Journal of Nuclear Medicine, published on May 10, 2017 as doi:10.2967/jnumed.117.189894

Analysis of prognostic values of various PET metrics in preoperative FDG PET for early stage bronchial carcinoma for progression free and overall survival: significantly increased glycolysis is a predictive factor.

Seraina Steiger[1], Michael Arvanitakis[2], Beate Sick[3], Walter Weder[2], Sven Hillinger[2]*, Irene A. Burger[1]*

1 Nuclear medicine, University Hospital Zurich, Zürich, Switzerland
2 Thoracic Surgery, University Hospital Zurich, Zürich, Switzerland
3 Department of Biostatistics, University of Zurich, Zürich, Switzerland

*Contributed equally.

Corresponding Author:
Irene A. Burger, Nuclear Medicine Department
University Hospital Zurich, Ramistr. 100
CH-8091 Zurich, Switzerland
Tel. +41 44 255 35 55 / Fax. +41 44 255 44 14
Email: irene.burger@usz.ch

Runing title: PET metrics predict outcome in NSCLC.

CONFLICT OF INTEREST:

There is no potential conflict of interest relevant to this article.

WORD COUNT: 4979

ABSTRACT

Purpose

To assess various volume based PET quantification metrics including: metabolic tumor volume (MTV) and total lesion glycolysis (TLG) with different thresholds as well as background activity based PET metrics (Background Subtracted Lesion activity (BSL) and Volume (BSV)) as prognostic markers for progression free and overall survival (PFS, OS) in early stage I and II non-small cell lung cancer (NSCLC) after resection.

Patients and Methods

133 patients received an adequate FDG PET/CT scan prior to surgery between January 2003 and December 2010. All PET activity metrics showed a skewed distribution and were log-transformed before calculating the Pearson correlation coefficients (PCC). Survival tree analysis was used to discriminate between high and low risk patients and to select the most important prognostic markers. Akaike information criterion (AIC) was used to compare two uni-variate models.

Results

Within the study time 36 patients died from NSCLC and 26 patients from other causes. At the end of follow up 70 patients were alive, with 67 patients being free of disease. All log-transformed PET metrics showed a very strong linear association with a PCC between 0.703-0.962. After multiple testing corrections only one prognostic marker contributed a significant split point in the survival tree analysis. Out of 10 potential predictors including 7 PET metrics BSL>6852 (p = 0.017) was chosen as split point assigning 13 patients into a high risk group. If BSL was removed from the set of predictors $TLG_{42\%}$ >4204 (p = 0.023) was chosen as split point. Using a dichotomized BSL or $TLG_{42\%}$ variable for an uni-variate Cox model the AIC-difference of both models was smaller than 2, therefore the data do not provide evidence that one of the two prognostic factors is superior.

Conclusion

Volume based PET metrics do correlate with PFS and OS and could be used for risk assessment in stage I-II NSCLC. The different PET metrics assessed in this study showed a

very high correlation; therefore, it is not surprising that there was no significant difference to predict PFS or OS within this study. Overall, Patients with large and metabolically active tumors should be considered high risk and might need further treatment after resection. Since all analysis steps were done with the same data these results should be validated on new patient data.

INTRODUCTION

The most common cancer and leading cause of cancer death worldwide is lung cancer, with non-small cell lung cancer (NSCLC) being by far the most common subtype. The therapeutic options depend on tumor stage, based on tumor size, tumor localization, infiltration of adjacent structures, lymph node involvement and distant metastasis[1]. In early-stage NSCLC a surgical procedure is considered the treatment of choice whereas the treatment in cases with local progression is more controversial[2]. After complete resection without positive margins of stage I-II NSCLC the need for adjuvant therapy still controversial, for nodal positive stage IIA-IIB adjuvant chemotherapy based on Cisplatin in combination with vinorelbine or gemcitabine is recommended according to the National Comprehensive Cancer Network (NCCN) guidelines 2016[3]. For patients without nodal involvement and stage I-II NSCLC observation is recommended, unless there are some factors considered to be indicative for a higher risk including poorly differentiated tumors, vascular invasion, wedge resection, tumors > 4 cm, visceral pleural involvement and unknown lymph node status[3]. Although FDG PET is an integral component of stating NSCLC since over 10 years[4, 5] and multiple studies showed already early on that the metabolic tumor activity on PET images is an independent prognostic factor for progression free survival (PFS) and overall survival (OS)[6-10], high metabolic tumor activity is not considered a risk factor for tumor staging and does not have an impact on treatment. This additional information about tumor aggressiveness beside the extent of metastasis on FDG PET/CT could be incorporated into clinical decisions regarding the need of adjuvant therapy.

There are several potential reasons why tumor metabolism still is not taken into account for tumor assessment; the first might be based on the methodology chosen to determine predictive measures. Most publications assess the median value of the PET metrics as a “virtual” cut off for Kaplan Meyer survival analysis, resulting in various different cut offs published in the literature (e.g. for SUV_{max} 15[6], 7.8[10] or 6[8]), being very variable and highly dependent on selected patient population.

