

# Measuring response to therapy using FDG PET: semi-quantitative and full kinetic analysis

Patsuree Cheebsumon · Linda M. Velasquez · Corneline J. Hoekstra · Wendy Hayes · Reina W. Kloet · Nikie J. Hoetjes · Egbert F. Smit · Otto S. Hoekstra · Adriaan A. Lammertsma · Ronald Boellaard

Received: 27 September 2010 / Accepted: 2 December 2010 / Published online: 6 January 2011  
© The Author(s) 2010. This article is published with open access at Springerlink.com

## Abstract

**Purpose** Imaging with positron emission tomography (PET) using  $^{18}\text{F}$ -2-fluoro-2-deoxy-D-glucose (FDG) plays an increasingly important role for response assessment in oncology. Several methods for quantifying FDG PET results exist. The goal of this study was to analyse and compare various semi-quantitative measures for response assessment with full kinetic analysis, specifically in assessment of novel therapies.

**Methods** Baseline and response dynamic FDG studies from two different longitudinal studies (study A: seven subjects with lung cancer and study B: six subjects with gastrointestinal cancer) with targeted therapies were reviewed. Quantification of tumour uptake included full kinetic methods, i.e. nonlinear regression (NLR) and Patlak analyses, and simplified measures such as the simplified kinetic method (SKM) and standardized uptake value

(SUV). An image-derived input function was used for NLR and Patlak analysis.

**Results** There were 18 and 9 lesions defined for two response monitoring studies (A and B). In all cases there was excellent correlation between Patlak- and NLR-derived response ( $R^2 > 0.96$ ). Percentage changes seen with SUV were significantly different from those seen with Patlak for both studies ( $p < 0.05$ ). After correcting SUV for plasma glucose, SUV and Patlak responses became similar for study A, but large differences remained for study B. Further analysis revealed that differences in responses amongst methods in study B were primarily due to changes in the arterial input functions.

**Conclusion** Use of simplified methods for assessment of drug efficacy or treatment response may provide different results than those seen with full kinetic analysis.

**Keywords** FDG PET · SUV · Patlak · NLR · Response monitoring

P. Cheebsumon · R. W. Kloet · N. J. Hoetjes · O. S. Hoekstra · A. A. Lammertsma · R. Boellaard (✉)  
Department of Nuclear Medicine and PET Research, VU University Medical Center,  
P.O. Box 7057, 1007 MB Amsterdam, The Netherlands  
e-mail: r.boellaard@vumc.nl

L. M. Velasquez · W. Hayes  
Bristol-Myers Squibb Co.,  
Princeton, NJ, USA

C. J. Hoekstra  
Department of Nuclear Medicine, Jeroen Bosch Hospital,  
's-Hertogenbosch, The Netherlands

E. F. Smit  
Department of Respiratory Medicine,  
VU University Medical Center,  
Amsterdam, The Netherlands

## Introduction

Positron emission tomography (PET) using  $^{18}\text{F}$ -2-fluoro-2-deoxy-D-glucose (FDG) is important for detection, staging, prognosis and assessment of treatment response in oncology [1–10]. Response monitoring using FDG PET has been implemented in a variety of tumours [9, 11]. Although criteria have been defined for evaluating PET studies using visual inspection [12, 13], quantification of FDG uptake is helpful for objective assessment of response to therapy.

The most common method for quantifying tumour FDG uptake is the use of standardized uptake values (SUV). Limitations of SUV include its dependency on patient preparation, scanning procedure, image reconstruction and

image analysis procedures [9, 14]. These limitations, however, can to a large extent be overcome by proper standardized procedures [9, 14, 15].

Apart from SUV, various other more quantitative measures can be used for analysis of FDG PET studies. Young et al. [16] first provided guidelines for the measurement of tumour response to therapy using FDG PET imaging. Quantitative approaches range from visual assessment, through semi-quantitative methods (e.g. SUV), to full kinetic analysis (Patlak). However, the European Organization for Research and Treatment of Cancer (EORTC) report did not address the accuracy or limitations of the various analytical methods for monitoring response. Hoekstra et al. [11, 17] compared various methods for quantifying FDG uptake in non-small cell lung cancer (NSCLC) such as nonlinear regression (NLR) of the data using a pharmacokinetic model, Patlak analysis and various simplified methods, e.g. the simplified kinetic method (SKM) [18], SUV normalized by body surface area and corrected for glucose (SUV<sub>BSA</sub>) and Sadato<sub>BSA</sub> [19]. Most measures correlated well with NLR ( $R^2 > 0.95$ ). Lammertsma et al. [20] discussed the relationship between these simplified methods and full kinetic analysis for monitoring response to classic cytotoxic drugs. Both groups of authors emphasized that care should be taken in using simplified methods (i.e. SUV-based methods) for evaluation of new drugs.

Potential discrepancies between SUV and full kinetic analysis results may be caused by changes in plasma glucose levels or differences in FDG plasma clearance among scans. Factors potentially affecting the relation between SUV and full kinetic analysis have been described by Freedman et al. [21, 22]. Moreover, lower SUV in patients with high plasma glucose levels has been reported [23].

Full kinetic approaches are considered most quantitative as they use the entire (measured) arterial input and tumour time-activity curves (TAC) in combination with a tracer kinetic model and plasma glucose levels to derive a more quantitative measure of the metabolic rate of glucose.

In order to correctly interpret SUV results, it has been recommended [15, 16] that a formal comparison with full kinetic approaches be included during early drug development trials. Although good correlation of the percentage change between SUV and Patlak over all tumour pairs has been reported [11, 21], there were differences when monitoring individual subject data [21, 24, 25]. These authors concluded that the two measures were not equivalent and that they may not equally be affected by changes occurring during treatment. Only after SUV responses have been validated against those seen using full quantitative analysis could SUV be used in a larger clinical setting [14, 20, 26]. The aim of the present study was to compare various semi-quantitative measures with full kinetic analy-

sis for assessing response to treatment. Although similar comparisons have been performed previously, these mainly involved classic cytotoxic drugs. In contrast, in the present study, simplified methods were evaluated for use with novel drugs.

## Materials and methods

### Subjects and study design

Data from two separate response monitoring studies, involving different drugs, were used, which will be referred to as studies A and B. These studies were selected based on targeted treatment. Written informed consent was obtained for all subjects and studies were approved by the Medical Ethics Review Committee of the VU University Medical Center.

### Study A

Seven subjects suffering from stage IIIB and IV NSCLC, enrolled in a response monitoring trial, were included in this study. Patients were treated with an orally active inhibitor of the epidermal growth factor receptor (EGFR) tyrosine kinase (erlotinib). Erlotinib (Tarceva, drug A) inhibits EGF-dependent proliferation of cells at submicromolar concentrations and blocks cell cycle progression in the G<sub>1</sub> phase, and is taken orally on a daily basis. All patients (four women and three men; mean body weight at baseline 74±16 kg, range 53–109 kg; mean body weight at response scan 79±16 kg, range 60–108 kg) underwent a dynamic FDG PET scan at baseline and after 3 weeks of therapy.

