Am J Nucl Med Mol Imaging 2018;8(1):50-61
www.ajnmmi.us /ISSN:2160-8407/ajnmmi0069658

Original Article

Assessment of very early response evaluation with ^{18}F-FDG-PET/CT predicts survival in erlotinib treated NSCLC patients-A comparison of methods

Joan Fledelius[1], Anne Winther-Larsen[2], Azza A Khalil[3], Karin Hjorthaug[4], Jørgen Frøkiær[4], Peter Meldgaard[3]

[1]Department of Nuclear Medicine, Herning Regional Hospital, Herning, Denmark; Departments of [2]Clinical Biochemistry, [3]Oncology, [4]Nuclear Medicine and PET Centre, Aarhus University Hospital, Aarhus, Denmark

Received November 23, 2017; Accepted December 25, 2017; Epub February 5, 2018; Published February 15, 2018

Abstract: We evaluated whether changes in ^{18}F-Fluoro-D-Glucose (^{18}F-FDG)-uptake evaluated early during erlotinib treatment predict survival in non-small cell lung cancer (NSCLC) patients. Positron emission tomography (PET)/CT scans from 56 NSCLC patients before and after 7-10 days of erlotinib treatment were analyzed with four different methods: Visual evaluation and percentage change in lean body mass corrected standardized uptake values (SULs): SUL_{peak}, SUL_{max} and total lesion glycolysis (TLG). The semi-quantitative parameters abilities to predict progression free survival (PFS) and overall survival (OS) were compared and we found that percentage change in SUL_{peak}, SUL_{max} and TLG all correlated with PFS and OS with the strongest correlation found for TLG (R=0.51, P < 0.001). The highest area under the curve (AUC) for predicting OS was for TLG (0.70 (0.56-0.85)) with a sensitivity of 0.68 and a specificity of 079. All methods except visual evaluation, SUL_{peak} at 15% and 30%, and TLG at 40% cut-off separates the survival curves for the response categories for PFS. For OS, visual evaluation and SUL_{max} did not, whereas TLG at 4 different cut-off levels and SUL_{peak} at the three lowest cut-off levels did. In conclusion: Early change in ^{18}F-FDG-uptake during erlotinib correlated to both PFS and OS. TLG, as suggested by PERCIST 1.0, shows the strongest correlation to survival, whereas visual evaluation seems to be less sensitive at this very early time-point, but lower cut-off levels for discriminating between response categories seem to be relevant as we find that 20-25% change for both response and progression is optimal.

Keywords: ^{18}F-FDG, PERCIST 1.0, early response evaluation, lung cancer

Introduction

Treatment with tyrosine kinase inhibitors (TKIs) in NSCLC has proven effective in certain subgroups of patients, in particular, but not exclusively, in epidermal growth factor receptor (EGFR) mutation positive (EGFR-mut) patients [1-5].

At our institution, we routinely establish the EGFR mutation status in adenocarcinoma patients, and erlotinib treatment is offered in the first-line to non-operable EGFR-mut patients. For the EGFR wild-type patients (EGFR-wt), erlotinib treatment is considered for second- or third-line treatment since it is known that a subgroup of EGFR-wt patients will respond to erlotinib treatment [6-8]. In order to identify this particular subgroup, it is important to find a reliable method to predict the response after a short treatment period.

Evaluating response with CT scans is not a particularly sensitive method, especially in EGFR-wt patients, because the anatomical changes are rather slow owing to the cytostatic nature of the response [9, 10]. However, alterations in glucose metabolism measured by the change in ^{18}F-FDG-uptake have been shown to happen very early, within days, in TKI sensitive cells and patients [6, 7, 9, 11, 12]. Furthermore, many studies have found that a change in the ^{18}F-FDG-uptake is predictive of the histopathological response of PFS and in some cases OS, but there is no agreement on how to measure this change, and various methods are presently used [13-16].

Therefore, we set out to identify the best way of predicting survival (PFS and OS) with an early ^{18}F-FDG-PET/CT for response evaluation by comparing various methods for quantification of change in ^{18}F-FDG uptake. Finding a method,

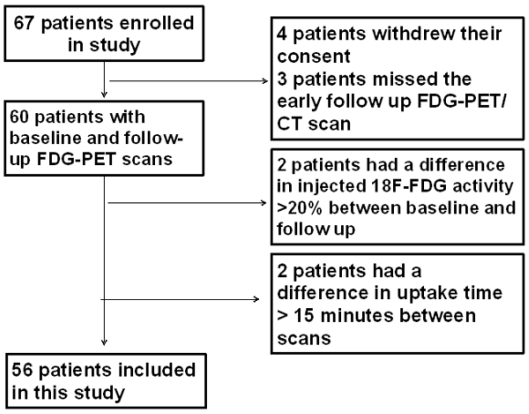

Figure 1. Patient selection for the present analysis from the original prospective study.

which will allow us to identify the subgroup likely to benefit from treatment very early in the course of treatment would enable us to test-treat patients and discontinue the treatment in case of no response. It has previously been demonstrated that the single lesion evaluations are not as sensitive as the more global TLG changes and visual evaluation for predicting CT response, by our group and others [17-19] and the present study was performed to evaluate if a similar pattern could be demonstrated for survival.

Materials and methods

Patients

This was a retrospective analysis of ^{18}F-FDG-PET/CT scans from a prospective single center study on advanced-stage (III-IV) NSCLC patients recruited from April 2013 until August 2015 at the Department of Oncology, Aarhus University Hospital, Denmark, the details on inclusion and exclusion criteria have been described previously [20]. In brief, we included all patients not eligible for curatively intended treatment who received erlotinib as first-, second- or third-line treatment, a flow chart of the inclusion in the present analysis is shown in **Figure 1**. Testing for EGFR mutations had been performed in all patients as part of the routine diagnostic work-up by use of the "Therascreen EGFR RGQ" PCR kit (QIAGEN, Manchester, UK) according to the manufacturer's protocol, based on this, patients were classified as either EGFR-wt or EGFR-mut. Informed consent was obtained from all individual participants and the study was approved by the Central Denmark Region Committees on Biomedical Research Ethics (no. 1-10-72-19-12).

^{18}F-FDG-PET/CT acquisition and evaluation

All patients had an ^{18}F-FDG-PET/CT scan performed before (baseline) and after 7-10 days of erlotinib treatment (follow-up) performed on a combined PET/CT scanner (Siemens Biograph TruePoint 40, Siemens Healthcare GMbH, Erlangen, Germany) at the Department of Nuclear Medicine and PET-Centre, Aarhus University Hospital, Denmark, with the same scanner model, acquisition- and reconstruction protocol, previously published in detail [20]. In brief, after a fasting period of at least 6 hours, assuring a glucose level < 11 mM, patients were injected with 5 MBq ± 10% ^{18}F-FDG/kg and scanned (3 min per bed position), after an uptake time of 60 ± 10 minutes, with a whole-body low-dose CT scan (50 mAs, 120 kVp).

