
Detection and Characterization of Tumor Changes in ^{18}F -FDG PET Patient Monitoring Using Parametric Imaging

Hatem Necib¹, Camilo Garcia², Antoine Wagner², Bruno Vanderlinden², Patrick Emonts³, Alain Hendlisz⁴, Patrick Flamen², and Irène Buvat¹

¹IMNC UMR 8165 CNRS, Paris 7 and Paris 11 Universities, Orsay, France; ²Department of Nuclear Medicine, Institut Jules Bordet, Université Libre de Bruxelles, Brussels, Belgium; ³Department of Radiology, Institut Jules Bordet, Université Libre de Bruxelles, Brussels, Belgium; and ⁴Department of Gastroenterology, Institut Jules Bordet, Université Libre de Bruxelles, Brussels, Belgium

In PET-based patient monitoring, metabolic tumor changes occurring between PET scans are most often assessed visually or by measuring only a few parameters (tumor volume or uptake), neglecting most of the image content. We propose and evaluate a parametric imaging (PI) method to assess tumor changes at the voxel level. **Methods:** Seventy-eight pairs of tumor images obtained from baseline and follow-up ^{18}F -FDG PET/CT for 28 patients with metastatic colorectal cancer were considered. For each pair, after CT-based registration of the PET volumes, the 2 PET datasets were subtracted. A biparametric graph of subtracted voxel values versus voxel values in the first PET scan was obtained. A model-based analysis of this graph was used to identify the tumor voxels in which significant changes occurred between the 2 scans and yielded indices characterizing these changes. The Response Evaluation Criteria in Solid Tumors (RECIST) based on the CT images obtained 5–8 wk after the second PET/CT scan were used to classify tumor masses as responding or progressive. On the basis of this classification, we compared the sensitivity and specificity of PI and an approach based on recommendations from the European Organization for Research and Treatment of Cancer (EORTC). **Results:** For tumor-based classification, the EORTC-based approach had a sensitivity and specificity of 85% and 52%, respectively, for detecting responding lesions, whereas PI had a sensitivity and specificity of 100% and 53%, respectively. None of responding tumors using RECIST was classified as progressive with the PI or EORTC-based criteria. Among the 14 progressive lesions according to RECIST, 12 were identified as progressive with PI whereas EORTC-based criteria classified only 1 as progressive and 13 as stable tumors. Considering the patient-based classification, none of the responders according to RECIST was classified as having progressive disease with the PI and EORTC-based criteria. PI has the advantage of showing a parametric image of the patient response to therapy, indicating potential heterogeneity in tumor response. **Conclusion:** The PI method has been successfully applied to characterize early metabolic tumor changes in 78 lesions from ^{18}F -FDG PET/CT scans of patients with metastatic colorectal cancer during chemotherapy. The PI findings correlated well with the standard RECIST-based response assessment.

Key Words: PET/CT; response monitoring; ^{18}F -FDG; advanced colorectal cancer

J Nucl Med 2011; 52:354–361

DOI: 10.2967/jnumed.110.080150

PET plays an increasing role in patient monitoring (1). Because metabolic changes always precede the morphologic changes detectable through conventional anatomic imaging such as CT or MRI, PET is expected to enable early assessment of response to therapy. Several studies have already reported earlier detection of tumor response or tumor recurrence with ^{18}F -FDG PET than with CT (2–5).

Changes in tumor uptake occurring between ^{18}F -FDG PET scans acquired before and during the course of therapy are most often assessed visually. This visual assessment is frequently associated with a quantitative characterization of the tumor uptake in each scan, by measuring the standardized uptake value (SUV) in each tumor mass. Recommendations about how to interpret SUV changes between scans have been published (4,6). These recommendations are based on using cutoff values of the SUV changes to classify tumor changes as complete response (CR), partial response (PR), stable disease, or progressive disease. Although SUV estimates are useful to objectively assess the change in tumor uptake between 2 scans, they suffer from limitations. First, several studies have shown that SUVs are highly dependent on several parameters, including the delay between injection and scan time (7), noise level and spatial resolution in the reconstructed images, and region used to estimate the SUV (8). Second, SUV provides either an average estimate of the tumor uptake, when calculated in a volume of interest (VOI) drawn around the tumor, or a local estimate when considering only the maximum voxel value in the tumor (maximum SUV [SUV_{max}]). As a result, complex changes in tumor uptake, for instance, heterogeneous tumor response or development of a necrotic area, can be overlooked when considering SUV changes only. The recommendations regarding the interpretation

Received Jun. 12, 2010; revision accepted Dec. 2, 2010.
For correspondence or reprints contact: Irène Buvat, IMNC, Campus d'Orsay, Bâtiment 440, 91406 Orsay Cedex, France.
E-mail: buvat@imnc.in2p3.fr
COPYRIGHT © 2011 by the Society of Nuclear Medicine, Inc.

of SUV changes between PET scans are thus of limited value (5). New criteria for classifying tumor changes in solid tumors based on PET/CT have been recently suggested (9) but still need to be widely assessed.

In this work, we propose a parametric imaging (PI) method to monitor tumor changes while overcoming some limitations of the current approaches. First, the method does not need a precise identification of the tumors to be analyzed. Second, it provides a view of the changes of tumor tracer uptake at the voxel level, showing evidence of intratumoral response heterogeneity if present. The performance of the method is studied using a set of 28 patients with metastatic colorectal cancer.

