
Reproducibility of Standardized Uptake Value Measurements Determined by ^{18}F -FDG PET in Malignant Tumors

Claude Nahmias¹ and Lindi M. Wahl²

¹Molecular Imaging and Translational Research, University of Tennessee, Knoxville, Tennessee; and ²Applied Mathematics, University of Western Ontario, London, Ontario

^{18}F -FDG PET is increasingly being used to monitor the early response of malignant tumors to chemotherapy. Understanding the reproducibility of standardized uptake values (SUVs) is an important prerequisite in estimating what constitutes a significant change. **Methods:** Twenty-six patients were studied on 2 separate occasions (mean interval \pm SD, 3 ± 2 d; range, 1–5 d). A static PET/CT scan was performed 94 ± 9 min after the intravenous injection of 383 ± 15 MBq of ^{18}F -FDG. Mean and maximum SUVs (SUV_{mean} and SUV_{max}, respectively) were determined for regions of interest drawn around the tumor on the first study and for the same regions of interest transferred to the second study. **Results:** SUV_{mean} in tumors ranged from 1.49 to 17.48 and SUV_{max} ranged from 2.99 to 24.09. The correlation between SUV_{mean} determined on the 2 separate visits was 0.99; the mean difference between the 2 measurements was 0.01 ± 0.27 SUV. The 95% confidence limits for the measurements were ± 0.53 . For SUV_{max}, the mean difference was -0.05 ± 1.14 SUV. **Conclusion:** Our study demonstrates that repeated measurements of SUV_{mean} performed a few days apart are highly reproducible. A decrease of 0.5 in the SUV is statistically significant.

Key Words: ^{18}F -FDG; cancer; SUV; reproducibility

J Nucl Med 2008; 49:1804–1808

DOI: 10.2967/jnumed.108.054239

PET is an imaging technique that allows the study of the spatial and temporal distribution of a radiopharmaceutical labeled with a positron emitter. ^{18}F -FDG is one such radiopharmaceutical that is applied to the study of glucose metabolism in vivo and plays an important role in the diagnosis, staging, and restaging of patients with cancer (1). The technique also has the potential to provide an accurate assessment of the early response to multicourse treatment with the ultimate goal of identifying responding and nonresponding tumors (2). ^{18}F -FDG PET has been shown to predict early response to therapy in patients with lymphoma (3), breast

(4), ovarian (5), gastric (6), or non-small-cell lung (7) cancers, among others.

The standardized uptake value (SUV), defined as ^{18}F -FDG retention normalized to injected dose and patient body weight, is an established index for quantifying glucose metabolic activity in tissues. The predominance of this index in the clinical literature was recently demonstrated in a review of 25 studies that have correlated the ^{18}F -FDG PET response with clinical outcome, involving nearly a thousand patients with lymphoma and lung, esophagus, head and neck, and other cancers. In 11 of the 25 studies, SUVs were used as the criterion to evaluate the results. However, the SUV threshold values used to determine patient outcome in these studies varied widely and were determined post hoc. In the other 14 studies, no quantitative measure was used; the images were evaluated visually (8).

To set an objective criterion with which to monitor success or failure of cancer therapy, the within-patient reproducibility of SUVs must be known. To date, a handful of studies have addressed this question, but no consensus has yet emerged. Although a fairly recent study considered the within-patient variability of ^{18}F -FDG uptake in normal tissue (9), estimates of SUV variability in cancer were measured on scanners using either sodium iodide crystals doped with thallium (9) or bismuth germanate (10–12) as detector material, and attenuation-correction factors were derived from transmission measurements. Current PET systems, based on lutetium orthosilicate-type detectors and using information from a CT study to correct for attenuation, should allow for a substantial improvement in accuracy and, thus, better reproducibility. Because the use of inappropriate thresholds or confidence intervals (CIs) has the potential to mask clinically significant change within a patient, the accurate quantification of SUV variability is a critical step in the clinical interpretation of longitudinal ^{18}F -FDG PET results.

The purpose of this study was to estimate metabolic activity in malignant tumors using SUVs as determined by ^{18}F -FDG PET/CT on 2 occasions, within at most 5 d of each other, to assess the reproducibility of the measurement and establish 95% CIs for the measurement.