### Study B

Six patients (one woman and five men) with advanced gastrointestinal malignancies were treated by BMS-582664 (brivanib alaninate) in combination with full-dose cetuximab (Erbiximab), a monoclonal antibody targeting EGFR. BMS-582664 (drug B) is a selective dual inhibitor of fibroblast growth factor and vascular endothelial growth factor signalling, and is taken orally on a daily schedule. All patients (mean body weight 84±17 kg, range 60–110 kg) underwent a dynamic baseline FDG PET scan within 2 weeks before start of treatment. Two dynamic response FDG PET scans were performed. Five patients (mean body weight 84±19 kg, range 63–110 kg) underwent an early response study during cycle 1 (day 15) of treatment and three patients (mean body weight 88±13 kg, range 73–108 kg) underwent a late response study after cycle 2 (day 56) of treatment. Thus, there were two subjects that underwent both an early and a late response study.

## PET scanning protocol

All patients fasted for at least 6 h before scanning. Patients were prepared in accordance with recently published guidelines for quantitative FDG PET studies [14, 27]. Patients were scanned in the supine position. Patients received an intravenous catheter for administration of FDG and for collection of additional blood samples to measure blood glucose (i.e. at 35, 45, 55 min post-injection during dynamic scanning) using validated and calibrated methods. Blood glucose levels were used for quantitative analysis, but because of the strict study design, patients could not be rescheduled in cases of high glucose levels. In addition, whole blood and plasma FDG concentrations were measured using a cross-calibrated well counter.

All studies were performed using an ECAT EXACT HR+ scanner (Siemens/CTI, Knoxville, TN, USA). The characteristics of this scanner have been described previously [28]. Each PET study started with a 10-min transmission scan. After completion of the transmission scan, a bolus of about 370 MBq FDG was injected intravenously and a dynamic emission scan was started simultaneously in a two-dimensional acquisition mode. This dynamic emission scan consisted of 40 frames with the following lengths:  $1 \times 30$ ,  $6 \times 5$ ,  $6 \times 10$ ,  $3 \times 20$ ,  $5 \times 30$ ,  $5 \times 60$ ,  $8 \times 150$ ,  $6 \times 300$  s.

All emission scans were reconstructed using filtered backprojection (FBP) with a 0.5 Hanning filter. During reconstruction all corrections needed for quantification, such as dead time, decay, scatter, random, attenuation and normalization corrections, were applied. In addition, the last three frames of the sinogram were summed and reconstructed using normalization and attenuation-weighted ordered subsets expectation maximization (OSEM) with 2 iterations and 16 subsets, followed by post-smoothing using a 0.5 Hanning filter [29]. Consequently, OSEM reconstructed images had the same dimensions and spatial resolution as FBP reconstructed images. OSEM reconstructions were used for volumes of interest (VOI) definition, because of its superior image quality compared with FBP reconstructions.

## VOI definition

To obtain image-derived input functions (IDIF), 3-D VOI were drawn manually in three vascular structures (i.e. left ventricle, aortic arch and ascending aorta) using an early frame highlighting the blood pool [30]. These VOI were then projected onto all frames yielding arterial whole blood TAC, i.e. image-derived input functions. A quality check of each IDIF was performed, as described previously [11, 17]. This involved comparison of the three IDIFs with the activity concentration measured in the three venous blood samples. The maximum discrepancy allowed was 7%,

which is two times the repeatability of the blood sample well counter measurements. Average input curves from the three VOI defined in the three vascular structures were used as input function during kinetic analysis. Tumour VOI were defined on the OSEM reconstructed images using a semi-automatic procedure [14, 27] based on a source to background-adapted 50% isocontour. These VOI were then projected onto all frames of the FBP reconstructed dynamic study to provide tumour TAC. Whenever possible, TAC for up to five lesions per patient was defined. The same lesions were selected for all longitudinal scans of a patient.

## Quantitative analysis and kinetic modelling

Four quantitative methods were investigated. These methods have been described in detail before [11] and a brief summary is described below:

1. NLR is the most quantitative method and involves fitting tumour TAC to the standard two-tissue irreversible compartment model (three rate constants plus blood volume fraction parameter) using an IDIF as arterial input function. The primary outcome parameters of this method are the overall influx rate constant ( $K_i$ ) in  $\text{min}^{-1}$  and the metabolic rate of glucose ( $\text{MR}_{\text{Glu}}$ ) in  $\text{mmol} \cdot \text{l}^{-1} \cdot \text{min}^{-1}$ , which equals  $K_i$  times blood glucose. The lumped constant was assumed to be equal to 1 in all studies.
2. Patlak graphical analysis is based on a linearization of the standard FDG model mentioned above, assuming that the free concentration of FDG in tissue reaches a steady state and that binding/trapping is irreversible. As the method is rapid and robust in the presence of noise, it has been recommended by the EORTC [16]. The primary outcome parameters of this method again are  $K_i$  and  $\text{MR}_{\text{Glu}}$ . In the Patlak method, linear regression is applied to a certain (user-defined) time interval of the collected data. Although, in the present study, various time intervals were investigated, only results for the interval 10–60 min post-injection will be presented, as these showed the best correlation ( $R^2=0.97$ ) with results obtained using NLR.
3. SKM: in contrast to NLR and Patlak, this method [18] only requires a static FDG scan and, in principle, a single venous blood sample obtained at 55 min post-injection. SKM can (partly) compensate for e.g. differences in FDG clearance from blood between various subjects or scans. SKM uses a tri-exponential function to describe the input function. The parameters of this function are partly based on population average value, but the third exponential function, describing the tail, is called using the blood concentration measured in a few blood samples. The primary outcome parameter of SKM is  $\text{MR}_{\text{Glu}}$ .

4. SUV: various SUV alternatives were investigated, which were based on different normalization procedures, e.g. body weight ( $SUV_{BW}$ ), body surface area ( $SUV_{BSA}$ ) and lean body mass ( $SUV_{LBM}$ ). In addition, all SUV parameters were calculated with and without a correction for blood glucose.

SKM and SUV are based on static PET images, i.e. they use the FDG concentration at a fixed time after FDG administration. In the present study dynamic scans were performed and thus the average tumour activity concentration observed at 50–60 min post-injection was used for calculations.

#### Comparison of methods

Patlak analysis has been recommended by the EORTC [16] as the method of choice for a full quantitative analysis of dynamic FDG PET studies. To assess whether this guideline is still valid for targeted therapies, Patlak-derived  $K_i$  and  $MR_{Glu}$  values were compared with those obtained using NLR fitting of TAC using a two-tissue compartment model with blood volume fraction correction.

