



Original article

Significance of and problems in adopting response evaluation criteria in solid tumor RECIST for assessing anticancer effects of advanced gastric cancer

SHIGEAKI YOSHIDA¹, YOSHINORI MIYATA², ATSUSHI OHTSU¹, NARIKAZU BOKU¹, KUNIAKI SHIRAO³, YASUHIRO SHIMADA³, and THE JAPAN CLINICAL ONCOLOGY GROUP

¹Gastrointestinal Oncology Division, National Cancer Center Hospital East, 6-5-1 Kashiwanoha Kashiwa, Chiba 277-8577, Japan

²Gastrointestinal Oncology Division, National Cancer Center Hospital, Tokyo, Japan

³Department of Medicine, Ibaraki Prefectural Hospital Cancer Center, Ibaraki, Japan

Abstract

Background. The response evaluation criteria in solid tumor (RECIST) exclude the use of barium meal studies. This will deeply affect the Japanese criteria for evaluating the response in the primary lesion of gastric cancer.

Methods. Of 280 patients with gastric cancer enrolled in a Japan Clinical Oncology Group (JCOG) phase III study, 255 had been assessed for response by the WHO and/or Japanese criteria. We selected these 255 patients as our subjects and reassessed their response outcomes by RECIST.

Results. Of the 255 patients, 32 (13%) had no evaluable lesion other than the primary site, and 171 (67%) had some measurable lesion defined by the WHO criteria. Because the lesions in 129 of these 171 patients were 20 mm or more in size, only 51% of the 255 subjects were eligible for assessment of the target lesion by RECIST. In 162 of the 171 patients who had a lesion of 10 mm or more, response rates by the old and new criteria were nearly equal, in spite of the different methods of measurement. The response rate in the primary lesions assessed by the Japanese criteria was lowest among all the groups examined.

Conclusion. The RECIST is simple and good for clinical practice. Nevertheless, the recommendation of 20 mm or more for target lesions, which may restrict the number of eligible patients, could be replaced by a size of 10 mm or more. Because the Japanese evaluation criteria are rigid and do not inflate the response rate, they can be used, as additional criteria for assessing the quality of response, when the RECIST is used.

Key words Gastric cancer · Response outcomes · RECIST · WHO criteria · Japanese criteria

Introduction

For evaluating anticancer effects in solid tumors, the WHO criteria (1979) [1] have been widely used

throughout the world, whereas for evaluating response in the primary lesion of gastric cancer, most Japanese studies have adopted the evaluation criteria proposed by the Japanese Research Society for Gastric Cancer in 1985 [2]. The Japanese criteria are based on the rate of reduction of the size of the primary lesion, determined by bidimensional measurement on roentgenograms taken before and after treatment, in regard to well demarcated lesions. According to them, a complete response (CR) is defined as disappearance of tumor for 4 weeks or more, a partial response (PR) as disappearance of tumor for less than 4 weeks, or a reduction of 50% or more for 4 weeks or more. Progression of disease (PD) is defined as an increase of tumor size of more than 25%, and no change (NC) as conditions other than CR, PR, and PD. For patients who have diffuse lesions such as scirrhous cancer, PR is defined as an increase of 50% or more in the square measure in the affected area, measured on a barium-filled photograph. For those lesions in which the size is difficult to measure, the Japanese criteria refer to endoscopic evaluation, in which a changing tumorous-to-flat appearance for 4 weeks or more is assessed as PR and tumor disappearance for 4 weeks or more as CR.