Received May 9, 2008; revision accepted Jul. 14, 2008.
For correspondence or reprints contact: Lindi Wahl, Applied Mathematics, Middlesex College, University of Western Ontario, London, ON N6A 5B7.
E-mail: lwahl@uwo.ca
COPYRIGHT © 2008 by the Society of Nuclear Medicine, Inc.

MATERIALS AND METHODS

A total of 26 patients (10 women, 16 men; mean age, 61 y; range, 25–72 y) were studied; 9 patients had lung cancer, 6 had metastatic breast cancer, 3 had esophageal cancer, and the remaining 8 patients had cancer in various other locations (Table 1). None of the patients was undergoing chemotherapy at the time of the study. The Institutional Review Board of the University of Tennessee Graduate School of Medicine approved the study protocol. Each patient was studied on 2 separate occasions (mean interval, 3 ± 2 d; range, 1–5 d); written informed consent was obtained before patients were enrolled into the study.

¹⁸F-FDG PET/CT

Each patient was advised to fast for at least 6 h before the examination. Blood glucose concentration was measured before each injection of ¹⁸F-FDG, using blood glucose reagent strips (OneTouch SureStep Blood Glucose Monitoring System; Lifescan). Each patient was studied on a Biograph-6 scanner (Siemens Medical Solutions) 90 min after the intravenous injection of 383 ± 15 MBq of ¹⁸F-FDG. Each patient was scanned from chin to pelvis (5–6 PET bed positions), and PET data were acquired for 3 min at each bed position. A CT scan covering the same area was performed using the following parameters: 130 kVp, 160 mAs, 0.6-s tube rotation, 6×2 mm collimation, table feed of 17.6 mm per rotation, CareDose (Siemens), reconstructed slice thickness of 5.0 mm, 5.0-mm interslice spacing, and a medium smooth convolution kernel. After compensation for random coincidences and scattered radiation and application of

CT-based attenuation correction, PET images were reconstructed onto a 256×256 matrix using an ordered-subset expectation maximization iterative algorithm (4 iterations and 16 subsets), with a 5-mm gaussian postprocessing filter, to a final image resolution of approximately 8 mm in full width at half maximum.

SUVs were obtained from regions of interest (ROIs) drawn around the tumor on the first study and from the same-size ROIs transferred to the second study, using the anatomic information from the CT scan and the metabolic information from the PET scan to ensure proper placement. In the case of multiple metastases, the most metabolically active lesion was selected. The ROIs were defined manually in the axial slice containing the most metabolically active portion of the tumor, using a circular ROI and a threshold of 30% of maximum as a rough guide for the size of the diameter. The radius of the circular ROI varied between 9 and 17 mm. The 9-mm radius corresponds to a diameter of about twice the spatial resolution of the scanner (reconstructed resolution, 8 mm).

The correlation between the 2 measurements was calculated using the Pearson product moment correlation coefficient. Because a high degree of correlation does not necessarily imply good agreement between the 2 measurements, particularly when the data span a large range (*13*), a Bland–Altman plot was constructed to assess this agreement. A Bland–Altman plot displays the difference between the 2 measurements versus their average as a scatter plot, on which each point represents 1 patient. The Kolmogorov–Smirnov test (*14*) was used to confirm that the distribution of the differences between each pair of SUV measurements was not significantly different from

