Visual Versus Quantitative Assessment of Intratumor ^{18}F-FDG PET Uptake Heterogeneity: Prognostic Value in Non–Small Cell Lung Cancer

Florent Tixier[1,2], Mathieu Hatt[2], Clemence Valla[1], Vincent Fleury[1], Corinne Lamour[3], Safaa Ezzouhri[1], Pierre Ingrand[4], Remy Perdrisot[1], Dimitris Visvikis[2], and Catherine Cheze Le Rest[1]

[1]Nuclear Medicine, CHU Milétrie, Poitiers, France; [2]INSERM, UMR 1101, LaTIM, Brest, France; [3]Department of Oncology, CHU Milétrie, Poitiers, France; and [4]Epidemiology and Biostatistics, CIC Inserm 1402, CHU Milétrie, Poitiers, France

The goal of this study was to compare visual assessment of intratumor ^{18}F-FDG PET uptake distribution with a textural-features (TF) automated quantification and to establish their respective prognostic value in non–small cell lung cancer (NSCLC). **Methods:** The study retrospectively included 102 consecutive patients. Only primary tumors were considered. Intratumor heterogeneity was visually scored (3-level scale [H_{visu}]) by 2 nuclear medicine physicians. Tumor volumes were automatically delineated, and heterogeneity was quantified with TF. Mean and maximum standardized uptake value were also included. Visual interobserver agreement and correlations with quantitative assessment were evaluated using the κ test and Spearman rank (ρ) coefficient, respectively. Association with overall survival and recurrence-free survival was investigated using the Kaplan–Meier method and Cox regression models. **Results:** Moderate correlations ($0.4 < \rho < 0.6$) between TF parameters and H_{visu} were observed. Interobserver agreement for H_{visu} was moderate ($\kappa = 0.64$, discrepancies in 27% of the cases). High standardized uptake value, large metabolic volumes, and high heterogeneity according to TF were associated with poorer overall survival and recurrence-free survival and remained an independent prognostic factor of overall survival with respect to clinical variables. **Conclusion:** Quantification of ^{18}F-FDG uptake heterogeneity in NSCLC through TF was correlated with visual assessment by experts. However, TF also constitutes an objective heterogeneity quantification, with reduced interobserver variability, and independent prognostic value potentially useful for patient stratification and management.

Key Words: ^{18}FDG-PET/CT; heterogeneity; textural features; prognosis; NSCLC

J Nucl Med 2014; 55:1235–1241
DOI: 10.2967/jnumed.113.133389

PET/CT imaging with ^{18}F-FDG is today a well-established tool for diagnostic oncology applications (*1*). Its exploitation for tumor delineation in the planning of radiotherapy treatment (*2*) or monitoring of response to therapy (*3,4*) is increasing. For non–small cell lung cancer (NSCLC), ^{18}F-FDG PET/CT image quantification has been shown to provide prognostic information. PET image–derived features, including metabolically active tumor volume (MATV), mean standardized uptake value (SUV_{mean}), and total lesion glycolysis (TLG, defined as MATV × SUV_{mean}), have been shown to provide an accurate assessment of tumor burden with potentially higher prognostic value than standard maximum SUV (SUV_{max}), for both surgical and nonsurgical patients (*5–9*).

Intratumor ^{18}F-FDG uptake heterogeneity has been associated with treatment failure (*10*), and its quantification has recently generated interest (*11–14*), including in lung cancer (*15*). Several methodologies have been proposed to assess intratumor heterogeneity, including visual evaluation (*16*), SUV coefficient of variation (SUV_{COV}) (*17*), area under the curve of the cumulative histogram (CH_{AUC}) (*18*), and textural-features (TF) analysis (*17,19*).

TF analysis can generate many features quantifying heterogeneity within a delineated MATV. Recent studies have identified a few of these features that are robust to the clinical range of reconstruction algorithms and acquisitions protocols (*20*), the delineation step, or the partial-volume effects (*21*), reaching similar or better physiologic reproducibility than SUV (*22*).

However, to date there has been no study investigating whether the quantitative heterogeneity assessment of the intratumor activity distribution through TF represents an added value relative to a visual assessment by nuclear medicine physicians. This study was therefore designed to compare a visual and qualitative tumor heterogeneity assessment with a characterization through TF analysis and to assess their respective prognostic value in NSCLC.

MATERIALS AND METHODS

Patients and PET/CT Imaging

One hundred eight consecutive nonmetastatic patients with newly diagnosed NSCLC between 2008 and 2011 were retrospectively considered. The institutional review board approved this retrospective study, and the requirement to obtain informed consent was waived. Only primary tumors with a MATV larger than 3 cm^3 (which, assuming a spheric shape, corresponds to an ~1.8-cm diameter) were included due to the inability of PET to characterize tracer distribution within smaller tumors because of its limited spatial resolution. As a result, 102 patients (79 men; mean age, 64 y) were included (Table 1). There were 49 squamous cell carcinomas and 53 adenocarcinomas.

