

Analysis of Prognostic Values of Various PET Metrics in Preoperative ^{18}F -FDG PET for Early-Stage Bronchial Carcinoma for Progression-Free and Overall Survival: Significantly Increased Glycolysis Is a Predictive Factor

Seraina Steiger¹, Michael Arvanitakis², Beate Sick³, Walter Weder², Sven Hillinger^{*2}, and Irene A. Burger^{*1}

¹Nuclear Medicine Department, University Hospital Zurich, Zurich, Switzerland; ²Thoracic Surgery, University Hospital Zurich, Zurich, Switzerland; and ³Department of Biostatistics, University of Zurich, Zurich, Switzerland

The purpose of this study was to assess various volume-based PET quantification metrics, including metabolic tumor volume and total lesion glycolysis (TLG) with different thresholds, as well as background activity-based PET metrics (background-subtracted lesion activity [BSL] and background-subtracted volume) as prognostic markers for progression-free and overall survival (PFS and OS, respectively) in early-stage I and II non-small cell lung cancer (NSCLC) after resection. **Methods:** Patients ($n = 133$) underwent an adequate ^{18}F -FDG PET/CT scan before surgery between January 2003 and December 2010. All PET activity metrics showed a skewed distribution and were log-transformed before calculation of the Pearson correlation coefficients. Survival tree analysis was used to discriminate between high- and low-risk patients and to select the most important prognostic markers. The Akaike information criterion was used to compare 2 univariate 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 strong linear association, with a Pearson correlation coefficient between 0.703 and 0.962. After multiple testing corrections, only 1 prognostic marker contributed a significant split point in the survival tree analysis. Of 10 potential predictors including 7 PET metrics, a BSL greater than 6,852 ($P = 0.017$) was chosen as split point, assigning 13 patients into a high-risk group. If BSL was removed from the set of predictors, a 42% TLG ($\text{TLG}_{42\%}$) of greater than 4,204 ($P = 0.023$) was chosen as split point. When a dichotomized BSL or $\text{TLG}_{42\%}$ variable was used for a univariate Cox model, the Akaike information criterion difference of both models was smaller than 2; therefore, the data do not provide evidence that 1 of the 2 prognostic factors is superior. **Conclusion:** Volume-based PET metrics 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 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. Because all analysis steps were done with the same data, these results should be validated on new patient data.

Key Words: oncology; lung; PET/CT; statistical analysis; BSL; MTV; PET segmentation; stage I–II; TLG

J Nucl Med 2017; 58:1925–1930

DOI: 10.2967/jnumed.117.189894

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). The need for adjuvant therapy is still controversial after complete resection without positive margins of stage I–II NSCLC patients. Although adjuvant chemotherapy based on cisplatin in combination with vinorelbine or gemcitabine is recommended for nodal-positive stage IIA–IIB, observation is recommended for nodal-negative patients, unless one of the factors considered to be indicative for a higher risk is present (poorly differentiated tumors, vascular invasion, wedge resection, tumors > 4 cm, visceral pleural involvement, and unknown lymph node status), according to the National Comprehensive Cancer Network guidelines 2016 (3).

Although ^{18}F -FDG PET has been an integral component of staging NSCLC for more than 10 y (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 in addition to the extent of metastasis on ^{18}F -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 considered 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 cutoff for Kaplan–Meyer survival analysis, resulting in various different cutoffs published in the literature (e.g., for SUV_{max}

Received Jan. 20, 2017; revision accepted Apr. 20, 2017.

For correspondence or reprints contact: Irene A. Burger, Nuclear Medicine Department, University Hospital Zurich, Ramistrasse 100, CH-8091 Zurich, Switzerland.

E-mail: irene.burger@usz.ch

*Contributed equally to this work.

Published online May 10, 2017.

COPYRIGHT © 2017 by the Society of Nuclear Medicine and Molecular Imaging.

15 (6), 7.8 (10), or 6 (8)) being 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 SUV_{max} , which reflects only the hottest voxel and is therefore 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 MTV and the SUV_{mean} measuring the uptake of the entire lesion (14). Several studies showed that TLG and MTV have a superior correlation to PFS and OS compared with 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 MTV, TLG, SUV_{max} (15). A study by Lee et al. described MTV, independent of other established prognostic factors (e.g., stage), to be highly prognostic for disease progression and death in lung cancer (18).