MATERIALS AND METHODS

Patients and Imaging Protocols

Twenty-eight patients (mean age, 62.8 y; age range, 23–83 y) with metastatic colorectal cancer treated at the Institut Jules Bordet, Brussels, Belgium, were considered. All patients were recruited as part of a prospective clinical trial on a larger cohort of patients assessing the clinical role of early ^{18}F -FDG PET/CT in chemotherapy-treated metastatic colorectal cancer (10). The study was approved by the ethics committee of the Institut Bordet and registered at clinicaltrials.gov (NCT00741481). All patients presented advanced metastatic colorectal cancer deemed able to receive chemotherapy. They were treated with 5-fluorouracil, leucovorin, and oxaliplatin (19 patients); folinic acid, fluorouracil, and irinotecan (9 patients); and capecitabine (1 patient) as first- (20) or second-line (9) treatments. The mean number of lesions per patient was 3 (range, 1–8). Seventy-eight lesions in total were analyzed (3 primary, 49 in the liver, 8 in the lungs, 10 in the peritoneum, and 8 in other locations).

In addition, to validate the PET/CT image registration method involved in the PI approach, 6 patients (mean age, 60 y; age range, 49–76 y) with inoperable non-small cell lung cancer (stage III or IV) treated at the Institut Jules Bordet were considered. Unlike the patients from the first group, these patients had 3 PET/CT scans, enabling a detailed evaluation of the registration accuracy that was not possible using 2 PET/CT scans only. These patients were retrospectively recruited and received gemcitabine, ifosfamide, and cisplatin; or cisplatin, docetaxel; or cisplatin, vinorelbine chemotherapy as first-, second-, or third-line treatments. The tumor masses were located in the lung (18), liver (8), mediastinum (3), bone (5), and lymph nodes (5).

CT. Each of the 28 patients had a first helical diagnostic CT scan with or without intravenous contrast injection (depending on the evaluated lesion) 8 d (range, 0–23 d) on average before the first ^{18}F -FDG PET/CT scan and after 5–8 wk on therapy or sooner in the case of clinical suspicion of progression. Axial slice thickness was 3 or 5 mm, depending on the scanner. A lesion-by-lesion analysis was performed. The target lesions were identified by a senior radiologist in a joint reading session with the nuclear medicine physician.

CT findings were interpreted according to Response Evaluation Criteria in Solid Tumors (RECIST) 1.0, with additional restrictions: lesions should be clearly individualized on both baseline PET and diagnostic CT scans and have a minimal diameter of 15 mm on the baseline diagnostic CT scan. Stable disease status was confirmed by an additional CT scan after 6–8 more weeks.

^{18}F -FDG PET/CT. Each patient underwent a baseline ^{18}F -FDG PET/CT scan just before the start of chemotherapy and a second scan at day 14.

Patient preparation, imaging, and reconstruction protocols were kept constant for serial scans of the same patient. All ^{18}F -FDG PET/CT images were obtained using a Discovery LS system (GE Healthcare), 60 min after injection of 4 MBq/kg. PET images were reconstructed with the built-in Advance software (GE Healthcare), using the ordered-subset expectation maximization algorithm (11) (2 iterations, 28 subsets) and gaussian postfiltering of 5.45 mm in full width at half maximum. The images were corrected for attenuation based on the CT scan and for scatter. CT was performed with a 4-slice multidetector helical scanner (Lightspeed; GE Healthcare). The tension was 120 kV, and the current was determined by the Auto-mA algorithm (GE Healthcare) and ranged from 30 to 200 mA. The other parameters were 0.5 s per CT rotation, a pitch of 1.5, and a table speed of 15 mm/rotation. The matrix of CT images was 512×512 (0.98×0.98 mm pixel size) with a 5-mm slice thickness. The PET volumes (128×128 pixels of 3.91×3.91 mm, 4.25-mm slice thickness) and the CT volumes (5-mm slice thickness, 512×512 pixels of 0.98×0.98 mm) were systematically coregistered using the LightSpeedAppscat_dst_dls_1.7_R2.9N. IRIX646.5 software.

Parametric Imaging of Tumor Changes

To compare PET scans acquired before and during antineoplastic therapy, the images were first converted to SUV units by normalizing the measured activity by the injected activity per body weight corrected for radioactive decay. Then, pairs of PET volumes acquired in the same patient were considered. For each patient, a set of slices of interest was manually selected in the baseline scan using a large parallelepiped VOI encompassing the organs including the tumor lesions.

For each pair of volumes, the comparison consisted of registering the 2 PET volumes, calculating a parametric image, and deriving quantitative indices from the parametric image.

Registration of PET Volumes. To compare 2 PET images at a voxel level, these scans first need to be registered so that a given voxel corresponds to the same volume element in each of the 2 scans. Because PET images suffer from modest spatial resolution and are systematically coregistered with CT volumes acquired in the same PET/CT session, the CT volumes were used to determine the transformation appropriate for realigning the PET VOI. The 2 CT VOIs were registered using a rigid transform (3 translations and 3 rotation parameters) derived from block-matching registration (12) as implemented in the Isogray software (Dosisoft). Local rigid transformation was assumed because we considered limited VOI including the tumor masses. The transformation mapping CT volume 2 on CT volume 1 was then used to register the second PET scan with the first PET scan.

