Robustness of intratumour ^{1} F-FDG PET uptake heterogeneity quantification for therapy response prediction in oesophageal carcinoma.

Mathieu Hatt, Florent Tixier, Catherine Cheze Le Rest, Olivier Pradier, Dimitris Visvikis

Robustness of intratumour 18F-FDG PET uptake heterogeneity quantification for therapy response prediction in ooesophageal carcinoma

Mathieu Hatt[1*], Florent Tixier[1*], Catherine Cheze Le Rest[2], Olivier Pradier[3], Dimitris Visvikis[1]

[1] INSERM, UMR 1101 LaTIM, Brest, France.

[2] Nuclear Medicine, CHU Milétrie, Poitiers, France.

[3] Radiotherapy, CHRU Morvan, Brest, France.

* contributed equally to this work

Running title: Robustness of FDG PET tumour heterogeneity quantification

Corresponding author: M. Hatt,
INSERM, UMR 1101, LaTIM
CHRU Morvan
2 avenue Foch
29609 Cedex, Brest, France
Tel: +33(0)2.98.01.81.11
Fax: +33(0)2.98.01.81.24
e-mail: hatt@univ-brest.fr.

Wordcount: ~5988

Financial support: French research ministry PhD grant

Conflict of interest: none

ABSTRACT:

PURPOSE: Intra-tumour uptake heterogeneity in PET quantified through textural features for response to therapy has been investigated in several studies, including assessment of their robustness with respect to reconstruction and physiological reproducibility. However, there has been no thorough assessment of the potential impact of pre-processing steps on the resulting quantification and its predictive value. The goal of this work was to assess the robustness of PET heterogeneity textural features with respect to the delineation of functional volumes and partial volume effects (PVE) correction (PVC).

METHODS: Fifty patients with oesophageal cancer were retrospectively analyzed. PVC of each PET image was performed. Tumour volumes were determined using fixed (FT) and adaptive thresholding (AT), and the fuzzy locally adaptive Bayesian (FLAB) algorithm and heterogeneity was quantified using local and regional textural features. Differences in the absolute quantification values of the image derived parameters considered were assessed using Bland-Altman analysis. The impact on their predictive value for the identification of patient non-responders was assessed by comparing areas under the receiver operating characteristic curves.

RESULTS: Heterogeneity parameters were more dependent on delineation than on PVC. The parameters most sensitive to delineation and PVC were among the regional ones (intensity variability and size zone variability), whereas local parameters such as entropy and homogeneity were the most robust. Despite the large differences in absolute values obtained from different delineation methods or after PVC, these differences did not necessarily translate into a significant impact on their predictive value.

CONCLUSIONS: Parameters such as entropy, homogeneity, dissimilarity (for local heterogeneity characterization) or zone percentage (for regional characterization) should be

privileged. This selection is based on a demonstrated high differentiation power in terms of predicting response, as well as a significant robustness with respect to the delineation method used and PVE.

Keywords: ^{18}FDG-PET/CT, heterogeneity, textural features, tumour delineation, partial volume effects correction, response to therapy prediction.

^{18}F-FDG Positron Emission Tomography/Computed Tomography (PET/CT) has been established as a powerful imaging technique mainly for diagnostic oncology applications [1]. It has been also increasingly considered for therapy applications, delineating gross tumour volumes in external beam radiotherapy treatment planning [2] and/or in monitoring response to therapy [3]. Within the specific context of response to therapy prediction or assessment, PET image derived indices such as metabolically active tumour volumes (MATV), mean standardized uptake values (SUV_{mean}) and derived total lesion glycolysis (TLG, equal to $MATV \times SUV_{mean}$) have been shown to provide a more accurate assessment of tumour burden with potentially higher predictive and prognostic value compared to standard maximum or peak SUV measurements (SUV_{max}, SUV_{peak}) for a variety of cancer models including lung [4], esophagus [5], head&neck [6], rectum [7], breast [8], pleural mesothelioma [9] and non-Hodgkin lymphomas [10].

More recently, intra-tumour uptake heterogeneity has been identified as a potential source of treatment failure [11] and its characterization through ^{18}F-FDG PET imaging is currently generating a substantial amount of interest [12-15]. Such characterization provides additional and complementary PET image derived quantitative indices with potential value as already demonstrated in predicting therapy response or as prognostic factors in several cancers including lung [16], sarcoma [17], oesophageal [18, 19] and rectal cancer [20]. A variety of methodologies has been proposed in order to assess intra-tumour uptake heterogeneity, including visual assessment [21], SUV coefficient of variation (SUV_{cov}) [20], area under the curve of the cumulative histogram (CH_{AUC}) [22] and textural features (TF) analysis [18].

TF analysis can potentially provide the most versatile range of indices that may be used to characterize uptake heterogeneity within the delineated tumour volume. However, recent studies have shown that only a limited number of these TF derived indices are robust with respect to the clinical range of image reconstruction algorithms and acquisitions protocols

[23], and reproducible with respect to physiological variability assessed on double baseline FDG PET scans [24]. Within the same context it has also been hypothesized that heterogeneity quantification could be dependent on the necessary functional tumour volume delineation method used, as well as be compromised by the low PET image spatial resolution and associated partial volume effects (PVE) [18, 22], although this has never been thoroughly investigated.

This study was therefore designed to assess the robustness of heterogeneity calculations based on TF analysis in terms of (i). absolute quantitative values and (ii). associated predictive value for therapy response assessment following concomitant radiochemotherapy in locally advanced oesophageal cancer patients (LAEC). The influence of both the tumour delineation methodology and the PVE was considered in this study, given their significant role in the process of determining heterogeneity indices.

