^{18}F-FDG PET/CT for Very Early Response Evaluation Predicts CT Response in Erlotinib-Treated Non–Small Cell Lung Cancer Patients: A Comparison of Assessment Methods

Joan Fledelius[1], Anne Winther-Larsen[2], Azza A. Khalil[3], Catharina M. Bylov[4], Karin Hjorthaug[5], Aksel Bertelsen[1], Jørgen Frøkiær[5], and Peter Meldgaard[3]

[1]Department of Nuclear Medicine, Herning Regional Hospital, Herning, Denmark; [2]Department of Clinical Biochemistry, Aarhus University Hospital, Aarhus, Denmark; [3]Department of Oncology, Aarhus University Hospital, Aarhus, Denmark; [4]Department of Radiology, Aarhus University Hospital, Aarhus, Denmark; and [5]Department of Nuclear Medicine and PET Centre, Aarhus University Hospital, Aarhus, Denmark

The purpose of this study was to determine which method for early response evaluation with ^{18}F-FDG PET/CT performed most optimally for the prediction of response on a later CT scan in erlotinib-treated non–small cell lung cancer patients. **Methods:** ^{18}F-FDG PET/CT scans were obtained before and after 7–10 d of erlotinib treatment in 50 non–small cell lung cancer patients. The scans were evaluated using a qualitative approach and various semiquantitative methods including percentage change in SUVs, lean body mass–corrected (SUL) SUL_{peak}, SUL_{max}, and total lesion glycolysis (TLG). The PET parameters and their corresponding response categories were compared with the percentage change in the sum of the longest diameter in target lesions and the resulting response categories from a CT scan obtained after 9–11 wk of erlotinib treatment using receiver-operating-characteristic analysis, linear regression, and quadratic-weighted κ. **Results:** TLG delineation according to the PERCIST showed the strongest correlation to sum of the longest diameter ($R = 0.564$, $P < 0.001$), compared with SUL_{max} ($R = 0.298$, $P = 0.039$) and SUL_{peak} ($R = 0.402$, $P = 0.005$). For predicting progression on CT, receiver-operating-characteristic analysis showed area under the curves between 0.79 and 0.92, with the highest area under the curve of 0.92 (95% confidence interval [CI], 0.84–1.00) found for TLG (PERCIST). Furthermore, the use of a cutoff of 25% change in TLG (PERCIST) for both partial metabolic response and progressive metabolic disease, which is the best predictor of the CT response categories, showed a κ-value of 0.53 (95% CI, 0.31–0.75). This method identifies 41% of the later progressive diseases on CT, with no false-positives. Visual evaluation correctly categorized 50%, with a κ-value of 0.47 (95% CI, 0.24–0.70). **Conclusion:** TLG (PERCIST) was the optimal predictor of response on later CT scans, outperforming both SUL_{peak} and SUL_{max}. The use of TLG (PERCIST) with a 25% cutoff after 1–2 wk of treatment allows us to safely identify 41% of the patients who will not benefit from erlotinib and stop the treatment at this time.

Key Words: FDG PET/CT; lung cancer; PERCIST 1.0; early response evaluation; TLG

J Nucl Med 2017; 58:1931–1937
DOI: 10.2967/jnumed.117.193003

Treatment with tyrosine kinase inhibitors (TKIs) has proven to be effective in non–small cell lung cancer (NSCLC) patients, with response rates of 10%–20% in unselected populations. Subgroups of patients have been identified with good and sometimes prolonged results (*1–4*).

At present, the selection of patients is done by detecting mutations in the epidermal growth factor receptor (EGFR) genes. EGFR mutation–positive patients (EGFR-mut) have higher response rates than EGFR wild-type (EGFR-wt) patients (*5–9*). However, a subgroup of EGFR-wt patients (*1,5,10*) also benefits from TKI treatment, suggesting that the selection of patients depending solely on mutation status will exclude some patients from a potential treatment benefit.

At our institution, erlotinib treatment is offered to nonoperable EGFR-mut, NSCLC patients as first-line treatment and to EGFR-wt patients as second- or third-line treatment. An important challenge with this treatment is how to evaluate the response, because TKIs are known to have mostly cytostatic effects (*11,12*) as opposed to cytotoxic effect. Therefore, a follow-up CT scan is routinely obtained 9–11 wk into the treatment for response evaluation. Considering that stage III–IV NSCLC patients in general have a short remaining life expectancy (*4,13,14*), it is essential to discontinue a futile treatment course as early as possible.

