Will There Be Resistance to the RECIST (Response Evaluation Criteria in Solid Tumors)?

Edmund A. Gehan, Mariella C. Teftt

From the earliest days of the Cancer Chemotherapy Cooperative Group Clinical Trials program in the 1950s, there have been concerns about the definition of response in solid tumors and the reliance on quantitative measurements. In the first clinical trial in solid tumors published by Zubrod et al. (1), “Treatment was considered to give a positive response if either the total measured tumor mass decreased, with no lesions increasing in size and no new lesions appearing... or the group of voting physicians considered that the treatment had been of benefit to the patient as a whole...” On the initiative of the World Health Organization (WHO) and following two meetings on the standardization of reporting results of cancer treatment in 1977 and 1979, Miller et al. (2) proposed uniform criteria for reporting response, recurrence, and disease-free interval and the grading of acute and subacute toxicity. These criteria have received wide acceptance and have become known as the WHO criteria in reporting the results of cancer treatment. In 1992, the Southwest Oncology Group (SWOG), in cooperation with the National Cancer Institute and other major cooperative oncology groups, participated in meetings to develop new toxicity criteria, end point definitions, and response criteria. A set of new guidelines was proposed because of “uncertainties in clinical trials objectives, limitations in the resolution of imaging methods, and demands for greater rigor in response and endpoint definitions” (3).

In this issue of the Journal, the criteria have been re-examined by a broad array of members from an international assortment of cooperative cancer study groups, and new guidelines for evaluating response have been proposed (4), which they refer to as RECIST (Response Evaluation Criteria in Solid Tumors). These guidelines are very similar in their definitions of response and progression to those proposed recently by James et al. (5) who have three co-authors in common with (4).

What, then, are these new guidelines and how do they differ from the WHO criteria? Table 1 gives a comparison of the WHO and RECIST guidelines (with differences in bold) in the definitions of the measurability of lesions at baseline, objective response, overall response, and duration of response. The major proposed change is that RECIST uses unidimensional measurements of the sum of the longest diameters (LDs) of tumors instead of the conventional bidimensional WHO method of the product of the longest diameter and that perpendicular to it, summed over all measured tumors.

Also, the criteria for progressive disease (PD) differ between RECIST and WHO guidelines. From Table 1, the definition of complete response (CR) is essentially the same between the guidelines; however, the definition of partial response (PR) differs. For PR, WHO requires a 50% decrease in the sum of the products of the perpendicular diameters from baseline, confirmed at 4 weeks, whereas RECIST requires at least a 30% decrease in the sum of LDs from baseline, confirmed at 4 weeks. These criteria are almost equivalent if one assumes spherical tumors and that the LD and the diameter perpendicular to the LD both decrease by at least 30% (although the latter was not measured by RECIST) because then the sum of the products of the diameters would decrease by approximately 50% or more. Therasse et al. (4) do not give a rationale for the choice of the 30% decrease criterion; however, James et al. (5) argue that measuring one dimension is simpler than two and that, for tumors of different volumes, the requirement for PR is more nearly linearly related to changes in LD than changes in the product of perpendicular diameters when tumors are spherical [Fig. 1 of (5)]. However, Hilsenbeck and Von Hoff (6) point out that measuring one dimension of a tumor is not really less laborious than measuring two, since usually multiple measurements are needed to make sure that one has the maximum diameter, especially for nonspherical tumors. There is general agreement in the literature that a spherical tumor 1 cm in diameter has approximately 10⁹ cells. Patients will have varying tumor volumes present at the start of study and, ideally, one would define PR based on the percent reduction in tumor volume at some time after the start of treatment. However, since tumor volumes cannot be accurately measured, is it better to use the LD or the product of the perpendicular diameters as a substitute for tumor volume? Skipper et al. (7) tested chemotherapeutic agents in animal systems, including L1210, CA755, and S180. Skipper et al. stated that “Basic studies in model leukemia systems indicate: the approximate average number of leukemic cells surviving therapy may be estimated by assuming logarithmic proliferation and a critical lethal number and that the percent (not the absolute number) of a leukemic cell population killed by a given dose of a given drug is reasonably constant.” In some animal and human tumor systems, a first approximation to a dose–response curve is that the logarithm of percent of cells surviving after treatment changes approximately linearly with log dose. If the assumption is made that tumors are spherical and that responding patients have equivalent percentage reductions in the measures of length, width, and depth of the tumor, then there would be essentially no difference in defining PR based on changes in LD or the product of perpendicular diameters. However, if for some tumors percent changes in LD do not reflect changes in the other dimensions, then surely measuring two dimensions would be better than one in estimating tumor volume.