MATERIALS AND METHODS

Patients

Fifty consecutive patients (table 1) with confirmed LAEC treated with exclusive concomitant radiochemotherapy between 2004 and 2008 were retrospectively included. All patients underwent an ^{18}F-FDG PET scan before initiating treatment as part of the staging procedure, and were subsequently treated with three courses of 5-fluorouracil/cisplatin and a median radiation dose of 60Gy given in 180cGy daily fractions delivered once daily, 5 days a week for 6–7 weeks. The analysis was carried out after an approval by the Institutional Ethics Review Board.

^{18}FDG-PET/CT Imaging

Patients fasted for at least 6 hours before injection of 5MBq/kg of ^{18}FDG, administered 60 minutes before data acquisition on a Philips GEMINI PET/CT scanner (Philips Medical Systems, Cleveland, OH, USA). CT data were acquired first (120kV and 100mAs, no contrast-enhancement). 3D PET data were acquired with 2min per bed position, and images were reconstructed using CT based attenuation correction (CT images were checked for artifacts that may lead to errors in quantification) and a 3D row-action maximum likelihood algorithm (RAMLA) using previously optimized protocol (2 iterations, relaxation parameter of 0.05, 5 mm 3-D Gaussian post-filtering, 4×4×4 mm^3 voxels grid sampling) [25].

Response evaluation

Response to therapy was evaluated one month after completion of treatment using CT and endoscopy based on Response Evaluation Criteria in Solid Tumours (RECIST) [26] with patients classified as non-responders including stable and progressive disease, or responders, including partial responders (PR) and complete responders (CR). In cases of PR or CR, patients also underwent fibroscopy. CR was confirmed by the absence of visible disease in the endoscopy and no viable tumour on biopsy. PR was confirmed by macroscopic residual (>10% viable) on biopsy. No discordance was observed between pathological, whenever available, and CT evaluation.

^{18}FDG-PET image analysis

Intra-tumour heterogeneity parameters

On the one hand, intra-tumour uptake heterogeneity was quantified considering only those indices that have been previously shown as robust considering image reconstruction and acquisition variability [23], and physiological reproducibility based on double baseline PET acquisitions [24]. Figure 1 illustrates the two types of heterogeneity parameters under

consideration. First, local heterogeneity parameters quantifying intensity variations between each voxel and its immediate neighbors only, averaged over the entire volume: entropy, homogeneity, dissimilarity. Second, regional heterogeneity parameters calculated through analysis at the level of groups of voxels and areas of various sizes and intensities: intensity variability (IV), size-zone variability (SZV), zone percentage (ZP), high intensity emphasis (HIE) [24].

On the other hand, these same parameters were also shown to provide significant predictive value in identifying non-responders to radiochemotherapy in LAEC [18]. The area under the curve of the cumulative SUV-volume histogram (CH_{AUC}) was included as an alternative global heterogeneity measure not previously evaluated for reproducibility and robustness. CH_{AUC} quantifies percentage of total tumour volume above percentage threshold of SUV_{max}, a higher AUC representing a more homogeneous uptake [22]. Other global parameters such as skewness, variance, kurtosis or SUV_{COV} were excluded considering their previously demonstrated poor physiological reproducibility [24]. Finally, other standard parameters not describing heterogeneity, such as SUV_{mean} and MATV were also included in this study for comparison purposes.

Delineation approaches and PVE correction

The heterogeneity parameters are extracted from functional tumour volumes delineated on the PET images. In order to investigate the dependency of heterogeneity quantification on tumour delineation, each oesophageal lesion was delineated on the baseline PET images using three different methods: a fixed threshold at 42% of the SUV_{max} (FT42%) [27], an adaptive threshold (AT) taking into account the measured contrast between the tumour (determined through a fixed threshold at 70% of the SUV_{max}) and the background uptake (determined through a ROI in the mediastinum) [28], and the Fuzzy Locally Adaptive Bayesian (FLAB)

algorithm [29]. The FLAB approach allows automatic tumour delineation by computing a probability of belonging to a given "class" (e.g. tumour or background) for each voxel within a 3D ROI containing the tumour and its surrounding background. This probability is calculated by taking into account the voxel's 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 its spatial correlation with neighboring voxels in 3D. This approach has been previously validated on simulated and clinical datasets showing high combined accuracy, robustness and reproducibility, for the delineation of both homogeneous and heterogeneous MATVs [30-32].

In order to investigate the dependency of heterogeneity quantification on PVE, all PET images were corrected for PVE using a voxel-based iterative deconvolution approach including wavelet-based denoising previously validated for PET imaging [33]. MATVs were subsequently delineated on the deconvolved PET images using FLAB and all image derived parameters considered were extracted (figure 2).

Statistical Analysis

Statistical analyses were performed using Medcalc^{TM} (MedCalc Software, Belgium). Correlation between parameters under investigation was assessed using Pearson's correlation coefficients applied to parameters derived from FLAB delineations on the non-PVE corrected PET images. Considering the parameters' values obtained with the FLAB delineation on the original non-PVE corrected PET images as reference, the variability of each PET image derived parameter with respect to either those derived using the two alternative delineation techniques considered, or derived from the corresponding PVE corrected images, was assessed using Bland-Altman analysis. This analysis provided mean and standard deviation (SD) of differences relative to the mean of the two measurements, as well as the respective

95% confidence intervals. Associated upper and lower reproducibility limits (URL and LRL respectively) defining the range of changes were calculated as mean±1.96×SD. Sub-analysis according to histology (adenocarcinoma *vs.* squamous cell carcinoma) was also carried out but no statistically significant differences were found and results were not included to reduce manuscript length.