Preclinical studies have shown that TKI-sensitive cells downregulate their glucose uptake early after exposure to TKI treatment (*1,11*). Clinically ^{18}F-FDG PET/CT performed early during TKI treatment has shown promise for predicting both anatomic response and survival (*1,5,15,16*), and for predicting histopathologic response (*7,17*).

The many methods used for response evaluation with ^{18}F-FDG PET include visual evaluation, change in SUV corrected for body weight (SUV) for maximum-intensity voxels, and mean value in a standardized volume of 1.2-cm (in diameter) volume of interest (SUV_{max} and SUV_{peak}), as well as more complex volume-based parameters such as total lesion glycolysis (TLG) with various ways of delineating the lesions.

PERCIST 1.0 from 2009 (*18*) uses the SUL_{peak} as the standard but also suggests TLG as a supplemental analysis. We have previously shown that percentage change in TLG is a promising predictor of CT response in a subgroup of EGFR-wt patients (*10*).

The aim of this study was therefore to identify which specific method of a selection of commonly used methods (including the

Received Mar. 20, 2017; revision accepted Apr. 25, 2017.
For correspondence or reprints contact: Joan Fledelius, Department of Nuclear Medicine, Herning Regional Hospital, DK-7400 Herning, Denmark.
E-mail: joan.fledelius@rm.dk
Published online May 10, 2017.

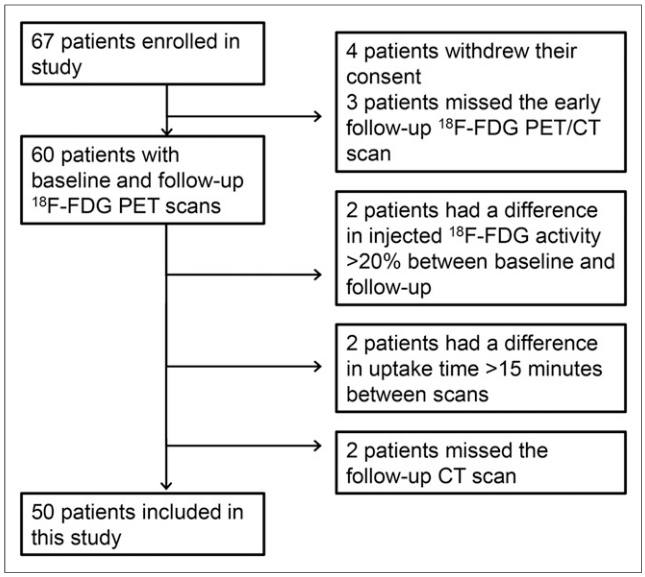

FIGURE 1. Patient selection.

PERCIST 1.0 methods) was the optimal for predicting the CT response, early during erlotinib treatment. We hypothesized that an evaluation of the total disease burden would improve response evaluation compared with single-hottest-lesion evaluation. We focused on safely selecting patients after 7–10 d of treatment, who will not have a treatment effect, enabling us to discontinue futile treatment at this early time point.

MATERIALS AND METHODS

The results are reported as according to Standards for Reporting Diagnostic Accuracy studies (STARDS) 2015.

Patients

This retrospective study evaluated ^{18}F-FDG PET/CT scans, in compliance with the PERCIST 1.0 to ensure comparability of images, from 50 patients enrolled in prospective single-center study originally including 67 consecutive patients with stage III–IV NSCLC between April 2013 and August 2015. The study was approved by the Central Denmark Region Committee on Biomedical Research Ethics (no. 1-10-72-19-12), and subjects signed an informed consent form. The study was reported to ClinicalTrials.gov (NCT02043002). All patients were candidates for palliative erlotinib treatment. Detailed inclusion criteria were previously published (*19*). The selection of patients included in the present study is presented in Figure 1.

All patients underwent an ^{18}F-FDG PET/CT scan before (baseline) and 7–10 d after initiation of erlotinib treatment (follow-up). CT scans of the chest and abdomen were acquired before and after 9–11 wk of treatment, or earlier on clinical indication. Testing for EGFR mutations had been performed in all patients with adenocarcinoma as part of the routine diagnostic workup by use of the Therascreen EGFR RGQ PCR kit (QIAGEN) according to the manufacturer's protocol, and on the basis of this, patients were categorized as either EGFR-wt or EGFR-mut.

^{18}F-FDG PET/CT Acquisition and Evaluation

All ^{18}F-FDG PET/CT scans were obtained on a combined PET/CT scanner (Siemens Biograph TruePoint 40; Siemens Healthcare GMbH) at the Department of Nuclear Medicine and PET-Centre, Aarhus University Hospital, Denmark, using the same scanner type and acquisition and reconstruction protocols as previously published in detail (*19*). In brief, all patients had a fasting period of at least 6 h, a blood glucose concentration of less than 11 mM, and an uptake time of 60 ± 10 min between injection of 5 MBq/kg ± 10% of ^{18}F-FDG and scan start (3 min per bed position). A whole-body low-dose CT scan (50 mAS, 120 kVp) was acquired.

