

Impact of tumor size and tracer uptake heterogeneity in ¹⁸F-FDG PET and CT Non–Small Cell Lung Cancer tumor delineation

Mathieu Hatt¹, Catherine Cheze-le Rest¹, Angela van Baardwijk³, Philippe Lambin³,
Olivier Pradier^{1,2}, Dimitris Visvikis¹

¹ INSERM, U650 LaTIM, CHRU Morvan, Brest, France

² Department of Radiotherapy, CHRU Morvan, Brest, France

³ MAASTRicht Radiation Oncology Clinic, Maastricht, the Netherlands

Running title: ¹⁸F-FDG PET and CT tumor delineation in NSCLC

Corresponding author:

Mathieu HATT

LaTIM, INSERM U650

CHU MORVAN

5 avenue Foch

29609 Brest

France

Tel.:+33298018111

Wordcount: 5981

Keywords: NSCLC, ¹⁸F-FDG, tumor delineation, tumor volumes, tumor size, uptake heterogeneity

ABSTRACT

The objectives of this study were to investigate the relationship between CT and 18F-FDG PET based tumor volumes in NSCLC and the impact of tumor size and uptake heterogeneity on various PET uptake delineation approaches.

Methods: 25 NSCLC cancer patients with 18F-FDG PET/CT were considered. 17 underwent surgical resection of their tumor and the maximum diameter was measured. Two observers performed manual delineation of the tumors on the CT images and delineated the tumor uptake on the corresponding PET images manually, using a fixed threshold at 50% of the maximum (T_{50}), an adaptive threshold methodology, and the Fuzzy Locally Adaptive Bayesian (FLAB) algorithm. Maximum diameters of the delineated volumes were compared to the histopathology reference when available. The volumes of the tumors were compared, and correlation between anatomical volume, PET uptake heterogeneity and the differences between delineations were investigated.

Results: All maximum diameters measured on PET and CT images significantly correlated with histopathology reference ($r>0.89$, $p<0.0001$). Significant differences were observed among the approaches: CT delineation resulted in large over evaluation ($+32\pm37\%$), whereas all delineations on PET images resulted in under evaluation (from $-15\pm17\%$ for T_{50} , to $-4\pm8\%$ for FLAB) except manual delineation ($+8\pm17\%$). Overall, CT volumes were significantly larger than PET volumes ($55\pm74\text{cm}^3$ for CT vs. from 18 ± 25 to $47\pm76\text{cm}^3$ for PET). A significant correlation was found between the anatomical tumor size and the heterogeneity (larger lesions are more heterogeneous). Finally, the more heterogeneous was the tumor uptake, the larger the under estimation of the PET volumes by threshold-based techniques.

Conclusion: Volumes based on CT were larger than those defined on PET images. Tumor size and tracer uptake heterogeneity have an impact on threshold-based methods that should not be used for the delineation of large heterogeneous NSCLC, as they tend to largely underestimate the functional tumor's spatial extent in such cases. For an accurate delineation of PET volumes in NSCLC, advanced image segmentation algorithms able to deal with tracer uptake heterogeneity should be preferred.

Introduction

Positron Emission Tomography (PET) using 2- [¹⁸F] fluoro-2-deoxy-D-glucose (¹⁸F-FDG), associated with Computed Tomography (CT) since the development of PET/CT devices, has been increasingly used for staging Non-small Cell Lung Cancer (NSCLC) (1). In addition, the use of ¹⁸F-FDG PET/CT in radiotherapy treatment planning for the definition of Gross Tumor Volume (GTV) has been similarly growing (2). Manual contouring of the tumor boundaries on the CT images is still the conventional methodology for target volume definition. On the other hand and despite a high spatial resolution, the delineation on CT alone may be biased by insufficient contrast between tumor and healthy tissues (e.g. in case of atelectasis, pleural effusion and fibrosis, or for tumors attached to the chest wall or mediastinum). Several studies have investigated the impact of delineation carried out on fused ¹⁸F-FDG PET/CT images, demonstrating significant modifications of the treatment plan (size, location or shape of the GTV) (3) as well as reduced inter- and intra-observer variability (4). Additional benefits from the use of PET imaging relative to CT include the potential to image cellular proliferation and tumor hypoxia using tracers such as ¹⁸F-FLT and ¹⁸F-FMISO or ⁶⁴Cu-ATSM respectively.