All patients underwent an ^{18}F-FDG PET/CT scan before initiating treatment as part of the routine staging procedure within a maximum of 2 wk from diagnosis. Patients fasted for at least 6 h, and glucose levels were less than 8 mmol/L before injection of ^{18}F-FDG (5 MBq/kg; 425 ± 95;

Received Dec. 5, 2013; revision accepted Mar. 25, 2014.
For correspondence or reprints contact: Mathieu Hatt, INSERM, UMR 1101, LaTIM, CHRU Morvan, 2 Avenue Foch, 29609, Brest, France.
E-mail: hatt@univ-brest.fr
Published online Jun. 5, 2014.

TABLE 1
Patient Characteristics (n = 102)

Characteristic	n
Sex	
Male	79 (78)
Female	23 (22)
Age (y)	
Range	48–84
Median ± SD	64 ± 8.9
Smoker	
No	16 (16)
Yes	86 (84)
Treatment	
Surgery only	18 (18)
Chemotherapy only	12 (12)
Radiotherapy only	0 (0)
Surgery + chemotherapy	20 (19)
Surgery + radiotherapy	1 (1)
Chemotherapy + radiotherapy	42 (41)
Chemotherapy + radiotherapy + surgery	9 (9)
TNM stage	
T1	13 (13)
T2	55 (53)
T3	21 (21)
T4	13 (13)
N0	32 (31)
N1	23 (23)
N2	29 (28)
N3	18 (18)
M0	102 (100)
M1	0 (0)
Clinical stage	
I	18 (18)
II	30 (29)
III	54 (53)
IV	0 (0)
Histology	
Adenocarcinoma	53 (52)
Squamous cell carcinoma	49 (48)

Data in parentheses are percentages.

range, 223–690), administered at 60 ± 4 min before data acquisition on a GEMINI PET/CT scanner (Philips). CT data were acquired first (120 kV and 100 mAs, no contrast enhancement). Three-dimensional PET data were subsequently acquired with 2 min per bed position, and images were reconstructed using CT-based attenuation correction and a 3-dimensional row-action maximum likelihood algorithm with a previously optimized protocol (2 iterations; relaxation parameter, 0.05; 5 mm in full width at half maximum 3-dimensional gaussian postfiltering; 4 × 4 × 4 mm voxels grid sampling) (*23*). SUVs were normalized using the patient body weight.

Treatment and Follow-up

Treatment consisted of surgery for 48 patients (surgery only [n = 18] or adjuvant chemotherapy [n = 20], radiochemotherapy [n = 9], or radiotherapy [n = 1] followed by surgery), whereas 54 patients did not undergo surgery (chemotherapy [n = 12] or concomitant chemoradiotherapy [n = 42]) (Supplemental Fig. 1; supplemental materials are available at http://jnm.snmjournals.org). Chemotherapy consisted of 2–6 courses (median, 4) of cisplatine–carboplatine in association with taxol, navelbine, gemcitabine, or pemetrexed. The mean total radiotherapy dose was 59.4 Gy.

Patients were followed with clinical examination and CT every 3–4 mo. Overall survival (OS) and recurrence-free survival (RFS) were defined as the time between diagnosis and death (or last follow-up) and between the end of treatment and recurrence, respectively.

Image Analysis

Only the primary tumors were analyzed. Intratumor heterogeneity was first scored by 2 observers (denoted H_{visu} from here onward) and quantified through TF analysis after MATV automatic delineation.

Two experienced nuclear medicine physicians were asked to review all PET images, masked to the survival information and heterogeneity quantification. The PET window settings were set to display the intratumor voxels' values without saturation and were the same for both observers. The nuclear medicine physicians assigned to each primary tumor a score on a 3-point scale: 1 for homogeneous distribution, 2 for moderately heterogeneous, and 3 for highly heterogeneous. An alternative scale was considered by adding 2 sublevels, A for diffuse or B for focalized heterogeneity, in scores 2 and 3, hence resulting in a 5-score scale: 1, 2A, 2B, 3A, and 3B. For instance, centrally necrotic tumors were scored as 3B (Fig. 1; Supplemental Fig. 2).

The quantitative heterogeneity analysis was performed using in-house software. MATVs were first delineated with the fuzzy locally adaptive Bayesian (FLAB) algorithm (*24,25*) previously validated for accuracy, robustness, and reproducibility using simulated and clinical datasets, including homogeneous and heterogeneous MATVs (*26–28*). FLAB was exploited in this work using 2 or 3 classes to adequately cover the entire MATV, including low-uptake regions.

Intratumor uptake heterogeneity was quantified using textural features that have been previously shown as robust considering variability in image reconstruction and acquisition protocols (*20*) and physiologic reproducibility based on test–retest acquisitions (*22*). Considered local heterogeneity parameters quantifying intensity variations between each voxel and its immediate neighbors only, averaged over the entire volume, were entropy, homogeneity, and dissimilarity. Regional heterogeneity parameters calculated through analysis at the level of groups of voxels and areas of various sizes and intensities were high-intensity emphasis (HIE), size-zone variability (SZV), and zone percentage (ZP). A 64-gray-level quantization was used, and local features were computed over 13 directions (*19,22*).