A second reason might be the wide variety of methods for PET quantification, including the most commonly reported maximum standard uptake value (SUV_{max}), which however does only reflect the hottest voxel and therefore is prone to high statistical noise and does not represent the overall tumor activity[11-13]. PET activity measures incorporating tumor volume have been developed, including the metabolic tumor volume (MTV) defined as the total number of voxels within a volume of interest having an uptake above a predetermined SUV threshold and the total lesion glycolysis (TLG) as a multiplication of the metabolic tumor volume and the average SUV (SUV_{mean}) measuring the uptake of the entire lesion[14]. Since then several studies showed that TLG and MTV have a superior correlation to PFS and OS compared to SUV_{max} for NSCLC[15-17]. Liao et al. investigated in stage IV NSCLC patients the prognostic value of baseline whole body tumor burden measuring the metabolic tumor volume (MTV), TLG, the maximum standard uptake value (SUV_{max})[15]. A study of Lee at al. described MTV independent of other established prognostic factors as for example stage to be highly prognostic for disease progression and death in lung cancer[18]. However, the SUV threshold has not been standardized yet, and several thresholds from SUV_{max} (40-50%) as well as absolute thresholds including all voxels with an SUV over 2.5 ($TLG_{2.5}$) were suggested.

A recent study however showed a systematic bias for PET volume quantification with absolute thresholds or relative thresholds based on SUV_{max}, showing that FDG-activity of lesions with a high SUV_{max} are underestimated by $TLG_{42\%}$ whereas the activity of lesions with low activity are underestimated by $TLG_{2.5}$ and that a background subtracted lesion activity (BSL) and background subtracted volume (BSV) was more accurate in both phantoms and humans[11]. Furthermore a retrospective analysis of therapy response assessment with FDG PET/CT before and after neo-adjuvant chemotherapy showed that the relative difference for BSL and BSV significantly correlated with the tumor regression grade on histopathology while PET volume metrics based on a SUV_{max} 42% threshold did not[19]. Others proposed background based thresholds to delineate MTV and TLG for SCLC and showed correlation to OS, however they selected liver activity as background, what could limit the analysis of

adenocarcinomas, known to have only mild FDG-activity that would be underestimated by the proposed method[20].

Therefore, it was the aim of our study to compare background based volume PET metrics with the commonly used quantification methods SUV_{max}, TLG and MTV with survival tree models, to predict early recurrence and OS after resection of stage I and II bronchial carcinomas in a large study cohort.

MATERIALS AND METHODS

This retrospective, single-center study was approved by the Institutional Review Board. The conduct of the study met all local legal and regulatory requirements and was in accordance with the ethical principles originating from the International Conference on Harmonization guideline E6: Good Clinical Practice (KEK-ZH 2014-0130). It included 133 consecutive patients with bronchial carcinoma stage I and II and a FDG PET/CT scan prior to surgery. Inclusion criteria were: (I) Histopathology or cytology confirming adeno or squamous cell carcinoma, (II) tumor stage I or II, (III) FDG PET/CT scan performed within 119 days prior to surgery. Exclusion criteria were: (I) tumor stage III or IV, (II) second malignancy, (III) malignant pleural or pericardial effusion, (IV) surrounding inflammatory infiltrate with increased FDG activity, that could not be separated from the tumor lesion (V) off-site PET/CT before surgery without quantitative adequate imaging, (VI) paravenous injection, (VII) patient not fasting for at least 4 hours, (VIII) elevated blood glucose (> 7 mmol/dl). PET/CT were acquired from January 2003 until September 2010, surgery followed from January 2003 until December 2010. Clinical follow-up was performed according to the European Society for Medical Oncology guidelines, with clinical and CT control every 3 months in the first year, every 6 months for the second and third year and annual checks for the fourth and fifth year. Follow up FDG PET/CT was performed for unclear findings.

PET/CT acquisition and analysis

All patients were examined using a routine clinical protocol in our institution on dedicated PET/CT scanners (GE Healthcare DSTX, 16-or 64-slices CT, 7-8 frames, frame time 2 minutes) with injection of 350 MBq FDG 45-60 minutes before examination. A low dose unenhanced CT-scan was performed for attenuation correction and used for anatomical localization (80 mA, 140 kV). Approximate total dose equivalent for the entire PET/CT examination = 10 mSv. Image analysis independently was performed by a dual board certified nuclear medicine physician and radiologist and a nuclear medicine trainee with 2 years of radiology experience.

A cubic volume of interest (VOI) was placed around the primary tumor, in a way that the entire tumor activity was within the VOI, but no physiologically increased activity (e.g. FDG-uptake in the heart) was included using the Advantage Window 4.6 software (GE Healthcare). Physiologic FDG uptake within the selected tumor VOI was manually segmented using the cut inside tool. Within the selected VOI the hottest voxel was measured (SUV_{max}), MTV and the corresponding TLG above 42% from SUV_{max} as well as MTV and TLG above 2.5 were assessed using either 42% of SUV_{max} or 2.5 as a cut off to delineate the volume. Furthermore, BSL and BSV were measured in the same VOI using a background adapted threshold for each lesion, which was determined in a separate VOI placed over the most active adjacent background activity (e.g. of the lung, thoracic wall or mediastinum) (Figure 1).

Statistical analysis

Inter-reader agreement is assessed with Bland-Altman agreement analysis and interclass correlation coefficients (ICC). To facilitate the comparability we report for all PET metrics. ICC-values where values < 0 are indicative for no agreement, values from 0–0.2 are interpreted as slight, 0.21–0.4 as fair, 0.41–0.6 as moderate, 0.61–0.8 as substantial and 0.81–1 as almost perfect agreement [21].

In a first descriptive analysis step all PET activity measuring predictors are compared with each other using scatter plots and Pearson correlation coefficients.

