Feasibility and performance of novel software to quantify metabolically active volumes and 3D partial volume corrected SUV and metabolic volumetric products of spinal bone marrow metastases on ^{18}F-FDG-PET/CT

Drew A. Torigian[1] MD, MA
Rosa Fernandez Lopez[2] MD
Sridevi Alapati[1] MBBS
Geetha Bodapati[1] MBBS
Frank Hofheinz[3] PhD
Joerg van den Hoff[4] PhD
Babak Saboury[1] MD, MPH
Abass Alavi[1] MD, MD (Hon.), PhD (Hon.), DSc (Hon.)

1. Department of Radiology, University of Pennsylvania School of Medicine, Philadelphia, PA, USA
2. HHUU Virgen del Rocvo, Seville, Spain
3. ABX advanced biochemical compounds GmbH, Radeberg, Germany
4. PET Center, Institute of Radiopharmacy, Research Center Dresden-Rossendorf, Dresden, Germany

***

Keywords: Bone marrow metastasis
- ^{18}F-FDG-PET/CT
- Partial volume correction
- Global disease assessment
- Quantitative analysis
- Software analysis

Correspondence address:
Drew A. Torigian, MD, MA
Department of Radiology
Hospital of the University of Pennsylvania 3400 Spruce Street
Philadelphia, PA 19104
Tel: 215-615-3805
Fax: 215-614-0033, Email:
Drew.Torigian@uphs.upenn.edu

Received:
22 December 2010
Accepted:
30 December 2010

Abstract

Our aim was to assess feasibility and performance of novel semi-automated image analysis software called ROVER to quantify metabolically active volume (MAV), maximum standardized uptake value-maximum (SUV_{max}), 3D partial volume corrected mean SUV ($cSUV_{mean}$), and 3D partial volume corrected mean MVP ($cMVP_{mean}$) of spinal bone marrow metastases on fluorine-18 fluorodeoxyglucose-positron emission tomography/computerized tomography (^{18}F-FDG-PET/CT). *We retrospectively studied* 16 subjects with 31 spinal metastases on FDG-PET/CT and MRI. Manual and ROVER determinations of lesional MAV and SUV_{max}, and repeated ROVER measurements of MAV, SUV_{max}, $cSUV_{mean}$ and $cMVP_{mean}$ were made. Bland-Altman and correlation analyses were performed to assess reproducibility and agreement. *Our results showed* that analyses of repeated ROVER measurements revealed MAV mean difference (D)=-0.03±0.53cc (95%CI(-0.22, 0.16)), lower limit of agreement (LLOA)=-1.07cc, and upper limit of agreement (ULOA)=1.01cc; SUV_{max} D=0.00±0.00 with LOAs=0.00; $cSUV_{mean}$ D=-0.01±0.39 (95%CI(-0.15, 0.13)), LLOA=-0.76, and ULOA=0.75; $cMVP_{mean}$ D=-0.52±4.78cc (95%CI(-2.23, 1.23)), LLOA=-9.89cc, and ULOA=8.86cc. Comparisons between ROVER and manual measurements revealed volume D= -0.39±1.37cc (95%CI (-0.89, 0.11)), LLOA=-3.08cc, and ULOA=2.30cc; SUV_{max} D=0.00±0.00 with LOAs=0.00. Mean percent increase in lesional SUV_{mean} and MVP_{mean} following partial volume correction using ROVER was 84.25±36.00% and 84.45±35.94%, respectively. *In conclusion*, it is feasible to estimate MAV, SUV_{max}, $cSUV_{mean}$, and $cMVP_{mean}$ of spinal bone marrow metastases from ^{18}F-FDG-PET/CT quickly and easily with good reproducibility via ROVER software. Partial volume correction is imperative, as uncorrected SUV_{mean} and MVP_{mean} are significantly underestimated, even for large lesions. This novel approach has great potential for practical, accurate, and precise combined structural-functional PET quantification of disease before and after therapeutic intervention.

Hell J Nucl Med 2011; 14(1): 8-14 *Published on line: 5 March 2011*

Introduction

Metastatic disease is the most common malignancy to involve the spinal bone marrow, and 2-deoxy-2-[^{18}F]fluoro-D-glucose (^{18}F-FDG) positron emission tomography/computed tomography (PET/CT) plays an important role in the detection, characterization, and treatment monitoring of such lesions in the setting of malignancy [1]. Typically, analysis of PET image datasets in this clinical setting is performed qualitatively and semi-quantitatively by measuring standardized uptake values (SUV) in a subset of lesions detected. However, such measurements are prone to various sources of error including partial volume effect (PVE) and sampling error due to fractional assessment of total disease burden present, which have consequences related to accurate assessment and management of patients with malignant disease [2-3].

Ideally, one would prefer to rapidly, accurately, and reproducibly measure the true volume and true metabolic activity of viable portions of all malignant lesions in the body. This would allow for accurate assessment of total viable tumor burden at any point in time for staging, forecasting prognosis, pretreatment planning, post-treatment response assessment, and restaging. As such, semi-automated computer-assisted user-interactive tools that are not technically demanding are essential for this purpose. Unfortunately, to our knowledge, there is no currently available software that allows one to simultaneously estimate metabolically active volume (MAV), 3D partial volume corrected standardized uptake value (cSUV), and 3D partial volume corrected metabolic volume product (cMVP) of malignant lesions in the bone marrow or other anatomic sites from the PET portion alone of ^{18}F-FDG-PET/CT.