Other global parameters such as skewness or kurtosis were excluded considering their previously demonstrated poor robustness (*20*) or physiologic reproducibility (*22*). CH_{AUC} and HIE were computed after edge-preserving filtering (*29*) and partial-volume effect correction (*30*) were applied, as it has been previously shown that such preprocessing is necessary for these parameters (*18,21*). Additional details about the FLAB automated algorithm and textural-features calculations have been added in the supplemental material sections. Finally, SUV_{max}, SUV_{mean}, MATV, and TLG were also included for comparison purposes as they have been previously shown to provide prognostic value in NSCLC (*5–9*).

Statistical Analysis

Statistical analyses were performed using Medcalc (MedCalc Software). Interobserver agreement regarding H_{visu} was estimated using

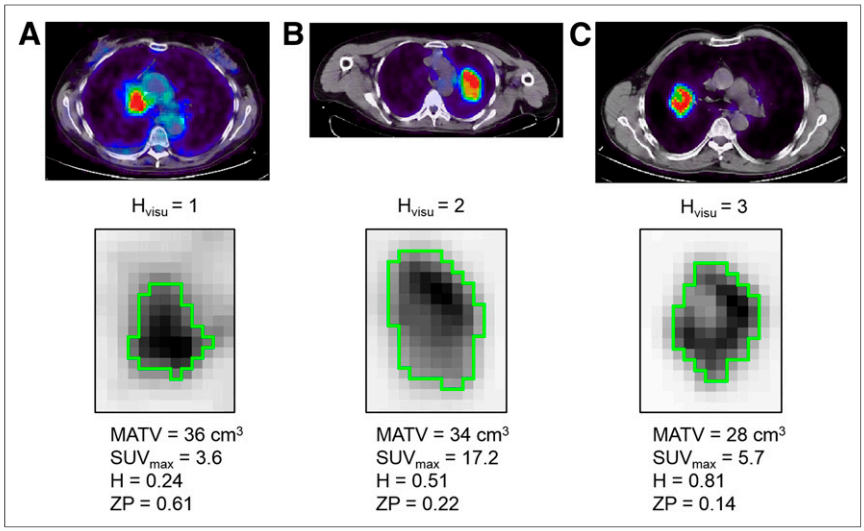

FIGURE 1. Illustration of primary lung tumors with H_{visu} values of 1, 2, and 3 (A, B, and C, respectively). Green contours are FLAB delineations, and examples of features are provided (values normalized between 0 and 1).

the weighted κ test with linear weights. Correlation between H_{visu} and quantitative features was assessed using Spearman ρ rank correlation. Variables with nonnormal distributions (e.g., MATV) were log-transformed.

Each feature's prognostic value for OS and RFS was assessed through univariate analysis using Kaplan–Meier curves and the log-rank test, with optimal cutoff values determined through receiver-operating-characteristic analysis. Statistically significant differences were considered for *P* values of less than 0.05 after Bonferroni adjustment for multiple testing.

Multivariate analysis was performed with Cox regression by including clinical variables along with image-derived features as continuous variables. Because there were 62 deaths, no more than 6 uncorrelated variables were included in separate models for OS.

RESULTS

Interobserver Agreement

Figure 1 illustrates 3 tumors with H_{visu} values of 1, 2, and 3. About half of the tumors were scored as highly heterogeneous. Interobserver agreement reached a moderately satisfactory level, with a weighted κ value of 0.64 (95% confidence interval [CI], 0.52–0.75), when the 3-point scale was used. The 2 observers were in agreement on the visual score for 74 of 102 tumors (73%). All discrepancies occurred between consecutive scores (1 and 2 or 2 and 3, never 1 and 3) (Supplemental Table 1). When the 5-level scale was used, interobserver agreement decreased to 0.58 (95% CI, 0.47–0.70), with the 2 observers agreeing on 62 of 102 (61%) (Supplemental Table 2). For the subsequent analysis, only the 3-point scale was considered, and only 1 value of H_{visu} obtained through consensus was used.

Correlation Between Visual Scoring and Quantitative Heterogeneity

Significant correlations were observed between H_{visu} and quantitative features (Fig. 2; Supplemental Table 3). The highest correlations were observed for local and regional TF, with ρ from 0.59 to 0.61 except ZP and HIE ($\rho = 0.44$, $P < 0.0001$, and $\rho = 0.20$, $P = 0.04$, respectively). CH_{AUC} was not correlated with H_{visu} ($\rho = 0.07$, $P = 0.5$), whereas SUV_{COV} showed a correlation similar to HIE ($\rho = 0.22$, $P = 0.027$) (Supplemental Fig. 3).

SUV_{max} and SUV_{mean} were not correlated with MATV ($r < 0.2$). On the other hand, TF showed variable levels of correlation with MATV, with an *r* of 0.6, −0.7, and 0.7 for entropy, dissimilarity, and homogeneity, respectively, and −0.5, −0.6, and −0.3 for SZV, ZP, and HIE, respectively. These correlations show that although heterogeneity is correlated with volume, such heterogeneity measurements could provide complementary information.