TABLE 1
Characteristics of Patients in Study

Patient no.	Age (y)	Sex	Location of lesion	SUVmean for...		Δt (d)	Cancer diagnosis
				First visit	Second visit		
1	42	F	Lung	10.72	10.59	1	NSCLC
2	58	F	Proximal femur	3.69	4.11	3	Breast
3	60	F	Mediastinum	10.94	11.14	1	Breast
4	65	F	Lung	1.49	1.78	5	Breast
5	71	M	Lung	4.75	4.76	1	NSCLC
6	48	F	Clavicle	5.82	6.11	2	Head and neck
7	65	M	Esophagus	7.11	6.66	2	Esophageal
8	54	M	Esophagus	8.74	8.69	4	Esophageal
9	25	M	Lung	4.54	4.74	4	Lymphoma
10	49	M	Neck	4.16	4.34	2	Esophageal
11	68	M	Mediastinum	2.67	2.40	1	Prostate
12	69	M	Mediastinum	17.48	17.35	2	NSCLC
13	61	M	Lung	11.11	10.86	1	NSCLC
14	62	F	Lung	5.10	5.22	1	NSCLC
15	66	F	Mediastinum	3.90	3.67	4	Breast
16	64	F	Pelvis	4.10	4.48	5	Breast
17	72	F	Lung	2.23	2.25	5	Melanoma
18	62	M	Lung	9.14	9.06	3	NSCLC
19	70	M	Liver	7.95	7.96	1	Colorectal
20	69	M	Mediastinum	4.28	4.03	1	Thyroid
21	53	M	Lung	1.86	2.23	3	Head and neck
22	69	M	Lung	8.07	8.36	4	NSCLC
23	64	M	Abdomen	6.08	6.40	5	Renal
24	55	F	Mediastinum	10.30	9.91	2	Breast
25	72	M	Mediastinum	5.14	4.82	1	NSCLC
26	53	M	Mediastinum	5.27	4.93	5	NSCLC

Δt = number of days between visits; NSCLC = non-small-cell lung cancer.

a normal distribution, and therefore the 95% CI for the difference between the 2 measurements was taken as the mean difference ± 1.96 times the SD of the difference.

To test whether variability was correlated with the intensity of ^{18}F -FDG uptake, the magnitude (absolute value) of the difference between the 2 measurements was plotted versus their average, as was the relative difference between the 2 measurements (percentage change). Again, the Pearson correlation coefficient, r , was determined, and a t statistic was then calculated as:

$$t = r \sqrt{\frac{n - 2}{1 - r^2}}$$

with $n = 26$. This value was compared with a 2-tailed t distribution with 24 degrees of freedom to determine whether the correlation differed significantly from zero. To investigate the use of either the mean SUV in the ROI (SUV_{mean}) or the maximum SUV in the ROI (SUV_{max}), the analysis was repeated for both cases.

RESULTS

The variation in patient weight between the 2 studies was 0.4 ± 1.0 kg (range, -3.1 to 1.4 kg). The plasma glucose concentrations were within the reference range; the mean difference in plasma glucose concentration was 3 ± 10 mg/dL (range, -19 to 32 mg/dL).

The mean ^{18}F -FDG uptake period was 94 ± 9 min; the mean difference between uptake periods was 0 ± 8 min (range, -28 to 16 min). SUV_{mean} in tumors ranged from 1.49 to 17.48, and SUV_{max} ranged from 2.99 to 24.09. The Kolmogorov-Smirnov test confirmed that the distribution of the differences between each pair of SUV measurements was not significantly different from a normal distribution ($P > 0.20$). The r between SUV_{mean} determined on the 2 separate visits was 0.99 ($n = 26$; $P < 0.0001$; 95% CI, 0.99 – 1.00) (Fig. 1).

The Bland-Altman plot revealed excellent agreement between the SUV_{mean} measured in the 2 studies (Fig. 2). The mean difference between the 2 measurements was 0.01 ± 0.27 SUV (SUV_{mean} range, -0.45 to 0.42). As illustrated in

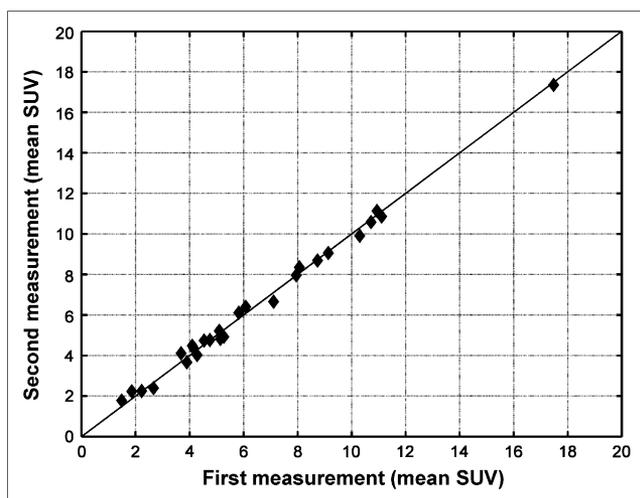


FIGURE 1. Correlation between 2 measurements of SUV_{mean} . Line represents line of identity.

