

# Comparison of RECIST, EORTC criteria and PERCIST for evaluation of early response to chemotherapy in patients with non-small-cell lung cancer

Jingjie Shang<sup>1</sup> · Xueying Ling<sup>1</sup> · Linyue Zhang<sup>1</sup> · Yongjin Tang<sup>1</sup> · Zeyu Xiao<sup>1</sup> · Yong Cheng<sup>1</sup> · Bin Guo<sup>1</sup> · Jian Gong<sup>1</sup> · Li Huang<sup>1</sup> · Hao Xu<sup>1</sup>

Received: 16 December 2015 / Accepted: 10 May 2016 / Published online: 28 May 2016  
© Springer-Verlag Berlin Heidelberg 2016

## Abstract

**Purpose** To compare the Response Evaluation Criteria in Solid Tumors (RECIST) 1.1, the European Organization for Research and Treatment of Cancer (EORTC) criteria and the Positron Emission Tomography Response Criteria in Solid Tumors (PERCIST) 1.0 using PET volume computer-assisted reading (PET VCAR) for response evaluation in patients with advanced non-small-cell lung cancer (NSCLC) treated with chemotherapy.

**Methods** A total of 35 patients with NSCLC were included in this prospective study. All patients received standard chemotherapy and underwent <sup>18</sup>F-FDG PET/CT scans before and after treatment. With the assistance of PET VCAR, the chemotherapeutic responses were evaluated according to the RECIST 1.1, EORTC criteria and PERCIST 1.0. Concordance among these protocols was assessed using Cohen's  $\kappa$  coefficient and Wilcoxon's signed-ranks test. Progression-free survival (PFS) was calculated using the Kaplan-Meier test.

**Results** RECIST 1.1 and EORTC response classifications were discordant in 20 patients (57.1 %;  $\kappa=0.194$ ,  $P<0.05$ ), and RECIST 1.1 and PERCIST 1.0 classifications were discordant in 22 patients (62.9 %;  $\kappa=0.139$ ,  $P<0.05$ ). EORTC and PERCIST 1.0 classifications were discordant in only 4 patients (11.4 %), resulting in better concordance ( $\kappa=0.804$ ,  $P>0.05$ ). Patients with a partial remission according to

RECIST 1.1 had significantly longer PFS ( $P<0.001$ ) than patients with progressive disease, but not significantly longer than patients with stable disease ( $P=0.855$ ). According to both the EORTC criteria and PERCIST 1.0, patients with a partial metabolic response had a significantly longer PFS than those with stable metabolic disease and those with progressive metabolic disease ( $P=0.020$  and  $P<0.001$ , respectively, for EORTC; both  $P<0.001$  for PERCIST 1.0).

**Conclusion** EORTC criteria and PERCIST 1.0 are more sensitive and accurate than RECIST 1.1 for the detection of an early therapeutic response to chemotherapy in patients with NSCLC. Although EORTC criteria and PERCIST 1.0 showed similar results, PERCIST 1.0 is preferred because detailed and unambiguous definitions are given. We also found that response evaluations with PERCIST 1.0 using a single lesion and multiple lesions gave similar response classifications.

**Keywords** RECIST · EORTC · PERCIST · Non-small cell lung cancer · Response evaluation

## Introduction

Non-small-cell lung cancer (NSCLC) is one of the most common cancers in the world. Chemotherapy is still the most widely used treatment for patients with advanced NSCLC [1]. Precise and early prediction of tumour response is of great value because it can help prevent unnecessary toxicity and the additional cost of administering ineffective treatment. Currently, the widely used approaches to monitoring therapeutic responses are based on anatomical changes identified on CT imaging or any other anatomical imaging modality. The first study to evaluate tumour responses to therapy was performed by Moertel and Hanley in 1976, and was followed by the World Health Organization (WHO) criteria in 1979,

Jingjie Shang and Xueying Ling contributed equally to the preparation of the manuscript.

✉ Hao Xu  
txh@jnu.edu.cn

<sup>1</sup> Department of Nuclear Medicine and PET/CT-MRI Centre, The First Affiliated Hospital of Jinan University, 510630 Guangzhou, China

Response Evaluation Criteria in Solid Tumors (RECIST) in 2000, and RECIST 1.1 in 2009 [2–5]. Although these response evaluation protocols have been updated and improved over time, they are still limited by their dependence on morphological changes to assess therapeutic response [6–9]. The assumption that changes in tumour dimensions are a true marker of therapeutic efficacy is prone to error because tumour tissue consists of variable substances and the regression of all these components may appear slowly and incompletely. This is especially true in NSCLC because fibrotic masses may remain even after cure.