Quantitative measures of FDG uptake in response scans were related to those obtained during baseline scans and fractional or percentage changes calculated. Percentage changes in tracer uptake due to therapy seen with various simplified methods and derived on a lesion per lesion basis were correlated against those obtained using  $MR_{Glu}$ -Patlak.

#### Comparison of input functions

As SUV assumes a constant relationship between injected radioactivity and body distribution, possible discrepancies in observed fractional changes (relative to full kinetic analysis) may be due to therapy-induced changes in shape and/or amplitude of the input function. Therefore, normalized areas under the curve (AUC) of observed input functions were calculated. Normalization was performed by dividing the observed AUC of the input function by the net amount of administered dose. Normalized AUC (nAUC) obtained during response scans were compared with those at baseline.

## Results

For study A, 18 lesions were defined over the 7 subjects. The mean ( $\pm$ SD) plasma glucose level was  $6.5 \pm 3.3$  (range 4.0–14.1)  $\text{mmol} \cdot \text{l}^{-1}$  for baseline studies and  $7.1 \pm 4.1$  (range 4.5–16.0)  $\text{mmol} \cdot \text{l}^{-1}$  for response studies ( $p$ =NS, Student's  $t$  test). Although the subjects were well prepared, high

plasma glucose levels ( $>8 \text{ mmol} \cdot \text{l}^{-1}$ ) were seen for two subjects at baseline and/or after treatment. All other subjects showed plasma glucose levels lower than  $6 \text{ mmol} \cdot \text{l}^{-1}$ .

Similarly, for study B, nine lesions from six subjects were analysed. However, for five subjects early response studies were available; of these five only two subjects successfully concluded the late response studies. For one subject only a late response study was successfully collected. The mean plasma glucose level was  $5.6 \pm 0.9$  (range 3.8–7.1)  $\text{mmol} \cdot \text{l}^{-1}$  for baseline and  $5.5 \pm 1.36$  (range 4.2–8.4)  $\text{mmol} \cdot \text{l}^{-1}$  for response ( $p$ =NS, Student's  $t$  test). All plasma glucose levels were within normal range [6, 11].

#### Validation of Patlak against NLR

A comparison of Patlak- versus NLR-derived  $MR_{Glu}$  is given in Fig. 1 for both studies. Correlations were excellent (all  $R^2 > 0.96$ ) for both studies, both at baseline and following therapy (Table 1). Moreover, slopes were close to 1 and did not significantly change between baseline and response studies. Although good correlations were found, there were small differences in  $MR_{Glu}$  values between those derived using NLR versus Patlak. These differences, however, correlated with fitted blood volume fractions, as shown in Fig. 2.

#### Correlation of methods

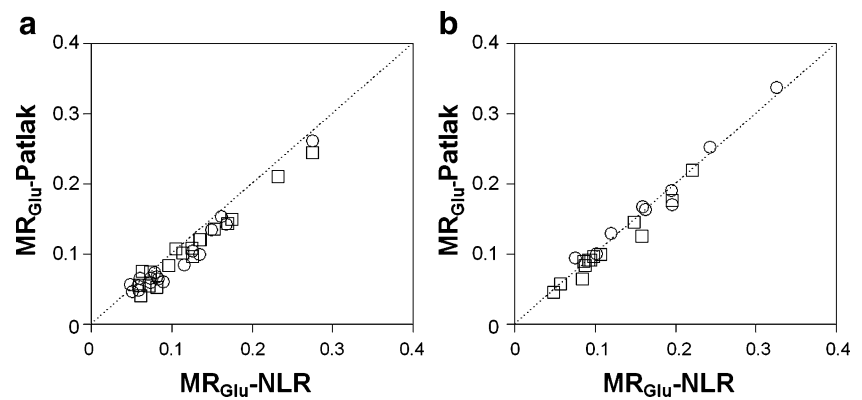
Results of the comparison of simplified methods against  $MR_{Glu}$ -Patlak are summarized in Tables 2 (study A) and 3 (study B).

In the case of study A, fair correlation (approximately  $R^2 > 0.6$ ) was seen at baseline for all simplified methods. In response studies, however, correlations worsened for those methods that do not correct for plasma glucose levels, i.e.  $SUV_{BW}$ ,  $SUV_{BSA}$  and  $SUV_{LBM}$ , and were significantly better for methods that correct for plasma glucose levels. When removing subjects with high blood glucose values ( $>11 \text{ mmol} \cdot \text{l}^{-1}$ ), correlation between SUV and  $MR_{Glu}$ -Patlak improved considerably for most SUV measures. In the case of study B, effects of plasma glucose corrections on correlations were not seen. In this particular study, however, (all) simplified methods (independent of plasma glucose correction) provided poorer correlations for baseline studies than for response studies.

#### Response assessment: comparison of methods

**Study A** The mean fractional change in  $MR_{Glu}$ -Patlak following therapy was  $26 \pm 35\%$  (range  $-24$  to  $87\%$ ), i.e. for most subjects and lesions an increase in  $MR_{Glu}$  was observed, possibly indicating progression of disease.

**Fig. 1** Correlation of Patlak-derived metabolic rate of glucose ( $MR_{Glu}$ ) versus nonlinear regression (NLR) for study A (**a**) and study B (**b**). Baseline data are indicated using *circles* and response studies using *squares*. Dashed line represents line of identity



Clinical interpretation of these results is beyond the scope of the present study. In Fig. 3 fractional changes obtained using simplified methods are plotted against those seen using  $MR_{Glu}$ -Patlak. Percentage changes obtained using various SUV measures without glucose correction may, at times, differ substantially from those seen using  $MR_{Glu}$ -Patlak. It should be noted, however, that the largest deviations were observed for two subjects with the highest plasma glucose values ( $>8 \text{ mmol}\cdot\text{l}^{-1}$ ). Figure 3a also illustrates that results for  $SUV_{BW}$ ,  $SUV_{BSA}$  and  $SUV_{LBM}$  were very similar. Fractional changes obtained using SUV with plasma glucose correction (Fig. 3b) and SKM (Fig. 3c) generally showed good agreement with those obtained using  $MR_{Glu}$ -Patlak.

**Study B** The mean fractional change in  $MR_{Glu}$ -Patlak following therapy was  $-34\pm 25\%$  (range  $-64$  to  $34\%$ ). In Fig. 4 fractional changes obtained using simplified methods are plotted against those obtained using  $MR_{Glu}$ -Patlak. Figure 4a illustrates results obtained with various SUV measures without glucose correction, whilst results after glucose correction are presented in Fig. 4b. Figure 4c shows data obtained using SKM. Note that in these figures we illustrated results for the primary lesion only to avoid unbalanced representation of subject data. From these figures it can be deduced that all simplified methods provide smaller fractional changes than  $MR_{Glu}$ -Patlak. Interestingly, in this case, plasma glucose correction did not reduce the discrepancy between SUV and  $MR_{Glu}$ -Patlak. Similar results were obtained for SUV measures with different normalization factors (BW, BSA, LBM).