All ^{18}F-FDG PET/CT scans were evaluated by 1 experienced nuclear medicine physician using Siemens Syngovia software (Siemens Healthcare GMbH). The evaluator was masked to the outcome and the result of the following CT scan. Evaluation of response was performed as visual evaluation described by Mac Manus et al. (*20*), considering both the overall change in ^{18}F-FDG uptake and the appearance of new ^{18}F-FDG–avid lesions (visual); the percentage change in SUL_{peak} in the lesion with the highest uptake at baseline and follow-up (not necessarily the same lesion) (SUL_{peak}) according to PERCIST 1.0 (*18*); and the percentage change in global TLG with various delineation methods: at SUL_{mean} + 2 SDs in a spheric 3-cm region of interest in the right lobe of the liver (SUL_{mean} [liver]) (TLG [PERCIST]) and at 30% (TLG 30), 40% (TLG 40), and 50% (TLG 50) of SUL_{max}. A lesion was considered evaluable for all semiquantitative methods if SUL_{peak} was $1.5 \times SUL_{mean}$ (liver) + 2 SDs. Delineation was performed semiautomatically after a manual rough outlining of each lesion, resulting in an SUL_{mean} and a metabolic tumor volume for the delineated area. TLG for each lesion was calculated as $SUL_{mean} \times$ metabolic tumor volume. Global TLG was the sum of all measurable lesions TLGs.

All the methods allocated patients into 4 different response categories: progressive metabolic disease (PMD), stable metabolic disease (SMD), partial metabolic response (PMR), and complete metabolic response. When the classification methods were used, for SUL_{peak} all patients were categorized as PMD, if new lesions had appeared, independent of SUL_{peak}. This was not the case for the TLG categories, because any new lesion was included in the TLG calculations, if measurable. Multiple cutoffs for these response categories were tested. An overview of the methods is presented in

TABLE 1
Time (Days) Between Scans and First Day of Treatment

Imaging method	Median baseline	>4 wk (*n*)	>2 wk (*n*)	Median follow-up	Early CT (*n*)
CT	15 (range, 4–56)	5 (10%)	26 (52%)	77 (range, 20–85)	14 (28%)
^{18}F-FDG PET/CT	1 (range, 0–21)			8 (range, 2–23)	

>4 and >2 wk = no. of patients with interval between baseline CT and first day of treatment above 4 (and 2) wk; early CT = no. of patients with a follow-up CT earlier than 9 wk.

TABLE 2
Compliance with PERCIST 1.0 Standardization Criteria

Parameter	Baseline	Follow-up	Numeric difference	PERCIST 1.0	Adherence to PERCIST 1.0
Injected activity (MBq)					
Mean ± SD	340 ± 90	335 ± 87	18 ± 14	Baseline ± 20%	100% (50/50)
Range	197–609	199–618	0–54		
Glucose level (mmol/L)					
Mean ± SD	6.3 ± 0.9	6.4 ± 0.9	0.4 ± 0.5	<11 mM	100% (50/50)
Range	4.6–8.8	4.7–9.0	0.0–1.6		
Uptake time (min)					
Mean ± SD	58.7 ± 4.0	59.3 ± 4.9	5.1 ± 4.19	60 ± 10 min	98% (98/100*)
Range	51–69	48–72	0–15	Baseline ± 15 min	100% (50/50)

*Uptake time at both baseline and follow-up. Two patients with 48- and 72-min uptake time at follow-up were included because difference between scans in both cases was within ±15 min.

Supplemental Table 1 (supplemental materials are available at http://jnm.snmjournals.org). SUL_{max} was measured for comparison but not used for classification into response categories.

CT Evaluation

The radiologic response was evaluated by 1 experienced radiologist on the first CT scan obtained after initiation of erlotinib according to the RECIST (version 1.1). Response evaluation included reporting of percentage change in the sum of the longest diameter (SLD) and the resulting response categories: progressive disease (PD), stable disease, partial response (PR), and complete response. We chose to dichotomize into PD versus non-PD (stable disease + PR) when appropriate, because this is used as the basis criteria to decide whether to continue or discontinue erlotinib treatment.

Statistical Analysis

Comparison of SUL_{peak}, SUL_{max}, and the TLGs to the SLD was performed using linear regression analysis; a significance level of 0.008 (Bonferroni adjustment for 6 methods) was used.

