

Repeatability of quantitative FDG-PET/CT and contrast enhanced CT in recurrent ovarian carcinoma: test retest measurements for tumor FDG uptake, diameter and volume

Running title: Repeatability of FDG-PET/CT in recurrent ovarian cancer. (tracking no. CCR-13-2634)

Andrea G. Rockall¹, Norbert Avril², Raymond Lam³, Robert Iannone³, P. David Mozley⁴, Christine Parkinson^{5,10}, Donald Bergstrom⁶, Evis Sala⁷, Shah-Jalal Sarker⁸, Iain A. McNeish^{8,9}, James D. Brenton^{5,10,11, 12}

1. Department of Radiology, Barts Health NHS Trust, London, UK
2. Department of Nuclear Medicine, Barts Cancer Institute, London, UK
3. Merck and Co, Whitehouse Station, NJ, USA
4. Merck Imaging, West Point, PA, USA
5. Department of Oncology, Cambridge University Hospitals Foundation Trust, Cambridge, UK
6. Sanofi, Bridgewater, NJ, USA
7. Department of Radiology, Cambridge University Hospitals Foundation Trust, Cambridge, UK
8. Centre for Experimental Cancer Medicine, Barts Cancer Institute, Queen Mary University London, UK
9. Department of Medical Oncology, St Bartholomew's Hospital/Barts Health NHS Trust, London, UK
10. Cancer Research UK Cambridge Institute, University of Cambridge, Li Ka Shing Centre, Cambridge CB2 0RE
11. NIHR Cambridge Biomedical Research Centre, Cambridge CB2 2QQ, UK
12. Cambridge Experimental Cancer Medicine Centre, Cambridge CB2 0RE, UK

KEY WORDS:

Test-retest, FDG-PET/CT, tumor volume, recurrent ovarian cancer

Financial support:

This study was funded by Merck and Co.

Corresponding author: Andrea G. Rockall, MRCP, FRCR, Division of Cancer and Surgery, Imperial College London, First Floor, ICTEM Building, Hammersmith Campus, Du Cane Road, London, W12 ONN, UK. Phone: +44 (0) 207 594 2793, Fax: +44 (0) 208 383 5830 Email: a.rockall@imperial.ac.uk

Conflict of interest: AR: honoraria for educational lectures from Novartis and Guerbet; several authors (RL, RI, PDM) work or worked for Merck and Co., DB works for Sanofi.

Abstract word count: 234

Manuscript word count: 3 992

Tables: 2

Figures: 4

STATEMENT OF TRANSLATIONAL RELEVANCE:

The detection of changes in tumor glucose metabolism, tumor diameter or tumor volume within a few weeks of commencing treatment has the potential to inform stratification of patient management. For patients with recurrent ovarian cancer, there is an urgent need to identify more effective therapies, and an imaging tool that can robustly identify early response would be of value both in the clinical setting and as a biomarker for drug development. Validation of imaging biomarkers is critical for effective and reliable use in clinical trials. Test/re-test data for measurements of FDG uptake, tumor diameter and tumor volume are essential in order to determine repeatability coefficients, thereby allowing the confident use of these techniques. However, these data have not previously been established for ovarian cancer. This study establishes robust repeatability coefficients for FDG measurements, enabling evidence-based use of PET/CT in stratification of patients into those with a metabolic or volumetric response to treatment.

ABSTRACT

Purpose:

Repeatability of baseline FDG-PET/CT measurements has not been tested in ovarian cancer. This dual-center, prospective study assessed variation in tumor FDG uptake, tumor diameter (TD) and tumor volume (TV) from sequential FDG-PET/CT and contrast-enhanced CT (CECT) in patients with recurrent platinum-sensitive ovarian cancer.

Methods:

Patients underwent two pre-treatment baseline FDG-PET/CT (n=21) and CECT (n=20) at 2 clinical sites with different PET/CT instruments. Patients were included if they had at least one target lesion (TL) in the abdomen with an SUV maximum (SUVmax) of ≥ 2.5 and a long axis diameter of ≥ 15 mm. Two independent reading methods were used to evaluate repeatability of TD and SUV uptake: on site and at an imaging clinical research organization (CRO). TV reads were only performed by CRO. In each reading set, TLs were independently measured on sequential imaging.

Results:

Median time between FDG-PET/CT was 2 days (range 1-7). For site reads, concordance correlation coefficient (CCC) for SUVmean, SUVmax and TD were 0.95, 0.94 and 0.99 respectively. Repeatability coefficients were 16.3%, 17.3% and 8.8% for SUVmean, SUVmax and TD respectively. Similar results were observed for CRO reads. TV CCC was 0.99 with a repeatability coefficient of 28.1%.

Conclusions:

There was excellent test/retest repeatability for FDG-PET/CT quantitative measurements across two sites and two independent reading methods. Cut-off values for determining change in SUVmean, SUVmax and TV establish limits to determine metabolic and/or volumetric response to treatment in platinum-sensitive relapsed ovarian cancer.

BACKGROUND

Epithelial ovarian cancer (EOC) remains the most lethal gynecological malignancy and overall survival has not changed significantly over the past 15 years (1). Most patients present with advanced stage disease and the primary treatment modality is surgical cytoreduction followed by platinum/taxane-based chemotherapy. Relapsed disease is classified by its likely response to further treatment with platinum-based chemotherapy, being either platinum-sensitive or platinum resistant (2). However, the majority of patients eventually develop progressive platinum-resistance. There is a clear unmet clinical need to identify new treatments for women with ovarian cancer.

Evaluation of the effectiveness of new drug treatments in EOC depends upon assessment of response and progression-free survival. Objective response measurement using RECIST 1.1 criteria (uni-dimensional tumor diameter) is highly validated across many cancer types, as well as having high utility in clinical practice (3;4). However, it is used inconsistently by regulatory authorities for the purposes of drug registration (5). There are inherent limitations using uni-dimensional measurements. Notably, the time taken for tumor shrinkage of 30% is typically in excess of 9 weeks and inevitably delays treatment decisions; particularly in ovarian cancer the shape of the mass may alter during treatment so that the long axis does not reflect change in volume and residual soft tissue along peritoneal or serosal surfaces or in complex masses may be difficult to measure and quantify or may not represent active disease (6-8). Despite these recognized limitations, contrast enhanced CT (CECT) is currently the standard of care technique for monitoring response to treatment in ovarian cancer, together with the serum cancer antigen 125 (CA 125) level. Conversely, individual patients with EOC would benefit from the development of more sensitive methods for determining non-responders. Earlier diagnoses of non-response or progressive disease will spare patients from the toxicities associated with futile treatments and access alternative therapies sooner.