Therefore, the aim of this study was to assess the feasibility and performance of a novel semi-automated image analysis software package called ROVER (Region of interest (ROI) visualization, evaluation, and image registration) (ABX advanced biochemical compounds

GmbH, Radeberg, Germany) to quantify MAV, SUV_{max}, $cSUV_{mean}$, and $cMVP_{mean}$ of spinal bone marrow metastases from the PET portion of ^{18}F-FDG-PET/CT scans in patients with known malignancy.

Subjects and methods

Study sample

Following Institutional Review Board (IRB) approval for retrospective data collection and image analysis along with Health Insurance Portability and Accountability Act (HIPAA) waiver prior to study initiation, we conducted a retrospective search for subjects with detectable spinal bone marrow metastases who had undergone ^{18}F-FDG-PET/CT and spinal magnetic resonance imaging (MRI) within 45 days of each other without intervening therapy. This study included 16 subjects (7 men and 9 women, mean age: 64.96±14.71 years old (range: 19-80)) with metastases from primary tumors listed in Table 1. We selected bone marrow lesions that showed ^{18}F-FDG uptake throughout their entirety (i.e., no necrosis was present) and that could easily be delineated visually on MRI. A total of 31 lesions were thus identified and evaluated.

Table 1. *Cancer diagnoses in 16 subjects*

Tumor type	Number of subjects
Lung cancer	7
Breast cancer	2
Malignant pleural mesothelioma	1
Anal cancer	1
Melanoma	1
Head and neck cancer	1
Gastric cancer	1
Ovarian and lung cancers	1
Pancreatic, breast, and colon cancers	1

MR image acquisition and analysis

Manual measurements of lesional volumes were performed on sagittal T1-weighted or T2-weighted MR images that had been obtained on 1.5T or 3T magnets (Siemens Healthcare USA Inc., Malvern, PA) with a dedicated spine coil using dedicated image visualization and analysis software (Terarecon Inc., San Mateo, CA). Bone marrow metastases generally had low signal intensity on T1-weighted images and high signal intensity on T2-weighted images relative to normal bone marrow. Freehand contours were manually drawn about the boundaries of lesions on each slice of visualization to estimate lesional areas (Fig. 1A). Lesional areas were then summed and multiplied by slice thickness to obtain lesional volumes.

PET/CT image acquisition

All ^{18}F-FDG-PET/CT scans were acquired using a 16 detector-row LYSO whole-body PET/CT scanner with time-of-flight capabilities (Gemini TF, Philips Healthcare, Bothell, WA). 3D PET data were acquired from the skull base to mid thighs ~60min after intravenous administration of ~555MBq of ^{18}F-FDG for 3min per bed position. Image reconstruction was performed using a list-mode maximum-likelihood expectation-maximization (ML-EM) algorithm with 33 ordered subsets and 3 iterations. The system model included time-of-flight as well as normalization, attenuation, randoms, and scatter corrections. Rescaled low-dose CT images were utilized for attenuation correction of PET images. PET and CT images were reconstructed at 5mm nominal slice thickness.

PET image analysis

Manual measurements

Standard manual SUV measurements were carried out by using dedicated image visualization and analysis software (Extended Brilliance Workstation, Philips Healthcare, Bothell, WA). 3D spherical volumes of interest (i.e., masks) were manually placed slightly outside of (~2-4 pixels beyond) the visual margin of each lesion. Maximum SUV (SUV_{max}) for each lesion was then recorded.

Semi-automated ROVER measurements

ROVER software was used to automatically delineate 3D ROI corresponding to ^{18}F-FDG avid spinal bone marrow metastatic lesions on PET images (following initial user guided gross visual identification of lesions) (Fig. 1B-1C). An initial threshold setting of 40% of the maximum lesional metabolic activity was used along with restriction to ROI with a minimum volume of 1cc. Reconstructed image resolution was estimated at 8mm. Other software settings were left on default. The software then automatically calculated MAV (in cc), SUV_{max}, SUV_{mean}, $cSUV_{mean}$, MVP_{mean} (in cc), and $cMVP_{mean}$ (in cc) for each ROI. (Please see Appendix for details regarding the ROVER software ROI delineation algorithm). All software measurements were made for all lesions by two separate readers on the same ^{18}F-FDG-PET/CT datasets with a time interval between measurements of approximately 2 months for purposes of inter-reader reproducibility assessment.

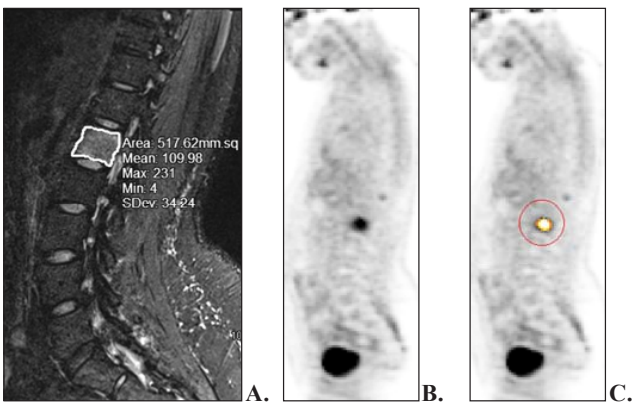

Figure 1. Image analysis of spinal bone marrow metastases on MR and ^{18}F-FDG-PET images. **A**, Sagittal fat-suppressed T2-weighted image demonstrates freehand manual outer boundary delineation of high signal intensity L1 lumbar spinal bone marrow metastasis. B, Sagittal PET image shows avid ^{18}F-FDG uptake in same lesion in A. **C**, Same sagittal PET image following placement of mask (red circle) about lesion and automatic delineation of 3D ROI (yellow) corresponding to metastasis using ROVER software.