In 1998, new response evaluation criteria in solid tumor (RECIST) were proposed by the RECIST working group in order to minimize the risk of measurement errors and prevent overestimation of response rates. These criteria were published in 2000 as *New guidelines to evaluate the response to treatment in solid tumors* [3]. In these guidelines, evaluation with computed axial tomography (CAT) scans or magnetic resonance imaging (MRI), is important because of the objectivity of these methods, and the use of ultrasonography is excluded because of its operator dependency. Although not stated, barium meal studies should be excluded for the same reason. Instead of radiography, the guidelines recommend the use of endoscopy with biopsy for assessing response outcomes in the digestive tract; however, with

Offprint requests to: S. Yoshida

Received: April 25, 2000 / Accepted: August 22, 2000

this method, the primary lesions are defined as non-measurable. The adoption of the new guidelines, therefore, will affect our daily clinical practice in evaluating response outcomes, particularly in Japan, where the incidence of gastric cancer is very much higher than that in the rest of the world.

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

Subjects and methods

During the period between 1992 and 1998, the Japan Clinical Oncology Group (JCOG) conducted a phase III trial for advanced gastric cancer treated with either 5-fluorouracil (5-FU) alone, uracil and futrafur (UFT) combined with mitomycin C (MMC), and 5-FU combined with cis-dichlorodiammine platinum (II) (CDDP) [4]. In this trial, 280 patients having measurable or evaluable lesions had been enrolled, but the response outcomes in 25 of the 280 patients (9%) could not be assessed because of early progression or drop-out from the evaluation. We selected the remaining 255 patients whose response outcomes had been evaluated by the WHO and/or Japanese criteria as the subjects of this study, and reassessed their response outcomes by RECIST. We compared the results obtained with those obtained from the original study, and examined the utility, influence on the response rate, and survival of the patients obtained in adopting the new evaluation criteria.

In the RECIST, it is recommended that the term “measurable lesions” refer to lesions 20 mm or more in size measured by conventional image diagnosis such as plain X-rays, CAT scans, and MRI, in order to rule out operator dependency. Accordingly, lesions less than 20 mm are all defined as non-measurable. All measurable lesions, up to a maximum of five lesions per organ, and ten lesions in total, representative of all involved organs, are identified as target lesions, and those evaluable lesions, other than the target lesions, as nontarget lesions. The response rate is calculated not with two-dimensional, but with unidimensional measurements, with the baseline sum longest diameter, i.e., the sum of the longest diameters, used for all target lesions. PR is defined as a decrease of 30% or more and PD as an increase of 20% or more in the diameter of the lesion. CR is defined as disappearance of tumor, as in the past, and stable disease (SD) as neither CR, PR, nor PD. Although the WHO criteria required a response duration of 4 weeks or more in PR and CR, the RECIST does not require any such duration, separately defining PR and CR with a duration of 4 weeks or more

Table 1. Target population of patients assessed by adopting the conventional WHO and Japanese criteria

	Number of patients	Percentage
Evaluable lesions only	52	20
No evaluable lesion other than primary site	32	13
Measurable lesions	171	67
Overall number of patients	255	100

as confirmed CR (cCR) and confirmed PR (cPR). In nontarget lesions, PR is not included in the definitions, and PD is defined as the appearance of one or more new lesions and/or unequivocal progression of existing nontarget lesions. Accordingly, the assessment of response outcomes in nontarget lesions is classified into three categories: CR, non-CR/non-PD, and PD.

Results

Patient eligibility by RECIST

Table 1 shows the characteristics of the patients. Of the 255 patients, 52 (20%) had only non-measurable lesions, and 32 (13%) had a neither measurable nor an evaluable lesion other than the primary site. Of the remaining 171 patients (67%), who had some measurable metastatic lesion according to the WHO criteria, 129 had lesions of 20 mm or more, 12 had lesions of 15–19 mm, 21 had lesions of 10–14 mm, and the remaining 9 had lesions of 9 mm or less (Table 2). In other words, 141 patients had lesions of 15 mm or more, and 162 had lesions of 10 mm or more. Because the RECIST recommend selecting only lesions over 20 mm as target lesions, only 51% (129/255) of these subjects are eligible for consideration with the RECIST.

