Repeatability of Quantitative FDG-PET/CT and Contrast-Enhanced CT in Recurrent Ovarian Carcinoma: Test–Retest Measurements for Tumor FDG Uptake, Diameter, and Volume

Andrea G. Rockall1, Norbert Avril2, Raymond Lam10, Robert Iannone10, P. David Mozley12, Christine Parkinson5,7, Donald Bergstrom11, Evis Sala6, Shah-Jalal Sarker4, Iain A. McNeish2,4, and James D. Brenton5,7,8,9

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 2[18F]fluoro-2-deoxy-D-glucose (FDG) uptake, tumor diameter, and tumor volume from sequential FDG-PET/CT and contrast-enhanced computed tomography (CECT) in patients with recurrent platinum-sensitive ovarian cancer.

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

Results: Median time between FDG-PET/CT was two days (range 1–7). For site reads, concordance correlation coefficients (CCC) for SUVmean, SUVmax, and tumor diameter were 0.95, 0.94, and 0.99, respectively. Repeatability coefficients were 16.3%, 17.3%, and 8.8% for SUVmean, SUVmax, and tumor diameter, respectively. Similar results were observed for CRO reads. Tumor volume 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. Cutoff values for determining change in SUVmean, SUVmax, and tumor volume establish limits to determine metabolic and/or volumetric response to treatment in platinum-sensitive relapsed ovarian cancer. Clin Cancer Res; 20(10); 1–10. ©2014 AACR.

Background

Epithelial ovarian cancer (EOC) remains the most lethal gynecologic malignancy, and overall survival (OS) 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 (PFS). Objective response measurement using Response Evaluation Criteria in Solid Tumors (RECIST) 1.1 criteria (unidimensional 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 unidimensional measurements. Notably, the time taken for tumor shrinkage of 30% is typically in excess of 9

Note: Supplementary data for this article are available at Clinical Cancer Research Online (http://clincancerres.aacrjournals.org/).

Corresponding Author: Andrea G. Rockall, Division of Cancer and Surgery, Imperial College London, First Floor, ICTEM Building, Hammersmith Campus, Du Cane Road, London W12 ONN, United Kingdom. Phone: 44-207-594-2793; Fax: 44-208-383-5880. E-mail: a.rockall@imperial.ac.uk

doi: 10.1158/1078-0432.CCR-13-2634
©2014 American Association for Cancer Research.
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–retest data for measurements of \( ^{18}\text{F} \)-fluoro-2-deoxy-D-glucose (FDG) uptake, tumor diameter, and tumor volume are essential 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.

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 computed tomography (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 nonresponders. Earlier diagnoses of nonresponse 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 and colleagues found that in the neoadjuvant setting, by using an \textit{a priori} stated cutoff value for decrease of standardized uptake value (SUV) from baseline of 20% after the first cycle, median OS was 38.3 months in metabolic responders compared with 23.1 months in metabolic nonresponders (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 OS. Importantly, standard clinical response criteria did not correlate with OS, suggesting that FDG-PET response may be a more powerful prognostic tool.

However, to adopt FDG-PET response into both clinical practice and drug development, the range of variability (or confidence interval; CI) surrounding the measurement of the SUV of \( ^{18}\text{F} \)-fluoro-2-deoxy-D-glucose (FDG) tumor uptake must be determined, to be able to set appropriate cutoff values for identifying true responses in tumor tissue. The repeatability of tumor FDG uptake in lung and other solid organ tumors has 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 about 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 cutoff 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 positron emission tomography (PET) measurements (SUV-mean, SUVmax) using a standardized volume of interest (VOI) as well as tumor diameter and tumor volume in a cohort of women with recurrent ovarian cancer treated at two sites using different PET/CT instruments.

Patients and Methods

The study protocol was reviewed and approved by Cambridgeshire 2 Research Ethics Committee, United Kingdom (09/H0308/129). Patients were recruited by two academic oncology centers in the United Kingdom 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 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 substudy. Center 1 recruited 8 patients and center 2 recruited 13 patients, mean age 60.4 years (median 61 years, range 38–74). Patients underwent two identical baseline imaging investigations before starting standard-of-care platinum-based chemotherapy. This study was performed according to the latest guidelines and recommendations from the Medicines and Healthcare Products Regulatory Agency (MHRA) and U.S. Food and Drug Administration for clinical trials as a prospective dual center study. All data

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–retest data for measurements of \( ^{18}\text{F} \)-fluoro-2-deoxy-D-glucose (FDG) uptake, tumor diameter, and tumor volume are essential 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.


Clinical Cancer Research

Published OnlineFirst February 26, 2014; DOI: 10.1158/1078-0432.CCR-13-2634

Downloaded from clinincancersres.aacrjournals.org on April 15, 2017. © 2014 American Association for Cancer Research.
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 enrollment, a baseline FDG-PET/CT scan was performed, immediately followed by a CECT. This was termed baseline 1 (BL1). Patients who did not have at least one lesion with both, 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 to 7 days later (baseline 2; BL2) before starting the treatment.