In cases where the assumptions for Cox models are violated when using the continuous versions of the predictors we use survival tree models to assess the association between the set of ten potential predictors and survival or time to recurrence package[22]. These models take censoring into account, do not assume a specific form of associations between predictors and hazards (such as proportional hazard), and moreover non-linear effects or higher order interactions of predictors are handled automatically. At each node of a survival tree a p-value adjusted log rank statistics is used to decide which variable- split leads for the current patient population to an optimal separation into a lower and higher risk groups. The p-values are adjusted for multiple testing and splitting is only continued if further significant splits are found. If a survival tree yield only one significant split of a continuous predictor this split reveals a data driven cut point which separates low from high risk patients. We use the Akaike information criterion (AIC) described by Brunham et al to compare two uni-variate models. To determine the AIC we use uni-variate Cox regression models on dichotomized variables using the cut-point revealed by the survival tree models. Models having an AIC-difference <= 2 than both models are equally good fitting the data. If the AIC-difference is larger than 10, than the model with the larger AIC should not be further considered[23].

To get further insight in the non-linear association between predictors and hazard and to confirm results from survival tree and Cox regression models we use also survival random forest model [24]. The modeled association between predictors and hazard is illustrated by partial dependency plots. The mortality on the y-axis depict the expected number of events in the setting of the investigated study, therefore absolute numbers of the mortality variable should be used to compare the hazard across different predictor values. The ticks at the x-axis depict the position of the observed predictor values and give an impression how well the curve is supported by observed data. A horizontal curve in the partial dependency plot implies that the predictor has no influence on the mortality. If a curve

steps from low values to high values, we would use the predictor value at the step position for separating between a lower and higher risk group.

Statistical analysis was performed using R version 3.1.0[25].

RESULTS

Our study population consisted of 133 patients, of which 45 female and 88 male with an age of 47 - 91 years, average 73 years. 62 were diagnosed with squamous cell carcinoma, 71 with adenocarcinoma. Patient characteristics and tumor histology are summarized in Table 1.

The mean and median follow up time for this study was 4.4 years (range 0.1-10.9 years). Patients that did not die during the study period were followed over at least 2.3 years. Within the study time 36 patients died from NSCLC or direct complications due to tumor progression. 26 patients died from other causes, among them 6 had evidence of recurrence, while 20 were free of diseases. At the end of follow up 70 patients were alive, 3 with known recurrence and 67 were free of disease (mean follow time up 5.8 years). Patient that died from NSCLC had a mean recurrence free time of 1.9 years (mean 676 days, range 48-1923 days) and an average survival time of 4.3 years (1578 days, range 34-3975 days). The distribution of events is illustrated in Figure 2.

Inter-reader agreement in ^{18}F-FDG PET

Inter-reader agreement is given in table 2. SUV_{max} of the primary tumor yielded very similar values for both readers and accordingly the ICC was very high (ICC of 1, $p<<0.0001$). Also for 6 investigated volume based PET metric the ICC showed almost perfect agreement (Table 2). Therefore, for all further analysis only the results of reader 1 were considered. The absolute values for all PET volume metrics are given in Table 3.

Correlation of ^{18}F-FDG PET volume metrics

All pairs of volume based PET-activity metrics showed after log-transformation a very strong linear association which was quantified by the Pearson correlation coefficients between 0.703-0.962 (Table 4).

Influence of the different predictors on mortality

The dependency plots are given in Figure 3. All volume based PET metrics yielded very similar curves. Up to a certain threshold the values of the predictors do not influence the hazard, however above this threshold the hazard increases almost linearly, however in the range above the threshold we have only few observations (less than 20), therefore the model is not so well supported in this range. Nevertheless, it indicates, that the threshold for all PET metrics was higher than the mean or median value.

Identification of high risk patients for survival

For the survival tree model we used 10 potential predictors, i.e. histology, stage, volume, SUVmax, $TLG_{42\%}$, $TLG_{2.5\%}$, $MTV_{42\%}$, $MTV_{2.5\%}$, BSL, BSV. After multiple testing corrections only BSL>6852 (p=0.017) was chosen as significant split point assigning patients in a low risk group (BSL≤6852, n=120) and high risk group (BSL>6852, n=13). No further split became significant in the fitted survival tree. If BSL as a possible variable was removed from the study population $TLG_{42\%}$ was selected with a split point at $TLG_{42\%}$>4204 (n=12) and $TLG_{42\%}$≤4204 (n=121) with a p of 0.023. The corresponding Kaplan-Meier curve for the split points are given in Figure 4. The high risk group revealed by $TLG_{42\%}$ is completely included in the high risk group revealed by BSL. The AIC-difference for both uni-variate models using BSL or $TLG_{42\%}$, respectively, was smaller than 2, therefore the data do not provide evidence for either event that one of the two prognostic factors (BSL or $TLG_{42\%}$) is superior (Figure 1). This was true for both types of events (dead of disease or early recurrence). All FDG PET metrics yielded similar results at various split points, further results are given in Table 5.

DISCUSSION

This study confirms previous reports suggesting an association between increased PET activity and OS as well as PFS for NSCLC[6]. Our survival tree analysis showed that volume based PET metrics were primarily selected to differentiate between low risk and high risk patients, confirming previous reports suggesting that TLG and MTV were superior predictive values for OS and PFS than SUV_{max}[26]. In our study BSL and $TLG_{42\%}$ showed very similar prognostic performance, resulting in the nearly exact same selection of high risk patients with a large, highly metabolically active tumor burden. We did this study in continuance to a former study with phantoms and 50 patients with lung tumors showing that the BSL correlated better with the total injected FDG activity than $TLG_{2.5}$ and $TLG_{42\%}$. The data showed that $TLG_{2.5}$ and $TLG_{42\%}$ have systematic errors: The activities of lesions with a high SUV_{max} are underestimated by $TLG_{42\%}$, the activity of lesions with low activity are underestimated by $TLG_{2.5}$[11]. A second study showed that therapy response assessment was indeed significantly limited by this systematic bias of $TLG_{42\%}$ while BSL and BSV were able to detect histopathologic response with a significantly higher accuracy[19]. In the current study population however, although BSL was the primary selected node, we could not show a significant improvement of predictability of OS or PFS compared to $TLG_{42\%}$. Given the high correlation of all volume based PET metrics this comes not by surprise and the grouping in high or low risk patients was therefore very similar for the various PET metrics. This further confirms that FDG PET is an independent predictive value for PFS and OS and therefore might be an important factor to be considered for more aggressive adjuvant therapy in early stag NSCLC.