Survival Analysis

Median follow-up was 36.6 ± 11.8 mo. After surgery, 32 patients had evidence of recurrence. At the time of last follow-up, 39 patients were alive. Median OS and RFS were 18.4 (range, 1–58; 95% CI, 14.5–23.1) mo and 11.4 (range, 1–58; 95% CI, 6.8–18.4) mo, respectively.

In the univariate analysis (Table 2), surgery ($P = 0.006$), sex ($P = 0.02$), age ($P = 0.03$), and stage ($P = 0.001$) were significantly associated with OS, as well as standard SUV or volume parameters (*P* from 0.009 for SUV_{max} to < 0.0001 for MATV). For instance, patients with a MATV greater than 35 cm³ had a median survival of 10 mo versus 49 mo for those with a MATV of 35 cm³ or less. Patients who underwent surgery survived almost twice as long (median survival, 28 mo) as those who did not (median survival, 15 mo).

Higher visual heterogeneity ($H_{visu} = 3$) was associated with poorer OS, although the trend was not statistically significant. Patients with an H_{visu} of 3 had a median survival of 17 mo, whereas those with an H_{visu} less than 3 had a median survival of 20 mo. Higher heterogeneity assessed by TF was associated with poorer OS ($P \leq 0.007$ for local TF and <0.0001 for regional TF, except HIE). For example, entropy greater than 0.75 was associated with a median survival of 11 mo versus 49 mo for those with entropy of 0.75 or less. SUV_{COV} and CHU_{AUC} were not associated with OS ($P = 0.4$ and 0.9, respectively). Figure 3 provides examples of Kaplan–Meier curves for the different parameters considered.

In the multivariate Cox models that included surgery, sex, stage, SUV_{max} (or SUV_{mean}), and either MATV, TLG, or 1 heterogeneity

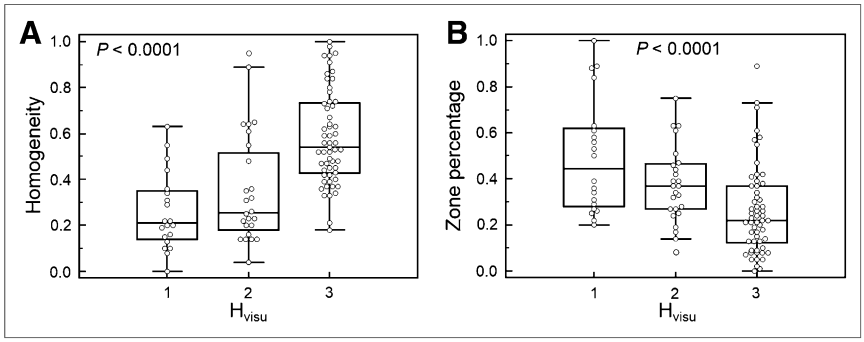

FIGURE 2. Illustration of distributions of homogeneity (A) and ZP (B) according to 3 levels of H_{visu}.

TABLE 2
OS Analysis (n = 102)

Parameter	Univariate P	Multivariate P
Clinical		
Surgery	0.006	—
Age	0.03	—
Sex	0.02	—
Smoker	0.9	—
Histology	0.3	—
Stage	0.001	<0.03
SUV and volumetric	0.009	—
SUV_{max}		
SUV_{mean}	0.008	—
MATV	<0.0001	0.0001
TLG	0.001	0.006
Global heterogeneity		
SUV_{COV}	0.4	—
CH_{AUC}	0.9	—
Visual heterogeneity		
$H_{visu} \geq 2$	1	—
$H_{visu} = 3$	1	—
Local heterogeneity (TF)		
Entropy	<0.0001	0.0001
Homogeneity	0.008	0.03
Dissimilarity	0.007	0.01
Regional heterogeneity (TF)		
HIE	0.9	—
SZV	<0.0001	0.0002
ZP	<0.0001	0.0001

TABLE 3
RFS Analysis (n = 48)

Parameter	Univariate P
Clinical	
Age	0.9
Sex	0.2
Smoker	0.1
Histology	0.7
Stage	0.1
SUV and volumetric	
SUV_{max}	1
SUV_{mean}	0.6
MATV	0.001
TLG	0.03
Visual heterogeneity	
$H_{visu} \geq 2$	1
$H_{visu} = 3$	0.3
Global heterogeneity	
SUV_{cov}	1
CH_{AUC}	0.8
Local heterogeneity (TF)	
Entropy	0.004
Homogeneity	0.005
Dissimilarity	0.001
Regional heterogeneity (TF)	
HIE	0.7
SZV	0.004
ZP	0.004

parameter (since these latter are correlated with each other), stage remained an independent prognostic factor but not surgery, sex, and SUV_{max} (or SUV_{mean}). MATV, as well as all heterogeneity quantification parameters obtained through TF, except HIE, remained independent prognostic factors with respect to stage (although not independent of each other).

