Response assessment in cancer clinical trials

The printing of this thesis was supported by the RECIST fund created in 2000 to help standardize response criteria in solid tumors.

Response Assessment in Cancer Clinical Trials

Evaluatie van respons in klinisch- oncologische studies

Thesis

to obtain the degree of doctor from the
Erasmus University Rotterdam
by command of the
rector magnificus

Prof.dr. S.W.J. Lamberts

and in accordance with the decision of the Doctorate Board
The public defence shall be held on

Thursday 8 June 2006 at 15.00 hours
by

Patrick Austin Jean-Pol Therasse
born at Amersham, United Kingdom

DOCTORAL COMMITTEE

Promotor:	Prof.dr. J. Verweij
Other members	Prof.dr. A.T. van Oosterom
	Prof.dr. J.G.M. Klijn
	Prof.dr. G. Stoter

CONTENTS

Chapter 1	General Introduction	page	7
Chapter 2	Measuring Response in Solid Tumors: Unidimensional Versus Bidimensional measurement	page	13
Chapter 3	New Guidelines to Evaluate the Response to Treatment in Solid Tumors	page	27
Chapter 4	Measuring the Clinical Response. What does it mean?	page	57
Chapter 5	RECIST vs WHO: Prospective Comparison of Response Criteria in an EORTC Phase II clinical trial investigating ET-743 in Advanced Soft Tissue Sarcoma	page	73
Chapter 6	Magnetic Resonance Imaging of the Axial Skeleton Enables Objective Measurement of Tumor Response on Prostate Cancer Bone Metastases	page	85
Chapter 7	RECIST Revisited. A Review of Validation Studies on Tumor Assessment	page	103
Chapter 8	Summary and Conclusions	Page	121
	Samenvatting	Page	127
Acknowledgements		page	133

Chapter 1

Response Assessment in Cancer Clinical Trials
General Introduction

Many patients suffering from some form of advanced or locally advanced cancer will at some point in time be treated with anticancer agents to try to stop the natural evolution of the disease. Most anticancer agents developed over the last four decades are cytotoxic drugs inducing tumor regression, sometimes also resulting in prolongation of survival. Historically, the direct therapeutic efficacy of such treatments has been monitored through successive evaluations of the tumor burden quantified through the measurement of the size of tumor lesions that were clinically or radiologically evaluable. This process of evaluation of response to treatment is nowadays integrated in the daily practice of every oncologist and is also used in clinical trials to determine the anti-tumor activity of new anticancer agents (or combination of agents). Sometimes, the level of anticancer activity may be associated with other indicators (duration of response, type of response, time to progression, etc) to document the therapeutic efficacy (clinical benefit) procured to the patients.

In the first clinical trials in solid tumors carried out in the 1950's, tumor response was already used as an endpoint based on the subjective evaluation reported by the physician [1]. By the end of the 1970's, a group of breast cancer specialists, under the auspices of International Union Against Cancer (UICC), set the principles under which response to treatment in breast cancer should be evaluated [2]. This work was subsequently adopted and integrated into the recommendations set by the World Health Organisation (WHO) for the evaluation of cancer treatment in solid tumors [3].

The principles of response evaluation, which are still valid today, can be summarized as follows:

- the overall cancer burden can be characterized by a qualitative evaluation of tumor lesions, which are measurable, and a quantitative evaluation of tumor lesions, which are not measurable;
- measurable lesions should be evaluated before the beginning of the treatment and at regular intervals thereafter. Non-measurable lesions should also be evaluated and reported without measurements;
- the same method of investigations should be used for the evaluation of lesions before, during and after treatment;
- the combination of the evaluations of measurable and non-measurable lesions provides an estimation of the treatment effect that can be characterized by one of the following four categories:
    - complete response: disappearance of all measurable and non-measurable tumor lesions
    - partial response: shrinkage of measurable tumor lesions beyond a pre-defined percentage
    - stable disease: no sufficient shrinkage of measurable tumor lesions
    - progression: increase of measurable tumor lesions beyond a certain percentage or appearance of one or more new lesion

- complete and partial response should be confirmed with a second examination, which takes place within a certain time after the responses have been observed;
- it is recommended that responses should be reviewed and confirmed by an independent panel of experts to ensure uniform response assessment within the trial.

After 1981, many new anticancer drugs have been developed, and many researchers have also started to investigate different combinations of treatments. The experience acquired over the years and the lack of details in the WHO recommendations has stimulated the development of modified versions of the WHO criteria. For example, the South West Oncology Group (SWOG) published their version of the WHO criteria in 1990 [4], promoting amongst others a large increase in tumor size (50%) to define the progression status. Also, in the early 1990's the EORTC developed its own version of the WHO criteria [5] defining minimum sizes for lesions from different organs to be considered as measurable.

Over the years, the use of the different versions (some of them published and others unpublished) of the original WHO criteria have rendered the accuracy of the relative comparison of the results of investigations based on the same therapy very unreliable. Numerous papers in the literature have also questioned the reliability of the methodology both in terms of intra-observer and inter-observer variability [6-9]. The evolution of cancer imaging, the importance given to "response rate" as an endpoint of early clinical trials and the rapidly increasing number of new anticancer agents to be tested (and therefore the number of centers to be involved in drug development) have also justified the initiative to revisit, update and possibly improve the existing response criteria. This thesis focuses on the development and the implementation of a new set of response criteria developed thanks to the collaborative effort of several cancer research groups.

REFERENCES

1. Zubrod CG, Schneiderman SM, Frei E III et al. Appraisal of methods for the study of chemotherapy of cancer in man: comparative therapeutic trial of nitrogen mustard and thio phosphamide. J Chronic Dis 1960;11:7-33.
2. Hayward JL, Carbone PP, Heuson JC et al. Assessment of response to therapy in advanced breast cancer: a project of the Programme on Clinical Oncology of the International Union Against Cancer, Geneva, Switzerland. Cancer 1977;39(3):1289-94.
3. Miller AB, Hogestraeten B, Staquet M, Winkler A. Reporting results of cancer treatment. Cancer 1981;47:207-14.
4. Green S, Weiss GR. Southwest Oncology Group standard response criteria, endpoint definitions and toxicity criteria. Invest New Drugs 1992;10:239-53.
5. Tumor eligibility and response criteria for phase II and III studies. EORTC Data Center Manuel, Brussels, 1992, monograph.

6. Warr D, McKinney S, Tannock I. Influence of measurement error on assessment of response to anticancer chemotherapy: proposal for new criteria of tumor response. J Clin Oncol 1984;2(9):1040-46.
7. Tonkin K, Tritchler D, Tannock I. Criteria of tumor response used in clinical trials of chemotherapy. J Clin Oncol 1985;3(6):870-75.
8. Warr D, McKinney S, Tannock I. Influence of measurement error on Response Rates. Cancer Treat Rep 1985;69(10):1127-32.
9. Thiesse Ph, Ollivier L, Di Stefano-Louineau D et al. Response rate accuracy in oncology trials: reasons for interobserver variability. J Clin Oncol 1997;15(12):3507-14.

Chapter 2

Measuring Response in Solid Tumors: Unidimensional versus Bidimensional Measurement

James K, Eisenhauer E, Christian M, Terenziani M, Vena D, Muldal A and Therasse P.

J Natl Cancer Inst. 17;91(6):523-8, 1999.

ABSTRACT

Background: Tumor shrinkage is a common endpoint used in screening new cytotoxic agents. The standard World Health Organization criterion for partial response is a 50% or more decrease in the sum of the products of two measurements (the maximum diameter of a tumor and the largest diameter perpendicular to this maximum diameter) of individual tumors. However, 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. It has been hypothesized that the calculation of bidimensional products is unnecessary, and a 30% decrease in the sum of maximum diameters of individual tumors (assuming spherical shape and equivalence to a 50% reduction in the sum of the bidimensional products) was proposed as a new criterion. We have applied the standard response and the new response criteria to the same data to determine whether the same number of responses in the same patients would result.

Methods: Data from 569 patients included in eight studies of a variety of cancers were reanalyzed. The two response criteria were separately applied, and the results were compared using the k statistic. The importance of confirmatory measurements and the frequency of nonspherical tumors were also examined. In addition, for a subset of 128 patients, a unidimensional criterion for disease progression (30% increase in the sum of maximum diameters) was applied and compared with the standard definition of a 25% increase in the sum of the bidimensional products.

Results: Agreement between the unidimensional and bidimensional criteria was generally found to be good. The k statistic for concordance for overall response was 0.95.

Conclusion: We conclude that one dimensional measurement of tumor maximum diameter may be sufficient to assess change in solid tumors.

Objective tumor shrinkage, or tumor response, has been adopted as a standard endpoint to select new anticancer agents for further study and has played a role in the development of all drugs approved for use in cancer treatment to date. Newer, noncytotoxic agents that are not anticipated to produce tumor shrinkage may require the development of a different intermediate endpoint to identify agents of promise for evaluation in large trials. However, since objective tumor response will continue to be of relevance in screening new cytotoxic anticancer agents and in comparing their relative merits, a standardized approach to tumor measurement and response criteria is important.