Calculation of Parametric Image of Significant Tumor Changes. The 2 registered PET scans, denoted PET1 and PET2, were first subtracted voxel-by-voxel. This yielded a volume $\text{DIFF} = \text{PET2} - \text{PET1}$ in which each voxel value $\text{DIFF}(i)$, where i denotes the voxel, corresponded to the difference in SUV in voxel i between the first and second scans. To identify the voxels i corresponding to actual changes in tumor uptake from this DIFF volume, a biparametric graph of $\text{DIFF}(i)$ (y -axis) against $\text{PET1}(i)$ (x -axis) was plotted (Fig. 1A). The number of points in this graph corresponds to the number of voxels in the analyzed volume (from 38,400 to 147,456 voxels, depending on the patient slice selection). In this graph,

voxels i that did not change much between the 2 scans are near the horizontal axis ($y = \text{DIFF}(i) = 0$), whereas voxels with substantial change have high $|\text{DIFF}(i)|$ values. In addition, tumor voxels will tend to have higher x (PET1) values than voxels corresponding to physiologic uptake. This plot can also be represented as a histogram showing the density of voxels with specific (PET1(i),DIFF(i)) coordinates, in which clouds of voxels can be clearly distinguished (Fig. 1B).

Voxels corresponding to changes in tumor uptake are identified by fitting this graph using a gaussian mixture model (GMM). GMMs have been shown to be useful for identifying clusters in structured data (13). The GMM assumes that the distribution of voxels in the biparametric plot can be expressed as a mixture of gaussian distributions, each characterized by a mean μ_k and a variance matrix V_k . The GMM reads (Supplemental Appendix 1; supplemental materials are available online only at <http://jnm.snmjournals.org>):

$$f(x_i) = \sum_{k=1}^K p_k f(x_i | \mu_k, V_k), \quad \text{Eq. 1}$$

where x_i is the 2-component vector (PET1(i),DIFF(i)), and p_k is the mixing proportion ($0 < p_k < 1$ for all $k = 1, \dots, K$ and $\sum_k p_k = 1$).

A stochastic expectation maximization algorithm was used to calculate the maximum likelihood estimates of the model parameters $\Pi = (p_1, \dots, p_{k-1}, \mu_1, \dots, \mu_k, V_1, \dots, V_k)$ (14). Given these estimates, the voxels were classified into K clusters, where each voxel is assigned to cluster k such that the conditional probability $t_k(x_i)$ to belong to cluster k is maximum (Supplemental Appendix 1). Figure 1C gives an example of identification of 3 clusters in the biparametric representation.

Solving the GMM requires an estimate of the number of clusters, an initialization of the model parameters, and a stopping criterion. Four clusters at most can be expected: 1 for voxels with no substantial SUV changes between the 2 PET scans (noise only), 1 for voxels in which physiologic changes not related to the tumor masses occurred, 1 for tumor voxels in which SUV increased between the 2 scans, and 1 for tumor voxels in which SUV decreased between the 2 scans. The initialization of these clusters is described in Supplemental Appendix 2. Given the initial clusters, the GMM is solved using stochastic expectation maximization (15). The stochastic expectation maximization algorithm is stopped when the absolute change in the loglikelihood is less than 0.001 between 2 successive iterations or after 150

iterations. The consistency of the resulting clusters is checked (Supplemental Appendix 2). In the case of inconsistency, the GMM is solved again using 1 cluster less. This is iterated until the final clusters (4, 3, or 2) are found to be consistent.

On the basis of this classification, a parametric volume is deduced. In this volume, all voxels belonging to the noise or physiologic changes cluster are set to zero. Only the voxels i belonging to the other clusters (increase or decrease in SUV) are assigned their $\text{DIFF}(i) = \text{PET2}(i) - \text{PET1}(i)$ values. To facilitate visual interpretation, negative values are coded on a green color scale, and positive values are coded on a red color scale.

Calculation of Quantitative Indices from Parametric Volume. For each cluster of voxels seen in the parametric volume, 2 indices were calculated. The tumor volume ΔV affected by changes in SUV was equal to the largest number of connected voxels in the cluster. The average change in SUV, ΔSUV , was the average $\text{PET2}(i) - \text{PET1}(i)$ in these connected voxels.

Each cluster of voxels corresponded to either a partial metabolic response (PMR, decrease in the cluster mean SUV) or a progressive or nonresponding lesion (progressive metabolic lesion [PML], increase in the cluster mean SUV). When there were multiple clusters within a lesion, if at least 1 cluster was progressive, the lesion was classified as PML. By definition, no stable tumor could be seen in the parametric volume.

Using the parametric volume, a patient was classified as having overall metabolic response if at least half of the lesions observed in the baseline PET scan showed a metabolic response, without any progressive or new lesion. Otherwise, the patient was classified as a nonresponder.

Evaluation of Image Registration Accuracy

The relevance of PI might strongly depend on the accuracy with which PET scans acquired weeks apart can be registered. To assess the registration accuracy, consistency measurements were applied (16) using the 6 patients with 3 PET/CT scans. This method consisted of registering 3 CT scans in a row and checking the consistency of the parameters produced by registration of each of the 3 scan pairs.

Evaluation of PI

After registration, PI was used to classify the 78 masses as responding or progressive tumor. Patients were also classified as responder or nonresponder.