The scans were evaluated by one experienced nuclear medicine specialist who was blinded to the outcome, the treatment response was evaluated by 4 different methods: 1) visual evaluation as described by Mac Manus *et al* [21], 2) percentage change in the highest intensity voxel (%SUL_{max}), 3) percentage change of the highest intensity 1 cm^3 (%SUL_{peak}) as according to PERCIST 1.0 [22], and 4) percentage change in TLG (%TLG) delineated at mean SUL + 2 standard deviations (SD) in a spherical 3 cm volume of interest in the right lobe of the liver (SUL_{mean} (liver)). For SUL_{max} and SUL_{peak}, the change between the "hottest" lesion at each time point was used, not necessarily the same lesion. A lesion was considered evaluable if SUL_{peak} was 1.5 × (SUL_{mean} (liver) + 2SD) according to PERCIST 1.0 [21], and the delineation was performed semi-automatically after manually roughly outlining each lesion, resulting in an SUL_{mean} and a metabolic tumor volume (MTV) of the delineated area, thus enabling a calculation of TLG for each lesion as SUL_{mean} × MTV and finally the %TLG was calculated as percentage change in the sum of TLGs from all evaluable lesions.

For all methods, various cut-offs were used for categorization of the treatment response into three response categories: Partial metabolic response (PMR), stable metabolic disease (SMD) and progressive metabolic disease (PMD). For %SUL_{max}, 25%, and 15% change was

Table 1. On overview of the evaluation methods used

Method	Parameter	SMD/PMR	SMD/PMD
Visual	Visual change	Significant decrease	Significant increase
SUL_{peak} (30%)*	$\%SUL_{peak}$	30% decrease	30% increase
SUL_{peak} (25%)	$\%SUL_{peak}$	25% decrease	25% increase
SUL_{peak} (20%)	$\%SUL_{peak}$	20% decrease	20% increase
SUL_{peak} (15%)	$\%SUL_{peak}$	15% decrease	15% increase
TLG (45/75%)*	% TLG	45% decrease	75% increase
TLG (50%)	% TLG	50% decrease	50% increase
TLG (40%)	% TLG	50% decrease	40% increase
TLG (30%)	% TLG	30% decrease	30% increase
TLG (25%)	% TLG	25% decrease	25% increase
TLG (20%)	% TLG	20% decrease	20% increase
SUL_{max} (25%)	$\%SUL_{max}$	25% decrease	25% increase
SUL_{max} (15%)	$\%SUL_{max}$	15% decrease	15% increase

*The two PERCIST 1.0 methods. SMD is stable metabolic disease, PMR is partial metabolic response and PMD is progressive metabolic disease.

Table 2. Patient- and tumor characteristics

Characteristics	Number (%)
Gender	
Female	27 (48)
Male	29 (52)
Performance status	
0-1	47 (84)
2	9 (16)
Smoking status	
Never or former*	41 (73)
Current	14 (25)
Unknown	1 (2)
Stage	
III	5 (9)
IV	51 (91)
Histology	
Adenocarcinoma	48 (86)
Squamous cell	8 (14)
EGFR mutation status	
EGFR-wt	48 (86)
EGFR-mut	8 (14)
Erlotinib treatment	
1st line of palliative treatment	10 (18)
2nd line of palliative treatment	38 (68)
3rd line of palliative treatment	8 (14)

*Former smoker was defined as having stopped smoking at time of diagnosis.

used, for $\%SUL_{peak}$, 30%, 25%, 20% and 15% change was used and for the %TLG, 45/75% change, 50%, 40%, 30%, 25% and 20% was used. An overview of all methods is found in **Table 1**.

Statistical analysis

Follow-up time was calculated using the reverse Kaplan-Meier method and the OS was measured from the day of inclusion until death of any cause, if patients were still alive on the last follow-up date (November 11th, 2016), they were censored at that day. The PFS was measured from the day of inclusion until progression on a CT scan, "clinical progression" or death, if patients stopped because of side effects or erlotinib treatment was still ongoing, they were censored. Estimates of median survival were calculated by the Kaplan-Meier method and the log rank test was used for overall- and pairwise comparison of the survival curves. All the survival data is reported as median (95% confidence interval (95% CI)) and a Bonferroni correction for the use of multiple methods was applied for the 13 methods in the Kaplan-Meier analyses resulting in a significance level of 0.004.

Correlations between the continuous variables ($\%SUL_{max}$, $\%SUL_{peak}$ and $\%_{TLG}$) and PFS and OS were evaluated using linear regression analysis and univariate Cox regression using a significance level of 0.017 (corrected for 3 methods). Receiver operating characteristics (ROC) analysis was used for evaluation of prediction of PFS < median and OS < median identifying the optimal cut-off visually by locating the data point nearest the top left corner on the ROC curve, when considering sensitivity and specificity equally important. Statistical analysis was performed using SPSS statistics version 23.0 for Macintosh (IBM SPSS Statistics, Chicago IL).

Results

In total, 56 patients were included in this study with a median age of 68 years (range: 44-83 years), the patient characteristics are presented in **Table 2**, no patients were lost to follow-up.

Table 3. Compliance with the PERCIST 1.0 standardization criteria

Parameter	Baseline	Follow-up	Numerical diff	PERCIST 1.0	Adherence
Injected FDG-activity					
Mean (SD)	348 (90)	343 (85)	19 (16)	Baseline ± 20%	100% (56/56)
Range	197-609	199-618	0-64		
Glucose level					
Mean (SD)	6.3 (0.9)	6.5 (0.9)	0.6 (0.5)	< 11 mM	100% (56/56)
Range	4.6-8.8	4.7-9.0	0.0-1.6		
Uptake time					
Mean (SD)	59.1 (4.4)	59.2 (5.0)	5.3 (4.1)	60 ± 10 min	97% (109/112*)
Range	51-74	48-72	0-15	Baseline ± 15 min	100% (56/56)

*The uptake time at both baseline and follow up, the three patients with 48, 72 and 74 minutes uptake time at one scan time were included because they all had a difference between the two scans within the allowed 15 min. SD is the standard deviation.

Table 4. Results from regression analysis for all ^{18}F-FDG-PET/CT continuous variables

	Correlation to PFS				Correlation to OS			
	R	p (lin)	HR (95% CI)	p (cox)	R	p (lin)	HR (95% CI)	p (cox)
%TLG (N=53)	0.510	< 0.001	1.021 (1.012-1.031)	< 0.001	0.458	0.001	1.018 (1.009-1.027)	< 0.001
$\%SUL_{max}$ (N=56)	0.387	0.003	1.022 (1.011-1.033)	< 0.001	0.346	0.009	1.013 (1.004-1.022)	0.003
$\%SUL_{peak}$ (N=56)	0.373	0.005	1.019 (1.008-1.031)	0.001	0.280	0.037	1.012 (1.001-1.023)	0.004

p (lin) and p (cox) are the *p*-value from the linear regression analysis and the univariate cox regression analysis respectively, N is the number of patients analyzed by each method, 95% CI is the 95% confidence interval, R is the correlation coefficient and HR is the hazard ratio from the univariate cox regression analysis.

