Comparison of Quantitative Methods on FDG PET/CT for Treatment Response Evaluation of Metastatic Colorectal Cancer

Ji-In Bang¹ · Yoojoo Lim² · Jin Chul Paeng¹ · Sae-Won Han² · Sohyun Park¹ ·
Jung Min Lee¹ · Hyun Joo Kim¹ · Gi Jeong Cheon¹ · Dong Soo Lee¹ ·
June-Key Chung¹ · Tae-You Kim² · Keon Wook Kang¹

Received: 1 June 2016 / Revised: 21 August 2016 / Accepted: 30 August 2016 / Published online: 13 September 2016
© Korean Society of Nuclear Medicine 2016

Abstract

Purpose FDG PET is effective in treatment response evaluation of cancer. However, there is no standard method for quantitative evaluation of FDG PET, particularly regarding cytostatic drugs. We compared various FDG PET quantitative methods in terms of response determination.

Methods A total of 39 refractory metastatic colorectal cancer patients who received a multikinase inhibitor treatment were included. Baseline and posttreatment FDG PET/CT scans were performed before and two cycles after treatment. Standardized uptake value (SUV) and total lesion glycolysis (TLG) values using various margin thresholds (30–70 % of maximum SUV with increment 10 %, twice mean SUV of blood pool, SUV 3.0, and SUV 4.0) were measured, with measurement target of the hottest lesion or a maximum of five hottest lesions. Treatment response by the PERCIST criteria was also determined. Predictive values of the PET indexes were evaluated in terms of the treatment response determined by the RECIST 1.1 criteria.

Results The agreement rate was 38 % between response determined by the PERCIST and the RECIST criteria ($\kappa = 0.381$). When patients were classified into disease control group (PR, SD) and non-control group (PD) by the RECIST criteria, percent changes of TLG with various margin thresholds

(particularly, 30–50 % of maximum SUV) exhibited significant differences between the two groups, and high diagnostic power for the response by the RECIST criteria. TLG-based criteria, which used a margin threshold of 50 % of maximum SUV, exhibited a high agreement with the RECIST criteria compared with the PERCIST criteria ($\kappa = 0.606$).

Conclusion In metastatic colorectal cancer, FDG PET/CT could be effective for treatment response evaluation by using TLG measured by margin thresholds of 30–50 % of maximum SUV. Further studies are warranted regarding the optimal cut-off values for this method.

Keywords Colorectal neoplasm · Fluorodeoxyglucose F18 · Positron-emission tomography · Therapy

Introduction

¹⁸F-Fluorodeoxyglucose (FDG) positron emission tomography (PET)/computed tomography (CT) is an effective imaging method in various cancers, and it is frequently used for response evaluation after anticancer therapy. Currently, the Response Evaluation Criteria in Solid Tumors (RECIST) 1.1 are most widely used for treatment response evaluation, in which size of a tumor lesion is used as a key index. FDG PET/CT can provide quantitative as well as qualitative information on metabolism of tumor cells, which is more sensitive for therapeutic effect, and it has been suggested that FDG PET/CT is an effective tool for treatment response evaluation even if FDG PET/CT is obtained in an interim phase. Additionally, recent targeted therapy agents such as multikinase inhibitors have different working mechanism and cytotoxic effects, and thus, FDG PET/CT is expected to play more important roles in response evaluation for such drugs.

✉ Jin Chul Paeng
paengjc@snu.ac.kr

✉ Sae-Won Han
saewon1@snu.ac.kr

¹ Department of Nuclear Medicine, Seoul National University Hospital, 101 Daehak-ro, Jongno-gu, Seoul 03080, Korea

² Department of Internal Medicine, Seoul National University Hospital, 101 Daehak-ro, Jongno-gu, Seoul 03080, Korea

On FDG PET/CT, metabolic activity is quantified as standardized uptake value (SUV), and changes of maximum SUV (SUV_{max}) or peak SUV (SUV_{peak}) during therapy have been investigated as indexes for evaluating treatment response and predicting prognosis [1–6]. There have been some guidelines to use SUV for treatment response evaluation, such as the European Organization for Research and Treatment of Cancer (EORTC) guideline, and the PET Response Criteria in Solid Tumors (PERCIST) [7, 8], among which the PERCIST guideline included specific methods for measurement [8]. However, there is still no consensus on how and what to measure on FDG PET/CT to evaluate treatment response.