Quantitative  $^{18}\text{F}$ -FDG PET is thought to overcome these limitations and is believed to be a more suitable assessment tool for evaluating therapeutic response [6–9]. Fundamentally, quantitative  $^{18}\text{F}$ -FDG PET treatment response assessment is based on the alteration in the standardized uptake value (SUV) between baseline and follow-up studies. However, the SUV is affected by various factors, such as technical, physical, and biological factors [10–12]. To facilitate reproducibility so that results across trials can be compared, a widely accepted standardized protocol is required. Currently, the following two protocols are used to quantify anticancer treatment responses in terms of metabolic changes: the European Organization for Research and Treatment of Cancer (EORTC) criteria, which were developed in 1999 [13], and the Positron Emission Tomography Response Criteria in Solid Tumors (PERCIST) 1.0, which were developed by American researchers in 2009 [9]. EORTC recommends using SUV normalized to body surface area (SUV<sub>bsa</sub>) to reduce the influence of body weight on SUV [13], but it does not propose a rigorous standardization protocol to ensure that the criteria are uniformly applied. PERCIST 1.0 recommends using SUV corrected for lean body mass (SUL) to avoid falsely high organ SUV in obese patients [14, 15]. It also recommends computing the SUL<sub>peak</sub> and total lesion glycolysis (TLG) values as response metrics. Furthermore, PERCIST 1.0 proposes a series of detailed and unambiguous regulations to standardize procedures to ensure reproducibility.

Many aspects of PERCIST 1.0 have been improved compared to the EORTC criteria, but PERCIST 1.0 involves the performance of some cumbersome tasks. For example, accurately measuring SUL<sub>peak</sub>, accurately registering the images from a series of examinations in the same patient, choosing the optimal reference region in the liver, comprehensively monitoring all the factors affecting SUV quantitation between two studies. Ideally a single application would be used to perform these functions automatically. PET volume computer-assisted reading (PET VCAR), an application of the Advantage Workstation (GE Healthcare), is one such program that can be used by the clinician to assist in monitoring treatment response [16].

Several studies have compared response evaluations using anatomical and metabolic criteria in patients with NSCLC

receiving chemotherapy [17, 18]. Ordu et al. [17] used SUV<sub>max</sub> instead of SUL<sub>peak</sub> when applying PERCIST 1.0. Ding et al. [18] manually calculated SUL<sub>peak</sub>, which was defined as the largest possible mean value of a spherical region of interest (ROI) of 1 cm<sup>3</sup> within a tumour. It is difficult manually to calculate SUL<sub>peak</sub> accurately. Furthermore, none of these reports described how the acquisition parameters of the two PET/CT examinations affecting SUV variability were monitored, how the images from serial examinations were registered, or how the reference region in the liver was chosen. The original intention of formulating PERCIST 1.0 was to create a standardized mechanism to enhance the reproducibility needed for the comparison of response rates between trials. If the standardized protocol is not strictly followed in practice, the most important component of PERCIST 1.0 will be lost. The aim of this study was to assess the concordance among RECIST 1.1, EORTC criteria and PERCIST 1.0 using PET VCAR for the early evaluation of response to chemotherapy in patients with NSCLC.

## Materials and methods

### Patients

From October 2012 to November 2014, patients with histopathologically proven advanced NSCLC who were scheduled to undergo chemotherapy treatment were prospectively enrolled in this study. Patients were excluded if they were diabetic, had brain metastasis, had undergone surgery, or had received previous chemotherapy or radiotherapy. All patients were given the following chemotherapy regimen: gemcitabine (1,000 mg/m<sup>2</sup>) and cisplatin (75 – 80 mg/m<sup>2</sup>; GP protocol). The baseline  $^{18}\text{F}$ -FDG PET/CT scan was performed during the week before the start of treatment, followed by an early  $^{18}\text{F}$ -FDG PET/CT scan within 10 to 14 days after two cycles of chemotherapy. Additional CT scans of the chest and other regions were performed as clinically indicated every 3 months or in patients with clinical suspicion of disease progression. Treatment was continued until progression was identified by CT according to RECIST 1.1, unacceptable toxicity, or patient withdrawal. Progression-free survival (PFS) was calculated as the time from the first chemotherapy administration to disease progression or death from any cause. The study was approved by the institutional review board of the First Affiliated Hospital of Jinan University, and was in compliance with national legislation and the Declaration of Helsinki guidelines. Written informed consent was obtained from all patients.

### PET/CT examinations

All  $^{18}\text{F}$ -FDG PET/CT examinations were performed with the same protocol using a GE Discovery PET/CT 690 system, and

the scanning procedure closely resembled the National Cancer Institute guidelines [19]. Patients were instructed to fast for at least 6 h prior to the  $^{18}\text{F}$ -FDG PET/CT scan. The blood glucose level was measured before the tracer was injected. Whole-body PET/CT images, generally from the top of the skull to mid-thigh, were acquired 50 to 70 min after intravenous injection of  $^{18}\text{F}$ -FDG at a dose of 0.08 – 0.10 mCi/kg body weight. Patients were asked to void their bladder immediately before scanning to minimize the presence of the tracer in the urinary tract.

CT data were acquired in breath-hold with 120 kV, 80 – 160 mA modulated using the GE AutomA technique with a noise index of 30, slice thickness of 3.75 mm, slice interval of 3.27 mm, pitch of 1.375, matrix size of  $512 \times 512$  and scan FOV of 50 cm. PET data were acquired in 3D time-of-flight (TOF) mode with a 2-min scan per bed position, slice thickness of 3.27 mm, slice interval of 3.75 mm, matrix size of  $192 \times 192$  and scan FOV of 70 cm. The PET data were attenuation-corrected (AC) by the integrated CTAC technology. The CT data were reconstructed in standard mode, window width/window level 400/40, advanced statistical iterative reconstruction 40 %. The PET data were then reconstructed in terms of the point spread function (PSF) together with TOF technology.