In the literature, various cut offs for MTV or TLG have been proposed. The largest cohort analyzing 529 patients with stage I or II NSCLC used the median value for MTV (16 cm^3) and TLG (70)[26]. The assumption that the median value of a cohort actually is the ideal discriminator between high- and low-risk patients is however questionable. Kim et al suggested MTV (11.6cm^3) and TLG (13.8)[27] using ROC analysis. However, this method

does not take the time to event “dead of disease” into account. Looking at the partial dependency plots from the survival random forest model we actually see that a nonlinear relation between risk and all volume based PET metrics were fitted, with a clear increase for high volumes and activities. This is also reflected by the relatively high split points in our cohort BSL=6852 and $TLG_{42\%}$=4024 when working with a single survival tree. This suggests that tumor large tumors with high FDG uptake on staging PET/CT should be considered high risk and therefore might profit form adjuvant therapy.

In our cohort $MTV_{2.5}$ and $TLG_{2.5}$ yielded 8 (6%) cases were tumor activity could not be assessed, since SUV_{max} was below 2.5 (Figure 5). However, all of these cases would be in the low risk category and therefore this limitation would not be as substantial for outcome prediction as it would be for therapy response assessment, if you cannot determine the baseline value. To use the back ground of the surrounding tissue instead of an absolute or relative cut off seams also to be a superior approach in other organs with variable background activities such as the liver[28]. First evidence showed more stable results of BSV compared to $MTV_{50\%}$ measurements between breath hold and free-breathing PET. This can be explained by the strong influence of blurring on the SUV_{max} and consequently the selected threshold, while back ground activity of the liver is less affected by motion, therefore the tumor volume between breath hold and free breathing PET scans show better correlation for BSV compared to $MTV_{50\%}$[28].

Our study has several limitations. It is a retrospective analysis of patients with bronchial carcinoma stage I or II who underwent surgery after imaging with PET/CT at a single institution. The retrospective setting allows a study with a large population with standardized clinical parameters such as histology, treatment regimen and long-lasting follow up period. However, there are also strong limitations given the heterogeneous time intervals of up to 119 days between imaging and surgery that could have led to a progression of the disease, the heterogeneous postsurgical therapy, and the long time

period with some heterogeneity also in PET data acquisition such as uptake time, injected dose or reconstruction algorithms.

Furthermore, some care was necessary in drawing the tumor VOI without including any physiological increased FDG accumulation. Furthermore the determination of the most active background is a further variable influencing the read out. However since most stage I or II tumors were either surrounded by lung parenchyma or close to the mediastinum, measurement of lung or mediastinal activity was straight forward. Also the prognostic performance of the dichotomized predictors with the data driven found cut-points are not confirmed yet.

CONCLUSION

Our data confirm that volume based PET metrics are predictive for both progression free survival and overall survival. We conclude that the ideal cut off between high- and low risk patients might not be the median of the respective predictor. In our data set volume based PET metrics including the total activity (BLS and $TLG_{42\%}$) showed the highest prognostic value among ten potential predictors. The high correlation between readers seems promising for further use of those metrics in clinical settings. However, the found cut offs (BSL>6852, $TLG_{42\%}$ >4204) need to be confirmed. A prospective study with standardized protocols in multiple institutions will be needed to delineate absolute cut offs that could be clinically used and applied to direct more aggressive therapy in patients with large tumors with high FDG uptake.

REFERENCES

1. Nair A, Klusmann MJ, Jogeesvaran KH, Grubnic S, Green SJ, Vlahos I. Revisions to the TNM staging of non-small cell lung cancer: rationale, clinicoradiologic implications, and persistent limitations. *Radiographics : a review publication of the Radiological Society of North America, Inc.* 2011;31:215-238.

2. Gounant V, Khalil A, Crequit P, et al. 2014 update on non-small cell lung cancer (excluding diagnosis). *Diagnostic and interventional imaging.* 2014.

3. Ettinger DS, Wood DE, Akerley W, et al. NCCN Guidelines (R) Insights: Non-Small Cell Lung Cancer, Version 4.2016 Featured Updates to the NCCN Guidelines. *J Natl Compr Canc Ne.* 2016;14:255-264.

4. De Leyn P, Lardinois D, Van Schil PE, et al. ESTS guidelines for preoperative lymph node staging for non-small cell lung cancer. *European journal of cardio-thoracic surgery : official journal of the European Association for Cardio-thoracic Surgery.* 2007;32:1-8.

5. Lardinois D, Weder W, Hany TF, et al. Staging of non-small-cell lung cancer with integrated positron-emission tomography and computed tomography. *The New England journal of medicine.* 2003;348:2500-2507.

6. Borst GR, Belderbos JS, Boellaard R, et al. Standardised FDG uptake: a prognostic factor for inoperable non-small cell lung cancer. *European journal of cancer.* 2005;41:1533-1541.

7. Pandit N, Gonen M, Krug L, Larson SM. Prognostic value of [18F]FDG-PET imaging in small cell lung cancer. *European journal of nuclear medicine and molecular imaging.* 2003;30:78-84.