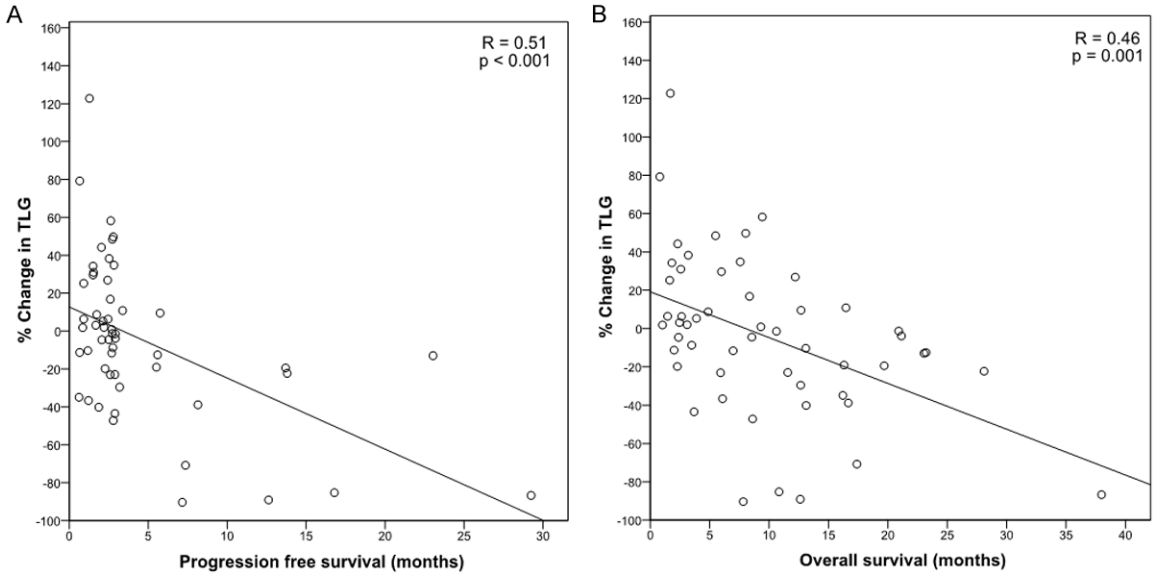

Figure 2. Scatterplots including the linear regression line for the strongest correlating parameter for the 53 patients analyzed by %TLG for a progression free survival (A) and overall survival (B).

Data on injected FDG-activity, glucose levels and uptake time are presented in **Table 3**, and the complete data for each patient is found in the supplementary file. There was (median (range)) 1 (0-21) days from baseline ^{18}F-FDG-PET/CT scan to the first day of treatment and 8 (2-23) days from the first day of treatment to follow-up ^{18}F-FDG-PET/CT scan.

Table 5. Receiver operating characteristics analyses on the PET parameters ability to predict progression free survival (PFS) < median and overall survival (OS) < median

Method	PFS < median			OS < median		
	AUC (95% CI)	Sens/Spec	Cut-off	AUC (95% CI)	Sens/Spec	Cut-off
%TLG (N=53)	0.74 (0.60-0.88)	0.68/0.81	-0.2%	0.70 (0.56-0.85)	0.68/0.79	1.4%
$\%SUL_{max}$ (N=56)	0.74 (0.61-0.87)	0.57/0.76	-1.4%	0.58 (0.43-0.73)	0.56/0.64	-4.6%
$\%SUL_{peak}$ (N=56)	0.70 (0.57-0.84)	0.61/0.71	-6.8%	0.58 (0.43-0.73)	0.56/0.54	-7.2%

AUC is the area under the curve, 95% CI is the 95 percent confidence interval, Sens/Spec is sensitivity/specificity and the Cut-off as the corresponding percentage change for this optimal sensitivity/specificity.

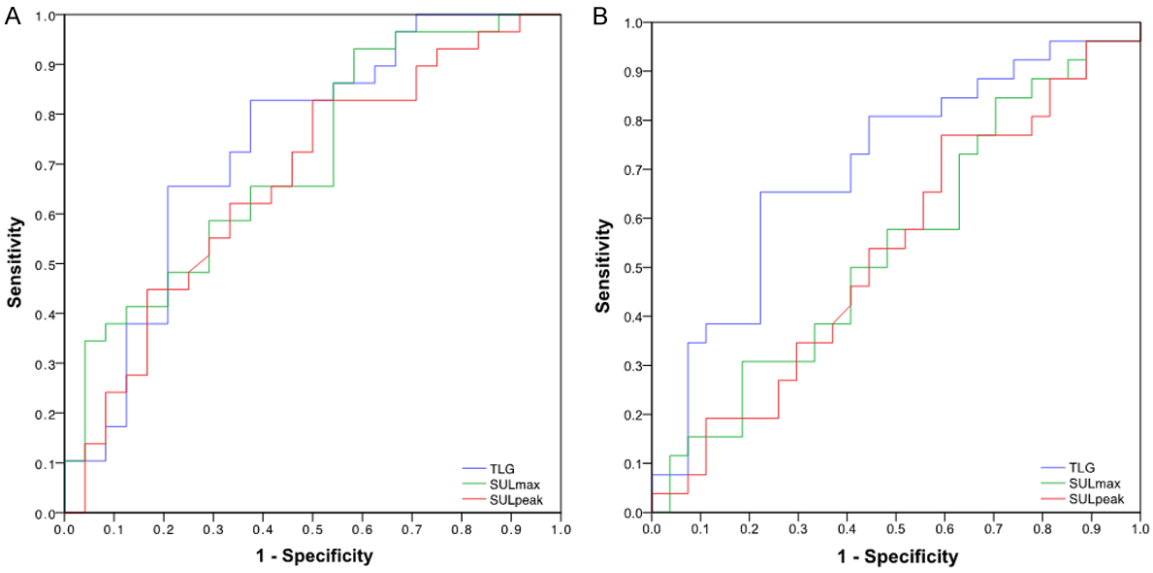

Figure 3. Receiver operating curves for predicting short progression free survival (< median) (A) and short overall survival (< median) (B) for all ^{18}F-FDG-PET/CT variables. The curves represent data from the 49 patients who were analyzable by all methods.

Table 6. Median progression free survival (months (95% confidence interval)) for the response categories for all 13 methods

	PMR	SMD	PMD	P
TLG (45/75%)	7.4 (0.8-13.9)	2.7 (2.5-2.8)	0.7 (Na)	< 0.001
TLG (50%)	14.6 (1.3-23.9)	2.7 (2.5-2.9)	1.3 (0.0-3.2)	0.003
TLG (40%)	7.4 (6.8-7.9)	2.7 (2.5-2.9)	2.0 (0.4-3.7)	0.005
TLG (30%)	7.4 (0.9-13.8)	2.7 (2.5-2.9)	2.0 (0.5-2.9)	< 0.001
TLG (25%)	7.4 (6.8-7.9)	2.7 (2.5-2.9)	2.4 (1.0-3.9)	< 0.001
TLG (20%)	7.2 0.1-14.2)	2.7 (2.4-3.0)	2.4 (1.0-3.9)	< 0.001
SUL_{max} (25%)	7.2 (4.7-9.7)	2.7 (2.4-2.9)	2.5 (1.9-3.2)	0.003
SUL_{max} (15%)	5.8 (2.9-8.6)	2.6 (2.4-2.8)	2.5 (1.9-3.2)	0.001
SUL_{peak} (30%)	7.4 (6.8-7.9)	2.7 (2.4-2.9)	2.5 (1.3-3.8)	0.009
SUL_{peak} (25%)	7.4 (6.8-7.9)	2.7 (2.5-3.0)	2.5 (1.4-3.7)	0.003
SUL_{peak} (20%)	7.4 (6.8-7.9)	2.7 (2.5-3.0)	2.5 (1.4-3.7)	0.003
SUL_{peak} (15%)	7.4 (0.7-14.1)	2.7 (2.4-2.9)	2.5 (1.4-3.7)	0.004
Visual	7.4 (6.8-7.9)	2.7 (2.5-2.9)	2.5 (1.7-3.3)	0.027

P is the *p*-value from the log rank test, PMD is progressive metabolic disease, PMR is partial metabolic response and SMD is stable metabolic disease.

