Marsden NHS Foundation Trust, Sutton, United Kingdom; and [13]Department of Medicine, Weill Cornell Medicine, New York, New York

18F-fluorodihydrotestosterone (18F-FDHT) is a radiolabeled analog of the androgen receptor's primary ligand that is currently being credentialed as a biomarker for prognosis, response, and pharmacodynamic effects of new therapeutics. As part of the biomarker qualification process, we prospectively assessed its reproducibility and repeatability in men with metastatic castration-resistant prostate cancer. **Methods:** We conducted a prospective multiinstitutional study of metastatic castration-resistant prostate cancer patients undergoing 2 (test/retest) 18F-FDHT PET/CT scans on 2 consecutive days. Two independent readers evaluated all examinations and recorded SUVs, androgen receptor–positive tumor volumes, and total lesion uptake for the most avid lesion detected in each of 32 predefined anatomic regions. The relative absolute difference and reproducibility coefficient (RC) of each metric were calculated between the test and retest scans. Linear regression analyses, intraclass correlation coefficients (ICCs), and Bland–Altman plots were used to evaluate repeatability of 18F-FDHT metrics. The coefficient of variation and ICC were used to assess interobserver reproducibility. **Results:** Twenty-seven patients with 140 18F-FDHT–avid regions were included. The best repeatability among 18F-FDHT uptake metrics was found for SUV metrics ($SUV_{max}$, $SUV_{mean}$, and $SUV_{peak}$), with no significant differences in repeatability among them. Correlations between the test and retest scans were strong for all SUV metrics ($R^2 \geq 0.92$; ICC $\geq 0.97$). The RCs of the SUV metrics ranged from 21.3% ($SUV_{peak}$) to 24.6% ($SUV_{max}$).

The test and retest androgen receptor–positive tumor volumes and TLU, respectively, were highly correlated ($R^2$ and ICC $\geq 0.97$), although variability was significantly higher than that for SUV (RCs > 46.4%). The prostate-specific antigen levels, Gleason score, weight, and age did not affect repeatability, nor did total injected activity, uptake measurement time, or differences in uptake time between the 2 scans. Including the most avid lesion per patient, the 5 most avid lesions per patient, only lesions 4.2 mL or more, only lesions with an SUV of 4 g/mL or more, or normalizing of SUV to area under the parent plasma activity concentration–time curve did not significantly affect repeatability. All metrics showed high interobserver reproducibility (ICC > 0.98; coefficient of variation < 0.2%–10.8%). **Conclusion:** Uptake metrics derived from 18F-FDHT PET/CT show high repeatability and interobserver reproducibility.

**Key Words:** prostate cancer; PET; FDHT; reproducibility; repeatability

Prostate cancer is driven by the androgen receptor (AR) signaling axis, including the terminal phase of the disease, metastatic castration-resistant prostate cancer (mCRPC). This AR addiction is the basis of numerous AR-targeted therapies for mCRPC that prolong survival and improve quality of life (*1,2*).

Given the central role the AR axis has in mCRPC and its treatment, there is a pressing need to credential noninvasive biomarkers capable of monitoring the pharmacologic targeting and effect of these drugs. 18F-fluorodihydrotestosterone (18F-FDHT) is a radiolabeled analog of dihydrotestosterone, the

**TABLE 1**
Patient Characteristics

| Characteristic | Center 1 | Center 2 | Center 3 | P |
|---|---|---|---|---|
| No. of patients | 13 | 14 | 5 | |
| Age (y) | 69 (52–88) | 65 (47–75)* | 69 (64–77) | 0.05* |
| Length (cm) | 176 (165–185) | 184 (164–194)* | 172 (164–177) | 0.03* |
| Weight (kg) | 83 (66–122) | 88 (65–125) | 90 (68–106) | 0.88 |
| Gleason score | 8 (7–10) | 8 (5–10) | 7.5 (6–9) | 0.31 |
| PSA (ng/mL) | 4.9 (0.5–1,298)† | 103 (11–1,602) | 107 (15–436) | 0.001† |
| Lesion (n) | | | | 0.06‡ |
|   Bone | 36 | 62 | 21 | |
|   Lymph node | 6 | 13 | 9 | |
|   Soft tissue | 2 | 0 | 0 | |
| Location (n) | | | | 0.99‡ |
|   Skull | 1 | 2 | 0 | |
|   Cervical vertebrae | 2 | 5 | 2 | |
|   Thoracic vertebrae | 4 | 7 | 2 | |
|   Lumbar vertebrae | 5 | 8 | 2 | |
|   Sacral vertebrae | 4 | 9 | 2 | |
|   Pelvis | 6 | 7 | 2 | |
| Ribs/sternum/clavicles | 10 | 17 | 7 | |
|   Extremities | 4 | 7 | 2 | |
|   Pelvic | 2 | 6 | 6 | |
|   Upper abdominal | 1 | 2 | 2 | |
|   Thoracic | 3 | 3 | 0 | |
|   Neck | 2 | 2 | 2 | |
| Scanner type | GE 690 or GE710 | Philips Gemini TF 64 | Philips Ingenuity TF 128 | |
| Uptake time (min) | | | | |
|   Test | 46 (36–53) | 45 (45–47)§ | 60 (42–60) | 0.06 |
|   Retest | 47 (38–59) | 45 (44–48)§ | 60 (57–67)‖ | 0.00‖ |
| Injected activity (MBq) | | | | |
|   Test | 306 (241–348) | 194 (152–216)* | 309 (234–319) | 0.00* |
|   Retest | 323 (298–355) | 193 (186–215)* | 295 (251–333) | 0.00* |
| Residual dose (MBq) | | | | |
|   Test | 16.4 (5.85–30.7) | 36.5 (26.0–62.5)* | 16.1 (14.4–31.8) | 0.00* |
|   Retest | 15.7 (6.68–28.7) | 35.9 (18.4–53.5)* | 20.5 (14.1–24.8) | 0.00* |

