New Response Evaluation Criteria in Solid Tumors (RECIST) Guidelines for Advanced Non–Small Cell Lung Cancer: Comparison With Original RECIST and Impact on Assessment of Tumor Response to Targeted Therapy

Mizuki Nishino[1], **David M. Jackman**[2], **Hiroto Hatabu**[3], **Beow Y. Yeap**[4], **Leigh-Anne Cioffredi**[2], **Jeffrey T. Yap**[1], **Pasi A. Jänne**[2], **Bruce E. Johnson**[2], and **Annick D. Van den Abbeele**[1]

[1]Department of Radiology, Dana-Farber Cancer Institute, 44 Binney St., Boston, MA 02116.

[2]Department of Medical Oncology, Dana-Farber Cancer Institute and Brigham and Women's Hospital, Boston, MA.

[3]Department of Radiology, Brigham and Women's Hospital, Boston, MA.

[4]Department of Medicine, Massachusetts General Hospital, Boston, MA.

Abstract

OBJECTIVE—The purpose of this article is to compare the recently published revised Response Evaluation Criteria in Solid Tumors (RECIST) guidelines (version 1.1) to the original guidelines (RECIST 1.0) for advanced non–small cell lung cancer (NSCLC) after erlotinib therapy and to evaluate the impact of the new CT tumor measurement guideline on response assessment.

MATERIALS AND METHODS—Forty-three chemotherapy-naive patients with advanced NSCLC treated with erlotinib in a single-arm phase 2 multicenter open-label clinical trial were retrospectively studied. CT tumor measurement records using RECIST 1.0 that were generated as part of the prospective clinical trial were reviewed. A second set of CT tumor measurements was generated from the records to meet RECIST 1.1 guidelines. The number of target lesions, best response, and time to progression were compared between RECIST 1.1 and RECIST 1.0.

RESULTS—The number of target lesions according to RECIST 1.1 decreased in 22 patients (51%) and did not change in 21 patients (49%) compared with the number according to RECIST 1.0 ($p < 0.0001$, paired Student's *t* test). Almost perfect agreement was observed between best responses using RECIST 1.1 and RECIST 1.0 (weighted $\kappa = 0.905$). Two patients with stable disease according to RECIST 1.0 had progressive disease according to RECIST 1.1 criteria because of new lesions found on PET/CT. There was no significant difference in time to progression between RECIST 1.1 and RECIST 1.0 ($p = 1.000$, sign test).

CONCLUSION—RECIST 1.1 provided almost perfect agreement in response assessment after erlotinib therapy compared with RECIST 1.0. Assessment with PET/CT was a major factor that influenced the difference in best response assessment between RECIST 1.1 and RECIST 1.0.

Keywords

CT; lung cancer; PET/CT; Response Evaluation Criteria in Solid Tumors (RECIST); tumor measurement

Objective assessment of the change in tumor burden is important for evaluating the tumor response to anticancer drugs as a surrogate for symptomatic response, and objective response has been shown to be closely correlated with prolonged survival with solid tumors. The Response Evaluation Criteria in Solid Tumors (RECIST) guideline was introduced in 2000 by an international working group to standardize and simplify tumor response criteria [1]. The key features of the original RECIST (version 1.0) included definitions of minimum CT size of measurable lesions, instructions on how many lesions to follow (up to 10, with a maximum of five per organ), and the use of unidimensional CT measures for overall evaluation of tumor burden [1, 2]. RECIST supplanted the bidimensional tumor size assessment that had been in common use.

RECIST has subsequently been widely accepted as a standardized measure of tumor response, especially in clinical trials where the primary end points are objective response or time to progression [2]. However, with rapid technical innovations in imaging techniques, such as MDCT and PET/CT, over the past decade, the limitations of the original RECIST and the need for revision have become clear [2, 3].

In January 2009, a revised RECIST guideline (version 1.1) was published by the RECIST Working Group, based in part on the investigations using the database consisting of more than 6,500 patients with more than 18,000 target lesions [2, 4–6]. Major imaging-related changes in RECIST 1.1 included a reduction in the number of lesions to be assessed, from a maximum of 10 to a maximum of five and from five per organ to two per organ; assessment of lymph node size (lesions ≥ 15 mm in the short axis are considered measurable and assessable as target lesions, and the short-axis measurement should be included in the sum of lesions in calculation of the tumor response); clarification of disease progression (in addition to the previous definition of progression in target disease of 20% increase in sum, a 5-mm absolute increase is now required); and inclusion of FDG PET assessment exclusively in the section on detection of new lesions [2] (Table 1).

It is anticipated that the new guideline improves feasibility and provides more accurate assessment of tumor response and time to progression while incorporating the current state-of-art imaging in the clinical oncologic practice. However, it remains to be seen how RECIST 1.1 will affect the selection and the CT measurement of the target lesions, assessment of tumor response to therapy, and time to progression compared with RECIST 1.0.

Lung cancer is a leading cause of cancer death in the United States and worldwide, accounting for over 150,000 deaths per year in the United States [7]. Eighty-five percent of subjects have non–small cell lung cancer (NSCLC), for which the 5-year survival rate is only 15% [7]. Newer therapeutic agents have been clinically applied for lung cancer, including erlotinib, an inhibitor of the tyrosine kinase domain of the epidermal growth factor receptor [8–10]. Erlotinib can be associated with dramatic clinical response in patients with epidermal growth factor receptor mutations. However, patients with an initial response to erlotinib eventually have a relapse due to development of acquired resistance [11–13]. With the emerging preclinical understanding of the mechanisms of erlotinib resistance, new therapies targeting these mechanisms are being developed. Therefore, the accurate CT tumor measurement and response assessment in advanced NSCLC treated with targeted therapy are critically important to determine the timing to adjust the therapeutic regimen for further prolongation of survival.

The purpose of this study is to compare CT tumor measurement and tumor response assessment based on RECIST 1.1 versus RECIST 1.0 in a phase 2 clinical trial of patients with advanced NSCLC treated with erlotinib.

Materials and Methods

Patients and Treatment

The original clinical trial was already performed with 80 eligible patients [14]. The study was approved by the Dana-Farber/Harvard Cancer Center institutional review board. Of 80 patients, complete CT tumor measurement records were available for 43 patients. Therefore, the study population of the current study consisted of 43 patients (22 men and 21 women; mean age, 77 years; range, 70–91 years) with histologically confirmed stage IIIB/IV NSCLC that had not been treated with chemotherapy. Patients enrolled in this phase 2 multicenter open-label study were treated with 150 mg of erlotinib by mouth each day as part of first-line therapy between March 2003 and May 2005 [14]. Thirty-nine patients were treated at the Dana-Farber Cancer Institute, and four patients were treated at the Massachusetts General Hospital. All patients gave written informed consent. Patients were treated without interruption until disease progression, severe or intolerable toxicity, or withdrawal of consent. Compliance was checked after each 28-day cycle with a treatment diary.