Fractional changes found with SKM, however, seemed to agree slightly better with those found with  $MR_{Glu}$ -Patlak (Fig. 4c). Tables 4 and 5 summarize all correlation coefficients and slopes for studies A and B, respectively.

#### nAUC of input functions

Figure 5 illustrates the correlation of nAUC of input functions between response and baseline studies. For study A (Fig. 5a), nAUC of the input function were similar for baseline and response studies. In contrast, for study B (Fig. 5b) differences were seen for some cases.

## Discussion

### Validation of Patlak analysis for response assessment

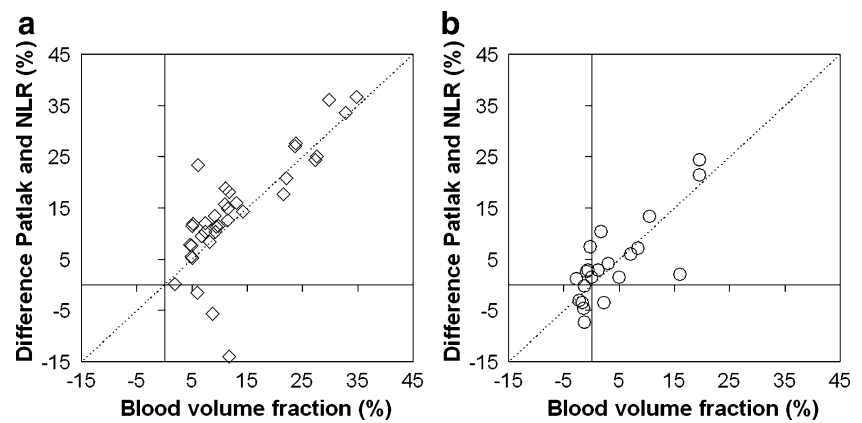
For both studies very good correlations ( $R^2 > 0.96$ ) between  $MR_{Glu}$ -NLR and  $MR_{Glu}$ -Patlak were found. This finding corresponds with previously published data [11] and supports the EORTC recommendation of using Patlak analysis for quantification of  $MR_{Glu}$  [16]. A possible limitation of Patlak analysis, however, is that a correction for blood volume is not included. This may explain the small differences in  $MR_{Glu}$  between both methods, especially for the lung cancer study (study A). The measured fractional difference in  $MR_{Glu}$  between both methods correlates with the fractional blood volume fraction or spillover fraction in those tumours located near large blood

**Table 1** Slope, standard error and correlation coefficient ( $R^2$ ) of  $MR_{Glu}$ -Patlak versus  $MR_{Glu}$ -NLR

Study	Baseline (fixed intercept to 0)		Response (fixed intercept to 0)		Baseline			Response		
	$R^2$	Slope $\pm$ SE	$R^2$	Slope $\pm$ SE	$R^2$	Slope $\pm$ SE	Intercept $\pm$ SE	$R^2$	Slope $\pm$ SE	Intercept $\pm$ SE
A	0.96	$0.87\pm 0.02$	0.97	$0.87\pm 0.02$	0.96	$0.92\pm 0.05$	$-0.01\pm 0.01$	0.97	$0.89\pm 0.04$	$-0.00\pm 0.01$
B	0.97	$1.01\pm 0.02$	0.96	$0.93\pm 0.02$	0.97	$0.99\pm 0.06$	$0.00\pm 0.01$	0.96	$0.92\pm 0.06$	$0.00\pm 0.01$



**Fig. 2** Relative difference between  $MR_{Glu}$ -Patlak and  $MR_{Glu}$ -NLR for both baseline and response scans against vascular volume fractions for study A (a) and study B (b). Dashed line represents line of identity



vessels (Fig. 2). In theory, Patlak could provide a slightly incorrect tumour response when there are large changes in blood volume fraction following therapy. For study A regression slopes between  $MR_{Glu}$ -NLR and  $MR_{Glu}$ -Patlak were  $0.87 \pm 0.02$  for both baseline and response studies. For study B a small difference in these regression slopes was seen, changing from  $1.01 \pm 0.02$  at baseline to  $0.93 \pm 0.02$  following therapy. Consequently, in this case responses measured using  $MR_{Glu}$ -Patlak would be biased by approximately 8%. This, however, should be balanced against the main advantages of Patlak analysis, its speed and its insensitivity to noise. The latter is relevant not only for high accuracy, but also a good test-retest variability [31] is needed to detect metabolic responses [11, 17]. Therefore, despite the small bias in measured response, Patlak analysis

was used for assessing the simplified methods also for study B.

#### Study A: correlation and validation of methods

For study A good correlations were found between several semi-quantitative (simplified) measures and  $MR_{Glu}$ -Patlak. These results are in agreement with previous reports [11, 21, 24]. However, post-treatment studies showed decreased correlation compared with baseline scans. A similar observation was made using SUV for response monitoring. A few subjects showed an SUV response that was substantially different from the  $MR_{Glu}$  response. This was primarily due to high values of and/or large changes in blood glucose levels, as correlations were higher for SKM

**Table 2** Slope, standard error and correlation coefficient ( $R^2$ ) of various semi-quantitative methods versus  $MR_{Glu}$ -Patlak for study A

Quantification methods	Baseline (fixed intercept to 0)		Response (fixed intercept to 0)		Baseline			Response		
	$R^2$	Slope $\pm$ SE	$R^2$	Slope $\pm$ SE	$R^2$	Slope $\pm$ SE	Intercept $\pm$ SE	$R^2$	Slope $\pm$ SE	Intercept $\pm$ SE
SUV <sub>BW</sub>	0.82 (0.82)	53.39 $\pm$ 3.30	0.39 (0.90)	41.76 $\pm$ 4.43	0.83 (0.83)	58.47 $\pm$ 6.66	-0.61 $\pm$ 0.69	0.41 (0.93)	34.47 $\pm$ 10.37	0.97 $\pm$ 1.25
SUV <sub>BWg</sub>	0.83 (0.87)	48.66 $\pm$ 2.79	0.76 (0.91)	44.24 $\pm$ 2.27	0.84 (0.88)	51.64 $\pm$ 5.71	-0.36 $\pm$ 0.59	0.78 (0.92)	38.90 $\pm$ 5.21	0.71 $\pm$ 0.63
SUV <sub>LBM</sub>	0.62 (0.60)	41.18 $\pm$ 2.53	0.03 (-0.21)	34.90 $\pm$ 3.21	0.77 (0.79)	29.94 $\pm$ 4.11	1.35 $\pm$ 0.43	0.36 (0.71)	18.78 $\pm$ 6.21	2.16 $\pm$ 0.75
SUV <sub>LBMg</sub>	0.58 (0.69)	37.83 $\pm$ 2.28	0.50 (0.14)	39.28 $\pm$ 2.16	0.80 (0.85)	25.97 $\pm$ 3.25	1.43 $\pm$ 0.34	0.77 (0.79)	25.56 $\pm$ 3.46	1.84 $\pm$ 0.42
SUV <sub>BSA</sub>	0.85 (0.85)	1,243.22 $\pm$ 58.60	0.33 (0.71)	1,013.89 $\pm$ 93.02	0.86 (0.86)	1,148.64 $\pm$ 118.01	11.42 $\pm$ 12.36	0.43 (0.92)	706.02 $\pm$ 204.51	41.36 $\pm$ 24.77
SUV <sub>BSAg</sub>	0.84 (0.91)	1,140.61 $\pm$ 52.76	0.79 (0.79)	1,111.32 $\pm$ 44.21	0.86 (0.92)	1,005.42 $\pm$ 101.84	16.33 $\pm$ 10.66	0.88 (0.91)	857.47 $\pm$ 78.34	34.11 $\pm$ 9.48
SKM	0.85 (0.90)	1.15 $\pm$ 0.05	0.78 (0.87)	1.10 $\pm$ 0.04	0.87 (0.91)	1.01 $\pm$ 0.09	0.01 $\pm$ 0.01	0.85 (0.94)	0.88 $\pm$ 0.09	0.02 $\pm$ 0.01