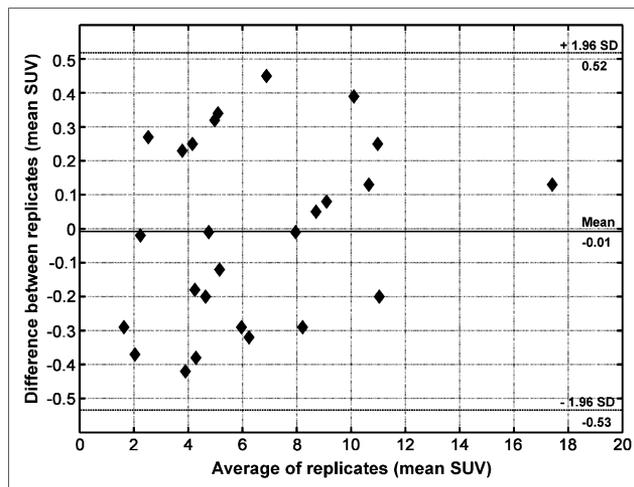


FIGURE 2. Bland-Altman plot on which the difference between 2 SUV_{mean} measurements is plotted against their average. The mean difference is -0.01 SUV; the 95% CI for SUV_{mean} is ± 0.53 SUV.

Figure 2, the 95% CIs for the measurement were ± 0.53 . When the SUVs were normalized to blood glucose concentration, the difference increased to 0.19 ± 0.83 (SUV_{mean} range, -2.17 to 2.73) (Supplemental Table 1; supplemental materials are available online only at <http://jnm.snmjournals.org>), and the 95% CIs increased to ± 1.63 (Supplemental Figure 1).

The Bland-Altman plot revealed a poorer agreement between the SUV_{max} measured in the 2 studies (Fig. 3). In this case, the mean difference between the 2 measurements was -0.05 ± 1.14 SUV (SUV_{max} range, -3.42 to 1.82), and the 95% CIs for the measurement were -2.3 and $+2.2$ SUV.

The magnitude of the variability between the 2 measurements did not correlate with SUV_{mean} (Fig. 4A); in this case, the linear least-squares slope was not different from zero

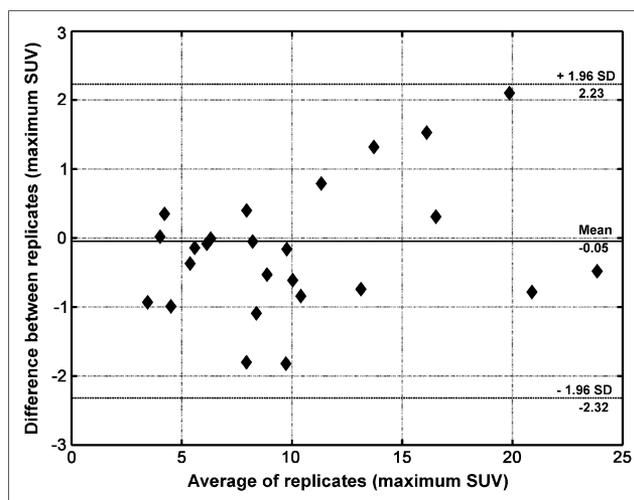


FIGURE 3. Bland-Altman plot on which the difference between 2 SUV_{max} measurements is plotted against their average. The mean difference is -0.05 SUV; the 95% CIs for SUV_{max} are -2.32 and $+2.23$ SUV.