RECIST (unidimensional) and WHO (bidimensional) criteria were applied to the same patients recruited in 14 different trials...
Table 1. Comparison of WHO and RECIST guidelines*

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>WHO</th>
<th>RECIST</th>
</tr>
</thead>
</table>
| Measurability of lesions at baseline   | 1. Measurable, bidimensional (product of LD and greatest perpendicular diameter)†  
  2. Nonmeasurable/evaluable (e.g., lymphangitic pulmonary metastases, abdominal masses) | 1. Measurable, unidimensional (LD only, size with conventional techniques >20 mm; spiral computed tomography >10 mm)  
  2. Nonmeasurable: all other lesions, including small lesions. Evaluate is not recommended. |
| Objective response                     | 1. Measurable disease (change in sum of products of LDs and greatest perpendicular diameters, no maximum number of lesions specified)  
  CR: disappearance of all known disease, confirmed at ≥4 wk  
  PR: ≥50% decrease from baseline, confirmed at ≥4 wk  
  PD: ≥25% increase of one or more lesions, or appearance of new lesions  
  NC: neither PR or PD criteria met  
  2. Nonmeasurable disease  
  CR: disappearance of all known disease, confirmed at ≥4 wk  
  PR: estimated decrease of ≥50%, confirmed at ≥4 wk  
  PD: estimated increase of ≥25% in existent lesions or appearance of new lesions  
  NC: neither PR or PD criteria met | 1. Target lesions (change in sum of LDs, maximum of 5 per organ up to 10 total [more than one organ])  
  CR: disappearance of all target lesions, confirmed at ≥4 wk  
  PR: ≥30% decrease from baseline, confirmed at 4 wk  
  PD: ≥20% increase over smallest sum observed, or appearance of new lesions  
  SD: neither PR or PD criteria met  
  2. Nontarget lesions  
  CR: disappearance of all target lesions and normalization of tumor markers, confirmed at ≥4 wk  
  PD: unequivocal progression of nontarget lesions, or appearance of new lesions  
  Non-PD: persistence of one or more nontarget lesions and/or tumor markers above normal limits |
| Overall response                       | 1. Best response recorded in measurable disease  
  2. NC in nonmeasurable lesions will reduce a CR in measurable lesions to an overall PR  
  3. NC in nonmeasurable lesions will not reduce a PR in measurable lesions | 1. Best response recorded in measurable disease from treatment start to disease progression or recurrence  
  2. Non-PD in nontarget lesion(s) will reduce a CR in target lesion(s) to an overall PR  
  3. Non-PD in nontarget lesion(s) will not reduce a PR in target lesion(s) |
| Duration of response                   | 1. CR  
  From: date CR criteria first met  
  To: date PD first noted  
  2. Overall response  
  From: date of treatment start  
  To: date PD first noted  
  3. In patients who only achieve a PR, only the period of overall response should be recorded | 1. Overall CR  
  From: date CR criteria first met  
  To: date disease recurrence first noted  
  2. Overall response  
  From: date CR or PR criteria first met (whichever status came first)  
  To: date recurrent disease or PD first noted  
  3. SD  
  From: date of treatment start  
  To: date PD first noted |

†Lesions that can only be measured unidimensionally are considered to be measurable (e.g., mediastinal adenopathy, malignant hepatomegaly).

[Table 4 of (4)], and there was almost no difference in the percentage of responders. This may mean that both the LD and the perpendicular diameter decreased by at least 30% from baseline so the criteria for PR were essentially equivalent between guidelines. There was some difference in the PR rates between RECIST and WHO criteria in 11 of the 14 trials and most of the trials were large (eight of the 14 having >50 patients). Consequently, though the differences in PR rates were small, this could influence the conduct of phase I and II trials of moderate size, since only small differences in the number of responders could affect the continuation or stopping of the trial. Therasse et al. (4) do not explore the subsequent prognosis of patients having PR by one criterion and NR by the other to determine which classification may have yielded more accurate predictions for the specific patients when there were disagreements.