### Evaluation of the therapeutic response by PET VCAR

We selected baseline and follow-up examinations from the same patient and launched the PET VCAR workflow. The images from the examinations were automatically coregistered based on the characteristics of the bone and soft tissue on the CT scan. The SUV-related parameters were automatically reviewed and compared using PET VCAR. PET

VCAR would highlight relevant parameters in the following scenarios: if the difference in injected doses of  $^{18}\text{F}$ -FDG was over 20 %, if the difference in uptake time was more than 15 min, if either the scan parameters or the reconstruction parameters varied between the baseline and follow-up scans, and if the blood glucose level was more than 200 mg/dL during one scan. If one of the SUV-related parameters was highlighted, the patient was excluded. All data were analysed separately by one experienced radiologist and one experienced nuclear physician to generate response assessments according to the RECIST 1.1, EORTC criteria and PERCIST 1.0. For RECIST 1.1 criteria, the size of the target lesion was manually measured by the reader, and PET VCAR automatically calculated the sum percentage variation in the longest diameters of the target lesion. For EORTC and PERCIST 1.0, the ROIs in the lesions on the baseline examination were automatically bookmarked and propagated onto the coregistered follow-up examination images. The propagated bookmarks were accepted or rejected using the confidence scale, and ROIs were manually adjusted by the reader as needed. The EORTC results are based on the percentage variation in the SUVmax of the target lesions; the PERCIST 1.0 results are classified as progressive metabolic disease (PMD), stable metabolic disease (SMD), partial metabolic response (PMR) and complete metabolic response (CMR).

### Response evaluation with RECIST, EORTC criteria and PERCIST

The details concerning the measurable baseline lesions and response assessment criteria for RECIST 1.1, EORTC and PERCIST 1.0 are shown in Table 1. For

**Table 1** Response evaluation criteria for RECIST, EORTC and PERCIST

Characteristic	RECIST 1.1	EORTC	PERCIST 1.0
Measurability of lesion at baseline	Lesions: longest diameter $\geq 10$ mm; lymph nodes: short axis $\geq 15$ mm	Lesions with high $^{18}\text{F}$ -FDG uptake	SULpeak of baseline lesions at least 1.5-fold greater than liver SULmean + $2 \times$ SD. If liver is abnormal, primary tumour should have uptake $> 2.0 \times$ SULmean of blood pool
Objective response	CR: disappearance of all target lesions	CMR: complete resolution of $^{18}\text{F}$ -FDG uptake within all lesions, making them indistinguishable from the surrounding tissue	CMR: complete resolution of $^{18}\text{F}$ -FDG uptake within all lesions to a level of less than or equal to that of the mean liver activity and indistinguishable from the background blood-pool levels
	PR: reduction of at least 30 % in the sum of diameters of target lesions	PMR: reduction of at least 25 % in the sum of SUV	PMR: reduction of at least 30 % in SULpeak and an absolute drop of 0.8 SULpeak units
	PD: increase of at least 20 % in the sum of diameters of target lesions or appearance of new lesions	PMD: increase of at least 25 % in the sum of SUV or appearance of new $^{18}\text{F}$ -FDG-avid lesions that are typical of cancer and not related to inflammation or infection	PMD: increase of at least 30 % in SULpeak and an absolute increase of 0.8 SULpeak units OR: 75 % increase in TLG, with no decrease in SUL, or appearance of new $^{18}\text{F}$ -FDG-avid lesions typical of cancer and not related to inflammation or infection
	SD: not CR, PR, or PD	SMD: not CMR, PMR, or PMD	SMD: not CMR, PMR, or PMD

RECIST 1.1, we chose up to five and no more than two lesions per organ as target lesions. For EORTC, we chose up to five lesions with the highest  $^{18}\text{F}$ -FDG uptake in as many involved organ systems as possible as target lesions. We chose the body surface area to normalize the SUV and chose SUVmax for the response calculation. For PERCIST 1.0, we chose the hottest lesion as the target lesion on the baseline and subsequent follow-up scans. The hottest lesion on the follow-up scan could be a lesion different from the previously measured lesion, on the assumption that it had been present since baseline. The SULpeak was used for the response calculation. The optimized reference region was automatically delineated in the right lobe of the liver. In patients with liver involvement, we chose the descending thoracic aorta as the background area. TLG was automatically computed by PET VCAR using an iterative adaptive segmentation algorithm to find a threshold value that separated the target volume from the background tissue by weighting the SULmax and SULmean within the target volume with a default weighting factor of 0.5 [20].

In addition, we also chose up to five of the hottest lesions (no more than two per organ) as target lesions using PERCIST 1.0. Each of the baseline target lesions had to meet the PERCIST 1.0 definition of measurable lesions. The sum of the variation in SUL from multiple lesions was measured and the response assessment was classified according to PERCIST 1.0. The response classifications based on evaluation of multiple lesions were compared with those from evaluation of a single lesion to determine whether there were differences in response assessment.