CT Examinations

CT scans of the chest were obtained at baseline and at every two cycles (8 weeks) of therapy to determine response to erlotinib or progression of disease. Each clinical trial site used their standard clinical chest CT protocol with iodinated IV contrast agent unless medically contraindicated. The protocol at the Dana-Farber Cancer Institute was as follows: a 4-MDCT scanner (Volume Zoom, Siemens Healthcare) was used. Patients were scanned in the supine position from the cranial to caudal direction from the clavicles to the adrenal glands at end-inspiration. One hundred microliters of iopromide (300 mg I/mL; Ultravist 300, Bayer HealthCare Pharmaceuticals) was injected IV with an automated injector (Stellant, Medrad) at a rate of 2–3 mL/s, with a scan delay of 30 seconds. Axial images (5 or 7 mm thickness) were reconstructed using standard and lung algorithms and were transferred to the PACS.

CT Tumor Measurement and Assessment of Tumor Response to Therapy

CT tumor measurement was prospectively performed by attending radiologists at Dana-Farber Cancer Institute and was recorded at the baseline and at every follow-up CT examination according to the original RECIST 1.0 criteria as a part of the ongoing phase 2 clinical trial [14]. The longest diameter of each target lesion was manually measured by radiologists on an axial CT image plane using calipers of a measurement tool on PACS. The CT tumor measurement record included the number of the treatment cycle, the date of assessment, the clinical site where imaging was performed, the method of imaging, the target lesion description and CT size measurement, the sum of the longest tumor diameters of target lesions, descriptions of nontarget lesions, the presence or absence of new lesions, the overall response for each imaging study, and the best response and time to progression for each patient.

The results of the RECIST 1.0 CT measurements for the completed clinical trial were retrospectively reviewed by a board-certified thoracic radiologist who was blinded to the patients' outcome, to generate a second set of CT tumor measurements that meet the RECIST 1.1 guideline. Briefly, the target lesions recorded in the original measurements were reassessed if they met the criteria of RECIST 1.1: lymph nodes less than 15 mm in the short axis were excluded from target lesions; when the number of target lesions exceeded the limits according to RECIST 1.1 (up to five in total and up to two per organ), smaller lesions were eliminated from target lesions; short-axis measurements were used for lymph nodes instead of long-axis measurements; and the PET/CT clinical reports were also

reviewed for the patients who underwent PET/CT during the treatment to determine whether any new lesion was detected on PET/CT meeting the RECIST 1.1 criteria for progression. The number of RECIST 1.1 target lesions and the sums of tumor diameters at baseline and follow-up was calculated and recorded. The percentage change in the sum of tumor diameters of target lesions was calculated and recorded at every two cycles of therapy. Tumor response to therapy was reassessed using the revised RECIST 1.1 for each CT measurement, and best response for each patient was assigned.

Time to progression and survival were calculated from the date of enrollment to the date of progression, using RECIST 1.1 and RECIST 1.0, or to the time of death, respectively. Time to progression was censored (i.e., losses from the sample before the observation of the final outcome were taken into account) if an end point was not reached by the time of last follow-up, or if a patient was lost to follow-up, or if it was not possible to determine the date of progression according to RECIST 1.1. Time to progression was estimated using the Kaplan-Meier method [15].

To assess interobserver variability, a new set of CT tumor measurements was obtained on the baseline CT scans of 39 patients treated at the Dana-Farber Cancer Institute using RECIST 1.0 and RECIST 1.1 by a board-certified thoracic radiologist. Of 43 patients, four were treated at an outside institution, and their CT scans were not available for this component of the study. The CT measurement was performed using a measurement tool on PACS workstation (Centricity, GE Healthcare). The sum of the longest diameters of the target lesions measured by this observer using RECIST 1.0 and RECIST 1.1 were compared with the CT measurements obtained during the clinical trial using RECIST 1.0 and RECIST 1.1, respectively. The CT measurement performed during the clinical trial (measurement 1) and that performed by the observer in the present study (measurement 2) were compared using Pearson's correlation and linear regression. Agreement in the two CT measurements was shown visually using Bland-Altman plots with 95% limits of agreement [16, 17].

Statistical Analysis

A paired Student's *t* test was used to assess the statistical significance of changes in the number of target lesions and the sum of lesion diameters at baseline between RECIST 1.1 and RECIST 1.0. The baseline CT measurements by RECIST 1.1 versus RECIST 1.0 as well as the percentage changes in follow-up CT measurements relative to baseline were compared using Pearson's correlation and linear regression.

The level of agreement between best responses by RECIST 1.1 versus RECIST 1.0 was assessed using a weighted kappa analysis based on the absolute difference of equally spaced scores. Agreement between the two assessments was categorized as poor (weighted $\kappa < 0$), slight (weighted $\kappa = 0\text{--}0.20$), fair (weighted $\kappa = 0.21\text{--}0.40$), moderate (weighted $\kappa = 0.41\text{--}0.60$), substantial (weighted $\kappa = 0.61\text{--}0.80$), and almost perfect (weighted $\kappa > 0.80$) [18].

Time to progression according to RECIST 1.1 versus RECIST 1.0 was compared using a sign test. All *p* values are based on a two-sided hypothesis. A *p* value of less than 0.05 was considered to be significant.

Results

The number of target lesions using RECIST 1.1 was significantly lower than that using RECIST 1.0 ($p < 0.0001$, paired Student's *t* test), with a decrease in the lesion number in 22 patients (51%) and no change in the remaining 21 patients (49%) (Fig. 1). The number of target lesions according to RECIST 1.0 ranged from one to nine (mean, 2.7; median, 2; mode, 2), whereas the number of target lesions according to RECIST 1.1 ranged from zero

to five (mean, 1.8; median, 2; mode, 1). The number of target lesions according to RECIST 1.1 was decreased by one in 15 patients, by two in two patients, by three in two patients, and by four in three patients, with a mean decrease of 0.86. The number of the target lesions was decreased as a result of the new definition of measurability of malignant lymph nodes at the baseline (a lymph node must be 15 mm in the short axis to be considered pathologically enlarged and measurable [2]), for 18 patients, and as a result of the reduction of the number of lesions required to assess tumor burden (from a maximum of 10 to a maximum of five total, and from five to two per organ, maximum), for four patients. There were three patients who had no target lesions when RECIST 1.1 was used. In these patients, all the target lesions according to RECIST 1.0 were lymph nodes smaller than 1.5 cm in the short axis and did not meet the RECIST 1.1 criteria for a target lesion.