FDG-PET/CT has been proposed as an imaging tool for the detection of response, by demonstrating metabolic changes in the tumor, early in the course of treatment (9;10). Early metabolic changes may also have prognostic value. In ovarian cancer, Avril et al found that in the neo-adjuvant setting, by using an a priori stated cut-off value for decrease of standardized uptake value (SUV) from baseline of 20% after the first cycle, median overall survival was 38.3 months in metabolic responders compared with 23.1 months in metabolic non-responders (11). There was a significant correlation between FDG-PET metabolic response after the first

($p=0.008$) and third ($p=0.005$) cycles of chemotherapy and overall survival. Importantly, standard clinical response criteria did not correlate with overall survival, suggesting that FDG-PET response may be a more powerful prognostic tool.

However, in order to adopt FDG-PET response into both clinical practice and drug-development, the range of variability (or confidence interval) surrounding the measurement of the standardized uptake value (SUV) of FDG tumor uptake must be determined, in order to be able to set appropriate cut-off values for identifying true responses in tumor tissue. The repeatability of tumor FDG uptake in lung and other solid organ tumors have recently been evaluated in a meta-analysis (12). However, FDG avid lesions in the abdomen and pelvis are often difficult to delineate and no test retest data regarding FDG uptake or tumor volume data have been published for EOC. More importantly, no previous study has compared the test retest variation in tumor FDG uptake with test retest variation in tumor volumes determined in combination with CECT. Repeatability data would allow the determination of robust cut-off values for defining metabolic response or metabolic progression in the absence of either complete disappearance of all lesions or appearance of one or more new lesions. Without this information, changes in SUV may be erroneously interpreted as response or progression and this could adversely affect patient care and clinical trial outcome. Also, although there is a general assumption that anatomical changes occur after several cycles of chemotherapy, there is little information available about early changes in tumor volumes and their ability to predict treatment response early during chemotherapy.

The purpose of this study was to establish the variation in the measurements of FDG uptake and tumor volumes in recurrent ovarian cancer. Our aims were to measure prospectively the test-retest repeatability of quantitative PET measurements (SUV_{mean}, SUV_{max}) using a standardized volume of interest as well as tumor diameter and tumor volume in a cohort of women with recurrent ovarian cancer treated at 2 sites using different PET/CT instruments.

PATIENTS AND METHODS

The study protocol was reviewed and approved by Cambridgeshire 2 Research Ethics Committee, UK (09/H0308/129). Patients were recruited by two academic oncology centers in the UK from a larger study cohort evaluating treatment response. All patients gave written informed consent. All screened patients had platinum-sensitive (defined as platinum-free interval of at least 6 months) ovarian cancer that had relapsed as confirmed by findings on

computed tomography (CT), with or without an elevated CA-125 level. The patient inclusion and exclusion criteria are listed in appendix 1. Of the cohort of 43 patients recruited into the main study, 21 patients agreed to take part in the test-retest sub-study. Center 1 recruited 8 patients and center 2 recruited 13 patients, mean age 60.4 years (median 61 years, range 38 to 74). Patients underwent two identical baseline imaging investigations prior to starting standard of care platinum-based chemotherapy. This study was performed according to the latest guidelines and recommendations from the MHRA and FDA for clinical trials as a prospective dual center study. All data were collected at the time of origin and collected in a secure database. The trial was funded and monitored by Merck and Co, which enabled us to conduct the study on the highest level of evidence possible.

Imaging techniques

Following enrolment, a baseline FDG-PET/CT scan was performed, immediately followed by a contrast-enhanced CT scan (CECT). This was termed baseline 1 (BL1). Patients who did not have at least one lesion with both, an SUVmax ≥ 2.5 at BL1 FDG-PET scan and a longest diameter lesion of ≥ 1.5 cm on the BL1 CECT scan were discontinued on the study. In those patients with at least one such lesion, imaging was repeated 1 - 7 days later (baseline 2 - BL2) prior to starting treatment.

FDG-PET imaging

Patients were imaged using a Gemini TF with a 64 channel CT (Philips Healthcare UK) in center 1 and a GE Discovery 690 (GE Healthcare UK) at center 2. Both PET/CT scanners were comparable in performance and both used time-of-flight technology. Both sites were qualified by the CRO (Perceptives Informatics, Billerica, MA, USA) and by the UK National Cancer Research Institute programme for PET sites involved in multi-centre trials. Daily quality control and regular standard calibration procedures were undertaken. The same PET/CT scanner was used for each patient throughout the study. All patients underwent the entire PET/CT imaging procedure twice within 7 days without any therapeutic interventions in-between.

Blood glucose levels were measured prior to the administration of the radiotracer (within 1 hour) for BL1 and BL2. Patients with a blood glucose level exceeding 150 mg/dL (8.3 mmol/L) were not injected with the radiotracer. The mean blood glucose level was 5.75mmol/l. The mean difference between BL1 and BL2 was 0.6 mmol/l, median 0.3 mmol/l and range 0 – 2.7 mmol/l.

The net dose of FDG injected was measured by placing the injection syringe in a dose calibrator before and after administration, with decay corrections factored into the calculation. All FDG doses were injected through a venous catheter. Patients rested for approximately 50 minutes in a comfortable recliner after FDG injection prior to PET/CT imaging. Patients were then asked to empty their bladder and positioned in prone on the scanner table. A scout scan was obtained to plan the imaging procedure. A transmission CT scan for attenuation correction was performed prior to the PET emission scan (at about 55 minutes post injection of FDG). PET emission scans started at 60 minutes (median = 60, mean = 61.4, range = 59 – 70 minutes) post-injection. If the 60-minute FDG uptake time target was missed, subsequent studies aimed for the actual uptake period at the first baseline FDG-PET/CT. The difference in FDG uptake time between the test and retest PET scans was between 0 to 3 minutes in all but three patients (in whom there was a difference of 6, 7 and 10 minutes respectively), with a mean difference of 1.9 minutes and median of 1 minute. The duration of all emission scans were identical for each PET/CT scanner. The acquisition parameters are given in appendix 2.