8. Takeda A, Yokosuka N, Ohashi T, et al. The maximum standardized uptake value (SUVmax) on FDG-PET is a strong predictor of local recurrence for localized non-small-cell lung cancer after stereotactic body radiotherapy (SBRT). *Radiotherapy and oncology : journal of the European Society for Therapeutic Radiology and Oncology.* 2011;101:291-297.

9. Kadota K, Colovos C, Suzuki K, et al. FDG-PET SUVmax combined with IASLC/ATS/ERS histologic classification improves the prognostic stratification of patients with stage I lung adenocarcinoma. *Annals of surgical oncology.* 2012;19:3598-3605.

10. Hanin FX, Lonneux M, Cornet J, et al. Prognostic value of FDG uptake in early stage non-small cell lung cancer. *European journal of cardio-thoracic surgery : official journal of the European Association for Cardio-thoracic Surgery.* 2008;33:819-823.

11. Burger IA, Vargas HA, Apte A, et al. PET quantification with a histogram derived total activity metric: superior quantitative consistency compared to total lesion glycolysis with absolute or relative SUV thresholds in phantoms and lung cancer patients. *Nuclear medicine and biology.* 2014;41:410-418.

12. Laffon E, Lamare F, de Clermont H, Burger IA, Marthan R. Variability of average SUV from several hottest voxels is lower than that of SUVmax and SUVpeak. *European radiology.* 2014;24:1964-1970.

13. Schwartz J, Humm JL, Gonen M, et al. Repeatability of SUV measurements in serial PET. *Medical physics.* 2011;38:2629-2638.

14. Larson SM, Erdi Y, Akhurst T, et al. Tumor Treatment Response Based on Visual and Quantitative Changes in Global Tumor Glycolysis Using PET-FDG Imaging. The Visual Response Score and the Change in Total Lesion Glycolysis. *Clinical positron imaging : official journal of the Institute for Clinical PET.* 1999;2:159-171.

15. Liao S, Penney BC, Zhang H, Suzuki K, Pu Y. Prognostic value of the quantitative metabolic volumetric measurement on 18F-FDG PET/CT in Stage IV nonsurgical small-cell lung cancer. *Academic radiology.* 2012;19:69-77.

16. 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.

17. Moon SH, Hyun SH, Choi JY. Prognostic significance of volume-based PET parameters in cancer patients. *Korean journal of radiology : official journal of the Korean Radiological Society.* 2013;14:1-12.

18. Lee P, Weerasuriya DK, Lavori PW, et al. Metabolic tumor burden predicts for disease progression and death in lung cancer. *International journal of radiation oncology, biology, physics.* 2007;69:328-333.

19. Burger IA, Casanova R, Steiger S, et al. 18F-FDG PET/CT of Non-Small Cell Lung Carcinoma Under Neoadjuvant Chemotherapy: Background-Based Adaptive-Volume Metrics

Outperform TLG and MTV in Predicting Histopathologic Response. *Journal of nuclear medicine : official publication, Society of Nuclear Medicine.* 2016;57:849-854.

20. Park SB, Choi JY, Moon SH, et al. Prognostic value of volumetric metabolic parameters measured by [18F]fluorodeoxyglucose-positron emission tomography/computed tomography in patients with small cell lung cancer. *Cancer imaging : the official publication of the International Cancer Imaging Society.* 2014;14:2.

21. Landis JR, Koch GG. Measurement of Observer Agreement for Categorical Data. *Biometrics.* 1977;33:159-174.

22. Hothorn T, Hornik K, Zeileis A. Unbiased recursive partitioning: A conditional inference framework. *J Comput Graph Stat.* 2006;15:651-674.

23. Burnham KP, Anderson DR. Multimodel inference - understanding AIC and BIC in model selection. *Sociol Method Res.* 2004;33:261-304.

24. U.B. IHaK. Random Forests for Survival, Regression and Classification (RF-SRC). *R package version 241.* 2016.

25. Team RC. R: A Language and Environment for Statistical Computing. . 2014.

26. Hyun SH, Choi JY, Kim K, et al. Volume-based parameters of (18)F-fluorodeoxyglucose positron emission tomography/computed tomography improve outcome prediction in early-stage non-small cell lung cancer after surgical resection. *Ann Surg.* 2013;257:364-370.

27. Kim K, Kim SJ, Kim IJ, 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. *Nuclear medicine communications.* 2012;33:613-620.

28. Li G, Schmidtlein CR, Burger IA, Ridge CA, Solomon SB, Humm JL. Assessing and accounting for the impact of respiratory motion on FDG uptake and viable volume for liver lesions in free-breathing PET using respiration-suspended PET images as reference. *Medical physics.* 2014;41:091905.

FIGURES:

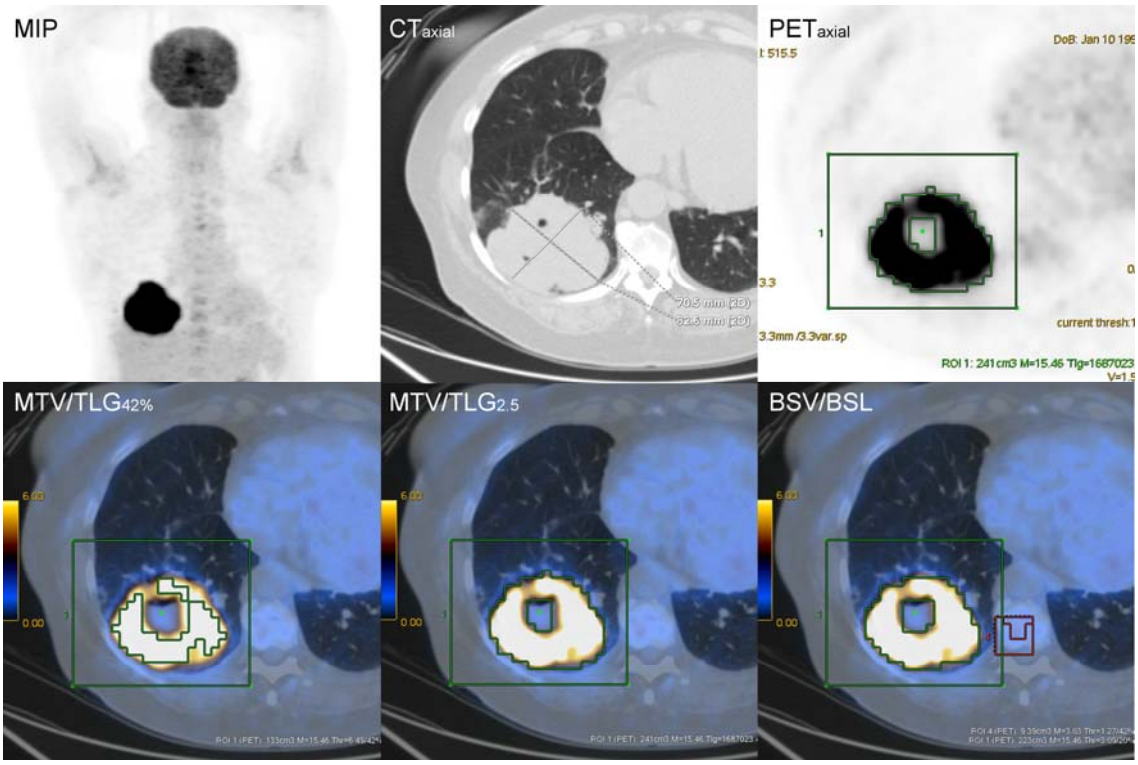

Figure1: PET/CT with 376 MBq FDG of a 57 y.o. woman with a large squamous cell cancer in the left lower lobe on the maximum intensity projection (MIP) with central necrosis on axial CT and high FDG uptake on axial PET (SUV_{max} 15.5). In the lower row the fused axial PET/CT image with the volume of interest placed around the tumor with the three cut offs are illustrated: showing that the 42% SUV_{max} based metrics underestimate the volume, however all measures were above the cut off for high risk according to the survival tree analysis. After lobectomy (pT2 pN0 cM0, G2-3, R0) she developed a recurrence after 1.2 years and died after 2.8 years with osseous, pulmonary and cerebral metastasis.

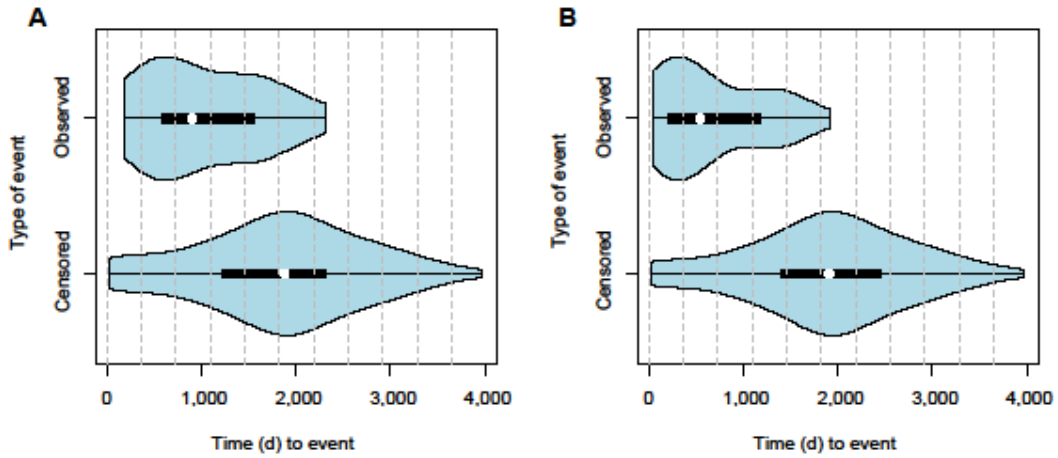

Figure 2: Distribution of events and censor times for A) Death of disease (DOD) and B) evidence of recurrence.