The criteria for both response and progression are necessarily arbitrary and have traditionally been expressed as percentage changes in tumor measurement from baseline to allow their application to all patients who have measurable disease. Several attempts have been made to harmonize the criteria for tumor response and progression [1–4], and those developed by the World Health Organization (WHO) [3] are the ones most frequently used. Four response categories are defined: complete response, partial response, stable disease, and progressive disease. Complete response is not problematic because regardless of the criteria employed, disappearance of all known disease is required. WHO defines partial response as a 50% or more decrease in "total tumor load of the lesions that have been measured." The definition states that, where possible, lesions should be measured bidimensionally (multiplying the largest diameter by its perpendicular – giving the "product") and, where there are multiple lesions, all the products should be summed. In contrast, progressive disease is defined as an increase of 25% in the size of one or more measurable lesions or the appearance of new lesion(s). In practice, many groups also define disease progression as a 25% increase in the sum of the products, rather than on the basis of change in a solitary lesion.

Because the process of measurement of two dimensions and the calculation of products and their sum is laborious, we were interested in determining whether an approach based on utilizing only one dimension was theoretically valid and practically feasible. In this article, we will discuss the mathematical justification for a unidimensional approach, a proposed partial response criterion based on unidimensional assessment, and finally will apply the new proposed criterion to a large dataset from several phase II and III trials to compare the study outcomes when unidimensional and bidimensional (WHO) criteria are applied to the same datasets.

THEORETICAL REASON FOR USING ONLY ONE DIMENSION FOR THE MEASUREMENT OF TUMOR RESPONSE

A theoretical reason that unidimensional measurement may be preferred to the bidimensional product is as follows: 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. If a fixed dose of cytotoxic agent kills a constant proportion of cancer cells, then the logarithm of cell kill is directly related to arithmetic dose increases. A consequence of this relationship is that the absolute number of cells killed depends on the

mass of cells present at the time of drug exposure and thus varies from patient to patient; therefore, a partial response must be defined as a proportionate reduction of this initial mass. Ideally, therefore, the change in tumor size should be directly (linearly) related to the logarithm of the number of cells killed. Assuming tumors are spherical, the theoretical relationship between cell number and bidimensional product, on the one hand, and maximum diameter, on the other, may be examined knowing that a tumor 1 cm in diameter contains 109 cells and the arithmetic formulas are $4/3\ \pi r^3$ (where r is the radius) for the volume, $2r$ for the diameter, and $(2r)^2$ for the bidimensional product. The legend for Fig. 1 demonstrates the calculation.

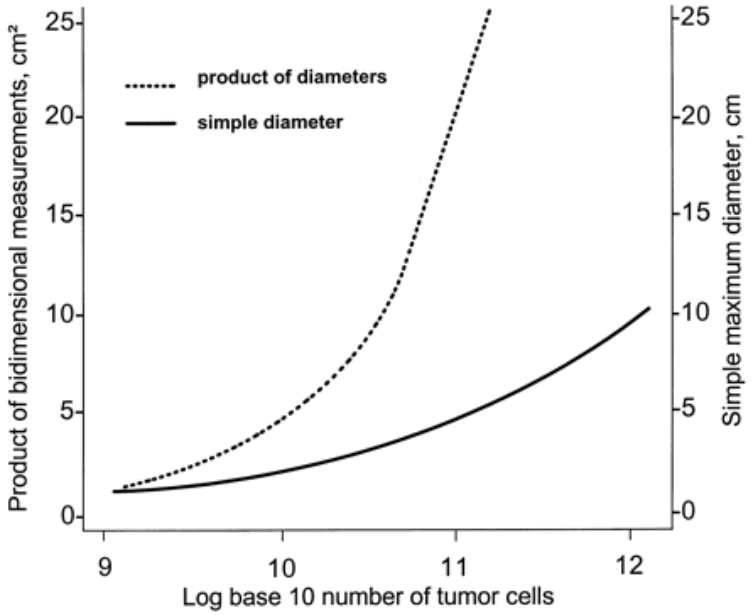

Fig. 1. Relationships between change in the number of tumor cells in a spherical tumor and simple maximum diameter and bidimensional product measurements. There is general agreement in the literature that a spherical tumor 1 cm in diameter contains 10^9 cells. Thus, in this case (given that K is a constant and the volume of a sphere is $4/3\pi.r^3$ where r is the radius), 10^9 4 K.4/3. π. $(0.5)^3$, and the general relationship between cell number, N, and radius of a tumor is $N = Q.r^3$ (where Q is a new constant equal to K.4/3. π). Because the absolute number of cells killed by a given dose of drug depends on the number of cells actually present at the time of drug exposure, attempts to measure the degree of lethality should relate to proportional reductions in tumor volume, i.e., to the log of the cell number killed. Specifically examining only the bidimensional product $(2r)^2$ and the unidimensional diameter, $2r$, Fig. 1. plots these functions of r against log $Q.r^3$. The x-axis is log $Q.r^3$, but (using the above conversion anchoring a cell number of 10^9 to an r of 0.5 cm) expressed as cell number rather than volume. The y-axes are in units of $(2r)^2$, the bidimensional product, and $2r$, the unidimensional diameter, on the left- and right-hand sides of the figure, respectively.

Fig. 1 shows this theoretical relationship for both simple diameter and product over the clinical range of tumor sizes, from 1 cm in diameter (assumed to contain 10^9 cells) to about 10 cm, assuming the tumors are spherical. Neither relationship is exactly linear but diameter is much more nearly proportional to the logarithm of cell number than is product and so changes expressed in units of diameter are approximately independent of the initial tumor sizes in different patients. Of course, the bidimensional product could be further mathematically transformed, by taking its square root, so as to have the same relationship to logarithm of cell number; the point is that simple maximum diameter already possesses this relation. The direct nature of the relationship of tumor diameter (in this case, the average diameter) and an exponent of cell number (in this case, the number of tumor cell doublings from a single cell) was, in fact, noted early in the history of clinical tumor measurement [5].

This mathematical consideration would favor the use of diameter rather than product but the strong intuitive feeling that two measurements must be better than one probably influenced the selection of the current response criteria. There is empiric evidence in the literature, however, that indicates that bidimensional measurement adds no further information than that provided by maximum diameter. Gurland and Johnson [6] demonstrated that maximum diameter correlates well with the greatest diameter perpendicular to it (correlation, 0.79–0.99, depending on the observer) and with the surface areas of various-shaped tumors (correlation, 0.85–0.99, depending on the observer) and with tumor perimeter (correlation, 0.98–0.99). Spears [7] has demonstrated that diameter becomes grossly inaccurate as an estimate of tumor size only when the length of the tumor mass is more than twice its width. The use of the bidimensional product is hallowed by many years of use, however, and has been successful in establishing many clinically useful drugs. Therefore, any suggestion for change would have to be accompanied by a demonstration that the new method is able to identify the same degree of response in the same patients as do current criteria.

PROPOSED CRITERIA FOR PARTIAL RESPONSE AND PROGRESSION USING ONE DIMENSION

Assuming spherical tumors, a 50% reduction in the product results in a decrease in volume of 65% in the tumor, as would a 30% reduction in the diameter (Table 1). Thus, we propose that the unidimensional criterion for partial response be a 30% decrease in the sum of the diameters.

In addressing the substitution of a single dimension for the product in developing partial response criteria, we also recognize that the WHO criteria for disease progression utilize changes in the bidimensional product and thus a change to unidimensional measurements in the assessment of progressive disease is also in order. We propose that disease progression be defined as a 30% increase in the sum of diameters. Although the change in volume to achieve this is much greater than the change in volume attributable to a 25% increase in products (Table 1), the risk of overcalling progression with the WHO criteria is high because of measurement error considerations. In fact, for small lesions, Lavin and Flowerdew [8] have shown that the current 25% increase in product results in a one in four chance of declaring that progression has occurred when, in fact, the tumor is unchanged.

Thus the current WHO criterion which determines that progression is achieved on the basis of only one (possibly small) lesion increasing in size by 25% [3] is very likely to result in a large number of false progressions. A further point is that, because such "progressions" are due to measurement error rather than to a real change in size, the more frequently an observation is made the greater the chance of a false progression being recorded [8]. A 30% increase in the largest diameter would represent slightly more than a doubling of tumor volume (120% increase) versus a 40% increase in tumor volume noted when there is a 25% increase in bidimensional product.