Comparison of the sum of tumor diameters of target lesions on baseline CT measurements by RECIST 1.1 and by RECIST 1.0 is shown in Figure 2A. A high correlation was noted between the baseline CT measurements by RECIST 1.1 and RECIST 1.0 ($r = 0.936$ and $r^2 = 0.876$, Pearson's correlation). The baseline CT measurement using RECIST 1.1 was significantly smaller than that using RECIST 1.0, with a mean difference of −1.6 cm and a regression slope of 0.73 ($p = 0.0001$, paired Student's t test). The distribution of the differences between baseline CT measurements by RECIST 1.0 and by RECIST 1.1 is shown in Figure 2B.

The percentage changes between RECIST 1.1 and RECIST 1.0 in the sum of tumor diameters of target lesions showed a high correlation ($r = 0.986$ and $r^2 = 0.972$ at cycle 2 [$n = 28$]; $r = 0.998$ and $r^2 = 0.996$ at cycle 4 [$n = 7$]; $r = 0.657$ and $r^2 = 0.432$ at cycle 6 [$n = 5$]; $r = 0.949$ and $r^2 = 0.901$ at cycle 8 [$n = 4$]; $r =$ not applicable and $r^2 =$ not applicable at cycle 10, [$n = 1$]; and $r = 0.979$ and $r^2 = 0.959$ at cycles 2–10 [$n = 45$], Pearson's correlation) (Figs. 3 and 4). One patient showed discrepant response assessment based on the percentage change in the sum of tumor diameters of target lesions according to RECIST 1.1 and RECIST 1.0 after the first two cycles of therapy (stable disease according to RECIST 1.1 and progressive disease according to RECIST 1.0; circle in Figure 3, asterisk in Figure 4). However, coincidently the patient also had a new lesion on CT performed at two cycles of therapy. Therefore, the overall response assessment of this patient at two cycles of therapy was progressive disease by both RECIST 1.1 and RECIST 1.0, resulting in the same best response and time to progression by RECIST 1.1 and by RECIST 1.0.

Of 43 patients, the best response according to RECIST 1.1 was the same as the best response according to RECIST 1.0 in 40 patients (93%) and was different in three patients (7%). Almost perfect agreement was observed between the best response according to RECIST 1.1 versus the best response according to RECIST 1.0 (weighted κ = 0.905) (Table 2). Among three patients with differences in best response according to RECIST 1.1 and RECIST 1.0, two patients had stable disease based on RECIST 1.0 but had progressive disease according to RECIST 1.1 because of new lesions noted on PET/CT. The other patient had progressive disease using RECIST 1.0 but was assessed as stable disease according to RECIST 1.1 because of a new lesion by RECIST 1.0 that was a lymph node measuring less than 10 mm in the short axis, which does not meet the new criteria of pathologic lymph node by RECIST 1.1. This patient was removed from the clinical trial and the erlotinib therapy when progressive disease was determined on the basis of RECIST 1.0, whereas the patient could have continued the erlotinib therapy if RECIST 1.1 were used for assessment.

Of 43 patients, only six (14%) underwent PET/CT during the course of the treatment. Of these six patients, four had no findings on PET/CT that influenced the assessment of response therapy; however, two (33%) of the six patients had new lesions on PET/CT, which changed the best response from stable disease to progressive disease. In the actual

trial utilizing RECIST 1.0, which does not include new lesions on PET/CT as a criterion for disease progression, these two patients were not considered to meet the criteria of progression at the time of PET/CT and continued with the trial. These two patients could have come off the clinical trial sooner if RECIST 1.1 criteria including PET/CT were used.

Time to progression did not show any significant difference (p = 1.000, sign test) between RECIST 1.1 and RECIST 1.0 (Fig. 5). Time to progression between RECIST 1.1 and RECIST 1.0 did not differ in 40 patients (93%). Time to progression according to RECIST 1.1 was shorter than that according to RECIST 1.0 in two patients whose best response changed from stable disease according to RECIST 1.0 to progressive disease according to RECIST 1.1 (54 days according to RECIST 1.1 vs 108 days according to RECIST 1.0; and 51 days according to RECIST 1.1 vs 92 days according to RECIST 1.0). In one patient whose best response changed from progressive disease by RECIST 1.0 to stable disease by RECIST 1.1, time to progression according to RECIST 1.1 was longer than that according to RECIST 1.0. For this patient, it was not possible to obtain the exact days of time to progression by RECIST 1.1 because the patient was removed from the protocol and the erlotinib therapy was discontinued at the time of documented progression by RECIST 1.0 (54 days) in the actual clinical trial.

A high correlation was noted between the CT measurement performed during the clinical trial (measurement 1) and that performed by the observer in the present study (measurement 2) for both RECIST 1.0 (r^2 = 0.9740, Pearson's correlation) and RECIST 1.1 (r^2 = 0.9846, Pearson's correlation) (Figs. 6A and 6B). Figures 6C and 6D show the Bland-Altman plots with the mean percentage of relative difference and the limits of agreement. For the two CT measurements using RECIST 1.0, the mean difference was −0.2% (95% limits of agreement, −30.8%, 30.4%) (Fig. 6C). For the two measurements using RECIST 1.1, the mean difference was 3.4% (95% limits of agreement, −18.6%, 25.4%) (Fig. 6D).

Discussion

The RECIST 1.1 criteria provided almost perfect agreement in the assessment of tumor response to therapy compared with the RECIST 1.0 criteria, with a decreased number of target lesions in patients with NSCLC treated with targeted therapy. To our knowledge, this is the first report that evaluated the impact of RECIST 1.1 on CT tumor measurement and the response assessment since the new guideline has been published [2].

The most common reason for the decrease in the number of target lesions using RECIST 1.1 was the new definition of measurability of malignant lymph nodes at the baseline (a lymph node must be at least 15 mm in the short axis to be considered pathologically enlarged and measurable as a target lesion [2]), which affected 18 patients. The new definition is intended to eliminate the inclusion of lymph nodes less than 15 mm in the short axis as target lesions [6]. Another reason for the decreased number of target lesions was the reduction in the number of lesions required to assess tumor burden (from a maximum of 10 to a maximum of five, and from five to two per organ), which influenced four patients [2]. Three of these four patients had three lesions in the lung according to RECIST 1.0. The smallest of the three lesions at the baseline was excluded from the target lesions. The other patient had widespread disease in the chest, upper abdomen, and ribs, with a total of nine target lesions according to RECIST 1.0. The largest five lesions by the longest diameter were chosen in keeping with the limit of two lesions per organ. Three patients (7%) had no target lesions according to RECIST 1.1, because all the target lesions according to RECIST 1.0 were lymph nodes smaller than 1.5 cm in the short axis, which does not meet the RECIST 1.1 criteria for a target lesion. Hence, these three patients would be nonassessable patients and would likely be excluded from participation in a prospective trial using RECIST 1.1. The

result indicates that RECIST 1.1 influences the eligibility of the patients for clinical trials, which requires measurable lesion by RECIST at baseline for enrollment.