Values in parentheses indicate correlation after removal of the two subjects with high blood glucose values ( $>11 \text{ mmol}\cdot\text{l}^{-1}$ )

**Table 3** Slope, standard error and correlation coefficient ( $R^2$ ) of various semi-quantitative methods versus  $MR_{Glu}$ -Patlak for study B

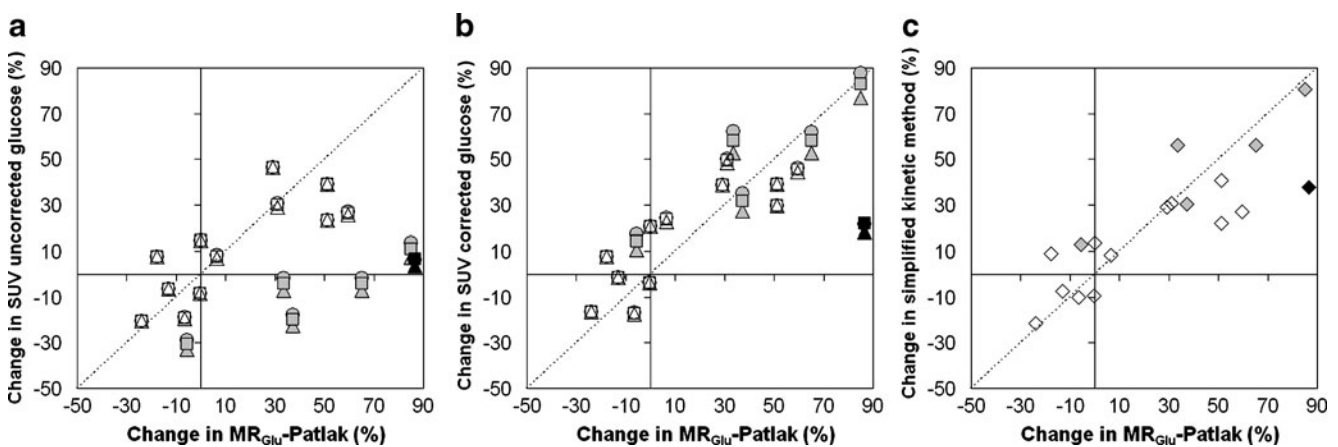
Quantification methods	Baseline (fixed intercept to 0)		Response (fixed intercept to 0)		Baseline			Response		
	$R^2$	Slope $\pm$ SE	$R^2$	Slope $\pm$ SE	$R^2$	Slope $\pm$ SE	Intercept $\pm$ SE	$R^2$	Slope $\pm$ SE	Intercept $\pm$ SE
SUV <sub>BW</sub>	0.11	37.64 $\pm$ 4.00	0.88	56.20 $\pm$ 2.73	0.43	20.95 $\pm$ 9.13	3.46 $\pm$ 1.75	0.88	61.95 $\pm$ 6.76	-0.73 $\pm$ 0.79
SUV <sub>BWg</sub>	0.22	37.37 $\pm$ 3.37	0.80	52.77 $\pm$ 3.16	0.57	21.67 $\pm$ 7.13	3.25 $\pm$ 1.37	0.80	53.21 $\pm$ 8.13	-0.06 $\pm$ 0.94
SUV <sub>LBM</sub>	0.15	32.77 $\pm$ 3.07	0.94	47.23 $\pm$ 1.31	0.54	18.34 $\pm$ 6.44	2.99 $\pm$ 1.24	0.94	44.59 $\pm$ 3.26	0.34 $\pm$ 0.38
SUV <sub>LBMg</sub>	0.26	32.49 $\pm$ 2.54	0.79	44.35 $\pm$ 2.29	0.71	18.69 $\pm$ 4.51	2.86 $\pm$ 0.87	0.81	38.82 $\pm$ 5.61	0.70 $\pm$ 0.65
SUV <sub>BSA</sub>	0.02	903.36 $\pm$ 95.19	0.92	1,355.20 $\pm$ 50.47	0.43	474.73 $\pm$ 206.97	88.83 $\pm$ 39.75	0.92	1,425.75 $\pm$ 127.85	-8.96 $\pm$ 14.85
SUV <sub>BSAg</sub>	0.16	892.40 $\pm$ 72.59	0.87	1,268.03 $\pm$ 55.81	0.68	492.08 $\pm$ 126.61	82.97 $\pm$ 24.32	0.87	1,238.16 $\pm$ 143.35	3.79 $\pm$ 16.65
SKM	0.57	0.92 $\pm$ 0.06	0.85	1.22 $\pm$ 0.06	0.74	0.64 $\pm$ 0.14	0.06 $\pm$ 0.03	0.85	1.21 $\pm$ 0.15	0.00 $\pm$ 0.02

and SUV measures with blood glucose correction. It should be noted, however, that these effects were caused by two of the seven subjects. Nevertheless, these findings demonstrate the need to correct for blood glucose levels if it cannot be guaranteed that they will be within the normal range.