Table 1. Equivalences for Spherical Tumors

	Change in Diameter (2r)	Change in product $(2r)^2$	Change in tumor volume (cell number change), $(4/3\ \pi r^3)$
Regression	Decrease	Decrease	Decrease
	30%	50%	65%
	50%	75%	87%
Progression	Increase	Increase	Increase
	12%	25%	40%
	15%	32%	52%
	21%	46%	77%
	25%	56%	95%
	26%	59%	100%
	30%	69%	120%

METHODS

Having postulated that a 30% decrease in the sum of longest diameters should produce partial response rates similar to a 50% decrease in the sum of the products, we decided to evaluate both criteria on the same dataset by reanalyzing eight studies of cytotoxic anticancer treatment that have shown overall response rates of greater than or equal to 15% (partial plus complete responses) in bidimensionally measurable disease. Included were seven National Cancer Institute of Canada Clinical Trials Group (NCIC CTG) phase II and phase III studies (411 total number of patients; 397 assessable patients) and one Treatment Referral Center trial of the National Cancer Institute (NCI) of the United States (all 172 patients assessable). Each study was evaluated individually because data on different clinical trials were not available centrally on one computerized database. The study details are shown in Table 2.

Table 2. Studies analyzed using two methods of response assessment

Tumor type	Prior therapy	No. of assessable patients[a]	Study type phase	Treatment	Endpoints reanalyzed[b]
Breast[c]	Adjuvant only	48	II	Docetaxel	CR, PR
Breast[d]	Yes	172	II	Paclitaxel	CR, PR
Brain (oligodendroglioma)[c]	Adjuvant only	31	II	Procarbazine, vincristine, lomustine	CR, PR
Melanoma[c]	No	190	III	Carmustine, dacarbazine, cisplatin with or without tamoxifen	CR, PR
Non-small-cell lung cancer[c]	No	24	II	Paclitaxel; ifosfamide	CR, PR, SD, PD
Colorectal[c]	Adjuvant only	31	II	LY231514	CR, PR, SD, PD
Soft tissue sarcoma[c]	No	28	II	Docetaxel	CR, PR, SD, PD
Ovary[c]	Yes	45	II	Topotecan	CR, PR, SD, PD

[a]Number of patients registered on study and with response information available at the time of this analysis. Some studies included in this article continued to accrue patients and had external response reviews, both of which may have altered the final reported response rates in the study publications.

[b]CR = Complete Response; PR = partial response; SD = stable disease; PD = progressive disease.

[c]National Cancer Institute of Canada clinical trial.

[d]National Cancer Institute, United States, trial.

Patients in all NCIC CTG studies had at least one bidimensionally measurable lesion greater than or equal to 1 × 1 cm in size, if measured by physical examination or chest x-ray, and greater than or equal to 2 × 2 cm, if measured by computed tomography (CT) scan, magnetic resonance imaging (MRI) scan, or ultrasound. For the NCI trial, the requirement for measurable disease was a lesion that could be measured on either physical examination or on x-ray film with ruler or calipers or be a CT or MRI lesion of at least 1.5 cm. Tumors were measured at baseline (prestudy) and at regular intervals during the trials and lesion measurements recorded on study-specific case report forms. The majority of patients on all trials (and all of the patients on the brain tumor study) had disease documented by radiologic evaluation (CT scan, ultrasound, or MRI). Each patient's tumor measurements, as derived from case report forms, were evaluated for partial response according to two criteria: 1) WHO—a greater than or equal to 50% decrease in the sum of the product of bidimensional measurements (i.e., the maximum diameter multiplied by the largest diameter at right angles to this, for each lesion) maintained for a minimum of 4 weeks; 2)

unidimensional— a greater than or equal to 30% decrease in the sum of the largest unidimensional measurements maintained for a minimum of 4 weeks. The criteria for complete response were the same for both definitions, i.e., disappearance of all known disease maintained for a minimum of 4 weeks.

In the first three NCIC CTG studies and the NCI trial shown in Table 2, partial response and complete response were calculated by both methods. In the second group of four NCIC CTG trials (128 patients), in addition to complete and partial response, patients were also categorized as having progressive disease or stable disease according to the following definitions:

Progressive disease was defined as 1) WHO—a greater than or equal to 25% increase in the sum of the products of bidimensional measurements or the appearance of any new lesion; or 2) unidimensional—a greater than or equal to 30% increase in the sum of the largest unidimensional measurements or the appearance of any new lesion.

Stable disease was defined as 1) WHO—change in the sum of the products insufficient for partial response and for progressive disease maintained for a minimum of 4 weeks from baseline; 2) Unidimensional—change in the sum of diameters insufficient for partial response and progressive disease maintained for a minimum of 4 weeks from baseline.

Patients without repeat measurements were classified as inassessable. At the time of this analysis, some patients remained on treatment and some trials had not completed accrual so the final reported response rate on some of these trials differs from those indicated here.

Analyses performed on the final set of four NCIC CTG trials (128 patients) shown in Table 2 in addition to the progression and stable disease determinations detailed above included the following: (*a*) an assessment of the need for confirmation of response by determining how many additional partial responders would be documented by both methods if only one set of measurements meeting partial response criteria were required; (*b*) the documentation of the lesion geometry: how often were lesions spherical or nonspherical (defined as a ratio of $\geq$1.5 : 1 in perpendicular diameters)?; and (*c*) determination of the frequency with which progressive disease was shown on the basis of new lesions as opposed to an increase in the sum of the products.

RESULTS

Results of the comparison of the standard WHO and new unidimensional criteria are shown in Table 3. There was, by definition, complete concordance with regard to complete response. Results for measuring partial responses also show an excellent agreement. In the following text, the observed proportions are followed by parentheses containing two percentage numbers. The first percentage is the observed proportional percentage and the second percentage following the ± sign represents the 95% confidence intervals of that proportion. There were 126 of 569 (22.1% ± 3.4%) partial responses to the 50% product (WHO) criterion and 126 of 569 (22.1% ± 3.4%) to the 30% diameter criterion. Only five of 569 (0.88% ± 0.8%) patients were judged partial responders by the 50% product criterion but not so by the 30% diameter criterion and only five of 569 (0.88% ± 0.8%) patients were judged partial responders by the unidimensional criterion but not by the bidimensional criterion. Thus, there was an agreement in 121 of 126 responses (96% ± 3%). Concordance

for overall response rate judged by the two criteria was tested using the k statistic, the calculation of which is given in the footnote to Table 4. This discounts for any agreement between the two criteria that might be due to chance alone. The κ statistic (κ=0.95) demonstrates excellent agreement.

As noted, we also examined the impact of requiring a confirmation of measurement change for the designation of partial response in a subset of four trials. When the 4-week confirmatory measurement for partial response was eliminated, an additional five responses were identified using WHO criteria (all had been designated as stable disease), giving an overall partial response rate of 21 + 5 = 26 of 128 (20% ± 7%). In the same patients with the use of the unidimensional approach, six additional responses were identified (all had been designated as stable disease) giving an overall partial response rate of 20 + 6 = 26 of 128 (20% ± 7%). Although some of these cases had no subsequent measurement available, when such data were available, it is of interest that three of the bidimensional partial responses and four of the unidimensional partial responses showed an increase in size of measurable disease.

As would be expected from the volume relationship between the WHO criterion for progressive disease (25% sum product increase; 40% volume change) and the unidimensional criterion (30% sum diameter increase; 120% volume change), there were more patients with stable disease using the latter definition than with the WHO definition because some patients with progressive disease measured by WHO criteria had insufficient increase in unidimensional sum to qualify for progressive disease according to the newly proposed criterion. The patients under discussion are those who met the criterion for progression without either having first shown a response or having met the time requirement that would classify them as having had stable disease. In the 128 patients studied for this endpoint, 18 were found to have disease progression because they developed new lesions. By an increase in measurement of pre-existing lesions, a further 24 were judged to have disease progression by the WHO bidimensional criterion, but only nine by the more stringent proposed unidimensional criterion. Thus, 42 (18 + 24) of the 128 patients (32.8%) had a “best response” of progression according to the WHO criterion, but only 27 (18 + 9) (21%) by the new proposal.

As would be expected in the evaluation of tumor masses assessed from real patient data, not all lesions were spherical in their geometry. In the last four trials, 128 patients had a total of 370 measurable lesions recorded, 351 of which were bidimensionally measurable. Of these bidimensional lesions, 69 (19.7%) were, in fact, nonspherical, as defined by a ratio of perpendicular diameters of greater than or equal to 1.5 : 1.

Table 3. Comparison of unidimensional (new) and World Health Organisation (WHO) standard response criteria applied to the same patients

	No. of patients				Response
Trial and crietria	CR	PR	SD	PD	Rate
Breast (n=48)[a]					
WHO	4	22			54%
New	4	22			54%
Breast (n=172)[b]					
WHO	4	36			23%
New	4	40			26%
Brain (n=31)[a]					
WHO	12	10			71%
New	12	10			71%
Melanoma (n=190)[a]					
WHO	9	37			24%
New	9	34			23%
NSCLC (n=24)[a]					
WHO	0	4	16	4	17%
New	0	4	19	1	17%
Colon (n=31)[a]					
WHO	1	6	15	9	23%
New	1	5	17	8	19%
Sarcoma (n=28)[a]					
WHO	1	4	13	10	18%
New	1	5	18	4	21%
Ovary (n=45)[a]					
WHO	0	7	19	19	16%
New	0	6	25	14	13%

[a] National Cancer Institute of Canada clinical trial

[b] National Cancer Institute, United States, trial

Table 4. Overall concordance of bidimensional and unidimensional criteria[1] in the assessment of overall (complete and partial) response rate

	Bidimensional criterion		
	Response	No response	Total
Unidimensional criterion			
Response	152	5	157
No response	5	407	412
Total	157	412	569

[1]Both criteria are testing the same categorical variable (a particular decrease in tumor volume), and the formal way of estimating their concordance is with the k statistic, which is positive if the agreement is more than would be expected by chance and would be unity if there were perfect concordance between the two criteria.

$$\kappa = (O - C)/(1 - C),$$

where O is the proportion of cases where agreement is observed and C is the proportion of cases where agreement would be expected by chance alone.