Recently, total lesion glycolysis (TLG) has been suggested as another promising index in oncological FDG PET/CT [9–12]. TLG represents both the metabolic and volumetric information of tumor, and the PERCIST guideline also proposed TLG as an exploratory index for treatment response evaluation although any specific method was not suggested regarding measurement target, how to measure, and cutoff value for classification [8]. In the case of multikinase inhibitors, several studies have reported that change of TLG is an effective biomarker to predict the response or survival [13, 14], whereas conventional imaging markers including SUV_{max} exhibited limited effectiveness [15]. Thus, more data are required to establish methods for evaluating treatment response using TLG, particularly regarding multikinase inhibitors.

The aim of this study was to compare quantitative methods of FDG PET/CT for treatment response evaluation, in patients with metastatic colorectal cancer who were treated with regorafenib, a multikinase inhibitor. FDG PET/CT scan was performed at an interim phase of two cycles after therapy; responses determined by the PERCIST criteria and various quantitative indexes on FDG PET were compared with those finally determined by the RECIST 1.1 criteria.

Materials and Methods

Patients

A subgroup of patients who were enrolled in a prospective clinical trial for identification of predictive biomarker of regorafenib in refractory colorectal cancer (NCT01996969) was included in this study. The main clinical trial enrolled patients with pathologically proven metastatic colorectal cancers in whom standard therapies failed. Patients were treated with regorafenib (160 mg/day) with 28-day cycles until disease progression, and FDG PET/CT was performed before and two cycles after the treatment (baseline and posttreatment scans, respectively). In the present study, patients were excluded in the case of (1) no measurable hypermetabolic lesion on FDG PET/CT according to the PERCIST guideline, (2)

difference in uptake time (time interval between FDG injection and image acquisition) between baseline and posttreatment scans over 30 min, and (3) extensively disseminated and conglomerated metastatic lesions.

Clinical information was obtained by medical record review, in terms of basic demographic data, lesion sites, starting date of the treatment, and number of metastatic organs. Contrast enhanced CT was performed every two cycles, and response was evaluated according to the RECIST 1.1 criteria.

FDG PET/CT

FDG PET/CT was performed using dedicated scanners (Biograph 40 or Biograph 64, Siemens Healthcare, Knoxville, TN, USA) from the skull base to the proximal thigh. Patients fasted for at least 6 h before FDG injection. After blood glucose level was confirmed to be below 200 mg/dL (mean 100 mg/dL, range 66–158 mg/dL), FDG (5.18 MBq/kg) was intravenously injected and images were acquired approximately 60 min (67 ± 9 min, range 51–92 min) after injection. CT images (50 mA, 120 kVp, 5-mm thickness) were acquired for attenuation correction and PET images were acquired in a 3-dimensional mode for 6–7 bed positions for 1 min per bed. PET images were reconstructed on 128×128 matrices using an iterative algorithm.

Image Analysis

Images were analyzed using an analysis software package (Syngo.via, Siemens Healthcare, Knoxville, TN, USA). For FDG uptake of a lesion, SUV_{max} and SUV_{peak} were measured; SUV_{peak} was defined as mean SUV measured in a 1-cm³-sized spherical volume-of-interest (VOI) of the highest uptake. SUV_{peak} normalized to the lean body mass (SUL_{peak}) was also calculated for PERCIST-based evaluation.

For volumetric analysis, a spherical VOI was drawn to encompass a whole target lesion and isoactivity contours were automatically drawn in the VOI with varying margin thresholds; 30–70 % of the SUV_{max} with an increment of 10 %, SUV 3.0, and SUV 4.0. Additionally, twice the mean SUV of the blood pool that was measured in the descending thoracic aorta was used as a margin threshold for an isoactivity contour. TLG was calculated by multiplying mean SUV and volume in each isoactivity-contoured VOI. TLG with a specific margin threshold was abbreviated by marking the threshold value as subscript, like $TLG_{30\%}$ and $TLG_{3.0}$. TLG with the margin threshold of twice the mean SUV of the blood pool was abbreviated as TLG_{2-BP} .