The baseline RECIST CT measurement was decreased in 25 patients (58%) when RECIST 1.1 was used. The mean difference in the baseline CT measurement of summed lesion diameters was 1.6 cm, and 19 (76%) of the 25 patients with decreases in baseline CT measurement by RECIST 1.1 had reductions of less than 3 cm. The decrease in the baseline CT size measurement was due to a decrease in the number of target lesions in 22 patients. For three patients, the baseline CT measurement by RECIST 1.1 was decreased because of the inclusion of the short-axis measurement for the lymph node, rather than the long axis, without a decrease in the number of target lesions. The decrease in the baseline CT measurement for these three patients was less than 1 cm. RECIST 1.1 provided concordant response assessment of target lesions with RECIST 1.0 at each measurement on CT scans at every two cycles of therapy except for one (2.2%) of 45 patients.

New lesions noted on PET/CT changed the best response for two patients from stable disease according to RECIST 1.0 to progressive disease according to RECIST 1.1. The results suggest that the inclusion of PET/CT in the assessment of new lesions may have a significant influence on response assessment by RECIST 1.1. In the current study, only six patients underwent PET/CT because PET/CT was not a part of the clinical trial protocol and was performed as a part of clinical care only if it was clinically indicated and requested by the referring physician. A study with a larger patient cohort with periodic PET/CT assessment is needed to further evaluate its impact.

Time to progression was different between RECIST 1.1 and 1.0 for only three patients (shortened for two and prolonged for one). In the cohort of 43 elderly patients with advanced NSCLC used in the present study, no patient showed complete or partial response. In a different clinical subset with dramatic initial response to erlotinib and subsequent slow growth of the tumor, as seen in women who are nonsmokers or former smokers, RECIST 1.1 with new criteria of progression (i.e., a 5-mm absolute increase is now required for progressive disease) may have more influence in the assessment of time to progression.

In the assessment of interobserver variability using two sets of CT measurements, the 95% limits of agreement were approximately within plus or minus 30% for both RECIST 1.0 and RECIST 1.1. The results are similar to the recently published interobserver variability data of NSCLC tumor measurements by Zhao et al. [17], indicating that the CT tumor measurements are reproducible within the partial response category (−30%); however, this finding raises a concern about placing patients in the progression disease category (+20%) too easily [18]. Of note, the present study showed that the 95% limits of agreement were better in the measurements using RECIST 1.1 (−18.6%, 25.4%) than in the measurements using RECIST 1.0 (−30.8%, 30.4%). The result suggests that the use of RECIST 1.1 may improve the reproducibility of CT tumor measurements and therefore may contribute to the adequate response assessment after therapy by reflecting the true tumor size changes, rather than the changes resulting from measurement variability. These findings require further investigation and validation with a larger cohort and a larger number of radiologists.

The limitations of the present study include a retrospective design and a relatively small number of patients. However, the CT tumor measurement records using RECIST 1.0 that we used in this study were prospectively performed during the clinical trial. We used these records to create a second set of CT tumor measurements meet RECIST 1.1 rather than reselecting and remeasuring the lesions. We used this approach to avoid differences of two RECIST measurements derived from inherent inconsistency and lack of reproducibility of CT tumor measurements [19]. In the present study, our main focus was the differences in CT

measurements resulting from changes in the RECIST 1.1 criteria. A study with independently performed RECIST 1.1 and RECIST 1.0 measurements in a larger study cohort is needed to further address the impact of RECIST 1.1 in the actual clinical trial and patient management. Although it is important to consider differences in survival using RECIST 1.1, this previously performed clinical trial used RECIST 1.0 for decision making, including when to stop the targeted therapy. Therefore, it is not possible to adequately compare survival for RECIST 1.1 and RECIST 1.0 using the data from this cohort. One would ideally need to compare survival in a prospective trial randomized by the two RECIST criteria (version 1.1 vs 1.0) for treatment decisions. Additional potential limitations include the exclusion of nearly half the original trial patients from this study because original measurements were not available and biases of lesion selection that might have occurred in the conversion to RECIST 1.1.

Assessment of tumor burden by CT tumor measurement using RECIST is an important feature of the clinical evaluation of cancer therapy and determines the end points of oncologic clinical trials. Radiologists, who interpret the oncology CT scans on daily basis, play a major role in CT tumor measurement by selecting, measuring, and reporting the lesions. It is important for radiologists to be aware of the changes in the new RECIST 1.1 guideline and its impact on CT tumor measurements as shown in the present study. First, a reduction in the maximum number of target lesions results in a decrease in the number of target lesions; thus, the time and efforts of radiologists are reduced when RECIST 1.1 is used. Second, the new definition of measurability of malignant lymph nodes contributes to reducing the number of target lesions and influences the eligibility for clinical trials in patients who have lymph nodes only as measurable lesions. Third, the inclusion of PET/CT in the detection of new lesions has a major effect on altering the best response assessment and time to progression.

In conclusion, RECIST 1.1 provided almost perfect agreement in the assessment of tumor response to therapy compared with RECIST 1.0, with decreased numbers of target lesions. Assessment with PET/CT was a major factor that influenced the difference in best response assessment between RECIST 1.1 and RECIST 1.0.

Acknowledgments

References

1. Therasse P, Arbuck SG, Eisenhauer EA, et al. New guidelines to evaluate the response to treatment in solid tumors (RECIST guidelines). J Natl Cancer Inst. 2000; 92:205–216. [PubMed: 10655437]

2. Eisenhauer EA, Therasse P, Bogaerts J, et al. New response evaluation criteria in solid tumors: revised RECIST guideline (version 1.1). Eur J Cancer. 2009; 45:228–247. [PubMed: 19097774]

3. Suzuki C, Jacobsson H, Hatschek T, et al. Radiologic measurements of tumor response to treatment: practical approaches and limitations. RadioGraphics. 2008; 28:329–344. [PubMed: 18349443]

4. Bogaerts J, Ford R, Sargent D, et al. Individual patient data analysis to assess modifications to the RECIST criteria. Eur J Cancer. 2009; 45:248–260. [PubMed: 19095437]

5. Moskowitz CS, Jia X, Schwartz LH, Gönen M. A simulation study to evaluate the impact of the number of lesions measured on response assessment. Eur J Cancer. 2009; 45:300–310. [PubMed: 19095439]

6. Schwartz LH, Bogaerts J, Ford R, et al. Evaluation of lymph nodes with RECIST 1.1. Eur J Cancer. 2009; 45:261–267. [PubMed: 19091550]

7. Jemal A, Siegel R, Ward E, et al. Cancer statistics, 2008. CA Cancer J Clin. 2008; 58:71–96. [PubMed: 18287387]

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. [PubMed: 15118073]

9. Paez JG, Janne PA, Lee JC, et al. EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. Science. 2004; 304:1497–1500. [PubMed: 15118125]

10. Pao W, Miller V, Zakowski M, et al. EGF receptor gene mutations are common in lung cancers from "never smokers" and are associated with sensitivity of tumors to gefitinib and erlotinib. Proc Natl Acad Sci USA. 2004; 101:13,306–13,311.