The criteria for progressive disease (PD) also differ between the guidelines. WHO requires at least a 25% increase of one or more lesions (or the appearance of new lesions), whereas RECIST requires at least a 20% increase in the sum of LDs over the smallest sum subsequent to the start of treatment (or the appearance of new lesions). The latter increase of 20% represents a change from the 30% increase in the guidelines of James et al. (5). Assuming spherical tumors, a 20% or 30% increase in the LDs and equivalent percent increases in the perpendicular diameters would mean that the sum of the products of the diameters would increase by 44% or 69%, respectively. The change from a 30% to 20% increase brings the criterion for PD by RECIST closer to the SWOG requirement of a 50% increase in the sum of products of the perpendicular diameters or an increase of 10 cm² (whichever is smaller) for all measurable lesions over the smallest sum observed (or baseline if no decrease).

The WHO requirement of only a 25% increase in the sum of the products of the perpendicular diameters was criticized by Lavin and Flowerdew (8), who showed in 1980 that there was a one in four chance of declaring that PD had occurred when the tumor was unchanged because of the variability in measurements. When 234 patients with progressive disease defined by SWOG criteria were classified by WHO and RECIST guidelines, the date of progression was earlier by use of the WHO criteria in 17 of the 19 patients having different dates of pro-
Therasse et al. questioned whether James’ criterion (5) of a 30% increase in LD for progression was too stringent a requirement, since it might mean a delay in beginning second-line therapy when there was a reasonable chance of response. Perhaps this was the reason that Therasse et al. (4) now require only a 20% increase in LD. This requirement is close to the SWOG requirement of a 50% increase in the sum of the products of the perpendicular diameters for PD when both diameters increase by 20%.

The RECIST guidelines are more specific than the WHO criteria on a number of points that would aid those evaluating solid tumors. From Table 1 of this editorial, RECIST gives specific size requirements for measurable lesions at baseline, distinguishes target from non-target lesions, and gives the maximum number of target lesions to be followed up to a total of 10. Also, RECIST gives a baseline tumor burden (smallest sum of LDs from the treatment start) for determining progressive disease and states that all target lesions should be measured to determine progressive disease instead of “one or more measurable lesions” (WHO criteria). The SWOG standard response criteria (3) also clarify certain points by defining the number of lesions to be followed at each objective status evaluation; by defining the best response as a function of the sequence of objective status evaluations; and by defining measurable, evaluable, and nonevaluable disease.

An integral part of evaluating the status of patients on clinical trials is reporting both the response and the toxicity outcomes for each patient. Any cost–benefit-type analysis for a treatment regimen should consider the benefit (response rate, duration of response, and survival) for a given cost (incidence and degrees of major toxic effects). Both Miller et al. (2) in their presentation of WHO criteria and Green and Weiss (3) in giving SWOG criteria provide guidelines for both response and toxicity. Curiously, Therasse et al. (4) give guidelines for evaluating response but not toxicity. There should be some agreement among those evaluating solid tumor patients about criteria for both response and toxicity. Another difference between Therasse et al. (4) and Miller et al. (2, 3) is that the former investigators give substantial discussion to the types of clinical trials (early, definitive, clinical practice of oncology, uncontrolled trial, etc.) that detracts from the succinctness of the guidelines. Essentially, response could be evaluated according to the same objective criteria, whatever the type of clinical trial.

The RECIST guidelines and those given by James et al. (5) are essentially the same, except in the percentage increase required in LD for determining PD status. Hilsenbeck and Von Hoff (6) raise some issues concerning the guidelines of James et al. (5) that are also relevant in commenting about the RECIST guidelines. Hilsenbeck and Von Hoff “wonder whether the theoretical and practical reasons the authors give for adopting the new criterion are sufficient to warrant changing what has become an accepted method for evaluating the efficacy, or lack of efficacy, of anti-tumor agents.”

In summary, although clinical investigators should not “re-sist” considering new guidelines for evaluating response, there could be continuing evaluation of criteria, especially for patients when there is a disagreement about whether they have responded or not. Even small differences in response rate could affect the conduct of phase I and II trials. As we enter the new millennium when there will be an increased number of cancer clinical trials and large databases of patient outcomes, it will be of even greater importance to have consistent criteria for evaluating both response and toxicity, so that meaningful comparisons can be made between treatments in clinical trials and also between treatment regimens in different trials, both nationally and internationally.

REFERENCES