Contrast-enhanced CT

The CECT scan was performed directly following the baseline FDG-PET/CT scan including the abdomen and pelvis (and chest if clinically indicated). CECT was defined as a volumetric CT acquisition of the body using a multidetector spiral CT scanner in the portal venous phase following intravenous contrast administration (CECT acquisition parameters are available in supplementary table 2). Images were viewed on 5mm reformatted slices in the axial plane, as per RECIST 1.1 rules, with the option to view in reformatted sagittal or coronal planes. All target lesions were measured in the axial plane (the plane of acquisition). The CECT scans were of sufficient quality to enhance interpretation of FDG-PET scans, permit RECIST assessments, and enable tumor volume image analysis to be performed. For all CECT scans, intravenous iodinated contrast media were used according to local standards of care. If contrast media was contraindicated in a patient, then CECT scan was not performed and the test-retest measurements for RECIST and volumetric analysis could not be evaluated.

Image analysis

Measurement of SUV

FDG uptake in tumor lesions was quantitatively assessed using Standardized Uptake Values (SUV) as a measure for tumor glucose metabolism. Activity concentrations in the attenuation-corrected PET images were converted to SUV's by dividing the activity concentrations derived

from PET by the decay-corrected injected dose divided by the patient's body weight. The following SUV parameters were obtained within a volume of interest: maximum SUV (SUVmax), mean SUV (SUVmean) and the mean weighted SUV, which is defined as the sum of all counts in all of the volumes of interest (VOIs) representing all of the TLs divided by the sum of all the voxels in all of the VOIs representing the TLs (the SUV mean weighted average, SUVmwa).

Image reads

Analysis of the FDG-PET scan and the CECT was performed without knowledge of any specific clinical information apart from the inclusion and exclusion criteria. Two independent reads of the FDG-PET and CECT were made, one being a site read and the other an imaging CRO read, using two different methods. The reads were performed to reflect the practice of trial reporting whereby once targets have been chosen on the first baseline scan, measurements of the same targets are subsequently performed. Target selection was independent between site and CRO as two different reading methods were being evaluated for test retest repeatability.

Site reads

The baseline FDG-PET/CT and CECT were viewed simultaneously by the PET expert and the gynecologic oncology CT expert respectively. A maximum of 5 target lesions were selected from the CECT, maximum two per organ. Although the inclusion criteria required at least one lesion to be SUVmax ≥ 2.5 and diameter ≥ 15 mm, the criteria used for selecting other target lesions were that each target lesion was FDG avid and of minimal size criteria as defined by RECIST 1.1 (10mm long axis for non-nodal target and 15mm short axis for nodal target).

The longest diameter of each TL was then measured on the CECT according to RECIST rules. On the FDG-PET images, a spherical VOI with a diameter of 15mm was used to measure the SUVmean and SUVmax of each target, following manual identification of the most avid part of the tumor lesion. All measured parameters were recorded and screen shots of each selected target lesion were stored. The BL1 scan was then closed. The BL2 scan was then opened and each target lesion was measured using the same technique. Each target lesion was checked with BL1 to ensure that the same target lesions were used, but with blinding to the prior measurements.

Contract Research Organisation reads

CRO reads for SUV measurements and TD were considered the secondary reads. CECT volumes and FDG-PET images were read by a single independent radiologist with significant

experience in reading CT-volumes and FDG-PET. CECT images were assessed for target lesion diameter and volume. The CT lesion selection criteria followed the guidelines set forth by the RECIST allowing selection of up to 10 lesions $\geq 10\text{mm}$ to be selected with a maximum of 5 per organ. The FDG-PET images for this study were used to assess the SUVmax and SUVmwa for up to 10 lesions. As long as appropriate, the same target lesions were chosen on PET and CECT. However, when an FDG-avid lesion was not suitable for RECIST measurements or vice versa, measurements on the other modality were not enforced. Thirty-five percent of subjects had different numbers of targets in the two modalities (28% had additional FDG-PET targets, 7% had additional CECT targets). All but one PET avid target had an SUVmax ≥ 2.5 . CECT images were used to delineate the target lesions which were described as series of Regions of Interest (ROIs), drawn on each slice where present, to ensure the entire volume of the lesion was assessed (VOI). ROIs were created using a semiautomatic approach combining freehand and autosegmentation, which allowed adjustment by a radiologist. The ROIs for target lesions provided longest diameter for non-nodal lesions, longest short axis diameter for nodal lesion and volume assessment of the individual target lesions. The lesion locations from CECT images were used to follow consistently the target lesions on sequential imaging. FDG-PET images were viewed along with the CECT images to confirm the selection of the same lesion selected on CECT, up to the extent possible to meet lesion selection criteria. Metabolic volumes were determined by an isocontour of 25% of the SUVmax.

Statistical methods

Repeatability

When the number of target lesions increases, the sum of target lesions (in terms of SUV and tumor size) also increases. As a result, correlation between repeated scans may be inflated due to different numbers of target lesions across patients. Therefore, for repeatability assessment, the average was used to summarize measures across multiple target lesions within a scan.

The repeatability of SUV and tumor size measurements was assessed based on the two baseline scans. The Kendall tau and Shapiro-Wilk tests were performed on both original and log-transformed data and the log-transformed data were closer to normality and constant variance. Scatter plots (Scan 1 vs Scan 2) with a 45° line through the origin and Bland-Altman plots (difference vs mean) were generated. Concordance correlation coefficient (CCC), within subject standard deviation (SD), and within-subject coefficient of variation (CV) were derived. The difference between two baseline measurements for the same patient was considered to be

within the normal variation for 95% of pairs of observations. The repeatability coefficient was estimated to be twice the standard deviation of the paired differences. On a logarithmic scale, expressed as a percent change from baseline, the repeatability coefficient was $[1 - \exp(-2 * SD_{diff})] \times 100\%$.

RESULTS

21 patients underwent two baseline imaging studies. In one case, the diagnostic CECT component could not be done with intravenous contrast media on BL2, and therefore the CT components (TD and TV) were not evaluated for repeatability, with a final number of 20 patients evaluable for TD and TV. The number of target lesions in each data set is provided in table 1. For the primary site reads, the mean diameter of target lesions was 32mm and all target lesions were ≥ 15 mm except for 4 which were between 11.1 and 14.6mm.

