*Journal of BUON* **13:** 359-362, 2008


*ORIGINAL ARTICLE*

# Discrepancy in the assessment of tumor response in patients with pancreatic cancer: WHO versus RECIST criteria

S.H. Ahn[1], H.S. Garewal[1,2], T. Dragovich[1,3]

*[1]Department of Medicine, Section of Hematology-Oncology, University of Arizona, Tucson; [2]Section of Hematology-Oncology, Southern Arizona VA Health Care System, Tucson; [3]Arizona Cancer Center, Tucson, USA*

## Summary

***Purpose:*** *The Response Evaluation Criteria in Solid Tumors (RECIST) have largely replaced the World Health Organization (WHO) criteria as a preferred method for assessing tumor response in clinical trials. We hypothesized that due to frequent asymmetric growth pattern, as well as somewhat diffuse margins of pancreatic cancer, the use of WHO vs. RECIST criteria may result in significantly different tumor response assessments. The purpose of this retrospective study was to compare the WHO (bidimensional) to RECIST (unidimensional) in assessing treatment response in pancreatic cancer patients enrolled in clinical trials.*

***Materials and methods:*** *We have evaluated the contrast-enhanced computed tomography (CT) images from 12 pancreatic cancer patients with measurable disease enrolled in two phase I/II clinical trials at the Arizona Cancer Center, between July 2000 and July 2003. The tumor measurements were re-calculated by RECIST and WHO criteria and were compared.*

***Results:*** *In 3 out of the 12 patients (25%) there was discordant response categorization when WHO criteria were used instead of RECIST. Clinical presentations in all 3 patients were more consistent with WHO categorization.*

***Conclusion:*** *Our retrospective data analysis suggests that use of different tumor response criteria (RECIST vs. WHO) may result in different assessments of treatment efficacy in patients with pancreatic cancer on clinical trials. This finding warrants further confirmation in a larger prospectively designed trial.*

**Key words:** bidimensional, response evaluation, tumor, unidimensional

## Introduction

RECIST, relying on unidimensional measurements, have largely replaced the WHO criteria (bidimensional measurement) as a preferred method for assessing tumor response in clinical trials (Table 1). WHO criteria utilize bidimensional measurement of individual tumors, multiplying the longest diameter by the perpendicular to it. The product is summed over all measured tumors and pre and post treatment sums are compared. According to the response, patients are categorized into 4 groups (complete response/CR, partial response/PR, stable disease/SD, disease progression/PD). The RECIST proposed unidimensional measurement of the sum of the longest diameters of the measured tumors; these are input from two dimensions using the bidimensional of the product of the longest diameter and the perpendicular to it. Other relevant RECIST changes included slight changes in treatment response categorization of PR and PD.

The goal of RECIST was simplification. However, in the course of applying the RECIST to various solid malignancies, there have been concerns raised that the two criteria may differ significantly in evaluating and characterizing the treatment efficacy.

Some tumor types are affected more than others, often based on the size, location and characteristics of lesions, especially the degree of diffuseness of the

*Correspondence to:* Tomislav Dragovich, MD, PhD. Arizona Cancer Center, 1515 N. Campbell Ave, Tucson, AZ 85724, USA. Tel: +1 520 626 7725, Fax: +1 520 626 2225, E-mail: tdragovich@azcc.arizona.edu

Received 06-05-2008; Accepted 11-06-2008

**Table 1.** Definition of best response according to WHO or RECIST criteria [3]

| *Response* | *WHO change in sum of products (longest diameter and greatest perpendicular diameter)* | *RECIST change in sums of longest diameter* |
|---|---|---|
| CR | Disappearance of all target lesions without any residual lesion; confirmed at 4 weeks | Disappearance of all target lesions; confirmed at 4 weeks |
| PR | 50% or more decrease in target lesions, without a 25% increase in any one target lesion; confirmed at 4 weeks | At least 30% reduction in the sum of the longest diameter of target lesions, taking as reference the baseline study; confirmed at 4 weeks |
| SD | Neither PR nor PD criteria are met | Neither PR nor PD criteria are met, taking as reference the smallest sum of the longest diameter recorded since treatment started |
| PD | 25% or more increase in the size of measurable lesion or appearance of new lesions | At least 20% increase in the sum of the longest diameter of target lesions, taking as reference the smallest sum of the longest diameter recorded since treatment started or appearance of new lesions |

WHO: World Health Organization, RECIST: Response Evaluation Criteria in Solid Tumors, CR: complete response, PR: partial response, SD: stable disease, PD: progressive disease

margins, which limits the accuracy and reproducibility of measurements. The location and characteristics of the primary lesion of the disease make it difficult to measure reproducibly with millimeter accuracy. We hypothesized that due to measurement difficulties, plus the frequent asymmetric growth pattern of pancreatic cancer, the use of WHO vs. RECIST criteria may result in differing response assessments in a significant number of patients with pancreatic cancer.