*Significant difference between sites (1-way ANOVA).
†Prostate-specific antigen (PSA) levels were significantly lower for Center 1 (Kruskal–Wallis test).
‡$\chi^2$ test.
§Variability was significantly different from other 2 sites (Levene test).
‖Uptake time was significantly longer for Center 3 (Kruskal–Wallis test).
Data are median, with range given in parentheses.

primary ligand of the AR, which offers an innovative way of directly imaging the primary molecular engine of castration-resistant prostate cancer with PET/CT. Preliminary studies using ^{18}F-FDHT PET/CT in patients with castration-resistant prostate cancer have demonstrated safety, feasibility, favorable pharmacokinetic properties, accuracy at identifying tumor localizations, and associations with survival (3–7). Furthermore, ^{18}F-FDHT was instrumental for demonstrating AR targeting in the early-phase clinical trials of enzalutamide and apalutamide, 2 AR-directed therapies that have demonstrated substantial clinical activity in mCRPC (8,9).

This international collaboration was undertaken to assess the repeatability and reproducibility of ^{18}F-FDHT uptake measures, a crucial component of biomarker development (10,11). Repeatability is defined as the measurement precision under a set of repeatability conditions (e.g., repeated scans within 1 subject) and reproducibility as the measurement precision under a set of different conditions in similar subjects (e.g., different locations, operators, readers) (12,13).

## TABLE 2
### Descriptive Statistics of Several Uptake Measures

| Metric | Overall Test | Overall Retest | Center 1 Test | Center 1 Retest | Center 2 Test | Center 2 Retest | Center 3 Test | Center 3 Retest | P (between centers) Test | P (between centers) Retest |
|---|---|---|---|---|---|---|---|---|---|---|
| SUV$_{max}$ | 7.46 ± 3.37 | 7.70 ± 3.78 | 6.88 ± 3.30 | 7.18 ± 3.63 | 8.01 ± 3.82 | 8.27 ± 4.35 | 6.77 ± 1.73 | 6.90 ± 1.80 | 0.35 | 0.43 |
| SUV$_{peak}$ | 6.53 ± 2.88 | 6.80 ± 3.22 | 6.43 ± 3.01 | 6.78 ± 3.25 | 6.77 ± 3.10 | 7.00 ± 3.53 | 5.66 ± 0.99 | 5.97 ± 1.19 | 0. 88 | 0.90 |
| SUV$_{mean}$ | 5.24 ± 2.28 | 5.41 ± 2.55 | 4.92 ± 2.24 | 5.20 ± 2.52 | 5.57 ± 2.61 | 5.75 ± 2.93 | 4.77 ± 1.13 | 4.83 ± 1.14 | 0.57 | 0.70 |
| TLU | 47.1 ± 104.7 | 46.4 ± 95.5 | 30.8 ± 35.92 | 29.8 ± 33.17 | 62.9 ± 138.0 | 60.8 ± 124.3 | 26.7 ± 32.6 | 29.8 ± 42.1 | 0.001* | 0.001* |
| ARTV | 8.78 ± 15.87 | 8.39 ± 13.81 | 5.74 ± 5.73* | 5.36 ± 5.13* | 11.6 ± 20.65 | 10.8 ± 17.65 | 5.36 ± 6.14 | 5.81 ± 7.38 | 0.003* | 0.003* |

*Volumetric measures were significantly larger in Center 2 (Kruskal–Wallis test).

The aim of this study was to prospectively assess repeatability and reproducibility of whole-body $^{18}$F-FDHT uptake metrics of mCRPC metastases.

## MATERIALS AND METHODS

Patients were recruited prospectively from 3 tertiary academic centers: Memorial Sloan Kettering Cancer Center (United States), VU University Medical Center (The Netherlands), and Austin Health (Australia). Each site opened its own study and managed the regulatory requirements specific to each institution and country. The trials, by prospective intent, were to collect and combine data under a predefined statistical plan. The lead site (Memorial Sloan Kettering) holds a U.S. Food and Drug Administration Investigational New Drug application for $^{18}$F-FDHT (#66115) and provided letters of cross-reference to facilitate submission for regulatory approval for the other sites. The institutional review boards of each center approved the study, and all patients provided written informed consent before inclusion. The clinicaltrials.gov identifier is NCT00588185 (this number applies only to Memorial Sloan Kettering, the only U.S.-based site).

### Patient Eligibility and Study Design

Eligibility criteria included pathologically proven mCRPC, castrate serum testosterone (≤50 ng/dL), 4 wk or more since patients' last anticancer pharmacologic therapy, and progressive disease based on a rise in prostate-specific antigen or on RECIST 1.1 imaging evidence of progressive disease or 2 or more new metastatic lesions on bone scan not attributable to the flair phenomenon.