Because no method for creating parametric images of tumor changes has been described for PET, we compared PI with a

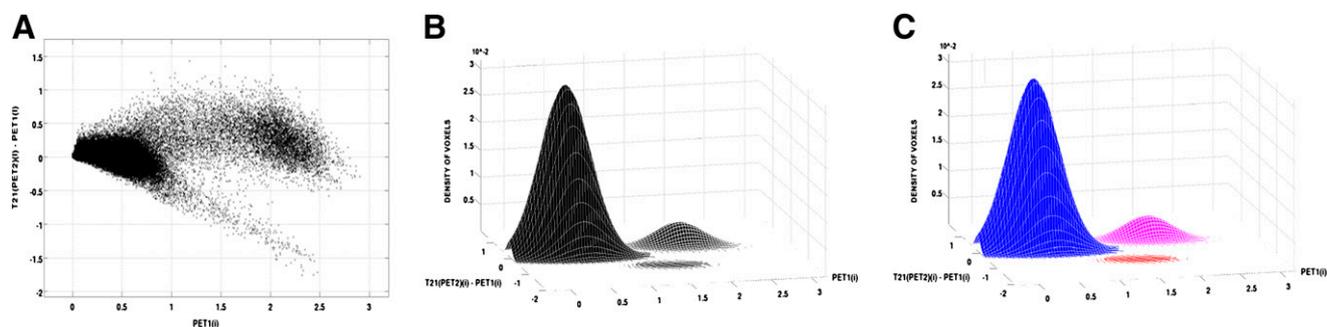


FIGURE 1. Patient with tumor in liver. (A) Biparametric graph of $\text{DIFF}(i)$ against $\text{PET1}(i)$: each point in graph corresponds to voxel i . (B) Histogram showing density of voxels with specific (PET1(i), $\text{DIFF}(i)$) coordinates. (C) Biparametric graph of $\text{DIFF}(i)$ against $\text{PET1}(i)$ (same as graph B) showing identification of GMM parameters with 3 clusters. Blue = voxels affected by changes in noise only; pink = voxels corresponding to physiologic uptake (liver); red = tumor voxels with substantial SUV changes.

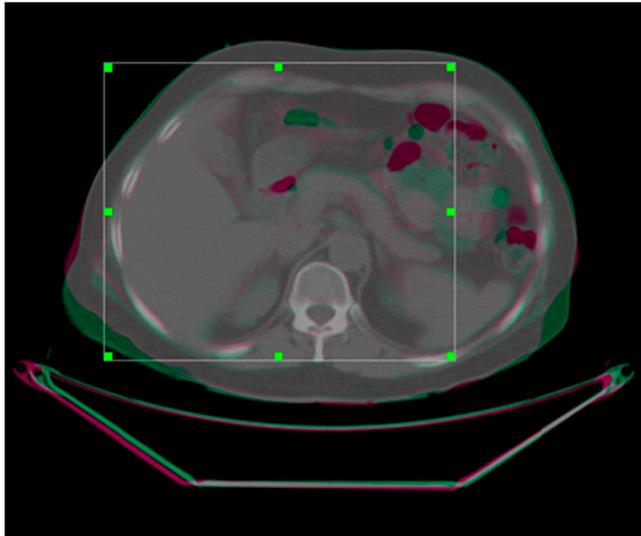


FIGURE 2. Slice of CT1 (green) superimposed with CT2 (red) after coregistration of 2 CT acquired 3 wk apart. Box shows region of interest considered for image registration.

regional approach for longitudinal study based on the recommendations by the European Organization for Research and Treatment of Cancer (EORTC) (6). The tumor VOIs were defined manually on the baseline PET scan and then positioned on the second scan as close to the original tumor volume as possible. Uptake was measured using the SUVmax in the tumor region. For each tumor, after 1 cycle of chemotherapy, the so-called EORTC tumor response was assessed as follows: PML, SUVmax increase of at least 25%; stable metabolic lesion, SUVmax change between -15% and $+25\%$; PMR, SUVmax reduction of a minimum of 15%; and complete metabolic response, complete resolution, indistinguishable from surrounding normal tissue.

PML and stable metabolic lesions were considered nonresponding lesions, whereas PMR and complete metabolic responses were considered responding lesions.

We also classified tumors by considering only the change in SUVmax, without setting any cutoff value, as follows: PML corresponded to lesions with an SUVmax increase ($\Delta\text{SUVmax} > 0$), whereas PMR corresponded to lesions with an SUVmax decrease. These classification rules are named SUVmaxSIGN.

Patient classification based on the EORTC criteria includes 2 groups only. A patient was classified as having overall metabolic responsive disease if most of the lesions observed on the baseline

PET scan showed a metabolic response, without any progressive lesion (PL) (new or $\geq 25\%$ increase of SUVmax). Otherwise, the patient was classified as a nonresponder. The same rules were applied to categorize patients based on the SUVmaxSIGN lesion classification.

To estimate the sensitivity and specificity of early response assessment using the PI, SUVmaxSIGN, and EORTC-based approaches, we used RECIST 1.0 derived from the late CT images to obtain the gold standard classification (17), including CR, PR, PL, and stable lesion.

CRs and PRs were considered responding lesions, whereas PLs and stable lesions were considered nonresponding lesions.