However, the integration of PET within the radiotherapy planning is complex, especially because there is neither consensus or guidelines regarding the delineation of ¹⁸F-FDG PET tumor uptake and/or how to subsequently use the delineated functional volumes. Most of the previously published studies have investigated the use of a specific threshold of the PET uptake to define the metabolically active tumor volume (MATV, defined as the tumor volume that can be seen and delineated on an ¹⁸F-FDG PET image) or spatial extent, with a large variability in the recommended threshold and resulting volumes (5-8). A few recent studies have investigated the

correlation between tumor histopathology measurements and the threshold of PET uptake (4, 9-12). For example, the study of Yu et al (12) on 15 patients proposed an “optimal” threshold of $31\% \pm 11\%$ of the PET maximum uptake within the tumor, for a good correlation with the corresponding histopathology derived tumor maximum diameter. Considering 3D reconstructed histopathology volumes instead of only the maximum diameter, Stroom et al (10) recommended a fixed threshold of 42% of the maximum PET uptake based on their findings in a group of 5 patients with rather small tumors. Finally, in the study by Wu et al (11) on 31 patients, a threshold of 50% of the maximum was proposed as the best threshold for PET uptake delineation in NSCLC with respect to the histopathological maximum diameter. This conclusion was reached by comparing the results obtained using a range of different fixed threshold values (from 20% to 55%), although only non-statistically significant differences were found with the other tested values. The same authors have subsequently shown that such a threshold was less appropriate than manual delineation which led to incorrect delineation in some cases (13). Manual contouring is however far from ideal, as it suffers from large intra and inter-observer variability (14), while it is also a tedious and time consuming procedure, especially in 3D.

Alternatively, other authors have considered the use of adaptive thresholding approaches taking into account the tumor-to-background ratio instead of fixed threshold but requiring the determination of a background region of interest as well as optimization for a given scanner model, acquisition protocol and image reconstruction using phantom acquisitions (8, 15-16). Using such an approach, van Baardwijk et al (4) obtained a significant correlation with histopathology measurements for 23 NSCLC tumors, as well as reduced inter observer variability. Finally, the use of more advanced image segmentation methodologies to

automatically delineate MATV has been proposed in several studies (17-24), with variable levels of validation. For example, we have already demonstrated that such automated image segmentation approaches can offer higher accuracy (18, 21), robustness (25) and reproducibility (14) than threshold-based (fixed or adaptive) methods.

Some of the previous studies investigating NSCLC tumor delineation on PET/CT hypothesized a significant influence of the anatomical and/or metabolic lesion size and activity distribution heterogeneity on both the results and the observed differences between delineation methodologies (8). However they neither quantified this heterogeneity nor thoroughly investigated such correlation with respect to the anatomical tumor and functional uptake sizes. The main objective of our study was therefore to investigate the correlation between anatomical tumor size as determined on CT, ¹⁸F-FDG uptake level of heterogeneity and the differences between various automatic PET MATV delineation approaches.

MATERIAL AND METHODS

Patients studies

25 patients with confirmed Non-Small Cell Lung Cancer (NSCLC, stage Ib-IIIb) were included in this study. All patients underwent an ¹⁸F-FDG PET/CT examination for staging purposes before treatment. Patients were instructed to fast a minimum of 6 hours before examination. Free-breathing PET and CT images were acquired 45 to 60 min after FDG injection. A total of seven 5-min bed positions with overlap were used for whole-body PET (Biograph PET/CT, Siemens, Erlangen, Germany) acquisitions that were corrected for attenuation using the CT data and iteratively reconstructed using the OSEM algorithm (4 iterations, 8 subsets). Within a week after

PET/CT acquisitions, 17 patients out of the 25 underwent surgery (lobectomy), which allowed further macroscopic examination. All specimens were processed the same way; namely the fresh specimens were put on ice and examined by one pathologist in order to measure the maximum diameter of the tumor in three dimensions (4). Specimen shrinkage estimated at ~10% was not taken into account since the measurements were carried out before fixation in formalin, allowing subsequent immunohistochemical examination for which the biopsies were paraffin-embedded.

This study was approved by the Institutional Ethics Review Board, and informed written consent was obtained from all patients prior to their inclusion in the study.

PET and CT tumors delineation

PET images were first up sampled using a cubic B-Splines interpolation scheme (26), in such a way that the voxels were of the same size as the associated CT images (see figure 1 for illustration). As the goal of this study was to compare anatomical and MATV as they can be seen and delineated on CT and ^{18}F -FDG PET images respectively, manual delineation on fused PET/CT images was not considered. Only primary tumors were delineated on both CT and PET images independently. Tumor anatomical volumes were manually delineated on CT without knowledge of the PET information by two observers both with more than 10 years experience in PET and CT imaging. Functional tumor volumes were manually delineated on PET images by one of the observers (and checked by the second observer) (13), as well as using semi or fully automatic image segmentation tools. A fixed threshold at 50% of the maximum (T_{50}) as suggested by Wu et al (11) and an adaptive threshold taking into account the background uptake (8) were considered. The adaptive threshold approach was optimized on phantom acquisitions carried out on the same PET/CT scanner that was used for the patient acquisitions. The method

requires the definition of a manual background region of interest (ROI) defining the background uptake in order to compute a first approximation of the tumor-to-background (T/B) contrast. Both observers were therefore instructed to place this background ROI in the lungs, at a distance of several centimeters from the boundaries of the tumors. They were however free to choose the actual size and position of the ROI, which led to two different results denoted A1 and A2. Finally, the Fuzzy Locally Adaptive Bayesian (FLAB) algorithm (18, 21) was also used. This algorithm allows automatic tumor delineation by computing a probability of belonging to a given “class” (e.g. tumor or background) for each voxel. This probability is determined by taking into account the voxel intensity with respect to the statistical distributions (characterized by their mean and variance) of the voxels in the various regions of the image, as well as spatial correlation with neighboring voxels. FLAB has demonstrated its ability to accurately differentiate if necessary both the overall tumor spatial extent from its surrounding background as well as tumor sub-volumes with different uptakes (18).