Statistical analysis

Tabulations of means, standard deviations, and ranges of lesional volumes, as well as medians and interquartile ranges (IQR) of SUV and MVP were performed. Correlation analyses between software and manual measurements of lesional

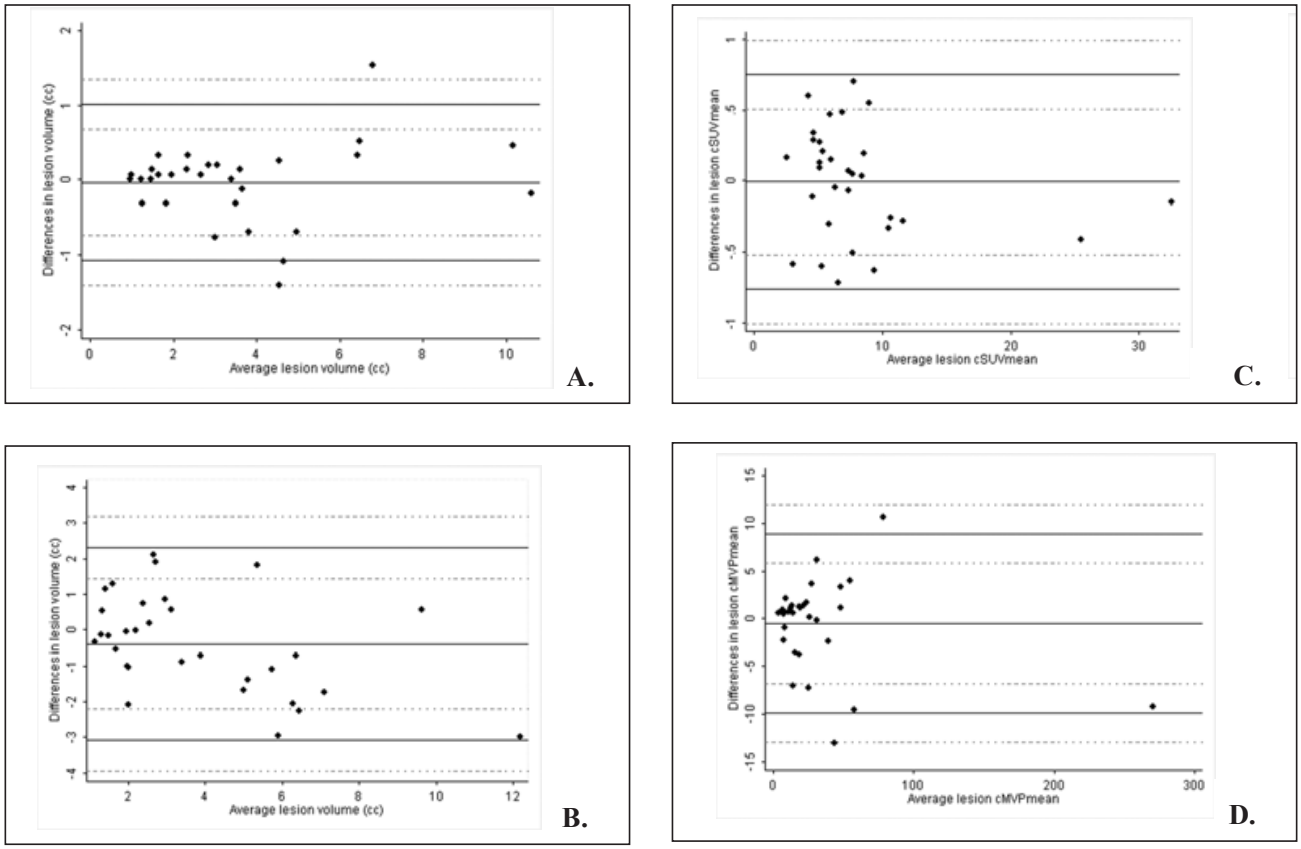

Figure 2. Bland-Altman scatterplots of reproducibility of ROVER measurements and of agreement between ROVER and manual measurements. Mean differences, LLOA, and ULOA are depicted as solid lines, whereas 95%CI of LOA are shown as dotted lines. **A**, Reproducibility of ROVER volume measurements. , Agreement between ROVER and manual MRI volume measurements. **C**, Reproducibility of ROVER $cSUV_{mean}$ measurements. **D**, Reproducibility of ROVER $cMVP_{mean}$ measurements.

volume, SUV, and MVP were performed. P values < 0.05 were considered as statistically significant.