### Statistical analysis

The paired *t* test was used to evaluate the differences in the same parameters between before and after chemotherapy. The response classifications of RECIST 1.1, EORTC and PERCIST 1.0 were graded on a four-point ordinal scale as follows: progressive disease (PD) or PMD =1, stable disease (SD) or SMD =2, partial remission (PR) or PMR =3, and complete response (CR) or CMR =4. Concordance was assessed using Cohen's  $\kappa$  coefficient. Agreement between the two assessments was categorized as poor (weighted  $\kappa < 0$ ), slight (weighted  $\kappa = 0 - 0.20$ ), fair (weighted  $\kappa = 0.21 - 0.40$ ), moderate (weighted  $\kappa = 0.41 - 0.60$ ), substantial (weighted  $\kappa = 0.61 - 0.80$ ), and almost perfect (weighted  $\kappa > 0.80$ ) [21]. The differences among the results of the three protocols were assessed using Wilcoxon's signed-ranks test. PFS was calculated using the Kaplan-Meier test and statistically evaluated using the log-rank test. *P* values less than 0.05 were considered statistically significant.

## Results

### Patient characteristics

A total of 46 consecutive eligible patients with advanced NSCLC were enrolled in this prospective study, of whom 11 were excluded for the following reasons: technical issues (three patients), patient's own wish (four patients), and early discontinuation of chemotherapy due to a significant decline in performance status (four patients). A total of 35 patients (26 men and 9 women; mean age 61 years, range, 48 – 84 years) were therefore included in this study, including 22 with adenocarcinoma and 13 with squamous cell carcinoma. Fourteen patients had stage III disease, and 21 had stage IV disease. The mean time between completion of the two cycles chemotherapy and the follow-up PET/CT scan was 12 days (range 10 – 14 days). Patient characteristics are shown in Tables 2 and 3. There were no significant differences in blood glucose levels, injected doses of  $^{18}\text{F}$ -FDG, and uptake times between before and after chemotherapy ( $P > 0.05$ ).

### Treatment response assessments with RECIST, EORTC criteria and PERCIST

The RECIST 1.1, EORTC criteria and PERCIST 1.0 results are shown in Table 4. With RECIST 1.1 the reduction in tumour diameter was  $28.8 \pm 1.6$  % mm (mean  $\pm$  SD), with 8

**Table 2** Clinical characteristic of the 35 evaluated patients

Parameter	Value
Sex	
Male	26
Female	9
Histology	
Adenocarcinoma	22
Squamous cell carcinoma	13
Location of primary tumour	
Central type	25
Peripheral type	10
Location of metastases	
Lung	13
Lymph nodes	35
Liver	7
Adrenal gland	11
Bone	15
Stage	
III	14
IV	21

**Table 3** Parameters before and after chemotherapy for NSCLC cancer

Parameter	Before chemotherapy	After two cycles of chemotherapy
Height (cm)		
Mean ± SD	167.6 ± 5.6	167.6 ± 5.6
Range	159 – 175	159 – 175
Weight (kg)		
Mean ± SD	63.3 ± 9.6	62.6 ± 10.1
Range	40 – 75	40 – 72
Blood glucose level (mg/dL)		
Mean ± SD	102.6 ± 15.6	106.2 ± 13.6
Range	86.4 – 124.2	88.2 – 126.0
Dose of <sup>18</sup> F-FDG (mCi)		
Mean ± SD	6.7 ± 0.7	6.5 ± 0.7
Range	5.4 – 8.0	5.1 – 7.5
Uptake time (min)		
Mean ± SD	60.5 ± 4.3	60.4 ± 4.5
Range	55 – 68	52 – 70

patients showing PD, 20 SD, 7 PR and 0 CR, giving a response rate (PR + CR) of 20.0 %. With EORTC criteria the reduction in SUVmax was 40.1 % ± 2.5 %, with 8 patients showing PMD, 8 SMD, 19 PMR and 0 CMR, giving a response rate (PMR + CMR) of 54.3 %. With PERCIST 1.0 the reduction in SULpeak was 36.1 % ± 2.5 %, with 8 patients showing PMD, 6 SMD, 21 PMR and 0 CMR, giving a response rate (PMR + CMR) of 60.0 %.

**Table 4** Therapy response assessments with RECIST, EORTC criteria and PERCIST

Criteria	Before chemotherapy	After two cycles of chemotherapy
<b>RECIST</b>		
Longest diameter (mm)	39.9 ± 2.2	33.5 ± 2.3
Tumour diameter reduction (%)	–	28.8 ± 1.6
Objective response (PD/SD/PR/CR) (n)	–	8/20/7/0
<b>EORTC criteria</b>		
Tumour SUVmax	18.7 ± 2.0	10.5 ± 1.9
Tumour SUVmax reduction (%)	–	40.1 ± 2.5
Objective response (PMD/SMD/PMR/CMR) (n)	–	8/8/19/0
<b>PERCIST</b>		
Tumour SULpeak	9.4 ± 1.2	6.1 ± 3.2
Background	2.01 ± 0.19	2.11 ± 0.17
Tumour SULpeak reduction (%)	–	36.1 ± 2.5
Objective response (PMD/SMD/PMR/CMR) (n)	–	8/6/21/0

**Comparison of treatment response assessments between RECIST and EORTC criteria**

RECIST 1.1 and EORTC classifications were discordant in 20 patients (57.1 %), with slight agreement in the response classification between the two assessments ( $\kappa = 0.194$ ,  $P < 0.05$ ; Table 5). Of eight patients classified as PD according to RECIST 1.1, four were classified according to EORTC as SMD, as the increase in the sum of the longest diameters of the target lesions was over 20 %, while the increase in the sum of SUVmax was less than 25 %. Of 20 patients classified as SD according to RECIST 1.1, 4 were classified as PMD according to EORTC, as new lesions were detected on PET, but not on CT. Twelve patients were classified as PMR, as the decrease in the sum of the diameters of the target lesions was less than 30 %, while the decrease in the sum of SUVmax was more than 25 %.