Investigated parameters and analysis

Firstly, for the 17 patients for whom macroscopic measurements were available, the maximum diameters were measured as the largest dimension in any orientation considering the different volume delineations (manual on CT and PET, T_{50} , A1&2 and FLAB) and compared to the histopathology reference. We reported both absolute (in cm) and relative (%) errors with respect to the maximum diameter in order to establish a hierarchy between the different methods. Secondly, for all patients the anatomical tumor volumes defined on CT images and the MATV obtained by each delineation approach were compared to each other. Delineations on original non upsampled PET images were performed to verify that the upsampling would not bias

the results of the various methods. Finally, the ^{18}F -FDG uptake heterogeneity was estimated using the coefficient of variation (COV) defined as the ratio between the standard deviation of SUV values and the mean SUV value within the delineated MATV. Two different volumes were used to calculate COV. The first was the one obtained using FLAB (COV_{FLAB}), since it was found to be the most accurate with respect to histology measurements (see section 3.1), while the second was the CT-based volume (COV_{CT}) copied onto the PET images.

Summary statistics are expressed as mean \pm 1 standard deviation (SD). Pearson coefficients were used to estimate correlation between parameters. Paired t-tests were used to assess the differences between the tumor volumes distributions obtained with the various delineation approaches. As most distributions were not normally distributed, they were log-transformed before analysis. All tests were two-sided and p values <0.05 were considered statistically significant.

RESULTS

Comparison with maximum diameter (histopathology reference)

Table 1 presents the results of maximum diameters measured on the 17 tumors based on either macroscopic examination or PET and CT images. All measured diameters were highly correlated with macroscopic measurements for all delineation approaches considered (r from 0.89 for T_{50} to 0.99 for FLAB, $p < 0.0001$) (figure 2A-C). Despite high correlations with maximum diameter for all methodologies as shown in table 1 and figure 2, significant differences were observed among the delineations (figure 3A-B). On the one hand, CT delineation consistently overestimated the maximum diameter of all tumors ($+32\pm 37\%$) with errors up to 3.5cm ($+140\%$). Manual delineation on PET images led to no significant bias but high standard deviation (mean error of $8\pm 17\%$), with maximum errors of -1.5cm (-17%) and +1.2 cm ($+39\%$).

On the other hand, PET automatic delineations mostly led to under evaluation of the real diameter. T_{50} led to the largest under estimation ($-15\pm 17\%$) with errors up to $+1.8\text{cm}$ ($+32\%$) and -2.3 cm (-46%). Adaptive thresholding led to a better accuracy, with similar results for both observers ($-11\pm 17\%$ and $-12\pm 16\%$ for A1 and A2 respectively) and errors up to -2cm (-40%). FLAB was associated with the most accurate results ($-4\pm 8\%$) with no error above $\pm 0.4\text{cm}$ ($\pm 13\%$).

Comparison of tumor volumes

Table II summarizes the results of tumor volumes for all patients. No significant differences in terms of volume determination on CT were measured between the two observers ($p > 0.08$). Therefore only the results for one observer will be considered from here onwards. No significant difference was observed between volumes obtained on original or up sampled PET images.

Anatomical tumor volumes delineated on CT images were the largest ($55\pm 74\text{cm}^3$) and were significantly different from all volumes defined on PET images ($p < 0.0001$). In addition, all PET based methodologies resulted in volumes being significantly different from each other ($p < 0.0001$). Among the PET defined tumor volumes, and consistent with what was observed according to the maximum diameters, the smallest volumes were obtained with T_{50} ($18\pm 25\text{cm}^3$) followed by the adaptive threshold ($23\pm 33\text{cm}^3$) and finally FLAB ($40\pm 71\text{cm}^3$) and manual ($47\pm 76\text{cm}^3$).

Regarding the overlap of the delineated volumes, the larger CT volumes almost systematically enclosed the PET based volumes, except for eight cases where small regions of PET uptake were just outside the anatomical volume, involving small margins comprising only a few voxels. The smallest PET uptake volumes generated with T_{50} were also almost systematically enclosed within the volumes defined by the adaptive thresholding, which in turn were mostly enclosed within the FLAB based

volumes. Figure 4 illustrates three different cases representative of the various situations encountered.

