Commentary

The RECIST criteria: implications for diagnostic radiologists

[1]A R PADHANI, MRCP, FRCR and [2]L OLLIVIER, MD

[1]*The Paul Strickland Scanner Centre, Mount Vernon Hospital, Northwood, Middlesex HA6 2RN, UK and* [2]*Department of Radiology, Institut Curie, Paris 75005, France*

Monitoring response of tumours to treatment is an integral and increasingly important function of radiologists working in oncological imaging. Imaging studies play a pivotal, objective role in quantifying tumour response to a variety of physical and pharmaceutical treatments. Objective tumour shrinkage has been widely adopted as a standard end-point to select new anti-cancer drugs for future study, as a prospective end-point for definitive clinical trials designed to estimate the benefit of treatment in a specific group of patients, and is widely used in everyday clinical practice to guide clinical decision-making. In the late 1970s it became apparent that a common language was necessary to report the results of cancer treatments in a consistent manner. Standardized criteria for measuring therapeutic response were adopted in 1981 but have been modified by various cancer organizations [1–3]. The World Health Organisation (WHO), the National Cancer Institute and the European Organisation for Research and Treatment of Cancer have recently adopted a new set of tumour response criteria (Response Evaluation Criteria in Solid Tumours (RECIST)) [4]. The RECIST criteria have been introduced to unify response assessment criteria, to define how to choose evaluable lesions and to enable the use of new imaging technologies (spiral CT and MRI). The RECIST documentation goes beyond lesion selection, measurement and assessment of response. It also makes specific recommendations on the usage of imaging techniques. The CT protocols are particularly detailed (imaging parameters for incremental and spiral machines, use of contrast enhancement and the presentation of images). The implications of this document are wide ranging and are likely to have cost and manpower implications for radiology departments in cancer treatment centres. This Commentary highlights these issues.

The RECIST response criteria are largely based on a retrospective statistical evaluation of measurements (not original imaging data) obtained in eight pharmaceutical-sponsored clinical trials where 569 patients were assessed for tumour response [5]. The data analysed were selected by their "availability" but did have a broad range of tumours and serial measurements, with outcomes recorded. The quality control of the measurements themselves is unknown. In an attempt to simplify tumour measurements, unidimensional and bidimensional evaluations were compared and the new criteria selected are chosen because of the link between change in diameter, product and volume of spherical lesions (Table 1). It has recently been noted that the two measurement methods continue to show good concordance for 4613 patients [6]. The new RECIST response criteria are designed to replace existing WHO criteria; the two sets of criteria are compared in Table 2.

It is important to note that the RECIST criteria still rely on size change of lesions to make response assessments. RECIST acknowledges that tumour shrinkage may not be an appropriate end-point in the investigation of new cytostatic agents currently in phase 1 and 2 clinical trials [7]. RECIST guidance defers the issues relating to functional tumour response and unique complexities of specific tumours or anatomical sites. There

Table 1. Relationship between change in diameter, product and volume for spherical lesions [4]

	Diameter ($2r$)	Product $[(2r)^2]$	Volume $(4/3\pi r^3)$
Response	Decrease	Decrease	Decrease
	30%	50%	65%
	50%	75%	87%
Disease progression	Increase	Increase	Increase
	12%	25%	40%
	20%	44%	73%
	25%	56%	95%
	30%	69%	120%

Shaded areas represent the response evaluation criteria in solid tumours (diameter) and WHO product criteria for change in tumour size to meet response and disease progression definitions.

Received 9 February 2001 and in revised form 20 April 2001, accepted 14 May 2001.

Table 2. Definition of best response according to WHO or RECIST criteria

Best response	WHO change in sum of products	RECIST change in sum longest diameter
Complete response (CR)	Disappearance of all target lesions without any residual lesion; confirmed at 4 weeks	Disappearance of all target lesions; confirmed at 4 weeks
Partial response (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
Stable disease (SD)	Neither PR or 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
Progressive disease (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 longest diameter recorded since treatment started or appearance of new lesions

WHO, World Health Organisation [2]; RECIST, Response Evaluation Criteria in Solid Tumours [4].

are many recognized limitations of size as a tumour response variable [8]. Size changes for both response and progression remain arbitrary. The measurement of lesions is laborious. Numerous errors occur when obtaining tumour measurements. These arise from observer variations of the estimated position of the boundary of lesions. The edges of irregular or infiltrating lesions are often difficult to define and, indeed, some tumours are impossible to measure. The difficulty of distinguishing peritumoral fibrosis from tumour spread further confounds attempts at measurement. RECIST now excludes cystic or necrotic lesions when evaluating response. Measurement errors in estimating change in the size of small lesions can result in misclassification of response. Lavin and Flowerdew [9] showed that the WHO criteria of a 25% increase in the product of bidimensional diameters results in a one in four chance of declaring that progression has occurred when, in fact, the tumour is unchanged. So serious are these errors that ''independent review panels'' are often employed by pharmaceutical companies to standardize the reporting of tumour response in clinical trials. Independent review panel reports can disagree with ''home radiologists'' in 50% of cases, with major disagreements in up to 40% of cases [10]. The causes for such disagreements include variations in examination technique, lesion selection and siting of edges of target lesions. The need to tackle these discrepancies appears to be the primary motivation for revising the WHO criteria.