Patients without surgical or medical castration remained on androgen depletion therapy with gonadotropin-releasing hormone analogs/inhibitors. Patients on enzalutamide or other antiandrogens within 4 wk were excluded, as this therapy directly competes with $^{18}$F-FDHT uptake. The design included means to evaluate the effect of time between the test and retest $^{18}$F-FDHT injections on the uptake measurements. Up to 3 cohorts were planned for test–retest scans (cohort 1: days 1 and 2; cohort 2: days 1 and 8; and cohort 3: days 1 and 22). Initially, patients would be studied in cohort 1. If unstable test–retest $^{18}$F-FDHT uptake (defined as a relative difference > 0.15) was present in 5 or more patients at any time, the study would proceed to the subsequent cohort. However, as a relative difference greater than 0.15 was not observed in 5 or more patients in cohort 1, there was no indication to proceed to subsequent cohorts, and all patients underwent $^{18}$F-FDHT PET/CT scans on 2 consecutive days.

### Image Acquisition

Images were acquired using a GE690 or GE710 (GE Healthcare) or Gemini TF64 or Ingenuity TF128 (both from Philips) PET/CT scanner. For each scan, a low-dose CT scan (120–140 kV, 80 mA) was obtained, followed by a dynamic 30-min PET scan over the thorax after intravenous $^{18}$F-FDHT administration. All scans were

## TABLE 3
### Mean Relative Differences and RCs on Lesion Level for Several Uptake Metrics

| Normalization factor | Quantitative tracer uptake measures | Overall Mean difference (%) | Overall RC (%) | Center 1 Mean difference (%) | Center 1 RC (%) | Center 2 Mean difference (%) | Center 2 RC (%) | Center 3 Mean difference (%) | Center 3 RC (%) |
|---|---|---|---|---|---|---|---|---|---|
| Body weight | SUV$_{max}$ | 2.5 | 24.6 | 3.8 | 23.3 | 2.3 | 27.3 | 1.8 | 18.9 |
| | SUV$_{peak}$ | 3.3 | 21.3 | 4.9 | 21.0 | 2.2 | 23.2 | 4.9 | 10.9 |
| | SUV$_{mean}$ | 2.8 | 24.2 | 4.9 | 24.2 | 2.5 | 26.7 | 1.3 | 16.4 |
| | TLU | 2.4 | 46.4 | 0.8 | 56.0 | 1.7 | 40.5 | 6.0 | 49.2 |
| | ARTV | −0.3 | 53.7 | −4.1 | 56.1 | −0.7 | 50.7 | 4.7 | 58.5 |
| Area under the parent plasma activity concentration curve | SUV$_{max/AUCpp}$ | 2.7 | 34.8 | 17.0 | 31.5 | 4.9 | 28.6 | −10.5 | 42.8 |
| | SUV$_{peak/AUCpp}$ | 3.9 | 24.3 | 18.7 | 37.4 | 3.1 | 21.8 | 0.7 | 24.8 |
| | SUV$_{mean/AUCpp}$ | 2.7 | 36.0 | 15.5 | 33.6 | 5.1 | 30.2 | −11.5 | 42.5 |
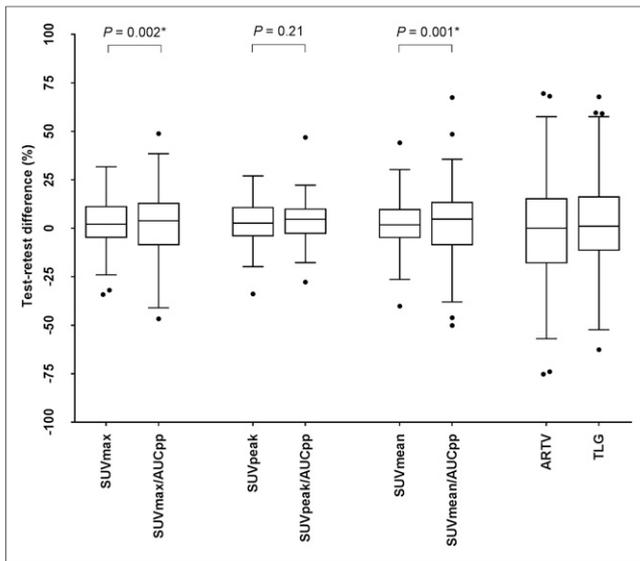
**FIGURE 1.** Box plots of percentage differences on lesion level between test and retest scans for several quantitative uptake values. Effect of normalizing to AUCpp is shown.

corrected for decay, scatter, random coincidences, and photon attenuation. During the dynamic scans, 3 intravenous samples were drawn at 5, 10, and 30 min after injection. Whole-blood activity concentration, plasma activity concentration, and parent and metabolite fractions (by high-pressure liquid chromatography) of $^{18}$F-FDHT were measured. A whole-body PET/CT (mid thigh to mid skull) followed, starting approximately 45 min after injection. A whole-body low-dose CT scan (120–140 kV, 80 mA) was acquired with a section thickness and reconstruction interval of 5 mm and pitch of 0.75–1.5. No oral or intravenous contrast material was administered.

### Data Management and Analysis

The Clinical Trials Network from the Society of Nuclear Medicine and Molecular Imaging provided both centralized data management and access to Imagys®, a web-based Imaging Clinical Trial management system by Keosys, for secure uploading, storage, downloading, and analysis of images.