The predictive accuracy of SUL_{peak}, SUL_{max}, and the TLGs was evaluated by receiver-operating-characteristic (ROC) analyses, predicting PD versus non-PD on CT. The optimal cutoff, considering sensitivity and specificity equally important, was identified by visually locating the data point nearest the top left corner on the ROC curve.

To compare the ^{18}F-FDG PET response categories with the CT response categories, quadratic weighted κ-analyses were performed (*21*). As many patients as practically possible were included; a power calculation was not performed.

Statistical analyses were performed with SPSS statistics for Macintosh (version 23.0; IBM SPSS Statistics).

RESULTS

Scan Times and Standardization Parameters

The time between scans and treatment start is presented in Table 1. The indication for an early CT scan was in all cases suspected clinical progression. Ten patients showed PD and 4 stable disease; the 4 scanned earlier than 4 wk all showed PD on CT and PMD on ^{18}F-FDG PET/CT (TLG [PERCIST (25%)]). Data on injected ^{18}F-FDG activity, glucose level, and uptake time on population basis are presented in Table 2. A full description of glucose, uptake time, and injected ^{18}F-FDG activity is available in Supplemental Table 2.

All patients were analyzable by visual evaluation and SUL_{peak} (and SUL_{max}), whereas the TLG measurements were not feasible to perform in 3–7 patients depending on the delineation method. When the PERCIST delineation method was used, 3 patients were not analyzable: in one patient because of an uptake at follow-up close to the background level; another patient had myriads of small ^{18}F-FDG–avid lesions; and the last patient showed intense, diffuse uptake in the lung tissue impairing delineation of the tumor. TLG (50, 40, and 30) showed in total a further 4 nonevaluable patients because of relatively low uptake or proximity to the liver (higher background level). For 2 patients, SLD was not available (nonmeasurable lesions). None of the patients was classified as complete responders on either CT or ^{18}F-FDG PET/CT.

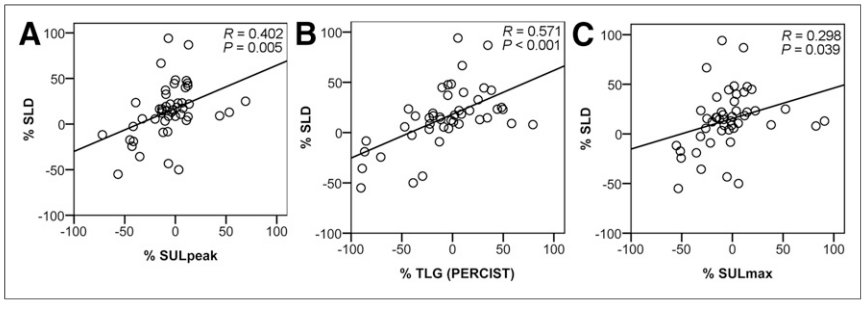

FIGURE 2. Correlation between SUL_{peak} (A), TLG (PERCIST) (B), and SUL_{max} (C) at 7–10 d and SLD measured on CT scans performed after 9–11 wk.

Comparing of SUL_{peak}, SUL_{max}, and TLGs with SLD

We found that all the measured parameters except SUL_{max} were significantly correlated to SLD, but TLG (PERCIST) showed the strongest correlation with an R of 0.571 ($P < 0.001$). Plots for SUL_{peak}, TLG (PERCIST), and SUL_{max} are presented in Figure 2; all variables are available in Supplemental Figure 1.

Dichotomizing the CT response into progression (PD) and non-PD resulted in 28 PD and 22 non-PD patients. For the

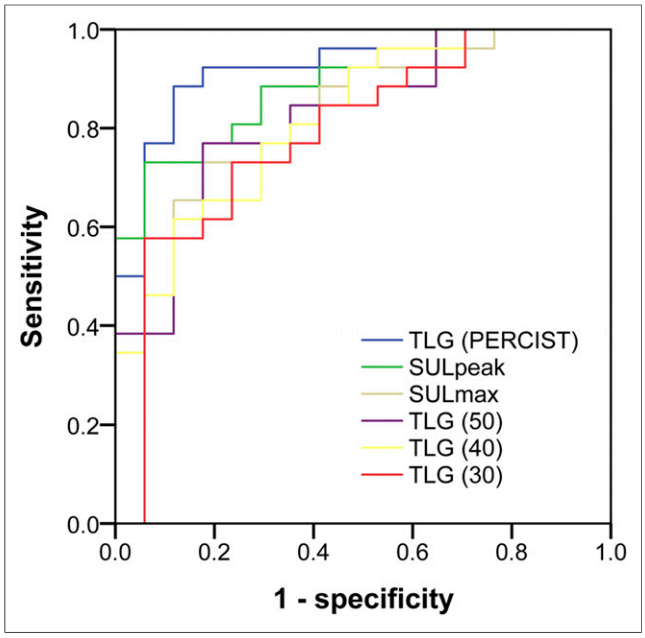

FIGURE 3. ROC curves for SUL_{peak}, SUL_{max}, and various TLG variations. Curves illustrate parameters' ability to predict PD on later CT scan obtained after 9–11 wk of treatment. Curves represent data from 43 patients who were analyzable by all methods.