Response outcomes

Table 3 shows the response outcomes by size of metastatic lesions, as assessed by the WHO criteria and RECIST. Because the definitions of response are very different with the two criteria, we basically adopted RECIST in the information shown in Table 3. Therefore, “CR” and “PR” in Table 3 indicate only the grade of tumor reduction corresponding to CR and PR, respectively, for those assessed by the conventional criteria (with no consideration given to response duration). In brackets in Table 3 we note CR and PR for the conventional criteria, with confirmed CR and PR in the RECIST noted as “cCR” and “cPR”, respectively. As to the response rate, we have shown the results calculated

Table 2. Size distribution of the 171 measurable lesions assessed by adopting the conventional WHO and/or Japanese criteria

Size	Number of cases	Cumulative number of cases	
20mm or more	129	20mm or more	129 (51)
15–19mm	12	15mm or more	141 (55)
10–14mm	21	10mm or more	162 (64)
9mm or less	9	All measurable lesions	171 (67)
Total	171		

Numbers in parentheses indicate percentages of total number of patients ($n = 255$)

Table 3. Response outcomes by size in measurable lesions assessed by adopting the WHO criteria and RECIST

Size of lesion	Response outcomes				Response rate (RR) (%)	
	CR [cCR]	PR [cPR]	NC	PD	RR ^a	cRR ^a
20mm or more ($n = 129$)	1 [1] 1 [1]	34 [25] 31 [24]	70 76	24 21	27 25	20 19
15–19mm ($n = 12$)	0 0	5 [3] 5 [2]	3 6	4 1	41 41	25 17
10–14mm ($n = 21$)	1 [1] 1 [1]	6 [6] 7 [7]	10 10	4 3	33 38	33 38
9mm or less ($n = 9$)	0 0	3 [3] 5 [4]	6 4	0 0	33 56	33 44

See text for definitions of CR, cCR, PR, cPR, NC, and PD

Upper columns indicate results with WHO criteria, and lower columns results with response evaluation criteria in solid tumor (RECIST)

^aRR: CR + PR/total; cRR = cCR + cPR/total

Table 4. Response outcomes in patients who had no measurable lesions other than the primary site and in the patients overall assessed by adopting the Japanese and WHO evaluation criteria

Lesions	Response outcomes				Response rate (%)	
	CR [cCR]	PR [cPR]	NC	PD	RR ^a	cRR ^a
Primary site only ^b ($n = 32$)	0 [0]	4 [4]	17	11	13	13
Number of subjects overall ($n = 255$)	2 [2]	64 [49]	124	65	26	19

Upper columns indicate results with WHO criteria, and lower columns, results with RECIST

^aRR = CR+PR/total; cRR = cCR+cPR/total

^bThose having neither measurable nor evaluable lesions other than the primary site

from the CR+PR patients as “RR”, and those from the cCR +cPR patients as “cRR”.

As shown in Table 3, the response rates assessed by the two different evaluation criteria were almost equal for lesions of 20mm or more in regard to both RR and cRR, though the RR was higher than the cRR. In contrast, the number of NC (SD) patients was higher, and that of PD patients was lower in those evaluated by RECIST than in those evaluated by the conventional criteria. In spite of the limited number of patients, similar results were also obtained for lesions of 15–19mm

and those of 10–14mm, but in lesions of 9mm or less, there was a discrepancy between the response rates evaluated by the old and new criteria. On the other hand, the patients with smaller lesions showed constantly higher response rates than those with lesions of 20mm or more, except for the cRR assessed by RECIST in the patients with lesions of 15–19mm, although the actual difference in cPR between old and new criteria represented only one patient.

