Jpn J Clin Oncol 2003;33(10)533–537

Measuring Response in Solid Tumors: Comparison of RECIST and WHO Response Criteria

Joon Oh Park[1], Soon Il Lee[1], Seo Young Song[1], Kihyun Kim[1], Won Seog Kim[1], Chul Won Jung[1], Young Suk Park[1], Young-Hyuk Im[1], Won Ki Kang[1], Mark Hong Lee[1], Kyung Soo Lee[2] and Keunchil Park[1]

[1]Division of Hematology/Oncology, Department of Medicine and [2]Department of Radiology, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Korea

Received June 22, 2003; accepted September 1, 2003

Background: Objective tumor response is a common endpoint in daily practice as well as in clinical trials to evaluate the efficacy of anti-cancer agents. Traditionally, the standard World Health Organization (WHO) criteria has been adopted in these contexts. However, the recent development of new classes of anti-cancer agents and progress in imaging technology have required new methodology to evaluate response to treatment. Recently, the Response Evaluation Criteria in Solid Tumors Group (RECIST) proposed new guidelines using unidimensional measurement. Theoretically, the simple sum of the maximum diameters of individual tumors is more linearly related to cell kill than is the sum of the bidimensional products. To validate these new guidelines, we have compared the standard WHO response criteria with the new RECIST guidelines in the same patient population.

Methods: Data from 79 patients enrolled in eight prospective phase II studies at Samsung Medical Center were retrospectively re-analyzed to determine the concordance between the two response criteria. The two response criteria were applied separately, and the results were compared using the κ statistic to test concordance for overall response rate.

Results: The overall response rate according to the WHO criteria was 31.6%. Using the RECIST criteria, nine patients were reclassified and the overall response rate was 30.4%. There was excellent agreement between the unidimensional and bidimensional criteria in 23 of 25 responses (92%). The κ statistic for concordance for overall response was 0.91.

Conclusions: We conclude that the new RECIST guidelines are comparable to the old response criteria in evaluating response in solid tumors. Moreover, the new guidelines are just as simple and reproducible in the measurement of response in daily practice as they are in clinical trials.

Key words: tumor response – RECIST – unidimensional measurement

INTRODUCTION

The activity of anti-cancer drugs is evaluated by measuring changes in tumor size in response to treatment (1). Tumor size has traditionally been estimated from bidimensional measurements (the product of the longest diameter and its longest perpendicular diameter for each tumor) (2). In the early 1980s, the World Health Organization (WHO) developed recommendations in an attempt to standardize criteria for response assessment, and the WHO response criteria were adopted as the standard method for evaluating tumor response (3).

For reprints and all correspondence: Keunchil Park, Division of Hematology/Oncology, Department of Medicine, Samsung Medical Center, Sungkyunkwan University School of Medicine, 50 Ilwon-dong Kangnam-ku, Seoul 135-710, Korea. E-mail: kpark@smc.samsung.co.kr

However, measuring in two dimensions and then calculating their products and their sums is laborious and has the risk of error. In theory, the changes in diameter relate more closely to the fixed proportion of cells killed by a standard dose of chemotherapy than do changes in the bidimensional product (1,2). Since the measuring methods and selection of target lesions were not clearly described in the WHO guidelines, assessment of tumor response is shown to be poorly reproducible between one investigator, or group of investigators, and another (1).

Furthermore, the arrival of new imaging technologies and recent progress in the development of new classes of anti-cancer agents required the establishment of a new methodology, and has led to a number of different modifications of WHO criteria (1,4).

In 1998, new response evaluation criteria in solid tumors (RECIST) were proposed by the RECIST working group in

Table 1. Studies analyzed using the WHO and RECIST criteria

Tumor type	No. of assessable patients	Study type	Treatment	Endpoints
Gastric cancer	10	Phase II	Docetaxel/CDDP	Response rate
Gastric cancer	19	Phase II	Oxaliplatin/5FU/LV	Response rate
Breast cancer	8	Phase II	Docetaxel/Doxorubicin	Response rate
Breast cancer	11	Phase II	Genexol^{R} (Paclitaxel)	Response rate
Non-small cell lung cancer	11	Phase II	Docetaxel	Response rate
Non-small cell lung cancer	5	Phase II	Genexol^{R} (Paclitaxel)/CDDP	Response rate
Non-small cell lung cancer	12	Phase II	Gemcitabine/CDDP	Response rate
Hepatocellular carcinoma	3	Phase II	Gemcitabine	Response rate

order to minimize the risk of measurement error and prevent overestimation of response rates (5). Several research groups were interested in determining whether an approach based only on measurements in one dimension was theoretically valid and practically feasible (1,5).