PET methods involving a continuous variable, the ROC curves for the 43 patients analyzed by all methods are presented in Figure 3. As can be seen, TLG (PERCIST) has the highest AUC of 0.923 (95% confidence interval [CI], 0.840–1.00) (Table 3), although the confidence intervals overlap between methods. TLG (30) had the lowest AUC of 0.790 (95% CI, 0.698–0.949).

The highest sensitivity and specificity for predicting PD on CT when considering the sensitivity and the specificity equally important (sensitivity of 0.89 and specificity of 0.88) was seen for TLG (PERCIST). For a specificity of 1.00, the sensitivity for TLG (PERCIST) was 0.50, whereas SUL_{peak} had a slightly higher sensitivity of 0.58. Again, TLG (30) showed the lowest values. Interestingly, SUL_{peak} identified more anatomic responders than TLG (PERCIST), but fewer anatomic progressions, as shown in Figure 4. Plots for SUL_{max} and TLG (30–50) are found in Supplemental Figure 2.

Metabolic Response Categories Compared with RECIST

TLG (PERCIST [25%]) resulted in correct classification in 25 of 47 patients (53%) according to the CT response. One patient was classified more than 1 level different (early PMR and later PD on CT). Noticeably, all of the PMDs identified were classified as PD on later CT scans, identifying 11 of 27 PDs (41%). Even though SUL_{peak} provided the highest sensitivity at a specificity of 1 in the ROC analysis and all patients were analyzable by this method, it only identified 23 of 50 patients correctly (46%) when the response categories for the optimal cutoff (20%) were used. The visual method identified 25 of 50 patients correctly (50%), and both SUL_{peak} and visual found one (the same patient) more than 1 level different, possibly because of bone flare. The results for TLG (PERCIST [25%]), SUL_{peak} (20%), and visual evaluation are presented in Table 4.

In general, the κ-values were rather low, ranging from 0.23 to 0.53, presented in Table 5. The method with the highest value 0.53 (95% CI, 0.31–0.75) was TLG (PERCIST [25%]).

Dichotomizing the ^{18}F-FDG PET classification into PMD and non-PMD (PMR + SMD) for comparison to the similar dichotomization for CT (PD vs. non-PD) resulted for TLG (PERCIST [25%]) in a sensitivity of 0.41, a specificity of 1.00, a positive predictive value of 1.00, and a negative predictive value of 0.56. For the SUL_{peak} (20%), a slightly lower sensitivity of 0.29, a specificity of 1.00, a positive predictive value of 1.00, and a negative predictive value of 0.52 were found. The categorization tables, sensitivity, specificity, and predictive values for all methods are presented in Supplemental Table 3.

DISCUSSION

This study was performed to identify which PET assessment method was the most optimal for predicting the later CT response early during erlotinib treatment in unselected NSCLC patients. Early assessment of progression would enable discontinuation of nonbeneficial treatment after a few days of treatment (7–10 d).

The main result of this study was that TLG (PERCIST) had the strongest correlation to SLD, statistically significant even when

TABLE 3
AUCs and Sensitivity and Specificity from ROC Analyses

Method	AUC	95% CI		Sensitivity and specificity equally important			Specificity as high as possible		
		Lower	Upper	Sensitivity	Specificity	Cutoff	Sensitivity	Specificity	Cutoff
TLG (PERCIST)	0.923	0.842	1.000	0.89	0.88	−6.6%	0.50	1.00	9.2%
SUL_{peak}	0.887	0.791	0.983	0.77	0.82	−6.9%	0.58	1.00	2.1%
SUL_{max}	0.835	0.715	0.954	0.73	0.82	−4.5%	0.35	1.00	13.2%
TLG 50	0.824	0.698	0.949	0.77	0.82	3.0%	0.39	1.00	30.0%
TLG 40	0.821	0.696	0.946	0.77	0.71	−1.4%	0.35	1.00	24.6%
TLG 30	0.790	0.645	0.934	0.73	0.76	0.0%	0.58	0.94	11.9%

95% CI = 95% confidence interval; cutoff = corresponding optimal cutoff value for percentage change.
AUCs optimized by identifying point on curve closest to upper left corner (considering sensitivity and specificity equally important [middle columns]) and considering specificity of major importance, thereby avoiding any false progressions [right columns]).