All images were evaluated independently by a dually trained radiologist/nuclear medicine physician and a nuclear medicine resident (8 and 3 y experience in PET/CT, respectively). Lesions were considered suggestive of metastases when uptake was visually higher than blood-pool activity measured in the thoracic aorta or background tissue specific to the site of the lesion and separate from known physiologic uptake (blood pool, biliary, urinary, and gastrointestinal

tracts). Lesion type (bone, nodal, or other soft tissue) and anatomic site (grouped into 11 regions for bone, 11 regions for nodes, and 10 regions for other soft tissue) were recorded (Supplemental Fig. 1; supplemental materials are available at http://jnm.snmjournals.org). The most visually prominent $^{18}$F-FDHT–avid lesion in each predefined anatomic region was delineated and a volume of interest generated semiautomatically using a 50% isocontour of $SUV_{max}$ corrected for local background. The following $^{18}$F-FDHT uptake metrics were recorded: $SUV_{max}$, $SUV_{peak}$ (1.2 cm$^3$ spheric region positioned within the lesion to maximize its mean value), and $SUV_{mean}$ (all voxels within the lesion) corrected for body weight. Additionally, these metrics were normalized to the area under the parent plasma time–activity concentration curve (AUC) at 30 min ($SUV_{AUCpp}$) (*14*). Androgen receptor–positive tumor volume ([ARTV] derived using a 50% threshold of $SUV_{max}$ corrected for local background) and total lesion uptake ([TLU] defined as $SUV_{mean} \times$ ARTV) of $^{18}$F-FDHT were calculated.

### Statistical Analysis

Repeatability and interobserver reproducibility were determined by calculating the relative absolute difference in $^{18}$F-FDHT uptake metrics between the test and retest scans, and between the values of the uptake metrics measured by the 2 readers. The relative absolute difference was computed as:

$$\%\text{Difference} = \frac{\text{Uptake metric day 2} - \text{uptake metric day 1}}{(\text{Uptake metric day 1} + \text{uptake metric day 2})/2} \times 100$$

If no lesion was identified in a patient, the absolute change was set to zero but was not considered when calculating quantitative repeatability coefficients (RCs). The RC was calculated as 1.96*SD of the relative absolute differences per lesion and per patient for all uptake metrics. Normality was evaluated visually using a quantile-quantile plot and histogram analyses. Significance of differences in uptake metrics between the 2 scans and between the 2 readers was assessed using a paired *t* test. To assess differences in RCs, a Levene test was performed; differences were deemed significant if the *P* value was less than 0.05. Linear regression analyses, intraclass correlation coefficients (ICCs), and Bland–Altman plots were used to evaluate repeatability. Additionally, the coefficient of variation (COV) and ICC were used to investigate interobserver reproducibility.

A Levene test was performed to assess the effect of various lesion selection strategies on repeatability and reproducibility: lesions of 4.2 mL or more (diameter ≥ 2 cm), SUV of 4.0 g/mL or more, and up to the 5 most radiotracer-avid lesions, as suggested by the PERCIST guidelines (*15*). In addition, the uptake values of these 5 individual target lesions were averaged per patient to obtain mean uptake values. A post hoc linear regression analysis was performed to evaluate the influence of prostate-specific antigen levels, Gleason score, weight, and differences in total injected activity and uptake time between both scans on a per-patient basis. On the basis of previous reports on repeatability of $^{18}$F-FDG uptake in malignant tumors, 30% or less variability between the test and retest was considered acceptable (*15,16*). All statistical analyses were performed using SPSS 22.0 (SPSS).

Additional details on study design, image acquisition and processing, radio–high-performance liquid chromatography, and analysis of $^{18}$F-FDHT metabolism are available on request.
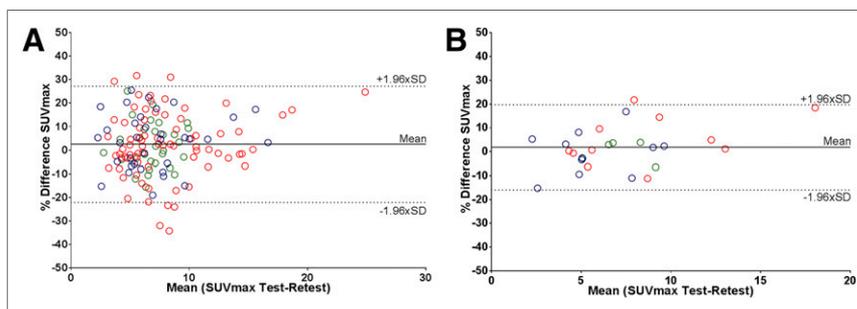


**FIGURE 2.** Bland–Altman plots showing repeatability $SUV_{max}$ on lesion (A) and patient (B) level. Blue = center 1; red = center 2; green = center 3.
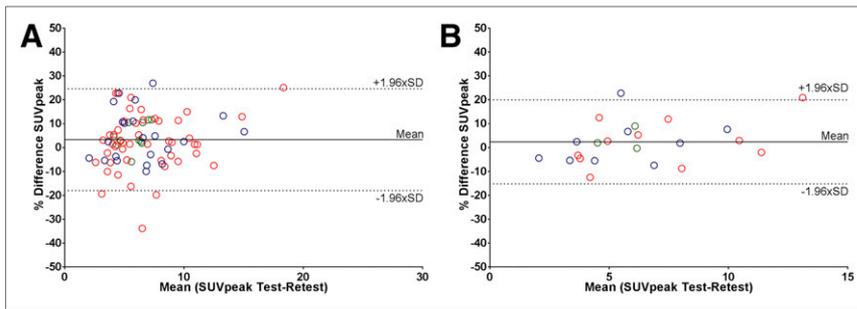
**FIGURE 3.** Bland–Altman plots showing repeatability of $SUV_{peak}$ on lesion (A) and patient (B) level. Blue = center 1; red = center 2; green = center 3.

## RESULTS

Thirty-two patients were included. The minimum number of paired evaluations per patient (i.e., per the anatomic regions described in the "Materials and Methods" section) was 1; the maximum was 12. Five patients were excluded from the RC calculations, because no lesions were detected on PET. Overall, 27 patients with a total of 140 $^{18}$F-FDHT–avid lesions were evaluated. No significant differences in patient characteristics were observed between the test and retest scans. The total injected activities at center 2 were significantly lower than those of centers 1 and 3; however, no systematic differences were found in the SUVs from centers 1 and 3 (Tables 1 and 2).