11. Pao W, Miller VA, Politi KA, et al. Acquired resistance of lung adenocarcinomas to gefitinib or erlotinib is associated with a second mutation in the EGFR kinase domain. PLoS Med. 2005; 2:e73. [PubMed: 15737014]

12. Engelman JA, Zejnullahu K, Mitsudomi T, et al. MET amplification leads to gefitinib resistance in lung cancer by activating ERBB3 signaling. Science. 2007; 316:1039–1043. [PubMed: 17463250]

13. Kobayashi S, Boggon TJ, Dayaram T, et al. EGFR mutation and resistance of non-small-cell lung cancer to gefitinib. N Engl J Med. 2005; 352:786–792. [PubMed: 15728811]

14. Jackman DM, Yeap BY, Lindeman NI, et al. Phase II clinical trial of chemotherapy-naive patients > or = 70 years of age treated with erlotinib for advanced non-small-cell lung cancer. J Clin Oncol. 2007; 25:760–766. [PubMed: 17228019]

15. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. J Am Stat Assoc. 1958; 53:457–481.

16. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet. 1986; 1:307–310. [PubMed: 2868172]

17. Zhao B, James LP, Moskowitz CS, et al. Evaluating variability in tumor measurements from same-day repeat CT scans of patients with non-small cell lung cancer. Radiology. 2009; 252:263–272. [PubMed: 19561260]

