



HAL
open science

Development of a nomogram combining clinical staging with (18)F-FDG PET/CT image features in non-small-cell lung cancer stage I-III.

Marie-Charlotte Desseroit, Dimitris Visvikis, Florent Tixier, Mohamed Majdoub, Rémy Guillevin, Rémy Perdrisot, Catherine Cheze Le Rest, Mathieu Hatt

► To cite this version:

Marie-Charlotte Desseroit, Dimitris Visvikis, Florent Tixier, Mohamed Majdoub, Rémy Guillevin, et al.. Development of a nomogram combining clinical staging with (18)F-FDG PET/CT image features in non-small-cell lung cancer stage I-III.. *European Journal of Nuclear Medicine and Molecular Imaging*, 2016, 43 (8), pp.1477-85. 10.1007/s00259-016-3325-5 . inserm-01285687

HAL Id: inserm-01285687

<https://inserm.hal.science/inserm-01285687>

Submitted on 9 Mar 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Development of a nomogram combining clinical staging with ¹⁸F-FDG PET/CT image features in Non-Small Cell Lung Cancer stage I-III

Marie-Charlotte Desseroit^{1,2,3}, MSc., Dimitris Visvikis², PhD, Florent Tixier^{1,3}, PhD, Mohamed Majdoub², MSc., Rémy Perdrisot^{1,3}, MD, PhD, Rémy Guillevin^{3,4}, MD, PhD, Catherine Cheze Le Rest^{1,3*}, MD, PhD, Mathieu Hatt^{2*}, PhD.

¹ University Hospital, Nuclear Medicine, Poitiers, France.

² INSERM, UMR 1101, LaTIM, University of Brest, France.

³ University of Poitiers, Medical school, EE DACTIM, Poitiers, France

⁴ University hospital, Radiology, Poitiers, France

* equally contributed

Corresponding author:

Marie-Charlotte Desseroit INSERM, UMR 1101, LaTIM

CHRU Morvan, 2 avenue Foch

29609, Brest, France

Tel: +33(0)2.98.01.81.75 - Fax: +33(0)2.98.01.81.24

e-mail: Marie-Charlotte.Desseroit@etudiant.univ-brest.fr

Abstract:

Purpose: Our goal was to develop a nomogram by exploiting intra-tumor heterogeneity in CT and PET images of routine ^{18}F -FDG PET/CT acquisitions to identify patients with poorest prognosis.

Methods: 116 stage I-III NSCLC patients with staging ^{18}F -FDG PET/CT imaging were retrospectively included. Primary tumor volumes were delineated using the FLAB algorithm and 3D SlicerTM on PET and CT images respectively. PET and CT heterogeneities were quantified using texture analysis. Reproducibility of the CT features was assessed on a separate test-retest dataset. The stratification power of the PET/CT features was evaluated using Kaplan-Meier method and the log-rank test. The best standard metric (functional volume) was combined with the least redundant and prognostic PET/CT heterogeneity features to build the nomogram.

Results: PET entropy and CT zone percentage had the highest complementary value with clinical stage and functional volume. The nomogram increased the stratification amongst stage II and III patients, allowing the identification of patients with poorest prognosis (clinical stage III, large tumor volume, high PET heterogeneity and low CT heterogeneity).

Conclusion: Intra-tumor heterogeneity quantified through textural features on both CT and PET images from routine staging ^{18}F -FDG PET/CT acquisitions can be used to create a nomogram with high stratification power compared to staging alone.

Keywords: PET/CT, textural features, heterogeneity, prognosis, NSCLC

Introduction

Lung cancer is the first cause of cancer death for men and the second for women[1]. Fluorodeoxyglucose (^{18}F -FDG) Positron Emission Tomography/Computed Tomography (PET/CT) imaging is used routinely for diagnosis and staging in Non-Small Cell Lung Cancer (NSCLC)[2]. Tumor activity accumulation is currently assessed on PET images using standard uptake value (SUV) metrics such as maximum, peak or mean SUV. In the same context, the higher resolution CT images from the PET/CT acquisitions are currently exploited only for attenuation correction purposes and to localize PET tracer uptake. More recently, prognostic models based on different PET and CT image derived features have been designed[2]. The value of intra-tumor heterogeneity in NSCLC has been assessed either on ^{18}F -FDG[3–7], or on unenhanced/contrast-enhanced CT images[8–11] with promising results in terms of prediction of response, recurrence, or survival, often with higher and/or complementary value compared to standard volume or intensity metrics.

The term heterogeneity conveys a different meaning depending on the image modality. On ^{18}F -FDG PET, it refers to the spatial distribution of the radiotracer uptake, which could reflect at least partly underlying processes such as metabolism, hypoxia and cellular proliferation[12]. On the other hand, on unenhanced CT, it refers to the variability of tissue densities that could result from spatially varying vascularization and/or necrosis[9]. Finally, on contrast-enhanced CT, heterogeneity refers to the spatial variability of micro-vessels' density [8]. This heterogeneity of image voxel intensities can be quantified by different numerical methods including textural features (TF), intensity histogram or filtering combined with statistical and frequency-based methods[15].

In the vast majority of recent studies, intra-tumor heterogeneity characterization has been considered separately either on CT or PET images, but rarely simultaneously on both modalities within the context of assessing their complementary prognostic value. In one such study, ^{18}F -FDG heterogeneity was quantified using histogram-derived entropy only, combined with parameters extracted from both contrast-enhanced perfusion CT images and the attenuation CT component of PET/CT [8]. Another study built a model on a small dataset of 27 patients, combining PET and CT image derived features for predicting post-radiotherapy tumor progression [16]. In addition, previous studies investigating CT derived features exploited dosimetry [9] or diagnostic CT images [17] rather than the CT component of PET/CT acquisitions.

In NSCLC, patients with stage I disease have much better outcome than those with stage II and III. It is expected that PET/CT derived features are unlikely to change the evaluation of stage I patients, as these usually have small tumor volumes. There is therefore a limited amount of information that can be extracted from these. On the contrary, tumors in stage II and III exhibit a much higher range of volumes and heterogeneity, and it is thus expected these features could potentially improve prognosis stratification for these patients.

The purpose of this work was thus to investigate the potential added prognostic value compared to staging alone, of tumor and heterogeneity features extracted from both PET and CT components of routinely acquired PET/CT scans in NSCLC by developing a multi-parametric nomogram with better stratification among patients with stage II-III, compared to stage I.

Materials and methods

Patient cohort and imaging

116 patients with stage I-III NSCLC, diagnosed between 2008 and 2012 in the University Hospital of Poitiers, France, were retrospectively included. Treatment consisted of surgery and/or (chemo)radiotherapy. Fourteen patients had palliative chemotherapy only. Radiotherapy was with curative intent for all patients (mean dose 59.4 Gy).

A maximum of two weeks after diagnosis, all patients underwent an ^{18}F -FDG PET/CT scan on a Philips GEMINI PET/CT scanner (Philips Medical Systems, USA) following standard routine protocol: image acquisition began after 6 hours of fasting and 60 ± 5 min after injection of 5 MBq/kg of ^{18}F -FDG (424 ± 97 MBq, range 220-690 MBq). Non-contrast enhanced, non-respiratory gated (free breathing) CT images were acquired (120 kV, 100 mAs), with an in-plane resolution of $0.853 \times 0.853 \text{ mm}^2$ and a 5 mm slice thickness. PET data were acquired using 2 min per bed position and images were reconstructed using a 3D row-action maximum likelihood algorithm (RAMLA) (2 iterations, relaxation parameter equal to 0.05, 5 mm full-width-at-half-maximum 3-D Gaussian post-filtering, $4 \times 4 \times 4 \text{ mm}^3$ voxels). All PET images were corrected for attenuation using the acquired CT.