One strategy to avoid inaccurate quantification using SUV would be to reschedule the study when high blood glucose levels are observed. Current guidelines [9, 14, 15] recommend postponement of a FDG PET study if the blood glucose level is elevated ( $>11 \text{ mmol}\cdot\text{l}^{-1}$ ). Moreover, this exclusion criterion still implies that subjects with a blood glucose level of  $10 \text{ mmol}\cdot\text{l}^{-1}$ , i.e. a value that is almost twofold higher than the average, would still be included in

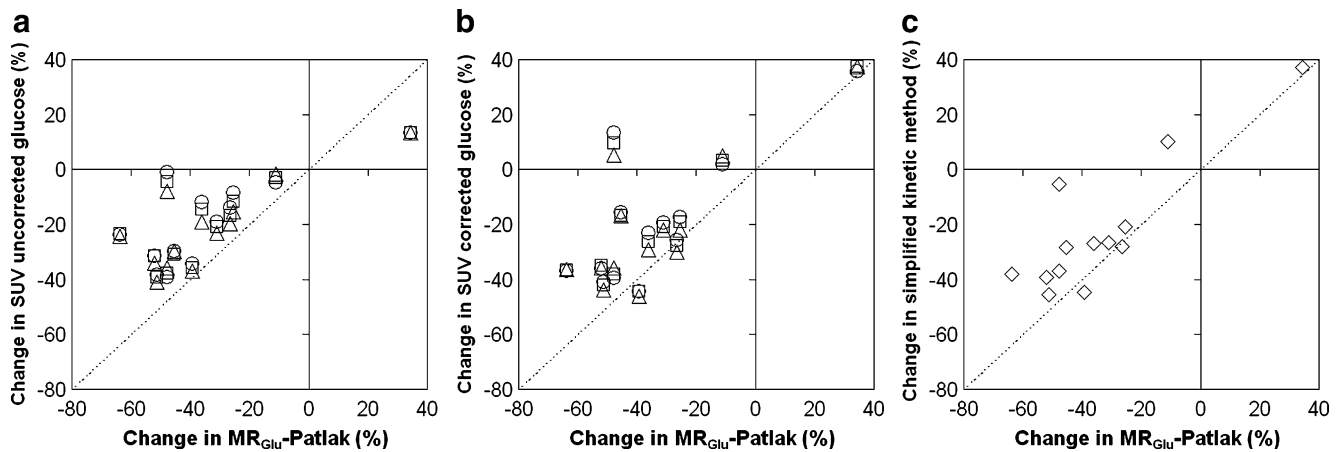
a trial. Such a high value would likely result in lower SUV values [32].

Another option, supported by the present results, is to include a glucose correction within the SUV calculation. However, use of blood glucose corrections is a matter of debate. Data have been published that either favour or discourage use of blood glucose-corrected SUV. Recently, Boellaard et al. [14, 27] and Velasquez et al. [31] reported results from a multi-centre trial, showing that test-retest variability of SUV worsened due to blood glucose correction. Interestingly, a subset of the same data, collected at a single centre, showed improved test-retest variability following blood glucose correction. The possible explanation for this discrepancy is that some of the centres



**Fig. 3** Relative percentage changes in SUV and SKM due to therapy compared with corresponding changes in  $MR_{Glu}$ -Patlak on a lesion per lesion basis for SUV<sub>BW</sub> (triangles), SUV<sub>LBM</sub> (circles) and SUV<sub>BSA</sub> (squares) (a) uncorrected for blood glucose and (b) corrected for blood glucose, and (c) for SKM in study A. Grey symbols represent

data of subjects with high blood glucose in the response scan only ( $<11 \text{ mmol}\cdot\text{l}^{-1}$ ). Black symbols represent subject data with high blood glucose values in both scans ( $>11 \text{ mmol}\cdot\text{l}^{-1}$ ). The open symbols represent data from subjects having a normal blood glucose level. Dashed line represents line of identity



**Fig. 4** Relative percentage changes in SUV and SKM due to therapy compared with corresponding changes in  $MR_{Glu}$ -Patlak on a lesion per lesion basis for  $SUV_{BW}$  (triangles),  $SUV_{LBM}$  (circles) and  $SUV_{BSA}$

(squares) (a) uncorrected for blood glucose and (b) corrected for blood glucose, and (c) for SKM in study B. Dashed line represents line of identity

used a bedside method for measuring blood glucose, i.e. a method that provided accurate results within 5% in 46.9% and within 10% in 78.3% of the cases, respectively [33]. Clearly, a more accurate (laboratory) method should be considered for response monitoring studies.

#### Study B: correlation and validation of methods

In contrast to study A, correlations between Patlak and all SUV methods, including those that incorporate a blood glucose correction, were poor for study B. During therapy, correlations between simplified methods (SUV or SKM) and Patlak improved to values previously found in other studies. A relatively consistent difference between fractional SUV and  $MR_{Glu}$ -Patlak responses was seen. All simplified methods showed a substantially smaller fractional change than Patlak analysis. This was observed for both early and late response studies and for all lesions (data not shown). SKM performed better than SUV, which was likely due to therapy-induced changes in FDG plasma clearance. In theory, there are two

factors that affect SUV calculations: (1) the FDG from plasma that is available for the tumour over time (input function) and (2) the concentration of free FDG in intravascular and tissue spaces. For example, tumours with low FDG uptake, having a relatively high fraction of free FDG, may show larger discrepancies between SUV and Patlak analysis [21, 24, 34]. Results shown in Fig. 5b indicate that for some patient studies there were substantial differences between nAUC of the input functions during baseline and response scans. However, the number of patients included is small and, therefore, any explanation for the observed differences between SUV and  $MR_{Glu}$  responses requires further validation. Nevertheless, nearly all subjects showed differences between SUV and  $MR_{Glu}$  responses.

#### Limitations

Full quantitative dynamic PET studies for response monitoring studies are complex resulting in discomfort for the

**Table 4** Slope, standard error and correlation coefficient ( $R^2$ ) of percentage change in various semi-quantitative methods versus those obtained with  $MR_{Glu}$ -Patlak for study A

Quantification methods	Fractional change (fixed intercept to 0)		Fractional change		
	$R^2$	Slope $\pm$ SE	$R^2$	Slope $\pm$ SE	Intercept $\pm$ SE
$SUV_{BW}$	0.14 (0.62)	0.21 $\pm$ 0.11	0.15 (0.64)	0.24 $\pm$ 0.14	-1.96 $\pm$ 6.26
$SUV_{BWg}$	0.49 (0.61)	0.71 $\pm$ 0.10	0.60 (0.71)	0.57 $\pm$ 0.12	10.34 $\pm$ 4.96
$SUV_{LBM}$	0.21 (0.63)	0.26 $\pm$ 0.10	0.21 (0.65)	0.28 $\pm$ 0.14	-0.97 $\pm$ 5.83
$SUV_{LBMg}$	0.48 (0.60)	0.79 $\pm$ 0.11	0.60 (0.71)	0.62 $\pm$ 0.13	11.85 $\pm$ 5.44
$SUV_{BSA}$	0.18 (0.62)	0.25 $\pm$ 0.11	0.19 (0.64)	0.27 $\pm$ 0.14	-1.42 $\pm$ 5.99
$SUV_{BSAg}$	0.50 (0.60)	0.76 $\pm$ 0.10	0.61 (0.71)	0.60 $\pm$ 0.12	11.16 $\pm$ 5.13
SKM	0.68 (0.70)	0.72 $\pm$ 0.08	0.71 (0.72)	0.63 $\pm$ 0.10	6.01 $\pm$ 4.32