The four categories of response have been retained to enable comparison of results of future treatments with those from the past. Although there are no major discrepancies in the meaning or the concept of the response categories, the definition of progressive disease has changed between the WHO and RECIST criteria. WHO criteria require that an increase of 25% should be present in the product of the bidimensional diameters to document disease progression. For a sphere, this would be an increase in tumour volume of approximately 40% (Table 1). The RECIST criteria require a 20% increase in the sum of the longest diameters, which is equivalent to a 73% increase in volume of a lesion similarly measured. The primary motivation for this change is to minimize the contribution of enlargement of small lesions [9]. As a result, it will be more difficult to categorize patients with progressive disease because greater volume increases will be required. The precision of measurement estimates has not been altered because there is no inherent biological meaning for an individual patient if there is a 30% or 40% change in tumour burden.

At baseline, lesions are to be categorized as measurable or non-measurable. Measurable lesions are defined as those that can be measured accurately in at least one diameter, that is $\geqslant 20$ mm using conventional imaging techniques (including incremental CT) or $\geqslant 10$ mm using spiral CT equipment. Non-measurable lesions are discrete lesions with smaller dimensions. Non-measurable lesions also include bony metastases, leptomeningeal disease, ascites, pleural/pericardial effusions, inflammatory breast cancer, lymphangitis carcinomatosa, and heavily calcified and cystic/necrotic lesions. Interestingly, tumour lesions situated in a previously irradiated area may also not be considered as measurable disease. The term ''evaluable'', which refers to lesions that can be viewed but cannot be measured, has been dropped. After establishing that measurable disease exists, it is necessary to document ''target'' and ''non-target'' lesions. Measurable lesions, up to a maximum of five lesions per organ and ten lesions in total, representative of all involved organs, should be identified as ''target lesions''. These target lesions should be selected on the basis of size and suitability for accurate repeated

measurements. A sum of the longest diameter of all target lesions constitutes the ''baseline sum longest diameter''. Changes in sum longest diameter is to be used to categorize ''target tumour response''. Non-target lesions need not be measured on follow-up studies but any change should be noted. Final response categorization should take into account changes in both target and non-target lesions as well as noting the presence or absence of new disease (Table 3). Note that for stable disease and progressive disease categories, the pre-treatment examination no longer serves as the baseline study. Instead, the reference study from which to make an evaluation is one where the smallest sum longest diameter was recorded (Table 2).

Baseline evaluations are to be performed as close as possible to the beginning of treatment, but no more than 4 weeks before treatment starts. There is flexibility on the frequency of re-evaluation studies. However, it is recommended that follow-up every other cycle (every 6–8 weeks) should be performed in the context of phase 2 studies where the beneficial effect of a therapy is unknown. An end of treatment examination enables overall treatment response assignment. In those patients with partial response (PR) or complete response (CR), confirmatory imaging is required at 4 weeks after the criteria for CR or PR have been met (this is also required for WHO criteria).

Imaging recommendations

The role of radiography in assessing tumour response is discounted, except for the chest radiograph. The consistency of the film-to-tube distance and projection has been stressed in this technique. Lesions adjacent to the chest wall or mediastinum should be preferentially assessed by CT or MRI. Radiographs cannot be used to assess bone lesions because bony metastases are classified as non-measurable lesions.

Table 3. Overall responses for all possible combinations of tumour responses in target and non-target lesions with or without the appearance of new lesions [4]

Target lesions	Non-target lesions	New lesions	Overall response
CR	CR	No	CR
CR	Incomplete response/SD	No	PR
PR	Non-PD	No	PR
SD	Non-PD	No	SD
PD	Any	Yes or no	PD
Any	PD	Yes or no	PD
Any	Any	Yes	PD

CR, complete response; PR, partial response; SD, stable disease; PD, progressive disease.

Ultrasound should not be used routinely to assess response of lesions that are not superficial because the examination is operator independent. Ultrasound examinations cannot be reproduced for independent review at a later date because there is the implicit assumption that hard copy films truly reflect lesion dimensions. However, note that clinical palpation, endoscopic evaluation and even pathological evaluation are highly operator dependent. Ultrasound may be used as a possible alternative to clinical measurement, particularly for superficially palpable lymph nodes, subcutaneous lesions and thyroid nodules. There are no specific comments regarding ultrasound evaluation of breast cancer where presumably the palpating hand remains an acceptable method of evaluating response!