Median time between sequential FDG-PET/CT was 2 days (range 1 to 7, mean 2.4 days). The repeatability of SUV and tumor size measurements were plotted in Figures 1 – 4. The paired values from the same subject fell near the solid line. The Pearson correlation between SUVmean and SUVmax was 0.95 for the site read. Concordance correlation coefficient (CCC) and repeatability cut-off values are given in table 2. Concordance correlation coefficient (and 80% confidence intervals) for SUVmean (average), SUVmax (average) and TD (average) were 0.95 (0.92-0.98), 0.94 (0.90- 0.97) and 0.99 (0.98 – 1.00) respectively. Repeatability cut-off values (indicating the lower limit of the 95% CI for % change between two baseline scans) were 16.3% for SUVmean (average), 17.3% for SUVmax (average) and 8.8% for TD (average). The repeatability results from the two reading methods were similar. Tumor volume (average) CCC was 0.99 (0.98 – 1.00) with a repeatability cut-off value of 28.1%.

The repeatability results are plotted in figures 1- 4.

DISCUSSION

FDG-PET is increasingly being used as a biomarker for treatment monitoring in cancer patients (13-15). In order to identify a metabolic response it is necessary to establish the normal variation of tumor FDG uptake prior to therapeutic intervention. We found a high test-retest repeatability of quantitative measures of tumor glucose metabolism (FDG uptake) using different parameters derived from Standardized Uptake Values (SUV) in ovarian cancer. This is in line with previous reports who observed a good repeatability of FDG uptake measurements in other tumors, as reported in a recent meta-analysis (12). The first study addressing this issue dates

back in 1999 and included 16 patients with different tumor types who underwent FDG-PET within 10 days without anti-cancer treatment in-between (16). In that study, the differences of repeated measurements were approximately normally distributed for all SUV parameters with a SD of the mean percentage difference of about 10%. The authors concluded that changes of a SUV parameter outside the 95% normal range may be used to define a metabolic response to therapy.

The PET technology has advanced since then from stand-alone PET scanners to combined PET-CT and subsequent studies have shown a high repeatability for measuring tumor FDG uptake using FDG-PET/CT in a limited number of patients (17;18). This is the first study to address this issue in ovarian cancer. The abdomen is particularly difficult to quantitatively evaluate for tumor FDG uptake due to physiologic excretion of FDG via the urinary tract and due to variable physiologic FDG uptake within bowel structures. Ovarian cancer often presents with serosal implants, which can be subject to motion, potentially compromising longitudinal FDG-PET imaging for treatment monitoring. To date, only few studies have addressed FDG-PET treatment monitoring in ovarian cancer. Our study is of potentially high clinical relevance as we showed that changes beyond 15-20% in tumor glucose metabolism allows the identification of treatment induced changes, which provides the basis for future prospective treatment monitoring studies in ovarian cancer. There are a number of targeted therapies currently in development or already being evaluated clinically. This includes drugs targeting the PI3K – Akt-mTOR pathway, the inhibition of angiogenesis as well as specific inhibitors of Interleukin-6 and Stat3 amongst other targets (19-23). A significant challenge is the evaluation of their therapeutic effectiveness as they are often cytostatic rather than cytotoxic and changes in tumor size occur late if at all. It is recognised that anti-angiogenic agents may therefore result in stable disease according to RECIST or PERSIST criteria although tumour necrosis may be seen following treatment. The data from the current study may contribute to support robust metabolic response criteria in the absence of change in tumor size when targeted therapy is being used.

Biomarkers that predict whether a drug will lengthen progression free survival (PFS) or overall survival (OS) are therefore needed, both for optimising the clinical management of patients and to accelerate decisions concerning novel drug efficacy in clinical trials. Bi-dimensional (WHO) or uni-dimensional (RECIST) measurements have been used for many years to monitor objective response to chemotherapy but certain limitations are recognised: 1. measurable changes take

time to allow for tumor shrinkage; 2. in some cases, clinical response to chemotherapy does not result in a change in tumor dimensions; 3. tumor response measured by RECIST does not always correlate well with PFS or OS (4;6;8;24). In addition, measuring the tumor dimension might be an effective criterion in the setting of spherical tumors that change size in a uniform fashion following therapy; however, tumor morphologies that do not manifest in this idealized shape can be challenging to evaluate for changes in size using uni-dimensional line lengths.

An important strength of our study is that we fully utilized the capability of FDG-PET/CT by performing a CECT after completion of the PET data acquisition. No such CECT repeatability study has previously been performed in ovarian cancer patients. Changes in tumor size are generally believed to occur at later time points after start of treatment as compared to metabolic changes. Tumor size measurements can be affected by partial volume artefacts particularly when the target lesion is small when using a CT slice thickness of 5mm and in addition, some bowel movement can occur during the time of CT acquisition further compounding partial volume artefacts. However, we found a test retest variation as low as 8.8% for tumor diameter when measured sequentially. We also assessed changes in tumor volume and found a test retest variation of 28.1%; despite this wider variation between repeat measurements in volume, changes in tumor volume following treatment may be more sensitive than changes in diameter and establishing the repeatability coefficient is thus highly relevant. It is important to point out that defining a tumor volume in the abdomen is particularly difficult and automated software algorithms for that purpose are currently under investigation. Such algorithms which render volumetric and tissue density measurements have been successfully used in the lungs: a study demonstrated that a semi-automated algorithm was able to accurately segment 14 out of 15 patient tumors imaged in thin section CT scans (7;25). Follow-up CT at 3 weeks after start of gefitinib showed that 73% of patients had an absolute change in tumor volume of at least 20%; in contrast, only 7% and 27% of patients showed similar changes in their tumor sizes using either uni-dimensional or bi-dimensional measurements, respectively.

An important strength, but also a potential weakness of our study is the highly standardized imaging environment with strong academic support. Highly experienced PET and CT reader have jointly interpreted the images, which might at least have partially contributed to the superiority of tumor size measurements over semi-automated volume measurements. Nevertheless, our data also demonstrate that a company based image analysis (CRO read) produced results comparable to the site reads. Our approach in this regard is novel as it directly allowed a comparison between an academic and a commercial setting for image analysis.

The FDG uptake measurements are affected by numerous factors and a meticulous quality assurance program needs to be place in order to achieve repeatable PET measurements. This includes patient preparation, obtaining blood glucose levels, measuring precisely the amount of injected FDG activity as well as scanner calibration amongst others. Of note, a recent study found much greater variance of SUV uptake measurements in a clinical PET setting when compared to ideal study settings (26). The variation in FDG uptake time is an important limitation in the repeatability of FDG-PET and specific efforts need to be place to ensure timely procedures. It is of crucial importance for oncologists to work closely together with their imaging group to ensure that procedures are in place to enable PET treatment monitoring studies.

A limitation of our study is that we did not independently repeat each of the two reading methods by a further reader, but rather we compared two independent reading methods. However, we have attempted to recreate the method used in standard trial sequential reporting to closely reflect clinical practice. Also, the selection of target lesions on the CECT was aided by simultaneous viewing of the PET images which could have resulted in an increase in the detection of lesions on CT.