All ^{18}F-FDG-PET/CT scans were evaluable by visual evaluation, $\%SUL_{peak}$ and $\%SUL_{max}$, but the TLG delineation was not reliable in 3 patients owing to inclusion of background tissue in most cases and in one case of myriads of very small FDG avid lesions.

The median PFS (95% CI) was 2.73 (2.58-2.89) months and median OS was 8.02 (6.02-10.03) months, and after a median follow up time of 24.3 (18.5-30.0) months, 7 patients were still alive and 2 patients were still treated with erlotinib (for 12.6 and 23.0 months respectively at the end of follow-up).

The median PFS for the 8 EGFR-mut patients was 15.1 (2.0-28.1)

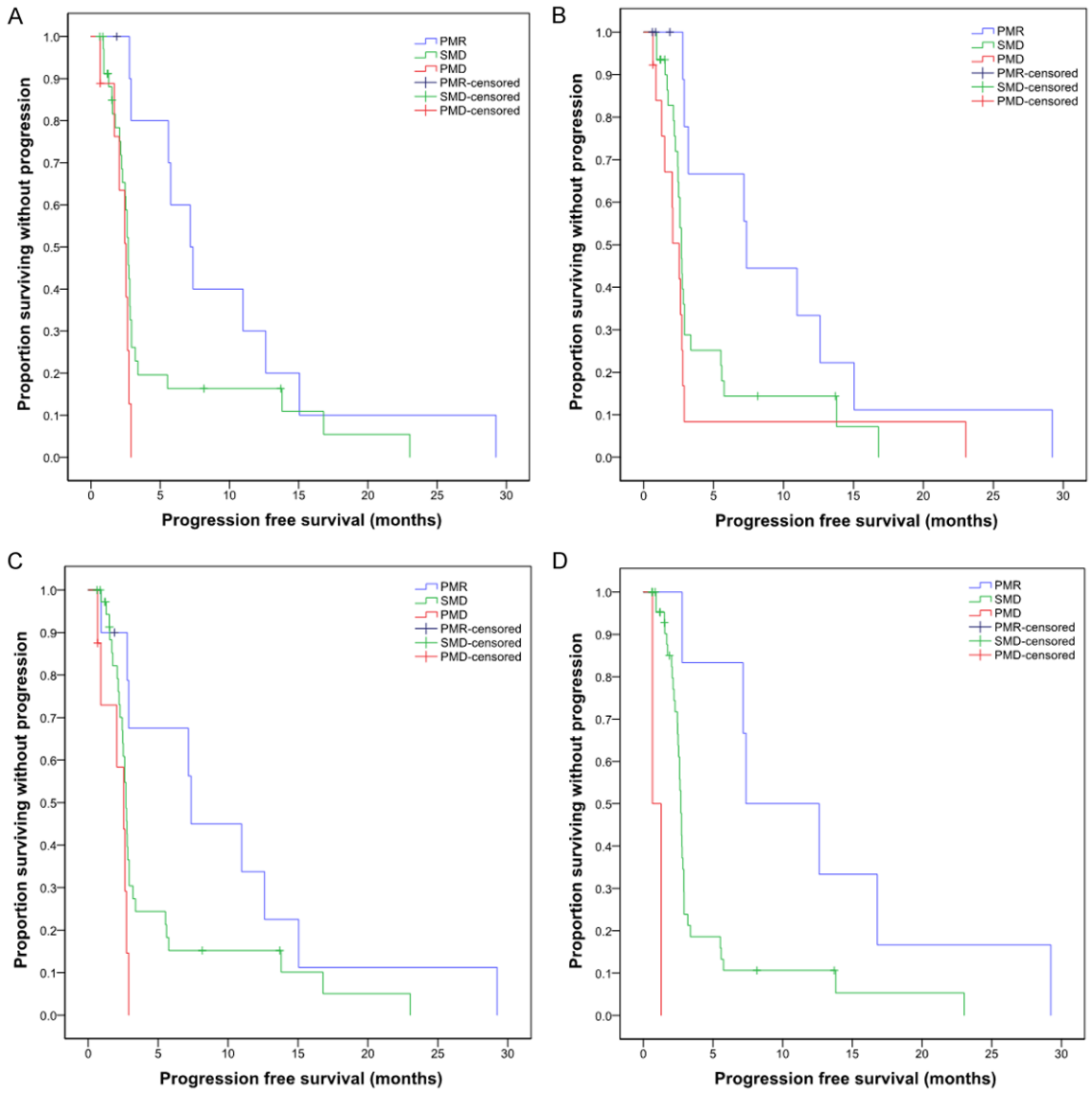

Figure 4. Kaplan-Meier curves for progression free survival for $\%SUL_{max}$ (25% cut-off) (A), visual evaluation (B) and the two PERCIST 1.0 methods: $\%SUL_{peak}$ (30% cut-off) (C) and %TLG (45/75% cut-off) (D). The median progression free survival for the different response categories are presented in **Table 6**.

months compared to the 2.6 (2.4-2.8) months, P < 0.001 for the EGFR-wt patients, and there was a highly significant difference between the median OS for the EGFR-mut patients of 16.7 months (95% CI not calculable) compared to the median OS of of 6.1 (2.9-9.3) months (P=0.008) for the EGFR-wt patients. We found responders among the EGFR-wt population, though, the number of EGFR-wt responders depended on the method used and the cut-off level applied for PMR, with the most sensitive method as an example, %TLG at 20% change identified 9 PMR in the EGFR-wt group (19.6%), and 7 PMR in the EGFR-mut group (87.5%).