From the 2 × 2 table above: O = (152 of 569) + (407 of 569) = 0.98; C = {(157 × 157/569) + (412 × 412/569)}/569 = 0.60; $\kappa = (0.98 - 0.60)/(1 - 0.60) = 0.95$.

DISCUSSION

Comparison of sequential tumor measurement data from eight phase II and phase III trials of the NCIC CTG and the U.S. NCI allowed partial response evaluation by both the criterion of a 50% decrease in the sum of the products (WHO criterion) and the criterion of a 30% decrease in the sum of the diameters (new criterion). A high degree of concordance was found between these two methods of evaluation: Regardless of the method used, the same conclusions about the efficacy of the regimen under study were reached. Furthermore, in general, the same patients were considered responders by either method. Of interest, a few additional patients would have been declared responders had there been no requirement for a 4-week confirmatory measurement. Some of these patients did, in fact, have subsequent measurements that failed to confirm that sufficient tumor shrinkage had been obtained to qualify for response. However, the majority of cases in which a measurement change sufficient for partial response was documented had tumor shrinkage confirmed by subsequent measurement.

Thus, it seems that, on the basis of the theoretical considerations presented above in which tumors were assumed to be spherical and our findings in a large set of actual patient data that included a range of both spherical and nonspherical lesions, the bidimensional measurement of solid tumors adds nothing to simple maximum diameter in assessing their response to treatment. This was first suggested by Gurland and Johnson [6] and was reaffirmed, on the basis of measuring experimental tumors in animals, by Watson [9]. In addition, there are practical reasons why diameter should be chosen over product. There is a

saving in calculation in that products are no longer required. The sum of diameters, because of its approximation to the logarithm of cell number (and thus, unlike the sum of the products, not exaggerating initial decreases in large tumors), is an ongoing indicator of how real tumor load is changing. The simplicity of the measurement also encourages the measurement of more lesions in an individual patient, and the greater the number of lesions measured, the less chance there is of falsely deciding that a partial response has occurred [10].

In terms of progression, the use of the WHO criterion (a 25% increase in product of a single lesion) creates a very high-risk situation for overcalling progression. As noted in our results, in many cases progression is obvious because of the appearance of new lesions, but for measurement-based progression, consideration could be given to ignoring small tumors (the minimum size depending on the number being measured) and to limiting the frequency of observations. A simpler solution, that a doubling of nadir size should be required for progression, was first proposed many years ago [11] and has been revived [10]. Measuring only maximum diameter and retaining the current 25% increase criterion would be consistent with this suggestion. We have tested the application of a 30% increase in the sum of diameters and found that, as would have been predicted, fewer patients are considered to have progression as their best response than when WHO criteria were used. In practice, however, the impact of this change is small: Patients who are truly progressing will declare that fact within another few weeks, and patients who truly have unchanging disease (and were incorrectly considered to have progressed) would continue to receive therapy. Since most decisions about the pursuit of new cytotoxic agents are based on the proportion of patients responding to therapy, small shifts in progression rate are unlikely to have an impact. The other advantage of utilizing a criterion of a 30% increase is that it is "symmetrical" with the partial response criterion of a 30% decrease and thus easy to remember.

In summary, we have shown that the simple maximum diameter of a tumor as well as the sum of such diameters is all that is required to determine tumor response, and we feel that this approach should replace response criteria utilizing the sum of the products in clinical cancer research.

Although not the primary intent of this article, it is also useful to raise the more philosophic issue about the "meaning" of partial response. As we have noted, any criterion of what should be called a response is arbitrary. Presumably the change of 50% in the sum of products (or 30% in sum of diameters) was chosen for its arithmetic convenience, but it may also be interesting to examine what happens when the criterion of a 50% reduction of the diameter rather than the product is applied. This criterion would certainly be more stringent: representing a reduction in tumor volume (a surrogate for cell number) of just over 87%. This is close to a one log reduction in cell number. Would this change give a greater biologic meaning to achieving a partial response? It would certainly change the numbers of responders and lower the response rates in many studies. It might also cause us to reject some of the agents as "inactive" that went on to further study following phase II evaluation. To determine if the adoption of a more stringent requirement for response was useful, a

similar type of retrospective analysis to the one performed here would be of value. Perhaps it would make us more efficient in discarding agents early on in drug development, but before suggesting this be adopted, it must be certain that regimens shown to have an impact on patient survival or palliation would not have been rejected because of lack of response activity in early phase trials.

REFERENCES

1. Breast Cancer Force Treatment Committee, National Cancer Institute: Report from the Combination Chemotherapy Trials Working Group. US Department of Health, Education, and Welfare. DHEW Publ No. (NIH) 77- 1192; 1977.
2. Hayward JL, Rubens RD, Carbone PP, Heusen JC, Kumaoka S, Segaloff A. Assessment of response to therapy in advanced breast cancer. Br J Cancer 1977;35:292–8.
3. Miller AB, Hoogstraten B, Staquet M, Winkler A. Reporting results of cancer treatment. Cancer 1981;47:207–14.
4. Oken MM, Creech RH, Tormey DC, Horton J, Davis TE, McFadden ET, et al. Toxicity and response criteria of the Eastern Cooperative Oncology Group. Am J Clin Oncol 1982;5:649–55.
5. Collins VP, Loeffler RK, Tivey H. Observations on growth rates of human tumors. Am J Roentgenol 1956;78:988–1000.
6. Gurland J, Johnson RO. Case for using only maximum diameter in measuring tumors. Cancer Chemother Rep 1966;50:119–24.
7. Spears CP. Volume doubling measurement of spherical and ellipsoidal tumors. Med Pediatr Oncol 1984;12:212–7.
8. Lavin PT, Flowerdew G. Studies in variation associated with the measurement of solid tumors. Cancer 1980;46:1286–90.
9. Watson JV.What does "response" in cancer chemotherapy really mean? Br Med J 1981;283:34–7.
10. Warr D, McKinney S, Tannock I. Influence of measurement error on assessment of response to anticancer chemotherapy: proposal for new criteria of tumor response. J Clin Oncol 1984;2:1040–6.
11. Gurland J, Johnson RO. How reliable are tumor measurements? JAMA 1965;194:973–8.

NOTES

We thank Kathy Bennett and Wendy Walsh in the National Cancer Institute of Canada Clinical Trials Group (NCIC CTG) office for their help in data review. We also thank the following individuals for their helpful comments on the manuscript: Susan Arbuck (National Cancer Institute [NCI], Bethesda, MD), Jantien Wanders (New Drug Development Office, European Organization for Research and Treatment of Cancer [EORTC], Amsterdam, The Netherlands), Jaap Verweij (Early Clinical Studies Group, EORTC, Rotterdam, The Netherlands), Richard Kaplan (NCI, Bethesda), and Stephen Gwyther (East Surrey Hospital, Redhill, Surrey, U.K.).

Chapter 3

New Guidelines to Evaluate the Response to Treatment in Solid Tumors

Therasse P, Arbuck SG, Eisenhauer EA, Wanders J, Kaplan RS, Rubinstein L, Verweij J, Van Glabbeke M, van Oosterom AT, Christian MC and Gwyther SG.

J Natl Cancer Inst. 2;92(3):205-16, 2000.

ABSTRACT

Anticancer cytotoxic agents go through a process by which their antitumor activity—on the basis of the amount of tumor shrinkage they could generate—has been investigated. In the late 1970s, the International Union Against Cancer and the World Health Organization introduced specific criteria for the codification of tumor response evaluation. In 1994, several organizations involved in clinical research combined forces to tackle the review of these criteria on the basis of the experience and knowledge acquired since then. After several years of intensive discussions, a new set of guidelines is ready that will supersede the former criteria. In parallel to this initiative, one of the participating groups developed a model by which response rates could be derived from unidimensional measurement of tumor lesions instead of the usual bidimensional approach. This new concept has been largely validated by the Response Evaluation Criteria in Solid Tumors Group and integrated into the present guidelines. This special article also provides some philosophic background to clarify the various purposes of response evaluation. It proposes a model by which a combined assessment of all existing lesions, characterized by target lesions (to be measured) and nontarget lesions, is used to extrapolate an overall response to treatment. Methods of assessing tumor lesions are better codified, briefly within the guidelines and in more detail in Appendix I. All other aspects of response evaluation have been discussed, reviewed, and amended whenever appropriate.