Correlation between ^{18}F -FDG uptake heterogeneity and tumor size, impact on delineation results

The calculated COV using the two different delineated tumor volumes (COV_{FLAB} and COV_{CT}) were highly correlated ($r=0.98$, $p<0.0001$). The heterogeneity of the PET uptake in these lung tumors was moderate to high, with a mean COV_{FLAB} of 0.26 ± 0.06 ranging from 0.21 to 0.48. COV_{CT} were higher with a mean of 0.37 ± 0.08 (range 0.3-0.6). 22 out of 25 tumors were characterized by a COV_{FLAB} between 0.2 and 0.3 (0.25 and 0.4 for COV_{CT}) and the three most heterogeneous were characterized by a COV_{FLAB} of 0.32, 0.36 and 0.48 (0.46, 0.48 and 0.69 for COV_{CT}). Figure 5 shows two different lesions with associated COV_{CT} and COV_{FLAB} values. A moderate but significant correlation was found between CT volumes and the PET heterogeneity, as larger anatomical volumes exhibited higher heterogeneity ($r=0.44$ and $r=0.5$ for COV_{CT} and COV_{FLAB} respectively, $p<0.03$). A similar correlation was found between MATVs and the corresponding heterogeneity, as larger functional volumes exhibited also significantly higher heterogeneity ($r=0.51$ and $r=0.58$ for COV_{CT} and COV_{FLAB} respectively, $p<0.002$).

Tumor size had an impact on the differences observed between the delineation results using the different images and segmentation approaches considered. A moderate ($r=0.44$) correlation was observed between anatomical tumor volumes and the differences between FLAB and T_{50} results (see figure 6A). The larger was the anatomical size of the lesion, the larger were the differences between FLAB and T_{50} volumes ($p=0.025$). Similar non significant trends were observed for differences between adaptive thresholding volumes or manual delineation and FLAB, ($r=<0.4$,

$p>0.08$). No correlation was found between anatomical tumor size and the differences observed between CT volumes and all of the PET volumes determined with the different segmentation approaches considered.

The impact of PET uptake heterogeneity was more significant than anatomical tumor size on the resulting MATV differences using the different PET delineation methodologies considered. As illustrated in figure 6B, differences between MATV obtained with T_{50} and FLAB were highly correlated ($r<-0.8$) with PET heterogeneity ($p<0.0001$) estimated either with COV_{CT} or COV_{FLAB} . The higher was the heterogeneity within the tumor, the smaller were the MATV obtained with T_{50} compared to that derived by FLAB. Similar correlation was observed for the differences between FLAB and A1 ($r<-0.7$, $p<0.0001$), as well as between FLAB and manual delineation ($r<0.6$, $p<0.001$).

DISCUSSION

MATV delineation on PET has been a subject of growing interest for NSCLC for several years, especially for radiotherapy applications such as dose redistribution, boosting and painting for which the MATV is not used in place of the anatomical one but rather as a complement to increase or redistribute dose within the lesion (27-29). These techniques are of interest especially for large tumors characterized by heterogeneous uptake within the MATV. However, the optimal MATV delineation methodology is still subject to debate, especially for these tumor cases.

Our results confirm that large discrepancies can be observed in image based determination of NSCLC tumor volumes according to the methodology used for tumor delineation. Using morphological imaging and manual delineation, we observed a large overestimation of TV as previously described by several authors (13). Using a fixed threshold of 50% as recommended by Wu et al (11), the

estimation of the maximum tumor diameter on PET images was not correct. We observed a constant underestimation of the maximum diameter which differs from Wu et al. since they reported mostly overestimations of the maximum diameter of the tumor. This difference is most probably related to the size of the tumors considered in the two studies. Wu et al. included mostly small tumors (median diameter of 2cm, range 1.1-6.5cm) while we considered larger tumors (4 ± 2 cm, range 1.5-9cm). The discordant results could be explained by the failure of binary threshold approaches to deal with heterogeneity which is more present in larger tumors.

On the other hand, we found similar differences between CT and PET volumes as Wu et al. in their subsequent study (13). CT volumes were significantly larger than PET based volumes in both studies, despite the differences in tumor sizes considered. In our group of patients, we mostly observed that the MATV was completely enclosed in the larger anatomical tumor volumes. Only in few cases elevated tracer uptake was observed outside the limits of the anatomical tumor, and this involved only of a few voxels. This marginal difference may be explained either by imperfect spatial registration between PET and CT modalities and/or the impact of respiratory motion.

Using the adaptive thresholding methodology as described by Nestle et al (8) PET tumor sizes did correlate well with the histopathology based gold standard, albeit with an underestimation of the maximum diameters in our group of lung tumors. Our results are in agreement with those of Van Baardjwick et al (4) who have previously investigated a slightly different semi-automatic methodology first proposed by Daisne et al (16).