For each of the image derived parameters considered, their predictive value with respect to identifying different categories of patient responders was assessed using receiver operating characteristics (ROC) analysis. The area under the curve (AUC) was considered as a figure of merit to quantify the predictive value of each PET image derived index. In order to assess whether or not this predictive value was dependent on the tumour delineation approach used or PVE, differences between AUCs were assessed using Delong et al methodology [34], assuming a statistically significant difference with $p<0.05$.

RESULTS

Correlation between parameters

Parameters under investigation exhibited different levels of cross correlation (table 2). The highest correlations were observed between MATV and intensity variability (r=-0.97), MATV and entropy (r=0.82), as well as between HIE and CH_{AUC} (r=-0.97), homogeneity and dissimilarity (r=-0.93), entropy and ZP (r=-0.90). Some other parameters were moderately correlated (r=0.5 and 0.8) whereas others were not at all correlated, for instance MATV and SZV, HIE or CH_{AUC} (r=-0.16, -0.22 and 0.07 respectively).

Impact of tumour delineation method

Using FT42% and AT led to substantially different functional volumes than using FLAB, with ~7±50% and -18±49% differences respectively (table 3). The associated upper and lower

reproducibility limits were very high (~75-100%), highlighting the fact that different delineation techniques led for some patients to highly different functional volumes.

Several heterogeneity parameters derived from these largely different volumes were relatively independent on these differences. Such parameters included CH_{AUC}, entropy, homogeneity, ZP and HIE with mean differences around 2% to -5% associated with a SD of ±4% to ±25% and upper and lower reproducibility limits from [-7%, +9%] for CH_{AUC} to [-39%, +48%] for homogeneity. By comparison, SUV_{mean} was equally dependent on the tumour delineation approach, with mean differences of about -1% to 5% and associated SD of ±21%. On the contrary, heterogeneity parameters such as IV, dissimilarity and SZV were the most sensitive to the delineated functional volume, resulting in mean differences of between 18% and -16% associated with a SD of ±20% to ±50% and higher upper and lower limits of 50-100%.

Impact of partial volume effects

Considering the FLAB delineation on the original non-corrected PET images as reference, the functional volume delineations on the PVC images led to relatively smaller volumes (mean differences of -14%) (table 3). Extracted heterogeneity parameters were overall less dependent on PVE corrections than they were on tumour delineation, most of the parameters exhibiting mean differences <10±15%, with the exception of IV, SZV and HIE. Similarly, upper and lower reproducibility limits were smaller than the corresponding values associated with the use of different tumour delineation methods, with the exception of SZV. The observed hierarchy in terms of dependence on either tumour delineation or PVE was also almost similar; the least sensitive parameters being the global and local heterogeneity characterization indices (CH_{AUC}, entropy, homogeneity, dissimilarity) and a single regional heterogeneity parameter (ZP). The remaining regional heterogeneity parameters (IV, HIE, and SZV) were the most sensitive (URL and LRL of ~±95%). The impact on standard parameters

such as MATV (upper and lower limits of about ±20-30%) was similar to that of some heterogeneity parameters (HIE and IV).

Impact on predictive value of baseline PET images

According to RECIST there were 14 non-responders and 36 responders (24 partial and 12 complete responders). Only results regarding the identification of non-responders *vs.* complete and partial responders are presented for readability purposes (table 4), as similar results were observed for the identification of complete responders *vs.* partial and non-responders.

All parameters under investigation when derived from FLAB delineated tumour volumes on the original baseline PET images (the chosen reference in this study) demonstrated high predictive value with AUC from 0.80 to 0.90, except for SUV_{mean}, SZV, HIE and CH_{AUC} (AUC of 0.60-0.72). This predictive value was mostly independent of the delineation method, although in absolute terms it was systematically higher when parameters were derived using FLAB delineated tumour volumes compared to the use of threshold-based ones (except for SUV_{mean}), or the use of PVE corrected images. The only statistically significant differences regarding the delineation method were found for homogeneity and dissimilarity indices. More specifically, FLAB MATV derived homogeneity and dissimilarity measures resulted in AUCs of 0.86 and 0.85 respectively, whereas FT42% MATV derived homogeneity and dissimilarity measures resulted in statistically significantly lower AUC of 0.74 ($p<0.05$). In the case of PVE correction, only two heterogeneity parameters were significantly affected, with statistically significant improvements of the AUCs for CH_{AUC} and HIE from 0.60 and 0.65 to 0.77 and 0.83 respectively ($p<0.04$).

DISCUSSION

There is currently increasing interest in the use of new image derived indices and more particularly in those parameters which allow quantification of intra-tumour activity distribution heterogeneity. Our work brings new results concerning the robustness of such PET image derived indices with respect to the inherent low PET image resolution and the determination of functional tumour volumes within which the heterogeneity analysis is performed. The results are complementary to the previously established robustness evaluations with respect to reconstruction parameters [23], and physiological reproducibility on double baseline scans [24]. Such characterization in terms of robustness and reproducibility for these new indices to all these factors is a requirement in order to reliably propose their use and assess their potential impact in clinical research trials. The current study is the first to our knowledge, to investigate the dependence of several intra-tumour uptake heterogeneity quantification parameters on functional tumour volume delineation techniques and PVE correction. Within this context, we assessed the potential impact of such factors on the predictive value of these PET image derived parameters in terms of response classification for oesophageal carcinoma patients undergoing concomitant chemoradiotherapy.