CONCLUSION:

We have shown excellent test/retest repeatability for FDG-PET/CT quantitative measurements in recurrent ovarian cancer across two independent reading methods. The repeatability coefficients suggest that a decrease in FDG uptake (SUV) of 15-20% from baseline and decrease in tumor size between 10-15% could be used to determine early tumor response.

ACKNOWLEDGEMENTS

We would like to acknowledge our patients and their care-givers; the help and support of the research nurses, trial staff and the staff in the PET centers. In addition, we acknowledge the help of Mark Utley, Gary Herman, Jeffrey Evelhoch, Eric Rubin (Merck and Co); Faith Dzumbunu, Craig Copland, Iain Murray (Barts Cancer Institute); Nick Bird, Charlotte Hodgkin (Addenbrooke's Hospital). We acknowledge additional support from Cancer Research UK, National Institute for Health Research (NIHR) Cambridge Biomedical Research Centre, Cambridge Experimental Cancer Medicine Centre and Hutchison Whampoa Limited.

Reference List

- (1) Coleman MP, Forman D, Bryant H, Butler J, Rachet B, Maringe C et al. Cancer survival in Australia, Canada, Denmark, Norway, Sweden, and the UK, 1995-2007 (the International Cancer Benchmarking Partnership): an analysis of population-based cancer registry data. *Lancet* 2011; 377(9760):127-138.
- (2) Chan S, Griffin M, Stewart J, Gregory K, Hughes A, Awwad S et al. Modern chemotherapy management of recurrent ovarian cancer: a multicentre study. *Clin Oncol (R Coll Radiol)* 2007; 19(2):129-134.
- (3) Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer* 2009; 45(2):228-247.
- (4) Eisenhauer EA. Optimal assessment of response in ovarian cancer. *Ann Oncol* 2011; 22 Suppl 8:viii49-viii51.
- (5) Buckler AJ, Schwartz LH, Petrick N, McNitt-Gray M, Zhao B, Fenimore C et al. Data sets for the qualification of volumetric CT as a quantitative imaging biomarker in lung cancer. *Opt Express* 2010; 18(14):15267-15282.
- (6) Villaruz LC, Socinski MA. The clinical viewpoint: definitions, limitations of RECIST, practical considerations of measurement. *Clin Cancer Res* 2013; 19(10):2629-2636.
- (7) Zhao B, Schwartz LH, Moskowitz CS, Ginsberg MS, Rizvi NA, Kris MG. Lung cancer: computerized quantification of tumor response--initial results. *Radiology* 2006; 241(3):892-898.
- (8) Shankar LK, Van den AA, Yap J, Benjamin R, Scheutze S, Fitzgerald TJ. Considerations for the use of imaging tools for phase II treatment trials in oncology. *Clin Cancer Res* 2009; 15(6):1891-1897.
- (9) Weber WA. Positron emission tomography as an imaging biomarker. *J Clin Oncol* 2006; 24(20):3282-3292.
- (10) 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-150S.
- (11) Avril N, Sassen S, Schmalfeldt B, Naehrig J, Rutke S, Weber WA et al. Prediction of response to neoadjuvant chemotherapy by sequential F-18-fluorodeoxyglucose positron emission tomography in patients with advanced-stage ovarian cancer. *J Clin Oncol* 2005; 23(30):7445-7453.
- (12) de Langen AJ, Vincent A, Velasquez LM, van Tinteren H, Boellaard R, Shankar LK et al. Repeatability of 18F-FDG uptake measurements in tumors: a metaanalysis. *J Nucl Med* 2012; 53(5):701-708.
- (13) Mghanga FP, Lan X, Bakari KH, Li C, Zhang Y. Fluorine-18 fluorodeoxyglucose positron emission tomography-computed tomography in monitoring the response of breast

- cancer to neoadjuvant chemotherapy: a meta-analysis. *Clin Breast Cancer* 2013; 13(4):271-279.
- (14) Dupas B, Augeul-Meunier K, Frampas E, Bodet-Milin C, Gastinne T, Le Gouill S. Staging and monitoring in the treatment of lymphomas. *Diagn Interv Imaging* 2013.
 - (15) Hoekstra CJ, Hoekstra OS, Stroobants SG, Vansteenkiste J, Nuyts J, Smit EF et al. Methods to monitor response to chemotherapy in non-small cell lung cancer with 18F-FDG PET. *J Nucl Med* 2002; 43(10):1304-1309.
 - (16) Weber WA, Ziegler SI, Thodtmann R, Hanauske AR, Schwaiger M. Reproducibility of metabolic measurements in malignant tumors using FDG PET. *J Nucl Med* 1999; 40(11):1771-1777.
 - (17) Velasquez LM, Boellaard R, Kollia G, Hayes W, Hoekstra OS, Lammertsma AA et al. Repeatability of 18F-FDG PET in a multicenter phase I study of patients with advanced gastrointestinal malignancies. *J Nucl Med* 2009; 50(10):1646-1654.
 - (18) van Velden FH, Nissen IA, Jongsma F, Velasquez LM, Hayes W, Lammertsma AA et al. Test-Retest Variability of Various Quantitative Measures to Characterize Tracer Uptake and/or Tracer Uptake Heterogeneity in Metastasized Liver for Patients with Colorectal Carcinoma. *Mol Imaging Biol* 2013.
 - (19) Coward J, Kulbe H, Chakravarty P, Leader D, Vassileva V, Leinster DA et al. Interleukin-6 as a therapeutic target in human ovarian cancer. *Clin Cancer Res* 2011; 17(18):6083-6096.
 - (20) Perren TJ, Swart AM, Pfisterer J, Ledermann JA, Pujade-Lauraine E, Kristensen G et al. A phase 3 trial of bevacizumab in ovarian cancer. *N Engl J Med* 2011; 365(26):2484-2496.
 - (21) Ledermann JA, Hackshaw A, Kaye S, Jayson G, Gabra H, McNeish I et al. Randomized phase II placebo-controlled trial of maintenance therapy using the oral triple angiokinase inhibitor BIBF 1120 after chemotherapy for relapsed ovarian cancer. *J Clin Oncol* 2011; 29(28):3798-3804.
 - (22) Hall M, Gourley C, McNeish I, Ledermann J, Gore M, Jayson G et al. Targeted anti-vascular therapies for ovarian cancer: current evidence. *Br J Cancer* 2013; 108(2):250-258.
 - (23) Baumann KH, Wagner U, du BA. The changing landscape of therapeutic strategies for recurrent ovarian cancer. *Future Oncol* 2012; 8(9):1135-1147.
 - (24) Choi H, Charnsangavej C, Faria SC, Macapinlac HA, Burgess MA, Patel SR et al. Correlation of computed tomography and positron emission tomography in patients with metastatic gastrointestinal stromal tumor treated at a single institution with imatinib mesylate: proposal of new computed tomography response criteria. *J Clin Oncol* 2007; 25(13):1753-1759.