A. PREAMBLE

Early attempts to define the objective response of a tumor to an anticancer agent were made in the early 1960s [1,2]. In the mid- to late 1970s, the definitions of objective tumor response were widely disseminated and adopted when it became apparent that a common language would be necessary to report the results of cancer treatment in a consistent manner.

The World Health Organization (WHO) definitions published in the 1979 WHO Handbook [3] and by Miller et al. [4] in 1981 have been the criteria most commonly used by investigators around the globe. However, some problems have developed with the use of WHO criteria: 1) The methods for integrating into response assessments the change in size of measurable and "evaluable" lesions as defined by WHO vary among research groups, 2) the minimum lesion size and number of lesions to be recorded also vary, 3) the definitions of progressive disease are related to change in a single lesion by some and to a change in the overall tumor load (sum of the measurements of all lesions) by others, and 4) the arrival of new technologies (computed tomography (CT) and magnetic resonance imaging (MRI)) has led to some confusion about how to integrate three-dimensional measures into response assessment.

These issues and others have led to a number of different modifications or clarifications to the WHO criteria, resulting in a situation where response criteria are no longer comparable among research organizations—the very circumstance that the WHO publication had set out to avoid. This situation led to an initiative undertaken by representatives of several research groups to review the response definitions in use and to create a revision of the WHO criteria that, as far as possible, addressed areas of conflict and inconsistency.

In so doing, a number of principles were identified:

1) Despite the fact that "novel" therapies are being developed that may work by mechanisms unlikely to cause tumor regression, there remains an important need to continue to describe objective change in tumor size in solid tumors for the foreseeable future. Thus, the four categories of complete response, partial response, stable disease, and progressive disease, as originally categorized in the WHO Handbook [3], should be retained in any new revision.

2) Because of the need to retain some ability to compare favorable results of future therapies with those currently available, it was agreed that no major discrepancy in the meaning and the concept of partial response should exist between the old and the new guidelines, although measurement criteria would be different.

3) In some institutions, the technology now exists to determine changes in tumor volume or changes in tumor metabolism that may herald shrinkage. However, these techniques are not yet widely available, and many have not been validated. Furthermore, it was recognized that the utility of response criteria to date had not been related to precision of measurement. The definition of a partial response, in particular, is an arbitrary convention—there is no inherent meaning for an individual patient of a 50% decrease in overall tumor load. It was not thought that increased precision of measurement of tumor volume was an important goal for its own sake. Rather, standardization and simplification of methodology were desirable. Nevertheless, the guidelines proposed in this document are not meant to discourage the development of new tools that may provide more reliable surrogate endpoints than objective tumor response for predicting a potential therapeutic benefit for cancer patients.

4) Concerns regarding the ease with which a patient may be considered mistakenly to have disease progression by the current WHO criteria (primarily because of measurement error) have already led some groups such as the Southwest Oncology Group to adopt criteria that require a greater increase in size of the tumor to consider a patient to have progressive disease [5]. These concerns have led to a similar change within these revised WHO criteria (see Appendix II).

5) These criteria have not addressed several other areas of recent concern, but it is anticipated that this process will continue and the following will be considered in the future:

- Measures of antitumor activity, other than tumor shrinkage, that may appropriately allow investigation of cytostatic agents in phase II trials;
- Definitions of serum marker response and recommended methodology for their validation; and
- Specific tumors or anatomic sites presenting unique complexities.

B. BACKGROUND

These guidelines are the result of a large, international collaboration. In 1994, the European Organization for Research and Treatment of Cancer (EORTC), the National Cancer Institute (NCI) of the United States, and the National Cancer Institute of Canada Clinical Trials Group set up a task force (see Appendix III) with the main objective of reviewing the existing sets of criteria used to evaluate response to treatment in solid tumors. After 3 years of regular meetings and exchange of ideas within the task force, a draft revised version of the WHO criteria was produced and widely circulated (see Appendix IV). Comments received (response rate, 95%) were compiled and discussed within the task force before a second version of the document integrating relevant comments was issued. This second version of the document was again circulated to external reviewers who were also invited to participate in a

consensus meeting (on behalf of the organization that they represented) to discuss and finalize unresolved problems (October 1998). The list of participants to this consensus meeting is shown in Appendix IV and included representatives from academia, industry, and regulatory authorities. Following the recommendations discussed during the consensus meeting, a third version of the document was produced, presented publicly to the scientific community (American Society for Clinical Oncology, 1999), and submitted to the Journal of the National Cancer Institute in June 1999 for official publication.

Data from collaborative studies, including more than 4000 patients assessed for tumor response, support the simplification of response evaluation through the use of unidimensional measurements and the sum of the longest diameters instead of the conventional method using two measurements and the sum of the products. The results of the different retrospective analyses (comparing both approaches) performed by use of these different databases are described in Appendix V. This new approach, which has been implemented in the following guidelines, is based on the model proposed by James et al. [6].

C. RESPONSE EVALUATION CRITERIA IN SOLID TUMORS (RECIST) GUIDELINES

1. INTRODUCTION

The introduction explores the definitions, assumptions, and purposes of tumor response criteria. Below, guidelines that are offered may lead to more uniform reporting of outcomes of clinical trials. Note that, although single investigational agents are discussed, the principles are the same for drug combinations, noninvestigational agents, or approaches that do not involve drugs.

Tumor response associated with the administration of anticancer agents can be evaluated for at least three important purposes that are conceptually distinct:

- Tumor response as a prospective endpoint in early clinical trials. In this situation, objective tumor response is employed to determine whether the agent/regimen demonstrates sufficiently encouraging results to warrant further testing. These trials are typically phase II trials of investigational agents/ regimens (see section 1.2), and it is for use in this precise context that these guidelines have been developed.
- Tumor response as a prospective endpoint in more definitive clinical trials designed to provide an estimate of benefit for a specific cohort of patients. These trials are often randomized comparative trials or single-arm comparisons of combinations of agents with historical control subjects. In this setting, objective tumor response is used as a surrogate endpoint for other measures of clinical

benefit, including time to event (death or disease progression) and symptom control (see section 1.3).

- Tumor response as a guide for the clinician and patient or study subject in decisions about continuation of current therapy. This purpose is applicable both to clinical trials and to routine practice (see section 1.1), but use in the context of decisions regarding continuation of therapy is not the primary focus of this document.

However, in day-to-day usage, the distinction among these uses of the term "tumor response" can easily be missed, unless an effort is made to be explicit. When these differences are ignored, inappropriate methodology may be used and incorrect conclusions may result.

1.1. Response Outcomes in Daily Clinical Practice of Oncology

The evaluation of tumor response in the daily clinical practice of oncology may not be performed according to predefined criteria. It may, rather, be based on a subjective medical judgment that results from clinical and laboratory data that are used to assess the treatment benefit for the patient. The defined criteria developed further in this document are not necessarily applicable or complete in such a context. It might be appropriate to make a distinction between "clinical improvement" and "objective tumor response" in routine patient management outside the context of a clinical trial.

1.2. Response Outcomes in Uncontrolled Trials as a Guide to Further Testing of a New Therapy

"Observed response rate" is often employed in single-arm studies as a "screen" for new anticancer agents that warrant further testing. Related outcomes, such as response duration or proportion of patients with complete responses, are sometimes employed in a similar fashion. The utilization of a response rate in this way is not encumbered by an implied assumption about the therapeutic benefit of such responses but rather implies some degree of biologic antitumor activity of the investigated agent.

For certain types of agents (i.e., cytotoxic drugs and hormones), experience has demonstrated that objective antitumor responses observed at a rate higher than would have been expected to occur spontaneously can be useful in selecting anticancer agents for further study. Some agents selected in this way have eventually proven to be clinically useful. Furthermore, criteria for "screening" new agents in this way can be modified by accumulated experience and eventually validated in terms of the efficiency by which agents so screened are shown to be of clinical value by later, more definitive, trials.

In most circumstances, however, a new agent achieving a response rate determined a priori to be sufficiently interesting to warrant further testing may not prove to be an effective treatment for the studied disease in subsequent randomized phase III trials. Random variables and selection biases, both known and unknown, can have an overwhelming effect in small, uncontrolled trials. These trials are an efficient and economic step for initial evaluation of the activity of a new agent or combination in a given disease setting. However, many such trials are performed, and the proportion that will provide false-positive results is necessarily substantial. In many circumstances, it would be appropriate to perform a second small confirmatory trial before initiating large resource-intensive phase III trials.

Sometimes, several new therapeutic approaches are studied in a randomized phase II trial. The purpose of randomization in this setting, as in phase III studies, is to minimize the impact of random imbalances in prognostic variables. However, randomized phase II studies are, by definition, not intended to provide an adequately powered comparison between arms (regimens). Rather, the goal is simply to identify one or more arms for further testing, and the sample size is chosen so to provide reasonable confidence that a truly inferior arm is not likely to be selected. Therefore, reporting the results of such randomized phase II trials should not imply statistical comparisons between treatment arms.