On the basis of the patient's tumor classification, the patient was categorized as a responder if at least half of the tumors were CR or PR without any PL. Otherwise, the patient was called a nonresponder.

PI Robustness to Registration Accuracy

As a preliminary assessment of the robustness of PI with respect to the accuracy of image registration, we considered 12 baseline PET scans that included 12 tumors from the colorectal cancer patients. Each PET volume was misaligned randomly in the 3 directions so that the magnitude of the displacement was equal to 1, 2, and 3 times the mean registration error (MRE) found using the consistency criterion. We then applied PI, assuming that the misaligned and initial (before misalignment) volumes were the first and second PET scans to be compared. This approach mimicked the comparison of the same PET volumes (i.e., no metabolic change), differing only because of incorrect registration.

RESULTS

Registration Accuracy

In the 6 lung cancer patients, volumes consisting of 31 slices (15.5 cm) were registered. No obvious misregistration artifacts were seen by overlapping the registered CT images using 2 different color scales (Fig. 2). The consistency measurements led to an MRE ($\pm\text{SD}$) of 1.60 ± 0.67 mm, which is, on average, less than the PET voxel size ($3.91 \times 3.91 \times 4.25$ mm).

Detection of Tumor Changes

In the 28 patients, volumes ranging from 9 to 26 slices (mean, 18 slices) were processed by PI. Figure 3 shows an example of a parametric image for a slice including 5 tumors and the results of the quantitative analysis. The results from the EORTC-based analysis are also shown.

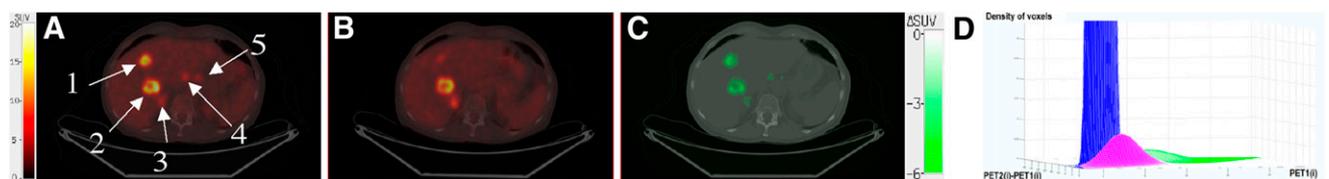


FIGURE 3. (A) PET1 showing 5 tumors, superimposed with CT1. (B) PET2 superimposed with CT2. (C) Parametric image (superimposed with CT1) showing only voxels with significant tumor changes between PET1 and PET2. These voxels are shown in green, meaning that SUV decreased between the 2 scans. For the 2 biggest tumors, the EORTC-based approach found a responding lesion (SUV decrease of 27% for tumor 1) and a stable lesion (SUV decrease of 10% for tumor 2). PI found 2 responding lesions ($\Delta\text{SUV} = -5.9$ and -2.6 for tumors 1 and 2, respectively), which were consistent with RECIST classification derived from late CT. (D) Biparametric graph fitted by GMM, for which 3 clusters can be distinguished: noise (blue), physiologic changes (pink), and tumor changes (green).

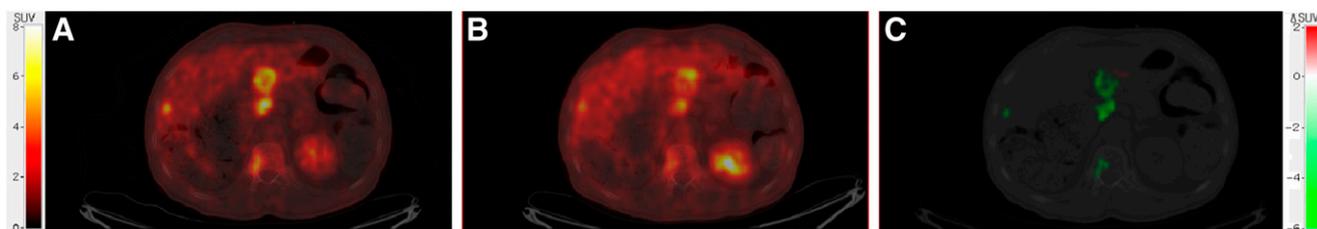


FIGURE 4. (A) PET1 superimposed with CT1. (B) PET2 superimposed with CT2. (C) Parametric image (superimposed with CT1) showing tumor with heterogeneous response.

Figure 4 shows another example in which a tumor had a heterogeneous evolution. Most of the tumor region decreased, except for a small region, which appeared in red in the parametric image. The green region had a ΔSUV of -3.5 , and the red region had a ΔSUV of $+1.4$. The tumor was characterized by an SUV decrease of 22.7% using the EORTC-based approach. The RECIST later classified this lesion as progressive disease.

Table 1 compares the lesion classification obtained using RECIST and the PET-based methods (EORTC-based, SUVmaxSIGN, and PI). None of the PRs seen by RECIST was classified as PML, except 1 with SUVmaxSIGN. RECIST did not find any CR, consistent with PI and SUVmaxSIGN, whereas the EORTC-based approach found 4 complete metabolic responses. Twelve of the 14 lesions seen as PLs with RECIST were also PMLs with PI, whereas EORTC classified only 1 of them as PML and SUVmaxSIGN classified 7.