Values in parentheses indicate correlation after removal of the two subjects with high blood glucose values ( $>11 \text{ mmol}\cdot\text{l}^{-1}$ )



**Table 5** Slope, standard error and correlation coefficient ( $R^2$ ) of percentage change in various semi-quantitative methods versus those obtained with MR<sub>Glu</sub>-Patlak for study B

Quantification methods	Fractional change (fixed intercept to 0)		Fractional change		
	$R^2$	Slope $\pm$ SE	$R^2$	Slope $\pm$ SE	Intercept $\pm$ SE
SUV <sub>BW</sub>	0.63	0.58 $\pm$ 0.06	0.65	0.50 $\pm$ 0.11	-4.03 $\pm$ 4.63
SUV <sub>BWg</sub>	0.61	0.66 $\pm$ 0.09	0.62	0.76 $\pm$ 0.18	4.94 $\pm$ 7.39
SUV <sub>LBM</sub>	0.56	0.53 $\pm$ 0.07	0.57	0.48 $\pm$ 0.13	-2.17 $\pm$ 5.31
SUV <sub>LBMg</sub>	0.56	0.61 $\pm$ 0.10	0.58	0.72 $\pm$ 0.19	5.45 $\pm$ 7.77
SUV <sub>BSA</sub>	0.59	0.55 $\pm$ 0.07	0.61	0.49 $\pm$ 0.12	-3.03 $\pm$ 4.96
SUV <sub>BSAg</sub>	0.59	0.63 $\pm$ 0.10	0.61	0.74 $\pm$ 0.18	5.64 $\pm$ 7.49
SKM	0.72	0.71 $\pm$ 0.08	0.74	0.83 $\pm$ 0.15	5.63 $\pm$ 6.08

subjects. Consequently, only a small number of patients are usually recruited (and/or approval is given for a small number of subjects). Additionally, in the present study a limited number of successfully completed studies were available. Despite the small number of subjects, however, both studies illustrate potential difficulties associated using SUV for response monitoring purposes. In the case of study B all observed fractional changes in SUV were different (smaller) from those seen with Patlak analysis. Therefore, the results of this study confirm that SUV responses may differ substantially from those obtained using full kinetic analysis [21, 24, 25].

It should be noted that most simplified methods can be applied in a routine clinical setting (i.e. using whole-body scanning), while full kinetic analysis requires a dynamic scan. An alternative could be to first perform a 60-min dynamic scan covering the primary tumour only followed with a whole-body acquisition to study presence and tracer uptake in distant metastases. However, in practice this strategy would seriously affect patient comfort and throughput. Moreover, apart from the limited coverage of the patient, a dynamic scan is also more sensitive to patient motion which could affect the accuracy and precision of full kinetic analysis.

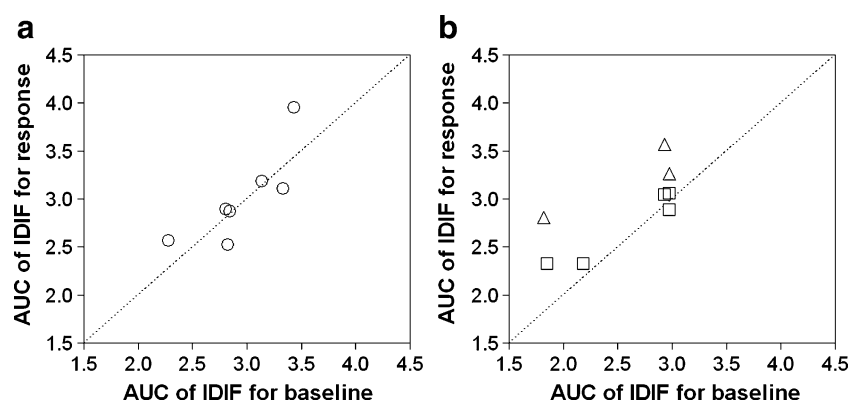
Finally, in our paper image-derived input functions were used. The quality of these input functions might be affected by partial volume effects, i.e. spillover of signal to and from

surrounding tissues. Therefore, in the present studies three blood samples were drawn to perform a quality check of the image-derived input function to avoid that these are affected by partial volume effects and/or patient motion. Therefore, it is expected that use of image-derived input functions in our studies did not affect the main conclusions of this paper. Yet, it is acknowledged that image-derived input functions should be used with care and that at least a quality check on the appropriateness of the input function should be performed, e.g. as described in [11, 17].

## Conclusion

In this study use of simplified methods for quantification of treatment response with FDG PET was compared with full kinetic analysis. Fractional changes seen with SUV differ from those seen using full quantification based on dynamic scans and Patlak analysis due to e.g. elevated plasma glucose values or changes in FDG plasma clearance. SUV may thus provide different response measures than those observed with metabolic rate of glucose. For clinical trials this could imply that use of SUV for assessment of drug efficacy may under- or overestimate treatment effects compared with full kinetic analysis and thus it would be recommended that SUV responses are initially characterized prior to implementation in larger clinical trials.

**Fig. 5** **a** Area under the curve (AUC) of image-derived input functions (IDIF), normalized for injected dose, of response versus baseline studies for study A. **b** Same for study B; early response data are indicated using squares and late response studies using triangles. Dashed line represents line of identity



**Acknowledgements** The authors would like to thank the Department of Nuclear Medicine & PET Research for FDG production and acquisition of PET studies. The work described in this paper was in part financed by Bristol-Myers Squibb Co. Patsuree Cheebsumon was supported by a scholarship from the National Science and Technology Development Agency of the Royal Thai Government.

**Conflicts of interest** The Department of Nuclear Medicine & PET Research of the VU University Medical Center was financially sponsored by Bristol-Myers Squibb Co. for participating in the trial and for providing consultancy.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