Comparison of the $\%SUL_{peak}$, $\%SUL_{max}$ and %TLG to PFS and OS

The correlation analysis of $\%SUL_{peak}$, $\%SUL_{max}$, and %TLG with PFS and OS is presented in **Table 4**. All the variables showed a linear correlation to both PFS and OS, but %TLG provides the best correlation for both PFS (R=0.51 and P < 0.001) and OS (R=0.46 and P=0.001), the

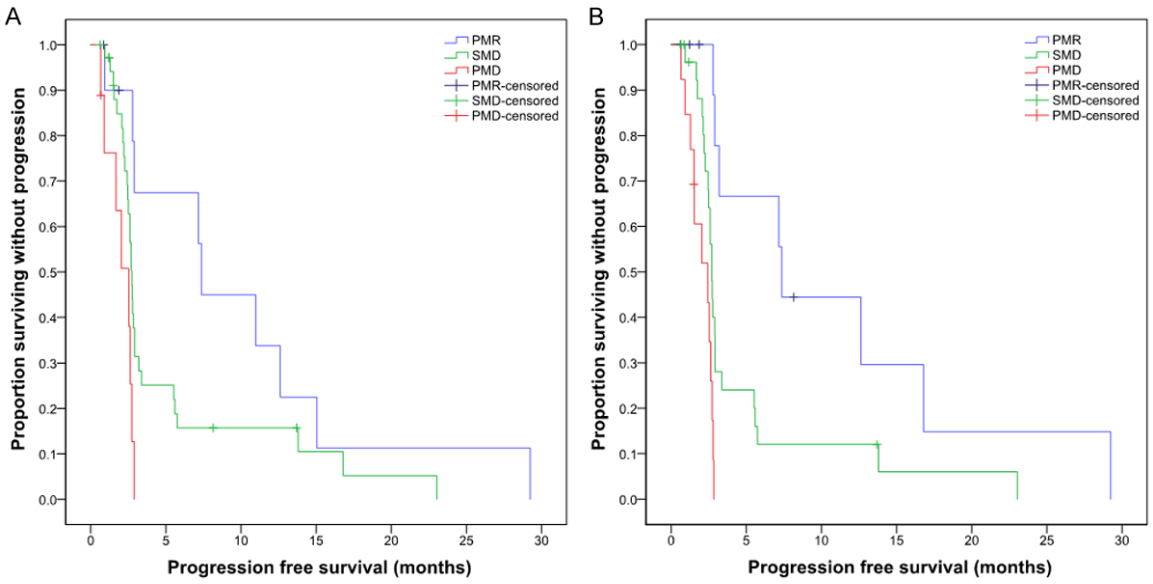

Figure 5. Kaplan-Meier curves for progression free survival for the optimal cut-off levels for $\%SUL_{peak}$ (20%) (A) and TLG (25%) (B), the median progression free survival for the different response categories are presented in **Table 6**.

scatterplots for %TLG are presented in **Figure 2**. The univariate cox regression analyses confirmed the significant correlation for all variables, again %TLG showed the highest hazard ratios (HRs) for predicting both PFS and OS (**Table 4**).

ROC analyses were performed using the median PFS and OS as divider (**Table 5**) with the highest AUC (95% CI) for PFS < median: 0.74 (0.60-0.88) for %TLG. For OS < median it was a similar situation, here with an AUC of 0.70 (0.56-0.85) for %TLG. Interestingly, $\%SUL_{max}$ and $\%SUL_{peak}$ did not predict OS < median (**Figure 3**).

Response categories predicting PFS and OS

PFS: All methods except visual evaluation, $\%SUL_{peak}$ at 15% and 30% cut-off and TLG at 40% cut-off showed an overall statistically significant difference at the 0.004 level. Comparing pairwise, considering both the difference between PMR/SMD and SMD/PMD, only %TLG showed $P < 0.05$ for PFS at all cut-off levels except 40% change, but no method was able to discriminate between both groups at the corrected 0.004 significance level. Data from all methods are presented in **Table 6**. Examples are presented in **Figure 4** for $\%SUL_{max}$ (25% cut-off), visual evaluation and the PERCIST methods, and in **Figure 5** for the most optimal cut-off levels for the PERCIST variables.

OS: Visual evaluation and $\%SUL_{max}$ failed to show different survival curves at the 0.004 level, but an overall difference in survival curves was found for %TLG at the 45/75%, 30%, 25% and 20% cut-off levels, and for $\%SUL_{peak}$ at 25%, 20% and 15% cut-off, though all methods failed to discriminate between both PMR/SMD and SMD/PMD even at the 0.05 level. The Kaplan-Meier curves for visual evaluation, $\%SUL_{peak}$ (20% cut-off) and %TLG (20% cut-off) are presented in **Figure 6** and data from all the methods are presented in **Table 7**, finally examples of metabolic response and progression are presented in **Figures 7** and **8**.

Discussion

The main results of the present study demonstrated that in this setting of very early response evaluation during erlotinib treatment in mainly EGFR-wt patients, we demonstrated that %TLG is the PET variable with the strongest correlation to PFS and OS compared to other often used single lesion variables. %TLG showed a significant difference between the response categories for both PFS and OS and the most optimal ROC analysis results. Hence, we consider this to be the most optimal method for predicting survival in this setting and strongly recommend adherence to the PERCIST 1.0 guidelines in order to reach agreement of choice of measurement parameter. This will hopefully lead us towards a higher comparability of