The addition of risk factors allowed a better differentiation of patients' outcome. Patients with a large MATV combined with high local and regional tracer heterogeneity had a median survival of 9 mo and a 3-y survival rate of 0%, whereas the other group had a median survival of 49 mo and a 3-y survival rate of 50% (Fig. 4). The complementary value of TF heterogeneity to MATV can be shown by comparing the finer stratification of patients into 4 groups with statistically different outcomes (Fig. 5). MATV combined with entropy (Fig. 5B) led to different survival curves, compared with the use of MATV alone (Fig. 5A). MATVs greater than 50 cm^3 were associated with longer survival than MATVs between 35 and 50 cm^3. However, among volumes above 35 cm^3, those with entropy greater than 0.95 had significantly shorter survival (Fig. 5B).

Concerning RFS, none of the patients treated with chemoradiotherapy achieved complete response; therefore, only patients who underwent surgery were included (Table 3). None of the clinical variables were associated with RFS. MATV (P = 0.001) and TLG (P = 0.03) were significant prognostic factors of RFS, in contrast to SUV measurements. An H_{visu} of 3 was associated with a median RFS of 7 mo whereas median RFS was 25 mo for those with an H_{visu} of less than 3, although this trend was not statistically significant (P = 0.3). Higher TF-based heterogeneity was significantly associated with poorer RFS ($P \leq 0.004$), except when using HIE. For example, patients with dissimilarity greater than 0.57 had a median RFS of 6 mo versus 25 mo for those with dissimilarity of 0.57 or less. No multivariate analysis was performed for RFS because of the lack of uncorrelated variables statistically significant in the univariate analysis. Figure 6 provides examples of corresponding Kaplan–Meier curves.

DISCUSSION

There is currently an increasing interest in the use of PET image–derived features allowing the quantification of intratumor heterogeneity (*11,12*). Visual assessment may be considered as a simple and valuable way of scoring intratumor tracer distribution. In the present study, visual/qualitative and quantitative assessment of heterogeneity were simultaneously considered in the same NSCLC patient cohort and compared in terms of prognostic value.

First, our results suggest that quantitative parameters obtained through TF analysis are correlated with the visual assessment by nuclear medicine physicians. Our study also demonstrated an added value for TF analysis over visual assessment. The first advantage is that, because MATV and heterogeneity determination is fully

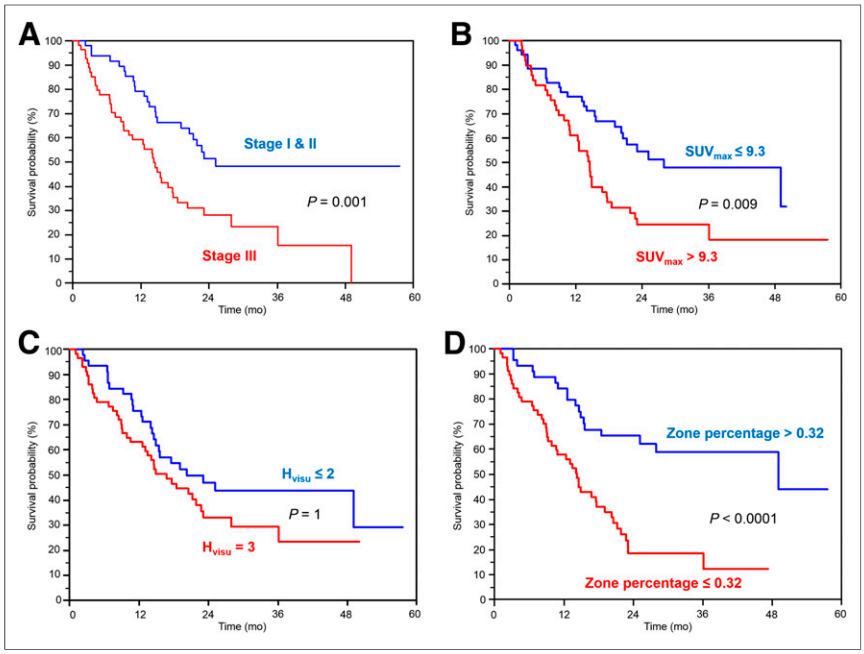

FIGURE 3. Examples of survival curves for OS (n = 102) according to stage (A), SUV_{max} (B), H_{visu} (C), and ZP (D).

automatic, the only inter- or intraobserver variability that might occur lies in the tumor location identification. Automated characterization is therefore likely to reduce interobserver variability associated with visual assessment, which as shown in this study was an issue for 27% of the cases. This was even worse when a larger visual scoring scale was considered, with a substantial decrease of interobserver agreement from 0.64 to 0.58 (from 73% to 61% of the 102 tumors). As a result, this finer scale was not further exploited, clearly demonstrating the difficulty in a fine visual characterization of intratumor tracer distribution.