As far as clinical variables are concerned, in addition to clinical stage (taking into consideration T and N stage), treatment modality, age, gender, histology and smoker status were considered.

PET and CT images derived parameters

Only primary tumors were analyzed. MATVs were automatically delineated in 3D on PET images with the Fuzzy Locally Adaptive Bayesian (FLAB) algorithm. This procedure has been previously extensively validated for repeatability, accuracy and robustness in delineating both homogeneous and heterogeneous MATV [18–22]. A similarly validated and recommended approach exploiting the 3D Slicer™ software was used to semi-automatically delineate morphological tumor volumes in 3D on CT images [23].

Quantitative characterization of the tumors was subsequently performed by extracting several metrics from the delineated volumes: anatomical tumor volume (ATV) and associated intensity measurements from CT, as well as metabolically active tumor volume (MATV) and associated standard SUV metrics (SUV_{max} , SUV_{mean}) from PET were calculated. In both PET and CT volumes, intra-tumor heterogeneity was quantified in 3D using TF analysis after a quantization into 64 grey levels, as previously recommended [11,24–26]. Quantization was performed using equation 1:

$$I_q = q \times \frac{I - I_{min}}{I_{max} - I_{min}} \quad (1)$$

I_{max} and I_{min} denotes max and min Hounsfield units in CT and SUV in PET respectively and $q=64$. Second order TF (e.g. entropy, dissimilarity) were derived from a single co-occurrence matrix taking into consideration all 13 orientations simultaneously, as it was previously shown to lead to parameters with higher complementary value with respect to volume [25]. For PET heterogeneity features, only TF demonstrated in previous studies as robust [27], reproducible [24], and with potential complementary value to MATV [25] were considered: namely local entropy and dissimilarity, high intensity large area emphasis and zone percentage. CT heterogeneity histogram-derived metrics and TF were also calculated (supplemental

table 1). As a first step, their reproducibility was assessed in the present work using a specific dataset of test-retest acquisitions of patients with NSCLC, from the Reference Image Database to Evaluate Therapy Response (RIDER) study available at The Cancer Imaging Archive (TCIA) [28]. We have used this external test-retest database with unenhanced CT scans, given that our study is retrospective and as such it was not possible to have test-retest CT scans for the patients of the present cohort. To the best of our knowledge, the RIDER dataset is the only publicly available CT test-retest dataset.

CT features reproducibility was assessed with Bland-Altman analysis by reporting the mean and standard deviation (SD) of the differences between the two measurements. Lower (LRL) and upper (URL) reproducibility limits were also calculated as $\pm 1.96 \times \text{SD}$. Before being considered for inclusion in the nomogram, CT features were first selected based on their reproducibility. Features demonstrating $\text{SD} > 10\%$, corresponding to twice the variability of tumor volume measurements, were excluded from further analysis.

Nomogram construction

Statistical analyses were performed using MedCalcTM (MedCalc Software, Belgium). Correlations between features were quantified using Spearman rank (rs) coefficients. In order to reduce false positive rates and inflation of type I errors due to multiple testing, we reduced the number of features that were divided in four categories: 1) clinical variables, 2) volume and standard metrics, 3) PET heterogeneity and 4) CT heterogeneity. Our goal was to build a nomogram combining the best features of each category in order to improve the stratification provided by stage alone (figure 1A). Staging was considered a baseline on which to improve,

since staging was previously shown to provide significant stratification especially for stage I relative to stages II and III. Therefore in this study, PET/CT features were investigated for patients with stage II-III (N=87) in order to better identify amongst them those with the poorest prognosis, *i.e.* cumulating factors of poor prognosis. To this end, the selection of features to exploit was based on both their stratification power and their redundancy: a feature was considered for inclusion in the nomogram only if it had shown a prognostic value and had a correlation with other features <0.7 [29].

The stratification power of the different parameters was assessed using the Kaplan-Meier method and log-rank test, with cut-off thresholds determined through receiver operating characteristics (ROC) curve analysis according to the Youden's index [30]. The parameters were ranked according to their discriminating power based on hazard ratio (HR) and associated p-value. Correction for multiple testing was performed using the false discovery rate Benjamini-Hochberg step-up procedure, consisting in declaring positive discoveries at level α (here $\alpha=0.05$), among the $k=1\dots K$ tested variables ordered according to their p-values (p in increasing order), those ranked above the one satisfying the condition $p(k) \leq \frac{k}{K} \times \alpha$ [31].

In order to evaluate the improvement in prognosis stratification amongst stage II-III patients, by adding PET and/or CT heterogeneity features compared to clinical staging alone, the resulting Kaplan-Meier curves were compared using median OS in each group, hazard ratios (HR) and associated 95% confidence intervals (CI). Higher values of HR, with 1 being excluded of the 95% CI, indicated models with better stratification power.

Results

Survival and standard clinical staging

Mean follow-up was 26 months (range 1.5-74 months). Mean and median overall survival (OS) were 36.7 [95% CI 31.1-42.2] and 22.3 months [95% CI 15.6-36.2] respectively (figure 1-a). At last follow-up 71 patients were dead and 45 alive.

Clinical staging showed significantly better survival for patients with stage I (median survival not reached as only 20.7% of patients died) compared to stage II (HR 4.7) and III (HR 6.6) ($p < 0.0001$) (table 2). Stage III patients had worse survival than stage II patients (median OS of 21.2 vs. 14.5 months, HR=1.4) (figure 1-b). Amongst patients with stage II and III, N0 patients ($n=15$) had slightly better prognosis than N+ patients ($n=72$), although the trend was not significant ($p=0.5$), and N stage was highly correlated ($r=0.84$) with clinical staging. Finally, stage IIIA ($n=30$) patients had slightly better survival than IIIB ($n=27$), although the trend was not significant ($p=0.1$).

CT features reproducibility

In the test-retest dataset, the reproducibility of the ATV measurement was $-1.7 \pm 4.9\%$, with LRL and URL of -11.2% and $+7.9\%$. Histogram-derived entropy was the most reproducible feature with a mean difference of $-1.2 \pm 2.0\%$. Seven heterogeneity features were found reproducible enough ($SD < 10\%$, Appendix, table 1) to be further considered.

Delineation and tumor volumes

All tumors in the 116 patients could be delineated successfully. The mean ATV was 55.9 cm^3 (median 27.8, range 4-630.3) whereas MATV was smaller with a mean of 50.7 cm^3 (median 23.7, range 1-414.9). Stage I had significantly smaller volumes

(ATV $4.2 \pm 18.7 \text{ cm}^3$ and MATV $5.6 \pm 9.7 \text{ cm}^3$) than stage II and III (ATV $42.1 \pm 90.4 \text{ cm}^3$ and MATV $35.6 \pm 81.2 \text{ cm}^3$).

Features selection for nomogram

The only clinical feature with high prognostic value was clinical stage. Amongst treatment modalities (chemotherapy, radiotherapy, surgery or combinations), lack of surgery was a prognostic factor of poorer OS, although it was also correlated with clinical stage: 26 out of 29 (90%) stage I and 72% of stage II patients underwent surgery, whereas 12 stage III patients (10/30 stage IIIA and 2/27 IIIB) could benefit from surgery. Patients receiving only chemotherapy with palliative intent (N=14) had significantly poorer survival, however, as for surgery, this prognostic factor was highly correlated with stage (amongst the 14 patients, 12 were stage III, 1 stage II and 1 stage I). Overall, treatment information was therefore already mostly contained in the stage data. None of the other considered clinical variables had a prognostic value except a trend for age and gender (table 2). As a result no additional clinical variables were incorporated in the nomogram.