In the current study, results obtained by the two observers using an adaptive thresholding were not significantly different, contrary to what was previously

observed in the case of esophageal tumors (30-31). This can be explained by the fact that contrary to the rather heterogeneous uptake that can be observed in the mediastinum surrounding esophageal tumors, the lung uptake is more homogeneous, thus leading to negligible variability in the manually determined background values. Manual delineation was less dependent on the heterogeneity within MATV than threshold-based methods, leading to satisfactory results with no significant bias (mean error <10%), although it was associated to a large standard deviation (17%) as some MATV were either largely overestimated (mostly the smaller lesions with lower contrast) or underestimated (some of the most heterogeneous ones with complex shapes). Overall, manual delineation was highly correlated with FLAB ($r=0.96$).

Automatic delineation on PET images using FLAB provided the best estimation of the tumor diameters in accordance with our previous evaluation of FLAB performance (18). Other advanced segmentation algorithms able to deal with heterogeneous MATV could potentially yield similar satisfactory results (22, 32). In our previous study FLAB was compared with a fixed threshold at 42%, instead of 50%, but with similar trends in the observed results. Furthermore, in our previous work the segmentation algorithms were applied to the original PET images without up-sampling and therefore with larger voxels. In the present study a resampling was carried out for an easier comparison with CT delineations and overlap estimation. This approach resulted in a more accurate estimation of the differences between PET and CT based delineation methodologies, without however significant differences in resulting volumes with respect to delineation performed on non-resampled images.

Tracer uptake heterogeneity within the MATV has been recognized as an important factor and a plausible explanation of failed cancer treatments (33). It has also been demonstrated in malignancies such as sarcomas, esophageal, cervix, as well as head and neck cancer that local and regional tracer uptake heterogeneity assessment with PET imaging can predict outcome (34-36). In NSCLC, Nestle et al already observed a larger variability between MATV delineations due to spatial tracer uptake heterogeneity, without however quantifying this heterogeneity and the associated correlation with the MATV results (8). The heterogeneity impact on the MATV delineation results can only be observed and reach statistically significant levels for objects larger than a few centimeters in diameter, since the limited PET spatial resolution cannot provide accurate imaging of tracer heterogeneity in smaller volumes of interest. These larger tumors are also most frequently encountered in radiotherapy treatment for which an accurate delineation of the overall MATV may be advantageous, particularly if one considers dose painting or boosting treatment scenarios.

Although limited by the small sample of patients with results that need to be confirmed with a larger group, our study adds several elements to the existing knowledge regarding the correlation between anatomical tumor size and ^{18}F -FDG PET uptake in NSCLC. Our results suggest that the larger the tumor, the more heterogeneous the ^{18}F -FDG PET uptake is likely to be. This is in agreement with the expected evolution of NSCLC since necrosis, hemorrhage or myxoid changes, known to cause areas of low attenuation in CT images, are more likely to appear in larger tumors. These large heterogeneous MATV are less likely to be accurately delineated using simple fixed or even adaptive binary threshold methods.

In this study we have used the COV to quantify the heterogeneity of the PET tracer uptake within the tumor. This heterogeneity factor does not offer any information regarding the spatial distribution of the heterogeneity and could potentially result in the same value for very different heterogeneous distributions. However, this simple parameter which provides a global measure of heterogeneity is sufficient for the purposes and objectives targeted in this study, allowing us to observe significant correlations between tracer uptake heterogeneity and differences in the MATV segmentation results, either with COV_{FLAB} or COV_{CT} . The most heterogeneous lesions were characterized by COV_{FLAB} values above 0.3, however values from 0.2 to 0.3 were distributed in a rather continuous fashion making it hard to set a threshold value allowing the differentiation of homogeneous from heterogeneous distributions. A more detailed characterization of the spatial distribution of tumor heterogeneity, which was outside the scope of this study, can be obtained using for instance local and regional textural features (35).

In the present study, as well as in previously published ones within the same context, one common limitation is the lack of respiratory gating. 4D PET imaging can provide solutions to improve sub-volumes delineation for dose painting applications (37). However, in our dataset the large size of the tumors should have reduced the potential impact of respiratory motion on the obtained results. In theory, the MATV could have been over estimated for the smallest lesions by both the respiratory motion and the partial volume effects. In practice, in our patients only a small fraction of the lesions (10-20%) were below 2-3cm.

Finally, a second limitation of our study is the determination of the tumor extent based on the measurement of the maximum diameter and not the entire volume. Errors regarding the maximum diameter may translate into significantly larger errors

with respect to the entire functional volume, especially when considering heterogeneous uptake distributions. It is indeed possible to obtain an accurate maximum diameter with inaccurate 3D delineations, especially for complex shapes. Unfortunately full volume histopathology datasets for which protocols and corresponding volume estimations are associated with numerous approximations and inaccuracies are not available yet for NSCLC. Hence, the maximum diameter measurements can be considered as a satisfactory surrogate and has been used in the vast majority of clinical studies.