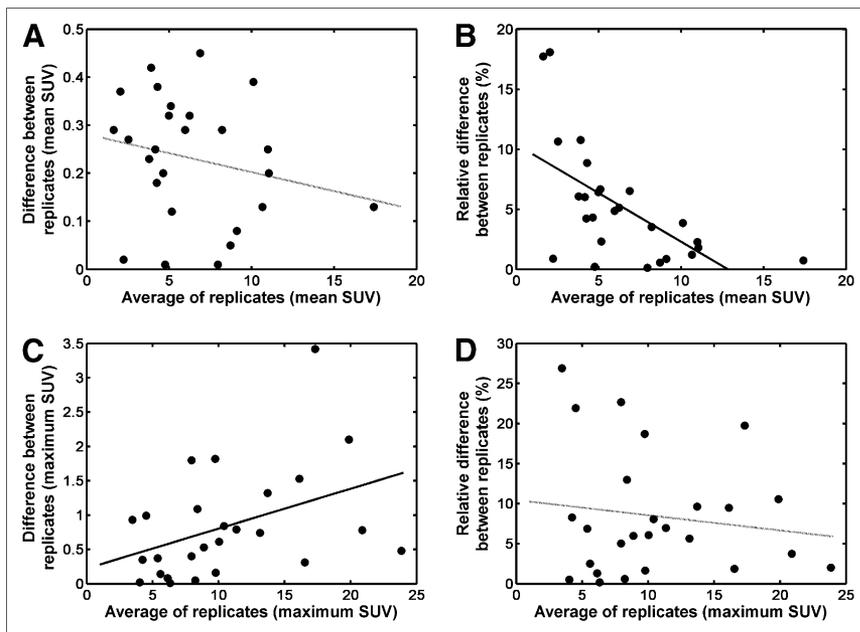


FIGURE 4. Plot of magnitude of difference between 2 measurements versus their average (A and C) and relative difference between 2 measurements (percentage change) (B and D). A and B show results for SUV_{mean} , and C and D show results for SUV_{max} . Lines represent trends in data (solid lines represent slopes that are statistically different from zero, and dotted lines represent slopes that are not different from zero).

($t = 1.1$, $P > 0.25$). Thus, the relative difference (percentage change) in SUV_{mean} clearly decreases with SUV_{mean} (Fig. 4B; $t = 3.75$, $P < 0.001$). In contrast, the magnitude of the variability increased with SUV_{max} (Fig. 4C; $t = 2.2$, $P < 0.05$). In this case, the relative difference was not correlated with SUV_{max} (Fig. 4D; $t = 0.69$, $P > 0.45$).

DISCUSSION

Our results confirm that repeated ^{18}F -FDG PET measurements can be performed with accuracy in patients with cancer being evaluated serially. Various approaches can be used to estimate activity in a tumor. These include qualitative, visual assessments; semiquantitative assessments, such as SUV, SUV corrected for plasma glucose concentration, and SUV corrected for body surface area or lean body mass (SUV-lean); and Patlak graphical analysis. Other approaches include quantitative assessments based on compartmental analysis (15). In our analysis, we have considered the semiquantitative measure of tumor metabolism given by the SUV. This choice was made because SUV is the most common parameter measured in a clinical setting. Its calculation is computationally simple, and most contemporary PET/CT scanners display the images in these units, provided the injected dose and the patient weight have been entered when setting up the PET acquisition. Also, because SUVs are confined to the measurement of radioactivity concentration at a fixed time after injection, they require considerably less scanner time than do the more involved compartmental analyses, which are predicated on dynamic data acquisitions.

After administration, ^{18}F -FDG is taken up and retained by various tissues at various rates. In malignant tumors, the accumulation of ^{18}F seldom reaches a plateau by 2 h after injection, whereas in benign tumors it may reach a plateau within 30 min (16). This observation has led several inves-

tigators to attempt to differentiate between inflammation and neoplasm by performing measurements at 2 time points, such as 60 and 120 min after injection (17,18). If the evaluation of metabolic activity is to be restricted to 1 time point, as would be the case in most clinical settings, this time point should be selected as late after injection as practically feasible. Ideally, the measurement would be performed when the activity concentration in the tissue of interest is changing little with time. In our study, we have chosen 90 min as a reasonable compromise between the ^{18}F activity reaching equilibrium and the physical half-life of the radioisotope, ensuring that sufficient activity remains for an accurate measurement. When using SUV as an estimate of tumor metabolic activity in serial studies, it is important that the measurements be performed at the same time after injection.