Among all considered tumour volume derived parameters, MATV was found to be the most sensitive parameter to the tumour delineation method used (about ±75-115% reproducibility limits). As already demonstrated in previous studies [5, 32], AT and FT42% led to smaller volumes than FLAB for larger, more heterogeneous uptakes (see figure 2 for an example), whereas much larger volumes were obtained for small, low uptake contrast lesions. As a consequence, FLAB delineations of large and heterogeneous tumours included more variability in voxel intensities than threshold-based methods which incorporated only high intensity areas. On the contrary for small, low uptake contrast tumours, threshold-based methods overestimated functional volumes by incorporating background voxels, leading to additional voxel intensity variability.

The respective robustness of the intra-tumour PET uptake heterogeneity parameters may be related to the type of heterogeneity information they quantify. For instance, entropy, homogeneity and dissimilarity are local heterogeneity parameters quantifying variations in the intensity between consecutive voxels that are then averaged over the entire volume. This should vary only slightly if one considers smaller or larger overall tumour volumes. These parameters were indeed among the least impacted by the delineation method (upper and lower reproducibility limits between <±10% to ±50%). On the other hand, regional heterogeneity parameters (ZP, HIE, IV, and SZV) quantify relationships between areas (groups of voxels) with different size and intensity. Among these, IV and SZV were found to be very sensitive to delineation (upper and lower limits between ±75% to ±110%), whereas ZP and HIE were more robust (upper and lower limits between ±20% to ±50%). There are different reasons that can account for the improved robustness of ZP and HIE over IV and SZV. In the case of ZP, the size of the characterized MATV is included in its calculation, normalizing over the differences of tumour volumes associated with different delineation approaches. Since the HIE measurement focuses on high intensity uptake regions, differences between the various MATV delineations is attenuated since all of them include at least the most active parts of the functional tumour volume. On the contrary, the least robust indices such as IV and SZV involve measurements which are neither focusing on high intensity regions nor normalized with respect to the size of the MATV volumes to characterize. Depending on the delineation result, the respective sizes and average intensity of each sub-region might be totally different on underestimated or over-estimated tumour volumes, leading to the low robustness of these two parameters. Finally, the only 'global' heterogeneity parameter included in this study (CH_{AUC}) was the least sensitive to MATV delineation, explained by the fact that smaller or larger volume delineations would only lead to changes in the extremity of the cumulative histogram, subsequently not significantly altering the overall area under the histogram used in

the determination of the CH_{AUC}, which may also be related to the limited ability of CH_{AUC} to quantify actual uptake heterogeneity.

Overall, heterogeneity characterization parameters that were the least dependent on the tumour delineation were CH_{AUC}, a global estimation of intra-tumour heterogeneity, and entropy, characterizing local changes in consecutive voxels. Although slightly more dependent on the two factors considered, other robust heterogeneity parameters for local and regional intra-tumour heterogeneity characterization include homogeneity as well as ZP and HIE respectively. Consequently these indices could be used to quantify heterogeneity of MATVs delineated using less reliable methods with a small impact on both their absolute and predictive value within the context of classifying patient response to therapy. Given the current lack of consensus regarding the method of choice for PET image based MATV determination these parameters should be privileged since they are robust with respect to the delineation method employed. On the contrary, the use of the less robust heterogeneity parameters (dissimilarity for local characterization, SZV and IV for regional characterization) should be carefully considered particularly within the context of the MATV delineation method employed.

On the other hand, it is worth noting that despite some high absolute differences between heterogeneity parameters derived from largely different MATVs due to different delineation approaches, only non-statistically significant differences were obtained concerning the associated predictive value of these parameters in terms of response to therapy prediction. This might be explained by the initial large differentiation between patient groups of response for most of the heterogeneity parameters considered in the current oesophageal cancer cohort. Clearly, this may not be the case in other cancer models and associated therapy regimes where a more significant impact on the response predictive value of the PET intra-tumour heterogeneity parameters might be observed. Another explanation would be that almost all

observed differences occurred in the same direction for most of the patients, leading to the same discrimination between distributions of response.

Regarding the impact of PVE correction, the results were somehow different. Firstly, in absolute terms the impact of the PVE correction on the intra-tumour heterogeneity parameters was smaller compared that of the tumour delineation, with most of the parameters exhibiting lower and upper reproducibility limits of <±30%. Overall, parameters characterizing local heterogeneity at the scale of few voxels (entropy, homogeneity, dissimilarity) within the entire tumour were less sensitive than regional heterogeneity parameters (IV, ZP, SZV and HIE) that exhibited larger differences on the corrected images. This suggests that regional intra-tumour heterogeneity measurements are more sensitive to PVC related intensity changes within the corrected images. Amongst the regional parameters, ZP exhibited the lowest sensitivity to PVC, mainly because this measurement is mostly related to the size of the sub-regions and not their intensity, which is the main aspect modified by PVC. Similarly with the robustness to the tumour delineation approach, the parameters that were least affected by the PVC were the CH_{AUC} and entropy.

Similarly to the delineation approach impact on response to therapy classification, no significant differences were found between the AUC based on the intra-tumour heterogeneity parameters derived from PVE corrected and those from the original images. The only exceptions were CH_{AUC} and HIE for which their original limited predictive value (AUC of 0.60 and 0.65) significantly improved after PVC (AUC of 0.77 and 0.83 respectively) ($p<0.04$). These results confirm the findings of the study that firstly introduced CH analysis for heterogeneity characterization suggesting that PVC improves the CH_{AUC} curves [22]. On the other hand, the difference observed for the HIE parameter can be explained by the focus of this measurement on the high intensity tumour uptake areas that are the most altered by the PVE correction. Finally, this result could be also partly explained by the fact that these two

parameters derived from the original baseline PET images (AUC below 0.7) were amongst the least predictive leaving more potential for improvement, in contrast to the other parameters that were already characterized by AUC values superior to 0.80.