However, the SUV threshold has not yet been standardized, 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} , demonstrating that ^{18}F -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 a background-subtracted volume (BSV) were more accurate in both phantoms and humans (11). Furthermore, a retrospective analysis of therapy response assessment with ^{18}F -FDG PET/CT before and after neoadjuvant chemotherapy showed that the relative difference for BSL and BSV significantly correlated with the tumor regression grade on histopathology whereas PET volume metrics based on an SUV_{max} 42% threshold did not (19). Others proposed background-based thresholds to delineate MTV and TLG for NSCLC and showed correlation to OS, using liver activity as background. This, however, could limit the analysis of adenocarcinomas, known to have only mild ^{18}F -FDG activity, and therefore 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). One hundred thirty-three consecutive patients were included with bronchial carcinoma stage I and II who had undergone ^{18}F -FDG PET/CT before surgery. Inclusion criteria were histopathology or cytology confirming adeno or squamous cell carcinoma, tumor stage I or II, and ^{18}F -FDG PET/CT performed within 119 d before surgery. Exclusion criteria were tumor stage III or IV; second malignancy; malignant pleural or pericardial effusion; surrounding inflammatory infiltrate

with increased ^{18}F -FDG activity that could not be separated from the tumor lesion; off-site PET/CT before surgery without quantitative adequate imaging; paravenous injection; patient not fasting for at least 4 h; and elevated blood glucose (>7 mmol/dL). PET/CT images were acquired from January 2003 until September 2010; surgery followed from January 2003 until December 2010. Clinical follow-up was performed according to guidelines of the European Society for Medical Oncology, with clinical and CT control every 3 mo in the first year, every 6 mo for the second and third years, and annual checks for the fourth and fifth years. Follow-up ^{18}F -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-slice CT, 7–8 frames, frame time of 2 min) with injection of 350 MBq of ^{18}F -FDG 45–60 min before examination. A low-dose unenhanced CT scan was obtained for attenuation correction and used for anatomic localization (80 mA, 140 kV). Approximate total dose equivalent for the entire PET/CT examination was 10 mSv. Image analysis was performed independently by a dual-board-certified nuclear medicine physician and radiologist and a nuclear medicine trainee with 2 y 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., ^{18}F -FDG uptake in the heart) was included, using the Advantage Window 4.6 software (GE Healthcare). Physiologic ^{18}F -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} ($MTV_{42\%}$ and $TLG_{42\%}$) or 2.5 ($MTV_{2.5}$ and $TLG_{2.5}$) as a cutoff 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) (Fig. 1).

Statistical Analysis

Interreader agreement is assessed with Bland–Altman agreement analysis and interclass correlation coefficients (ICCs). To facilitate the comparability, we report all PET metrics: ICC values < 0 indicate no agreement, values from 0 to 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 were compared with each other using scatterplots and Pearson correlation coefficients.

In cases in which the assumptions for Cox models were violated when the continuous versions of the predictors were used, we used survival tree models to assess the association between the set of 10 potential predictors and survival or time to recurrence package (22). These models took censoring into account and did not assume a specific form of associations between predictors and hazards (such as proportional hazard); moreover, nonlinear effects or higher-order interactions of predictors were handled automatically. At each node of a survival tree, P -value-adjusted log rank statistics were 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 were adjusted for multiple testing, and splitting was continued only if further significant splits were found. If a survival tree yielded only 1 significant split of a continuous predictor, this split revealed a data-driven cut point, which separated low- from high-risk patients. We used the Akaike information criterion (AIC) described by Brunham et al. to compare 2 univariate models. To