**Comparison of treatment response assessments between RECIST and PERCIST**

RECIST 1.1 and PERCIST 1.0 classifications were discordant in 22 patients (62.9 %), with slight agreement in the response classification between the two assessments ( $\kappa = 0.139$ ,  $P < 0.05$ ; Table 6). Of eight patients classified as PD according to RECIST 1.1, four were classified as SMD according to PERCIST 1.0, as the increase in the sum of the longest diameters of the target lesions was more than 20 %, while the increase in the SULpeak of the target lesions was less than 30 %. Of 20 patients classified as SD according to RECIST 1.1, four were classified as PMD according to PERCIST 1.0, as new lesions were detected on PET, but not on CT. Fourteen patients were classified as PMR, as the decrease in the sum of the longest diameters of the target lesions was less than 30 %, while the decrease in the SULpeak of the target lesions was more than 30 %.

**Table 5** Comparison of treatment response assessments according to RECIST and EORTC

RECIST	EORTC				
	PMD	SMD	PMR	CMR	Total
PD	4	4	0	0	8
SD	4	4	12	0	20
PR	0	0	7	0	7
CR	0	0	0	0	0
Total	8	8	19	0	35

$\kappa = 0.194$ , indicating minimal agreement between the two assessments  $p = 0.042$ , Wilcoxon’s signed-ranks test, indicating a significant difference between the two assessments

**Table 6** Comparison of treatment response assessments between RECIST and PERCIST

RECIST	PERCIST				Total
	PMD	SMD	PMR	CMR	
PD	4	4	0	0	8
SD	4	2	14	0	20
PR	0	0	7	0	7
CR	0	0	0	0	0
Total	8	6	21	0	35

$\kappa=0.139$ , indicating minimal agreement between the two assessments  
 $p=0.018$ , Wilcoxon's signed-ranks test, indicating a significant difference between the two assessments

### Comparison of treatment response assessments between EORTC criteria and PERCIST

EORTC and PERCIST 1.0 classifications were discordant in four patients (11.4 %), with an almost perfect agreement in response classification between the two assessments ( $\kappa=0.804, P>0.05$ ; Table 7). Of eight patients classified as PMD according to EORTC, one was classified as SMD according to PERCIST 1.0, as the increase in the sum of SUVmax of the multiple lesions was more than 25 %, while the increase in the SULpeak of the single lesion was less than 30 %. Of eight patients classified as SMD according to EORTC, one was classified as PMD according to PERCIST 1.0, as the increase in the sum of the SUVmax of the multiple lesions was less than 25 %, while the increase in the TLG was more than 75 %. Two patients were classified as PMR, as the decrease in the sum the SUVmax of the multiple lesions was less than 25 %, while the decrease in the SULpeak of the single lesion was more than 30 %.

**Table 7** Comparison of treatment response assessments between EORTC and PERCIST

PERCIST	EORTC				Total
	PMD	SMD	PMR	CMR	
PMD	7	1	0	0	8
SMD	1	5	0	0	6
PMR	0	2	19	0	21
CMR	0	0	0	0	0
Total	8	8	19	0	35

$\kappa=0.804$ , indicating perfect agreement between the two assessments  
 $p=0.722$ , Wilcoxon's signed-ranks test, indicating no significant differences between the two assessments

### Progression-free survival analysis

The median overall PFS was 7 months (range 3–21 months). With RECIST 1.1, the median PFS was 10 months in patients with PR, 9 months in those with SD, and 3 months in those with PD. The median PFS in patients with PR was significantly longer than in those with PD ( $P<0.001$ ), but was not significantly longer than in those with SD ( $P=0.855$ ; Fig. 1a). With EORTC criteria, the median PFS was 11 months in patients with PMR, 4 months in those with SMD, and 3 months in those with PMD. The median PFS in patients with PMR was significantly longer than in those with the SMD and in those with PMD ( $P=0.020$  and  $P<0.001$ , respectively; Fig. 1b). With PERCIST 1.0, the median PFS was 11 months in patients with PMR, 3 months in those with SMD, and 3 months in those with PMD. The median PFS in the patients with PMR was significantly longer than in those with SMD and in those with PMD ( $P<0.001$  for both; Fig. 1c).