A study comparing visual heterogeneity scoring, SUV_{COV}, and CH_{AUC} found high correlations (0.72 and 0.87 for SUV_{COV} and CH_{AUC}, respectively) (*31*). In our study, lower correlations were found between H_{visu} and TF, whereas SUV_{COV} and CH_{AUC} were not correlated with H_{visu}. Several factors may explain this difference. This previous study included only 9 gastrointestinal stromal tumors (GIST) and 12 malignant lymphomas (ML), manually delineated within a single 2-dimensional slice. A 4-value scale was used for visual scoring, and interobserver variability was not reported. The authors did not take into account the difference in volumes between GIST and ML. These volumes were also much larger (119 ± 102 cm^3) than in our study (48 ± 58 cm^3), suggesting that CH_{AUC} and SUV_{COV} might be appropriate to characterize high heterogeneity levels as found in large GIST and ML lesions but may not be sufficient to quantify finer heterogeneity differences found in smaller NSCLC tumors. This is also supported by the distributions of CH_{AUC} values, with a small range (SD = 0.05) in our study (0.32 ± 0.05), compared with those found in Watabe et al. (0.41 ± 0.14 for GIST and 0.64 ± 0.08 for ML) (*31*). Finally, FLAB delineation excluded areas with uptake similar to the background or lower, contrary to manual delineation, as in Watabe et al. However, only 4 cases presented areas with uptake low enough to be excluded and already presented large volumes and high heterogeneity. It is thus unlikely that the exclusion of the low-uptake region in these 4 cases might have had an impact on either the survival analysis or the correlation between H_{visu} and CH_{AUC}.

The second advantage of TF analysis is that it leads to additional prognostic value that may be useful for patient management. A stronger differentiation between groups of patients with different outcome was highlighted by combining parameters, although the parameters correlated with each other. As an example, patients with an MATV greater than 35 cm^3 combined with entropy greater than 0.74 and ZP of 0.32 or less had a median survival of 9 mo and a 3-y survival of 0%, whereas the others had a median survival of 60 mo and a 3-y survival of 50% (Fig. 4). As shown in Figure 5, despite the correlation between TF parameters and MATV, patient outcome could not be fully described using the MATV alone (Fig. 5A). This is illustrated in Figure 5B, showing that smaller but more heterogeneous lesions were associated with poorer OS than larger and more homogeneous ones.

A recent study suggested that entropy is unable to predict tracer uptake heterogeneity for tumors with a MATV less than 45 cm^3 (*32*). In our cohort, volumes were 48 ± 58 cm^3 (range, 3–415; median, 34). Our results only partly confirm this, because entropy provided additional information with respect to MATV in larger volumes (Fig. 5). MATV between 30 and 45 cm^3 exhibited an entropy between 0.55 and 0.81 (Supplemental Fig. 4A), with a weak

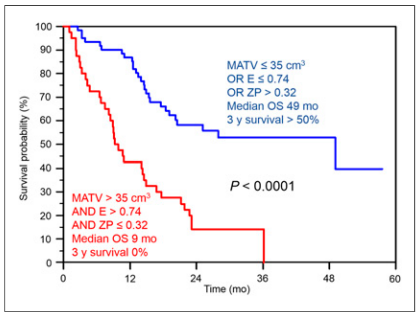

FIGURE 4. Survival curves for OS (n = 102) with stratification obtained according to combination of MATV, entropy, and ZP.

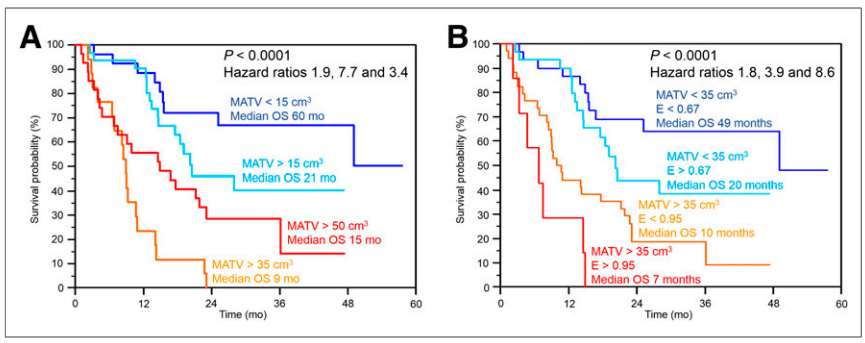

FIGURE 5. Differentiation of 4 different OS groups using MATV alone (A) or MATV (B) and substratification with entropy.