### Repeatability

The best repeatability of $^{18}$F-FDHT PET/CT uptake metrics was found for SUV, where the predefined threshold of variability of 30% or less was met (Table 3; Fig. 1). No significant differences in variability were found between $SUV_{max}$, $SUV_{mean}$, and $SUV_{peak}$, and correlations between the test and retest scans were strong ($R^2 \geq 0.92$; ICC $\geq 0.97$). Bland–Altman graphs did not show skewness of the data (Figs. 2 and 3). The RCs of the overall SUV metrics ranged from 21.3% ($SUV_{peak}$) to 24.6% ($SUV_{max}$). Significantly smaller RCs were found between $SUV_{mean}$ and $SUV_{peak}$ at center 3 and those of centers 1 and 2 ($P = 0.03–0.04$). Only for $SUV_{max}$, the variability was significantly less in soft tissue versus bone lesions (RCs 18.2% vs. 26.1%; $P = 0.04$). Repeatability of the uptake metrics showed a trend toward dependency on lesion size, but not on absolute SUVs (Fig. 4).

Test and retest TLU and ARTV values also showed good correlation ($R^2$ and ICC $\geq 0.97$), although variability was significantly larger than for SUV, and the predefined variability threshold of 30% or less was not met (RCs > 46.4%) (Fig. 5). Mean TLU was significantly larger in patients from center 2, yet variability was only significantly lower than that of center 1 (40.5 vs. 56.0%; $P = 0.02$). Even when evaluated on a per-region basis, RCs remained significantly higher compared with those from the SUV metrics and were not influenced by lesion type.

Assessing variability of the $^{18}$F-FDHT uptake metrics on a per-patient basis improved repeatability of all uptake metrics (Table 3; Fig. 6). RCs of SUV decreased 6% on average, which was significant for $SUV_{max}$ and $SUV_{mean}$. The improvement of volumetric measures was larger, with changes in RCs of TLU and ARTV being 12.7 and 23.1%, respectively. This was mainly caused by a large decrease in variability of ARTV of centers 2 and 3 after averaging the data. Prostate-specific antigen level, Gleason score, weight, and age did not affect repeatability, nor did differences in total

injected activity or uptake time after injection between both scans ($R^2$: < 0.08) (Fig. 7).

### Normalization to Parent Plasma Input Curve

Adequate blood samples were available from 21 of the 27 patients with a total of 103 lesions. Normalizing SUV to AUC significantly decreased the overall repeatability on both lesion and patient bases for centers 1 and 3 (Tables 3 and 4). This was mainly due to large differences (>50%) in whole-blood activity concentrations between samples in the test and retest samples from 2 patients. When these outliers were removed, the repeatability for centers 1 and 3 improved and only a slight change in RCs on an overall lesional basis was observed after normalization ($SUV_{max}$: 29.9%; $SUV_{mean}$: 30.3%; $SUV_{peak}$: 21.6%). This was also seen for RCs on a per-patient level ($SUV_{max}$: 25.6%; $SUV_{mean}$: 23.8%; and $SUV_{peak}$: 16.3%).

### Lesion Selection

Inclusion of up to the 5 most avid lesions per patient did not significantly affect repeatability for any of the uptake metrics. If these lesions were assessed on a per-patient basis, RCs were similar to those before lesion selection. Likewise, only including the single most avid lesion, lesions of 4.2 mL or more, or lesions with an SUV of 4 g/mL or more did not significantly affect repeatability. Decrease in RCs ranged from 0% to 6.5% for all uptake metrics.

### Reproducibility

Reproducibility between readers was excellent for $SUV_{max}$ and $SUV_{peak}$, with discrepancies in measurements between the readers found in only 2 of 300 measurements and 12 of 140 lesions for $SUV_{max}$ and $SUV_{peak}$, respectively. Lesions showing discrepancies were close to regions of high physiologic uptake (e.g., liver, urinary tract, or vascular structures) or showed diffuse uptake (e.g., diffuse disease in the pelvis). Both metrics showed high reproducibility (ICC: 1.00) and a low COV ($\leq$0.20%).

The remaining semiquantitative uptake measures were more dependent on volume-of-interest definition. The correlation between
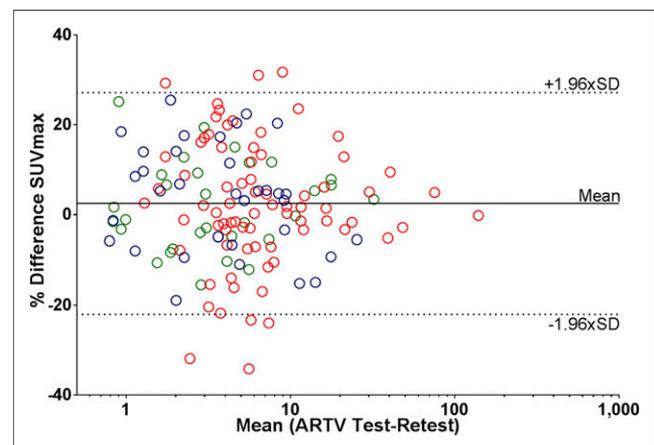