**FDG-PET imaging.** Patients were imaged using a Gemini TF with a 64 channel CT (Philips Healthcare) in center 1 and a GE Discovery 690 (GE Healthcare) at center 2. Both PET/CT scanners were comparable in performance and both used time-of-flight technology. Both sites were qualified by the clinical research organization (CRO: Perceptives Informatics) and by the UK National Cancer Research Institute programme for PET sites involved in multicentre 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 before 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 5.75 mmol/L. The mean difference between BL1 and BL2 was 0.6 mmol/L, median 0.3 mmol/L, and range 0 to 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 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 S2). Images were viewed on 5 mm 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 were 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 SUV as a measure for tumor glucose metabolism. Activity concentrations in the attenuation-corrected PET images were converted to SUVs 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 VOI: maximum SUV (SUVmax), mean SUV (SUVmean), and the mean weighted SUV, which is defined as the sum of all counts in all of the VOIs representing all of the target lesions divided by the sum of all the voxels in all of the VOIs representing the target lesions (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 five 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 ≥ 15mm, 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 (10 mm long axis for non-nodal target and 15 mm short axis for nodal target).
The longest diameter of each target lesion was then measured on the CECT according to RECIST rules. On the FDG-PET images, a spherical VOI with a diameter of 15 mm 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 to 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 organization reads.** CRO reads for SUV measurements and tumor diameter 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 to be selected with a maximum of five 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 percentage 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 SUVmax >2.5. CECT images were used to delineate the target lesions which were described as series of regions of interest (ROI), drawn on each slice where present, to ensure the entire volume of the lesion was assessed (VOI). ROIs were created using a semi-automatic approach combining freehand and auto-segmentation, 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 the lesion selection criteria. Metabolic volumes were determined by an isocontour of 25% of the SUVmax.

**Statistical analysis**

**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 because of different numbers of target lesions across the 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 on the basis of 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 SD, and within-subject coefficient of variation 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 SD of the paired differences. On a logarithmic scale, expressed as a percentage change from baseline, the repeatability coefficient was \[1 - \exp(-2 \times SD_{diff}) \times 100\%\].

**Results**

Twenty-one 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 (tumor diameter and tumor volume) were not evaluated for repeatability, with a final number of 20 patients evaluable for tumor diameter and tumor volume. The number of target lesions in each dataset is provided in Table 1. For the primary site reads, the mean diameter of target lesions was 32 mm, and all target lesions were ≥15 mm except for four that were between 11.1 mm and 14.6 mm.

The Pearson correlation between SUVmean and SUVmax was 0.95 for the site read. CCC and repeatability cutoff values are given in Table 2. CCCs (and 80% CIs) for SUVmean (average), SUVmax (average), and tumor diameter (average) were 0.95 (0.92–0.98), 0.94 (0.90–0.97), and 0.99 (0.98–1.00), respectively. Repeatability cutoff 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 tumor diameter (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 cutoff value of 28.1%.

The repeatability results are plotted in Figs. 1–4.

**Discussion**

FDG-PET is increasingly being used as a biomarker for treatment monitoring in patients with cancer (13–15). To identify a metabolic response, it is necessary to establish the normal variation of tumor FDG uptake before therapeutic intervention. We found a high test–retest repeatability of quantitative measures of tumor glucose metabolism (FDG uptake) using different parameters derived from 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 anticancer 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

<table>
<thead>
<tr>
<th>Number of target lesions (site and CRO reads)</th>
<th>Site read Combined PET and CT target lesions</th>
<th>CRO read PET target lesions</th>
<th>CRO read CT target lesions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of patients</td>
<td>21</td>
<td>21</td>
<td>21</td>
</tr>
<tr>
<td>Mean number of target lesions</td>
<td>4.0</td>
<td>4.4</td>
<td>3.8</td>
</tr>
<tr>
<td>SD</td>
<td>1.3</td>
<td>2.4</td>
<td>2.2</td>
</tr>
<tr>
<td>Median</td>
<td>5.0</td>
<td>5.0</td>
<td>3.0</td>
</tr>
<tr>
<td>Range</td>
<td>2–5</td>
<td>1–8</td>
<td>1–8</td>
</tr>
<tr>
<td>Total target lesions</td>
<td>85</td>
<td>93</td>
<td>80</td>
</tr>
</tbody>
</table>

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 anticancer 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