Amongst the image-derived “standard” metrics, there was none highly correlated with stage. Standard intensities from CTs had no prognostic value, whereas SUV_{max} and SUV_{mean} were not significant after correction for multiple testing. ATV and MATV were both prognostic factors, but were highly correlated ($r_s=0.80$, table 2, figure 2A). MATV was selected for tumor volume quantification and included in the nomogram with the cut-off value of 35 cm^3 , as it provided higher stratification than ATV (HR 2.4 vs. 1.9, table 2).

Amongst the PET heterogeneity features, co-occurrence entropy and HILAE had a correlation <0.7 with both MATV and stage. The entropy was selected to quantify PET heterogeneity in the nomogram (cut-off value of 7.3), with an associated HR of 1.9 ($p=0.01$) since HILAE was not significant ($p=0.2$). Amongst the 7 reproducible CT heterogeneity features, all exhibited a correlation <0.7 with MATV, clinical stage or PET entropy and five had a statistically significant prognostic value (table 2). CT Zone percentage (ZP) (cut-off value of 0.8) was selected since it had the best stratification power (HR=2.1, $p=0.003$).

The nomogram combining stage and MATV constituted our “standard” nomogram (*i.e.*, without any CT or PET heterogeneity consideration). It resulted in median OS of 21.8 and 7.5 months for lower and higher risk groups respectively, with HR of 4.3 and 11.1 compared to stage I patients (table 3, supplemental figure 2a), and HR of 2.6 between lower and higher risk groups. Adding heterogeneity parameters in the modelled to increased stratification, with OS of 21.8 vs. 8.5 (HR=2.9) and 21.2 vs. 6.5 (HR=2.7) for PET entropy and CT ZP respectively (table 3, supplemental figures 2b-c). Hazard ratios with respect to stage I patients similarly increased for PET entropy and CT ZP to 4.5 and 4.8 respectively for lower risk group, and to 13.1 and 12.8 for higher risk group. The best stratification was obtained by including all 4 parameters in the model (table 3, figure 1c) with a median OS of 21.2 months in lower risk group vs. 6.5 months for the HRG, resulting in a HR of 3.6 between the two groups. Hazard ratios compared to stage I patients increased to 4.9 and 17.4.

As features were added, the number of patients in higher risk group decreased from 57 using stage alone to 15 using the proposed nomogram. The percentage of events in higher risk group increased to reach 100% (*vs.* 77% with clinical stage only), whereas the median OS decreased from 14.5 to 6.5 months. The proposed

nomogram therefore led to a more useful stratification, patients with stage II-III being well stratified (HR=3.6) between a larger intermediate risk group (N=72) and a smaller higher risk group (N=15) with very short OS (95% CI 3.9 to 9.9 months) and an 18-months survival of 0%.

Discussion

Tumor heterogeneity quantification from either CT or PET/CT has been extensively considered in order to predict outcome for patients with NSCLC. However, most of previous studies focused on either CT or PET image derived features, rarely on both modalities simultaneously to evaluate their complementary value. In addition, recent studies exploiting CT for intra-tumor heterogeneity characterization have used dosimetry or diagnostic CTs, rather than the CT component of PET/CT [9,17]. The present study investigated the combination of CT and PET heterogeneity features using the two modality images acquired on a PET/CT. Several TFs computed on anatomical (CT) and functional (PET) derived tumor volumes were reproducible and associated with OS. Firstly, our results confirm the established prognostic value of clinical staging and primary tumor volume. Secondly, their addition to higher order PET and CT heterogeneity metrics led to a nomogram with higher stratification power.

With respect to tumor heterogeneity characterization from non-enhanced CT images, we evaluated the reproducibility of features using test-retest non-enhanced CT acquisitions of lung tumors. Bland-Altman analysis showed that the semi-automatic morphological tumor volume delineation step on CT images had high test-retest reproducibility ($-1.7 \pm 4.9\%$), a much lower variability than the $\pm 15\%$ previously shown for PET [22,32]. The reproducibility of histogram-derived and TFs computed on the same test-retest CT images has already been evaluated with concordance

correlation coefficients (CCC)[11,26], concluding that all histogram-derived parameters were reproducible, in contrast to our findings that identified only standard deviation, uniformity and entropy to be reproducible. This might be explained by the fact that Bland-Altman method could be more conservative than CCC. The majority of the investigated features (27 out of 34) in this work exhibited insufficient reproducibility.

Most CT derived TFs were correlated with the anatomical tumor volume, and some also moderately correlated with MATV. We emphasize that our goal was to add features with complementary value to parameters that already have established prognostic value (e.g. stage and volume). Therefore, adding to the nomogram PET and CT features correlated with volume or stage would not improve stratification. The chosen PET and CT features also had to be non-redundant: tissue density heterogeneity in CT (ZP) was not correlated with ^{18}F -FDG heterogeneity in PET (entropy) with $r_s = -0.11$ (supplemental figure 3).

Patients exhibiting higher intra-tumor ^{18}F -FDG uptake heterogeneity had poorer outcome irrespective of treatment modality, corroborating recent results [4,6,7,16]. ^{18}F -FDG accumulation is assumed to be mediated by multiple physiological processes: metabolism, vascularization, perfusion, aggressiveness, and hypoxia [33,34], which can be responsible for tumor heterogeneity. It has been recently shown that PET tumor heterogeneity can be associated with tumor vascularization [35]. It is also known that most hypoxic cells are more resistant [36].

CT tissue density heterogeneity within the primary tumor was also found to be associated with survival, lower heterogeneity (contrary to higher ^{18}F -FDG heterogeneity) being associated with poorer outcome (figure 3). Tumor heterogeneity

in CT is due to the presence of different types of cells (fibroblast, collapse, mucin, carcinoma cells, etc.) and tumor growth is accompanied by destruction of collagen fibrosis, which implicates decreasing of the air-containing region and increasing of the solid region[37]. On the one hand, our findings contradict other studies that associated higher tissues heterogeneity with poorer prognosis [9,17,38] on dosimetry or diagnostic CT with different image characteristics. On the other hand, our results are in agreement with another recent study that showed lower heterogeneity measured in the attenuation CT component of PET/CT was associated with poorer survival[8]. This result was obtained using entropy calculated in the intensity histogram, which was also included in our study and showed certain prognostic power, lower values being associated with worse survival. Although we have chosen for the nomogram zone percentage instead given that it had higher stratification power, the overall results between the two studies are indeed comparable. However, given the limited evidence associating higher homogeneity in the attenuation CT with poorer prognosis in NSCLC currently available, further studies are needed to draw definitive conclusions.

Compared to our previous works[4,5], 15 additional patients were included and OS was updated with follow-up until September 2014, and the entire statistical analysis was redone. It was shown in the present cohort that combining an increasing number of features improved the stratification of patients, resulting in a nomogram combining clinical stage, tumor functional volume, PET heterogeneity (local entropy) and CT heterogeneity (zone percentage), thus suggesting that complementary prognostic value can be extracted not only from ^{18}F -FDG PET, but also using the attenuation CT from routine PET/CT. This should be emphasized, as previous studies have focused on contrast-enhanced, diagnostic or dosimetry CT, not

the CT component of PET/CT. Our results contribute to the accumulating evidence that quantification of tumor heterogeneity in both PET and CT can help identifying patients with the poorest outcome in NSCLC. Our nomogram identified patients with very poor prognosis, who could be offered alternative strategies (targeted therapies or treatment intensification) at diagnosis using baseline PET/CT images that are acquired as a standard for staging in routine clinical management of NSCLC. The present nomogram was “trained” and fitted to the present patient cohort and will thus require validation in prospectively recruited patients.