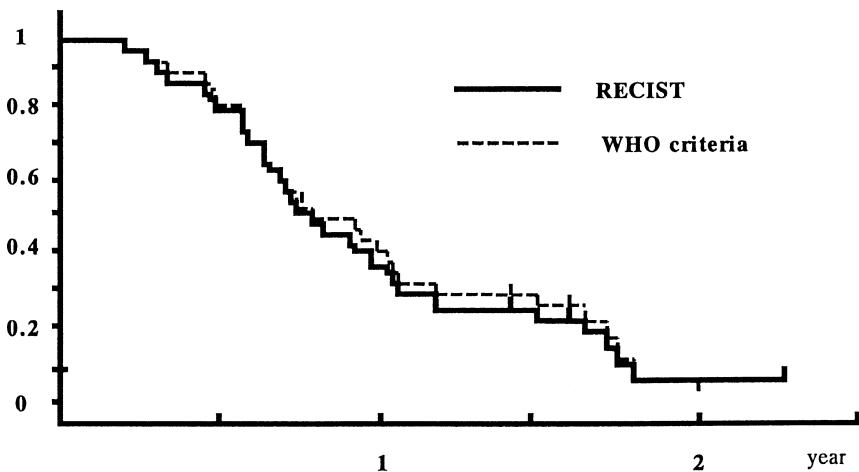
Table 4 shows the response outcomes in the 32 patients who had no measurable (or evaluable) lesions

Table 5. Response outcomes of five responders whose evaluation results assessed by conventional criteria and RECIST were inconsistent

Case number	Conventional criteria				RECIST		
	Overall	Stomach	LN	Liver	Overall	Target	Non-target
1	PR	PR	—	NC	SD	SD	Non-CR/non-PD
2	PR	NC	PR	—	SD	SD	Non-CR/non-PD
3	PR	NC	—	PR	SD	SD	Non-CR/non-PD
4	PR	—	PR	NC	SD	SD	Non-CR/non-PD
5	NC	NC	PR	NC	PR	PR	Non-CR/non-PD

LN, Lymph node

See text for definitions of SD and non-CR/non-PD

**Fig. 1.** Survival curves of patients with partial response (PR) evaluated by response evaluation criteria in solid tumor (RECIST) and WHO criteria. See text for definition of PR

other than the primary site, and the response outcomes in the patients overall, assessed by the conventional criteria. RR and cRR in these 32 patients were exactly the same (13%), and this rate was much lower than that of any other groups examined.

In 5 (4%) of the 129 patients having measurable lesions over 20 mm, the evaluation results as assessed by the conventional criteria and RECIST were inconsistent, as listed in Table 5. Of these 5 patients (cases 1 to 5 in Table 5), 4 were evaluated as PR and the remaining 1 as NC with the conventional criteria. But with the RECIST, the 4 PR patients' evaluation results were changed from PR to SD, and the 1 NC patient's results were changed to PR. The reasons for these were: the exclusion of response in the primary site (in case 1), unidimensional measurement (cases 2 and 3) and non-organ-specific sampling for target lesions (cases 4 and 5) in the RECIST. In case 1, for example, the PR in the primary lesion, assessed by the Japanese criteria, was changed to non-CR/non-PD with the RECIST. In cases 2 and 3, the tumor reduction rate was 51%–52% by two-dimensional measurement, but it was less than 30% by unidimensional measurements in regard to the baseline sum longest diameter. With the conventional criteria,

measurable lesions are assessed in regard to the organ. For example, multiple liver metastasis was regarded as one measurable lesion, the reduction rate of which was calculated as that in the total of the representative metastatic foci. In the RECIST, in contrast, up to ten measurable lesions are selected, in order of size, to define the target lesion, regardless of the organ. Because of this, the multiple liver metastases in case 4, which showed no response to chemotherapy, were regarded as the target lesions with the RECIST, and the response in the lymph node was not assessed. Conversely, the responses in the lymph nodes in case 5 were regarded as responses to the target lesions and the response outcome of NC in the liver was ruled out.