- (25) Zhao B, Oxnard GR, Moskowitz CS, Kris MG, Pao W, Guo P et al. A pilot study of volume measurement as a method of tumor response evaluation to aid biomarker development. *Clin Cancer Res* 2010; 16(18):4647-4653.
- (26) Kumar V, Nath K, Berman CG, Kim J, Tanvetyanon T, Chiappori AA et al. Variance of SUVs for FDG-PET/CT is greater in clinical practice than under ideal study settings. *Clin Nucl Med* 2013; 38(3):175-182.

Legends for all figures (Figures 1 to 4)

Left side (a,c) on original scale and Right side (b,d) on log scale. Top (a, b): Scan 1 value versus Scan 2 value. The paired values from the same subject are plotted against each other (the first observation on the y-axis and the second on the x-axis). If the paired values from the same subject are similar, the points will fall near the solid line. Bottom (c,d): Difference versus Average. The differences and means between the paired values are plotted to see if a trend exists. Horizontal lines: dashed lines=mean difference, dotted lines=lower and upper 95% confidence limits for difference.

Figure 1. Repeatability of SUVmean at Baseline (Site read)

Figure 2. Repeatability of SUVmax at Baseline (Site read)

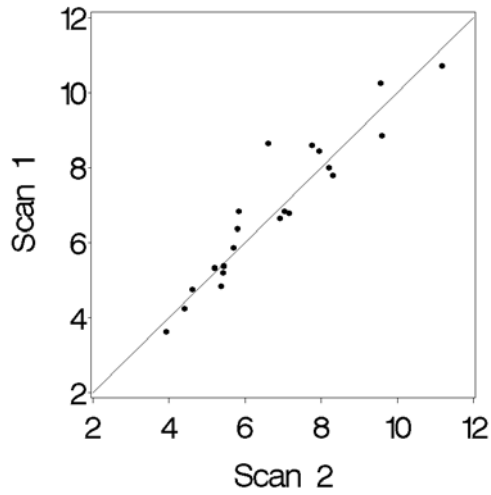
Figure 3. Repeatability of Tumor (lesion longest) Diameter (mm) at Baseline (Site read)

Figure 4. Repeatability of Tumor Volume (cc) at Baseline (CRO read)

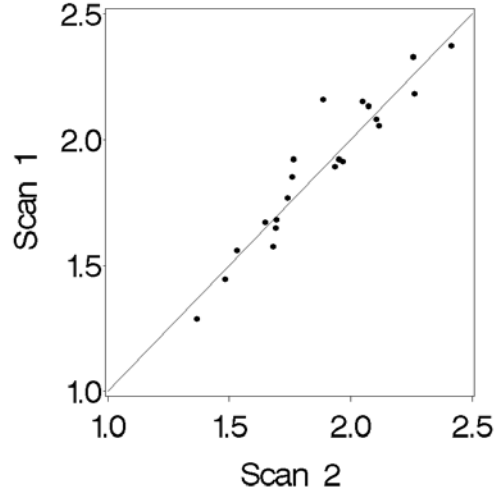
Figure 1. Repeatability of SUVmean at Baseline (Site read)

SUV mean (average of target lesions)

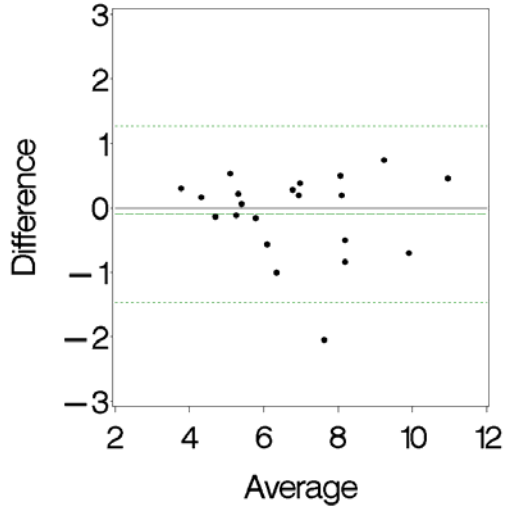
(a)



(b)



(c)



(d)

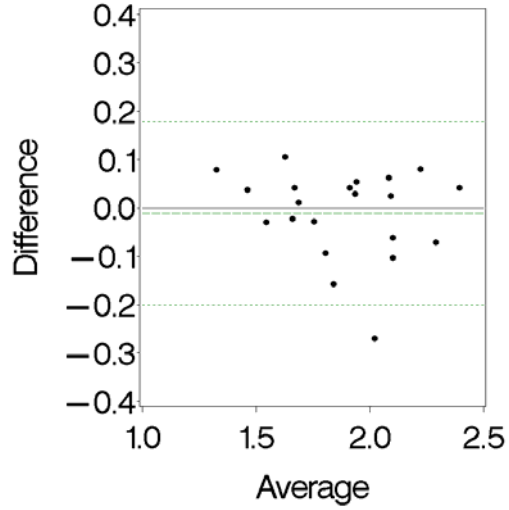
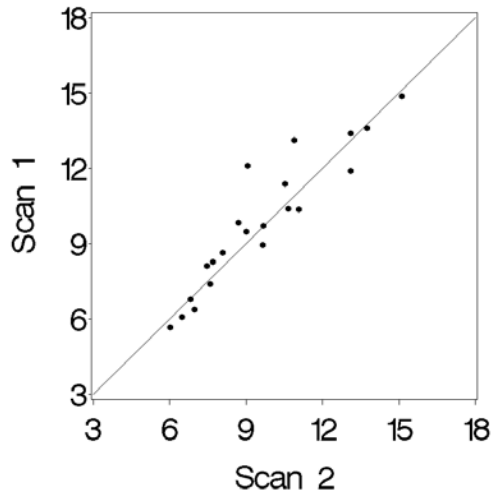