CONCLUSION

Volumes based on CT were systematically and significantly larger than those defined on PET images. In addition, tumor size and PET uptake heterogeneity had significant impact on the MATV PET delineation results using semi or fully automatic image segmentation tools. Based on our results, fixed and adaptive thresholding should not be used for the MATV delineation of large heterogeneous NSCLC ^{18}F -FDG PET uptake distributions. These methods inherently assume homogeneous uptake in both background and MATV, and therefore tend to largely underestimate the functional tumor's spatial extent in such cases. The use of thresholding approaches should be restricted to smaller lesions with sufficient tumor to background contrast, or for larger tumors exhibiting homogeneous uptake distributions. For an accurate automatic delineation of MATV in NSCLC, advanced image segmentation algorithms able to deal with tracer uptake heterogeneity should be used.

Acknowledgments

This work was partly funded by the French National Research Agency under the contract ANR-08-E TEC-005-01.

Table captions

Table I: Maximum diameters measurements on pathology and image delineations for all 17 patients. Pearson correlation (r) coefficients are given along with P values and 95% confidence intervals (CI) for r .

Table II: Tumor volumes measured on CT and PET images (n=25).

Figure captions

Figure 1: Illustration of up-sampled PET images (central axial slice). (A) original PET image with voxel size ($5.31 \times 5.31 \times 5 \text{ mm}^3$) and (B) up-sampled with voxel size equal to CT ($0.98 \times 0.98 \times 5 \text{ mm}^3$) using cubic B-spline interpolation.

Figure 2: (A) and (B) show correlations with manual delineations on CT and PET respectively, (C) correlations with FLAB delineations on PET.

Figure 3: (A) absolute (in cm) differences and (B) relative (%) errors between pathology measurements and image-based delineations.

Figure 4: white is for manual on CT, blue for T_{50} , purple for A1 and green for FLAB. (A) small ($< 2 \text{ cm}$ in diameter) and (B,C) larger lesions with (B) moderate ($\text{COV}_{\text{FLAB}} = 0.23$) and (C) higher ($\text{COV}_{\text{FLAB}} = 0.30$) heterogeneity. For readability purposes, A1 contours are not shown in B-C and manual PET contours are not shown in B as they were similar to FLAB and T_{50} .

Figure 5: illustration of the heterogeneity estimation for two tumors.

Figure 6: correlation between (A) anatomical size or (B) uptake heterogeneity, and differences between T_{50} and FLAB volumes.

		Measurements (cm) according to each delineation							
		Pathologic	CT1(Manual)	CT2(Manual)	PET(Manual)	PET(T ₅₀)	PET(A1)	PET(A2)	PET(FLAB)
Patients	1	6.2	6.6	6.7	5.7	4.6	5	4.8	5.8
	2	2.7	3.3	3.3	3.4	2.8	3.1	2.8	3
	3	9	10.5	10.1	8.9	7	7.5	7.7	9.2
	4	1.5	1.8	1.9	2.1	1.3	1.6	1.3	1.5
	5	1.8	3.4	3.4	2	1.2	1.4	1.3	1.6
	6	3.1	4	3.9	3.2	2.4	2.6	2.5	2.8
	7	4.3	5	5.1	4.5	3.8	3.9	3.8	3.9
	8	3.1	5.7	5.7	5.1	2.8	4	3.7	3.5
	9	3.5	3.9	4	3.4	2.7	2.9	3	3.1
	10	5.7	7.6	7.7	7.4	7.5	4.7	6.7	5.4
	11	5	5.1	5.3	4.7	2.7	3	2.9	4.6
	12	2.8	3.5	3.2	3.2	2.4	2.5	2.6	2.8
	13	4.1	5.2	5.1	4.3	3.2	3.3	3.3	4
	14	4	4.8	4.9	3.7	3.2	3.4	3.2	3.9
	15	7	7.4	7.4	5.8	6.2	6.5	6.3	6.7
	16	2.3	2.3	2.4	2.1	1.8	1.7	1.9	2.1
	17	2.5	6	5.9	4.5	2.5	2.7	2.6	2.2
Statistics	Mean±STD	4.0±2.0	5.1±2.2	5.1±2.1	4.2±1.9	3.4±1.9	3.5±1.6	3.6±1.8	3.9±2.0
	Median	3.5	5.0	5.1	3.7	2.8	3.1	3.0	3.5
	Range	1.5-9	1.8-10.5	1.9-10.1	1.9-8.9	1.2-7.5	1.4-7.5	1.3-7.7	1.5-9.2
	Pearson r	-	0.90	0.91	0.95	0.89	0.95	0.93	0.99
	95% CI for r	-	0.74-0.96	0.76-0.96	0.86-0.98	0.72-0.96	0.85-0.98	0.81-0.98	0.98-1.00
	P	-	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001