MRI is accepted as a method for obtaining measurements provided that the same anatomical plane is used on subsequent examinations. If possible, the same imager should be used for serial evaluations. RECIST recognize that images produced by scanners at different field strengths will vary in quality. There are no specific sequence recommendations.

When choosing measurable lesions on CT, the basic rule to be followed is that the minimum size should be no less than double the slice thickness. This is to minimize partial volume averaging effects that can lead to an underestimation of lesion size. The longest diameter of the target lesion should be obtained in the axial plane only. The type of CT machine is important with regard to selection of the minimum size of lesions. For spiral/helical CT scanners, the minimum size of a lesion may be 10 mm provided that a 10 mm collimation is used and reconstructions are performed at 5 mm intervals. For conventional, incremental CT scanners, the minimum size of lesions should be 20 mm with the use of contiguous 10 mm thick slices. In parts of the body where slice thicknesses used are less than 10 mm (for example examinations on small sized patients), the minimum size of measurable lesions will differ, bearing in mind the rule above.

The RECIST document recommends routine use of oral contrast medium. Many studies have shown that water is a better contrast medium when evaluating stomach and bowel lesions. Routine use of iv contrast medium is also recommended despite the fact that it may add little to a clinical study when objective response rate is based on measurable disease as the end-point. The RECIST guidance notes that some lesions become measurable only after administration of iv contrast medium. Contrast medium can be avoided when evaluating discrete lung disease. Furthermore, ''an adequate volume of suitable contrast agents should be given so that metastases

are demonstrated to their best effect and a consistent method of delivery is used on subsequent examinations for a given patient''.

Another key recommendation is that all images should be filmed, not just selected images of target lesions. Imaging departments with multislice CT scanners may thus incur increased film costs. This is intended to ensure that independent reviewers can satisfy themselves that no other co-existing abnormalities are present. All appropriate window settings should be included, particularly in the thorax. It is recommended that lesions should be measured on the same window settings at each examination. It is not acceptable to measure a lesion at lung windows in one examination and soft tissue windows on another.

The introduction of RECIST criteria is a faît accompli. RECIST criteria will replace WHO criteria for the evaluation of objective tumour response in anti-cancer drug trials. These criteria will eventually play an increasing role in routine clinical practice. Response evaluation still uses size change as the primary tumour response variable. Categories of response have not been altered. What has changed is recognition of the importance of imaging and the method by which lesions are assessed (unidimensional measurement instead of bidimensional diameters). An overall response category will require assessment of changes in all categorized lesions with or without the appearance of new lesions. Larger volume changes will be required to document progressive disease. These RECIST criteria and imaging recommendations have important implications for imaging departments in cancer centres. Participation in clinical research is a time and resource intensive process. Research protocols demand resources in excess of normal clinical demands. Under RECIST, CT examinations will be required at increased frequency during treatment and an additional examination is required to confirm response in patients achieving PR or CR. The reduced use of plain radiographs and ultrasound with the emphasis on CT is further likely to result in increased workloads for CT personnel. The need to measure and assess changes in multiple lesions in different categories, before an overall response assignment is made, is likely to have implications on the time spent by radiologists when evaluating patients participating in clinical trials. Radiologists should be enthusiastic about formulating and participating in clinical research, otherwise non-trained staff will undertake these evaluations. An increased share of pharmaceutical resources (for equipment, time and manpower) is vital for successful implementation of these recommendations.

References

1. Miller AB, Hoogstraten B, Staquet M, Winkler A. Reporting results of cancer treatment. Cancer 1981;47:207–14.
2. WHO Handbook for reporting results of cancer treatment. Geneva, Switzerland: World Health Organisation, 1979:48.
3. Hawthorn J. A practical guide to EORTC studies. Brussels: European Organisation for the Research and Treatment of Cancer (EORTC), 1994.
4. 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 Nat Cancer Inst 2000;92:205–16.
5. James K, Eisenhauer E, Christian M, Terenziani M, Vena D, Muldal A, et al. Measuring response in solid tumors: unidimensional versus bidimensional measurement. J Nat Cancer Inst 1999;91:523–8.
6. James K, Eisenhauer E, Therasse P. Measure once or twice — does it really matter? J Natl Cancer Inst 1999;91:1780–1.
7. Padhani AR. Are current tumour response criteria relevant for the 21st century? Br J Radiol 2000; 73:1031–3.
8. Husband J. Imaging of treated cancer. Br J Radiol 1995;68:1–12.
9. Lavin PT, Flowerdew G. Studies in variation associated with the measurement of solid tumors. Cancer 1980;46:1286–90.
10. Thiesse P, Ollivier L, Di Stefano-Louineau D, Negrier S, Savary J, Pignard K, et al. Response rate accuracy in oncology trials: reasons for interobserver variability. J Clin Oncol 1997;15: 3507–14.