Overall, despite large differences observed on the absolute values of derived heterogeneity parameters depending on the delineation method used and PVE correction, the resulting predictive values in identifying non responders were mostly unchanged, certainly because the different groups of response were characterized by largely different distributions. However, one issue that can be raised from the large observed variability on the absolute values of PET image derived heterogeneity parameters is the need for a rigorous standardization of image pre-processing techniques and associated analysis to allow reliable cut-off values in multi-center studies using these indices. The physiological reproducibility of parameters such as entropy, homogeneity, dissimilarity or zone percentage providing quantification of local and regional intra-tumour uptake heterogeneity has been previously demonstrated. The present study has further highlighted the robustness of these parameters within the context of response to therapy studies, especially those involving multi-center analysis, showing that their predictive value is on the one hand not affected by PVE while on the other is relatively independent on the method used to delineate the tumour volumes to be analysed.

We did not investigate the robustness of heterogeneity indices with respect to various reconstruction algorithms, which would complement the study of Galavis et al [23] and will require a full separate analysis of tumours list mode data reconstructed with different algorithms. This may be carried out in a future work.

CONCLUSIONS

In this study, the robustness of PET image derived intra-tumour heterogeneity parameters with respect to the tumour volume where the analysis is performed and the PVE was investigated. Although differences in the absolute values of these parameters were obtained using different tumour delineation methods or after PVE correction, these differences did not always translate into a significant impact on their predictive value of response in an oesophageal cancer patient population undergoing concomitant chemoradiotherapy. Some parameters demonstrated superior robustness, although for some of them, such as for instance the global heterogeneity parameter CH_{AUC}, these small differences did result in significant changes in the associated predictive value. In conclusion, parameters such as entropy and homogeneity or zone percentage for local or regional heterogeneity characterization respectively should be preferred, as they provide both a high differentiation power in terms of patient response prediction, and are robust with respect to the delineation method and PVE.

Disclosure of Conflicts of Interest: No potential conflicts of interest were disclosed.

Parameter	Number of patients (%)
Gender	
Male	45(90)
Female	5(10)
Age	
Range	45-84
Median	69
Site	
Upper esophagus	13(26)
Middle esophagus	20(40)
Lower esophagus	17(34)
Histology type	
Adenocarcinoma	14(28)
Squamous cell carcinoma	36(72)
Histology differentiation	
Well differentiated	14(28)
Moderately differentiated	12(24)
Poorly differentiated	5(10)
Unknown	19(38)
TNM Stage	
T1	7(14)
T2	8(16)
T3	24(48)
T4	11(22)
N0	20(40)
N1	30(60)
M0	34(68)
M1	16(32)
AJCC Stage	
I	4(8)
IIA	8(16)
IIB	6(12)
III	16(32)
IVA	16(32)

Table 1. Patients characteristics

Parameter	SUV_{mean}	MATV	Entropy	Homogeneity	Dissimilarity	Intensity variability (IV)	Size-zone variability (SZV)	Zone percentage (ZP)	High intensity emphasis (HIE)	Area under the curve of the cumulative histogram (CH_{AUC})
SUV_{mean}	**1.00**	0.20	**0.30**	-0.10	-0.02	0.08	0.09	**-0.40**	**0.40**	**-0.50**
MATV		**1.00**	**0.82**	**0.69**	**-0.77**	**0.97**	-0.16	**-0.70**	-0.22	0.07
Entropy			**1.00**	**0.60**	**-0.80**	**0.77**	-0.25	**-0.90**	-0.08	-0.07
Homogeneity				**1.00**	**-0.93**	**0.80**	**-0.36**	**-0.42**	**-0.67**	**0.59**
Dissimilarity					**1.00**	**-0.83**	**0.41**	**0.60**	**0.58**	**-0.45**
Intensity variability (IV)						**1.00**	-0.25	**-0.62**	**-0.41**	**0.28**
Size-zone variability (SZV)							**1.00**	0.24	**0.43**	**-0.32**
Zone percentage (ZP)								**1.00**	-0.18	**0.32**
High intensity emphasis (HIE)									**1.00**	**-0.97**
Area under the curve of the cumulative histogram (CH_{AUC})										**1.00**

Table 2. Correlations (Pearson coefficients) between parameters derived from FLAB delineations on non-corrected PET images. Bold text denotes significant correlations.

Parameter			Difference with respect to FLAB delineation on non-PVE corrected image								
			FT42%			AT			PVC		
			mean±SD (%)	LRL (%)	URL (%)	mean±SD (%)	LRL (%)	URL (%)	mean±SD (%)	LRL (%)	URL (%)
Standard		SUV_{mean}	-1±20.7	-41.5	39.6	5.2±21.1	-36.2	46.7	16.9±8.7	-0.2	34
		MATV	7.1±52	-94.9	109.1	-18.4±49.4	-115.1	78.3	-14.4±19.2	-32.0	23.2
Heterogeneity quantification	Local	Entropy	1.5±9.4	-16.9	19.9	-4.3±12.6	-28.9	20.3	-4.6±5.3	-15.0	5.7
		Homogeneity	4.7±22.1	-38.7	48	-3.7±19.9	-42.8	35.4	4.2±10.9	-17.1	25.5
		Dissimilarity	-5.3±28	-60.1	49.5	4.9±22.6	-39.3	49.1	-3.2±11.4	-25.5	19.1
	Regional	Intensity variability (IV)	7±50.1	-91.1	105.1	-16.2±47	-108.4	76	0.2±19.9	-38.8	39.2
		Size-zone variability (SZV)	-6.5±48.9	-102.3	89.4	17.8±46.9	-74.2	109.8	3.3±47.9	-90.7	97.3
		Zone percentage (ZP)	-2.4±20.4	-42.3	37.5	5.3±14.1	-22.4	32.9	10.3±11.6	-12.5	33
		High intensity emphasis (HIE)	-4.6±19.8	-43.5	34.3	3.7±24.7	-44.6	52.1	-20.6±18.8	-57.5	16.3
	Global	Area under the curve of the cumulative histogram (CH_{AUC})	1.2±4.1	-6.8	9.2	-1.1±6	-12.8	10.6	5±5.6	-5.9	15.9

Table 3. Bland-Altman analysis (mean and standard deviation (SD), lower (LRL) and upper (URL) reproducibility limits, calculated as mean±1.96×SD) for differences between parameters derived from volumes delineated with fixed (FT42%) or adaptive (AT) thresholding, or on PVE corrected images, with respect to those derived from volumes delineated with FLAB.