One limitation is the variability in treatment modalities, although patients with metastatic disease were not considered. Yet, it allows suggesting that PET and CT heterogeneity may be survival predictors independently of the treatment modality: the nomogram obtained by excluding patients treated with palliative chemotherapy only showed similar stratification (supplemental figure 4). A second limitation concerns the evaluation of the CT features reproducibility and the comparison with the prognostic value extracted from perfusion or contrast-enhanced CTs. We could not perform the reproducibility evaluation in our cohort, nor could we compare with other CT image modalities, since in the present cohort, only routine PET/CT acquisitions were available. Regarding the reproducibility, we used a publicly available dataset that has been previously used [11,26,28].

Another limitation is that we did not characterize lymph nodes, only the primary tumor. Note there was no significant difference ($p > 0.2$) between the distributions of the PET or CT heterogeneity features of the primary tumor according to N stage (supplemental figure 5). Because staging incorporates lymph nodes status, the information is contained in the nomogram. It could be interesting to include the PET/CT image features of lymph nodes in a future work, although their volumes are

often smaller than primary tumors, which would limit the amount of exploitable texture information[25].

Another improvement could be the use of PET/CT images with more appropriate properties for textural features analysis thanks to improved reconstruction schemes (smaller voxels, higher resolution...) [39] and respiratory-gated acquisitions [40].

The last limitation concerns the combination of features and the statistical evaluation of the increased stratification. We considered a simple addition of risk factors to design the proposed nomogram, dichotomizing patients with stage II and III into a more specific higher risk group (patients having all four risk factors: stage III, large functional volume, high ^{18}F -FDGPET heterogeneity and low CT heterogeneity) and a larger, lower risk group (patients having some but not all risk factors). The use of methods from the field of machine learning[41], may allow building more robust nomograms, and could allow statistical comparison of the performance of the models, although this would require larger groups of patients. This will be the focus of future investigations, along with the validation of the proposed nomogram in a prospective cohort, which is currently being recruited.

Conclusions

Tumor heterogeneity quantified with textural features in the non-contrast enhanced CT and PET components of routine ^{18}F -FDGPET/CT images can provide complementary prognostic value in NSCLC. A 4-variable nomogram was designed and showed higher stratification power compared to standard clinical staging. Amongst patients with stage II and III, those with stage III, larger functional tumor volume, higher ^{18}F -FDGPET heterogeneity, and lower tissue density CT heterogeneity had very poor outcome, with an 18-months survival of 0% and a

median OS of 6.5 months only. These results should now be validated in a prospective study.

Compliance with Ethical Standards

Funding: This work has received a French government support granted to the CominLabs excellence laboratory and managed by the National Research Agency in the "Investing for the Future" program under reference ANR-10-LABX-07-01, and support from the city of Brest.

Disclosure of Conflicts of Interest: The authors declare that they have no conflict of interest.

Research involving Human Participants and/or Animals: All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. For this retrospective study formal consent is not required.

References

1. Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. *CA. Cancer J. Clin.* 2011;61:69–90.
2. Sauter AW, Schwenzler N, Divine MR, Pichler BJ, Pfannenberger C. Image-derived biomarkers and multimodal imaging strategies for lung cancer management. *Eur. J. Nucl. Med. Mol. Imaging.* 2015;
3. Cook GJR, Yip C, Siddique M, Goh V, Chicklore S, Roy A, et al. Are pretreatment 18F-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.
4. Tixier F, Hatt M, Valla C, Fleury V, Lamour C, Ezzouhri S, et al. Visual Versus Quantitative Assessment of Intratumor 18F-FDG PET Uptake Heterogeneity: Prognostic Value in Non-Small Cell Lung Cancer. *J Nucl Med.* 2014;55:1235–41.
5. Hatt M, Majdoub M, Vallières M, Tixier F, Le Rest CC, Groheux D, et al. 18F-FDG PET Uptake Characterization Through Texture Analysis: Investigating the Complementary Nature of Heterogeneity and Functional Tumor Volume in a Multi-Cancer Site Patient Cohort. *J Nucl Med.* 2015;56:38–44.
6. Cook GJR, O'Brien ME, Siddique M, Chicklore S, Loi HY, Sharma B, et al. Non-Small Cell Lung Cancer Treated with Erlotinib: Heterogeneity of (18)F-FDG Uptake at PET-Association with Treatment Response and Prognosis. *Radiology.* 2015;141309.
7. Pyka T, Bundschuh RA, Andratschke N, Mayer B, Specht HM, Papp L, et al. Textural features in pre-treatment [F18]-FDG-PET/CT are correlated with risk of local recurrence and disease-specific survival in early stage NSCLC patients receiving primary stereotactic radiation therapy. *Radiat. Oncol. Lond. Engl.* 2015;10:100.
8. Win T, Miles KA, Janes SM, Ganeshan B, Shastry M, Endozo R, et al. Tumor Heterogeneity and Permeability as Measured on the CT Component of PET/CT Predict Survival in Patients with Non-Small Cell Lung Cancer. *Clin. Cancer Res.* 2013;19:3591–9.
9. Aerts HJWL, Velazquez ER, Leijenaar RTH, Parmar C, Grossmann P, Cavalho S, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat. Commun.* 2014;5:4006.
10. Ganeshan B, Goh V, Mandeville HC, Ng QS, Hoskin PJ, Miles KA. Non-small cell lung cancer: histopathologic correlates for texture parameters at CT. *Radiology.* 2013;266:326–36.
11. Fried DV, Tucker SL, Zhou S, Liao Z, Mawlawi O, Ibbott G, et al. Prognostic value and reproducibility of pretreatment CT texture features in stage III non-small cell lung cancer. *Int. J. Radiat. Oncol. Biol. Phys.* 2014;90:834–42.
12. O'Connor JPB, Rose CJ, Waterton JC, Carano RAD, Parker GJM, Jackson A. Imaging Intratumor Heterogeneity: Role in Therapy Response, Resistance, and

Clinical Outcome. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* 2015;21:249–57.

13. Weber WA, Schwaiger M, Avril N. Quantitative assessment of tumor metabolism using FDG-PET imaging. *Nucl. Med. Biol.* 2000;27:683–7.

14. Willaime JMY, Turkheimer FE, Kenny LM, Aboagye EO. Quantification of intra-tumour cell proliferation heterogeneity using imaging descriptors of ¹⁸F fluorothymidine-positron emission tomography. *Phys. Med. Biol.* 2013;58:187.

15. Ganeshan B, Miles KA. Quantifying tumour heterogeneity with CT. *Cancer Imaging Off. Publ. Int. Cancer Imaging Soc.* 2013;13:140–9.

16. Vaidya M, Creach KM, Frye J, Dehdashti F, Bradley JD, El Naqa I. Combined PET/CT image characteristics for radiotherapy tumor response in lung cancer. *Radiother. Oncol. J. Eur. Soc. Ther. Radiol. Oncol.* 2012;102:239–45.

17. Grove O, Berglund AE, Schabath MB, Aerts HJWL, Dekker A, Wang H, et al. Quantitative computed tomographic descriptors associate tumor shape complexity and intratumor heterogeneity with prognosis in lung adenocarcinoma. *PloS One.* 2015;10:e0118261.

18. 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.