Table I

Tumor volume (cm ³) (n=25)	Mean±STD	Median	Range
CT1(Manual)	54.5±74.0	28.2	1.9-338.9
CT2(Manual)	55.1±74.8	29.1	1.8-339.4
PET(Manual)	47.3±76.4	21.3	2.1-356.2
PET(T ₅₀)	17.7±25.1	9.2	8.5-125.8
PET(A1)	22.6±33.2	11.9	1.2-166.9
PET(A2)	21.8±33.9	11.3	0.9-172.4
PET(FLAB)	39.5±70.5	15.8	1.1-345.1

Table II

REFERENCES

1. Hicks RJ, Kalff V, MacManus MP, et al. (18)F-FDG PET provides high-impact and powerful prognostic stratification in staging newly diagnosed non-small cell lung cancer. *J Nucl Med.* Nov 2001;42(11):1596-1604.
2. MacManus M, Nestle U, Rosenzweig KE, et al. Use of PET and PET/CT for radiation therapy planning: IAEA expert report 2006-2007. *Radiother Oncol.* Apr 2009;91(1):85-94.
3. Chiti A, Kirienko M, Gregoire V. Clinical use of PET-CT data for radiotherapy planning: what are we looking for? *Radiother Oncol.* Sep 2010;96(3):277-279.
4. van Baardwijk A, Bosmans G, Boersma L, et al. PET-CT-based auto-contouring in non-small-cell lung cancer correlates with pathology and reduces interobserver variability in the delineation of the primary tumor and involved nodal volumes. *Int J Radiat Oncol Biol Phys.* Jul 1 2007;68(3):771-778.
5. Biehl KJ, Kong FM, Dehdashti F, et al. 18F-FDG PET definition of gross tumor volume for radiotherapy of non-small cell lung cancer: is a single standardized uptake value threshold approach appropriate? *J Nucl Med.* Nov 2006;47(11):1808-1812.
6. Hellwig D, Graeter TP, Ukena D, et al. 18F-FDG PET for mediastinal staging of lung cancer: which SUV threshold makes sense? *J Nucl Med.* Nov 2007;48(11):1761-1766.
7. Yaremko B, Riauka T, Robinson D, Murray B, McEwan A, Roa W. Threshold modification for tumour imaging in non-small-cell lung cancer using positron emission tomography. *Nucl Med Commun.* May 2005;26(5):433-440.
8. Nestle U, Kremp S, Schaefer-Schuler A, et al. Comparison of different methods for delineation of 18F-FDG PET-positive tissue for target volume definition in radiotherapy of patients with non-Small cell lung cancer. *J Nucl Med.* Aug 2005;46(8):1342-1348.
9. Yu HM, Liu YF, Hou M, Liu J, Li XN, Yu JM. Evaluation of gross tumor size using CT, 18F-FDG PET, integrated 18F-FDG PET/CT and pathological analysis in non-small cell lung cancer. *Eur J Radiol.* Oct 2009;72(1):104-113.
10. Stroom J, Blaauwgeers H, van Baardwijk A, et al. Feasibility of pathology-correlated lung imaging for accurate target definition of lung tumors. *Int J Radiat Oncol Biol Phys.* Sep 1 2007;69(1):267-275.
11. Wu K, Ung YC, Hornby J, et al. PET CT thresholds for radiotherapy target definition in non-small-cell lung cancer: how close are we to the pathologic findings? *Int J Radiat Oncol Biol Phys.* Jul 1 2010;77(3):699-706.
12. Yu J, Li X, Xing L, et al. Comparison of tumor volumes as determined by pathologic examination and FDG-PET/CT images of non-small-cell lung cancer: a pilot study. *Int J Radiat Oncol Biol Phys.* Dec 1 2009;75(5):1468-1474.
13. Wu K, Ung YC, Hwang D, et al. Autocontouring and manual contouring: which is the better method for target delineation using 18F-FDG PET/CT in non-small cell lung cancer? *J Nucl Med.* Oct 2010;51(10):1517-1523.
14. Hatt M, Cheze-Le Rest C, Aboagye EO, et al. Reproducibility of 18F-FDG and 3'-deoxy-3'-18F-fluorothymidine PET tumor volume measurements. *J Nucl Med.* Sep 2010;51(9):1368-1376.
15. Schaefer A, Kremp S, Hellwig D, Rube C, Kirsch CM, Nestle U. A contrast-oriented algorithm for FDG-PET-based delineation of tumour volumes for the radiotherapy of lung cancer: derivation from phantom measurements and validation in patient data. *Eur J Nucl Med Mol Imaging.* Nov 2008;35(11):1989-1999.