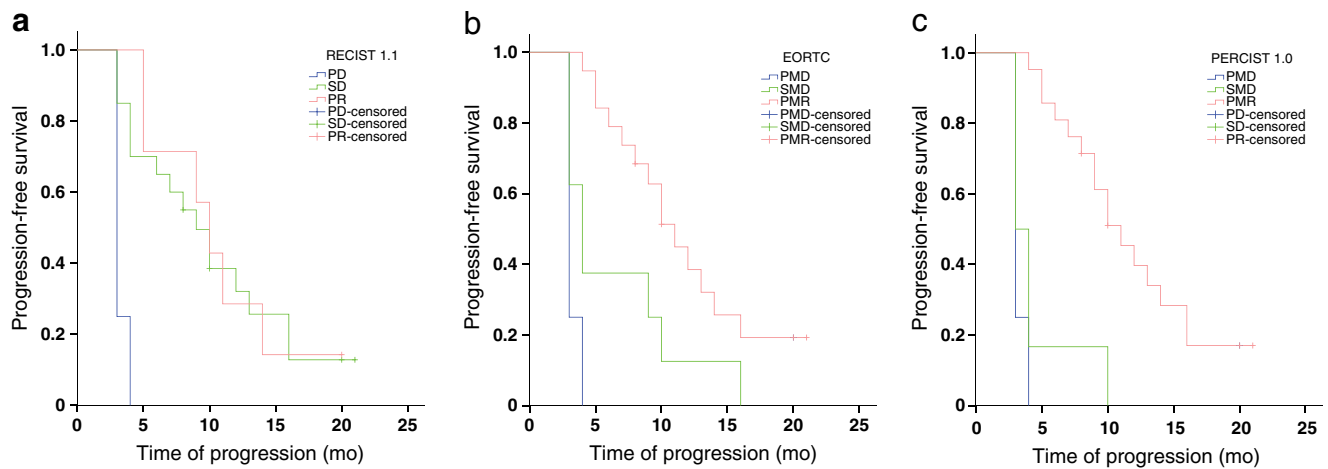
### Comparison of PERCIST treatment response assessments between multiple lesions and a single lesion

The therapeutic responses according to PERCIST 1.0 with multiple lesions were as follows: 8 patients had PMD, 5 had SMD, 22 had PMR, and 0 had CMR. PERCIST 1.0 classifications with multiple lesions and with a single lesion were discordant in three patients (8.6 %), with an almost perfect agreement in response classification between the two assessments ( $\kappa=0.803, P>0.05$ ). One patient was reclassified from PMR to SMD, as the reduction in summed SULpeak of multiple lesions was less than the reduction in the single lesion SULpeak. Two patients were reclassified from SMD to PMR, as the reduction in summed SULpeak of multiple lesions was greater than the reduction in the single lesion SULpeak.

### Discussion

In this study, we used an automatic application to evaluate the early therapeutic response to chemotherapy in patients with advanced NSCLC determined using RECIST 1.1, EORTC criteria and PERCIST 1.0. We found that EORTC criteria and PERCIST 1.0 were more sensitive and accurate than RECIST 1.1 for the detection of an early therapeutic response. Furthermore, EORTC criteria and PERCIST 1.0 showed better concordance for the detection of an early therapeutic response.

Comparison of RECIST and PERCIST 1.0 in patients with NSCLC has already been reported. In a study by Ding et al. [18], 44 patients with NSCLC who received cisplatin-based combination chemotherapy were evaluated using RECIST and PERCIST 1.0 after two or four to six cycles of treatment. The differences between RECIST and PERCIST 1.0 evaluations were significant if new lesions were not identified on



**Fig. 1** Progression-free survival in relation to response assessments according to RECIST 1.1 (a), EORTC criteria (b), and PERCIST 1.0 (c). *PR* partial remission, *PMR* partial metabolic response, *SD* stable

disease, *SMD* stable metabolic disease, *PD* progressive disease, *PMD* progressive metabolic disease

PET images when using RECIST. There have been no published studies comparing response evaluations between RECIST and EORTC criteria in patients with NSCLC treated with chemotherapy. In other tumours, Monteil et al. [22] found a discrepancy between RECIST and EORTC evaluations in patients with colorectal cancer receiving chemotherapy. In this study, we found that the therapeutic response rate was 20.0 %, 54.3 % and 60.0 % using RECIST 1.1, EORTC criteria and PERCIST 1.0, respectively. RECIST 1.1 and EORTC response classifications were discordant in 20 patients (57.1 %), and RECIST 1.1 and PERCIST 1.0 classifications were discordant in 22 patients (62.9 %), with slight agreements in response classification between the two assessments. The results of this study are similar to those of previous studies [18, 22]. Of the 20 patients classified as having SD according to RECIST 1.1, 12 were reclassified as having PMR according to EORTC criteria and 14 were reclassified as having PMR according to PERCIST 1.0 because metabolic changes exceeded the threshold, but morphological changes did not.

Through the follow-up, we found that the median PFS in patients with PR according to RECIST 1.1 was significantly longer than in patients with PD, but not in patients with SD. On the other hand, the median PFS in patients with PMR according to EORTC criteria and PERCIST 1.0 was significantly longer than that in patients with SMD and PMD. The patients who shifted from SD with RECIST 1.1 to PMR when evaluated with EORTC criteria or PERCIST 1.0 had a longer PFS. One possible reason for this is that metabolic changes occur earlier than anatomical changes after chemotherapy. Similar results have been shown in several studies [23–26]. These results indicate that EORTC criteria and PERCIST are more sensitive and accurate than RECIST for the detection of an early therapeutic response. This advantage can potentially allow earlier therapy modification based on the  $^{18}\text{F}$ -FDG PET response.