In addition, inter-reader reproducibility of software measurements and agreement between software and manual measurements were assessed using the Bland-Altman graphical approach [4-5]. Briefly, this approach involves creation of scatterplots of the difference between two measurements plotted against the average of the two measurements. Upper and lower limits of agreement (LOA) were defined as the symmetric range encompassing 95% of the data, where upper limit of agreement (ULOA)=mean difference (D)+1.96x[standard deviation of the differences (SD)] and lower limit of agreement (LLOA)=D–1.96xSD. Confidence intervals of 95% (95%CI) of D were calculated as $D\pm t_{n-1,0.025}xsqrt((SD)^2/n)$, where n=number of lesions assessed. 95%CI of ULOA were calculated as $ULOA\pm t_{n-1,0.025}xsqrt(3x(SD)^2/n)$ and 95%CI of LLOA were calculated as $LLOA\pm t_{n-1,0.025}xsqrt(3x(SD)^2/n)$. Please note that this approach was utilized only after confirmation of approximate normality of distribution of difference measurements via visual assessment of histograms, and after visual assessment of Bland-Altman scatterplots for uniform bias and variability of difference measurements throughout the range of measurement.

Lastly, assessment of percent increase in SUV_{mean} and MVP_{mean} of lesions following partial volume correction through ROVER was performed. All statistical analyses were performed using Stata software (Stata/IC Version 10.1, StataCorp, College Station, TX).

Results

Assessments of lesion volume

Mean volume of lesions determined by ROVER analysis was 3.63±2.40cc (range: 0.96-10.69cc). Mean volume of lesions determined by manual MRI analysis was 4.02±3.04cc (range: 0.82-13.69cc). The Pearson correlation coefficient between repeated sets of ROVER volumetric measurement was 0.98 (P<0.0001). The Pearson correlation coefficient between ROVER and manual volumetric measurements was 0.90 (P<0.0001).

Bland-Altman analysis of repeated ROVER volume measurements revealed D of -0.03±0.53cc (95%CI (-0.22, 0.16)), LLOA of -1.07cc (95%CI (-1.41, -0.73)), and ULOA of 1.01cc (95%CI (0.67, 1.34)) (Fig. 2A). Bland-Altman analysis of agreement between ROVER and manual volume measurements revealed D (ROVER minus MRI) of -0.39±1.37cc (95%CI (-0.89, 0.11)), LLOA of -3.08cc (95%CI (-3.95, -2.21)), and ULOA of 2.30cc (95%CI (1.43, 3.17)) (Fig. 2B).

Assessments of lesion $cSUV_{mean}$ and $cMVP_{mean}$

Median $cSUV_{mean}$ of lesions determined by ROVER PET image analysis was 6.63 (IQR: 5.02-8.44). The Spearman correlation coefficient between repeated sets of ROVER $cSUV_{mean}$ measurement was 0.98 (P<0.0001). Bland-Altman analysis of repeated ROVER $cSUV_{mean}$ measurements revealed D of -0.01±0.39 (95%CI (-0.15, 0.13)), LLOA of -0.76 (95%CI (-1.01, -0.52)), and ULOA of 0.75 (95%CI (0.50, 0.99)) (Fig. 2C).

Median $cMVP_{mean}$ of lesions determined by ROVER PET image analysis was 20.4cc (IQR: 10.55-40.03cc). The Spearman correlation coefficient between repeated sets of ROVER $cMVP_{mean}$ measurement was 0.97 (P<0.0001). Bland-Altman analysis of repeated ROVER $cMVP_{mean}$ measurements revealed D of -0.52±4.78cc (95%CI (-2.23, 1.23)), LLOA of -9.89cc (95% CI (-12.92, -6.86)), and ULOA of 8.86cc (95%CI (5.82, 11.89)) (Fig. 2D).

Assessments of lesion SUV_{max}

Median SUV_{max} of lesions determined by ROVER PET image analysis was 4.72 (IQR: 3.93-8.03). Median SUV_{max} of lesions determined by manual PET image analysis was 4.72 (IQR: 3.93-8.03). The Spearman correlation coefficient between repeated sets of ROVER SUV_{max} measurement was 1.00 (P<0.0001). The Spearman correlation coefficient between ROVER and manual SUV_{max} measurements was 1.00 (P<0.0001).

Bland-Altman analysis of repeated sets of ROVER lesional SUV_{max} measurement revealed a mean difference of 0.00±0.00 with LOA of 0.00. Bland-Altman analysis of agreement between ROVER and manual SUV_{max} measurements revealed a mean difference of 0.00±0.00 with LOA of 0.00.

Assessment of the effects of partial volume correction on SUV_{mean} and MVP_{mean}

Mean percent increase in SUV_{mean} and MVP_{mean} of lesions following partial volume correction by ROVER software was 84.25±36.00% (range: 7.86-164.53%) and 84.45±35.94% (range: 8.07-165.08%), respectively.

Discussion

Imaging via ^{18}F-FDG-PET/CT has revolutionized the diagnostic and therapeutic evaluation of patients with cancer, as it provides for noninvasive quantitative assessment of malignant disease in the entire body with high sensitivity and specificity [6-7]. In particular, ^{18}F-FDG-PET often detects metastases to bone marrow that are invisible on CT due to insufficient destruction or sclerosis of the bone, and allows one to determine early tumor response assessment during or following therapy [1, 7-9]. Although MRI has high sensitivity for detection of bone marrow metastases, it is nonspecific for separating viable from non-viable tumor, and quantitative assessment is typically achieved by manual 1D or 2D measurement of a subset of lesions.