Response and survival

Of the 129 patients having measurable metastatic lesions of over 20 mm, 34 were assessed as showing PR (without consideration of response duration) and 24 as showing PD by the conventional criteria, and in the same way, 31 were assessed as PR and 21 as PD by the RECIST. Figure 1 shows the survival curves in these 34 and 31 responders, and Fig. 2 shows the survival curves

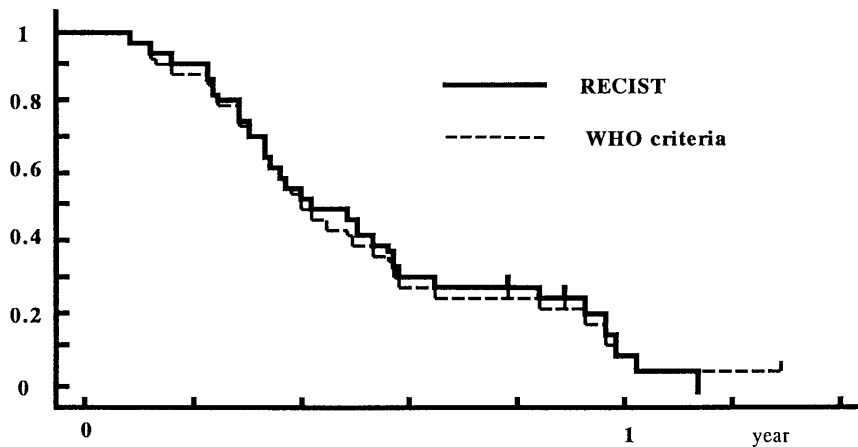


Fig. 2. Survival curves of patients with progressive disease (PD) evaluated by RECIST and WHO criteria. See text for definition of PD

of the PD patients. Regardless of the evaluation criteria, the survival curves were nearly equal in both the responders and those with progression.

Discussion

In the conventional evaluation criteria proposed by WHO in 1979 [1], the primary lesions of gastric cancer were excluded from both measurable and evaluable lesions. In Japan, where gastric cancer is very common and diagnostic skill is well advanced, various changes in the roentgenographic, endoscopic and histopathological appearances of gastric cancer observed after chemotherapy were being watched with keen interest. As a result, quite a few studies proposed tentative criteria for assessing the anticancer effects on the primary lesion. In this situation, the Japanese Research Society for Gastric Cancer made a comprehensive survey of past studies, in 1985, and proposed the Japanese evaluation criteria, using barium meal study and/or endoscopy [2]. At the beginning, the objectivity of these criteria was controversial, because barium meal studies involve much operator dependency. Actually, different air volumes can easily change the images on double-contrast photographs. But as it was easy for Japanese radiologists to assess the equality in photographs of conditions before and after chemotherapy, the evaluation results came to be reproducible soon after, and the response outcomes in the primary lesion were shown to be well correlated with the survival period [5]. In addition, recent Japanese studies demonstrate that the response rate in the primary lesion is lower than that in the metastasis, without exception, as found in the present study. Moreover, since responders in the primary lesion usually combine this with response in the metastases, their long-term survival rate is significantly higher than that in those having a response only in the metastatic lesion. In other

words, response in the primary site can be thought of as a kind of predictor for estimating the long-term survival of the patient [6,7].

As stated in the RECIST, their main purpose is to provide objectivity in response evaluation without the use of complicated measurements. Actually, the response evaluation with RECIST is extremely simple and we could not show any significant differences between the evaluation results (overall response rate and survival) with the WHO criteria and those with RECIST, in measurable lesions of 20mm or more. This should clearly indicate the validity of the RECIST and the uselessness of two-dimensional measurements, as pointed out by James et al. [8].

The considerations that we must take into account in adopting RECIST, however, are how to regard the response outcomes in the primary lesion and how to select an adequate criterion for measurable lesions. As mentioned, the local response rate in the primary lesion is the least in all of the target populations. Hence, it is true, that the response outcome of the primary lesion assessed by the Japanese criteria leads to a decrease in the overall response rate, and the use of the Japanese criteria does not inflate the overall response rate. Adopting the Japanese criteria, on the other hand, requires the level of diagnostic skill of Japanese radiologists; however, the Japanese criteria also include endoscopic evaluation criteria, in which a changing tumorous-to-flat appearance can be assessed as PR. It appears that this could be taken into consideration in the response evaluation of RECIST as an additional criterion, although, in our series, the number of patients having a non-evaluable lesion other than the primary site was not high, accounting for 13% (32/255) of our subjects.