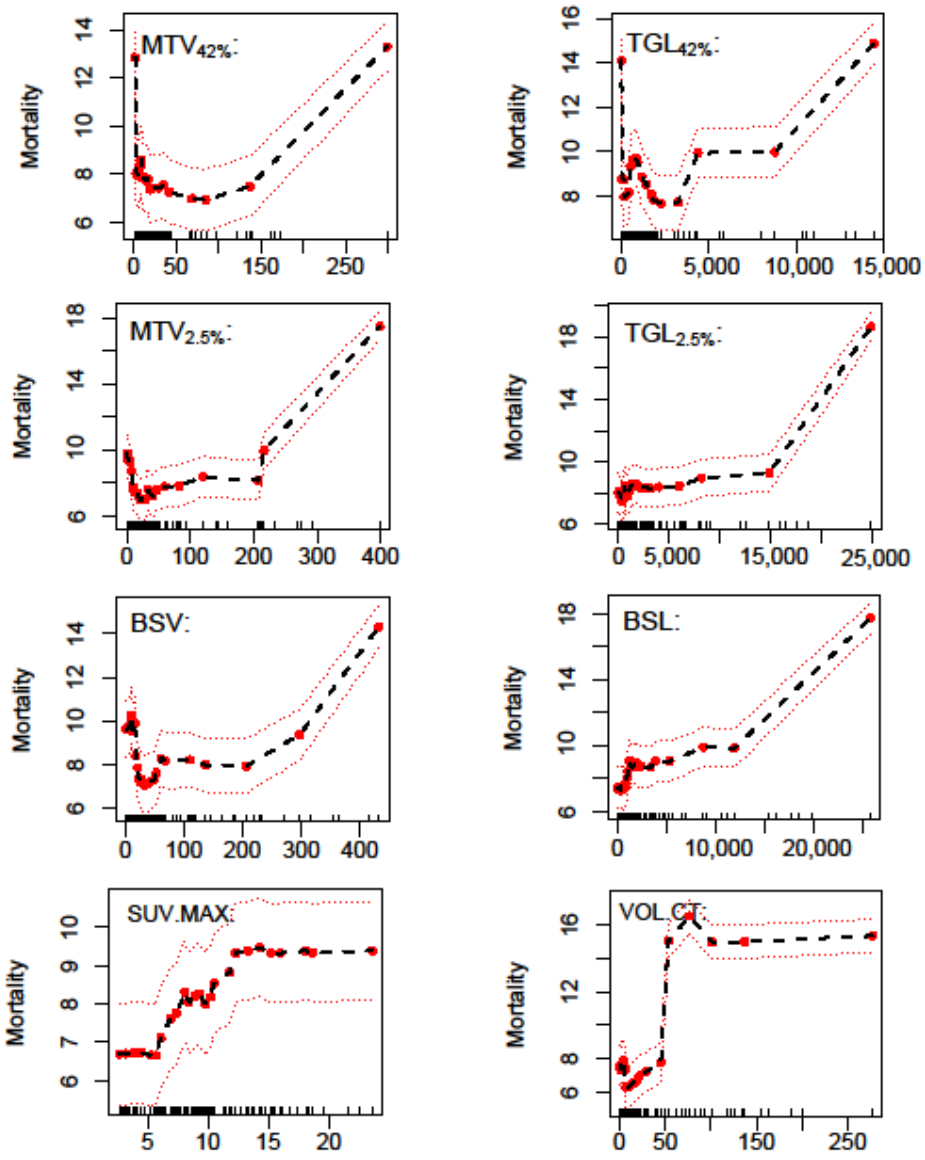

Figure 3: Dependency plots for all volume based PET metrics to illustrate the influence on mortality. The mortality on the y-axis depicts the expected number of events in the setting of the investigated study. Hence, absolute numbers of the mortality variable can be used to compare the hazard across different predictor values. The ticks at the x-axis depict the position of the observed predictor values and give an impression how well the curve is supported by observed data. A horizontal curve in the partial dependency plot implies that the predictor has no influence on the mortality. If a curve steps from low values to high values we can use the predictor value at the step position for separating between a lower and higher risk group.

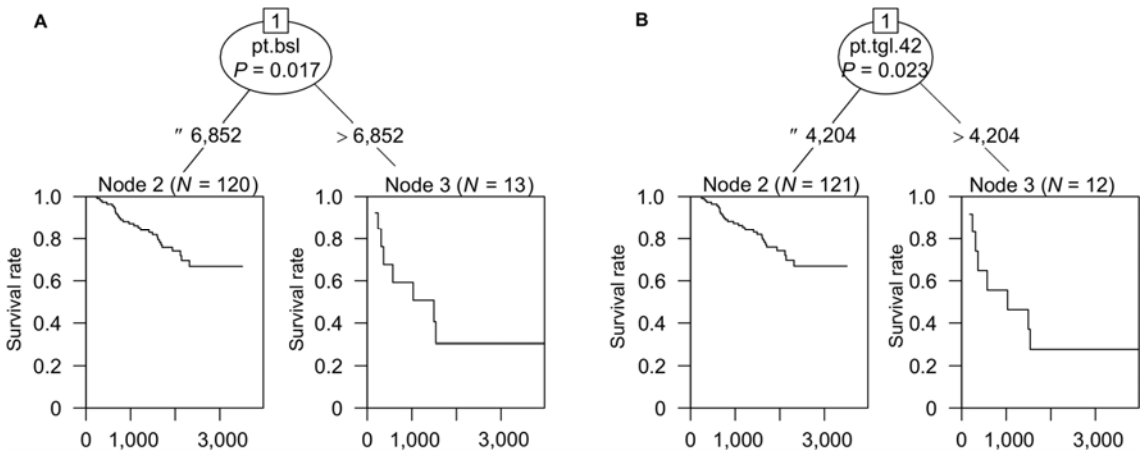

Figure 4: Kaplan-Meier curves for the two split points according to the Cox-regression survival tree model showing that patients with very high values for the PET metrics BSL and $TLG_{42\%}$ did significantly worse than patients with values below the split point of (6852 for BSL and 4204 for $TLG_{42\%}$), there was no significant difference for the predictive value of both models.