The most commonly used approach for semi-quantitative analysis of ^{18}F-FDG-PET is measurement of lesional SUV. Typically, SUV_{max}, the SUV of an ROI of 1 pixel with the maximum pixel value in a lesion, and SUV_{mean}, the average SUV of all pixels in an ROI, are measured and reported using commercially available software platforms [10]. However, these parameters are susceptible to various sources of error.

First, in general, for heterogeneous ROI SUV_{max} inherently involves intralesional sampling error (given its definition), and for homogeneous ROI SUV_{max} typically overestimates the true mean value of a lesion (usually by ~2-3 times the standard deviation of the noise level as long as recovery effects are small). In contrast, SUV_{mean} can be measured based on pixel data representative of the entire lesion, leading to more accurate statistical estimation of the true mean value. However, SUV_{mean} measurement in practice is much more variable due to operator-dependent factors including size and shape of mask delineation and location of mask placement within or about a lesion, as well as presence of non-uniformity of lesional and background ^{18}F-FDG activity [2, 11]. Second, SUV is often measured by manual placement of a 2D mask at a particular level through a lesion, also leading to intralesional sampling error depending on the level selected. Moreover, SUV is generally measured only for a subset of lesions present, leading to sampling error based on which particular lesions are selected. Third, SUV_{max} and SUV_{mean} are susceptible to PVE due to image blurring and image sampling, generally leading to underestimation of lesional metabolic activity and overestimation of lesional size, especially for small lesions with size ≤2-3 times full-width-at-half-maximum (FWHM) of the reconstructed image resolution [2, 11-13]. These errors can accumulate and result in significant effect on the accuracy and reproducibility of regional and global disease assessment and quantification. This is particularly true for measuring tumor response where, for example, any lesion that decreases in size may appear to have a reduction in ^{18}F-FDG uptake greater than that which has actually occurred. Thus, there is a need for an accurate, reproducible, and practical standardized approach to minimize these sources of error for routine use in patients with cancer and other disorders.

Furthermore, currently used methods to quantify disease activity on ^{18}F-FDG-PET/CT do not routinely assess measures of global disease burden such as the MVP (also called total lesional glycolysis (TLG), total glycolytic volume (TGV), and whole-body metabolic burden (WBMB)), which combines volumetric and metabolic measurements into a summary measure of global disease burden. Global disease assessment will likely have great importance for improved pretreatment planning, patient selection for clinical trials, prognostication of patient outcome, prediction of treatment response, and response assessment. Several reports in the literature (as listed below) reveal the potential utility of global disease assessment on ^{18}F-FDG-PET or ^{18}F-FDG-PET/CT.

Alavi et al, in 1993, calculated total brain metabolism as the product of brain volume (measured through computer-assisted segmentation of cerebrospinal fluid and brain on MRI) and average metabolic rate (measured manually within a 50% isocontour on ^{18}F-FDG-PET), corrected for atrophy, in patients with Alzheimer's disease and in age-matched controls. They reported that partial volume corrected metabolic rates per unit weight of brain were not significantly different in these cohorts, but that total brain metabolism was significantly lower in Alzheimer's patients. This study was the first in the literature to introduce and apply the concepts of partial volume correction along with global metabolic assessment to clinical ^{18}F-FDG-PET disease quantification, while simultaneously highlighting their importance [14-15]. Subsequently, Larson et al, in 1999, proposed the use of TLG, defined as the product of lesional volume (determined by an automatic adaptive image thresholding technique [16] on ^{18}F-FDG-PET) and lesional SUV_{mean} (measured manually in 2D on ^{18}F-FDG-PET), and a relative response index defined by the percentage change in TLG following therapy, to perform response assessment in 41 patients with malignant disease. They observed that changes in TLG following treatment were generally greater than changes in SUV, and agreed more closely with assessments using a visual response score [17]. This was the first report in the literature to demonstrate

the application of global metabolic assessment for ^{18}F-FDG-PET quantification in the clinical oncologic setting. Similarly, Akhurst et al, in 2000, reported use of TLG in 6 patients with metastatic renal cell carcinoma, showed high correlation between CT and ^{18}F-FDG-PET volume estimates for lesions larger than 5cc, and demonstrated that TLG correlated better with CT-derived tumor volume than did ^{18}F-FDG-PET volume alone [18]. Guillem et al, in 2004, measured absolute change in TLG (with TLG defined as the product of MAV and SUV_{max} on ^{18}F-FDG-PET) in 15 patients with locally advanced primary rectal cancer following preoperative chemoradiation therapy, and correlated this with various measures of clinical outcome. They showed that patients with absolute change in TLG of $\geq$69.5 had significantly improved disease-specific and recurrence-free survival, and therefore may predict long-term outcomes [19].