Here, we attempt to clarify the significance and problems with the RECIST criteria in a retrospective comparison of the assessment results obtained using the WHO and RECIST criteria in the same patient population.

PATIENTS AND METHODS

During the period 1996–2001, our center conducted eight phase II studies for several types of cancer as part of multicenter trials. These were all pharmaceutical company-sponsored studies for the approval of new treatments in Korea. The study details are shown in Table 1. A total of 79 patients with measurable and evaluable lesions according to WHO criteria had been enrolled in these studies. All patients in these trials had at least one bidimensionally measurable lesion greater than or equal to 2×1 cm in size measured by computed tomography (CT) scan. All CT scans were performed on a helical CT scanner (HiSpeed Advantage scanner; General Electric Medical Systems, Milwaukee, WI, USA) with intravenous administration of non-ionic contrast materials, and slice collimation was 7 mm in all CT scans used in this study. The CT image data were reconstructed with 7-mm thickness and were directly displayed on monitors of picture archiving and communications system (PACS) (General Electric Medical Systems Integrated Imaging Solutions, Mt Prospect, IL, USA). Tumor measurements were performed with electronic calipers at baseline and at regular intervals during the trials, and were recorded on study-specific case report forms. Disease had been documented in all patients by radiological evaluation (CT scan). We reviewed each patient's medical records and case report forms, and we re-confirmed tumor measurements from original imaging studies using the WHO and RECIST criteria respectively as previously described (Table 2) (3,5). One medical oncologist and one radiologist retrospectively reviewed all CT images together and decisions on the tumor measurements were reached by consensus. The results were compared using the κ statistic to test concordance for overall response rate.

RESULTS

As previously mentioned, all patients had at least one measurable lesion and were therefore eligible according to WHO and RECIST criteria. Results of the comparison between standard WHO and new RECIST criteria are shown in Tables 3 and 4. There were 25 of 79 (31.6%) partial responses (PR) to the WHO criteria and 24 of 79 (30.4%) to the RECIST criteria. Only one of 79 (1.3%) patients had PR according to the WHO criteria but not according to the RECIST criteria, whilst two of 79 (2.5%) patients were assessed as PR according to the

Table 2. WHO and RECIST criteria for tumor response

	WHO	RECIST
Measurability	Measurable, bidimensional	Measurable, unidimensional: Conventional method ≥20 mm; Spiral CT ≥10 mm; Target versus non-target lesion
	Non-measurable/evaluable	Non-measurable
Objective response		
Complete response (CR)	Disappearance of all known lesion(s); confirmed at 4 weeks	Disappearance of all known lesion(s); confirmed at 4 weeks
Partial response (PR)	At least 50% decrease; confirmed at 4 weeks	At least 30% decrease; confirmed at 4 weeks
Stable disease (SD)	Neither PR nor PD criteria met	Neither PR nor PD criteria met
Progressive disease (PD)	25% increase; no CR, PR or SD documented before increased disease, or new lesion(s)	20% increase; no CR, PR, or SD documented before increased disease, or new lesion(s)

Table 3. Comparison of the WHO and new RECIST criteria applied to the same patients

Trial and criteria	No. patients				Response rate (%)	Progression rate (%)
	CR	PR	SD	PD		
Gastric cancer (Docetaxel/CDDP, *n* = 10)						
WHO	0	0	3	7	0	70.0
RECIST	0	0	5	5	0	50.0
Gastric cancer (Oxaliplatin/5FU/LV, *n* = 19)						
WHO	0	4	7	8	21.1	42.1
RECIST	0	4	9	6	21.1	31.6
Breast cancer (Docetaxel/Doxorubicin, *n* = 8)						
WHO	0	5	1	2	62.5	25.0
RECIST	0	5	1	2	62.5	25.0
Breast cancer (Genexol^{R}, *n* = 11)						
WHO	0	4	4	3	36.4	27.3
RECIST	0	4	5	2	36.4	18.2
Non-small cell lung cancer (Docetaxel, *n* = 11)						
WHO	0	3	2	6	27.3	54.5
RECIST	0	3	2	6	27.3	54.5
Non-small cell lung cancer (Genexol^{R}/CDDP, *n* = 5)						
WHO	0	3	2	0	60.0	0
RECIST	0	3	2	0	60.0	0
Non-small cell lung cancer (Gemcitabine/CDDP, *n* = 12)						
WHO	0	6	4	2	50.0	16.7
RECIST	0	5	6	1	41.7	8.3
Hepatocellular carcinoma (Gemcitabine, *n* = 3)						
WHO	0	0	1	2	0	66.7
RECIST	0	0	1	2	0	66.7
Total (*n* = 79)						
WHO	0	25	24	30	31.6	38.0
RECIST	0	24	31	24	30.4	30.4