19. Hatt M, Cheze le Rest C, Descourt P, Dekker A, De Ruyscher 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.

20. 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.

21. Arens AIJ, Troost EGC, Hoeben BAW, Grootjans W, Lee JA, Grégoire V, et al. Semiautomatic methods for segmentation of the proliferative tumour volume on sequential FLT PET/CT images in head and neck carcinomas and their relation to clinical outcome. *Eur. J. Nucl. Med. Mol. Imaging.* 2014;41:915–24.

22. Heijmen L, de Geus-Oei LF, de Wilt JH, Visvikis D, Hatt M, Visser EP, et al. Reproducibility of functional volume and activity concentration in ¹⁸F-FDG PET/CT of liver metastases in colorectal cancer. *Eur J Nucl Med Mol Imaging.* 2012;39:1858–67.

23. Parmar C, Rios Velazquez E, Leijenaar R, Jermoumi M, Carvalho S, Mak RH, et al. Robust Radiomics feature quantification using semiautomatic volumetric segmentation. *PloS One.* 2014;9:e102107.

24. 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 ¹⁸F-FDG PET. *J Nucl Med.* 2012;53:693–700.

25. Hatt M, Majdoub M, Vallières M, Tixier F, Le Rest CC, Groheux D, et al. 18F-FDG PET Uptake Characterization Through Texture Analysis: Investigating the Complementary Nature of Heterogeneity and Functional Tumor Volume in a Multi-Cancer Site Patient Cohort. *J. Nucl. Med. Off. Publ. Soc. Nucl. Med.* 2015;56:38–44.
26. Hunter LA, Krafft S, Stingo F, Choi H, Martel MK, Kry SF, et al. High quality machine-robust image features: Identification in nonsmall cell lung cancer computed tomography images. *Med. Phys.* 2013;40:121916.
27. Hatt M, Tixier F, Cheze Le Rest C, Pradier O, Visvikis D. Robustness of intratumour ¹⁸F-FDG PET uptake heterogeneity quantification for therapy response prediction in oesophageal carcinoma. *Eur. J. Nucl. Med. Mol. Imaging.* 2013;40:1662–71.
28. Armato SG 3rd, Meyer CR, Mcnitt-Gray MF, McLennan G, Reeves AP, Croft BY, et al. The Reference Image Database to Evaluate Response to therapy in lung cancer (RIDER) project: a resource for the development of change-analysis software. *Clin. Pharmacol. Ther.* 2008;84:448–56.
29. Dancy CP, Reidy J. *Statistics Without Maths for Psychology*. 5 edition. Harlow, England ; New York: Prentice Hall; 2011.
30. Youden WJ. Index for rating diagnostic tests. *Cancer.* 1950;3:32–5.
31. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B Methodol.* 1995;57:289–300.
32. 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 tumor volume measurements. *J Nucl Med.* 2010;51:1368–76.
33. Vesselle H, Schmidt RA, Pugsley JM, Li M, Kohlmyer SG, Vallires E, et al. Lung cancer proliferation correlates with [F-18]fluorodeoxyglucose uptake by positron emission tomography. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* 2000;6:3837–44.
34. Kunkel M, Reichert TE, Benz P, Lehr H-A, Jeong J-H, Wieand S, et al. Overexpression of Glut-1 and increased glucose metabolism in tumors are associated with a poor prognosis in patients with oral squamous cell carcinoma. *Cancer.* 2003;97:1015–24.
35. Tixier F, Groves AM, Goh V, Hatt M, Ingrand P, Le Rest CC, et al. Correlation of Intra-Tumor 18F-FDG Uptake Heterogeneity Indices with Perfusion CT Derived Parameters in Colorectal Cancer. *PLoS ONE.* 2014;9:e99567.
36. Rajendran JG, Wilson DC, Conrad EU, Peterson LM, Bruckner JD, Rasey JS, et al. [(18)F]FMISO and [(18)F]FDG PET imaging in soft tissue sarcomas: correlation of hypoxia, metabolism and VEGF expression. *Eur. J. Nucl. Med. Mol. Imaging.* 2003;30:695–704.

37. Ikehara M, Saito H, Yamada K, Oshita F, Noda K, Nakayama H, et al. Prognosis of small adenocarcinoma of the lung based on thin-section computed tomography and pathological preparations. *J. Comput. Assist. Tomogr.* 2008;32:426–31.
38. Ganeshan B, Panayiotou E, Burnand K, Dizdarevic S, Miles K. Tumour heterogeneity in non-small cell lung carcinoma assessed by CT texture analysis: a potential marker of survival. *Eur. Radiol.* 2012;22:796–802.
39. Yan J, Chu-Shern JL, Loi HY, Khor LK, Sinha AK, Quek ST, et al. Impact of Image Reconstruction Settings on Texture Features in 18F-FDG PET. *J Nucl Med.* 2015;56:1667–73.
40. Yip S, McCall K, Aristophanous M, Chen AB, Aerts HJWL, Berbeco R. Comparison of texture features derived from static and respiratory-gated PET images in non-small cell lung cancer. *PloS One.* 2014;9:e115510.
41. Parmar C, Grossmann P, Bussink J, Lambin P, Aerts HJWL. Machine Learning methods for Quantitative Radiomic Biomarkers. *Sci. Rep.* 2015;5:13087.

Table 1: Patients characteristics

Characteristic		No. of patients (N=116)
Gender	Male	88
	Female	28
Age (y)	Range	40-84
	Mean + SD	63±9
Treatment	Surgery only	28
	Chemotherapy only	14
	Radiotherapy only	0
	Surgery + chemotherapy	21
	Surgery + radiotherapy	1
	Chemotherapy + radiotherapy	43
	Chemotherapy + radiotherapy + surgery	9
Clinical stage	I	29
	II	30
	III	57
	IV	0

Table 2: selection of features according to prognostic value and redundancy

Features		Prognostic p-value [hazard ratio]	Rank correlation with		
			Clinical Stage	MATV	PET entropy
Clinical	Age	0.029*			
	Gender	0.034*			
Standard metrics	1 st order features	<u>MATV</u>	0.0002 [2.5]	0.12	
		ATV	0.02 [1.9]	0.09	0.80
		SUV _{max}	0.043*		
		SUV _{mean}	0.04*		
PET Heterogeneity	Co-occurrence matrix	<u>Local entropy</u>	0.01 [1.9]	-0.01	0.31
		Dissimilarity	0.02 [1.8]	-0.14	-0.90
	Intensity size- zone matrix	Zone percentage	0.0016 [2.3]	-0.08	-0.81
CT Heterogeneity	1 st order (Intensity histogram)	Entropy	0.019 [1.8]	-0.08	-0.41
		Uniformity	0.016 [1.8]	0.08	0.43
		Standard Deviation	0.028 [1.7]	-0.06	-0.51
	Co-occurrence matrix	Entropy	0.01 [1.9]	-0.18	-0.38
	Intensity size- zone matrix	<u>Zone percentage</u>	0.003 [2.1]	-0.01	-0.39

* not significant after correction for multiple testing, significant p-value are in **bold**

Features in the nomogram are underlined.