Bural et al, in 2006, reported the use of aortic atheroburden (AB) in 18 patients of variable age to quantify the extent of atherosclerosis. Aortic atheroburden was calculated as the product of aortic wall volume (measured by manual 2D ROI delineation on contrast-enhanced CT) and aortic wall SUV (measured manually in 2D on ^{18}F-FDG-PET). In each aortic wall segment, AB, SUV, and wall volume statistically significantly increased with age, and total aortic AB appeared to increase with age in a nonlinear fashion [20]. This report was the first in the literature to apply the concept of global metabolic assessment to clinical ^{18}F-FDG-PET quantification of cardiovascular disease. Francis et al, in 2007, measured tumor TGV on ^{18}F-FDG-PET, defined as the product of MAV and SUV_{mean}, in 22 patients with malignant pleural mesothelioma before and after chemotherapy via a semi-automated 3D region-growing iterative algorithm with adaptive thresholding. Changes in TGV, SUV_{max}, and CT-modified response evaluation criteria in solid tumors (RECIST) were then correlated with clinical outcome measures. Percentage change in TGV after chemotherapy was strongly predictive of survival, whereas percentage change in SUV_{max} and CT-modified RECIST were not. Furthermore, there was a significant relationship between the magnitude of TGV percentage change and survival [21]. Benz et al, in 2008, measured changes in lesional CT-derived volume, SUV_{mean}, SUV_{max}, and TLG (where TLG was defined as the product of manually measured CT-derived volume and SUV measured on ^{18}F-FDG-PET/CT) in 20 patients with locally advanced high grade soft tissue sarcomas following neoadjuvant therapy, correlated them with presence or absence of histopathological tumor response, and performed receiver operator characteristic (ROC) curve analysis to determine the accuracy of changes in these parameters to predict histopathological response. TLG measurements were less accurate in predicting tumor response to therapy than SUV measurements, likely due to use of CT-derived volume rather than PET-derived MAV for calculation of TLG, given that CT-derived volumes did not discriminate between responders and non-responders [22]. Lastly, Berkowitz et al, in 2008, measured the WBMB of 19 patients with non-Hodgkin's lymphoma from ^{18}F-FDG-PET and CT before and after anti-CD20 radioimmunotherapy by taking the product of lesional volume (determined by manual 2D ROI delineation on CT or by 40% SUV_{max} isocontour on PET) and $cSUV_{mean}$ (calculated by manual 2D SUV_{mean} measurements and a recovery coefficient lookup curve [23]) to generate lesional metabolic burden (MB), and summing over all lesions. Of 10 patients with partial response, MB showed the most consistent and largest percentage change compared to SUV_{max} and SUV_{mean}. In addition, preliminary analysis suggested that high WBMB is correlated with high lactate dehydrogenase, high lactate, and lower cholesterol levels, and seemed to confer a short survival [24].

In the current study, we have shown the feasibility of novel semi-automated software called ROVER to automatically delineate 3D ROI corresponding to ^{18}F-FDG avid spinal bone marrow metastatic lesions on the PET portion of ^{18}F-FDG-PET/CT scans, and to calculate lesional MAV with high accuracy and reproducibility even when invisible on CT. This will therefore decrease errors in SUV measurement related to variations in manual mask delineation and intralesional sampling using traditional 2D approaches (i.e., interobserver variability of measurements will be minimized). Furthermore, we have shown the feasibility to calculate $cSUV_{mean}$ and $cMVP_{mean}$ of metastatic lesions with good reproducibility, leading to a significant decrease in errors related to PVE. To our knowledge, use of semi-automated image analysis software to calculate $cSUV_{mean}$ and $cMVP_{mean}$ of malignant lesions in 3D based on the PET portion of ^{18}F-FDG-PET/CT imaging has not previously been reported in the literature. In addition, this approach to partial volume correction does not make assumptions about lesional shape, size, volume, or radiotracer uniformity, and does not require registration with or segmentation of structural imaging datasets, in contrast to many other available methods of partial volume correction [2, 23, 25-28]. Interestingly, we observed that partial volume correction affected SUV_{mean} even for large lesions, likely as a consequence of image blurring from motion during imaging (related to physiological, respiratory, or patient movements) since motion essentially increases the width of the local point spread function thus decreasing the spatial resolution, further emphasizing the utility of partial volume correction even for the quantification of large lesions. Overall, in our experience, the software was fast and easy to use, which are important features if in the future such image analysis approaches are to be utilized routinely in clinical practice or for research application.

This study has several limitations. First, it was performed retrospectively utilizing a small cross-sectional sample of subjects with various primary malignancies, potentially leading to selection and other unknown biases. Despite this, the preliminary results of the current study appear promising. Future studies are necessary to assess software performance according to specific anatomic site (other than spinal bone marrow) and specific tumor histology. Second, metastatic lesions selected for analysis were well visualized on ^{18}F-FDG-PET images. Software performance will likely decrease to some degree when applied to lesions with lower target-to-background contrast ratios. Further research will be necessary to study this in more detail, and may require input of structural imaging cues for accurate segmentation of such lesions. Third, only non-necrotic lesions were analyzed in this study so that we could directly compare MRI based volumes to PET based volumes, as it is often difficult or impossible to accurately measure the volumes of non-necrotic portions of tumor on MRI via manual measurement. Fourth, partial volume correction will in general have less accuracy for very small lesions compared to larger lesions. However, this will have minimal impact upon quantification of total disease burden unless all lesions present are very small. Fifth, histopathological measurement of the actual volume of bone marrow metastases

was not used as the reference standard for assessment of lesion volume as this would be impracticable to perform and would be prone to errors related to tissue specimen processing and gross visual assessment of lesion margins. However, MRI has already been extensively validated in the literature as an accurate and reproducible means to assess structural volumes of lesions and normal tissues in vivo, and is often used clinically for this purpose. Therefore, we feel that use of MRI as a reference standard of lesion volume is adequate in this preliminary study. This is corroborated by the high correlation between PET and MRI based volume assessments.