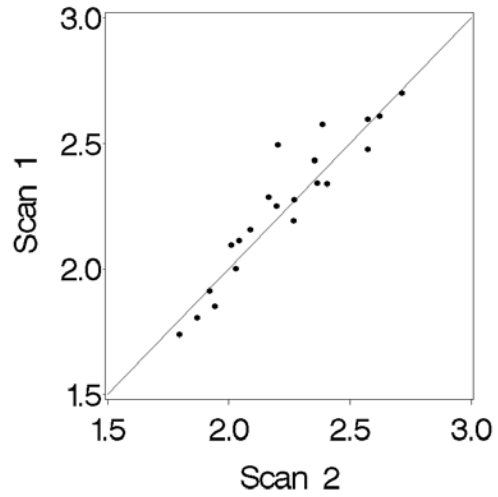
Figure 2. Repeatability of SUVmax at Baseline (Site read)

SUV max (average of target lesions)

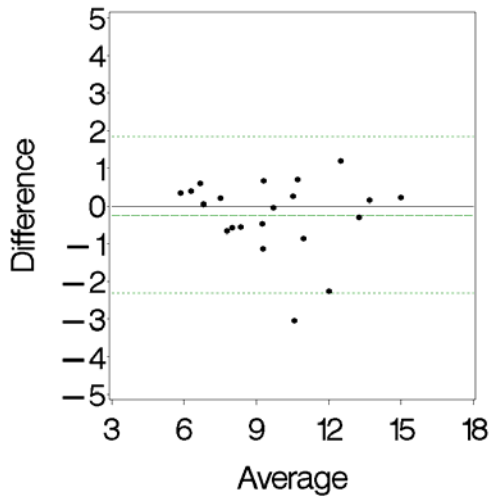
(a)



(b)



(c)



(d)

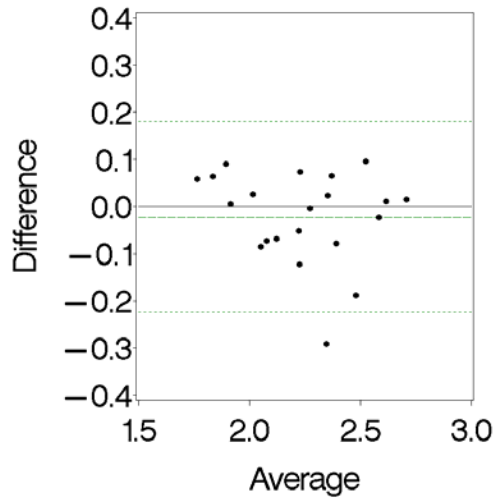


Figure 3. Repeatability of Tumor (lesion longest) Diameter (mm) at Baseline (Site read)

Tumor Diameter (average of target lesions)

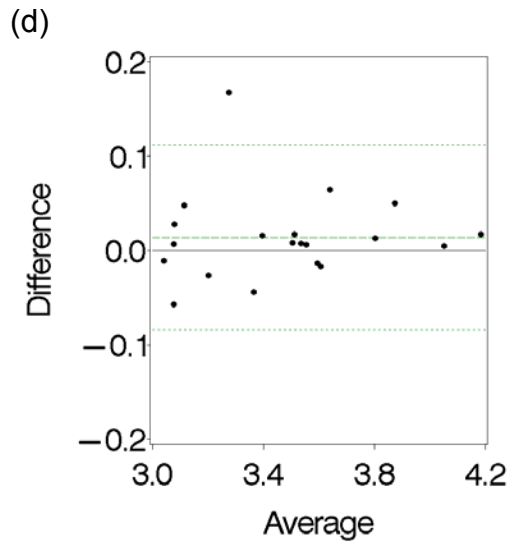
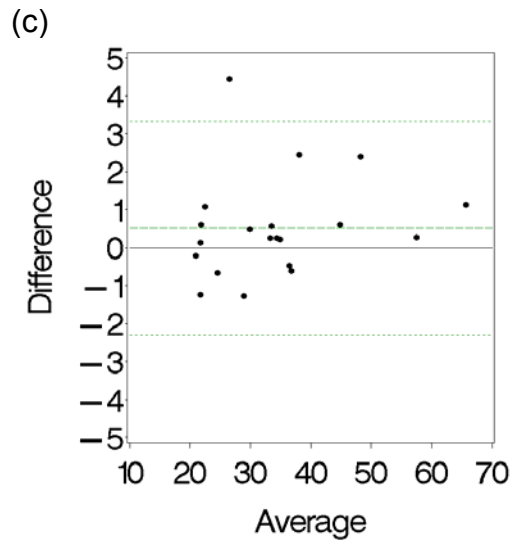
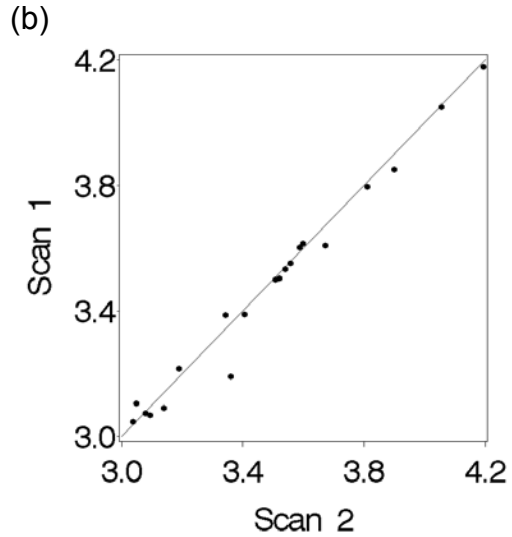
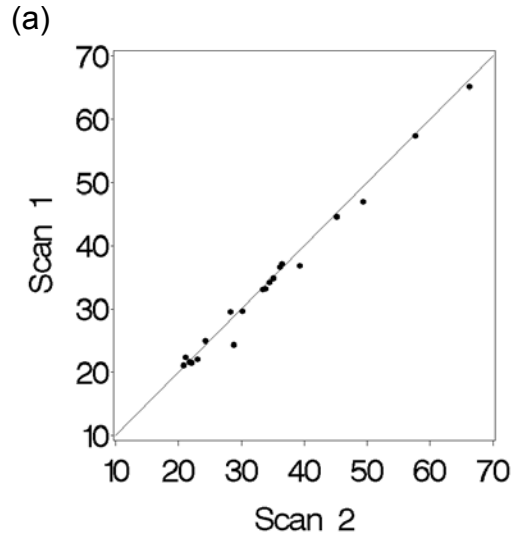
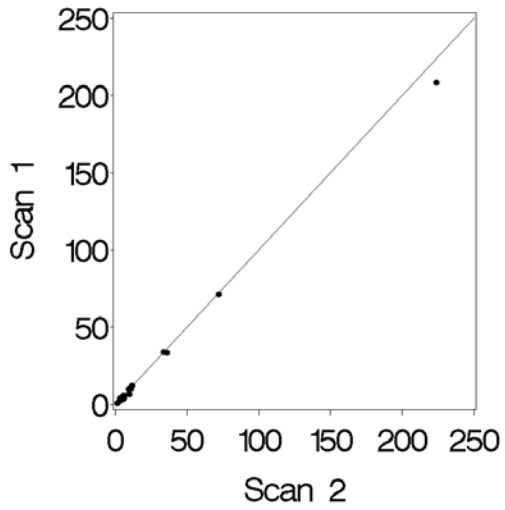