Our interest stemmed from a pilot case, described below, in which treatment would have been prematurely discontinued if the RECIST criteria were strictly adhered to. A 46-year-old patient with unresectable pancreatic cancer enrolled in a study had a discrepancy between the WHO and RECIST criteria at a point where it would affect further therapy. During the course of treatment, the dominant tumor mass was described as "stable" by the radiologist, but the reported measurements were 41 × 19 mm to 47 × 20 mm (Figure 1). Based on RECIST this was PD. However, based on the conventional WHO criteria, it was SD. The CA 19.9 tumor marker level decreased from 44 to 13 U/mL, liver function tests improved, and the patient felt better clinically during the same period. At this point a decision to discontinue the patient from the study treatment would result from the strict application of the planned RECIST criteria. However, an assessment of the overall clinical picture was inconsistent with tumor progression. In this case, the clinical presentation was consistent with WHO criteria correctly classifying this as SD and allowing continuation of treatment.

Since there were no published studies on different measurement criteria in pancreatic cancer patients we have decided to compare the results of WHO criteria (bidimensional) vs. RECIST in assessing treatment response in pancreatic cancer patients enrolled in two phase I/II clinical trials at the Arizona Cancer Center & Southern Arizona VA Health Care System (SAVAHCS).

## Materials and methods

All CT scans were performed on a helical CT scanner (HiSpeed ZX/I; General Electric Medical Systems, Milwaukee, WI, USA) with intravenous administration of non-ionic contrast materials, and slice collimation was 3 mm in arterial and 5 mm in the venous phase in all CT scans used in this study. The CT image data were reconstructed with 5-mm thickness and directly displayed on monitors of picture archiving and communications system (Fuji Synapse PACS). Twelve patients with measurable pancreatic cancer enrolled in phase I/II clinical trials at the Arizona Cancer Center between July 2000 and July 2003 were randomly selected and evaluated using both RECIST and WHO criteria. Tumor measurements were performed with electronic calipers and re-evaluated by two or more radiologists, and disease categorizations were compared. For the RECIST, response to treatment was categorized into the following 4 categories depending on the change in the sum of the longest diameter of target lesions: CR, indicating disappearance of all target lesions; PR, indicating at least 30% reduction; PD, indicating at least 20% increase, and SD, indicating anything in between PR and PD. WHO criteria differed and showed that a change in the sum of the product was evaluated instead of the longest diameter. The WHO categorization of PR (at least 50% reduction) and PD (at least 25% increase) also differs from RECIST.

## Results

Overall findings are presented in Table 2. In 3 out of the 12 patients (25%) we found discordant response categorization when WHO criteria were used instead of RECIST. In all 3 patients, the response assessments were changed from SD to PD when WHO criteria were used instead of the RECIST.

Two of 3 patients had an increase in serum CA 19.9 as well as worsened liver function. The first patient's CA 19.9 increased from 17,800 to 35,000 U/mL and total bilirubin level from 1.5 to 7.6 mg/ml. The second patient also had an increase in CA 19.9 from 5,900 to 17,300 U/mL and total bilirubin level from 0.5 to 2.2 mg/ml. The third patient had relatively stable biochemical values but developed increased abdominal cancer-related pain.

In our experience, the discordance between RECIST and WHO criteria tended to be more pronounced when the tumor lesions had indistinct margins and/or had undergone asymmetric growth changes (Figure 1).