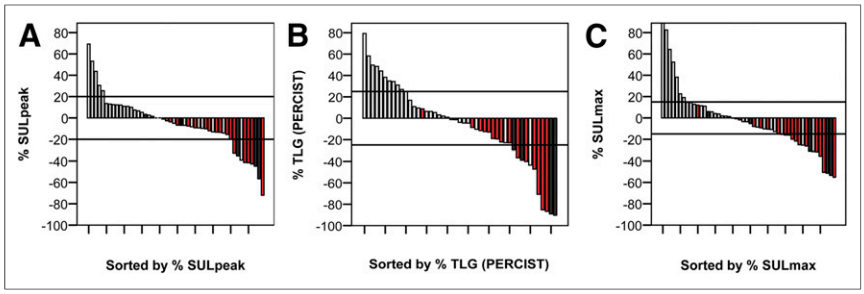

FIGURE 4. Waterfall plots for SUL_{peak} (A), TLG (PERCIST) (B), and SUL_{max} (C) showing distribution of CT categories (light gray = PD, red = SD, and black = PR). Horizontal reference lines represent optimal cutoff for PMR and PMD. For SUL_{max}, the horizontal reference lines represent 15% change suggested by European Organisation for Research and Treatment of Cancer criteria for early evaluation.

applying a significance level of 0.008. It was the best predictor of PD versus non-PD on CT, though the difference between methods was not statistically significant. A 25% change was the optimal cutoff and avoided false PMDs. It performed better than the PERCIST-suggested 75% increase and 45% decrease, possibly because of the early time point for evaluation.

There are many assessment methods for evaluating the treatment response with ^{18}F-FDG PET/CT. The most commonly used parameter is SUV_{max}, but SUL_{peak} is also frequently used, especially because this was recommended as the standard method in the PERCIST 1.0 from 2009 (*18*) and is also recommended for response evaluation in the latest European guidelines, together with the mentioning of the increasing interest of TLG (*22*). However, few studies have compared the performance of these methods simultaneously.

Previously, a comparison of 6 different parameters including SUV_{peak}, SUV_{max}, and SUV_{mean} using different delineation methods was performed for residual activity after 1 wk of erlotinib treatment (*23*). In this study, SUVs other than SUV_{max} were not superior at predicting progression on CT but they did not consider the percentage change; moreover, they did not include TLG.

In another study including 34 erlotinib-treated patients, it was demonstrated that percentage change in SUV_{peak} after 1 wk of treatment was predictive of PD versus non-PD on CT after 6 wk of treatment (*5*). However, there was no comparison to changes in TLG or other parameters.

Moon et al. compared SUV_{max}, SUV_{peak}, and TLG in 52 stage IV NSCLC patients before and after 4 cycles of platinum-based chemotherapy (*24*); consistent with the results of the present study, they found that change in TLG was a predictor of progression-free survival whereas change in SUV_{peak} and SUV_{max} were not, thus supporting our finding that TLG outperforms SUL_{peak} and SUL_{max}.

We found that a 25% change in TLG separated the response categories better than the 45%/75% cutoff from PERCIST 1.0. This is consistent with previous findings by Kahraman et al. demonstrating that a cutoff between 20% and 30% change is superior for predicting progression-free survival (*25*). Our results suggest that the same is true for predicting PD/non-PD on CT. When using 25% change in TLG (PERCIST), we still identified a large group of SMD patients who later showed PD on CT. To reduce the size of this category, we could use the ROC-determined cutoff value of 9.2% increase (or 10%), which reduced this group from 15 to 12. However, because the cutoff values are determined from the data one must expect the estimated sensitivity and specificity to be too optimistic, and the day-to-day variation (*26–28*) should also be considered. Taking this into account, we prefer to use the 25% cutoff.

The visual evaluation method performed well and was comparable to the more sensitive of the semiquantitative methods, SUL_{peak} and TLG (PERCIST). We have previously demonstrated in a chemotherapy-treated population of locally advanced NSCLC patients that there is a strong interobserver agreement for this method, but the agreement is stronger for SUL_{peak} change; however, TLG was not included in that study (*29*). This suggests that visual evaluation is a reliable alternative method for evaluation of the cases that are nonevaluable with the TLG methods.

The strengths of the present study are the strict adherence to standardization as according to PERCIST and the head-on comparison of 6 different methods for evaluating response, including the PERCIST 1.0 recommendations and the analysis on categorization using various cutoff levels, not only dichotomizing by ROC optimization.