**FIGURE 4.** Bland–Altman plots showing influence of ARTV on repeatability of $SUV_{max}$ on lesion level. Log scale is used on $x$-axis. Blue = center 1; red = center 2; green = center 3.
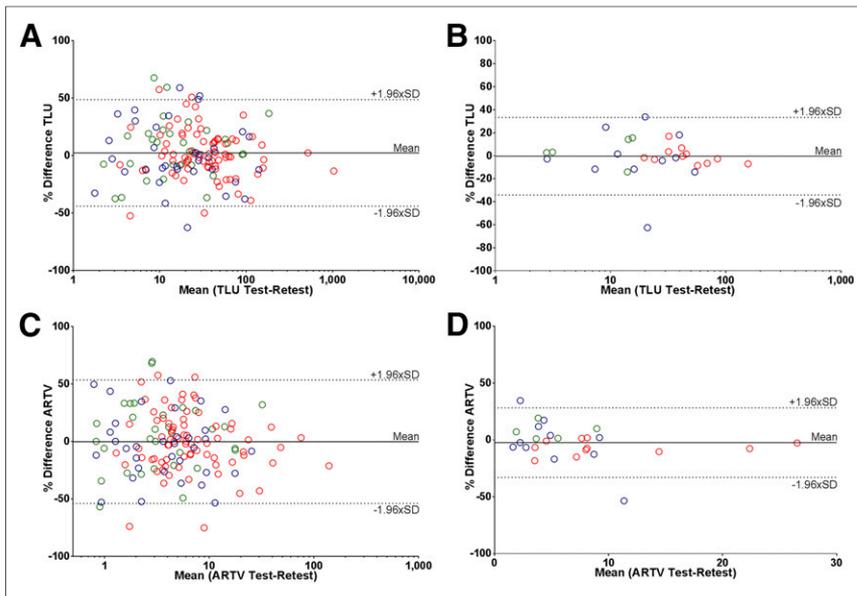
**FIGURE 5.** Bland–Altman plots showing repeatability of TLU (A and B) and ARTV (C and D) on lesion (A and C) and patient (B and D) level. For TLU and lesion-level ARTV: log scale is used on *x*-axis. Blue = center 1; red = center 2; green = center 3.

both readers for $SUV_{mean}$ was still excellent (ICC: 0.99), but the variation was significantly higher (COV: 2.3%). TLU and ARTV were less reproducible than all SUV metrics (COV: 10.8% and 10.4%, respectively), yet the ICCs remained above 0.98.

## DISCUSSION

This multicenter prospective study assessed repeatability and reproducibility of $^{18}$F-FDHT, both of which are key components of the tracer's analytic validation as a clinical biomarker. Repeatability of SUV metrics was superior to that of volumetric metrics,
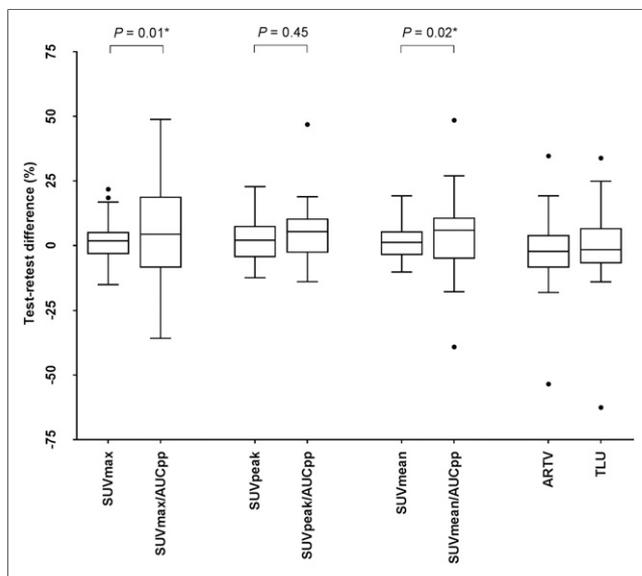
with repeatability coefficients ranging between 16.4% and 17.8% on a patient basis and 21.3%–24.6% on a region basis. As a necessary step in biomarker development, this study demonstrated the feasibility of $^{18}$F-FDHT PET/CT imaging in a multiinstitutional setting and satisfied the requirement to evaluate the biomarker's test–retest repeatability (*17*). In the current era of AR-directed mCRPC drug development, such biomarkers can serve as a pharmacodynamic, a prognostic, and a response indicator (*6–9*).

Most studies on test–retest repeatability in PET/CT have evaluated $^{18}$F-FDG uptake. The PERCIST guidelines recommend a more than 30% change in SUV to define a meaningful change in clinical status for both disease response and progression (*15*). Weber et al. evaluated $^{18}$F-FDG PET/CT imaging in 74 patients with non–small cell lung cancer in a multiinstitutional (*n* = 9) clinical trial and reported thresholds of 28%/32% decrease and 39%/47% increase in $SUV_{max}$ and $SUV_{peak}$, respectively, to be most indicative of actual therapeutic effects (*16*). However, multiple technical and logistic factors can affect these measurements, including differences in volume of interest, delineation, magnitude of uptake metrics, and uptake time after intravenous injection, as well as difficulties related to adherence to protocol design in a multiinstitutional setting (*18,19*). Similar studies in patients with prostate cancer have been conducted with other radiotracers. Variation coefficients of 14% and 7% were