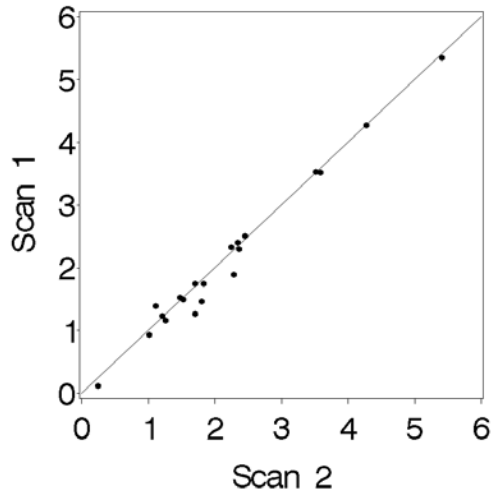
Figure 4. Repeatability of Tumor Volume (cc) at Baseline (CRO read)

Tumor Volume (average of target lesions)

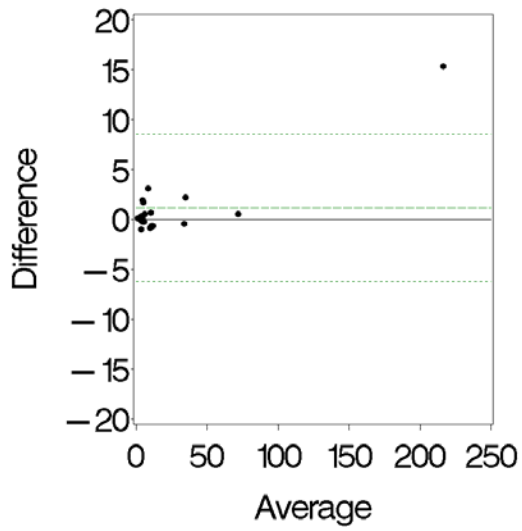
(a)



(b)



(c)



(d)

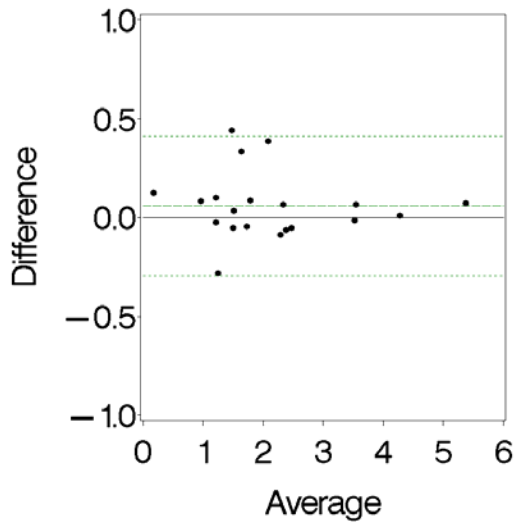


Table 1. Number of target lesions (Site and CRO reads)

	Site Read Combined PET and CT target lesions	CRO Read PET target lesions	CRO Read CT target lesions
Number of patients	21	21	21
Mean number of target lesions	4.0	4.4	3.8
SD	1.3	2.4	2.2
Median	5.0	5.0	3.0
Range	(2, 5)	(1, 8)	(1, 8)
Total target lesions	85	93	80
Lesion numbers per patient			
1 lesion	0	2 (9.5%)	3 (14.3%)
2 lesions	4 (19.0%)	5 (23.8%)	5 (23.8%)
3 lesions	4 (19.0%)	2 (9.5%)	3 (14.3%)
4 lesions	0	1 (4.8%)	0
5 lesions	13 (61.9%)	4 (19.0%)	6 (28.6%)
6 lesions	0	1 (4.8%)	1 (4.8%)
7 lesions	0	3 (14.3%)	2 (9.5%)
8 lesions	0	3 (14.3%)	1 (4.8%)

Table 2. Repeatability of FDG SUV and Tumor Size Measures at Baseline

PET and CT	n	CCC and 80% CI [†]	Geo. Mean [‡]	Geo. SD and 80% CI [§]	Geo. CV and 80% CI [*]	Repeatability Cut-off [¶] % Decrease
Site Read, up to 5 target lesions						
SUV mean (avg)	21	0.95 (0.92,0.98)	6.66	1.07 (1.06,1.08)	6.6 (5.5,8.5)	16.3
SUV max (avg)	21	0.94 (0.90,0.97)	9.69	1.07 (1.06,1.09)	7.1 (5.9,9.1)	17.3
TD (mm) (avg)	20	0.99 (0.98,1.00)	31.6	1.03 (1.03,1.04)	3.4 (2.8,4.3)	8.8
CRO Read, up to 10 target lesions						
SUV mwa	21	0.98 (0.97,0.99)	4.97	1.06 (1.05,1.08)	6.3 (5.3,8.1)	15.6
SUV max (avg)	21	0.98 (0.96,0.99)	7.85	1.07 (1.06,1.09)	7.2 (6.0,9.2)	17.6
TD (mm) (avg)	20	0.97 (0.95,0.99)	27.7	1.06 (1.05,1.08)	6.0 (5.0,7.7)	14.8
TV (cc) (avg)	20	0.99 (0.98,1.00)	7.77	1.13 (1.10,1.16)	12.7 (10.5,16.4)	28.1
[†] Concordance correlation coefficient and 80% confidence interval [‡] Geometric grand mean [§] Geometric within-patient standard deviation and 80% confidence interval [*] Geometric within-patient coefficient of variation and 80% confidence interval [¶] Repeatability Cut-off = the lower limit of the 95% CI for % change between two baseline scans TD: tumor diameter; TV: tumor volume; avg: average						