Currently, no studies have been published comparing response evaluations between EORTC criteria and PERCIST 1.0 in patients with NSCLC receiving chemotherapy. In this study, we found that EORTC criteria and PERCIST 1.0 had excellent concordance in response classification. Similar results have been obtained in previous studies. Ziai et al. [27] found perfect concordance between EORTC and PERCIST 1.0 evaluations in patients with small-cell lung cancer who were receiving chemotherapy and radiotherapy. Skougaard et al. [28] found that EORTC criteria and PERCIST 1.0 gave similar evaluation responses in patients with metastatic colorectal cancer treated with a combination of irinotecan and cetuximab. A reasonable explanation for the excellent concordance is that all of these studies focused on the most metabolically active part of the patients' tumours, which is regarded as the most aggressive disease fragment, determining the disease development according to cancer stem cell theory [29, 30].

On the other hand, EORTC and PERCIST 1.0 response classifications were discordant in only four patients (11.4 %). Similar results were obtained by Skougaard et al. [28], who found that EORTC and PERCIST 1.0 evaluations were discordant in 13.1 % of patients (8/61). Although EORTC and PERCIST 1.0 evaluations are rarely discordant, they differ in several other aspects. With EORTC criteria, SUV<sub>bsa</sub> is recommended for assessing tumour  $^{18}\text{F}$ -FDG uptake. Definitions of the conditions under which SUV<sub>max</sub> or SUV<sub>mean</sub> are used for response calculation were not given. Details of the selected standard for target lesions were also not given. Thus, applying EORTC criteria may have generated different outcomes by several different mechanisms. While with PERCIST 1.0, SUL<sub>peak</sub> is used to assess tumour  $^{18}\text{F}$ -FDG uptake. The number and selected standard of target lesions are

explicitly defined. TLG is recommended as another measure of progression. An absolute and rigorous standardization of the PET protocol is given to guarantee reproducible SUL. These definitions mean that PERCIST 1.0 is more consistently applied among international observers. This is crucial, because if no consistent criteria are available, comparisons between studies will be inaccurate and advances in cancer treatment will be hampered [7, 9]. Thus, although EORTC criteria and PERCIST 1.0 had excellent concordance in response classifications, PERCIST 1.0 showed the potential for better consensus and comparability. The computer-assisted reading application that was fully customized for use with PERCIST 1.0 makes the procedure of assessment simpler and helps ensure more accurate results.

Regarding the number of target lesions, PERCIST 1.0 recommends choosing the hottest single lesion as the target lesion. It also recommends that multiple lesions (no more than two per organ and up to five of the hottest lesions) be measured as a secondary analysis. In this study, although the results of response evaluations with multiple lesions and a single lesion were similar, they were still discordant in three patients. Further study with a larger sample size is needed to determine the optimal number of target lesions for response assessment.

The current study had some limitations. First, we have not yet evaluated the corresponding pathological information with these imaging findings. Further study is needed to investigate the correlation of these findings with pathological examinations. Second, the precision of SUL was not determined in this study for ethical reasons. Animal studies are needed to quantify the precision of SUL in future research.

## Conclusion

This study demonstrated that EORTC criteria and PERCIST 1.0 are more sensitive and accurate than RECIST 1.1 for the evaluation of the early therapeutic response to chemotherapy in patients with NSCLC. Compared with the EORTC protocol, less ambiguous definitions of standardization help simplify response evaluations with PERCIST 1.0.

**Acknowledgments** We would like to thank the staff members of the Department of Nuclear Medicine and PET/CT-MR Centre, First Affiliated Hospital of Jinan University, for their excellent technical support.

## Compliance with ethical standards

**Conflicts of interest** None.

**Ethical approval** All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the principles of the 1964 Declaration of Helsinki and its later amendments or

comparable ethical standards. This article does not describe any studies with animals performed by any of the authors.