Data Analysis

Three sets of lesion(s) were selected as measurement targets on baseline FDG PET/CT; (1) the hottest lesion (target A), (2)

the hottest lesions up to a maximum of five lesions total (target B), and (3) the hottest lesions up to a maximum of five lesions total with a maximum of two lesions per organ (target C). In the cases of target B and target C, quantitative values of target lesions were summed to be a single value. Additionally, response was determined by the PERCIST response criteria, which are based on SUL_{peak} values measured in target A, as metabolic complete remission (M-CR), metabolic partial remission (M-PR), metabolic stable disease (M-SD), or metabolic progressive disease (M-PD) [8]. From baseline and posttreatment FDG PET/CT scans, percent difference ($\% \Delta$) in each quantitative index was calculated and tested as a determining factor for treatment response.

For comparison of predictive value of PET indexes, response assessment according to the RECIST 1.1 criteria on CT after four cycles of treatment was used, unless PD was determined on CT after two cycles of treatment. With regard to the response assessment on posttreatment CT scan by the RECIST criteria, complete remission (CR), partial response (PR) or stable disease (SD) was classified as disease control group, and progressive disease (PD) was classified as non-control group. Also, M-CR, M-PR, and M-SD were classified as disease control group, and M-PD was classified as non-control group.

Statistical Analysis

All values were expressed as mean $\pm$ SD. The agreement between the RECIST and the PERCIST response criteria was analyzed by κ -statistic. $\% \Delta$ of PET indexes between baseline and posttreatment scans in disease control group and non-control group were compared with Mann–Whitney U test and two-tailed P -values were calculated. The efficacy of response evaluation by PET indexes was tested using receiver operating characteristic (ROC) curve analysis. ROC curves were characterized by the area under the curve (AUC). The same analyses were repeated in patients without new lesion on posttreatment scan, because presence of any new lesion results in classification into PD regardless of other factors. All statistical analyses were performed using a commercial statistical software package (MedCalc Ver. 15.8; MedCalc Software bvba, Ostend, Belgium) and P -values < 0.05 were considered statistically significant.

Results

Patients

Among 117 patients who were enrolled in the main clinical trial, a total of 39 patients were included in the final analysis (M:F = 29:10, age 61 ± 9 years). Patient and lesion characteristics are summarized in Table 1. Baseline FDG PET/CT was

Table 1 Patient characteristics

Characteristics	Value
Number of patients	39
Sex	
Male	29 (74 %)
Female	10 (26 %)
Number of metastatic organs	
1	21 (54 %)
2	13 (33 %)
≥ 3	5 (13 %)
Metastatic organs	
Lung	25
Liver	20
Lymph node	19
Peritoneal seeding	12
Bone	6
Adrenal gland	4
Others (GI tract, pancreas, pleura)	4
Number of tumor lesions	
1	1
2	2
3	2
4	0
≥ 5	34

performed 6 ± 4 days (range 2–21 days) before the initiation of treatment and posttreatment PET/CT was performed 52 ± 7 days (range 35–72 days) after two cycles of treatment. The time interval between baseline and posttreatment PET/CT was 58 ± 6 days (range 49–77 days). Baseline enhanced CT was performed 17 ± 9 days (range 2–36 days) before the initiation of treatment and, posttreatment CT was performed 97 ± 24 days (range 38–132 days) after the initiation of treatment. By the RECIST criteria, three patients were PR, 21 were SD, and 15 were PD; and thus, 24 were disease control group and 15 were non-control group.

Comparison of PERCIST and RECIST Criteria

By the PERCIST criteria, 19 patients were M-PR, eight were M-SD, and 12 were M-PD (Table 2). The agreement between responses determined by the PERCIST and the RECIST was not high; the agreement rate was 38 % (15/39) and weighted κ -coefficient was 0.381. A more favorable response was determined by the PERCIST criteria in 20 patients (51 %), and by the RECIST criteria in four patients (10 %). In two patients who were determined as PD by the PERCIST criteria, new metastatic lesions were not detected on CT and determined as SD by the RECIST criteria. When patients were classified into disease control group and non-control group, the

Table 2 Comparison of responses determined by the PERCIST and the RECIST criteria

Response by PERCIST criteria	Response by RECIST criteria			
	CR	PR	SD	PD
M-CR	0	0	0	0
M-PR	0	2	14	3
M-SD	0	1	4	3
M-PD	0	0	3	9

agreement rate between the RECIST and the PERCIST criteria was 78 %, and κ -coefficient was 0.494.