unidimensional criteria but not according to the bidimensional criteria. The overall response rate according to the WHO criteria was 31.6%. Using the RECIST criteria, the overall response rate was 30.4%, with nine patients being reclassified (Fig. 1); six were switched from the PD (progressive disease) to the SD (stable disease) group, one from the SD to the PR group and two from the PR to the SD group. The number of SD group patients was higher, and the number in the PD group was lower in those assessed using the RECIST criteria.

Concordance for overall response rate judged according to the two criteria was tested using the κ statistic. There was excellent agreement between the unidimensional and bidimensional criteria in 23 of 25 responses (92%). The κ statistic for concordance in overall response was 0.91. However, when the responses were subdivided into PR, SD and PD, the concordance for the response rate was 0.83. This phenomenon was the result of the largest re-categorization of patients from PD to SD.

DISCUSSION

Evaluating the efficacy of anti-cancer treatment is important for medical decisions in practice as well as in clinical trials (1). The methodology used to evaluate the response to treatment has evolved substantially over the past decades (1,5). In 1981, Miller et al. published 'Reporting results of cancer treatment', based on the WHO recommendations, in order to standardize the response assessment (3). Such standardization of guide-

Table 4. Overall concordance between WHO and RECIST criteria in the assessment of overall response rate

RECIST criteria	WHO criteria		Total
	Response	Non-response	
Response	23	1	24
Non-response	2	53	55
Total	25	54	79

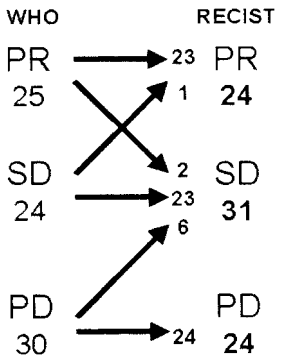

Figure 1. Changes in response after the WHO and RECIST criteria were separately applied to the same patients.

lines for tumor response evaluation has helped greatly in the methodology for screening new drugs, as well as in the comparison of drug efficacy in randomized trials (1). However, most investigators and research groups have faced some problems using the WHO criteria due to their complexity. Since no precise indication was given as regards to selection of target lesions, a great deal of time-consuming effort has been made repeatedly to assess the bidimensional measurements of the majority of clinically and radiologically detected abnormalities (3). These issues have led to a number of different modifications of the WHO criteria, resulting in a situation where response criteria are no longer comparable between research groups (5). To clarify and simplify the tumor response assessment rules, the RECIST group suggested and published new criteria where bidimensional measurement and exhaustive measurement of all lesions was abandoned (1,5). However, the previous recommendations have been followed for almost 20 years in thousands of cancer trials and still appear as reference criteria for any historical comparison. Thus, the recently proposed RECIST criteria raise the question of whether a simple unidimensional measurement (the sum of the longest diameter of the tumors) is equivalent to the more complicated bidimensional measurements with regard to tumor response. For this reason, several researchers were interested in determining whether the approach based on the RECIST criteria was theoretically valid and practically feasible. In 1999, James et al. (2) proposed the use of unidimensional measurement of tumor lesions for the first time. In this work, retrospective analysis of 14 different studies demonstrated that bidimensional or unidimensional measurement of tumor lesions did not change the response rate in each individual study (2). Recently, several types of solid tumor including lung, breast, colon and gastric cancer were reassessed to validate the RECIST criteria, and this analysis concluded that the WHO and RECIST criteria were equivalent in terms of response rate (6–10). Furthermore, it was reported that the unidimensional measurement correlated well with the three-dimensional volume measurement by helical CT (11).