18. Kundel HL, Polansky M. Measurement of observer agreement. Radiology. 2003; 228:303–308. [PubMed: 12819342]

19. Erasmus JJ, Gladish GW, Broemeling L, et al. Interobserver and intraobserver variability in measurement of non-small-cell carcinoma lung lesions: implications for assessment of tumor response. J Clin Oncol. 2003; 21:2574–2582. [PubMed: 12829678]
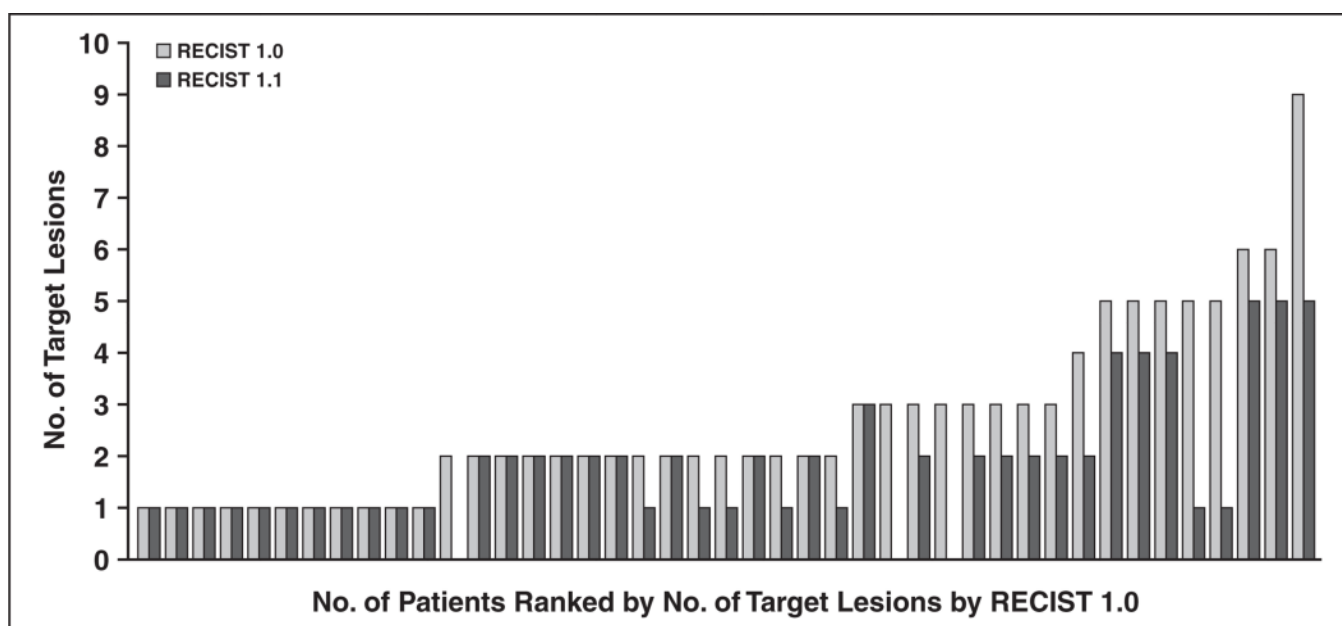
Fig. 1.
Number of target lesions according to Response Evaluation Criteria in Solid Tumors (RECIST) 1.1 versus that according to RECIST 1.0. Number of target lesions calculated using RECIST 1.1 was significantly lower than that calculated using RECIST 1.0 ($p < 0.0001$, paired Student's t test).
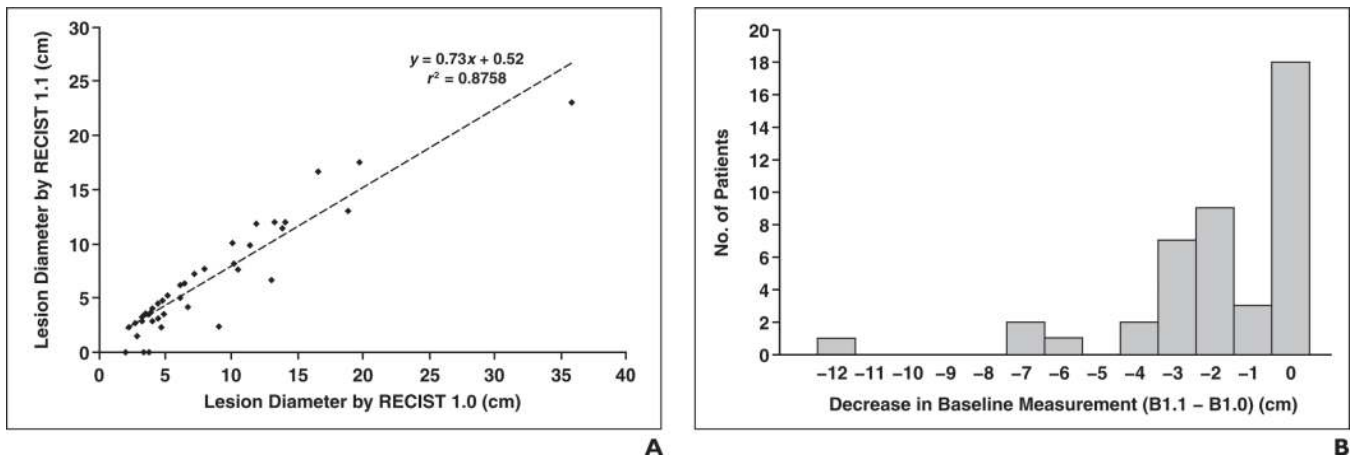
Fig. 2.
Comparison of diameters of target lesions by Response Evaluation Criteria in Solid Tumors (RECIST) 1.1 and 1.0.
**A**, Comparison of sum of longest diameters of target lesions on baseline CT measurements by RECIST 1.1 and RECIST 1.0 showed high correlation between two CT measurements ($r = 0.936$ and $r^2 = 0.8758$, Pearson's correlation).
B, Decrease in sum of longest diameters of target lesions on baseline CT measurements by RECIST 1.1 (B1.1) compared with baseline CT measurement RECIST 1.0 (B1.0) ($p = 0.0001$, paired Student's t test).
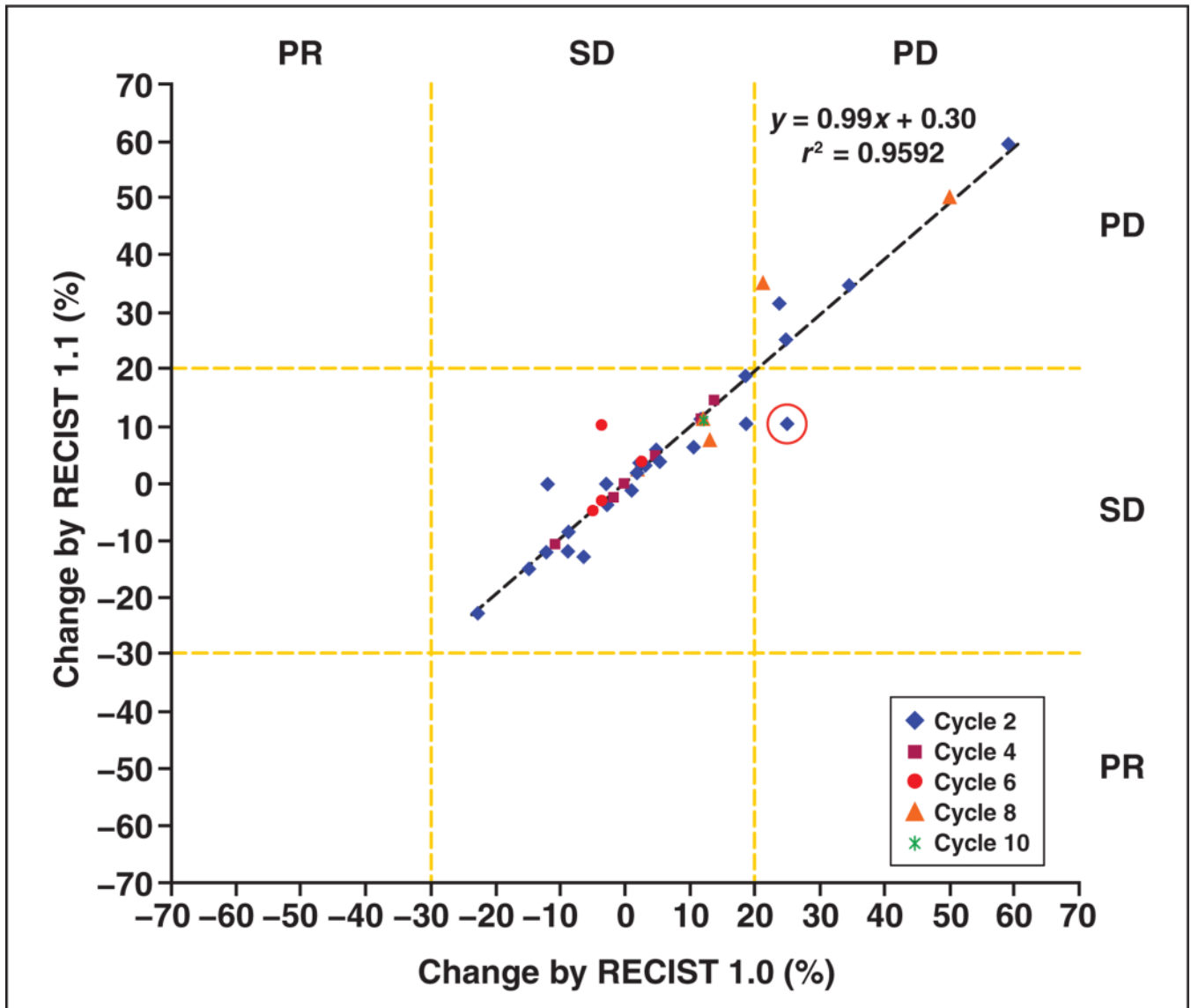
Fig. 3.