Normalization of SUVs to blood glucose concentration has been suggested as a means of providing a more stable parameter that is independent of blood glucose variations (19). Although this may be true, one should consider the error in this measurement that may offset any benefit that may come from glucose normalization, especially when there are only small differences in glucose levels at the time of each study. The coefficient of variation for repeated measurements of blood glucose concentration that is reported in the literature for comparable devices (20) is similar to that determined in our laboratory (5.2%).

One study has considered the within-patient variability of ^{18}F -FDG uptake in normal tissue in 70 patients who underwent two ^{18}F -FDG PET studies, but these studies were performed 271 ± 118 d apart (9). Despite this long interval, the authors of this study concluded that SUVs measured in the liver and mediastinum of patients who were cancer free were stable over time. They also concluded that correcting for blood glucose levels increased the variability of the values, a result

similar to ours, in which we observed the SD of the distribution of the difference between the 2 measurements increase from 0.27 to 0.83. Furthermore, normalizing for body surface area or lean body mass did not improve the reproducibility of the measurements.

To our knowledge, the reproducibility of SUV measurements in malignant tumors has been determined by only 3 groups. One group studied 10 patients with untreated lung cancer. Each patient was studied on 2 separate occasions within 1 wk. They used a single-scan method, SUV normalized to SUV-lean, and a kinetic approach, calculating the Patlak influx constant, K_i , to characterize ^{18}F -FDG uptake in tumors (10). They reported a mean difference of 0.58 SUV and an SD of 0.91. Another group studied 16 patients examined twice within 10 d. They calculated several parameters, including SUV and K_i ; they reported the 95% reference range for spontaneous fluctuations in SUV to be ± 0.91 (11). A third group studied 11 patients with non-small-cell cancer on 2 consecutive days (12). They used a single-scan method and SUV-lean and plasma glucose concentration. They used a variety of methods to draw volumes of interest. For manually drawn regions, they reported a mean difference of 0.04 SUV and an SD of 0.56; for the maximum pixel value approach, they reported a mean difference of 0.25 ± 1.33 SUV. The authors concluded that the highest reproducibility was found for manually determined regions, whereas the poorest reproducibility was for the maximum pixel approach (12).

The single maximum pixel (SUV_{max}) value within the ROI, as opposed to the mean of the pixel values (SUV_{mean}), is often used because it is suggested that the maximum value is less dependent on the placement and drawing of the ROI. Our results indicate that the use of the maximum pixel value has worse reproducibility ($3\% \pm 11\%$) than does the mean value ($1\% \pm 7\%$), a trend similar to that reported by Krak et al. (12).

The SUV difference between the 2 measurements does not depend on the average value of the replicates when considering SUV_{mean} , whereas the difference increases with increasing SUV when SUV_{max} is considered. Consequently, relative changes in SUV_{mean} are dependent on the average SUV, but the relative change in SUV_{max} is not.

Between successive PET studies, a number of factors other than the natural history of the tumor may cause variability in the measured SUV. These factors include fluctuations in plasma glucose and patient weight, errors in repositioning ROIs or image registration, and variations in the uptake period. Our study demonstrates that in total, these diverse sources of variability contribute less than 0.5 SUV in 95% of repeated studies.

We have previously shown that effective chemotherapy can cause a significant reduction in tumor metabolism within 14 d and that a decrease of 0.5 SUV between studies performed 1 and 3 wk after the initiation of chemotherapy was predictive of those patients who survived more than 6 mo and in whom chemotherapy was presumably successful (7). Our study confirms that decreases in SUV that are larger than 0.5 SUV may be used to define a successful metabolic response to therapy.

CONCLUSION

Repeated measurements of SUV performed a few days apart are reproducible. A decrease of 0.5 SUV is a statistically significant decrease that may be considered when establishing thresholds to predict success of chemotherapy in patients with cancer.

ACKNOWLEDGMENTS

This study was presented, in part, at the Annual Meeting of the Society of Nuclear Medicine, June 2008, New Orleans, Louisiana.