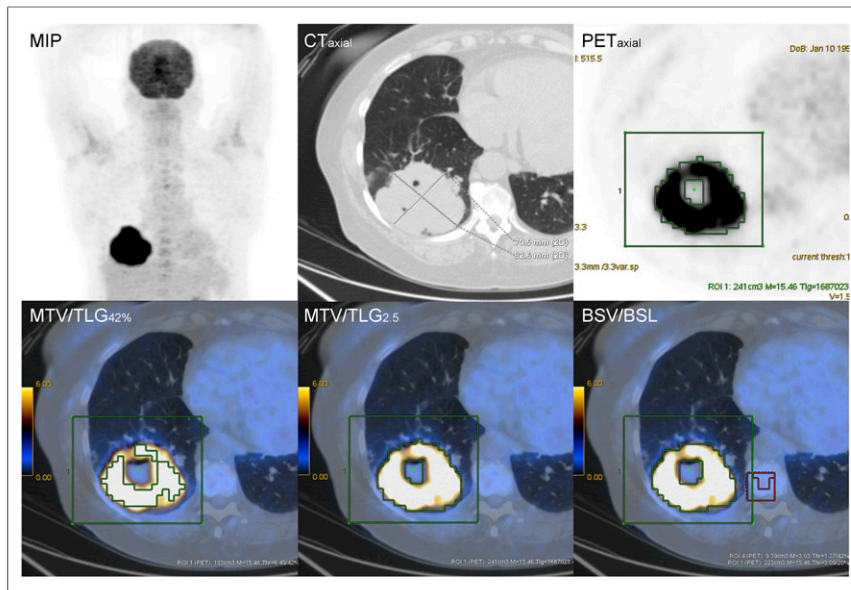


FIGURE 1. PET/CT image with 376 MBq of ^{18}F -FDG of a 57-y-old woman with large squamous cell cancer in left lower lobe on maximum-intensity projection (MIP) with central necrosis on axial CT and high ^{18}F -FDG uptake on axial PET (SUV_{max} , 15.5). In lower row, fused axial PET/CT image with VOI placed around tumor with 3 cutoffs is illustrated, showing that 42% SUV_{max} -based metrics underestimate volume. However, all measures were above cutoff for high risk according to survival tree analysis. After lobectomy (pT2 pN0 cM0, G2–3, R0), she developed a recurrence after 1.2 y and died after 2.8 y with osseous, pulmonary, and cerebral metastasis.

TABLE 1

Patient Characteristics and Tumor Histology

Characteristic	n	%
Median age at diagnosis (y)	73 (range, 47–91)	
Sex		
Female	45	34
Male	88	66
Histology		
Squamous cell carcinoma	62	47
	Female, 18/male, 44	Female, 14/male, 33
Adenocarcinoma	71	53
	Female, 27/male, 44	Female, 20/male, 33
Stage		
IA	20	15
IB	53	40
IIA	21	16
IIB	39	29%
Lymph node metastases		
N0	87	
N1	36	

determine the AIC, we used univariate Cox regression models on dichotomized variables with the cut point revealed by the survival tree models. When models had an AIC difference of less than 2, then both models were equally good fitting the data. If the AIC difference was larger than 10, then the model with the larger AIC was not considered further (23).

To get further insight into the nonlinear association between predictors and hazard and to confirm results from survival tree and Cox regression models, we also used a survival random forest model (24). The modeled association between predictors and hazard is illustrated by partial dependency plots. The mortality on the y-axis depicts 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 software (version 3.1.0) (25).

RESULTS

Our study population consisted of 133 patients, of whom 45 were women and 88 men (age range, 47–91 y; average, 73 y). Sixty-two were diagnosed with squamous cell carcinoma and 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 y (range, 0.1–10.9 y). Patients who did not die during the study period were followed over at least 2.3 y. Within the study time, 36 patients died from NSCLC or direct complications due to tumor progression. Twenty-six patients died from other causes, among them 6 had evidence of recurrence, whereas 20 were free of diseases. At the end of follow-up, 70 patients were alive, 3 had

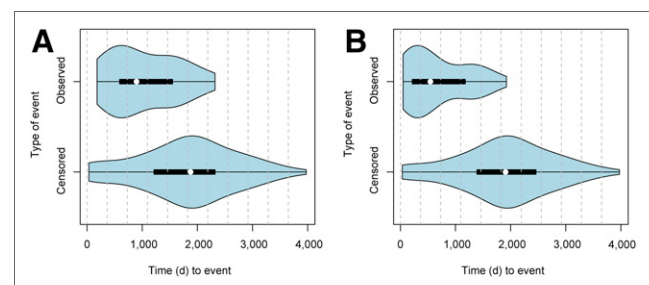


FIGURE 2. Distribution of events and censor times for death of disease (A) and evidence of recurrence (B).