The main limitation is the large variation in the interval between treatment and CT scans both at baseline and at follow-up. We find the variation in the time intervals for ^{18}F-FDG PET/CT scans acceptable, and comparable to other studies (*5*). The early progressions on CT within 4 wk after initiating treatment were considered true progressions and as such should not significantly

TABLE 4
TLG (PERCIST [25%]), SUL_{peak} (20%), and Visual Response Compared with RECIST 1.1

	TLG (PERCIST [25%])				SUL_{peak} (20%)				Visual			
CT	PMR	SMD	PMD	Total	PMR	SMD	PMD	Total	PMR	SMD	PMD	Total
PR	4	0	0	4	3	2	0	5	4	1	0	5
SD	6	10	0	16	5	12	0	17	5	11	1	17
PD	1	15	11	27	1	19	8	28	1	17	10	28
Total	1	25	11	47	9	33	8	50	10	29	11	50

SD = stable disease.

TABLE 5
κ-Values for All Methods

Method	κ (95% CI)
Visual	0.47 (0.24–0.70)*
SUL_{peak} 30%	0.39 (0.19–0.59)
SUL_{peak} 25%	0.41 (0.20–0.62)
SUL_{peak} 20%	0.41 (0.20–0.62)
SUL_{peak} 20%	0.42 (0.22–0.61)*
TLG (PERCIST) 45%/75%	0.25 (NC)
TLG (PERCIST) 50%	0.29 (0.20–0.39)
TLG (PERCIST) 40%	0.32 (0.16–0.47)
TLG (PERCIST) 30%	0.45 (0.24–0.67)
TLG (PERCIST) 25%	0.53 (0.31–0.75)*
TLG (PERCIST) 20%	0.49 (0.29–0.69)
TLG 50, 45%/75%	0.23 (NC)
TLG 50, 50%	0.27 (0.17–0.36)
TLG 50, 40%	0.25 (0.16–0.33)
TLG 50, 30%	0.35 (0.16–0.53)
TLG 50, 25%	0.33 (0.12–0.53)
TLG 50, 20%	0.38 (0.16–0.61)*
TLG 40, 45%/75%	0.27 (NC)
TLG 40, 50%	0.34 (0.16–0.52)
TLG 40, 40%	0.33 (0.16–0.50)
TLG 40, 30%	0.42 (0.18–0.66)
TLG 40, 25%	0.44 (0.20–0.69)*
TLG 40, 20%	0.43 (0.20–0.66)
TLG 30, 45%/75%	0.24 (NC)
TLG 30, 50%	0.33 (0.12–0.54)
TLG 30, 40%	0.38 (0.14–0.62)
TLG 30, 30%	0.32 (0.12–0.53)
TLG 30, 25%	0.38 (0.15–0.61)
TLG 30, 20%	0.41 (0.18–0.64)*

*Highlight of cutoff resulting in highest κ values for each method.

95% CI = 95% confidence interval; NC = not calculable by method used, described by Fleiss et al. (*21*).

Quadratic weighted κ-values (95% CIs) for all methods analyzed including various cutoffs for response and progression.

influence the PET results. The 5 cases of prolonged interval (>4 wk) between the baseline CT and treatment would potentially attenuate the CT response and thereby increase the risk for PET to show a better response than reflected on CT, but in 4 of the 5 cases TLG (PERCIST [25%]) agreed with the CT response. For a comparison of methods, we consider it usable because it applies to all methods. However, it needs closer attention in the future, because standardization is particularly important in the early response evaluation setting.

CONCLUSION

The present study demonstrates that using percentage change in global TLG delineated according to PERCIST tends to be a more sensitive method for early response evaluation of NSCLC patients during erlotinib treatment than highest intensity lesion evaluation (SUL_{peak}) and visual evaluation. The method allows for prediction of a later PD on CT identifying 41% of the PDs. We intend to use this finding in future studies and in the clinical setting, supplementing with visual evaluation in the few cases not evaluable by the TLG method.