REFERENCES

1. Juweid ME, Cheson BD. Positron-emission tomography and assessment of cancer therapy. *N Engl J Med*. 2006;354:496–507.
2. Weber WA. Use of PET for monitoring cancer therapy and for predicting outcome. *J Nucl Med*. 2005;46:983–995.
3. Kostakoglu L, Coleman M, Leonard JP, Kuji I, Zoe H, Goldsmith SJ. PET predicts prognosis after 1 cycle of chemotherapy in aggressive lymphoma and Hodgkin's disease. *J Nucl Med*. 2002;43:1018–1027.
4. Dose Schwarz J, Bader M, Jenicke L, Hemminger G, Jänicke F, Avril N. Early prediction of response to chemotherapy in metastatic breast cancer using sequential ^{18}F -FDG PET. *J Nucl Med*. 2005;46:1144–1150.
5. Avril N, Sassen S, Schmalfeldt B, et al. Prediction of response to neoadjuvant chemotherapy by sequential F-18-fluorodeoxyglucose positron emission tomography in patients with advanced-stage ovarian cancer. *J Clin Oncol*. 2005;23:7445–7453.
6. Ott K, Fink U, Becker K, et al. Prediction of response to preoperative chemotherapy in gastric carcinoma by metabolic imaging: results of a prospective trial. *J Clin Oncol*. 2003;21:4604–4610.
7. Nahmias C, Hanna WT, Wahl LM, Long MJ, Hubner KF, Townsend DW. Time course of early response to chemotherapy in non-small cell lung cancer patients with ^{18}F -FDG PET/CT. *J Nucl Med*. 2007;48:744–751.
8. Kelloff GJ, Hoffman JM, Johnson B, et al. Progress and promise of FDG-PET imaging for cancer patient management and oncologic drug development. *Clin Cancer Res*. 2005;11:2785–2808.
9. Paquet N, Albert A, Foidart J, Hustinx R. Within-patient variability of ^{18}F -FDG: standardized uptake values in normal tissues. *J Nucl Med*. 2004;45:784–788.
10. Minn H, Zasadny KR, Quint LE, Wahl RL. Lung cancer: reproducibility of quantitative measurements for evaluating 2-[^{18}F]-fluoro-2-deoxy-D-glucose uptake at PET. *Radiology*. 1995;196:167–173.
11. Weber WA, Ziegler SI, Thödtmann R, Hanauske AR, Schwaiger M. Reproducibility of metabolic measurements in malignant tumors using FDG PET. *J Nucl Med*. 1999;40:1771–1777.
12. Krak N, Boellaard R, Hoekstra OS, Twisk JWR, Hoekstra CJ, Lammertsma AA. Effects of ROI definition and reconstruction method on quantitative outcome and applicability in a response monitoring trial. *Eur J Nucl Med Mol Imaging*. 2005;32:294–301.
13. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986;8476:307–310.
14. Press WH, Teukolsky SA, Vetterling WT, Flannery BP. *Numerical Recipes in C: The Art of Scientific Computing*. 2nd ed. Cambridge, U.K.: Cambridge University Press; 1992:620–628.
15. Hoekstra CJ, Paglianiti I, Hoekstra OS, et al. Monitoring response to therapy in cancer using [^{18}F]-2-fluoro-2-deoxy-D-glucose and positron emission tomography: an overview of different analytical methods. *Eur J Nucl Med*. 2000;27:731–743.
16. Lodge MA, Lucas JD, Marsden PK, Cronin BF, O'Doherty MJA. PET study of ^{18}F FDG uptake in soft tissue masses. *Eur J Nucl Med*. 1999;26:22–30.
17. Hustinx R, Smith RJ, Benard F, et al. Dual time point fluorine-18 fluorodeoxyglucose positron emission tomography: a potential method to differentiate malignancy from inflammation and normal tissue in the head and neck. *Eur J Nucl Med*. 1999;26:1345–1348.
18. Mavi A, Urhan M, Yu JQ, et al. Dual time point ^{18}F -FDG PET imaging detects breast cancer with high sensitivity and correlates well with histologic subtypes. *J Nucl Med*. 2006;47:1440–1446.
19. Huang SC. Anatomy of SUV. *Nucl Med Biol*. 2000;27:643–646.
20. Chen ET, Nichols JH, Duh SH, Hortin G. Performance evaluation of blood glucose monitoring devices. *Diabetes Technol Ther*. 2003;5:749–768.