Quantitative Indexes of FDG PET

% Δ values of SUV_{max} , SUV_{peak} , and SUL_{peak} were not significantly different between disease control group and non-control group (Table 3). All tested PET indexes except TLG_{2-BP} of target A were also not significantly different. However, % Δ of TLG values measured in target B and target C were significantly different between disease control group and non-control group, with all tested margin thresholds ($P=0.001$ – 0.035 , Table 3). $TLG_{30\%}$, $TLG_{40\%}$, and $TLG_{50\%}$ exhibited higher significances than other indexes; the highest significance was obtained with $TLG_{50\%}$ of target B and $TLG_{30\%}$ of target C.

In 29 patients without new lesion on posttreatment scan, the results were similar; % Δ of $TLG_{30\%}$, $TLG_{40\%}$, and $TLG_{50\%}$ measured in target B and target C were significantly different between disease control group and non-control group ($P=0.002$ – 0.021 , Table 4)

Comparison of Quantitative PET Indexes and RECIST Criteria

In ROC curve analyses, AUC values of quantitative indexes of target B and target C were higher than those of target A (Table 5). In target B and target C, $TLG_{30\%}$, $TLG_{40\%}$, and $TLG_{50\%}$ exhibited higher AUC than those of SUV_{max} , SUV_{peak} , and SUL_{peak} . The highest AUC was obtained with $TLG_{30\%}$ of target C in all patients (-10.6 as the optimal cutoff value), and with $TLG_{50\%}$ of target B in patients without new lesions (6.4 as the optimal cutoff value) (Fig. 1).

Based on the results, a set of TLG-based criteria for treatment response evaluation was designed, in which a patient was classified as a disease control group in the case % Δ of $TLG_{50\%}$ measured in target B is <6.4 and there is no new lesion on FDG PET/CT. The κ -coefficient between the TLG-based criteria and the RECIST criteria was 0.606, which was higher than that of the PERCIST criteria. The κ -coefficient between the TLG-based criteria and the PERCIST criteria was 0.376.

Discussion

In this study, it was demonstrated that TLG values are more effective quantitative indexes for treatment response evaluation than SUV and SUL, in patients with metastatic colorectal cancer. On baseline and posttreatment scans, % Δ of TLG values exhibited more significant differences between disease control group and non-control group than SUV_{max} , SUV_{peak} , and SUL_{peak} . Among TLG values, $TLG_{30\%}$, $TLG_{40\%}$, and $TLG_{50\%}$ measured in target B or target C (the hottest lesions up to a maximum of five lesions with/without organ

Table 3 Comparison of quantitative PET indexes between disease control group and non-control group in all patients

Parameters (% Δ values)	Target A			Target B			Target C		
	Disease control	Non-control	<i>P</i>	Disease control	Non-control	<i>P</i>	Disease control	Non-control	<i>P</i>
SUV_{max}	-32.7 ± 31.0	-18.9 ± 38.9	0.202	-26.8 ± 32.4	-15.0 ± 24.1	0.086	-27.7 ± 28.3	-16.7 ± 24.6	0.163
SUV_{peak}	-32.9 ± 26.3	-17.5 ± 36.6	0.246	-24.8 ± 31.3	-11.8 ± 24.7	0.097	-26.6 ± 26.0	-14.1 ± 24.5	0.123
SUL_{peak}	-34.3 ± 26.8	-18.0 ± 38.0	0.182	-26.3 ± 31.8	-12.5 ± 25.8	0.110	-27.7 ± 25.2	-12.9 ± 31.6	0.131
$TLG_{30\%}$	-13.4 ± 48.3	49.3 ± 143.2	0.110	-8.7 ± 51.0	56.8 ± 68.9	0.002*	-10.6 ± 43.8	52.9 ± 68.4	0.001*
$TLG_{40\%}$	-11.5 ± 51.8	31.5 ± 111.8	0.309	-9.0 ± 45.2	55.1 ± 77.9	0.002*	-10.0 ± 37.3	53.7 ± 79.1	0.003*
$TLG_{50\%}$	-6.4 ± 73.2	31.8 ± 117.0	0.352	-9.0 ± 42.0	63.4 ± 104.2	0.001*	-10.2 ± 37.6	65.4 ± 108.4	0.003*
$TLG_{60\%}$	-0.1 ± 96.7	36.1 ± 117.0	0.502	-6.5 ± 50.4	75.1 ± 138.4	0.006*	-6.8 ± 48.9	80.2 ± 153.5	0.009*
$TLG_{70\%}$	-11.2 ± 89.5	58.2 ± 158.4	0.309	-13.5 ± 51.1	92.2 ± 183.3	0.007*	-14.5 ± 52.4	97.8 ± 206.4	0.016*
$TLG_{3,0}$	-45.6 ± 42.9	16.0 ± 132.6	0.081	-29.8 ± 59.7	138.1 ± 414.5	0.015*	-32.6 ± 51.7	134.2 ± 414.7	0.016*
$TLG_{4,0}$	-51.9 ± 48.4	11.7 ± 147.4	0.103	-27.4 ± 89.1	70.9 ± 214.9	0.019*	-29.8 ± 82.4	65.4 ± 207.0	0.035*
TLG_{2-BP}	-54.0 ± 45.9	1235.2 ± 4835.7	0.028*	-31.0 ± 74.9	97.7 ± 220.2	0.009*	-29.1 ± 77.6	90.9 ± 220.1	0.010*