TABLE 2
Interreader Agreement

PET metric	ICC	95% confidence interval	P
SUV _{max}	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

known recurrence, and 67 were free of disease (mean follow-up time, 5.8 y). Patients who died from NSCLC had a mean recurrence-free time of 1.9 y (mean, 676 d; range, 48–1,923 d) and an average survival time of 4.3 y (1,578 d; range, 34–3,975 d). The distribution of events is illustrated in Figure 2.

Interreader Agreement in ¹⁸F-FDG PET

Interreader agreement is given in Table 2. The SUV_{max} of the primary tumor yielded similar values for both readers, and accordingly the ICC was high (ICC, 1; $P < 0.0001$). Also, for 6 investigated volume-based PET metrics, 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 ¹⁸F-FDG PET Volume Metrics

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

Influence of Different Predictors on Mortality

The dependency plots are given in Figure 3. All volume-based PET metrics yielded 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. In the range above the threshold, though, we have only few observations (<20); therefore, the model is not well supported in this range. Nevertheless, this result 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, that is, histology, stage, volume, SUV_{max}, TLG_{42%}, TLG_{2.5}, MTV_{42%},

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

MTV_{2.5}, BSL, and BSV. After multiple testing corrections only a BSL greater than 6,852 ($P = 0.017$) was chosen as a significant split point, assigning patients in a low-risk group (BSL ≤ 6,852, $n = 120$) and high-risk group (BSL > 6,852, $n = 13$). 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 a TLG_{42%} greater than 4,204 ($n = 12$) and TLG_{42%} of 4,204 or less ($n = 121$), with a P of 0.023. The corresponding

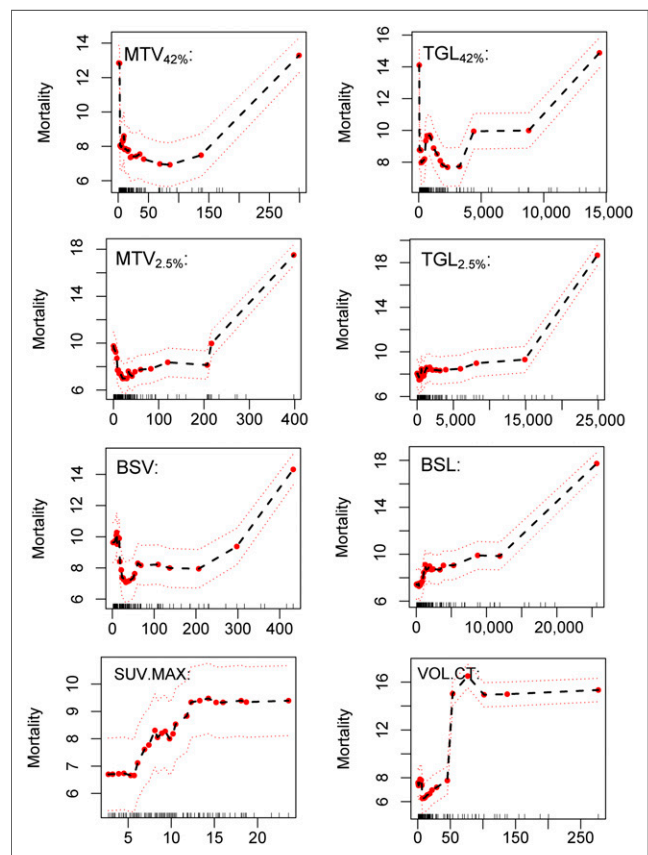


FIGURE 3. Dependency plots for all volume-based PET metrics to illustrate influence on mortality. Mortality on y-axis depicts expected number of events in setting of investigated study. Hence, absolute numbers of mortality variable can be used to compare hazard across different predictor values. Ticks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