Percentage changes in sum of long diameter CT measurements by Response Evaluation Criteria in Solid Tumors (RECIST) 1.1 versus RECIST 1.0 at every two cycles of therapy. High correlation in percentage changes and high concordance of response assessments were observed between CT measurements by RECIST 1.1 and by RECIST 1.0. ($r = 0.986$ and $r^2 = 0.972$ at cycle 2 [$n = 28$]; $r = 0.998$ and $r^2 = 0.996$ at cycle 4 [$n = 7$]; $r = 0.657$ and $r^2 = 0.432$ at cycle 6 [$n = 5$]; $r = 0.949$ and $r^2 = 0.901$ at cycle 8 [$n = 4$]; $r =$ not applicable and $r^2 =$ not applicable at cycle 10 [$n = 1$]; and $r = 0.979$ and $r^2 = 0.959$ at cycles 2–10 [$n = 45$], Pearson's correlation). PD = progressive disease, PR = partial response, SD = stable disease.
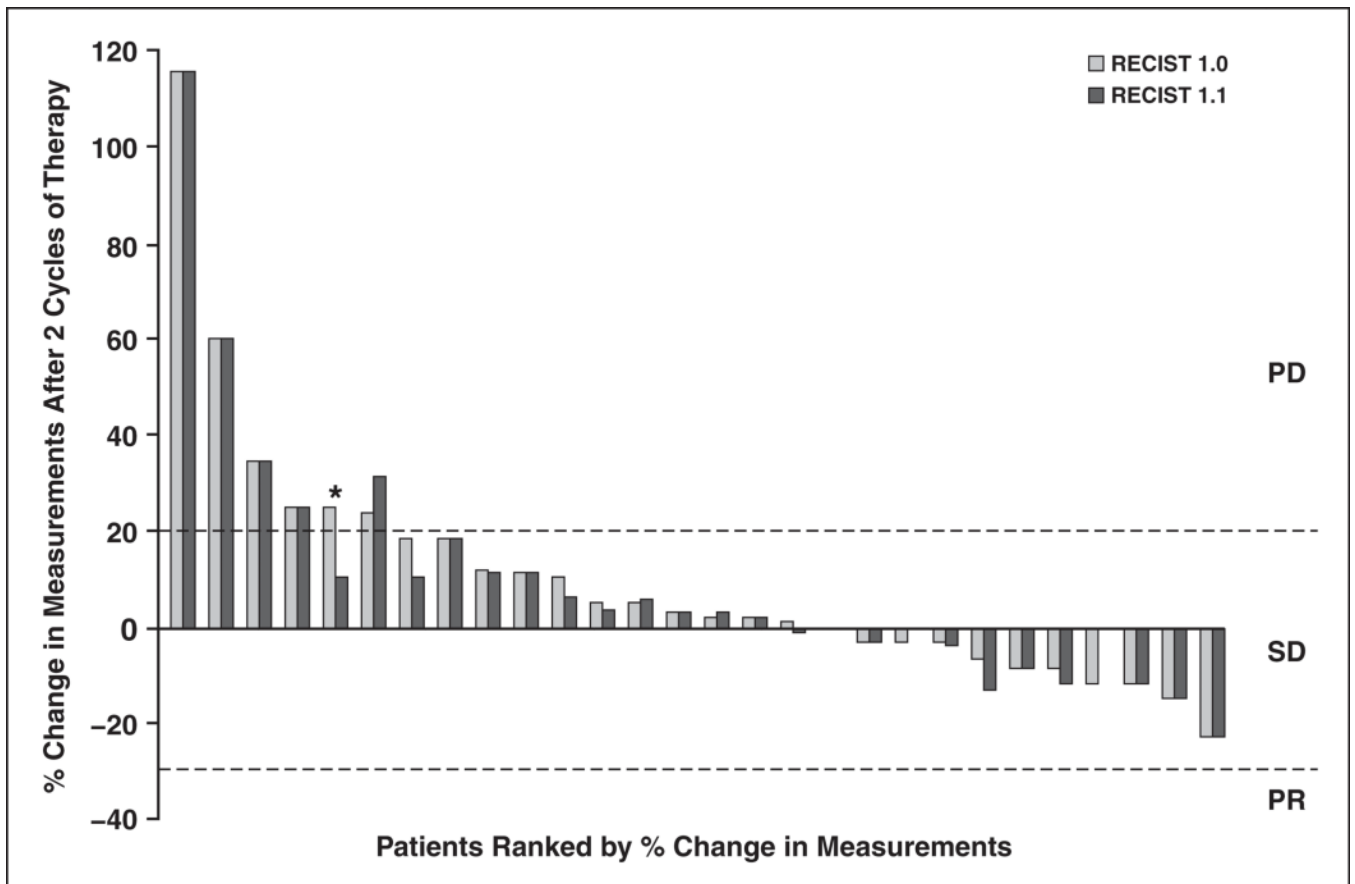
Fig. 4.
Percentage changes in sum of longest diameter CT measurements by Response Evaluation Criteria in Solid Tumors (RECIST) 1.1 and RECIST 1.0 after first two cycles of therapy. Response assessments of target lesions at first two cycles of therapy by RECIST 1.1 and RECIST 1.0 were concordant in 27 (96%) of 28 patients and discordant in one patient (4%; *asterisk*). PD = progressive disease, PR = partial response, and SD = stable disease.
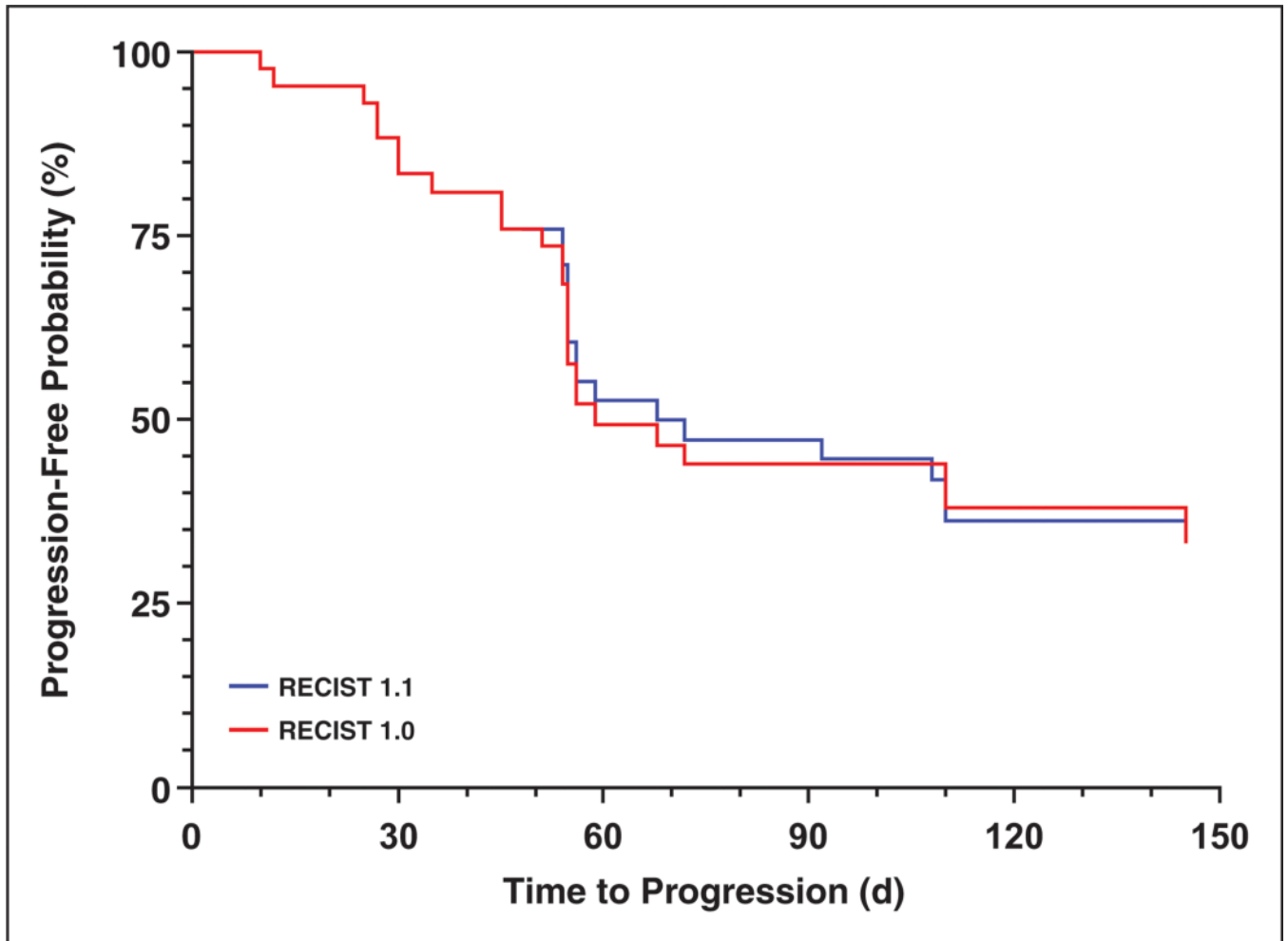
Fig. 5.
Time to progression by Response Evaluation Criteria in Solid Tumors (RECIST) 1.1 versus
RECIST 1.0. Difference in time to progression was not significant (p = 1.000, sign test).
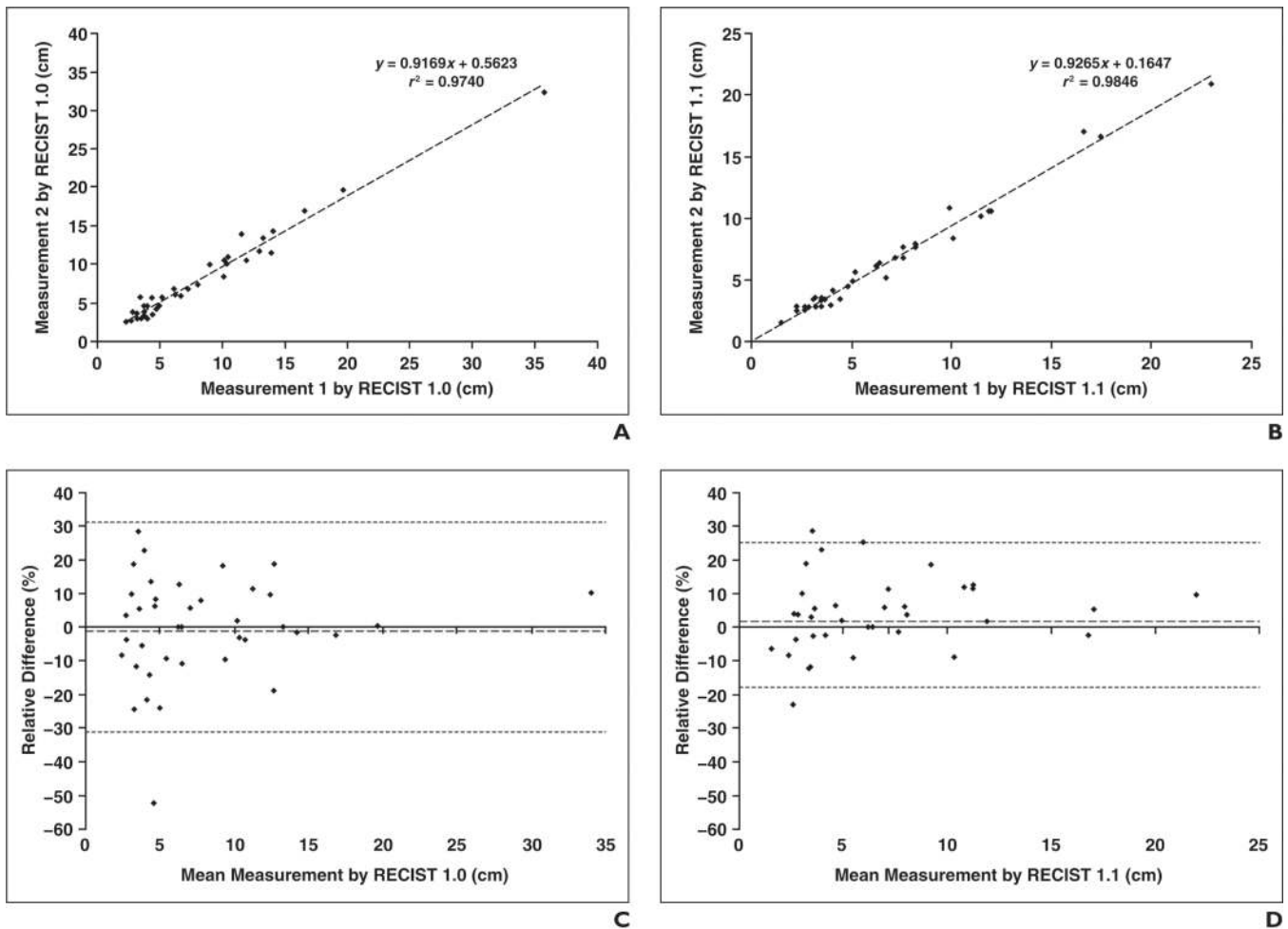
Fig. 6.
Comparison of sum of longest diameters of target lesions on baseline CT measurements performed in clinical trial (measurement 1) and by observer in present study (measurement 2).