* Statistically significant; % Δ , percent difference between baseline and posttreatment scans

Table 4 Comparison of quantitative PET indexes between disease control group and non-control group in patients without new lesion

Parameters (%Δ values)	Target A			Target B			Target C		
	Disease control	Non-control	P	Disease control	Non-control	P	Disease control	Non-control	P
SUV _{max}	-32.8 ± 32.4	-14.9 ± 55.6	0.556	-30.5 ± 27.0	-15.4 ± 29.4	0.217	-29.5 ± 28.1	-16.0 ± 30.3	0.304
SUV _{peak}	-33.1 ± 27.4	-16.2 ± 49.7	0.600	-28.3 ± 25.6	-11.5 ± 29.2	0.258	-28.2 ± 25.2	-13.5 ± 28.9	0.237
SUL _{peak}	-34.6 ± 27.9	-15.8 ± 51.6	0.500	-29.9 ± 26.0	-11.3 ± 30.5	0.217	-26.9 ± 25.8	-11.3 ± 43.4	0.500
TLG _{30%}	-18.1 ± 41.8	35.3 ± 68.6	0.070	-13.9 ± 42.1	57.4 ± 68.8	0.006*	-14.5 ± 38.2	52.2 ± 63.8	0.005*
TLG _{40%}	-15.4 ± 47.7	23.8 ± 67.7	0.217	-13.7 ± 32.3	61.8 ± 90.1	0.008*	-13.4 ± 31.5	56.6 ± 90.7	0.011*
TLG _{50%}	-8.8 ± 75.0	16.8 ± 70.6	0.438	-12.7 ± 34.9	72.4 ± 123.7	0.002*	-12.7 ± 34.9	67.9 ± 131.8	0.021*
TLG _{60%}	-4.2 ± 98.8	22.0 ± 78.0	0.258	-9.9 ± 45.4	99.2 ± 215.3	0.015*	-11.1 ± 43.5	98.4 ± 188.8	0.070
TLG _{70%}	-15.5 ± 89.0	46.7 ± 145.2	0.149	-19.3 ± 40.0	141.6 ± 257.5	0.018*	-18.2 ± 46.3	147.6 ± 295.0	0.110
TLG _{3,0}	-49.7 ± 40.2	0.4 ± 76.5	0.122	-37.6 ± 46.2	49.4 ± 160.7	0.088	-38.6 ± 44.3	45.2 ± 154.8	0.098
TLG _{4,0}	-55.0 ± 48.4	-11.9 ± 85.2	0.258	-34.6 ± 83.9	84.6 ± 295.4	0.135	-35.0 ± 81.4	77.3 ± 281.1	0.181
TLG _{2-BP}	-56.6 ± 46.4	-4.0 ± 84.8	0.181	-33.5 ± 78.0	67.4 ± 224.6	0.078	-37.9 ± 70.5	71.0 ± 226.4	0.098

* Statistically significant; %Δ, percent difference between baseline and posttreatment scans

limitation) appeared to be the most appropriate quantitative index, as they exhibited highest statistical significances.