Stratification of stage II-III patients (N=87) w.r.t. stage I (N=29, group 1)

Models	Lower risk (group 2)				Higher risk (group 3)				
	N	Median OS [95% CI]	Events (%)	HR w.r.t group 1 [95% CI]	N	Median OS [95% CI]	Events (%)	HR w.r.t group 1 [95% CI]	HR w.r.t group 2 [95% CI]
Clinical stage only	30	21.8 [13.1-36.2]	21 (70.0%)	4.7 [2.6-8.6]	57	14.5 [9.9-18.4]	44 (77.2%)	6.6 [3.8-11.3]	1.4 [0.8-2.5]
Stage & primary tumor MATV	56	21.8 [14.9-36.2]	37 (66.1%)	4.3 [2.6-7.2]	31	7.5 [4.7-14.1]	28 (90.3%)	11.1 [5.4-22.8]	2.6 [1.3-5.1]
Stage & primary tumor MATV & PET Entropy	63	21.8 [14.7-31.3]	42 (66.7%)	4.5 [2.7-7.4]	24	8.5 [4.0-14.1]	23 (95.8%)	13.1 [5.7-29.9]	2.9 [1.3-6.5]

Stage & primary tumor MATV & CT Zone			47				18		
Percentage	67	21.2 [14.5-30.4]	(70.1%)	4.8 [2.9-7.9]	20	6.5 [3.9-14.1]	(90.0%)	12.8 [5.2-31.7]	2.7 [1.1-6.3]
Stage & primary tumor MATV & PET Entropy & CT Zone			50				15		
Percentage	72	21.2 [14.5-30.4]	(69.4%)	4.9 [3.0-8.0]	15	6.5 [3.9-9.9]	(100%)	17.4 [5.7-52.8]	3.6 [1.2-10.5]

Table 3: Kaplan-Meier analysis demonstrating increasing stratification power by adding PET/CT features (metabolic volume >35cm³, PET entropy >7.3, CT zone percentage <0.8) to stage (III vs. II). w.r.t.: with respect to.