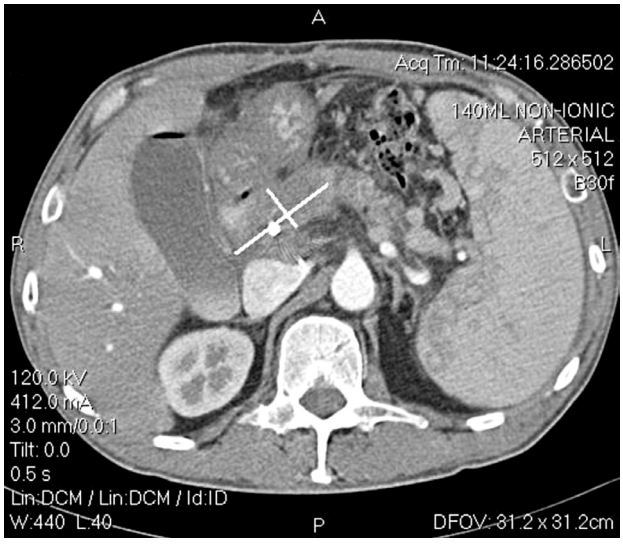


**Figure 1.** A CT image showing asymmetrical pancreatic head mass measuring 41 × 19 mm. On follow-up imaging study it was classified as progressive disease based on RECIST. However, when the WHO bi-dimensional criteria were applied, it was classified as stable disease. It is fairly obvious that aiming for millimeter accuracy would be futile in this lesion.

## Discussion

Accurate evaluation of the objective response to anticancer treatment is crucial in medical decision-making process and also when evaluating efficacy of drugs in clinical trials. Since the early 1980's, WHO criteria were accepted and widely used for the description of the effects of anticancer treatment for solid tumors [1,2]. As an effort to standardize tumor response measurements, a new guideline for evaluating tumor response was proposed in 2000, known as RECIST [3].

Combined imaging and clinical data from over 4,000 patients recruited in 14 different trials supported

**Table 2.** Comparison of WHO and RECIST criteria in pancreatic cancer patients

| | *WHO bidimensional measurement ($cm^2$)* | | | *Response by WHO criteria* | *RECIST unidimensional measurement (cm)* | | | *Response by RECIST criteria* |
|---|---|---|---|---|---|---|---|---|
| *Patient no.* | *Pre* | *Post* | *Change (%)* | | *Pre* | *Post* | *Change (%)* | |
| 1 | 31.55 | 34.00 | 7.76 | SD | 7.5 | 6.8 | -9.34 | SD |
| 2 | 31.57 | 57.50 | 82.13 | PD | 8.1 | 11.5 | 41.97 | PD |
| 3 | 25.04 | 34.10 | 36.18 | PD | 10.7 | 12.0 | 12.14 | SD |
| 4 | 16.29 | 21.15 | 29.83 | PD | 5.4 | 6.0 | 11.11 | SD |
| 5 | 23.50 | 7.88 | -66.47 | PR | 11.9 | 6.9 | -42.02 | PR |
| 6 | 10.70 | 13.41 | 25.32 | PD | 5.0 | 5.7 | 14.00 | SD |
| 7 | 65.25 | 51.96 | -20.37 | SD | 19.8 | 19.3 | -2.53 | SD |
| 8 | 9.00 | 12.58 | 39.77 | PD | 3.0 | 3.7 | 23.33 | PD |
| 9 | 35.88 | 26.00 | -27.54 | SD | 9.1 | 7.5 | -17.59 | SD |
| 10 | 13.56 | 13.59 | 0.22 | SD | 5.3 | 5.4 | 1.88 | SD |
| 11 | 11.16 | 11.61 | 4.03 | SD | 7.7 | 8.6 | 11.68 | SD |
| 12 | 1.43 | 1.00 | -30.07 | SD | 1.3 | 1.0 | -23.08 | SD |

WHO: Bidimensional measurement of individual tumors by multiplying the longest diameter and that perpendicular to it. The product is summed over all measured tumors.
RECIST: Unidimensional measurement of the sum of the longest diameters of tumors. Measurement was limited to 5 lesions per organ (10 in total) instead of all.
Treatment response categorization of partial disease response (PR) and progression (PD) were changed. PR: partial response, SD: stable disease, PD: progressive disease.

the use of the RECIST, showing that there was no difference in the percentage of responders [3]. The main "benefit" of RECIST is said to be easier application. On the contrary, Hilsenbeck and Von Hoff [4] suggested that measuring one dimension of a tumor is not necessarily less laborious than measuring two, since multiple measurements are generally needed to assure that one has assessed the maximum diameter, which is especially true for non-spherical tumors. For example, Prasad et al. [5] compared the WHO and RECIST criteria in patients with breast cancer metastatic to lung and liver. The results were discordant when there was asymmetric tumor growth/shrinkage or when the length exceeded twice its width. A retrospective analysis of several large datasets including 130 patients on clinical trials at a cancer center also revealed that response assessment by RECIST often resulted in different categorization of response compared to WHO [6]. Erasmus et al. [7] suggested that interobserver variability might be another problem for both of these criteria with inconsistency of reading between readers.