.

Parameter	AUC of ROC curves (NR *vs.* PR+CR)			
	FT42%	**AT**	**FLAB**	**FLAB PVC**
SUV_{mean}	0.66	0.65	0.64	0.60
MATV	0.84	0.87	0.89	0.87
Entropy	0.84	0.86	0.88	0.85
Homogeneity	**0.74***	0.82	**0.86***	0.87
Dissimilarity	**0.74***	0.81	**0.85***	0.88
Intensity variability (IV)	0.85	0.87	0.90	0.88
Size-zone variability (SZV)	0.66	0.70	0.72	0.86
Zone percentage (ZP)	0.74	0.78	0.81	0.82
High intensity emphasis (HIE)	0.59	0.65	**0.65***	**0.83***
Area under the curve of the cumulative histogram (CH_{AUC})	0.56	0.60	**0.60***	**0.77***

* denotes statistically significant difference between the two AUCs.

Table 4. Area under the ROC curve (AUC), regarding the identification of non-responders for each parameter depending on the delineation method and PVC.

Figure captions

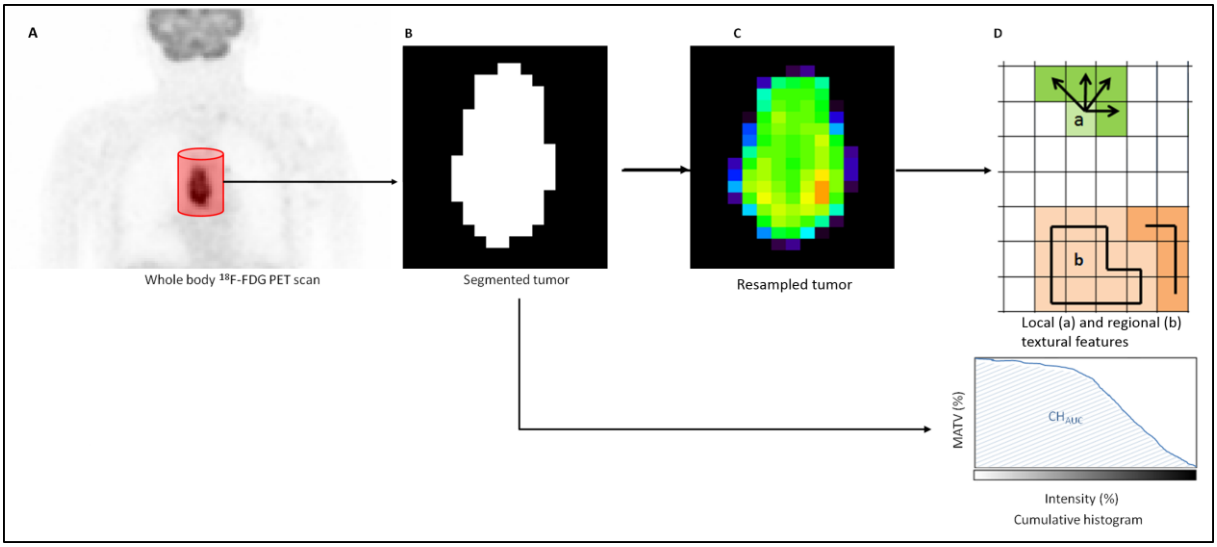


Figure 1. After delineation, MATV was resampled and extraction of local (a) and regional (b) heterogeneity parameters was performed. Local parameters were obtained by analysis of voxels along a direction (a) and areas of voxels sharing similar intensity (b). CH_{AUC} was also computed.