In conclusion, it is feasible to estimate MAV, SUV_{max}, $cSUV_{mean}$, and $cMVP_{mean}$ of ^{18}F-FDG avid spinal bone marrow metastases from the PET portion of ^{18}F-FDG-PET/CT scans through use of a novel semi-automated image analysis software package called ROVER. The reproducibility of measurements provided by this software is good, and there is a high level of agreement between ROVER and manual lesional volumetric measurements indicating reasonable accuracy as well. This computer-assisted approach has great potential to enable practical, accurate, and precise combined structural-functional PET quantification of disease.

The authors declare that they have no conflicts of interest.

Bibliography

1. Basu S, Torigian D, Alavi A. Evolving concept of imaging bone marrow metastasis in the twenty-first century: critical role of FDG-PET. *Eur J Nucl Med Mol Imaging* 2008; 35: 465-71.
2. Soret M, Bacharach SL, Buvat I. Partial-volume effect in PET tumor imaging. *J Nucl Med* 2007; 48: 932-45.
3. Basu S, Alavi A. Feasibility of automated partial-volume correction of SUVs in current PET/CT scanners: can manufacturers provide integrated, ready-to-use software? *J Nucl Med* 2008; 49: 1031-2; authors reply 2-3.
4. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; 1: 307-10.
5. Bland JM, Altman DG. Measuring agreement in method comparison studies. *Stat Methods Med Res* 1999; 8: 135-60.
6. Alavi A, Lakhani P, Mavi A et al. PET: a revolution in medical imaging. *Radiol Clin North Am* 2004; 42: 983-1001, vii.
7. Torigian DA, Huang SS, Houseni M, Alavi A. Functional imaging of cancer with emphasis on molecular techniques. *CA Cancer J Clin* 2007; 57: 206-24.
8. Song JW, Oh YM, Shim TS et al. Efficacy comparison between ^{18}F-FDG PET/CT and bone scintigraphy in detecting bony metastases of non-small-cell lung cancer. *Lung Cancer* 2009; 65: 333-8.
9. Basu S, Alavi A. Bone marrow and not bone is the primary site for skeletal metastasis: critical role of [^{18}F]fluorodeoxyglucose positron emission tomography in this setting. *J Clin Oncol* 2007; 25: 1297; authors reply -9.
10. Huang SC. Anatomy of SUV. Standardized uptake value. *Nucl Med Biol* 2000; 27: 643-6.
11. Basu S, Zaidi H, Houseni M et al. Novel quantitative techniques for assessing regional and global function and structure based on modern imaging modalities: implications for normal variation, aging and diseased states. *Semin Nucl Med* 2007; 37: 223-39.
12. Hoffman EJ, Huang SC, Phelps ME. Quantitation in positron emission computed tomography: Effect of object size. *J Comput Assist Tomogr* 1979; 3: 299-308.
13. Hickeson M, Yun M, Matthies A et al. Use of a corrected standardized uptake value based on the lesion size on CT permits accurate characterization of lung nodules on FDG-PET. *Eur J Nucl Med Mol Imaging* 2002; 29: 1639-47.
14. Alavi A, Newberg AB, Souder E, Berlin JA. Quantitative analysis of PET and MRI data in normal aging and Alzheimer's disease: atrophy weighted total brain metabolism and absolute whole brain metabolism as reliable discriminators. *J Nucl Med* 1993; 34: 1681-7.
15. Chawluk JB, Alavi A, Dann R et al. Positron emission tomography in aging and dementia: effect of cerebral atrophy. *J Nucl Med* 1987; 28: 431-7.
16. Erdi YE, Mawlawi O, Larson SM et al. Segmentation of lung lesion volume by adaptive positron emission tomography image thresholding. *Cancer* 1997; 80: 2505-9.
17. Larson SM, Erdi Y, Akhurst T et al. Tumor Treatment Response Based on Visual and Quantitative Changes in Global Tumor Glycolysis Using PET-FDG Imaging. The Visual Response Score and the Change in Total Lesion Glycolysis. *Clin Positron Imaging* 1999; 2: 159-71.
18. Akhurst T, Ng VV, Larson SM et al. Tumor Burden Assessment with Positron Emission Tomography with. *Clin Positron Imaging* 2000; 3: 57-65.
19. Guillem JG, Moore HG, Akhurst T et al. Sequential preoperative fluorodeoxyglucose-positron emission tomography assessment of response to preoperative chemoradiation: a means for determining longterm outcomes of rectal cancer. *J Am Coll Surg* 2004; 199: 1-7.
20. Bural GG, Torigian DA, Chamroonrat W et al. Quantitative assessment of the atherosclerotic burden of the aorta by combined FDG-PET and CT image analysis: a new concept. *Nucl Med Biol* 2006; 33: 1037-43.
21. Francis RJ, Byrne MJ, van der Schaaf AA et al. Early prediction of response to chemotherapy and survival in malignant pleural mesothelioma using a novel semiautomated 3-dimensional volume-based analysis of serial ^{18}F-FDG PET scans. *J Nucl Med* 2007; 48: 1449-58.
22. Benz MR, Allen-Auerbach MS, Eilber FC et al. Combined assessment of metabolic and volumetric changes for assessment of tumor response in patients with soft-tissue sarcomas. *J Nucl Med* 2008; 49: 1579-84.
23. Srinivas SM, Dhurairaj T, Basu S et al. A recovery coefficient method for partial volume correction of PET images. *Ann Nucl Med* 2009; 23: 341-8.
24. Berkowitz A, Basu S, Srinivas S et al. Determination of whole-body metabolic burden as a quantitative measure of disease activity in lymphoma: a novel approach with fluorodeoxyglucose-PET. *Nucl Med Commun* 2008; 29: 521-6.
25. Boussion N, Hatt M, Lamare F et al. A multiresolution image based approach for correction of partial volume effects in emission tomography. *Phys Med Biol* 2006; 51: 1857-76.
26. Quarantelli M, Berkouk K, Prinster A et al. Integrated software for the analysis of brain PET/SPECT studies with partial-volume-effect correction. *J Nucl Med* 2004; 45: 192-201.
27. Rousset OG, Ma Y, Evans AC. Correction for partial volume effects in PET: principle and validation. *J Nucl Med* 1998; 39: 904-11.
28. Muller-Gartner HW, Links JM, Prince JL et al. Measurement of radiotracer concentration in brain gray matter using positron emission tomography: MRI-based correction for partial volume effects. *J Cereb Blood Flow Metab* 1992; 12: 571-83.
29. Poetzsch C, Hofheinz F, van den Hoff J. Fast user guided segmentation and quantification of volumes in 3D datasets. *Mol Imaging Biol* 2005; 7: 152.