**FIGURE 6.** Box plots of percentage differences on patient level between test and retest scans for several quantitative uptake values. Effect of normalizing to AUC is shown.
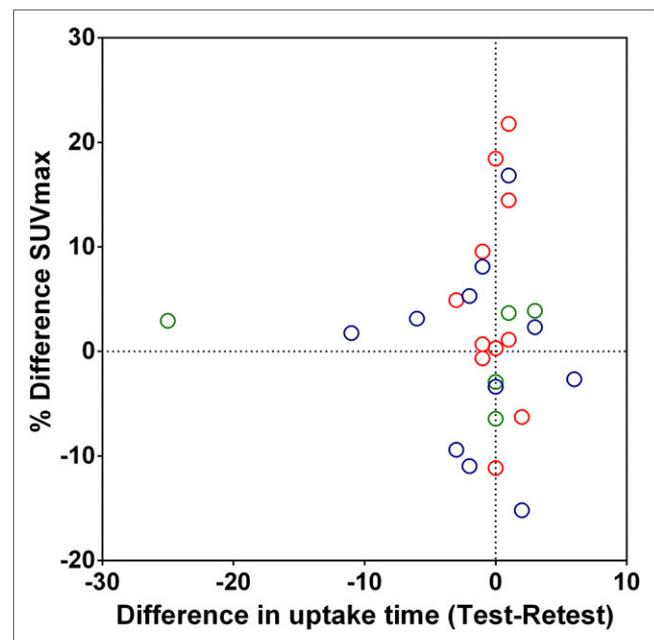


**FIGURE 7.** Scatterplot showing effect of differences in uptake time (mins) between test and retest scans on differences in uptake at patient level. Similar patterns are seen for other quantitative uptake metrics. Blue = center 1; red = center 2; green = center 3.

## TABLE 4
### Mean Relative Differences and RCs on Patient Level for Several Uptake Metrics

| Normalization factor | Quantitative tracer uptake measures | Overall | | Center 1 | | Center 2 | | Center 3 | |
|---|---|---|---|---|---|---|---|---|---|
| | | Mean difference (%) | RC (%) | Mean difference (%) | RC (%) | Mean difference (%) | RC (%) | Mean difference (%) | RC (%) |
| Body weight | $SUV_{max}$ | 1.9 | 17.8 | −0.4 | 18.1 | 4.8 | 20.1 | −2.4 | 14.8 |
| | $SUV_{peak}$ | 2.4 | 17.6 | 2.1 | 18.7 | 2.3 | 19.5 | 3.6 | 9.5 |
| | $SUV_{mean}$ | 2.1 | 16.4 | 0.7 | 17.7 | 4.7 | 17.5 | −0.3 | 8.6 |
| | TLU | −0.4 | 33.7 | −2.8 | 49.7 | −0.1 | 14.3 | 4.3 | 23.3 |
| | ARTV | −2.3 | 30.6 | −2.5 | 43.8 | −6.7 | 12.5 | 7.8 | 14.5 |
| Area under the parent plasma activity concentration curve | $SUV_{max/AUCpp}$ | 3.7 | 37.3 | 18.8 | 59.4 | 6.3 | 19.6 | −14.9 | 38.5 |
| | $SUV_{peak/AUCpp}$ | 6.2 | 26.5 | 21.1 | 43.9 | 3.8 | 14.9 | −2.8 | 31.2 |
| | $SUV_{mean/AUCpp}$ | 4.0 | 36.9 | 22.2 | 44.6 | 6.2 | 21.2 | −15.9 | 39.3 |

reported on ^{18}F-NaF PET/CT in patients with mCRPC for $SUV_{max}$ and $SUV_{mean}$, respectively (20). In a study using ^{18}F-fluoromethylcholine in patients with mCRPC, repeatability coefficients ranging between 22% and 26% were reported for different SUV metrics (21). Additionally, this study also reported that RCs of metabolically active tumor volume and TLU were significantly larger than those for SUV (36% and 33%, respectively). Other studies also using $SUV_{max}$-based thresholds showed similar results (22,23), yet a significant decrease in repeatability was seen when only lesions of 4.2 mL or less were included in the analysis. Studies have also shown decreased variability when evaluating repeatability on a per-patient (as opposed to a per-lesion) basis (14,21).

### Normalization to Parent Plasma Input Curve

Two studies have shown a correlation ($R^2$: 0.6–0.7) between nonlinear regression analysis of dynamic ^{18}F-FDHT data and SUV (5,14). Additionally, preliminary results showed a near-perfect correlation when the SUV was normalized to the AUC ($R^2$: 0.99). A potential advantage of normalization to the parent plasma input curves is that the uptake metrics are corrected for any treatment-induced or other changes in the radiotracer's metabolism, albeit at the expense of an additional dynamic PET scan, venous blood samples, and metabolite analysis. Moreover, including an additional variable into uptake metric calculations can increase uncertainty (14,21), although in the present study, $SUV_{AUCpp}$ did not significantly affect overall variability of any of the SUV metrics on a lesion level. One outlier was seen with unexplained large differences in whole-blood activity concentrations between test and retest scans, which could not be accounted for by sample measurement errors, suggesting the need for caution in the case of response assessment.

Our study had limitations. To overcome possible confounders in our study, all lesions were delineated by 2 independent readers. For $SUV_{max}$ and $SUV_{peak}$, reproducibility was nearly perfect, and differences in $SUV_{mean}$ between readers were small. Moreover, differences in uptake time between the test and retest scans did not affect repeatability, suggesting that the influence of this factor was minor. However, repeatability data from 2 readers are insufficient to make strong statements about agreement across a larger pool of readers and will require validation. Patients with castration-resistant prostate cancer often present with numerous metastatic lesions and, ideally, each lesion should be delineated and assessed. However, this is impractical in routine clinical scenarios and therefore we predefined anatomic regions. Yet, this still resulted in 10 or more evaluable regions in 20% of the patients. Several other (simpler) lesion selection criteria were also investigated; those regions did not result in a change in variability.