A and **B**, High correlation was noted between two measurements for both Response Evaluation Criteria in Solid Tumors (RECIST) 1.0 (**A**) and RECIST 1.1 (**B**) ($r^2 = 0.9740$ and $r^2 = 0.9846$, respectively; Pearson's correlation).

C and **D**, Bland-Altman plots of two sets of baseline CT measurements for RECIST 1.0 (**C**) and RECIST 1.1 (**D**) are shown. Interobserver variability is shown as function of mean of two measurements. Dotted lines indicate upper and lower 95% limits of agreement. Dashed lines indicate mean relative difference.

TABLE 1

Summary of Major Changes in Response Evaluation Criteria in Solid Tumors (RECIST) 1.1 Compared With RECIST 1.0

RECIST Guideline	RECIST 1.1	RECIST 1.0
No. of target lesions	Up to 2 per organ; up to 5 in total	Up to 5 per organ; up to 10 in total
Assessment of lymph nodes	Short-axis measurements should be used and recorded; ≥ 15 mm, target lesions; ≥ 10 mm but < 15 mm, nontarget lesions; < 10 mm, nonpathological	No clear guideline provided
Clarification of disease progression	20% increase in the sum of target lesions and 5-mm absolute increase are required	20% increase in the sum of target lesions (no minimum absolute size increase is required)
FDG PET scan	Included only in the detection of new lesions[a]	Not included

[a]New lesions on the basis of FDG PET can be identified according to the following algorithm [2]: a negative FDG PET at baseline with a positive FDG PET at follow-up is a sign of progressive disease based on a new lesion. For no FDG PET at baseline and a positive FDG PET at follow-up, if the positive FDG PET at follow-up corresponds to a new site of disease confirmed by CT, this is progressive disease. If the positive FDG PET at follow-up is not confirmed as a new site of disease on CT, additional follow-up CT scans are needed to determine whether there is truly progression occurring at that site. If the positive FDG PET at follow-up corresponds to a preexisting site of disease on CT that is not progressing on the basis of the anatomic images, this is not progressive disease.

TABLE 2

Best Response Assessment by Response Evaluation Criteria in Solid Tumors (RECIST) 1.1 Versus RECIST 1.0

Best Response by RECIST 1.0	Best Response by RECIST 1.1		
	Progressive Disease	Stable Disease	Nonevaluable
Progressive disease	17	1	0
Stable disease	2	17	0
Nonevaluable	0	0	6

Note—Data are number of patients. Weighted κ = 0.905.