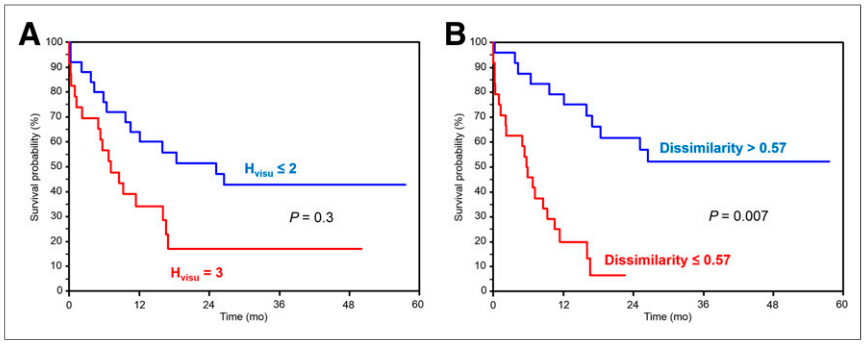

FIGURE 6. Examples of survival curves for RFS ($n = 48$) according to H_{visu} (A) and dissimilarity (B).

correlation for r of 0.57. For tumors with volumes less than 30 cm^3, this correlation was equal to 0.95, suggesting that for a MATV less than 30 cm^3 the information provided by entropy was indeed similar to that of volume. Finally, the proposed threshold value of 45 cm^3 may not be applicable for other heterogeneity measurements, because in our study different values of homogeneity or ZP were observed for similar volumes down to the lower limit of 3 cm^3 (Supplemental Figs. 4B and 4C).

Regarding the prognostic value of PET parameters in NSCLC, including SUV, MATV, and derived TLG (*5–9*) and TF (*15*), our findings are in line with previous results. The only study that investigated the prognostic value of ^{18}F-FDG PET heterogeneity in NSCLC was conducted on 53 patients with stage 3 and above, exclusively treated with combined chemoradiotherapy (nonsurgical patients) (*15*). Their results regarding MATV and TLG, showing no significant association with outcome, might be confounded by the fact that all their patients were stage 3 and 4 combined with a fixed-thresholding delineation approach. However, considering tracer distribution they found that heterogeneity quantified through busyness, coarseness, contrast, and complexity was significantly associated with RFS and OS. None of these parameters were included in our study because we have previously demonstrated their lower reproducibility (*22*). However, although our parameters were different and extracted from MATVs delineated using a more robust segmentation method, we also found that intratumor ^{18}F-FDG heterogeneity was significantly associated with outcome independently of sex, age, stage, and surgery. These results strengthen the assumption that higher intratumor activity distribution heterogeneity is associated with more complex vascularization and proliferation, properties that are currently not considered by standard treatments. Because higher heterogeneity was also a prognostic factor for the surgical group, heterogeneity may be also considered in identifying tumors with higher global aggressiveness.

Our study is limited by its retrospective nature, and the proposed automated quantification should be tested prospectively in a larger patient population. It included a mixture of different treatments (with or without surgery, radiotherapy, chemotherapy), leading to a heterogeneous cohort. However, the relatively large number of patients and events allowed taking into account surgery in the survival analysis. The large number of squamous cell carcinomas in our cohort was due to the exclusion of metastatic disease (mostly adenocarcinomas), because related patient management and prognosis differ greatly. We considered it more interesting to assess new parameters to refine patient stratification beyond the well-established TNM classification. A last limitation of our study is that we focused on the primary tumor. Including lymph nodes in the analysis could be of importance because of their impact on prognosis (*33*) but was outside the scope of the present study focusing on tracer distribution heterogeneity, which is meaningless to assess on small structures such as lymph nodes, considering the limited spatial resolution in PET imaging. The main objective of this study was to determine whether automated quantification through TF analysis could relate to visual assessment of PET images, and primary tumor analysis was sufficient within this context.

CONCLUSION

Automated quantification of intratumor ^{18}F-FDG uptake heterogeneity in NSCLC through textural-features analysis provides parameters that are correlated with visual analysis by experts. Nevertheless, our results also highlight several advantages of using automated quantification, including objective heterogeneity evaluation with reduced interobserver variability, and more clinically pertinent stratification through independent prognostic factors in NSCLC.

DISCLOSURE

The costs of publication of this article were defrayed in part by the payment of page charges. Therefore, and solely to indicate this fact, this article is hereby marked "advertisement" in accordance with 18 USC section 1734. No potential conflict of interest relevant to this article was reported.

REFERENCES

1. Krause BJ, Schwarzenbock S, Souvatzoglou M. FDG PET and PET/CT. *Recent Results Cancer Res.* 2013;187:351–369.
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–S35.
3. Wahl RL, Jacene H, Kasamon Y, Lodge MA. From RECIST to PERCIST: evolving considerations for PET response criteria in solid tumors. *J Nucl Med.* 2009;50 (suppl 1):122S–150S.
4. Herrmann K, Benz MR, Krause BJ, Pomykala KL, Buck AK, Czernin J. ^{18}F-FDG-PET/CT in evaluating response to therapy in solid tumors: where we are and where we can go. *Q J Nucl Med Mol Imaging.* 2011;55:620–632.
5. Chen HH, Chiu NT, Su WC, Guo HR, Lee BF. Prognostic value of whole-body total lesion glycolysis at pretreatment FDG PET/CT in non-small cell lung cancer. *Radiology.* 2012;264:559–566.
6. Kahraman D, Holstein A, Scheffler M, et al. Tumor lesion glycolysis and tumor lesion proliferation for response prediction and prognostic differentiation in patients with advanced non-small cell lung cancer treated with erlotinib. *Clin Nucl Med.* 2012;37:1058–1064.
7. Kim K, Kim S-J, Kim I-J, Kim YS, Pak K, Kim H. Prognostic value of volumetric parameters measured by F-18 FDG PET/CT in surgically resected non-small-cell lung cancer. *Nucl Med Commun.* 2012;33:613–620.
8. Liao S, Penney BC, Wroblewski K, et al. Prognostic value of metabolic tumor burden on ^{18}F-FDG PET in nonsurgical patients with non-small cell lung cancer. *Eur J Nucl Med Mol Imaging.* 2012;39:27–38.
9. Hyun SH, Ahn HK, Kim H, et al. Volume-based assessment by ^{18}F-FDG PET/CT predicts survival in patients with stage III non-small-cell lung cancer. *Eur J Nucl Med Mol Imaging.* 2014;41:50–58.
10. 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;38:987–991.
11. 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;39:1670–1672.