### CONCLUSION

Metrics derived from ^{18}F-FDHT PET/CT show high repeatability and interobserver reproducibility. Among ^{18}F-FDHT uptake metrics, SUV had the best repeatability, and although ARTV and TLU showed good correlation, variability was higher.

### DISCLOSURE

### REFERENCES

1. Huggins C, Hodges CV. Studies on prostatic cancer: I—the effect of castration, of estrogen and of androgen injection on serum phosphatases in metastatic carcinoma of the prostate. 1941. J Urol. 2002;168:9–12.
2. Graham L, Schweizer MT. Targeting persistent androgen receptor signaling in castration-resistant prostate cancer. Med Oncol. 2016;33:44.
3. Larson SM, Morris M, Gunther I, et al. Tumor localization of 16beta-^{18}F-fluoro-5alpha-dihydrotestosterone versus ^{18}F-FDG in patients with progressive, metastatic prostate cancer. J Nucl Med. 2004;45:366–373.
4. Zanzonico PB, Finn R, Pentlow KS, et al. PET-based radiation dosimetry in man of ^{18}F-fluorodihydrotestosterone, a new radiotracer for imaging prostate cancer. J Nucl Med. 2004;45:1966–1971.
5. Beattie BJ, Smith-Jones PM, Jhanwar YS, et al. Pharmacokinetic assessment of the uptake of 16beta-^{18}F-fluoro-5alpha-dihydrotestosterone (FDHT) in prostate tumors as measured by PET. J Nucl Med. 2010;51:183–192.
6. Vargas HA, Wassberg C, Fox JJ, et al. Bone metastases in castration-resistant prostate cancer: associations between morphologic CT patterns, glycolytic activity, and androgen receptor expression on PET and overall survival. Radiology. 2014;271:220–229.
7. Fox JJ, Gavane SC, Blanc-Autran E, et al. Positron emission tomography/computed tomography-based assessments of androgen receptor expression and glycolytic activity as a prognostic biomarker for metastatic castration-resistant prostate cancer. JAMA Oncol. 2018;4:217–224.

8. Rathkopf DE, Morris MJ, Fox JJ, et al. Phase I study of ARN-509, a novel antiandrogen, in the treatment of castration-resistant prostate cancer. *J Clin Oncol.* 2013;31:3525–3530.

9. Scher HI, Beer TM, Higano CS, et al. Antitumour activity of MDV3100 in castration-resistant prostate cancer: a phase 1-2 study. *Lancet.* 2010;375:1437–1446.

10. Amur SG, Sanyal S, Chakravarty AG, et al. Building a roadmap to biomarker qualification: challenges and opportunities. *Biomark Med.* 2015;9:1095–1105.

11. Woodcock J, Buckman S, Goodsaid F, Walton MK, Zineh I. Qualifying biomarkers for use in drug development: a US Food and Drug Administration overview. *Expert Opin Med Diagn.* 2011;5:369–374.

12. Raunig DL, McShane LM, Pennello G, et al. Quantitative imaging biomarkers: a review of statistical methods for technical performance assessment. *Stat Methods Med Res.* 2015;24:27–67.

13. Kessler LG, Barnhart HX, Buckler AJ, et al. The emerging science of quantitative imaging biomarkers terminology and definitions for scientific studies and regulatory submissions. *Stat Methods Med Res.* 2015;24:9–26.

14. Kramer G, Yaqub M, Schuit R, et al. Assessment of simplified methods for quantification of [$^{18}$F]FDHT uptake in patients with metastasized castrate resistant prostate cancer [abstract]. *J Nucl Med.* 2016;57(suppl 2):464.

15. O JH. Lodge MA, Wahl RL. Practical PERCIST: a simplified guide to PET response criteria in solid tumors 1.0. *Radiology.* 2016;280:576–584.

16. Weber WA, Gatsonis CA, Mozley PD, et al. Repeatability of $^{18}$F-FDG PET/CT in advanced non-small cell lung cancer: prospective assessment in 2 multicenter trials. *J Nucl Med.* 2015;56:1137–1143.

17. O'Connor JP, Aboagye EO, Adams JE, et al. Imaging biomarker roadmap for cancer studies. *Nat Rev Clin Oncol.* 2017;14:169–186.

18. Lodge MA. Repeatability of SUV in oncologic $^{18}$F-FDG PET. *J Nucl Med.* 2017;58:523–532.

19. Kaneta T, Sun N, Ogawa M, et al. Variation and repeatability of measured standardized uptake values depending on actual values: a phantom study. *Am J Nucl Med Mol Imaging.* 2017;7:204–211.

20. Lin C, Bradshaw TJ, Perk TG, et al. Repeatability of quantitative $^{18}$F-NaF PET: a multicenter study. *J Nucl Med.* 2016;57:1872–1879.

21. Oprea-Lager DE, Kramer G, van de Ven PM, et al. Repeatability of quantitative $^{18}$F-fluoromethylcholine PET/CT studies in prostate cancer. *J Nucl Med.* 2016;57:721–727.

22. Hatt M, Cheze-Le Rest C, Aboagye EO, et al. Reproducibility of $^{18}$F-FDG and 3′-deoxy-3′-$^{18}$F-fluorothymidine PET tumor volume measurements. *J Nucl Med.* 2010;51:1368–1376.

23. Frings V, de Langen AJ, Smit EF, et al. Repeatability of metabolically active volume measurements with $^{18}$F-FDG and $^{18}$F-FLT PET in non-small cell lung cancer. *J Nucl Med.* 2010;51:1870–1877.