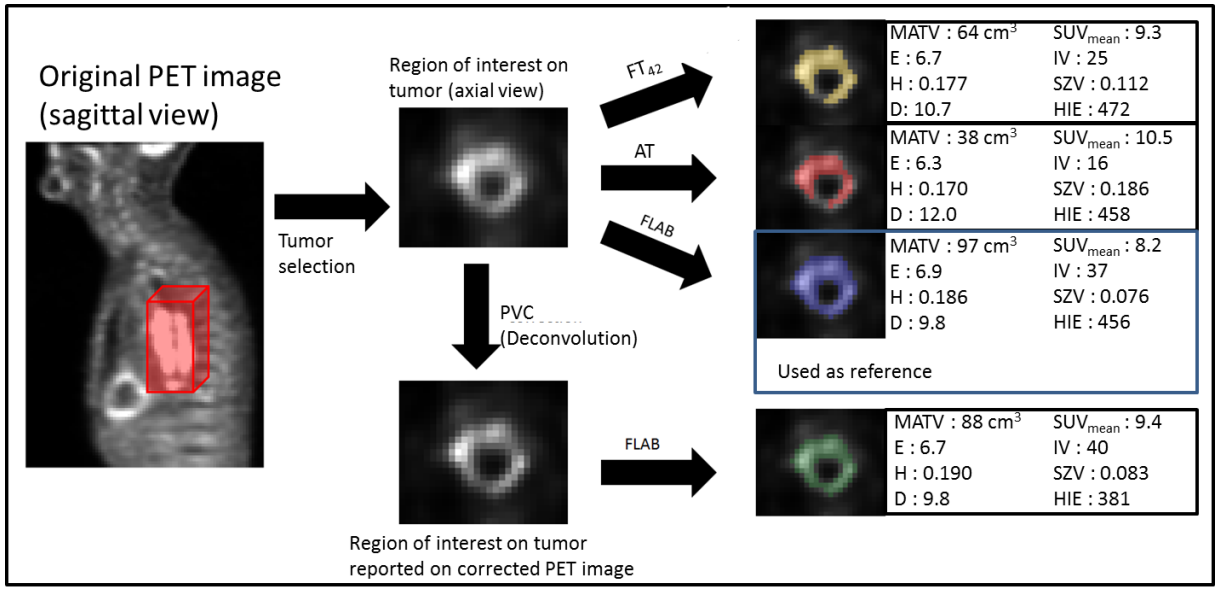
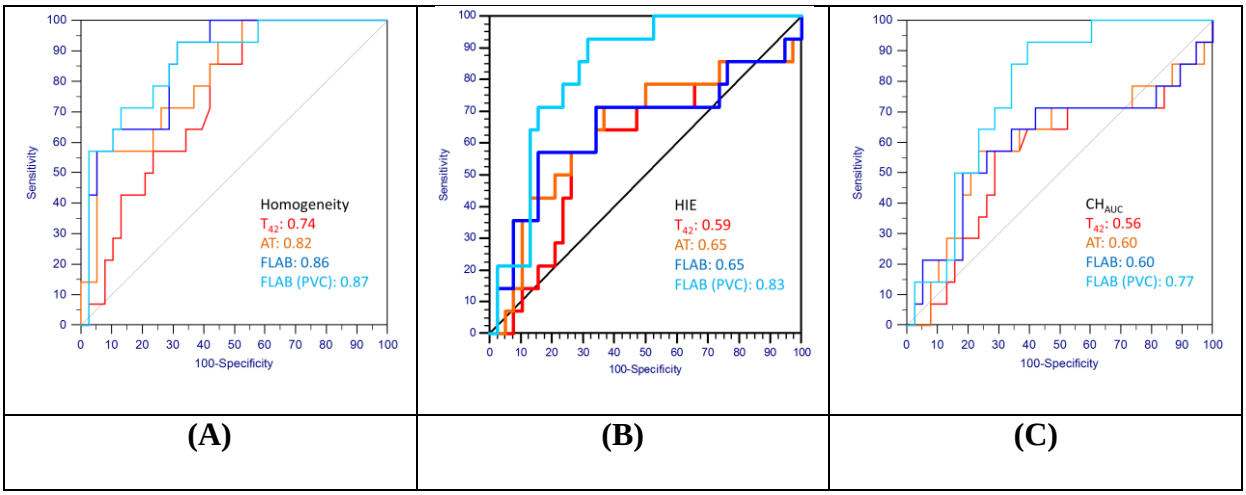


Figure 2. Workflow for the comparison of measurements depending on delineation method and PVC.



(A) **(B)** **(C)**

Figure 3. Examples of ROC curves for the identification of non-responders obtained using the three delineations and the PVC image, for (A) homogeneity, (B) SZV and (C) CH_{AUC}.

REFERENCES

1. Krause BJ, Schwarzenbock S, Souvatzoglou M. FDG PET and PET/CT. Recent Results Cancer Res. 2013;187:351-69.
2. Jarritt PH, Carson KJ, Hounsell AR, Visvikis D. The role of PET/CT scanning in radiotherapy planning. Br J Radiol. 2006;79 Spec No 1:S27-35.
3. Herrmann K, Benz MR, Krause BJ, Pomykala KL, Buck AK, Czernin J. (18)F-FDG-PET/CT in evaluating response to therapy in solid tumours: where we are and where we can go. Q J Nucl Med Mol Imaging. 2011;55:620-32.
4. Liao S, Penney BC, Wroblewski K, Zhang H, Simon CA, Kampalath R, et al. Prognostic value of metabolic tumour burden on 18F-FDG PET in nonsurgical patients with non-small cell lung cancer. Eur J Nucl Med Mol Imaging. 2012;39:27-38.
5. Hatt M, Visvikis D, Pradier O, Cheze-le Rest C. Baseline (18)F-FDG PET image-derived parameters for therapy response prediction in ooesophageal cancer. Eur J Nucl Med Mol Imaging. 2011;38:1595-606.
6. Deron P, Mertens K, Goethals I, Rottey S, Duprez F, De Neve W, et al. Metabolic tumour volume. Prognostic value in locally advanced squamous cell carcinoma of the head and neck. Nuklearmedizin. 2011;50.
7. Melton GB, Lavely WC, Jacene HA, Schulick RD, Choti MA, Wahl RL, et al. Efficacy of preoperative combined 18-fluorodeoxyglucose positron emission tomography and computed tomography for assessing primary rectal cancer response to neoadjuvant therapy. J Gastrointest Surg. 2007;11:961-9; discussion 9.
8. Hatt M, Groheux D, Martineau A, Espie M, Hindie E, Giacchetti S, et al. Comparison Between 18F-FDG PET Image-Derived Indices for Early Prediction of Response to Neoadjuvant Chemotherapy in Breast Cancer. J Nucl Med. 2013.
9. Lee HY, Hyun SH, Lee KS, Kim BT, Kim J, Shim YM, et al. Volume-based parameter of 18)F-FDG PET/CT in malignant pleural mesothelioma: prediction of therapeutic response and prognostic implications. Ann Surg Oncol. 2010;17:2787-94.
10. Cazaentre T, Morschhauser F, Vermandel M, Betrouni N, Prangere T, Steinling M, et al. Pre-therapy 18F-FDG PET quantitative parameters help in predicting the response to radioimmunotherapy in non-Hodgkin lymphoma. Eur J Nucl Med Mol Imaging. 2010;37:494-504.
11. 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. 2011.
12. Visvikis D, Hatt M, Tixier F, Cheze Le Rest C. The age of reason for FDG PET image-derived indices. Eur J Nucl Med Mol Imaging. 2012.
13. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RG, Granton P, et al. Radiomics: Extracting more information from medical images using advanced feature analysis. Eur J Cancer. 2012.
14. Chicklore S, Goh V, Siddique M, Roy A, Marsden PK, Cook GJ. Quantifying tumour heterogeneity in 18F-FDG PET/CT imaging by texture analysis. Eur J Nucl Med Mol Imaging. 2013;40:133-40.
15. Davnall F, Yip CS, Ljungqvist G, Selmi M, Ng F, Sanghera B, et al. Assessment of tumour heterogeneity: an emerging imaging tool for clinical practice? Insights Imaging. 2012;3:573-89.
16. Cook GJ, Yip C, Siddique M, Goh V, Chicklore S, Roy A, et al. Are Pretreatment 18F-FDG PET Tumour Textural Features in Non-Small Cell Lung Cancer Associated with Response and Survival After Chemoradiotherapy? J Nucl Med. 2013;54:19-26.