## References

- Avril NE, Weber WA. Monitoring response to treatment in patients utilizing PET. *Radiol Clin North Am* 2005;43:189–204.
- Bastiaannet E, Groen H, Jager PL, Cobben DC, van der Graaf WT, Vaalburg W, et al. The value of FDG-PET in the detection, grading and response to therapy of soft tissue and bone sarcomas; a systematic review and meta-analysis. *Cancer Treat Rev* 2004;30:83–101.
- Erdi YE. The use of PET for radiotherapy. *Curr Med Imaging Rev* 2007;3:3–16.
- Fletcher JW, Djulbegovic B, Soares HP, Siegel BA, Lowe VJ, Lyman GH, et al. Recommendations on the use of 18F-FDG PET in oncology. *J Nucl Med* 2008;49:480–508.
- de Geus-Oei LF, van der Heijden HF, Corstens FH, Oyen WJ. Predictive and prognostic value of FDG-PET in nonsmall-cell lung cancer: a systematic review. *Cancer* 2007;110:1654–64.
- Hoekstra CJ, Stroobants SG, Smit EF, Vansteenkiste J, van Tinteren H, Postmus PE, et al. Prognostic relevance of response evaluation using [18F]-2-fluoro-2-deoxy-D-glucose positron emission tomography in patients with locally advanced non-small-cell lung cancer. *J Clin Oncol* 2005;23:8362–70.
- Juweid ME, Stroobants S, Hoekstra OS, Mottaghy FM, Dietlein M, Guerazzi A, et al. Use of positron emission tomography for response assessment of lymphoma: consensus of the Imaging Subcommittee of International Harmonization Project in Lymphoma. *J Clin Oncol* 2007;25:571–8.
- Vansteenkiste JF, Stroobants SG. The role of positron emission tomography with 18F-fluoro-2-deoxy-D-glucose in respiratory oncology. *Eur Respir J* 2001;17:802–20.
- Weber WA. Use of PET for monitoring cancer therapy and for predicting outcome. *J Nucl Med* 2005;46:983–95.
- Borst GR, Belderbos JS, Boellaard R, Comans EF, De Jaeger K, Lammertsma AA, et al. Standardised FDG uptake: a prognostic factor for inoperable non-small cell lung cancer. *Eur J Cancer* 2005;41:1533–41.
- 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;43:1304–9.
- Cheson BD, Pfistner B, Juweid ME, Gascoyne RD, Specht L, Horning SJ, et al. Revised response criteria for malignant lymphoma. *J Clin Oncol* 2007;25:579–86.
- Juweid ME, Cheson BD. Positron-emission tomography and assessment of cancer therapy. *N Engl J Med* 2006;354:496–507.
- Boellaard R. Standards for PET image acquisition and quantitative data analysis. *J Nucl Med* 2009;50 Suppl 1:11S–20S.
- Shankar LK, Hoffman JM, Bacharach S, Graham MM, Karp J, Lammertsma AA, et al. Consensus recommendations for the use of 18F-FDG PET as an indicator of therapeutic response in patients in National Cancer Institute Trials. *J Nucl Med* 2006;47:1059–66.
- Young H, Baum R, Cremerius U, Herholz K, Hoekstra O, Lammertsma AA, et al. Measurement of clinical and subclinical tumour response using [18F]-fluorodeoxyglucose and positron emission tomography: review and 1999 EORTC recommendations. European Organization for Research and Treatment of Cancer (EORTC) PET Study Group. *Eur J Cancer* 1999;35:1773–82.
- Hoekstra CJ, Paglianiti I, Hoekstra OS, Smit EF, Postmus PE, Teule GJ, et al. Monitoring response to therapy in cancer using [18F]-2-fluoro-2-deoxy-D-glucose and positron emission tomography: an overview of different analytical methods. *Eur J Nucl Med* 2000;27:731–43.
- Hunter GJ, Hamberg LM, Alpert NM, Choi NC, Fischman AJ. Simplified measurement of deoxyglucose utilization rate. *J Nucl Med* 1996;37:950–5.
- Sadato N, Tsuchida T, Nakaumra S, Waki A, Uematsu H, Takahashi N, et al. Non-invasive estimation of the net influx constant using the standardized uptake value for quantification of FDG uptake of tumours. *Eur J Nucl Med* 1998;25:559–64.
- Lammertsma AA, Hoekstra CJ, Giacccone G, Hoekstra OS. How should we analyse FDG PET studies for monitoring tumour response? *Eur J Nucl Med Mol Imaging* 2006;33 Suppl 1:16–21.
- Freedman NM, Sundaram SK, Kurdziel K, Carrasquillo JA, Whatley M, Carson JM, et al. Comparison of SUV and Patlak slope for monitoring of cancer therapy using serial PET scans. *Eur J Nucl Med Mol Imaging* 2003;30:46–53.
- Huang SC. Anatomy of SUV. Standardized uptake value. *Nucl Med Biol* 2000;27:643–6.
- Kim CK, Gupta NC. Dependency of standardized uptake values of fluorine-18 fluorodeoxyglucose on body size: comparison of body surface area correction and lean body mass correction. *Nucl Med Commun* 1996;17:890–4.
- McDermott GM, Welch A, Staff RT, Gilbert FJ, Schweiger L, Semple SI, et al. Monitoring primary breast cancer throughout chemotherapy using FDG-PET. *Breast Cancer Res Treat* 2007;102:75–84.
- Doot RK, Dunnwald LK, Schubert EK, Muzi M, Peterson LM, Kinahan PE, et al. Dynamic and static approaches to quantifying 18F-FDG uptake for measuring cancer response to therapy, including the effect of granulocyte CSF. *J Nucl Med* 2007;48:920–5.
- Mankoff DA, Muzi M, Krohn KA. Quantitative positron emission tomography imaging to measure tumor response to therapy: what is the best method? *Mol Imaging Biol* 2003;5:281–5.
- Boellaard R, Oyen WJ, Hoekstra CJ, Hoekstra OS, Visser EP, Willemsen AT, et al. The Netherlands protocol for standardisation and quantification of FDG whole body PET studies in multi-centre trials. *Eur J Nucl Med Mol Imaging* 2008;35:2320–33.
- Brix G, Zaers J, Adam LE, Bellemann ME, Ostertag H, Trojan H, et al. Performance evaluation of a whole-body PET scanner using the NEMA protocol. National Electrical Manufacturers Association. *J Nucl Med* 1997;38:1614–23.
- Boellaard R, van Lingen A, Lammertsma AA. Experimental and clinical evaluation of iterative reconstruction (OSEM) in dynamic PET: quantitative characteristics and effects on kinetic modeling. *J Nucl Med* 2001;42:808–17.
- van der Weerd AP, Klein LJ, Boellaard R, Visser CA, Visser FC, Lammertsma AA. Image-derived input functions for determination of MRGlu in cardiac (18F)-FDG PET scans. *J Nucl Med* 2001;42:1622–9.

31. Velasquez LM, Boellaard R, Kollia G, Hayes W, Hoekstra OS, Lammertsma AA, et al. Repeatability of  $^{18}\text{F}$ -FDG PET in a multicenter phase I study of patients with advanced gastrointestinal malignancies. *J Nucl Med* 2009;50:1646–54.
32. Eary JF, Mankoff DA. Tumor metabolic rates in sarcoma using FDG PET. *J Nucl Med* 1998;39:250–4.
33. Dai KS, Tai DY, Ho P, Chen CC, Peng WC, Chen ST, et al. Accuracy of the EasyTouch blood glucose self-monitoring system: a study of 516 cases. *Clin Chim Acta* 2004;349:135–41.
34. Sundaram SK, Freedman NM, Carrasquillo JA, Carson JM, Whatley M, Libutti SK, et al. Simplified kinetic analysis of tumor  $^{18}\text{F}$ -FDG uptake: a dynamic approach. *J Nucl Med* 2004;45:1328–33.