12. Lambin P, Rios-Velazquez E, Leijenaar R, et al. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer.* 2012;48:441–446.
13. Chicklore S, Goh V, Siddique M, Roy A, Marsden PK, Cook GJ. Quantifying tumour heterogeneity in ^{18}F-FDG PET/CT imaging by texture analysis. *Eur J Nucl Med Mol Imaging.* 2013;40:133–140.
14. Davnall F, Yip CS, Ljungqvist G, et al. Assessment of tumor heterogeneity: an emerging imaging tool for clinical practice? *Insights Imaging.* 2012;3:573–589.
15. Cook GJ, Yip C, Siddique M, et al. Are pretreatment ^{18}F-FDG PET tumor textural features in non-small cell lung cancer associated with response and survival after chemoradiotherapy? *J Nucl Med.* 2013;54:19–26.
16. Miller TR, Pinkus E, Dehdashti F, Grigsby PW. Improved prognostic value of ^{18}F-FDG PET using a simple visual analysis of tumor characteristics in patients with cervical cancer. *J Nucl Med.* 2003;44:192–197.
17. El Naqa I, Grigsby P, Apte A, et al. Exploring feature-based approaches in PET images for predicting cancer treatment outcomes. *Pattern Recognit.* 2009;42: 1162–1171.
18. van Velden FH, Cheebsumon P, Yaqub M, 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;38:1636–1647.
19. Tixier F, Le Rest CC, Hatt M, et al. Intratumor heterogeneity characterized by textural features on baseline ^{18}F-FDG PET images predicts response to concomitant radiochemotherapy in esophageal cancer. *J Nucl Med.* 2011;52:369–378.
20. 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;49:1012–1016.
21. Hatt M, Tixier F, Cheze Le Rest C, Pradier O, Visvikis D. Robustness of intratumour ^{18}F-FDG PET uptake heterogeneity quantification for therapy response prediction in oesophageal carcinoma. *Eur J Nucl Med Mol Imaging.* 2013;40: 1662–1671.
22. Tixier F, Hatt M, Le Rest CC, Le Pogam A, Corcos L, Visvikis D. Reproducibility of tumor uptake heterogeneity characterization through textural feature analysis in ^{18}F-FDG PET. *J Nucl Med.* 2012;53:693–700.
23. Visvikis D, Turzo A, Gouret A, et al. Characterisation of SUV accuracy in FDG PET using 3-D RAMLA and the Philips Allegro PET scanner [abstract]. *J Nucl Med.* 2004;45:103P–104P.
24. 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;28:881–893.
25. 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.* 2010;77:301–308.
26. 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;38:663–672.
27. Hatt M, Cheze-le Rest C, van Baardwijk A, Lambin P, Pradier O, Visvikis D. Impact of tumor size and tracer uptake heterogeneity in ^{18}F-FDG PET and CT non-small cell lung cancer tumor delineation. *J Nucl Med.* 2011;52:1690–1697.
28. Hatt M, Cheze-Le Rest C, Aboagye EO, et al. Reproducibility of ^{18}F-FDG and 3′-deoxy-3′-^{18}F-fluorothymidine PET tumor volume measurements. *J Nucl Med.* 2010;51:1368–1376.
29. Le Pogam A, Hanzouli H, Hatt M, Cheze Le Rest C, Visvikis D. Denoising of PET images by combining wavelets and curvelets for improved preservation of resolution and quantitation. *Med Image Anal.* 2013;17:877–891.
30. 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;36:1064–1075.
31. Watabe T, Tatsumi M, Watabe H, et al. Intratumoral heterogeneity of F-18 FDG uptake differentiates between gastrointestinal stromal tumors and abdominal malignant lymphomas on PET/CT. *Ann Nucl Med.* 2012;26:222–227.
32. Brooks FJ, Grigsby PW. The effect of small tumor volumes on studies of intratumoral heterogeneity of tracer uptake. *J Nucl Med.* 2014;55:37–42.
33. Cuaron J, Dunphy M, Rimner A. Role of FDG-PET scans in staging, response assessment, and follow-up care for non-small cell lung cancer. *Front Oncol.* 2012;2:208.