FDG PET/CT has been proven to be an effective tool for treatment response monitoring in advanced colorectal cancers [16–21]. Posttreatment changes of SUV or kinetic parameters on FDG PET were effective indexes for treatment response evaluation, and were significantly related to prognosis. However, most previous studies included colorectal cancer patients only with hepatic metastasis, who were treated with standard chemotherapy. Because recently used targeted therapy agents have a different mechanism from those of conventional chemotherapeutic agents, methods for response evaluation should be reassessed. Currently, the EORTC criteria and the PERCIST criteria are often used as standard methods. In a recent study, it was reported that the EORTC criteria and the PERCIST criteria have similar diagnostic and predictive values for treatment response evaluation in metastatic

colorectal cancer treated with irinotecan and cetuximab [22]. However, the methods for treatment response evaluation still require more refinement and validation, particularly with regard to recently developed targeted therapy agents such as a mutikinase inhibitor used in the current study [23].

In the present study, %Δ of SUV_{max} was not effective for predicting the response, which is consistent with a previous study [15]. Instead, %Δ of TLG with various margin thresholds was effective to predict the response. Contrary to SUV_{max} or SUV_{peak} that represents the most malignant cell component of a lesion, TLG reflects both metabolic activity and tumor burden of a lesion. Thus, TLG is expected to be more appropriate to assess overall progression status of malignancy in the case of multiple metastatic diseases. The PERCIST guideline suggested that two or three times SD above SUL of the normal liver may be a less variable margin threshold for measuring TLG than that defined by a percentage of maximal uptake [8].

Table 5 AUC values of quantitative PET indexes in ROC curve analysis

Parameters (%Δ)	All patients			Patients without new lesion		
	Target A	Target B	Target C	Target A	Target B	Target C
SUV _{max}	0.626	0.667	0.636	0.578	0.662	0.636
SUV _{peak}	0.614	0.661	0.650	0.571	0.649	0.656
SUL _{peak}	0.631	0.656	0.647	0.591	0.662	0.591
TLG _{30%}	0.656	0.797	0.808*	0.734*	0.838	0.844*
TLG _{40%}	0.600	0.792	0.783	0.662	0.831	0.818
TLG _{50%}	0.592	0.803*	0.781	0.604	0.877*	0.792
TLG _{60%}	0.567	0.758	0.747	0.649	0.805	0.734
TLG _{70%}	0.600	0.756	0.731	0.688	0.799	0.708
TLG _{3,0}	0.669	0.731	0.658	0.701	0.721	0.714
TLG _{4,0}	0.658	0.725	0.703	0.649	0.695	0.675
TLG _{2-BP}	0.711*	0.750	0.744	0.675	0.727	0.714

* The highest AUC in each target set

Fig. 1 Optimal cutoff value for $\% \Delta$ of TLG_{30%} of target C in all patients (a), and $\% \Delta$ of TLG_{50%} of target B in patients without new lesions (b)

However, the blood pool activity was used as a reference value in the present study, because most patients had hepatic metastasis and reliability of the normal hepatic SUV was limited. Intriguingly, margin threshold of 30 to 50 % of SUV_{max} resulted in higher statistical significance in predicting response than margin thresholds based on the blood pool activity or a specific SUV. This is a promising result, because 30 to 50 % of SUV_{max} is frequently used as a margin threshold for measuring metabolic tumor volume or TLG due to its convenience.

Another issue in treatment response evaluation is target of measurement. In the RECIST 1.0, measurement target was a maximum of ten lesions total with a maximum of five lesions per organ, which was changed in the RECIST 1.1 into a maximum of five lesions total with a maximum of two lesions per organ [24]. In the PERCIST guideline, it was suggested that the hottest single tumor lesion should be a target lesion; however, a maximum of five measurable target lesions could be recorded as exploratory data. Recently, it was reported that the number of lesions assessed by the PERCIST guideline does not affect the result of response evaluation in metastatic breast cancer patients [25]. In the present study, target lesions were up to a maximum of five lesions total in target B and target C, which resulted in higher statistical significance than target A, the single hottest lesion. Thus, it can be suggested that sampling of several targets would be a better method than sampling a single lesion. However, the limitation in sampling number per organ does not appear to be critical, because the difference between results from target B and target C was minimal. The effect of the limitation in sampling number per organ needs to be investigated in further studies.

Most of the patients enrolled in this study had more than four tumor lesions on baseline scans. In the case of interlesional heterogeneity of treatment response, the hottest target lesions could be changed between baseline and post-treatment scans. In this study, $\% \Delta$ was assessed in the same target lesions as those of the baseline scan. If the target lesions were determined by the order of high uptake on each scan, measurement target for targets A and C would have been

changed in more than 50 % of cases. However, even in such cases, the results exhibited a similar tendency (data not shown). Therefore, it appears possible to determine measurement target by the order of high uptake on each scan.