**Informed consent** Informed consent was obtained from all individual participants included in the study.

## References

- Rossi A, Di Maio M. Platinum-based chemotherapy in advanced non-small-cell lung cancer: optimal number of treatment cycles. *Expert Rev Anticancer Ther*. 2016. doi:10.1586/14737140.2016.1170596
- Moertel CG, Hanley JA. The effect of measuring error on the results of therapeutic trials in advanced cancer. *Cancer*. 1976;38:388–94.
- Miller AB, Hoogstraten B, Staquet M, Winkler A. Reporting results of cancer treatment. *Cancer*. 1981;47:207–14.
- Therasse P, Arbuck SG, Eisenhauer EA, Wanders J, Kaplan RS, Rubinstein L, et al. New guidelines to evaluate the response to treatment in solid tumors. European Organization for Research and Treatment of Cancer, National Cancer Institute of the United States, National Cancer Institute of Canada. *J Natl Cancer Inst*. 2000;92:205–16.
- Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer*. 2009;45:228–47.
- Hicks RJ. Role of 18F-FDG PET in assessment of response in non-small cell lung cancer. *J Nucl Med*. 2009;50 Suppl 1:31S–42.
- Weber WA. Assessing tumor response to therapy. *J Nucl Med*. 2009;50 Suppl 1:1S–10S.
- Juweid ME, Cheson BD. Positron-emission tomography and assessment of cancer therapy. *N Engl J Med*. 2006;354:496–507.
- 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–50.
- Adams MC, Turkington TG, Wilson JM, Wong TZ. A systematic review of the factors affecting accuracy of SUV measurements. *AJR Am J Roentgenol*. 2010;195:310–20.
- Boellaard R. Standards for PET image acquisition and quantitative data analysis. *J Nucl Med*. 2009;50 Suppl 1:11S–20.
- Boellaard R. Need for standardization of 18F-FDG PET/CT for treatment response assessments. *J Nucl Med*. 2011;52 Suppl 2:93S–100.
- Young H, Baum R, Cremerius U, Herholz K, Hoekstra O, Lammertsma AA, et al. Measurement of clinical and subclinical tumour response using [18F]-fluorodeoxyglucose and positron emission tomography: review and 1999 EORTC recommendations. European Organization for Research and Treatment of Cancer (EORTC) PET Study Group. *Eur J Cancer*. 1999;35:1773–82.
- Zasadny KR, Wahl RL. Standardized uptake values of normal tissues at PET with 2-[fluorine-18]-fluoro-2-deoxy-D-glucose: variations with body weight and a method for correction. *Radiology*. 1993;189:847–50.
- Sugawara Y, Zasadny KR, Neuhoff AW, Wahl RL. Reevaluation of the standardized uptake value for FDG: variations with body weight and methods for correction. *Radiology*. 1999;213:521–5.
- Fox JJ, Autran-Blanc E, Morris MJ, Gavane S, Nehmeh S, Van Nuffel A, et al. Practical approach for comparative analysis of multilesion molecular imaging using a semiautomated program for PET/CT. *J Nucl Med*. 2011;52:1727–32.



17. Ordu C, Selcuk NA, Akosman C, Eren OO, Altunok EC, Toklu T, et al. Comparison of metabolic and anatomic response to chemotherapy based on PERCIST and RECIST in patients with advanced stage non-small cell lung cancer. *Asian Pac J Cancer Prev*. 2015;16:321–6.
18. Ding Q, Cheng X, Yang L, Zhang Q, Chen J, Li T, et al. PET/CT evaluation of response to chemotherapy in non-small cell lung cancer: PET response criteria in solid tumors (PERCIST) versus response evaluation criteria in solid tumors (RECIST). *J Thorac Dis*. 2014;6:677–83.
19. Shankar LK, Hoffman JM, Bacharach S, Graham MM, Karp J, Lammertsma AA, et al. Consensus recommendations for the use of 18F-FDG PET as an indicator of therapeutic response in patients in National Cancer Institute Trials. *J Nucl Med*. 2006;47:1059–66.
20. Sebastian TB, Manjeshwar RM, Akhurst TJ, Miller JV. Objective PET lesion segmentation using a spherical mean shift algorithm. *Med Image Comput Comput Assist Interv*. 2006;9:782–9.
21. Kundel HL, Polansky M. Measurement of observer agreement. *Radiology*. 2003;228:303–8.
22. Monteil J, Mahmoudi N, Leobon S, Roudaut PY, El Badaoui A, Verbeke S, et al. Chemotherapy response evaluation in metastatic colorectal cancer with FDG PET/CT and CT scans. *Anticancer Res*. 2009;29:2563–8.
23. Lee DH, Kim SK, Lee HY, Lee SY, Park SH, Kim HY, et al. Early prediction of response to first-line therapy using integrated 18F-FDG PET/CT for patients with advanced/metastatic non-small cell lung cancer. *J Thorac Oncol*. 2009;4:816–21.
24. Wahl RL, Zasadny K, Helvie M, Hutchins GD, Weber B, Cody R. Metabolic monitoring of breast cancer chemohormonotherapy using positron emission tomography: initial evaluation. *J Clin Oncol*. 1993;11:2101–11.
25. Novello S, Vavala T, Levra MG, Solitro F, Pelosi E, Veltri A, et al. Early response to chemotherapy in patients with non-small-cell lung cancer assessed by [18F]-fluoro-deoxy-D-glucose positron emission tomography and computed tomography. *Clin Lung Cancer*. 2013;14:230–7.
26. Mac MMP, Hicks RJ, Matthews JP, McKenzie A, Rischin D, Salminen EK, 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;21:1285–92.
27. Ziai D, Wagner T, El Badaoui A, Hitzel A, Woillard JB, Melloni B, et al. Therapy response evaluation with FDG-PET/CT in small cell lung cancer: a prognostic and comparison study of the PERCIST and EORTC criteria. *Cancer Imaging*. 2013;13:73–80.
28. Skougaard K, Nielsen D, Jensen BV, Hendel HW. Comparison of EORTC criteria and PERCIST for PET/CT response evaluation of patients with metastatic colorectal cancer treated with irinotecan and cetuximab. *J Nucl Med*. 2013;54:1026–31.
29. Huang EH, Heidt DG, Li CW, Simeone DM. Cancer stem cells: a new paradigm for understanding tumor progression and therapeutic resistance. *Surgery*. 2007;141:415–9.
30. Al-Hajj M, Becker MW, Wicha M, Weissman I, Clarke MF. Therapeutic implications of cancer stem cells. *Curr Opin Genet Dev*. 2004;14:43–7.