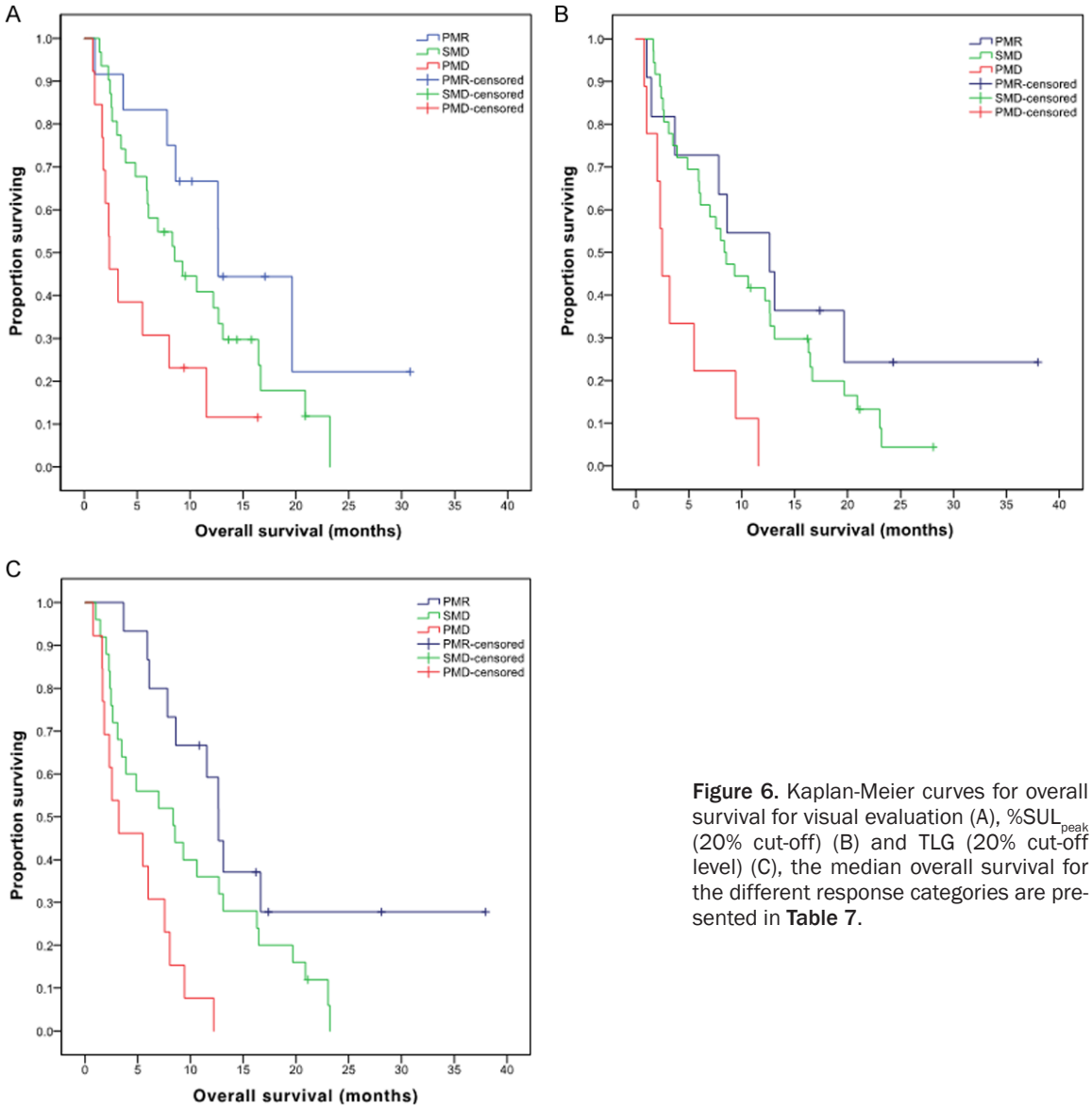

Figure 6. Kaplan-Meier curves for overall survival for visual evaluation (A), $\%SUL_{peak}$ (20% cut-off) (B) and TLG (20% cut-off level) (C), the median overall survival for the different response categories are presented in **Table 7**.

response evaluation studies in the future. We find that the 20-25% change defining PMR and PMD is the optimal level for predicting both PFS and OS in this early setting, consistent with what we have previously demonstrated for prediction of CT response.

We have previously shown that %TLG at a 25% cut-off level performs better than $\%SUL_{peak}$ and $\%SUL_{max}$ for prediction of response on CT scans performed after 9-11 weeks of erlotinib treatment in this population [19]. Previously, a number of ^{18}F-FDG- and ^{18}F-fluorothymidine-PET/CT studies tested various measurement variables and cut-off levels for response in a population comparable to ours including advanced NSCLC patients treated with erlotinib and scanned after 1 week of treatment [6, 15, 23]. These studies demonstrated that SUV_{max} and SUV_{peak} predicted both PFS and OS. In contrast to our present study TLG was not found to be superior to SUV_{max} or SUV_{peak}, considering early change in ^{18}F-FDG-uptake, in fact, they found that PMR/non-PMR by TLG was not associated with PFS at any cut-off levels (20%, 30% and 45% were tested), whereas SUV_{max} and SUV_{peak} were. In this present study we demonstrated statistically significant different survival curves for many of the methods we tested for PFS, including the TLGs where we also found the lowest *p*-values when comparing the response categories pairwise, indicating that TLG in fact is a strong predictor of PFS.

Table 7. Median overall survival (months (95% confidence interval)) for the response categories for all 13 methods

	PMR	SMD	PMD	P
TLG (45/75%)	12.6 (4.8-20.5)	7.6 (4.5-10.7)	0.7 (Na)	< 0.001
TLG (50%)	Not reached	7.6 (4.8-10.7)	1.3 (0.0-3.2)	0.017
TLG (40%)	12.6 (7.4-17.9)	2.7 (2.5-2.9)	2.0 (0.4-3.7)	0.019
TLG (30%)	13.1 (6.6-19.6)	2.7 (2.5-2.9)	2.0 (0.5-2.9)	0.001
TLG (25%)	12.7 (11.9-13.4)	2.7 (2.5-2.9)	2.4 (1.0-3.9)	0.001
TLG (20%)	12.7 (10.7-14.6)	2.7 (2.4-3.0)	2.4 (1.0-3.9)	< 0.001
SUL_{max} (25%)	13.1 (6.1-20.1)	2.7 (2.4-2.9)	2.5 (1.9-3.2)	0.008
SUL_{max} (15%)	12.7 (9.3-16.0)	2.6 (2.4-2.8)	2.5 (1.9-3.2)	0.041
SUL_{peak} (30%)	12.6 (5.6-19.6)	2.7 (2.4-2.9)	2.5 (1.3-3.8)	0.004
SUL_{peak} (25%)	12.6 (5.6-19.6)	2.7 (2.5-3.0)	2.5 (1.4-3.7)	0.002
SUL_{peak} (20%)	12.6 (6.9-18.3)	2.7 (2.5-3.0)	2.5 (1.4-3.7)	0.002
SUL_{peak} (15%)	12.6 (7.4-17.9)	2.7 (2.4-2.9)	2.5 (1.4-3.7)	0.003
Visual	12.6 (12.6-12.7)	2.7 (2.4-2.9)	2.5 (1.7-3.3)	0.021

P is the *p*-value from the log rank test, PMD is progressive metabolic disease, PMR is partial metabolic response and SMD is stable metabolic disease.

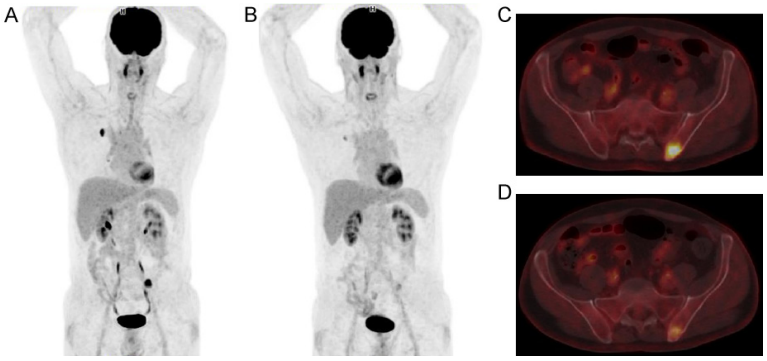

Figure 7. Patient example with whole body PET scan and trans axial fused PET/CT: A 58-year-old male before (A and C) and after 9 days of erlotinib treatment (B and D) with partial metabolic response for all methods.

Consistent with the results of our study, a similar Italian study including 53 stage IV NSCLC patients demonstrated a significant difference between the response categories according to the "EORTC" using SUV_{max}, evaluating the response as early as 2 days after initiating erlotinib treatment, they report median survival times for the three response groups, which are very similar to the values demonstrated in the present study [7].

Another study including 40 advanced NSCLC patients, mostly EGFR-mut, showed that TLG (40% of SUV_{max}), SUV_{peak}, SUV_{max} and SUV_{mean} after 6 weeks of treatment were associated with OS. The percentage change in SUV_{peak}, SUV_{max} and SUV_{mean} in the primary tumor was associated with OS, but percentage change in TLG was not [24]. Again, this is in contrast to our results, we believe that including all measurable lesions in the TLG evaluation (total tumor burden evaluation), though often tedious in advanced disease stages, could be responsible for the advantage we find with the TLG measurements.