1.3. Response Outcomes in Clinical Trials as a Surrogate for Palliative Effect

1.3.1. Use in nonrandomized clinical trials

The only circumstance in which objective responses in a nonrandomized trial can permit a tentative assumption of a palliative effect (i.e., beyond a purely clinical measure of benefit) is when there is an actual or implied comparison with historical series of similar patients. This assumption is strongest when the prospectively determined statistical analysis plan provides for matching of relevant prognostic variables between case subjects and a defined series of control subjects. Otherwise, there must be, at the very least, prospectively determined statistical criteria that provide a very strong justification for assumptions about the response rate that would have been expected in the appropriate “control” population (untreated or treated with conventional therapy, as fits the clinical setting). However, even under these circumstances, a high rate of observed objective response does not constitute proof or confirmation of clinical therapeutic benefit. Because of unavoidable and nonquantifiable biases inherent in nonrandomized trials, proof of benefit still requires eventual confirmation in a prospectively randomized, controlled trial of adequate size. The appropriate endpoints of therapeutic benefit for such a trial are survival, progression-free survival, or symptom control (including quality of life).

1.3.2. Use in randomized trials

Even in the context of prospectively randomized phase III comparative trials, "observed response rate" should not be the sole, or major, endpoint. The trial should be large enough that differences in response rate can be validated by association with more definitive endpoints reflecting therapeutic benefit, such as survival, progression-free survival, reduction in symptoms, or improvement (or maintenance) of quality of life.

2. MEASURABILITY OF TUMOR LESIONS AT BASELINE

2.1. Definitions

At baseline, tumor lesions will be categorized as follows: measurable (lesions that can be accurately measured in at least one dimension (longest diameter to be recorded) as ≥20 mm with conventional techniques or as ≥10 mm with spiral CT scan (see section 2.2)) or nonmeasurable (all other lesions, including small lesions (longest diameter <20 mm with conventional techniques or <10 mm with spiral CT scan) and truly nonmeasurable lesions).

The term "evaluable" in reference to measurability is not recommended and will not be used because it does not provide additional meaning or accuracy.

All measurements should be recorded in metric notation by use of a ruler or calipers. All baseline evaluations should be performed as closely as possible to the beginning of treatment and never more than 4 weeks before the beginning of treatment.

Lesions considered to be truly nonmeasurable include the following: bone lesions, leptomeningeal disease, ascites, pleural/pericardial effusion, inflammatory breast disease, lymphangitis cutis/pulmonis, abdominal masses that are not confirmed and followed by imaging techniques, and cystic lesions.

(Note: Tumor lesions that are situated in a previously irradiated area might or might not be considered measurable, and the conditions under which such lesions should be considered must be defined in the protocol when appropriate.)

2.2. Specifications by Methods of Measurements

The same method of assessment and the same technique should be used to characterize each identified and reported lesion at baseline and during follow-up. Imaging-based evaluation is preferred to evaluation by clinical examination when both methods have been used to assess the antitumor effect of a treatment.

2.2.1. Clinical examination

Clinically detected lesions will only be considered measurable when they are superficial (e.g., skin nodules and palpable lymph nodes). For the case of skin lesions,

documentation by color photography – including a ruler to estimate the size of the lesion – is recommended.

2.2.2. Chest x-ray

Lesions on chest x-ray are acceptable as measurable lesions when they are clearly defined and surrounded by aerated lung. However, CT is preferable. More details concerning the use of this method of assessment for objective tumor response evaluation are provided in Appendix I.

2.2.3. CT and MRI

CT and MRI are the best currently available and most reproducible methods for measuring lesions selected for response assessment. Conventional CT and MRI should be performed with contiguous cuts of 10 mm or less in slice thickness. Spiral CT should be performed by use of a 5-mm contiguous reconstruction algorithm; this specification applies to the tumors of the chest, abdomen, and pelvis, while head and neck tumors and those of the extremities usually require specific protocols. More details concerning the use of these methods of assessment for objective tumor response evaluation are provided in Appendix I.

2.2.4. Ultrasound

When the primary endpoint of the study is objective response evaluation, ultrasound should not be used to measure tumor lesions that are clinically not easily accessible. It may be used as a possible alternative to clinical measurements for superficial palpable lymph nodes, subcutaneous lesions, and thyroid nodules. Ultrasound might also be useful to confirm the complete disappearance of superficial lesions usually assessed by clinical examination. Justifications for not using ultrasound to measure tumor lesions for objective response evaluation are provided in Appendix I.

2.2.5. Endoscopy and laparoscopy

The utilization of these techniques for objective tumor evaluation has not yet been fully or widely validated. Their uses in this specific context require sophisticated equipment and a high level of expertise that may be available only in some centers. Therefore, utilization of such techniques for objective tumor response should be restricted to validation purposes in specialized centers. However, such techniques can be useful in confirming complete histopathologic response when biopsy specimens are obtained.

2.2.6. Tumor markers

Tumor markers alone cannot be used to assess response. However, if markers are initially above the upper normal limit, they must return to normal levels for a patient to be considered in complete clinical response when all tumor lesions have disappeared. Specific additional criteria for standardized usage of prostate-specific antigen and CA (cancer antigen) 125 response in support of clinical trials are being validated.

2.2.7. Cytology and histology

Cytologic and histologic techniques can be used to differentiate between partial response and complete response in rare cases (e.g., after treatment to differentiate between residual benign lesions and residual malignant lesions in tumor types such as germ cell tumors). Cytologic confirmation of the neoplastic nature of any effusion that appears or worsens during treatment is required when the measurable tumor has met criteria for response or stable disease. Under such circumstances, the cytologic examination of the fluid collected will permit differentiation between response or stable disease (an effusion may be a side effect of the treatment) and progressive disease (if the neoplastic origin of the fluid is confirmed). New techniques to better establish objective tumor response will be integrated into these criteria when they are fully validated to be used in the context of tumor response evaluation.

3. TUMOR RESPONSE EVALUATION

3.1. Baseline Evaluation

3.1.1. Assessment of overall tumor burden and measurable disease

To assess objective response, it is necessary to estimate the overall tumor burden at baseline to which subsequent measurements will be compared. Only patients with measurable disease at baseline should be included in protocols where objective tumor response is the primary endpoint. Measurable disease is defined by the presence of at least one measurable lesion (as defined in section 2.1). If the measurable disease is restricted to a solitary lesion, its neoplastic nature should be confirmed by cytology/histology.

3.1.2. Baseline documentation of "target" and "nontarget" lesions

All measurable lesions up to a maximum of five lesions per organ and 10 lesions in total, representative of all involved organs, should be identified as target lesions and recorded and measured at baseline. Target lesions should be selected on the basis of their size (those with the longest diameter) and their suitability for accurate repeated

measurements (either by imaging techniques or clinically). A sum of the longest diameter for all target lesions will be calculated and reported as the baseline sum longest diameter. The baseline sum longest diameter will be used as the reference by which to characterize the objective tumor response.

All other lesions (or sites of disease) should be identified as nontarget lesions and should also be recorded at baseline. Measurements of these lesions are not required, but the presence or absence of each should be noted throughout follow-up.

3.2. Response Criteria

3.2.1. Evaluation of target lesions

This section provides the definitions of the criteria used to determine objective tumor response for target lesions. The criteria have been adapted from the original WHO Handbook [3], taking into account the measurement of the longest diameter only for all target lesions: complete response—the disappearance of all target lesions; partial response—at least a 30% decrease in the sum of the longest diameter of target lesions, taking as reference the baseline sum longest diameter; progressive disease—at least a 20% increase in the sum of the longest diameter of target lesions, taking as reference the smallest sum longest diameter recorded since the treatment started or the appearance of one or more new lesions; stable disease—neither sufficient shrinkage to qualify for partial response nor sufficient increase to qualify for progressive disease, taking as reference the smallest sum longest diameter since the treatment started.

3.2.2. Evaluation of nontarget lesions

This section provides the definitions of the criteria used to determine the objective tumor response for nontarget lesions: complete response—the disappearance of all nontarget lesions and normalization of tumor marker level; incomplete response/stable disease—the persistence of one or more nontarget lesion(s) and/or the maintenance of tumor marker level above the normal limits; and progressive disease—the appearance of one or more new lesions and/or unequivocal progression of existing nontarget lesions.

(Note: Although a clear progression of “nontarget” lesions only is exceptional, in such circumstances, the opinion of the treating physician should prevail and the progression status should be confirmed later by the review panel (or study chair)).

3.2.3. Evaluation of best overall response

The best overall response is the best response recorded from the start of treatment until disease progression/recurrence (taking as reference for progressive disease the

smallest measurements recorded since the treatment started). In general, the patient's best response assignment will depend on the achievement of both measurement and confirmation criteria (see section 3.3.1). Table 1 provides overall responses for all possible combinations of tumor responses in target and nontarget lesions with or without the appearance of new lesions.

Table 1: Overall responses for all possible combinations of tumor responses in target and nontarget lesions with or without the appearance of new lesions[a]

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

[a]CR = complete response; PR = partial response; SD = stable disease; and PD = progressive disease. See text for more details.