The present study has several limitations. First, a relatively small number of patients were included in the study and the statistical power was limited. Additionally, the response was finally determined by CT images using the RECIST criteria. Considering that regorafenib therapy might not cause definite size changes on CT images [26], larger prospective trials with FDG PET/CT with regard to patient prognosis is required for validation. Lastly, while TLG measured by margin thresholds of 30 to 50 % exhibited the most significant statistical results in this study, optimal cutoff value for determining PR, SD, and PD were not investigated in this study due to small sample size. Further studies are warranted to determine optimal cutoff values of each TLG measurement for optimal classification of response groups.

Conclusion

In metastatic colorectal cancer, FDG PET/CT could be effective for treatment response evaluation by using TLG measured by margin thresholds of 30 to 50 % of SUV_{max}. Further studies are warranted regarding the optimal cutoff values for these methods.

Acknowledgments This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (009–0093820).

Compliance with Ethical Standards

Conflict of Interest Ji-In Bang, Yoojoo Lim, Jin Chul Paeng, Sae-Won Han, Sohyun Park, Jung Min Lee, Hyun Joo Kim, Gi Jeong Cheon, Dong Soo Lee, June-Key Chung, Tae-You Kim, and Keon Wook Kang declare that they have no conflict of interest.

Ethical Approval All procedures followed were performed in accordance with the ethical standards of the responsible committee on human experimentation and with the Helsinki Declaration of 1975, as revised in 2013. The study design of the prospective clinical trial was approved by the Institutional Review Board of the Seoul National University Hospital and all subjects in the study gave written informed consent (IRB number: 1307-144-507). The study design of the retrospective analysis and exemption of informed consent were approved by the Institutional Review Board of the Seoul National University Hospital (H-1605-063-761).

This manuscript has not been published before or is not under consideration for publication anywhere else and has been approved by all co-authors.