TABLE 3
PET Metrics

PET metric	Mean	Median	Minimum	Maximum
SUV _{max}	9.3	8.9	1.1	23.6
MTV _{42%}	27.3	10.7	1.0	299.0
TLG _{42%}	1,620.7	584.5	36.2	14,464.2
MTV _{2.5}	46.7	18.8	0.1	401.0
TLG _{2.5}	2,491.1	863.1	2.3	24,965.0
BSV	57.3	25.3	0.5	464.0
BSL	2,567.1	1,004.6	9.0	26,377.5

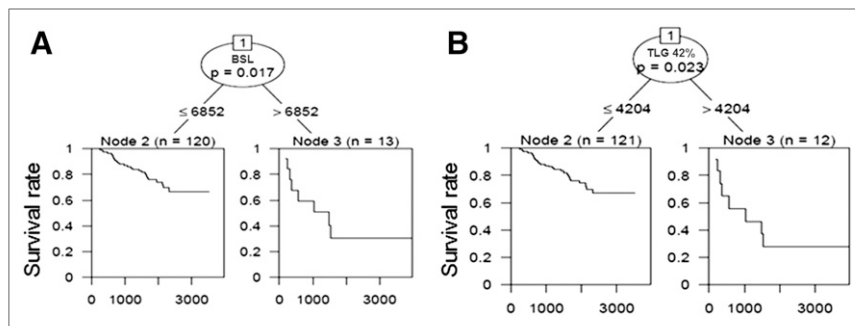


FIGURE 4. Kaplan–Meier curves for 2 split points according to Cox regression survival tree model showing that patients with high values for PET metrics BSL and $TLG_{42\%}$ did significantly worse than patients with values below split point (6,852 for BSL and 4,204 for $TLG_{42\%}$). There was no significant difference for predictive value for either model.

Kaplan–Meier curve for the split points are given in Figure 4. The high-risk group revealed by a $TLG_{42\%}$ was completely included in the high-risk group revealed by BSL. The AIC difference for both univariate models using BSL or $TLG_{42\%}$, respectively, was smaller than 2; therefore, the data do not provide evidence for either event that 1 of the 2 prognostic factors (BSL or $TLG_{42\%}$) was superior (Fig. 1). This was true for both types of events (dead of disease or early recurrence). All ^{18}F -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 better predictive values for OS and PFS than SUV_{max} (26). In our study, BSL and $TLG_{42\%}$ showed 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 of a former study with phantoms and 50 patients with lung tumors showing that the BSL correlated better with the total injected ^{18}F -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 $TLG_{42\%}$, and 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 $TLG_{42\%}$ whereas 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 with $TLG_{42\%}$. Given the high correlation of all volume-based PET metrics, this is not surprising and the grouping in high- or low-risk patients was therefore similar for the various PET metrics. This further confirms that ^{18}F -FDG PET is independently predictive for PFS and OS and therefore might be an important factor to be considered for more aggressive adjuvant therapy in early-stage NSCLC.

In the literature, various cutoffs for MTV or TLG have been proposed. The largest cohort, analyzing 529 patients with stage I or II NSCLC, used the median values 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 (11.6 cm^3) and TLG (13.8) (27) using receiver-operating-characteristic analysis. However, this method does not take the time to the event “dead of disease” into account. When the partial dependency plots from the survival random forest model are considered, 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 = 6,852 and $TLG_{42\%}$ = 4,024) when working with a single survival tree, suggesting that large tumors with high ^{18}F -FDG uptake on staging PET/CT should be considered high risk and therefore might profit from adjuvant therapy.