DISCLOSURE

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

REFERENCES

1. Tiseo M, Ippolito M, Scarlattei M, et al. Predictive and prognostic value of early response assessment using 18FDG-PET in advanced non-small cell lung cancer patients treated with erlotinib. *Cancer Chemother Pharmacol.* 2014;73:299–307.
2. Sharma SV, Bell DW, Settleman J, Haber DA. Epidermal growth factor receptor mutations in lung cancer. *Nat Rev Cancer.* 2007;7:169–181.
3. Pérez-Soler R, Chachoua A, Hammond LA, et al. Determinants of tumor response and survival with erlotinib in patients with non-small-cell lung cancer. *J Clin Oncol.* 2004;22:3238–3247.
4. Shepherd FA, Pereira JR, Ciuleanu T, et al. Erlotinib in previously treated non–small-cell lung cancer. *N Engl J Med.* 2005;353:123–132.
5. Zander T, Scheffler M, Nogova L, et al. 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.
6. Takahashi R, Hirata H, Tachibana I, et al. 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.
7. Aukema TS, Kappers I, Olmos RA, et al. 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.
8. Lynch TJ, Bell DW, Sordella R, et al. 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.
9. Tsao MS, Sakurada A, Cutz JC, et al. Erlotinib in lung cancer: molecular and clinical predictors of outcome. *N Engl J Med.* 2005;353:133–144.
10. Winther-Larsen A, Fledelius J, Demuth C, Bylov CM, Meldgaard P, 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.
11. Su H, Bodenstein C, Dumont RA, et al. 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.
12. Tuma RS. Sometimes size doesn't matter: reevaluating RECIST and tumor response rate endpoints. *J Natl Cancer Inst.* 2006;98:1272–1274.
13. Fukuoka M, Yano S, Giaccone G, et al. Multi-institutional randomized phase II trial of gefitinib for previously treated patients with advanced non-small-cell lung cancer. *J Clin Oncol.* 2003;21:2237–2246.
14. Thatcher N, Chang A, Parikh P, et al. Gefitinib plus best supportive care in previously treated patients with refractory advanced non-small-cell lung cancer: results from a randomised, placebo-controlled, multicentre study (Iressa survival evaluation in lung cancer). *Lancet.* 2005;366:1527–1537.
15. Sunaga N, Oriuchi N, Kaira K, et al. Usefulness of FDG-PET for early prediction of the response to gefitinib in non-small cell lung cancer. *Lung Cancer.* 2008;59:203–210.
16. Hachemi M, Couturier O, Vervueren L, et al. [^{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.
17. van Gool MH, Aukema TS, Schaake EE, et al. Timing of metabolic response monitoring during erlotinib treatment in non-small cell lung cancer. *J Nucl Med.* 2014;55:1081–1086.

18. Wahl RL, Jacene H, Kasamon Y, Lodge MA. From RECIST to PERCIST: evolving considerations for PET response criteria in solid tumors. *J Nucl Med.* 2009;50(suppl 1):122S–150S.
19. Winther-Larsen A, Fledelius J, Sorensen BS, 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.
20. Mac Manus MP, Hicks RJ, Matthews JP, et al. 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.
21. Fleiss JL, Levin B, Paik MC. *Statistical Methods for Rates and Proportions*. 3rd ed. New York, New York, NY: John Wiley & Sons; 2003.
22. Boellaard R, Delgado-Bolton R, Oyen WJG, et al. FDG PET/CT: EANM procedure guidelines for tumour imaging: version 2.0. *Eur J Nucl Med Mol Imaging.* 2015;42:328–354.
23. Kobe C, Scheffler M, Holstein A, et al. Predictive value of early and late residual ^{18}F-fluorodeoxyglucose and ^{18}F-fluorothymidine uptake using different SUV measurements in patients with non-small-cell lung cancer treated with erlotinib. *Eur J Nucl Med Mol Imaging.* 2012;39:1117–1127.
24. Moon SH, Cho S-H, Park LC, et al. Metabolic response evaluated by ^{18}F-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.
25. Kahraman D, Holstein A, Scheffler M, et al. Tumor lesion glycolysis and tumor lesion proliferation for response prediction and prognostic differentiation in patients with advanced non-small cell lung cancer treated with erlotinib. *Clin Nucl Med.* 2012;37:1058–1064.
26. Nakamoto Y, Zasadny KR, Minn H, Wahl RL. Reproducibility of common semiquantitative parameters for evaluating lung cancer glucose metabolism with positron emission tomography using 2-deoxy-2-[^{18}F]fluoro-D-glucose. *Mol Imaging Biol.* 2002;4:171–178.
27. Weber WA, Gatsonis CA, Mozley PD, et al. Repeatability of ^{18}F-FDG PET/CT in advanced non-small cell lung cancer: prospective assessment in 2 multicenter trials. *J Nucl Med.* 2015;56:1137–1143.
28. Kramer GM, Frings V, Hoetjes N, et al. Repeatability of quantitative whole body ^{18}F-FDG PET/CT uptake measures as function of uptake interval and lesion selection in non-small cell lung cancer patients. *J Nucl Med.* 2016;57:1343–1349.
29. Fledelius J, Khalil A, Hjorthaug K, 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 chemoradiotherapy. *EJNMMI Res.* 2016;6:71.