Interestingly, we did find responders in the EGFR-wt group, up to 19.6% with the most sensitive method, which is consistent with findings in previous studies [6, 13]. This is an important observation, because our regular selection for TKI treatment is based on the EGFR mutation status. If only EGFR-mut patients are offered this treatment, some wild-type patients miss the potential benefit. The reason why some EGFR-wt patients respond to the TKI treatment is less clear than for the EGFR-mut patients, but unknown or rare mutations we do not yet test for in the daily clinic could be responsible for the "wild type" responses.

Another important study in 19 gefitinib treated stage III-IV NSCLC patients demonstrated that %change in SUV_{max} was predictive of both PFS and OS but that the response categories according to EORTC was only associated with PFS, and not associated with OS [25]. However, the small number of patients included may explain their negative result. In contrast to their results, the present study in a larger population showed a significant association for the EORTC categories (%SUL_{max} at both 15% and 25% change) for both PFS and OS with regard to both the univariate cox regression (data not shown) and Kaplan-Meier analysis.

In a study of 22 patients, Benz *et al* found, in concordance with our results, a significant association between the response categories

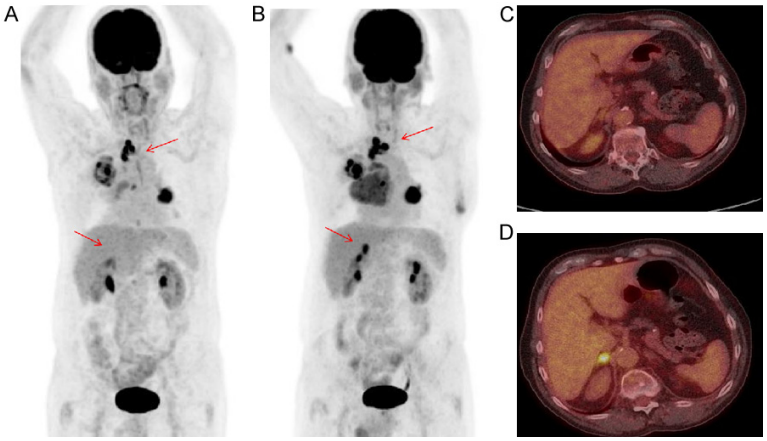

Figure 8. Patient example with whole body PET scan and trans axial fused PET/CT: A 69-year-old male before (A and C) and after 12 days of erlotinib treatment (B and D), with progression for all methods with new lesions in both the right adrenal gland (D) and a thoracic lymph node. Red arrows indicate the lesions with the highest impact on the response evaluations.

after 2 weeks of erlotinib treatment and PFS as well as OS using the PERCIST criteria, though on SUV_{max} values [26].

In a small population of 23 advanced NSCLC patients, it was studied and demonstrated that TLG evaluation in up to five lesions was superior in predicting PFS and OS to SUV_{max} in the hottest lesion (as SUL_{peak} according to PERCIST does), supporting the results of the present study [18]. Furthermore, the difference between the single value evaluation and the total disease evaluation was reported to be owing "bone flare" in some of the patients [18]. In the present study, we have one patient with suspected bone flare affecting the visual evaluation but not "hot" enough to affect the SUL_{peak} evaluation. Therefore, it should be considered to exclude seemingly progressive bone lesions in this population for evaluation, especially, if the rest of the lesions do not show progression.

We have previously studied the inter observer agreement for response evaluation for both semi-quantitative evaluation according to PERCIST and for visual evaluation in locally advanced NSCLC patient and found that the agreement is strong for both methods but stronger for the semi-quantitative method than for visual evaluation, allowing us to continue this larger study with one experienced observer only [27].

The strengths of our study are an overall strict adherence to the PERCIST recommendations for standardization and a head-to-head comparison of PERCIST, EORTC, total lesion TLG and visual evaluation combined with an analysis of various cut-off levels for metabolic response and progression in a reasonable size population including a large group of EGFR-wt patients. In these patients, an early response evaluation is essential if they are to benefit from treatment with TKIs. The main weakness of this study is the variation in times between the baseline scans and initiation of treatment, this time period should of course be very short when the response evaluation is performed so early into treatment. This could cause an underestimation of the ^{18}F-FDG-uptakeand result in fewer cases of PMR than would be detected if the time between baseline scan and treatment was short.

Disclosure of conflict of interest

None.

Address correspondence to: Dr. Joan Fledelius, Department of Nuclear Medicine, Herning Regional Hospital, DK-7400 Herning, Denmark. Tel: +45 31603316; Fax: +45 78437361; E-mail: joan.fledelius@rm.dk

References

[1] Sharma SV, Bell DW, Settleman J and Haber DA. Epidermal growth factor receptor mutations in lung cancer. Nat Rev Cancer 2007; 7: 169-181.

[2] Pérez-Soler R, Chachoua A, Hammond LA, Rowinsku EK, Huberman M, Karp D, Rigas J, Clark MG, Santabárbara P and Bonomi P. Determinants of tumor response and survival with erlotinib in patients with non-small-cell lung cancer. J Clin Oncol 2004; 22: 3238-3247.

[3] Shepherd FA, Pereira JR, Ciuleanu T, Eng HT, Hirsh V, Thongprasert S, Campos D, Maoleekoonpiroj S, Smylie M, Martins R, van Kooten M, Dediu M, Findlay B, Tu D, Johnston D,

Bezjak A, Clark G, Santabarbara P and Seymour L. Erlotinib in previously treated non-small-cell lung cancer. N Engl J Med 2005; 353: 123-132.

[4] Takahashi R, Hirata H, Tachibana I, Shimosegawa E, Inoue A, Nagatomo I, Takeda Y, Kida H, Goya S, Kijima T, Yoshida M, Kumagay T, Kumanogoh A, Okumura M, Hatazawa J and Kawase I. Early [^{18}F]fluorodeoxyglucose positron emission tomography at two days of gefitinib treatment predicts clinical outcome in patients with adenocarcinoma of the lung. Clin Cancer Res 2012; 18: 220-228.

[5] Lynch TJ, Bell DW, Sordella R, Gurubhagavatula RA, Okimoto S, Brannigan BW, Harris PL, Haserlat SM, Supko JG, Haluska FG, Louis DN, Christiani DC, Settleman J and Haber DA. Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. N Engl J Med 2004; 350: 2129-2139.

[6] Zander T, Scheffler M, Nogova L, Kobe C, Engel-Riedel W, Hellmich M, Papachristou I, Toepelt K, Draube A, Heukamp L, Buettner R, Ko YD, Ullrich RT, Smit E, Boellard R, Lammertsma AA, Hallek M, Jacobs AH, Schlesinger A, Schulte K, Querings S, Stoelben E, Neumaier B, Thomas RK, Dietlein M and Wolf J. Early prediction of nonprogression in advanced non-small-cell lung cancer treated with erlotinib by using [^{18}F]Fluorodeoxyglucose and [^{18}F]Fluorothymidine positron emission tomography. J Clin Oncol 2011; 29: 1701-1708.