In our retrospective analysis, there was excellent agreement between the unidimensional and bidimensional criteria, with a concordance rate of 0.91 using the κ statistic. When re-analyzed using the RECIST criteria, the overall response rate decreased from 31.6 to 30.4% in our study. The application of the RECIST criteria resulted in the reclassification of nine patients (11.4%) from the PD to SD group, from the SD to PR group and from the PR to SD group. The largest re-categorization of patients was from PD to SD, which would have an impact on a time to progression. However, this could not be assessed because patients were taken off study or treatment when they met the WHO criteria. Thus, this should be evaluated in a new prospective study of validating the RECIST criteria. As already described by Therasse et al. (5) in their retrospective comparison of six trials involving 795 patients, the switch from WHO to RECIST criteria resulted in a more favorable classification. In general, small shifts in progression rate are unlikely to have an impact on results of phase II trials (11). However, to evaluate the efficacy of new classes of non-cytotoxic agents that are unlikely to produce tumor shrinkage, investigators must consider this phenomenon (1,12). Thus, variations in assessment of efficacy resulting from this phenomenon should be considered in various clinical situations (10). Erasmus et al. (13) also noted that interobserver variability of measurements is greater than intraobserver variability; measurement differences are greatest when the edge of the lesion is irregular or spiculated and differences are smallest when the edge is well defined. Because these differences could lead to an incorrect interpretation of tumor growth or response and impact on results of clinical trial, the methodology for tumor measurements needs to be improved and clearly defined.

In conclusion, our work provides evidence for the accuracy and usefulness of the RECIST criteria and validates the use of unidimensional measurement. Although the new RECIST guidelines were valid for evaluating tumor response, the assessment of these criteria should be continued, and additional response parameters should be sought in order to overcome some of the limitations of the RECIST criteria. Because the sample size for this analysis is small and small measure error might have a larger impact on the outcome, this study would be strengthened by prospective validation of the RECIST criteria in new phase II or phase III trials.

References

1. Therasse P. Measuring the clinical response. What does it mean? *Eur J Cancer* 2002;38:1817–23.
2. James K, Eisenhauer E, Christian M, Terenziani M, Vena D, Muldal A, et al. Measuring response in solid tumors: unidimensional versus bidimensional measurement. *J Natl Cancer Inst* 1999;91:523–8.
3. Miller AB, Hoogstraten B, Staquet M, Winkler A. Reporting results of cancer treatment. *Cancer* 1981;47:207–14.
4. Padhani AR, Husband JE. Are current tumour response criteria relevant for the 21st century? *Br J Radiol* 2000;73:1031–3.
5. 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.
6. Yoshida S, Miyata Y, Ohtsu A, Boku N, Shirao K, Shimada Y. Significance of and problems in adopting response evaluation criteria in solid tumor RECIST for assessing anti-cancer effects of advanced gastric cancer. *Gastric Cancer* 2000;3:128–33.

7. Werner-Wasik M, Xiao Y, Pequignot E, Curran WJ, Hauck W. Assessment of lung cancer response after non-operative therapy: tumor diameter, bidimensional product, and volume. A serial CT scan-based study. *Int J Radiat Oncol Biol Phys* 2001;51:56–61.
8. Kimura M, Tominaga T. Outstanding problems with response evaluation criteria in solid tumors (RECIST) in breast cancer. *Breast Cancer* 2002;9: 153–9.
9. Trillet-Lenoir V, Freyer G, Kaemmerlen P, Fond A, Pellet O, Lombard-Bohas C, et al. Assessment of tumour response to chemotherapy for metastatic colorectal cancer: accuracy of the RECIST criteria. *Br J Radiol* 2002;75:903–8.
10. Sohaib SA, Turner B, Hanson JA, Farquharson M, Oliver RT, Reznek RH. CT assessment of tumor response to treatment: comparison of linear, cross-sectional and volumetric measures of tumour size. *Br J Radiol* 2000; 73:1178–84.
11. Gehan EA, Tefft MC. Will there be resistance to the RECIST (Response Evaluation Criteria in Solid Tumors)? *J Natl Cancer Inst* 2000;92:179–81.
12. Padhani AR, Ollivier L. The RECIST (Response Evaluation Criteria in Solid Tumors) criteria: implications for diagnostic radiologists. *Br J Radiol* 2001;74:983–6.
13. Erasmus JJ, Gladish GW, Broemeling L, Sabloff BS, Truong MT, Herbst RS, 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;21:2574–82.