17. O'Sullivan F, Wolsztynski E, O'Sullivan J, Richards T, Conrad E, Eary J. A Statistical Modeling Approach to the Analysis of Spatial Patterns of FDG-PET Uptake in Human Sarcoma. IEEE Trans Med Imaging. 2011.
18. Tixier F, Le Rest CC, Hatt M, Albarghach N, Pradier O, Metges JP, et al. Intratumour heterogeneity characterized by textural features on baseline 18F-FDG PET images predicts response to concomitant radiochemotherapy in oesophageal cancer. J Nucl Med. 2011;52:369-78.
19. Tan S, Kligerman S, Chen W, Lu M, Kim G, Feigenberg S, et al. Spatial-Temporal [(18)F]FDG-PET Features for Predicting Pathologic Response of Oesophageal Cancer to Neoadjuvant Chemoradiation Therapy. Int J Radiat Oncol Biol Phys. 2012.
20. El Naqa I, Grigsby P, Apte A, Kidd E, Donnelly E, Khullar D, et al. Exploring feature-based approaches in PET images for predicting cancer treatment outcomes. Pattern Recognit. 2009;42:1162-71.
21. Miller TR, Pinkus E, Dehdashti F, Grigsby PW. Improved prognostic value of 18F-FDG PET using a simple visual analysis of tumour characteristics in patients with cervical cancer. J Nucl Med. 2003;44:192-7.
22. van Velden FH, Cheebsumon P, Yaqub M, Smit EF, Hoekstra OS, Lammertsma AA, et al. Evaluation of a cumulative SUV-volume histogram method for parameterizing heterogeneous intratumoural FDG uptake in non-small cell lung cancer PET studies. Eur J Nucl Med Mol Imaging. 2011.
23. Galavis PE, Hollensen C, Jallow N, Paliwal B, Jeraj R. Variability of textural features in FDG PET images due to different acquisition modes and reconstruction parameters. Acta Oncol. 2010;49:1012-6.
24. Tixier F, Hatt M, Cheze Le Rest C, Le Pogam A, Corcos L, Visvikis D. Reproducibility of tumour uptake heterogeneity characterization through textural feature analysis in 18F-FDG PET imaging. Journal of Nuclear Medicine. 2012;in press.
25. Visvikis D, Turzo A, Gouret A, Damine P, Lamare F, Bizais Y, et al. Characterisation of SUV accuracy in FDG PET using 3-D RAMLA and the Philips Allegro PET scanner. Journal of Nuclear Medicine. 2004;45:103.
26. Therasse P, Arbuck SG, Eisenhauer EA, Wanders J, Kaplan RS, Rubinstein L, et al. New guidelines to evaluate the response to treatment in solid tumours. 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;92:205-16.
27. Erdi YE, Mawlawi O, Larson SM, Imbriaco M, Yeung H, Finn R, et al. Segmentation of lung lesion volume by adaptive positron emission tomography image thresholding. Cancer. 1997;80:2505-9.
28. Nestle U, Kremp S, Schaefer-Schuler A, Sebastian-Welsch C, Hellwig D, Rube C, 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. 2005;46:1342-8.
29. 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. 2009;28:881-93.
30. 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. 2011;38:663-72.
31. Hatt M, Cheze-Le Rest C, Aboagye EO, Kenny LM, Rosso L, Turkheimer FE, et al. Reproducibility of 18F-FDG and 3'-deoxy-3'-18F-fluorothymidine PET tumour volume measurements. J Nucl Med. 2010;51:1368-76.

32. Hatt M, Cheze le Rest C, Descourt P, Dekker A, De Ruysscher D, Oellers M, et al. Accurate automatic delineation of heterogeneous functional volumes in positron emission tomography for oncology applications. Int J Radiat Oncol Biol Phys. 2010;77:301-8.
33. Boussion N, Cheze Le Rest C, Hatt M, Visvikis D. Incorporation of wavelet-based denoising in iterative deconvolution for partial volume correction in whole-body PET imaging. Eur J Nucl Med Mol Imaging. 2009;36:1064-75.
34. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics. 1988;44:837-45.