In our cohort, $MTV_{2.5}$ and $TLG_{2.5}$ yielded 8 (6%) cases in which tumor activity could not be assessed, because the SUV_{max} was below 2.5 (Fig. 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 one cannot determine the baseline value. To use the background of the surrounding tissue instead of an absolute or relative cutoff also seems 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 with $MTV_{50\%}$ 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, whereas background activity of the liver is less affected by motion, therefore the tumor volume between breath-hold and free-breathing PET scans shows better correlation for BSV than $MTV_{50\%}$ (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 a long-lasting follow-up period. However, there are also strong limitations given the heterogeneous time intervals of up to 119 d between imaging and surgery that could have led to a progression of the disease, the heterogeneous postsurgical therapy, and the long 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 physiologic increased ^{18}F -FDG accumulation. Furthermore, the determination of the most active

TABLE 5
Split Points According to Survival Tree Analysis

PET metric	Cutoff	No. of selected high-risk/low-risk patients	P
$TLG_{42\%}$	4,204	12/121	0.023
BSV	231	7/126	0.012
BSL	6,852	13/120	0.013

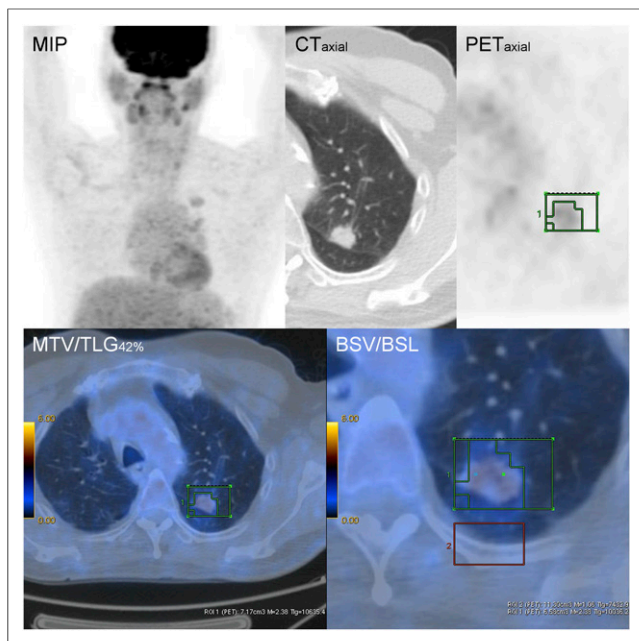


FIGURE 5. ^{18}F -FDG PET/CT with 351 MBq showing solitary pulmonary mass on maximum-intensity projection (MIP), spiculated on axial CT with only minimal ^{18}F -FDG uptake on axial PET (SUV_{max} , 2.3). Therefore, only PET metrics for $\text{MTV}_{42\%}/\text{TLG}_{42\%}$ and BSV/BSL could be measured.

background is a further variable influencing the read out. However, because most stage I or II tumors were either surrounded by lung parenchyma or close to the mediastinum, measurement of lung or mediastinal activity was straightforward. Also, the prognostic performance of the dichotomized predictors with the data-driven cut points has not yet been confirmed.

CONCLUSION

Our data confirm that volume-based PET metrics are predictive for both PFS and OS. The ideal cutoff between high- and low-risk patients might not be the median of the respective predictor. In our dataset, volume-based PET metrics including the total activity (BSL and $\text{TLG}_{42\%}$) showed the highest prognostic value among 10 potential predictors. The high correlation between readers seems promising for further use of those metrics in clinical settings. However, the cutoffs ($\text{BSL} > 6,852$, $\text{TLG}_{42\%} > 4,204$) need to be confirmed. A prospective study with standardized protocols in multiple institutions will be needed to delineate absolute cutoffs that could be used clinically and applied to direct more aggressive therapy in patients with large tumors with high ^{18}F -FDG uptake.

DISCLOSURE

No potential conflict of interest relevant to this article was reported.