(Notes:

- Patients with a global deterioration of health status requiring discontinuation of treatment without objective evidence of disease progression at that time should be classified as having "symptomatic deterioration." Every effort should be made to document the objective disease progression, even after discontinuation of treatment.
- Conditions that may define early progression, early death, and inevaluability are study specific and should be clearly defined in each protocol (depending on treatment duration and treatment periodicity).
- In some circumstances, it may be difficult to distinguish residual disease from normal tissue. When the evaluation of complete response depends on this determination, it is recommended that the residual lesion be investigated (fine-needle aspiration/biopsy) before confirming the complete response status.)

3.2.4. Frequency of tumor re-evaluation

Frequency of tumor re-evaluation while on treatment should be protocol specific and adapted to the type and schedule of treatment. However, in the context of phase II studies where the beneficial effect of therapy is not known, follow-up of every other cycle (i.e., 6–8 weeks) seems a reasonable norm. Smaller or greater time intervals than these could be justified in specific regimens or circumstances.

After the end of the treatment, the need for repetitive tumor evaluations depends on whether the phase II trial has, as a goal, the response rate or the time to an event (disease progression/ death). If time to an event is the main endpoint of the study, then routine re-evaluation is warranted of those patients who went off the study for reasons other than the expected event at frequencies to be determined by the protocol. Intervals between evaluations twice as long as on study are often used, but no strict rule can be made.

3.3. Confirmatory Measurement/Duration of Response

3.3.1. Confirmation

The main goal of confirmation of objective response in clinical trials is to avoid overestimating the response rate observed. This aspect of response evaluation is particularly important in nonrandomized trials where response is the primary endpoint. In this setting, to be assigned a status of partial response or complete response, changes in tumor measurements must be confirmed by repeat assessments that should be performed no less than 4 weeks after the criteria for response are first met. Longer intervals as determined by the study protocol may also be appropriate.

In the case of stable disease, measurements must have met the stable disease criteria at least once after study entry at a minimum interval (in general, not less than 6–8 weeks) that is defined in the study protocol (see section 3.3.3).

(Note: Repeat studies to confirm changes in tumor size may not always be feasible or may not be part of the standard practice in protocols where progression-free survival and overall survival are the key endpoints. In such cases, patients will not have "confirmed response." This distinction should be made clear when reporting the outcome of such studies.)

3.3.2. Duration of overall response

The duration of overall response is measured from the time that measurement criteria are met for complete response or partial response (whichever status is recorded first) until the first date that recurrent or progressive disease is objectively documented (taking as reference for progressive disease the smallest measurements recorded since the treatment started). The duration of overall complete response is measured from the time measurement criteria are first met for complete response until the first date that recurrent disease is objectively documented.

3.3.3. Duration of stable disease

Stable disease is measured from the start of the treatment until the criteria for disease progression is met (taking as reference the smallest measurements recorded since the

treatment started). The clinical relevance of the duration of stable disease varies for different tumor types and grades. Therefore, it is highly recommended that the protocol specify the minimal time interval required between two measurements for determination of stable disease. This time interval should take into account the expected clinical benefit that such a status may bring to the population under study.

(Note: The duration of response or stable disease as well as the progression-free survival are influenced by the frequency of follow-up after baseline evaluation. It is not in the scope of this guideline to define a standard follow-up frequency that should take into account many parameters, including disease types and stages, treatment periodicity, and standard practice. However, these limitations to the precision of the measured endpoint should be taken into account if comparisons among trials are to be made.)

3.4. Progression-Free Survival/Time to Progression

This document focuses primarily on the use of objective response endpoints. In some circumstances (e.g., brain tumors or investigation of noncytoreductive anticancer agents), response evaluation may not be the optimal method to assess the potential anticancer activity of new agents/regimens. In such cases, progression- free survival/time to progression can be considered valuable alternatives to provide an initial estimate of biologic effect of new agents that may work by a noncytotoxic mechanism. It is clear though that, in an uncontrolled trial proposing to utilize progression-free survival/time to progression, it will be necessary to document with care the basis for estimating what magnitude of progression-free survival/time to progression would be expected in the absence of a treatment effect. It is also recommended that the analysis be quite conservative in recognition of the likelihood of confounding biases, e.g., with regard to selection and ascertainment. Uncontrolled trials using progression- free survival or time to progression as a primary endpoint should be considered on a case-by-case basis, and the methodology to be applied should be thoroughly described in the protocol

4. RESPONSE REVIEW

For trials where the response rate is the primary endpoint, it is strongly recommended that all responses be reviewed by an expert or experts independent of the study at the study's completion. Simultaneous review of the patients' files and radiologic images is the best approach.

(Note: When a review of the radiologic images is to take place, it is also recommended that images be free of marks that might obscure the lesions or bias the evaluation of the reviewer(s)).

5. REPORTING OF RESULTS

All patients included in the study must be assessed for response to treatment, even if there are major protocol treatment deviations or if they are ineligible. Each patient will be assigned one of the following categories: 1) complete response, 2) partial response, 3) stable disease, 4) progressive disease, 5) early death from malignant disease, 6) early death from toxicity, 7) early death because of other cause, or 9) unknown (not assessable, insufficient data). (Note: By arbitrary convention, category 9 usually designates the "unknown" status of any type of data in a clinical database.)

All of the patients who met the eligibility criteria should be included in the main analysis of the response rate. Patients in response categories 4–9 should be considered as failing to respond to treatment (disease progression). Thus, an incorrect treatment schedule or drug administration does not result in exclusion from the analysis of the response rate. Precise definitions for categories 4–9 will be protocol specific.
All conclusions should be based on all eligible patients.
Subanalyses may then be performed on the basis of a subset of patients, excluding those for whom major protocol deviations have been identified (e.g., early death due to other reasons, early discontinuation of treatment, major protocol violations, etc). However, these subanalyses may not serve as the basis for drawing conclusions concerning treatment efficacy, and the reasons for excluding patients from the analysis should be clearly reported. The 95% confidence intervals should be provided.

6. RESPONSE EVALUATION IN RANDOMIZED PHASE III TRIALS

Response evaluation in phase III trials may be an indicator of the relative antitumor activity of the treatments evaluated but may usually not solely predict the real therapeutic benefit for the population studied. If objective response is selected as a primary endpoint for a phase III study (only in circumstances where a direct relationship between objective tumor response and a real therapeutic benefit can be unambiguously demonstrated for the population studied), the same criteria as those applicable to phase II trials (RECIST guidelines) should be used.

On the other hand, some of the guidelines presented in this special article might not be required in trials, such as phase III trials, in which objective response is not the primary endpoint. For example, in such trials, it might not be necessary to measure as many as 10 target lesions or to confirm response with a follow-up assessment after 4 weeks or more. Protocols should be written clearly with respect to planned response evaluation and whether confirmation is required so as to avoid post-hoc decisions affecting patient evaluability.

APPENDIX I. SPECIFICATIONS FOR RADIOLOGIC IMAGING

These notes are recommendations for use in clinical studies and, as such, these protocols for computed tomography (CT) and magnetic resonance imaging (MRI) scanning may differ from those employed in clinical practice at various institutions. The use of standardized protocols allows comparability both within and between different studies, irrespective of where the examination has been undertaken.

Specific Notes

- For chest x-ray, not only should the film be performed in full inspiration in the posteroanterior projection, but also the film to tube distance should remain constant between examinations. However, patients in trials with advanced disease may not be well enough to fulfill these criteria, and such situations should be reported together with the measurements.

Lesions bordering the thoracic wall are not suitable for measurements by chest x-ray, since a slight change in position of the patients can cause considerable differences in the plane in which the lesion is projected and may appear to cause a change that is actually an artifact. These lesions should be followed by a CT or an MRI. Similarly, lesions bordering or involving the mediastinum should be documented on CT or MRI.

- CT scans of the thorax, abdomen, and pelvis should be contiguous throughout the anatomic region of interest. As a rule of thumb, the minimum size of the lesion should be no less than double the slice thickness. Lesions smaller than this are subject to substantial "partial volume" effects (i.e., size is underestimated because of the distance of the cut from the longest diameter; such a lesion may appear to have responded or progressed on subsequent examinations, when, in fact, they remain the same size (Fig. 1)). This minimum lesion size for a given slice thickness at baseline ensures that any lesion appearing smaller on subsequent examinations will truly be decreasing in size. The longest diameter of each target lesion should be selected in the axial plane only.

The type of CT scanner is important regarding the slice thickness and minimum-sized lesion. For spiral (helical) CT scanners, the minimum size of any given lesion at baseline may be 10 mm, provided the images are reconstructed contiguously at 5-mm intervals. For conventional CT scanners, the minimum-sized lesion should be 20 mm by use of a contiguous slice thickness of 10 mm.