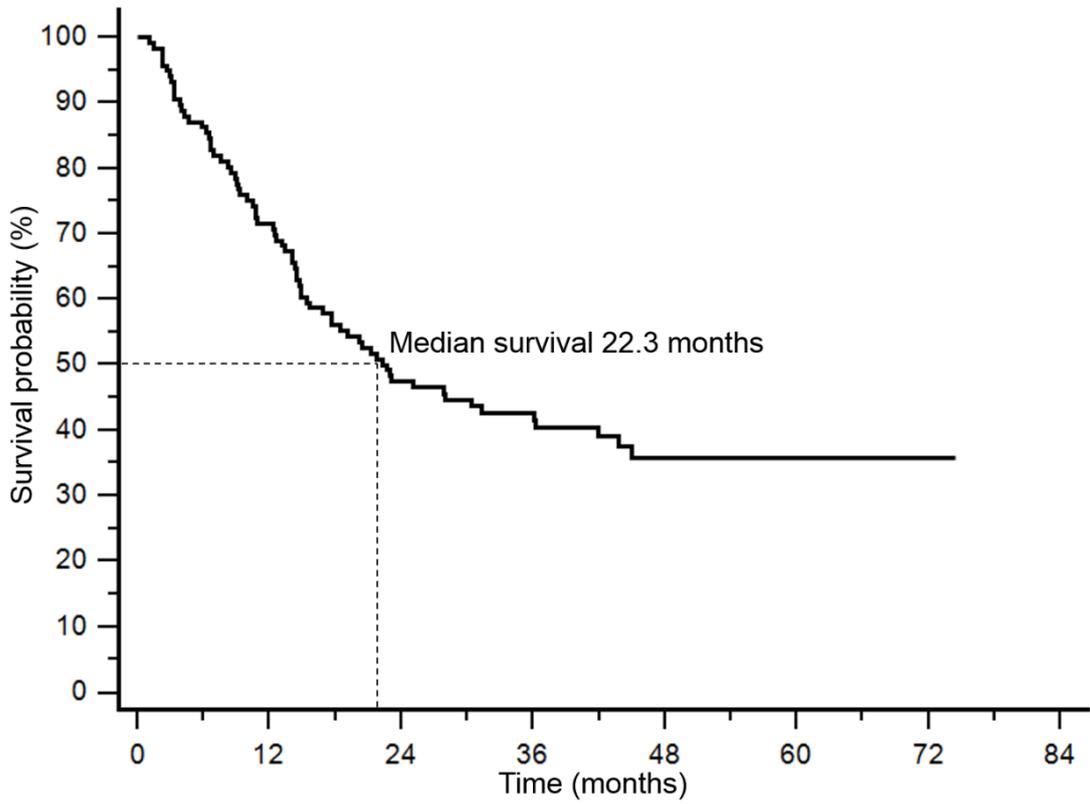


Figure 1-a

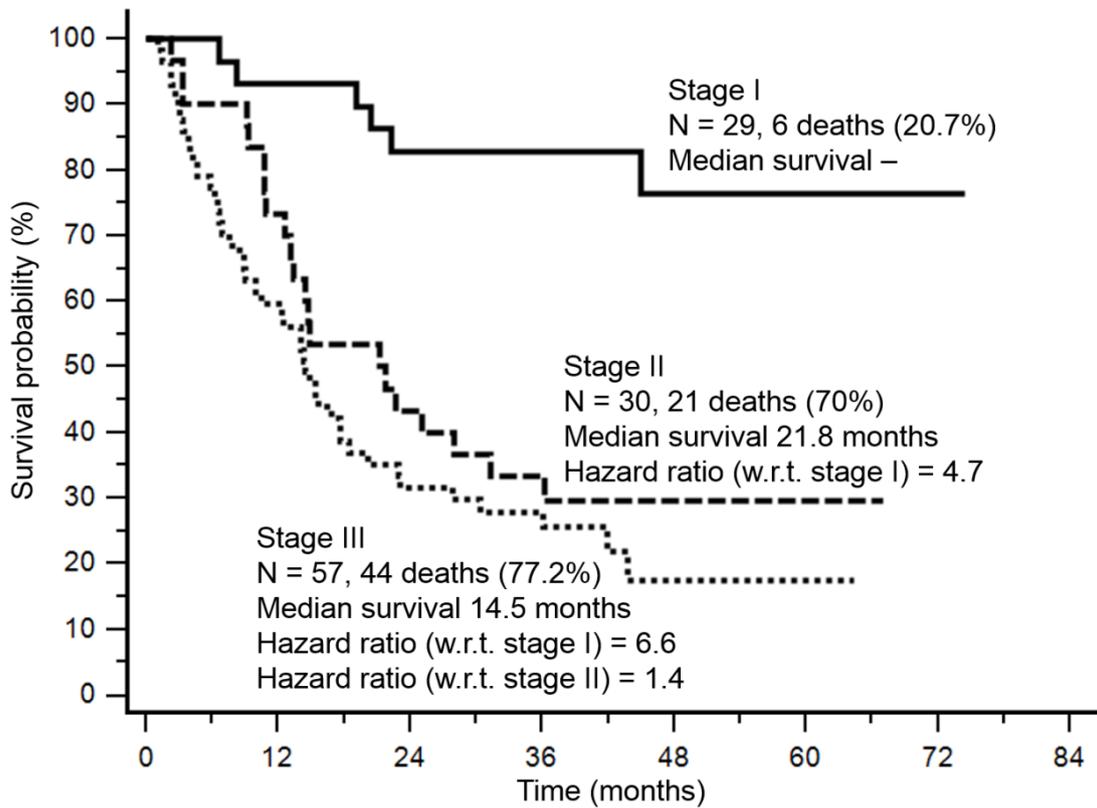


Figure 1-b

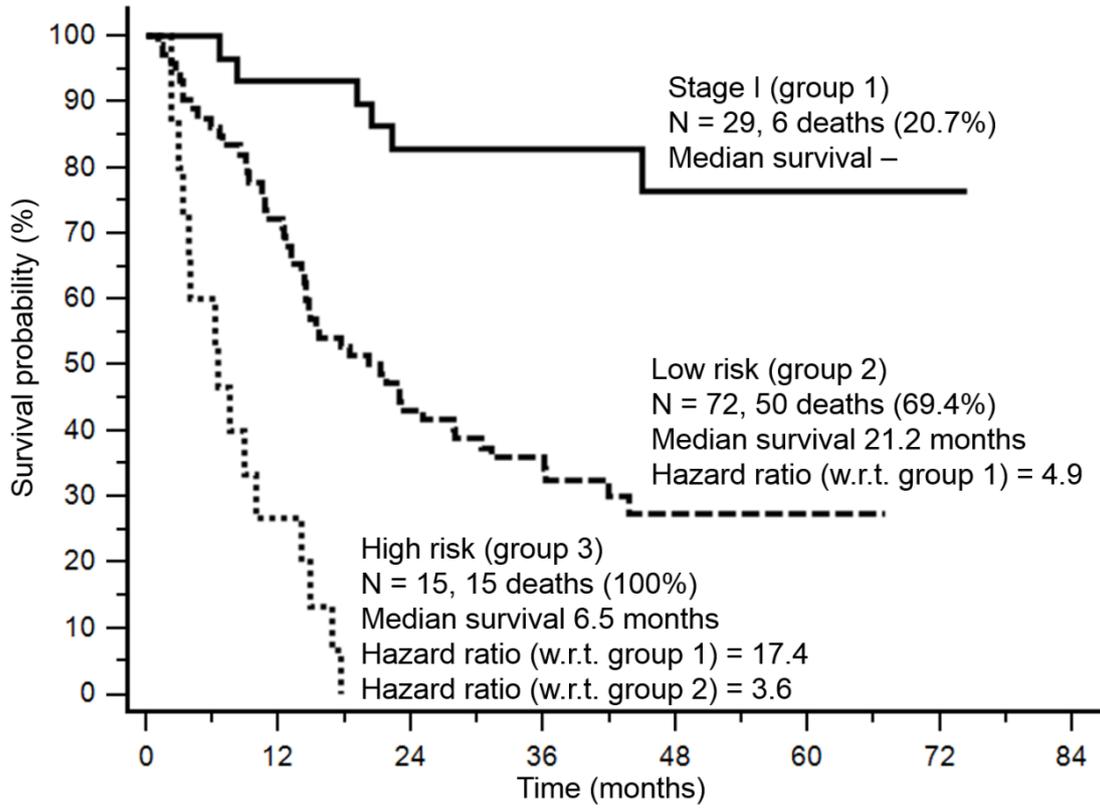


Figure 1-c

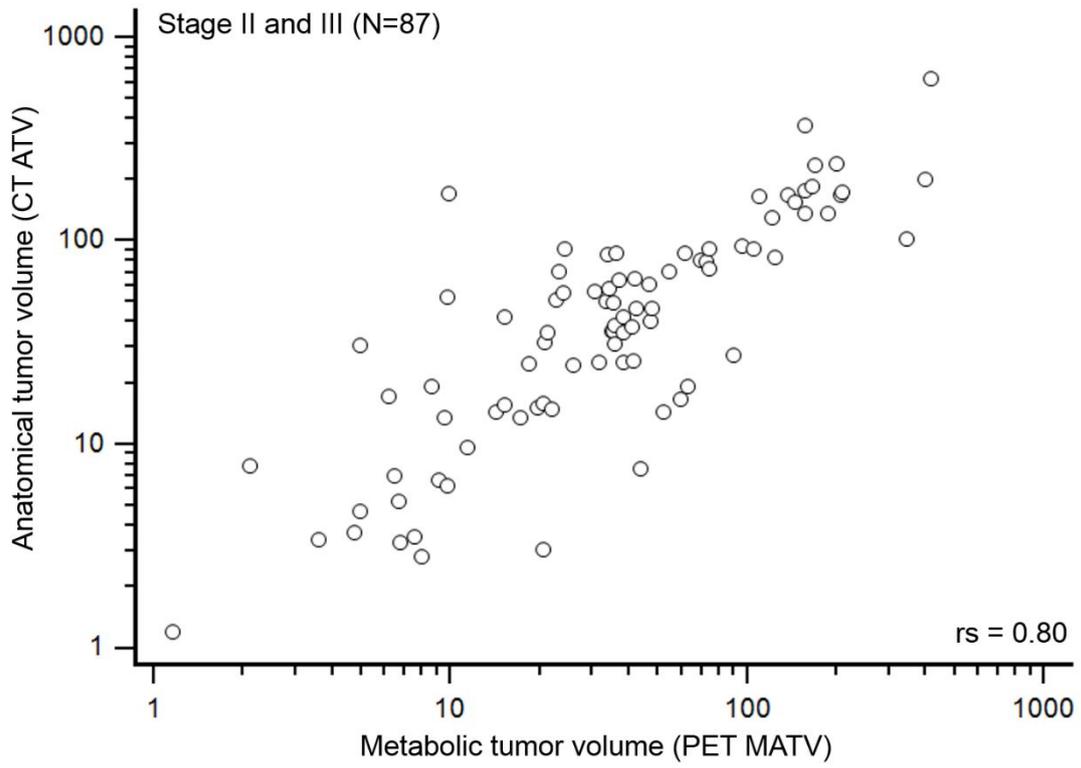


Figure 2-a

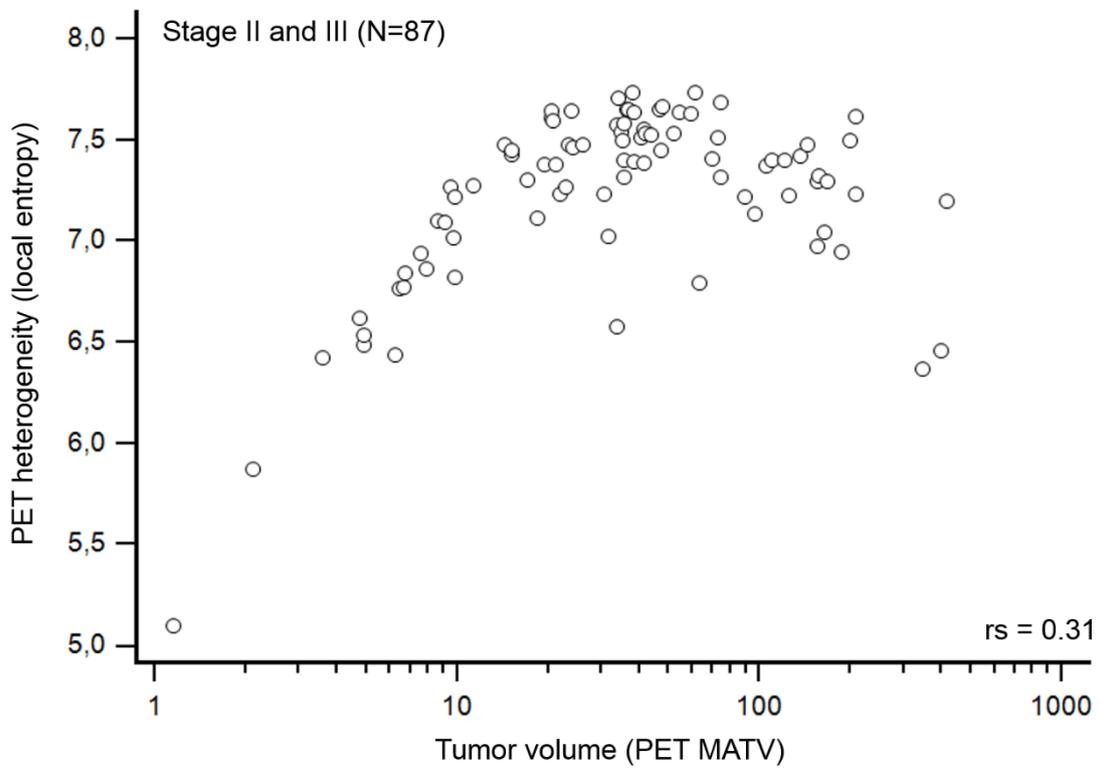


Figure 2-b

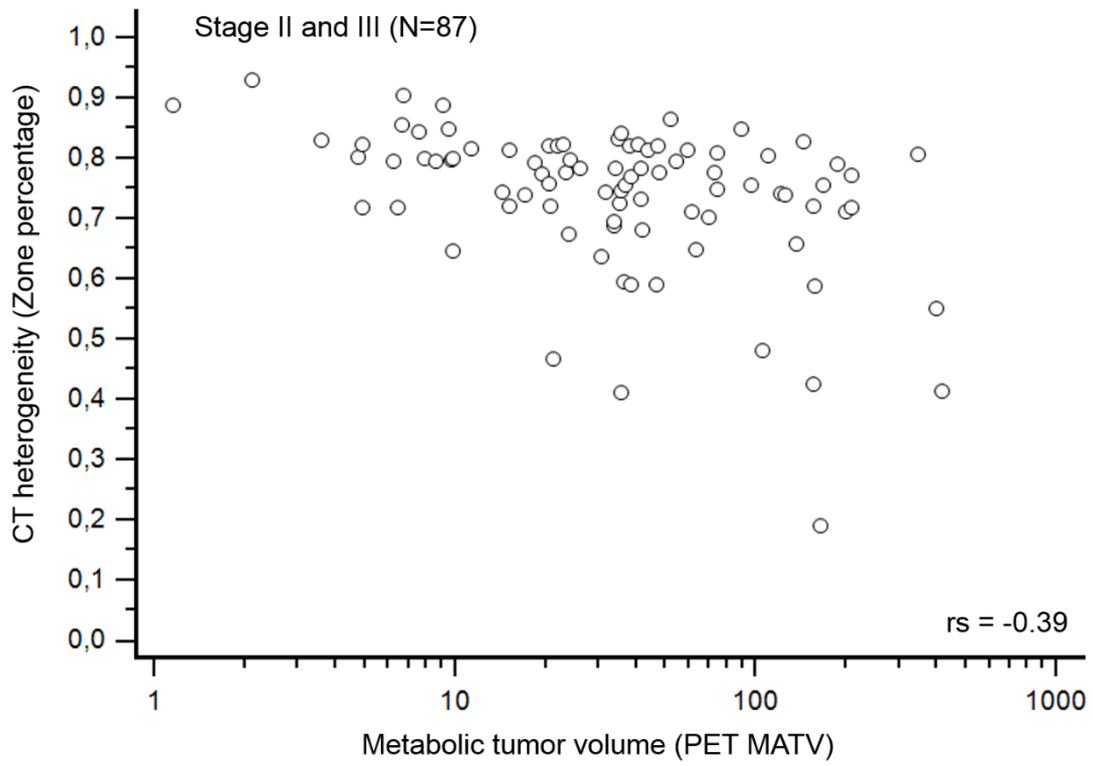


Figure 2-c

Figure 3

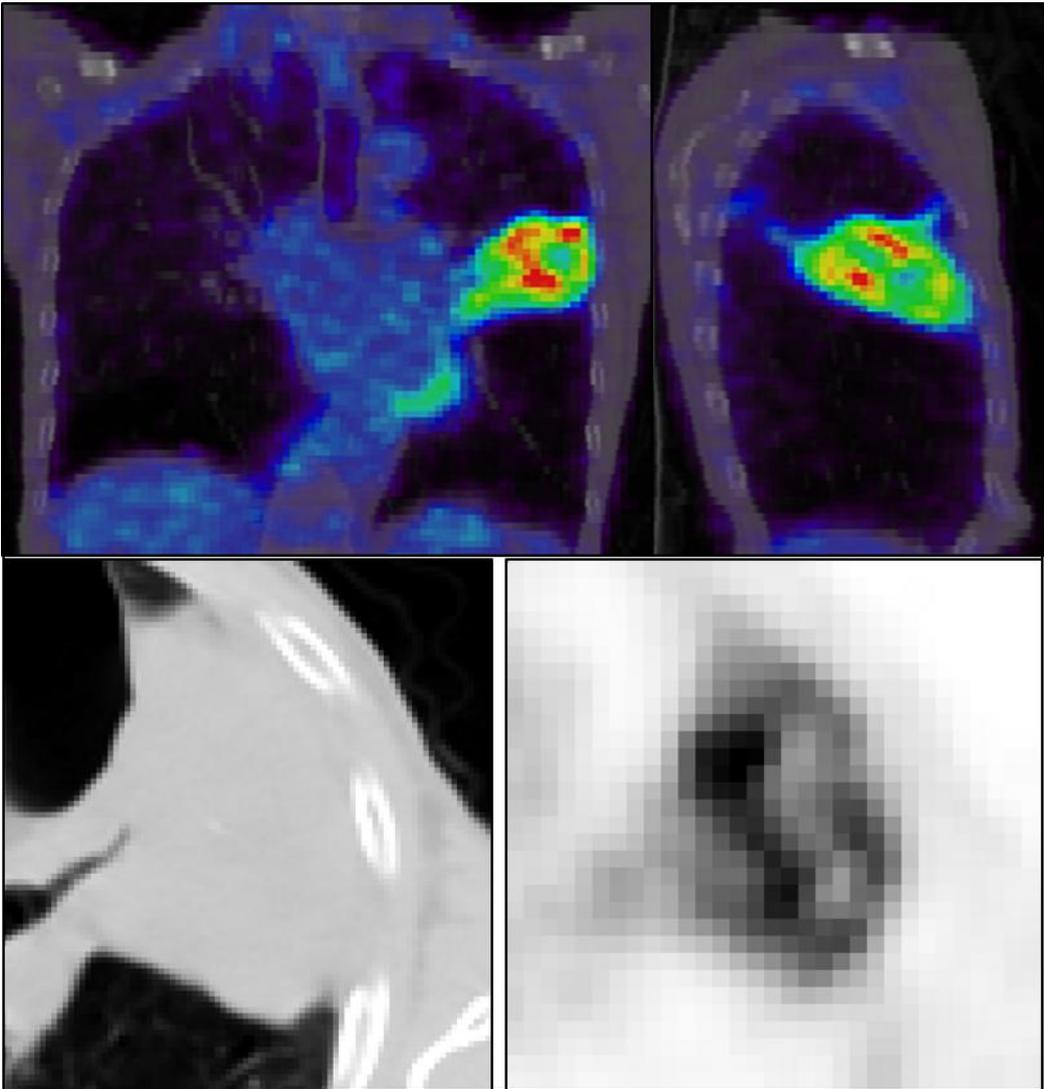


Figure captions

Figure 1: Kaplan-Meier curves for (a) the cohort and with (b) stage only and (c) the proposed nomogram. w.r.t.: with respect to.

Figure 2: Scatter diagrams showing the distributions of (a) metabolic and anatomical volumes, (b) PET heterogeneity and tumor volume, (c) CT heterogeneity and tumor volume.

Figure 3: Example of PET/CT images of a 6-month OS patient cumulating all risk factors. Top row, sagittal and coronal fused PET/CT, bottom row axial CT and PET.