As a measurable lesion, the RECIST recommends the selection of lesions of 20mm or more on conventional image diagnosis, because a great concern is that

assessment does not depend on differences in diagnostic skill between individuals. If we were to adopt this recommendation strictly, however, we would lose more than 25% (42/171) of measurable lesions as non-measurable lesions, as shown in our study. This would bring about a delay in the completion of phase II studies, and cause severe problems, particularly in investigational trials of new drugs. The recommendation of “over 20 mm” for selecting a target lesion was derived theoretically from the fact that CAT scan images are provided with contiguous cuts of 10 mm or less in slice thickness.

Our results here, however, have demonstrated that, whichever response criteria we adopted, there were no significant differences in the response rates in lesions of 10–14 mm and those of 15–19 mm, although there were differences in those less than 9 mm. On the other hand, the response rate was actually higher in smaller lesions than in lesions of 20 mm or more. This may occur because smaller lesions respond more readily to chemotherapy, with a resultant the similarity in the response rates evaluated by RECIST and by the WHO criteria which have been used as standard methodology for a long time. From these results, therefore, we could be allowed to select the criterion of “10 mm” for a measurable lesion, instead of “20 mm”. In such a situation, we would be able to cover the 95% (162/171) of measurable metastatic lesions according to the WHO criteria, as target lesions with the RECIST.

In conclusion, the response rates assessed by the RECIST and WHO criteria are equal in gastric cancer patients with measurable lesions over 10 mm in size. From this point of view, RECIST is valid and useful in lesions of 10 mm or more. As the Japanese evaluation

criteria do not inflate the response rate, it is thought that, when RECIST is used, they could still be useful as an additional methodology for assessing the quality of response to chemotherapy or the long-term survival of the patients treated.

Acknowledgments This work was supported in part by a Grant-in-Aid (11S-3) from the Ministry of Health and Welfare of Japan.

References

1. WHO handbook for reporting results of cancer treatment. Geneva (Switzerland): World Health Organization Offset Publication no. 48;1979.
2. Japanese Research Society for Gastric Cancer. General rules for gastric cancer study. 11th ed. Tokyo: Kanehara; 1985.
3. 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. *J Natl Cancer Inst* 2000;92:205–16.
4. Shimada Y, Shirao K, Ohtsu A, Hyodo I, Saito H, Yamamichi N, et al. Phase III study of UFT + MMC versus 5-FU + CDDP versus 5-FU alone in patients with advanced gastric cancer: JCOG study 9205. *Proc ASCO* 1043 (abstract). *J Clin Oncol* 1999;18:272.
5. Kurihara M, Izumi T, Ito T. Evaluation of the effect of chemotherapy in gastric cancer patients with an intact primary tumor according to criteria of the Japanese Research Society for Gastric Cancer. *J Jpn Soc Cancer Ther* 1991;26:644–54.
6. Ohtsu A, Boku N, Yoshida S, Miyata Y, Shirao K, Shimada Y, et al. Response of the primary lesion in gastric cancer to chemotherapeutic trials. *Int J Clin Oncol* 1998;3:3–6.
7. Koizumi W, Kurihara M, Tanabe S, Kondo I, Yamazaki I, Nonaka M, et al. Advantage of Japanese response criteria for estimating the survival of patients with primary gastric cancer. *Gastric Cancer* 1999;2:14–9.
8. James K, Eisenhauer E, Christian M, Terenziani M, Vena D, Muldul A, Therasse P. Measuring response in solid tumors: unidimensional versus bidimensional measurement. *J Natl Cancer Inst* 1999;91:523–8.