All 44 lesions classified as stable using RECIST were seen with PI either as PMR (25 tumors) or as PML (19 tumors). All clusters of voxels in PI corresponded to tumor regions.

Considering the RECIST classification as a standard, the EORTC-based approach had a sensitivity and specificity of 85% and 52%, respectively, for detecting responding lesions, the SUVmaxSIGN had sensitivity and specificity of 95% and 21%, respectively, and PI had sensitivity and specificity of 100% and 53%, respectively. These values corresponded to positive predictive values and negative predictive values of 38% and 91%, respectively, for the EORTC criteria, 29% and 92%, respectively, for SUVmaxSIGN, and 43% and 100%, respectively, for PI.

Table 2 compares the classification of the 28 patients between responders and nonresponders obtained using RECIST, EORTC, SUVmaxSIGN, and PI. None of the

patients seen as responders by RECIST was classified as having progressive disease with PI, SUVmaxSIGN, and the EORTC-based approaches. Only 6 patients were classified differently with PI, compared with RECIST, whereas 8 patients were classified differently between EORTC and RECIST and 10 between SUVmaxSIGN and RECIST. All these differences in classification corresponded to patients seen as nonresponders by RECIST whereas PET classified them as responders.

PI Robustness to Registration Accuracy

When analyzing the faked PET volume pairs in identical PET scans differing only by misregistration, the PI volume was always empty when the displacement magnitude was 1MRE ($=1.6$ mm). It was empty in 7 of 12 cases when the displacement magnitude was 2MRE and never empty when the displacement magnitude was 3MRE. For displacement magnitudes of 2MRE, when the parametric image was not empty, the mean ΔV of the clusters seen in PI was 70 voxels ± 30 , whereas it was 282 ± 293 in the 78 clusters seen in the real analysis of the 28 patients. The artifacts in the PI due to misregistration could be clearly identified as symmetric responding and nonresponding regions in the parametric image (Fig. 5).

DISCUSSION

The interpretation of serial ^{18}F -FDG PET/CT scans in treatment monitoring is most often based on the visual assessment of tumor mass tracer uptake, and measurements of SUV, but there is no consensus about the criteria to translate SUV changes into patient classification (9). When comparing scans acquired from the same patient, summarizing the complexity of the uptake distribution by only an

TABLE 1
Classification of Tumor Masses

RECIST	Complete metabolic response			PMR			Stable metabolic lesion			PML		
	EORTC	SUVmaxSIGN	PI	EORTC	SUVmaxSIGN	PI	EORTC	SUVmaxSIGN	PI	EORTC	SUVmax SIGN	PI
PR = 20	2	0	0	15	19	20	3	0	0	0	1	0
Stable lesion = 44	2	0	0	23	39	25	14	0	0	5	5	19
PL = 14	0	0	0	3	7	2	10	0	0	1	7	12

TABLE 2
Classification of Patients as Responders and Nonresponders

RECIST classification	EORTC		SUVmaxSIGN		PI	
	Responder	Nonresponder	Responder	Nonresponder	Responder	Nonresponder
Responders ($n = 6$)	6	0	6	0	6	0
Nonresponders ($n = 22$)	8	14	10	12	6	16
Total ($n = 28$)	14	14	16	12	12	16

SUV per tumor is extremely restrictive. This observation motivated the development of PI, which considers all changes in uptake between the 2 PET scans to be compared and then identifies those related to tumor masses. The approach is similar to the subtraction of ictal SPECT coregistered to MRI used to localize the epileptic zone from ictal and interictal SPECT scans (18). In our application, in which the PET scans to be compared are acquired weeks apart, a first step was to determine whether accurate registration of the 2 PET scans was possible.

We considered rigid registration of the CT scans already aligned with the PET scans. Rigid registration was used to avoid any uncontrolled distortion at the sites of the tumor masses through elastic deformations. To make rigid registration appropriate, it was applied to a VOI defined as a set of slices. This selection does not involve any precise delineation of a region and is fast. To assess the registration accuracy, we validated the approach based on consistency measurements (16) using a set of lung cancer patients for which 3 scans of each patient were available. The scans of these lung cancer patients were acquired using the same imaging and processing protocol as that used for patients with colorectal cancer. We thus hypothesized that if the registration procedure was accurate enough for these lung cancer patients, it would also be reliable for the 28 patients with colorectal cancer. The registration error between the CT scans (1.6 mm) was, on average, less than the PET voxel size ($3.91 \times 3.91 \times 4.25$ mm). Even when accounting for the variability of the error (SD, 0.67 mm) and assuming a normal distribution of the error, a 95% confidence interval still does not encompass a PET voxel, suggesting that in most cases, registration errors are less than 1 voxel in the PET VOI. This was confirmed visually on the 28 pairs of scans that were registered in the colorectal cancer patients

(results not shown). Accurate local registration of PET scans acquired several weeks apart is thus feasible.