In summary, our pilot retrospective data analysis suggests that the use of different tumor response criteria (RECIST vs. WHO) may result in different assessments of treatment efficacy, thus possibly influencing treatment decisions in a significant number of patients with pancreatic cancer on clinical trials. The results were discordant in 3 of 12 cases (25%) between the two criteria, suggesting that this occurs fairly commonly. Overall, clinical findings in those discordant cases were more consistent with WHO disease assessment. Changes in categorization from SD to PD or *vice versa* have an impact on treatment decision-making as effective treatment may be prematurely stopped or patients may continue with toxic and ineffective treatment. This hypothesis, however, requires validation in a larger, prospective study in patients with pancreatic adenocarcinoma.

The use of RECIST has already been associated with an inconsistent assessment of response in some tumor types such as mesotheliomas and gastrointestinal stromal tumors (GIST) [9,10]. Byrne and Nowak proposed an adaptation of RECIST to accommodate the specific growth characteristics of mesothelioma deserving specific criteria [11]. For GIST, the development of imatinib has changed the whole perspective of morphologic vs. functional evaluation of the response. FDG-PET identifies metabolic response before any morphological response (tumor shrinkage) could be reported utilizing either RECIST or WHO criteria [12]. In GIST, size changes do not always correlate with clinical benefit observed. Tumor density changes without change in size could reflect progression or response. Thus, new criteria have been proposed for assessing the imatinib-treated GIST tumors by Benjamin et al. [13]. With the development of numerous molecularly targeted therapeutic agents, the whole concept of morphologic changes as the sole criteria for assessing the response might have to be revisited. In many current clinical trials progression free survival (PFS) is the primary endpoint, not the response rate. Our study suggests that different tumor measurement criteria may be necessary when dealing with different tumor types and growth patterns. New techniques, such as 3D volumetric measurement for the assessment of tumor response to anticancer treatment, may help resolve this issue, however, they have yet to gain wide acceptance [8].

## References


1. World Health Organization. WHO Handbook for Reporting Results of Cancer Treatment. WHO Offset Publication No. 48. Geneva, Switzerland: World Health Organization, 1979.
2. Miller AB, Hoogstraten B, Staquet M, Winkler A. Reporting results of cancer treatment. Cancer 1981; 47: 207-214.
3. Therasse P, Arbuck SG, Eisenhauer EA et al. New guidelines to evaluate the response to treatment in solid tumors. J Natl Cancer Inst 2000; 92: 205-216.
4. Hilsenbeck S, Von Hoff DD. Measure once or twice-does it really matter? (Editorial). J Natl Cancer Inst 1999; 91: 494-495.
5. Prasad SR, Saini S, Sumner JE et al. Radiological measurement of breast cancer metastases to lung and liver: comparison between WHO (bidimensional) and RECIST (unidimensional) guidelines. J Comput Assist Tomogr 2003; 27: 380-384.
6. Mazumdar M, Smith A, Schwartz LH. A statistical simulation study finds discordance between WHO criteria and RECIST guideline. J Clin Epidemiol 2004; 57: 358-365.
7. Erasmus JJ, Gladish GW, Browemeling 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.
8. Hopper KD, Kasales CJ, Eggli KD et al. The impact of 2D versus 3D quantification of tumor bulk determination on current methods of assessing response to treatment. J Comput Assist Tomogr 1996; 20: 930-937.
9. van Klaveren RJ, Aerts JG, de Bruin H, Giaccone G, Manegold C, van Meerbeeck JP. Inadequacy of the RECIST criteria for response evaluation in patients with malignant pleural mesothelioma. Lung Cancer 2004; 43: 63-69.
10. Nowak AK. CT, RECIST, and malignant pleural mesothelioma. Lung Cancer 2005; 49: S37-S40.
11. Byrne MJ, Nowak AK. Modified RECIST criteria for assessment of response in malignant pleural mesothelioma. Ann Oncol 2004; 15: 257-260.
12. Stroobants S, Goeminne J, Seegers M et al. 18FDG-Positron emission tomography for the early prediction of response in advanced soft tissue sarcoma treated with imatinib mesylate (Glyvec). Eur J Cancer 2003; 39: 2012-2020.
13. Benjamin RS, Choi H, Homer AM et al. We should desist using RECIST, at least in GIST. J Clin Oncol 2007; 25: 1760-1764.