References

- Eschmann SM, Friedel G, Paulsen F, Reimold M, Hehr T, Budach W, et al. Repeat 18F-FDG PET for monitoring neoadjuvant chemotherapy in patients with stage III non-small cell lung cancer. *Lung Cancer*. 2007;55:165–71.
- Lin C, Itti E, Haioun C, Petegnief Y, Luciani A, Dupuis J, et al. Early 18F-FDG PET for prediction of prognosis in patients with diffuse large B-cell lymphoma: SUV-based assessment versus visual analysis. *J Nucl Med*. 2007;48:1626–32.
- Berghmans T, Dusart M, Paesmans M, Hossein-Foucher C, Buvat I, Castaigne C, et al. Primary tumor standardized uptake value (SUVmax) measured on fluorodeoxyglucose positron emission tomography (FDG-PET) is of prognostic value for survival in non-small cell lung cancer (NSCLC): a systematic review and meta-analysis (MA) by the European Lung Cancer Working Party for the IASLC Lung Cancer Staging Project. *J Thorac Oncol*. 2008;3:6–12.
- Benz MR, Czernin J, Allen-Auerbach MS, Tap WD, Dry SM, Elashoff D, et al. FDG-PET/CT imaging predicts histopathologic treatment responses after the initial cycle of neoadjuvant chemotherapy in high-grade soft-tissue sarcomas. *Clin Cancer Res*. 2009;15:2856–63.
- Yanagawa M, Tatsumi M, Miyata H, Morii E, Tomiyama N, Watabe T, et al. Evaluation of response to neoadjuvant chemotherapy for esophageal cancer: PET response criteria in solid tumors versus response evaluation criteria in solid tumors. *J Nucl Med*. 2012;53:872–80.
- Hatt M, van Stiphout R, le Pogam A, Lammering G, Visvikis D, Lambin P. Early prediction of pathological response in locally advanced rectal cancer based on sequential 18F-FDG PET. *Acta Oncol*. 2013;52:619–26.
- 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.
- 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.
- Byun BH, Kong CB, Park J, Seo Y, Lim I, Choi CW, et al. Initial metabolic tumor volume measured by 18F-FDG PET/CT can predict the outcome of osteosarcoma of the extremities. *J Nucl Med*. 2013;54:1725–32.
- Pak K, Cheon GJ, Nam HY, Kim SJ, Kang KW, Chung JK, et al. Prognostic value of metabolic tumor volume and total lesion glycolysis in head and neck cancer: a systematic review and meta-analysis. *J Nucl Med*. 2014;55:884–90.
- Rahim MK, Kim SE, So H, Kim HJ, Cheon GJ, Lee ES, et al. Recent trends in PET image interpretations using volumetric and texture-based quantification methods in nuclear oncology. *Nucl Med Mol Imaging*. 2014;48:1–15.
- Kurtipek E, Cayci M, Duzgun N, Esme H, Terzi Y, Bakdik S, et al. (18)F-FDG PET/CT mean SUV and metabolic tumor volume for mean survival time in non-small cell lung cancer. *Clin Nucl Med*. 2015;40:459–63.
- Minamimoto R, Barkhodari A, Harshman L, Srinivas S, Quon A. Prognostic value of quantitative metabolic metrics on baseline Pre-Sunitinib FDG PET/CT in advanced renal cell carcinoma. *PLoS One*. 2016;11, e0153321.
- Farnebo J, Gryback P, Harmenberg U, Laurell A, Wersall P, Blomqvist LK, et al. Volumetric FDG-PET predicts overall and progression-free survival after 14 days of targeted therapy in metastatic renal cell carcinoma. *BMC Cancer*. 2014;14:408.
- Wong AL, Lim JS, Sinha A, Gopinathan A, Lim R, Tan CS, et al. Tumour pharmacodynamics and circulating cell free DNA in patients with refractory colorectal carcinoma treated with regorafenib. *J Transl Med*. 2015;13:57.
- Findlay M, Young H, Cunningham D, Iveson A, Cronin B, Hickish T, et al. Noninvasive monitoring of tumor metabolism using fluorodeoxyglucose and positron emission tomography in colorectal cancer liver metastases: correlation with tumor response to fluorouracil. *J Clin Oncol*. 1996;14:700–8.
- Bender H, Bangard N, Metten N, Bangard M, Mezger J, Schomburg A, et al. Possible role of FDG-PET in the early prediction of therapy outcome in liver metastases of colorectal cancer. *Hybridoma*. 1999;18:87–91.
- Dimitrakopoulou-Strauss A, Strauss LG, Rudi J. PET-FDG as predictor of therapy response in patients with colorectal carcinoma. *Q J Nucl Med Mol Imaging*. 2003;47:8–13.
- Dimitrakopoulou-Strauss A, Strauss LG, Burger C, Ruhl A, Irgartinger G, Stremmel W, et al. Prognostic aspects of 18F-FDG PET kinetics in patients with metastatic colorectal carcinoma receiving FOLFOX chemotherapy. *J Nucl Med*. 2004;45:1480–7.
- de Geus-Oei LF, van Laarhoven HW, Visser EP, Hermens R, van Hoorn BA, Kamm YJ, et al. Chemotherapy response evaluation with FDG-PET in patients with colorectal cancer. *Ann Oncol*. 2008;19:348–52.
- de Geus-Oei LF, Vriens D, van Laarhoven HW, van der Graaf WT, Oyen WJ. Monitoring and predicting response to therapy with 18F-FDG PET in colorectal cancer: a systematic review. *J Nucl Med*. 2009;50 Suppl 1:43S–54.
- 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.
- Grothey A, Van Cutsem E, Sobrero A, Siena S, Falcone A, Ychou M, et al. Regorafenib monotherapy for previously treated metastatic colorectal cancer (CORRECT): an international, multicentre, randomised, placebo-controlled, phase 3 trial. *Lancet*. 2013;381:303–12.
- Chalian H, Tore HG, Horowitz JM, Salem R, Miller FH, Yaghmai V. Radiologic assessment of response to therapy: comparison of RECIST Versions 1.1 and 1.0. *Radiographics*. 2011;31:2093–105.
- Pinker K, Riedl C, Ong L, Jochelson M, Ulaner GA, McArthur H, et al. Impact of the number of lesions analyzed in metastatic breast cancer on response assessment by 18F-FDG PET/CT using PERCIST. *J Nucl Med*. 2016; doi:10.2967/jnumed.115.166629.
- Lim Y, Han SW, Yoon JH, Lee JM, Paeng JC, et al. Clinical implication of anti-angiogenic effect of regorafenib in metastatic colorectal cancer. *PLoS One*. 2015;10, e0145004.