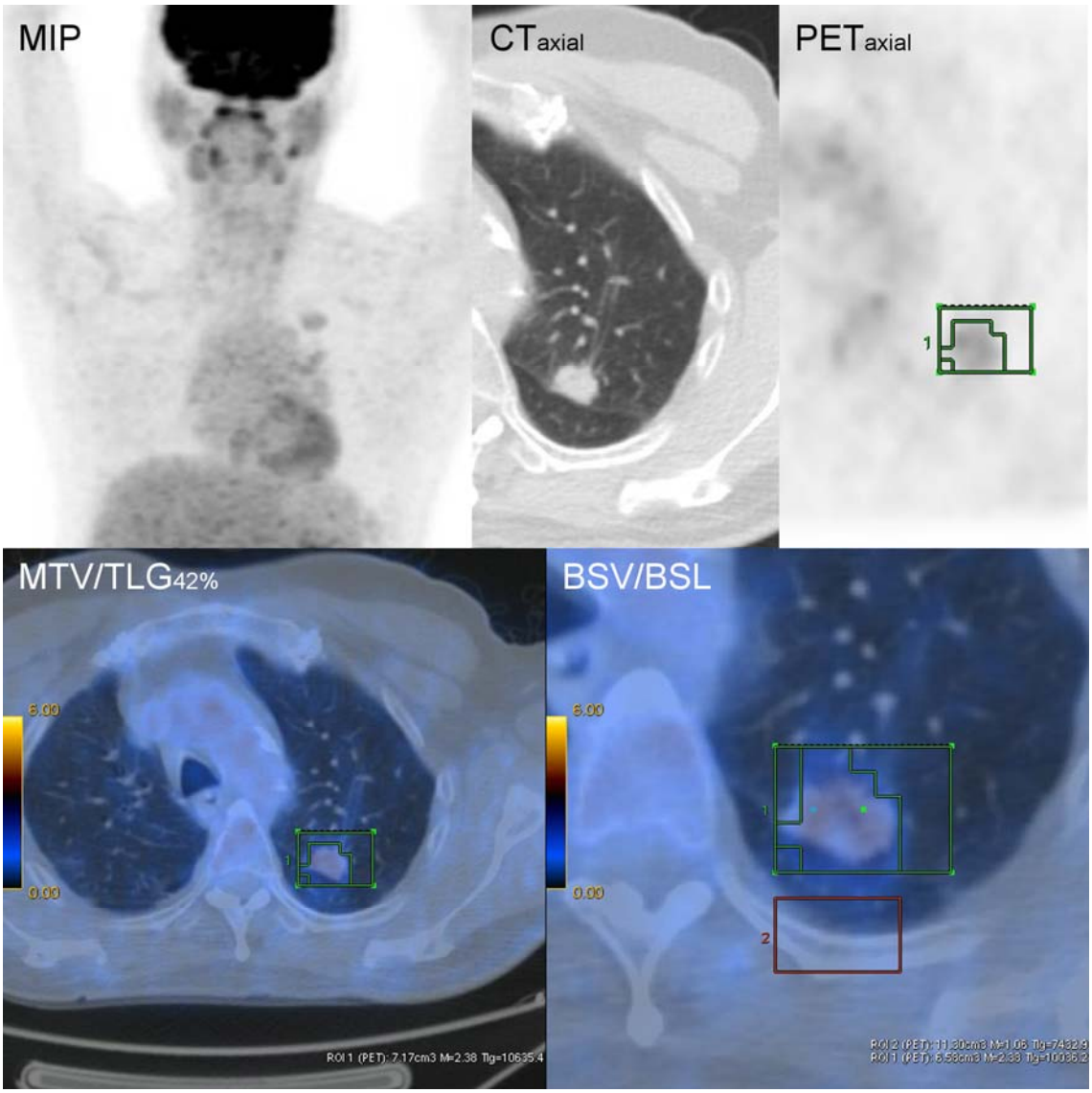

Figure 5: FDG PET/CT with 351 MBq showing a solitary pulmonary mass on maximum intensity projection (MIP), spiculated on axial CT with only minimal FDG uptake on axial PET (SUV_{max} 2.3). Therefore, only the PET metrics for MTV42%/TLG42% and BSV/BSL could be measured.

TABLES

Table 1: Patient characteristics and tumor histology

	Numbers	%
Patients		
Age, median and range at diagnosis (years)	73 (47-91)	
Gender:		
Female	45	34%
Male	88	66%
Histo:		
Squamous cell carcinoma	62	47 %
(female/male)	18/44	14%/33%
Adenocarcinoma	71	53 %
(female/male)	27/44	20%/33%
IA	20	15 %
IB	53	40 %
IIA	21	16 %
IIB	39	29 %
Lymphnode metastases		
N0	87	
N1	36	

Table 2: Interreader agreement

PET Metric:	ICC	95%-CI	p
SUVmax	1		2.56e-300
$MTV_{42\%}$	0.999	0.998<ICC<0.999	5.41e-173
$TLG_{42\%}$	1		3.94e-229
$MTV_{2.5}$	0.990	0.985<ICC<0.993	1.91e-106
$TLG_{2.5}$	0.998	0.997<ICC<0.999	4.21e-151
BSV	0.988	0.984<ICC<0.992	6.27e-110
BSL	0.997	0.996<ICC<0.998	2.33e-152

Table 3: PET metrics

PET Metric:	mean	median	min	max
SUVmax	9.3	8.9	1.1	23.6
MTV42%	27.3	10.7	1.0	299.0
TLG42%	1620.7	584.5	36.2	14464.2
MTV2.5	46.7	18.8	0.1	401.0
TLG2.5	2491.1	863.1	2.3	24965.0
BSV	57.3	25.3	0.5	464.0
BSL	2567.1	1004.6	9.0	26377.5

Table 4: Correlation of PET metrics

PET Metric:	**$MTV_{42\%}$**	**$TLG_{42\%}$**	**$MTV_{2.5}$**	**$TLG_{2.5}$**	**BSV**	**BSL**
$MTV_{42\%}$	1	0.885	0.766	0.703	0.751	0.706
$TLG_{42\%}$		1	0.896	0.891	0.881	0.925
$MTV_{2.5}$			1	0.932	0.837	0.911
$TLG_{2.5}$				1	0.813	0.920
BSV					1	0.962
BSL						1

Table 5: Split points according to the survival tree analysis

PET Metric:	Cut off	Number of selected high risk / low risk patients	p
TLG42%	4204	12 / 121	0.023
BSV	231	7 / 126	0.012
BSL	6852	13 / 120	0.013