Figure 1. Repeatability of SUVmean at Baseline (Site read). 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.
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% to 20% in tumor glucose metabolism allow the identification of treatment-induced changes, which provide 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 recognized that antiangiogenic agents may therefore result in stable disease according to RECIST or PERSIST criteria although tumor 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 PFS or OS are therefore needed, both for optimizing the clinical management of patients and to accelerate decisions concerning novel drug efficacy in clinical trials. Bidimensional (World Health Organization) or unidimensional (RECIST) measurements have been used for many years to monitor objective response to chemotherapy but certain limitations are recognized: (i) measurable changes take time to allow for tumor shrinkage, (ii) in some cases, clinical response to chemotherapy does not result in a change in tumor dimensions, and (iii) 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 the 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 unidimensional 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 patients with ovarian cancer. Changes in tumor size are generally believed to occur at later time points after start of the treatment as compared with 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 5 mm 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 semiautomated 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 unidimensional or bidimensional 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 semiautomated volume measurements. Nevertheless, our data also demonstrate that a company-based image analysis (CRO read) produced results comparable with 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 placed 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 with ideal study settings (26). The variation in FDG

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

Table 2. Repeatability of FDG SUV and tumor size measures at baseline

<table>
<thead>
<tr>
<th>PET and CT</th>
<th>n</th>
<th>CCC and 80% CIa</th>
<th>Geo. meanb</th>
<th>Geo. SD and 80% CIc</th>
<th>Geo. CV and 80% CId</th>
<th>Repeatability cutoff e decrease</th>
</tr>
</thead>
<tbody>
<tr>
<td>Site read, up to 5 target lesions</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SUVmean (avg.)</td>
<td>21</td>
<td>0.95 (0.92–0.98)</td>
<td>6.66</td>
<td>1.07 (1.06–1.08)</td>
<td>6.6 (5.5–8.5)</td>
<td>16.3</td>
</tr>
<tr>
<td>SUVmax (avg.)</td>
<td>21</td>
<td>0.94 (0.90–0.97)</td>
<td>9.69</td>
<td>1.07 (1.06–1.09)</td>
<td>7.1 (5.9–9.1)</td>
<td>17.3</td>
</tr>
<tr>
<td>TD (mm; avg.)</td>
<td>20</td>
<td>0.99 (0.98–1.00)</td>
<td>31.6</td>
<td>1.03 (1.03–1.04)</td>
<td>3.4 (2.8–4.3)</td>
<td>8.8</td>
</tr>
<tr>
<td>CRO read, up to 10 target lesions</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SUVmwa</td>
<td>21</td>
<td>0.98 (0.97–0.99)</td>
<td>4.97</td>
<td>1.06 (1.05–1.08)</td>
<td>6.3 (5.3–8.1)</td>
<td>15.6</td>
</tr>
<tr>
<td>SUVmax (avg.)</td>
<td>21</td>
<td>0.98 (0.96–0.99)</td>
<td>7.85</td>
<td>1.07 (1.06–1.09)</td>
<td>7.2 (6.0–9.2)</td>
<td>17.6</td>
</tr>
<tr>
<td>TD (mm; avg.)</td>
<td>20</td>
<td>0.97 (0.95–0.99)</td>
<td>27.7</td>
<td>1.06 (1.05–1.08)</td>
<td>6.0 (5.0–7.7)</td>
<td>14.8</td>
</tr>
<tr>
<td>TV (cc; avg.)</td>
<td>20</td>
<td>0.99 (0.98–1.00)</td>
<td>7.77</td>
<td>1.13 (1.10–1.16)</td>
<td>12.7 (10.5–16.4)</td>
<td>28.1</td>
</tr>
</tbody>
</table>

Abbreviations: TD, tumor diameter; TV, tumor volume; avg., average; CV, coefficient of variation.
aCCC and 80% CI.
bGeometric grand mean.
cGeometric within-patient SD and 80% CI.
dGeometric within-patient coefficient of variation and 80% CI.
eRepeatability cutoff = the lower limit of the 95% CI for % change between two baseline scans.
uptake time is an important limitation in the repeatability of FDG-PET and specific efforts need to be placed 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% to 20% from baseline and decrease in tumor size between 10% and 15% could be used to determine early tumor response.

Disclosure of Potential Conflicts of Interest

A.G. Rockall has received honoraria from the speakers' bureau of Novartis and from the speakers' bureau of Guerbet for educational lectures. R. Lam is a principal scientist in Merck. R. Iannone is a full-time employee in Merck and Co. Inc. and has ownership interest (including patents) in Merck. No potential conflicts of interest were disclosed by the other authors.

References


Clinical Cancer Research

Repeatability of Quantitative FDG-PET/CT and Contrast-Enhanced CT in Recurrent Ovarian Carcinoma: Test–Retest Measurements for Tumor FDG Uptake, Diameter, and Volume

Andrea G. Rockall, Norbert Avril, Raymond Lam, et al.

Clin Cancer Res Published OnlineFirst February 26, 2014.

Updated version
Access the most recent version of this article at:
doi:10.1158/1078-0432.CCR-13-2634

Supplementary Material
Access the most recent supplemental material at:
http://clincancerres.aacrjournals.org/content/suppl/2014/02/25/1078-0432.CCR-13-2634.DC1

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.