The PI approach is based on the subtraction of the 2 PET scans to be compared, each expressed in SUV, and on the interpretation of the subtracted values in light of the SUVs in the baseline PET scan through a biparametric graph. An automatic procedure identifies the clusters of voxels belonging to tumor masses. No voxel contiguity criterion was used when generating the biparametric graph and solving the GMM. Yet, the resulting clusters were spatially consistent: each cluster corresponded to one or several distinct groups of adjacent voxels that were associated with well-identified structures (Figs. 3 and 4). The method does not require the setting of some cutoff in changes in SUV, below which the change would not be considered significant. The identification of the voxels with significant changes is based only on the simultaneous analysis of all voxels in the biparametric graph using the GMM. Solving this model involved setting the number of clusters to be identified. Given the limited number of possible clusters (2–4 at the most), an iterative search automatically determined a meaningful solution. We checked that the final classification was robust with respect to the empiric initialization of the clusters (results not shown).

When applied to patient data, the method always yielded parametric volumes that were visually consistent. The clusters with significant change in SUV identified by PI always corresponded to tumor masses in which changes could be visually identified. The parametric volume also shows heterogeneous responses within tumor sites (Fig. 4) that are not described when considering a single SUV per tumor. Whether such heterogeneity brings useful information about the tumor response and final outcome will need further investigation. A limitation of the method is that it does not

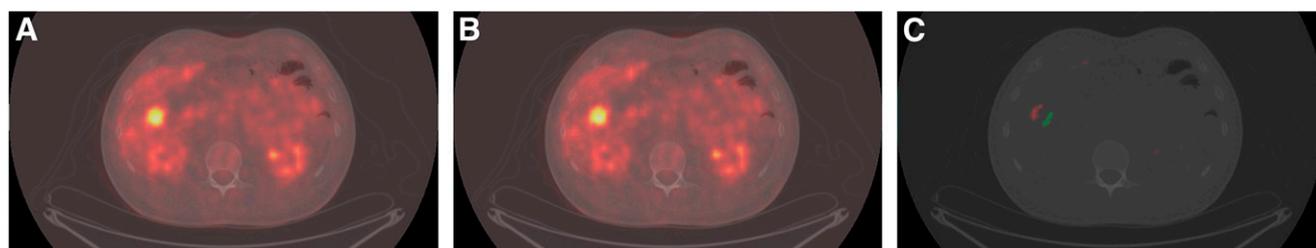


FIGURE 5. (A) PET1 superimposed with CT1. (B) Baseline PET/CT misaligned by 4.8 mm (faked PET2). (C) Parametric image (superimposed with CT1) showing artifacts due to imperfect image registration.

identify any stable tumor mass (in which SUV has not changed between the 2 scans). The method does not identify CRs either but classifies them as responding tumors. Finally, if a tumor appears in a region that is not part of the volumes selected to perform PI, it will be missed. Thus, the parametric volume should not be interpreted without viewing the 2 PET scans independently and should be regarded as an aid to quickly visualize the changes between the 2 PET scans and automatically get quantitative information regarding these changes.

In our patient population, none of the tumor sites classified as a PR using RECIST was seen as a progressive tumor in the early PET scan using PI. Any tumor seen as progressive in the early PET scan was later confirmed as resistant to the treatment by RECIST. The same conclusion could also be drawn when interpreting the early PET scans based on the EORTC criteria. However, the PI approach was more consistent with the RECIST classification regarding the identification of the PLs: among the 14 tumor masses seen as progressive by the late CT, 12 were classified as progressive using PI on the early PET scan, whereas only 1 using the EORTC-based criteria and 7 using SUVmaxSIGN were classified as progressive. The missed PLs with PI or SUVmaxSIGN were classified as PR, whereas the 13 missed lesions using the EORTC approach were classified as stable disease (10) or PR (3). In our study, 25 of 78 tumor masses were classified as stable disease with RECIST but PR with PI. This classification might suggest that PI is more sensitive to the changes due to therapy than RECIST, although we cannot prove it from this single study. Overall, the results obtained using PI were extremely promising in terms of sensitivity of detection of responding tumor masses (100% vs. 85% for the EORTC-based criterion). The specificity was lower (53% with PI and 52% with the EORTC-based criterion). This might not be due to the PET image analysis itself but rather to the inability of the early PET scan to detect all nonresponding patients. Another hypothesis may be an initial response detected by the early PET scan, with a rapid escape from chemotherapy detected by the late diagnostic CT scan. The negative predictive value of the early PET scan was high: with PI, all lesions identified as nonresponding were confirmed as nonresponding with the late CT scan versus 91% of the lesions identified as nonresponding with the EORTC-based criteria and 92% with SUVmaxSIGN.

Regarding the patient classification, the PI-, SUVmaxSIGN-, and EORTC-based approaches yielded the same classification for 21 of 28 patients. The 7 other patients were all classified as nonresponding with RECIST. Five of them were classified as nonresponding with PI, but 2 were classified as responding. Two were classified as nonresponding and 5 as responding with the EORTC criteria, whereas all 7 were classified as responding with SUVmaxSIGN. The agreement between the early PET scan and RECIST from the late CT scan was thus slightly more consistent when the early PET scan was analyzed with PI than when the analysis was based on the EORTC or SUVmaxSIGN criteria.

In our patients, when repeating the PI analysis starting from a different PET VOI, the repeatability coefficient of Δ SUV was 8.3%, with an intraclass correlation coefficient of 0.99 (95% confidence interval of 0.99–1), suggesting a high repeatability. The tumor and patient classifications (Tables 1 and 2) remained unchanged.