REFERENCES

- Nair A, Klusmann MJ, Jogevaran KH, Grubnic S, Green SJ, Vlahos I. Revisions to the TNM staging of non-small cell lung cancer: rationale, clinico-radiologic implications, and persistent limitations. *Radiographics*. 2011;31:215–238.
- Gounant V, Khalil A, Crequit P, et al. Update on non-small cell lung cancer (excluding diagnosis). *Diagn Interv Imaging*. 2014;95:721–725.
- 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 Netw*. 2016;14:255–264.
- De Leyn P, Lardinois D, Van Schil PE, et al. ESTS guidelines for preoperative lymph node staging for non-small cell lung cancer. *Eur J Cardiothorac Surg*. 2007;32:1–8.
- Lardinois D, Weder W, Hany TF, et al. Staging of non-small-cell lung cancer with integrated positron-emission tomography and computed tomography. *N Engl J Med*. 2003;348:2500–2507.
- Borst GR, Belderbos JS, Boellaard R, et al. Standardised FDG uptake: a prognostic factor for inoperable non-small cell lung cancer. *Eur J Cancer*. 2005;41:1533–1541.
- Pandit N, Gonen M, Krug L, Larson SM. Prognostic value of ^{18}F FDG-PET imaging in small cell lung cancer. *Eur J Nucl Med Mol Imaging*. 2003;30:78–84.
- Takeda A, Yokosuka N, Ohashi T, et al. The maximum standardized uptake value (SUV_{max}) on FDG-PET is a strong predictor of local recurrence for localized non-small-cell lung cancer after stereotactic body radiotherapy (SBRT). *Radiother Oncol*. 2011;101:291–297.
- Kadota K, Colovos C, Suzuki K, et al. FDG-PET SUV_{max} combined with IASLC/ATS/ERS histologic classification improves the prognostic stratification of patients with stage I lung adenocarcinoma. *Ann Surg Oncol*. 2012;19:3598–3605.
- Hanin FX, Lonneux M, Cornet J, et al. Prognostic value of FDG uptake in early stage non-small cell lung cancer. *Eur J Cardiothorac Surg*. 2008;33:819–823.
- 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. *Nucl Med Biol*. 2014;41:410–418.
- Laffon E, Lamare F, de Clermont H, Burger IA, Marthan R. Variability of average SUV from several hottest voxels is lower than that of SUV_{max} and SUV_{peak} . *Eur Radiol*. 2014;24:1964–1970.
- Schwartz J, Humm JL, Gonen M, et al. Repeatability of SUV measurements in serial PET. *Med Phys*. 2011;38:2629–2638.
- 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. *Clin Positron Imaging*. 1999;2:159–171.
- Liao S, Penney BC, Zhang H, Suzuki K, Pu Y. Prognostic value of the quantitative metabolic volumetric measurement on ^{18}F -FDG PET/CT in stage IV non-surgical small-cell lung cancer. *Acad Radiol*. 2012;19:69–77.
- 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.
- Moon SH, Hyun SH, Choi JY. Prognostic significance of volume-based PET parameters in cancer patients. *Korean J Radiol*. 2013;14:1–12.
- Lee P, Weerasuriya DK, Lavori PW, et al. Metabolic tumor burden predicts for disease progression and death in lung cancer. *Int J Radiat Oncol Biol Phys*. 2007;69:328–333.
- Burger IA, Casanova R, Steiger S, et al. ^{18}F -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. *J Nucl Med*. 2016;57:849–854.
- Park SB, Choi JY, Moon SH, et al. Prognostic value of volumetric metabolic parameters measured by ^{18}F fluorodeoxyglucose-positron emission tomography/computed tomography in patients with small cell lung cancer. *Cancer Imaging*. 2014;14:2.
- Landis JR, Koch GG. Measurement of observer agreement for categorical data. *Biometrics*. 1977;33:159–174.
- Hothorn T, Hornik K, Zeileis A. Unbiased recursive partitioning: a conditional inference framework. *J Comput Graph Stat*. 2006;15:651–674.
- Burnham KP, Anderson DR. Multimodel inference: understanding AIC and BIC in model selection. *Sociol Methods Res*. 2004;33:261–304.
- Ishwaran H, Kogalur UB. Random forests for survival, regression and classification (RF-SRC). Version 2.4.1. University of Miami website. <http://web.ccs.miami.edu/~hishwaran>. Accessed October 26, 2017.
- R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing website. 2014. <http://www.R-project.org/>. Accessed October 26, 2017.
- 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.
- Kim K, Kim SJ, Kim JJ, Kim YS, Pak K, Kim H. Prognostic value of volumetric parameters measured by ^{18}F -FDG PET/CT in surgically resected non-small-cell lung cancer. *Nucl Med Commun*. 2012;33:613–620.
- 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. *Med Phys*. 2014;41:091905.