16. Daisne JF, Sibomana M, Bol A, Doumont T, Lonneux M, Gregoire V. Tri-dimensional automatic segmentation of PET volumes based on measured source-to-background ratios: influence of reconstruction algorithms. *Radiother Oncol.* Dec 2003;69(3):247-250.
17. Yu H, Caldwell C, Mah K, et al. Automated radiation targeting in head-and-neck cancer using region-based texture analysis of PET and CT images. *Int J Radiat Oncol Biol Phys.* Oct 1 2009;75(2):618-625.
18. Hatt M, Cheze le Rest C, Descourt P, et al. Accurate automatic delineation of heterogeneous functional volumes in positron emission tomography for oncology applications. *Int J Radiat Oncol Biol Phys.* May 1 2010;77(1):301-308.
19. El Naqa I, Yang D, Apte A, et al. Concurrent multimodality image segmentation by active contours for radiotherapy treatment planning. *Med Phys.* Dec 2007;34(12):4738-4749.
20. Montgomery DW, Amira A, Zaidi H. Fully automated segmentation of oncological PET volumes using a combined multiscale and statistical model. *Med Phys.* Feb 2007;34(2):722-736.
21. Hatt M, Cheze le Rest C, Turzo A, Roux C, Visvikis D. A fuzzy locally adaptive Bayesian segmentation approach for volume determination in PET. *IEEE Trans Med Imaging.* Jun 2009;28(6):881-893.
22. Belhassen S, Zaidi H. A novel fuzzy C-means algorithm for unsupervised heterogeneous tumor quantification in PET. *Med Phys.* Mar 2010;37(3):1309-1324.
23. Dewalle-Vignion AS, Betrouni N, Lopes R, Huglo D, Stute S, Vermandel M. A New Method for Volume Segmentation of PET Images, Based on Possibility Theory. *IEEE Trans Med Imaging.* Feb 2011;30(2):409-423.
24. Sebastian TB, Manjeshwar RM, Akhurst TJ, Miller JV. Objective PET lesion segmentation using a spherical mean shift algorithm. *Med Image Comput Comput Assist Interv.* 2006;9(Pt 2):782-789.
25. Hatt M, Cheze Le Rest C, Albarghach N, Pradier O, Visvikis D. PET functional volume delineation: a robustness and repeatability study. *Eur J Nucl Med Mol Imaging.* Apr 2011;38(4):663-672.
26. Thevenaz P, Blu T, Unser M. Interpolation revisited. *IEEE Trans Med Imaging.* Jul 2000;19(7):739-758.
27. Thorwarth D, Geets X, Paiusco M. Physical radiotherapy treatment planning based on functional PET/CT data. *Radiother Oncol.* Sep 2010;96(3):317-324.
28. Petit SF, Aerts HJ, van Loon JG, et al. Metabolic control probability in tumour subvolumes or how to guide tumour dose redistribution in non-small cell lung cancer (NSCLC): an exploratory clinical study. *Radiother Oncol.* Jun 2009;91(3):393-398.
29. Lambin P, Petit SF, Aerts HJ, et al. The ESTRO Breur Lecture 2009. From population to voxel-based radiotherapy: exploiting intra-tumour and intra-organ heterogeneity for advanced treatment of non-small cell lung cancer. *Radiother Oncol.* Aug 2010;96(2):145-152.
30. Hatt M, Visvikis D, Pradier O, Cheze-le Rest C. Baseline (18)F-FDG PET image-derived parameters for therapy response prediction in oesophageal cancer. *Eur J Nucl Med Mol Imaging.* May 11 2011.
31. Hatt M, Visvikis D, Albarghach NM, Tixier F, Pradier O, Cheze-le Rest C. Prognostic value of (18)F-FDG PET image-based parameters in oesophageal cancer and impact of tumour delineation methodology. *Eur J Nucl Med Mol Imaging.* Mar 2 2011.
32. Nelson A, Brockway K, Nelson A, Piper J. PET Tumor Segmentation: Validation of a Gradient-based Method Using a NSCLC PET Phantom. Paper presented at: Society of Nuclear Medicine annual meeting, 2009.
33. Basu S, Kwee TC, Gatenby R, Saboury B, Torigian DA, Alavi A. Evolving role of molecular imaging with PET in detecting and characterizing heterogeneity of cancer tissue at the primary and metastatic sites, a

plausible explanation for failed attempts to cure malignant disorders. *Eur J Nucl Med Mol Imaging*. Jun 2011;38(6):987-991.

34. El Naqa I, Grigsby P, Apte A, et al. Exploring feature-based approaches in PET images for predicting cancer treatment outcomes. *Pattern Recognit*. Jun 1 2009;42(6):1162-1171.

35. Tixier F, Le Rest CC, Hatt M, et al. Intratumor heterogeneity characterized by textural features on baseline 18F-FDG PET images predicts response to concomitant radiochemotherapy in esophageal cancer. *J Nucl Med*. Mar 2011;52(3):369-378.

36. Eary JF, O'Sullivan F, O'Sullivan J, Conrad EU. Spatial heterogeneity in sarcoma 18F-FDG uptake as a predictor of patient outcome. *J Nucl Med*. Dec 2008;49(12):1973-1979.

37. Aristophanous M, Yap JT, Killoran JH, Chen AB, Berbeco RI. Four-Dimensional Positron Emission Tomography: Implications for Dose Painting of High-Uptake Regions. *Int J Radiat Oncol Biol Phys*. Oct 13 2010.