A preliminary study about the possible impact of small misregistration of the serial PET scans on PI was conducted. The results suggest that the method is robust with respect to realistic registration errors (i.e., within the mean \pm 2 SD of the MRE). In addition, artifacts in the PI images due to misregistration might be visually recognizable (Fig. 5) on the basis of the presence of responding and nonresponding clusters at almost symmetric positions with respect to the initial tumor location and the small volume of these clusters. We also found that the interpolation scheme (nearest-neighbor interpolation, trilinear, B-splines of the third or fourth order) used in the registration procedure did not modify at all the cluster volumes in the parametric image and only slightly modified the Δ SUVs, always by less than 0.5 SUV (results not shown). A comprehensive study of the robustness of the method in various situations was beyond the scope of this study.

CONCLUSION

We proposed a PI approach to compare serial 18 F-FDG PET scans in the context of patient monitoring. This approach yielded images of the tumor foci that had changed between PET scans, possibly highlighting heterogeneous tumor response, and provided quantitative parameters characterizing the tumor changes. The PI method was successfully applied to characterize early metabolic changes of 78 lesions from 18 F-FDG PET/CT scans of patients with metastatic colorectal cancer. The PI findings correlated well with the standard RECIST-based response assessment obtained 5–8 wk after the early follow-up PET scan.

REFERENCES

1. Ben-Haim S, Eil P. 18 F-FDG PET and PET/CT in the evaluation of cancer treatment response. *J Nucl Med*. 2009;50:88–99.
2. Ichiya Y, Kuwabara Y, Sasaki M, et al. A clinical evaluation of FDG-PET to assess the response in radiation therapy for bronchogenic carcinoma. *Ann Nucl Med*. 1996;10:193–200.
3. Choi N, Fischman A, Niemierko A, et al. Dose-response relationship between probability of pathologic tumor control and glucose metabolic rate measured with FDG PET after preoperative chemoradiotherapy in locally advanced non-small-cell. *Int J Radiat Oncol Biol Phys*. 2002;54:1024–1035.
4. Mac Manus M, Hicks R, Matthews J, et al. Positron emission tomography is superior to computed tomography scanning for response-assessment after radical radiotherapy or chemoradiotherapy in patients with non-small-cell lung cancer. *J Clin Oncol*. 2003;21:1285–1292.
5. Weber WA. Use of PET for monitoring cancer therapy and for predicting outcome. *J Nucl Med*. 2005;46:983–995.
6. Young H, Baum R, Cremerius U, et al. Measurement of clinical and subclinical tumour response using [18 F]-fluorodeoxyglucose and positron emission tomography: review and 1999 EORTC recommendations. *Eur J Cancer*. 1999;35:1773–1782.
7. Stahl A, Ott K, Schwaiger M, Weber WA. Comparison of different SUV-based methods for monitoring cytotoxic therapy with FDG PET. *Eur J Nucl Med Mol Imaging*. 2004;31:1471–1478.

8. Boellaard R, Krak NC, Hoekstra OS, Lammertsma AA. Effects of noise, image resolution, and ROI definition on the accuracy of standard uptake values: a simulation study. *J Nucl Med.* 2004;45:1519–1527.
9. Wahl R, Jacene H, Kasamon Y, Lodge M. From RECIST to PERCIST: evolving considerations for PET response criteria in solid tumors. *J Nucl Med.* 2009;50 (suppl 1):122S–150S.
10. Hendlisz A, Emonts P, Covas A, et al. Is positron emission tomography (PET) with FDG an early predictor of the RECIST morphological response to chemotherapy in metastatic colorectal cancer patients (mCRC) [abstract]? *J Clin Oncol.* 2009;27(suppl):15s.
11. Hudson H, Larkin R. Accelerated image reconstruction using ordered subsets of projection data. *IEEE Trans Med Imaging.* 1994;13:601–609.
12. Ourselin S, Roche A, Prima S, Ayache N. Blockmatching: a general framework to improve robustness of rigid registration of medical images. In: Delp SL, DiGioia AM, Jaramaz B, eds. *Medical Image Computing and Computer-Assisted Intervention.* Berlin, Germany: Springer-Verlag; 2000:557–566.
13. Celeux G. Gaussian parsimonious clustering models. *Pattern Recognit.* 1995;28: 781–793.
14. Celeux G, Diebolt J. The SEM algorithm: a probabilistic teacher algorithm derived from the EM algorithm for the mixture problem. *Comput Statist Quart.* 1985;2:73–82.
15. Biernacki C, Celeux G, Govaert G, Langrognet F. Model-based cluster and discriminant analysis with the MIXMOD software. *Comput Stat Data Anal.* 2006;51:587–600.
16. Holden M, Hill DLG, Denton ERE, Jarosz JM, Cox TCS, Hawkes DJ. Voxel similarity measures for 3D serial MR brain image registration. *IEEE Trans Med Imaging.* 2000;19:94–102.
17. Therasse P, Arbuck SG, Eisenhauer EA, et al. New guidelines to evaluate the response to treatment in solid tumors. European Organization for Research and Treatment of Cancer, National Cancer Institute of the United States, National Cancer Institute of Canada. *J Natl Cancer Inst.* 2000;92:205–216.
18. Zubal IG, Spencer SS, Imam K, Smith EO, Wisniewski G, Hoffer PB. Difference images calculated from ictal and interictal Technetium-99m-HMPAO SPECT scans of epilepsy. *J Nucl Med.* 1995;36:684–689.