[7] Tiseo M, Ippolito M, Scarlattei M, Spadaro P, Cosentino S, Latteri F, Ruffini L, Bartolotti M, Bortesi B, Fumarola C, Caffara C, Cavazzoni A, Alfieri RR, Petronini PG, Bordonaro R, Bruzzi P, Ardizzoni A and Soto Parra HJ. Predictive and prognostic value of early response assessment using ^{18}FDG-PET in advanced non-small cell lung cancer patients treated with erlotinib. Cancer Chemother Pharmacol 2014; 73: 299-307.

[8] Winther-Larsen A, Fledelius J, Demuth C, Bylov CM, Meldgaard P and Sorensen BS. Early change in FDG-PET signal and plasma cell-free DNA level predicts erlotinib response in EGFR wild-type NSCLC patients. Transl Oncol 2016; 9: 505-511.

[9] Su H, Bodenstein C, Dumont RA, Seimbille Y, Dubinett S, Phelps ME, Herschman H, Czernin J and Weber W. Monitoring tumor glucose utilization by positron emission tomography for the prediction of treatment response to epidermal growth factor receptor kinase inhibitors. Clin Cancer Res 2006; 12: 5659-5667.

[10] Tuma RS. Sometimes size doesn't matter: reevaluating RECIST and tumor response rate endpoints. J Natl Cancer Inst 2006; 98: 1272-1274.

[11] Sunaga N, Oriuchi N, Kaira K, Yanagitani N, Tomizawa Y, Hisada T, Ishizuka T, Endo K and Mori M. Usefulness of FDG-PET for early prediction of the response to gefitinib in non-small cell lung cancer. Lung Cancer 2008; 59: 203-210.

[12] Hachemi M, Couturier O, Vervueren L, Fosse P, Lacœuille F, Urban T and Hureaux J. [^{18}F]FDG positron emission tomography within two weeks of starting erlotinib therapy can predict response in non-small cell lung cancer patients. PLoS One 2014; 9: e87629.

[13] Mileshkin L, Hicks RJ, Hughes BG, Mitchell PL, Charu V, Gitlitz BJ, Macfarlane D, Solomon B, Amler LC, Yu W, Pirzkall A and Fine BM. Changes in ^{18}F-fluorodeoxyglucose and ^{18}F-fluorodeoxythymidine positron emission tomography imaging in patients with non-small cell lung cancer treated with erlotinib. Clin Cancer Res 2011; 17: 3304-3315.

[14] Aukema TS, Kappers I, Olmos RaV, Codrington HE, van Tinteren H, van Pel R and Klomp HM. Is ^{18}F-FDG PET/CT useful for the early prediction of histopathologic response to neoadjuvant erlotinib in patients with non-small cell lung cancer? J Nucl Med 2010; 51: 1344-1348.

[15] Kahraman D, Holstein A, Scheffler M, Zander T, Nogova L, Lammertsma AA, Boellaard R, Neumaier B, Dietlein M, Wolf J and Kobe C. Tumor lesion glycolysis and tumor lesion proliferation for response prediction and prognostic differentiation in patients with advanced non-small cell lung cancer treated with erlotinib. Clin Nucl Med 2012; 37: 1058-1064.

[16] van Gool MH, Aukema TS, Schaake EE, Rijna H, Valdés Olmos RA, van Pel R, Burgers SA, van Tinteren H, Klomp HM; NEL Study Group. Timing of metabolic response monitoring during erlotinib treatment in non-small cell lung cancer. J Nucl Med 2014; 55: 1081-1086.

[17] Moon SH, Cho SH, Park LC, Ji JH, Sun JM, Ahn JS, Park K, Choi JY and Ahn MJ. Metabolic response evaluated by 18F-FDG PET/CT as a potential screening tool in identifying a subgroup of patients with advanced non-small cell lung cancer for immediate maintenance therapy after first-line chemotherapy. Eur J Nucl Med Mol imaging 2013; 40: 1005-1013.

[18] Ho KC, Fang YD, Chung HW, Liu YC, Chang JW, Hou MM, Yang CT, Cheng NM, Su TP and Yen TC. TLG-S criteria are superior to both EORTC and PERCIST for predicting outcomes in patients with metastatic lung adenocarcinoma treated with erlotinib. Eur J Nucl Med Mol Imaging 2016; 43: 2155-2165.

[19] Fledelius J, Winther-Larsen A, Khalil AA, Bylov CM, Hjorthaug K, Bertelsen A, Frøkiær J and Meldgaard P. ^{18}F-FDG-PET/CT for very early response evaluation predicts CT response in Er-

lotinib treated NSCLC patients-A comparison of assessment methods. J Nucl Med 2017; 58: 1931-1937.

[20] Winther-Larsen A, Fledelius J, Sorensen BS and Meldgaard P. Metabolic tumor burden as marker of outcome in advanced EGFR wild-type NSCLC patients treated with erlotinib. Lung Cancer 2016; 94: 81-87.

[21] Mac Manus MP, Hicks RJ, Matthews JP, McKenzie A, Rischin D, Salminen EK and Ball DL. Positron emission tomography is superior to computed tomography scanning for response-assessment after radical radiotherapy or chemoradiotherapy in patients with non-small-cell lung cancer. J Clin Oncol 2003; 21: 1285-1292.

[22] Wahl RL, Jacene H, Kasamon Y and Lodge MA. From RECIST to PERCIST: evolving considerations for PET response criteria in solid tumors. J Nucl Med 2009; 50 Suppl 1: 122S-50S.

[23] Kahraman D, Scheffler M, Zander T, Nogova L, Lammartsmaa AA, Boellaard R, Neumaier B, Ullrich RT, Holstein A, Dietlein M, Wolf J and Kobe C. Quantitative analysis of response to treatment with erlotinib in advanced non-small cell lung cancer using 18F-FDG and 3'-deoxy-3'-18F-fluorothymidine PET. J Nucl Med 2011; 52: 1871-1877.

[24] Cook GJ, O'Brien ME, Siddique M, Chicklore S, Loi HY, Sharma B, Punwani R, Bassett P, Goh V and Chua S. Non-small cell lung cancer treated with erlotinib: heterogeneity of ^{18}F-FDG uptake at response and prognosis. Radiology 2015; 276: 883-893.

[25] Kanazu M, Maruyama K, Ando M, Asami K, Ishii M, Uehira K, Minomo Y, Matsuda Y, Kawaguchi, Atagi S, Ogawa Y, Kusunoki Y, Takada M and Kubo A. Early pharmacodynamic assessment using ^{18}F-fluorodeoxyglucose positron-emission tomography on molecular targeted therapy and cytotoxic chemotherapy for clinical outcome prediction. Clin Lung Cancer 2014; 15: 182-187.

[26] Benz MR, Herrmann K, Walter F, Garon EB, Reckamp KL, Figlin R, Phelps ME, Weber WA, Czernin J and Allen-Auerbach MS. ^{18}F-FDG PET/CT for Monitoring treatment responses to the epidermal growth factor receptor inhibitor erlotinib. J Nucl Med 2011; 52: 1684-1689.

[27] Fledelius J, Khalil AA, Hjorthaug K and Frøkiær J. Inter-observer agreement improves with PERCIST 1.0 as opposed to qualitative evaluation in non-small cell lung cancer patients evaluated with F-18-FDG PET/CT early in the course of chemo-radiotherapy. EJNMMI Res 2016; 6: 71.