30. Hofheinz F, Poetzsch C, van den Hoff J. Quantitative 3D ROI delineation in PET: algorithm and validation. *J Nucl Med* 2007; 48: 407P.
31. Hofheinz F, Dittrich S, Potzsch C, Hoff J. Effects of cold sphere walls in PET phantom measurements on the volume reproducing threshold. *Phys Med Biol* 2010; 55: 1099-113.

Appendix

The ROI delineation algorithm of ROVER accepts the selection of a subvolume of the PET dataset (i.e., a mask) by the user that is typically placed about a group of one or more visually identified ^{18}F-FDG avid lesions) as well as the selection of an initial threshold value by the user to define 3D ROI corresponding to those lesions for evaluation by the software. Any potential ROI which is contained within the mask and which has at least one voxel larger than the initial threshold value is considered. The algorithm itself is an iterative one where in each step the local background for each ROI inside the mask is determined. Subsequently, a background corrected threshold is applied to delineate the ROI. The iteration stops when all ROI volumes have converged [29-30]. The local background value ROVER uses is the average of all background voxels, where a background voxel is defined by: 1. having a location inside the mask, 2. not belonging to any ROI, 3. having a minimum distance of 1 x reconstruction resolution to the ROI, 4. having a maximum distance of 2.5 x reconstruction resolution to the ROI.

It should be emphasized that with this definition, only voxels in the immediate vicinity of an ROI contribute to the ROI's background, and inhomogeneities (if any) are very small. The background is thus locally approximately constant, and the effective background is therefore computed as the average of the background voxels. The actually applied absolute threshold is computed as:

$Threshold = (C_{75\%} - Bg) \times 0.43 + Bg,$

where Bg is the background activity concentration and $C_{75\%}$ is the average activity concentration within a 75% isocontour of the ROI's maximum. T is the background corrected threshold, which was determined from phantom measurements as T=0.43. This equation is inspired by the results presented by Hofheinz et al [31]. There it was shown that the volume reproducing background corrected threshold depends only on the relative object size and, to some extent, on the object geometry. For objects larger than the reconstructed image resolution, the background corrected threshold can be considered to be essentially constant at the value given above.

Figure 3 illustrates the final situation of the iterative process. The gray vertically oriented rectangle represents the intensity profile of the true ROI whereas the shape bounded by the curvilinear thick solid line represents the measured PET signal intensity. V_R is the ROI volume and V_G is the considered mask volume including the background. The total activity (A_{total}) inside the ROI including that from partial volume effects is given by:

$A_{total} = [V_G \cdot C(v)] - (V_G - V_R) \times Bg,$

where the summation is over all $v \in V_G$ and C(v) is the activity concentration at voxel v for all voxels inside the volume V_G. The partial volume corrected mean activity concentration of the ROI is then given by:

$C_{mean} = A_{total} / V_R$

where C_{mean} is in units of Bq/cc, and the partial volume corrected mean SUV is given by:

$$cSUV_{mean} = \frac{C_{mean}}{IA} \times BW$$

where IA is equal to injected activity in Bq, and BW represents patient body weight in g.

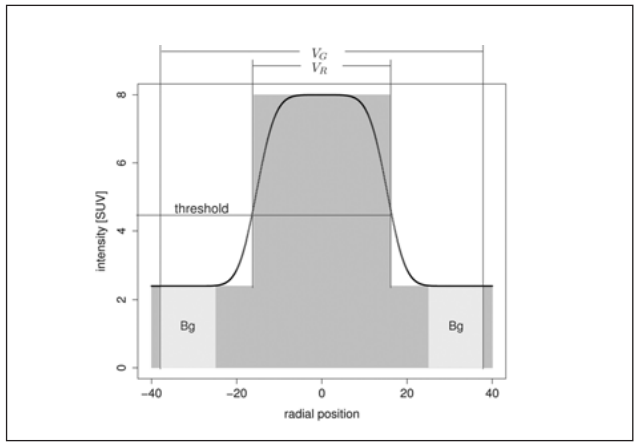

Figure 3. Illustration of ROI delineation algorithm of ROVER.