The fundamental difference between spiral and conventional CT is that conventional CT acquires the information only for the particular slice thickness scanned, which is then expressed as a two-dimensional representation of that thickness or volume as a gray scale image. The next slice thickness needs to be scanned before it can be imaged and so on. Spiral CT acquires the data for the whole volume imaged, typically

the whole of the thorax or upper abdomen in a single breath hold of about 20–30 seconds. To view the images, a suitable reconstruction algorithm is selected, by the machine, so the data are appropriately imaged. As suggested above, for spiral CT, 5-mm reconstructions can be made, thereby allowing a minimum-sized lesion of 10 mm.

Spiral CT is now the standard in most hospitals involved in cancer management in the United States, Europe, and Japan, so the above comments related to spiral CT are pertinent. However, some institutions involved in clinical trials will have conventional CT, but the number of these scanners will decline as they are replaced by spiral CT.

Other body parts, where CT scans are of different slice thickness (such as the neck, which is typically 5-mm thickness), or in the young pediatric population, where the slice thickness may be different, the minimum-sized lesion allowable for measurability of the lesion may be different. However, it should be double the slice thickness. The slice thickness and the minimum-sized lesion should be specified in the study protocol.

In patients in whom the abdomen and pelvis have been imaged, oral contrast agents should be given to accentuate the bowel against other soft-tissue masses. This procedure is almost universally undertaken on a routine basis.

Intravenous contrast agents should also be given, unless contraindicated for medical reasons such as allergy. This is to accentuate vascular structures from adjacent lymph node masses and to help enhance liver and other visceral metastases. Although, in clinical practice, its use may add little, in the context of a clinical study where objective response rate based on measurable disease is the endpoint, unless an intravenous contrast agent is given, a substantial number of otherwise measurable lesions will not be measurable. The use of intravenous contrast agents may sometimes seem unnecessary to monitor the evolution of specific disease sites (e.g., in patients in whom the disease is apparently restricted to the periphery of the lungs). However, the aim of a clinical study is to ensure that lesions are truly resolving, and there is no evidence of new disease at other sites scanned (e.g., small metastases in the liver) that may be more easily demonstrated with the use of intravenous contrast agent that should, therefore, also be considered in this context.

The method of administration of intravenous contrast agents is variable. Rather than try to institute rigid rules regarding methods for administering contrast agents and the volume injected, it is appropriate to suggest that an adequate volume of a suitable contrast agent should be given so that the metastases are demonstrated to best effect and a consistent method is used on subsequent examinations for any given patient.

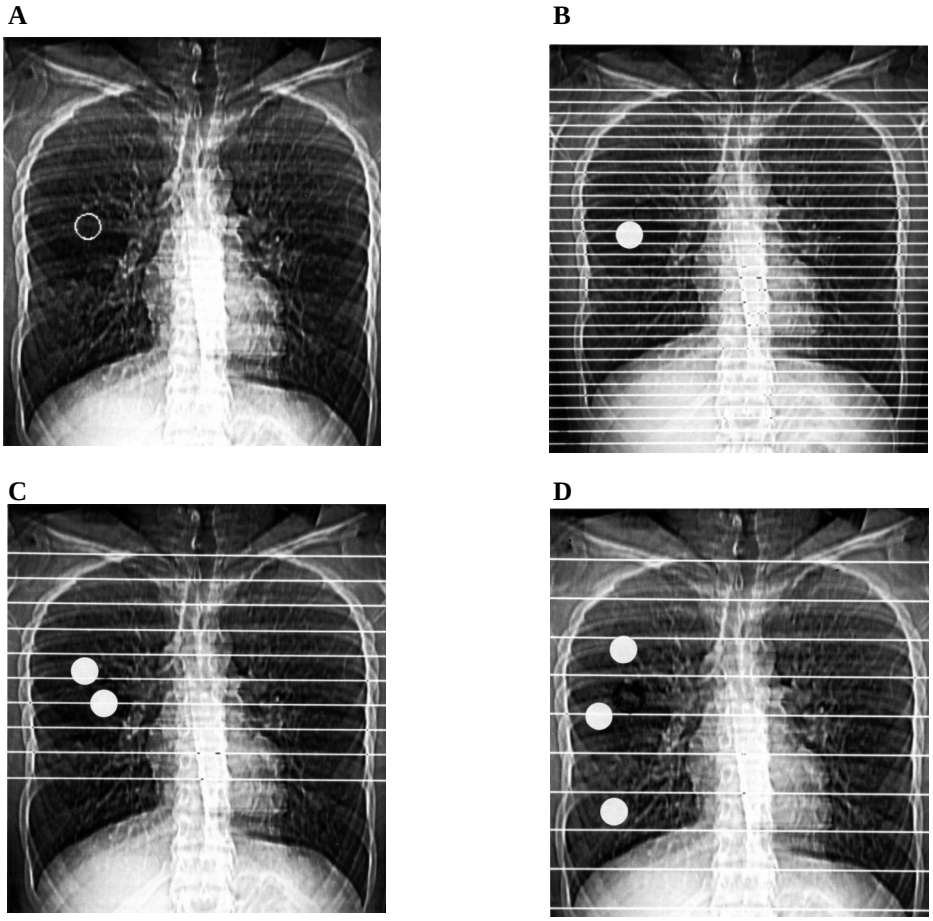

Fig 1. A) Computed tomography (CT) "scannogram" of the thorax with a simulated 20-mm lesion in the right mid-zone. B) CT "scannogram" of the thorax with contiguous slices of 10-mm thickness. Each volume within the slice thickness is scanned, and the average attenuation coefficient (i.e., density of multiple small cubes [voxels]) is represented spatially in two dimensions (pixels) as a cross-sectional image on a gray scale. It is important to note each line on the figure is a spatial representation of the average density for the structures that pass through that slice thickness, and the line does not represent a thin "cut" through it at that level. Therefore, a lesion of at least 20 mm will appear about its true diameter on at least one image because sufficient volume of the lesion is present so as not to average it down substantially. C) CT scannogram performed at 15-mm intervals. Depending on how much of the tumor is within the slice thickness, the average density may be substantially underestimated, as in the upper of the two lesions, or it may approximate the true tumor diameter, lower lesion. This is an oversimplification of the process but illustrates the point without going into the physics of CT reconstruction. D) CT scannogram performed at 24-mm intervals and of 10-mm thickness. The lesion may be imaged through its diameter, it may be partially imaged, or it may not be imaged at all. This is the equivalent of imaging a very small lesion and trying to determine whether its true diameter has changed from one examination to the next.

All images from each examination should be included and not "selected" images of the apparent lesion. This distinction is intended to ensure that, if a review is undertaken, the reviewer can satisfy himself/ herself that no other abnormalities coexist. All window settings should be included, particularly in the thorax, where the lung and soft-tissue windows should be considered.

Lesions should be measured on the same window setting on each examination. It is not acceptable to measure a lesion on lung windows on one examination and on soft-tissue settings on the next (Fig. 2). In the lung, it does not really matter whether lung or soft-tissue windows are used for intraparenchymal lesions, provided a thorough assessment of nodal and parenchymal disease has been undertaken and the target lesions are measured as appropriate by use of the same window settings for repeated examinations throughout the study.

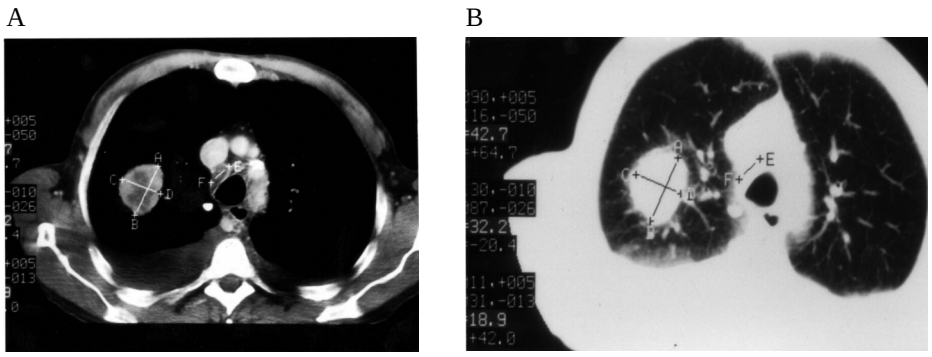

Fig 2. A) Computed tomography (CT) scan of the thorax at the level of the carina on "soft-tissue" windows. Two lesions have been measured with calipers. The intraparenchymal lesion has been measured bidimensionally, using the greatest diameter and the greatest perpendicular distance. Unidimensional measurements require only the greatest diameter to be measured. The anterior–carinal lymph node has been measured using unidimensional criteria. B) The same image as above imaged on "lung" windows, with the calipers remaining as they were for the soft-tissue measurements. The size of the lung lesion appears different. The anterior–carinal lymph node cannot be measured on these windows. The same windows should be used on subsequent examinations to measure any lesions. Some favor soft-tissue windows, so paratracheal, anterior, and subcarinal lesions may be followed on the same settings as intraparenchymal lesions.

Use of MRI is a complex issue. MRI is entirely acceptable and capable of providing images in different anatomic planes. It is, therefore, important that, when MRI is used, lesions must be measured in the same anatomic plane by use of the same imaging sequences on subsequent examinations. MRI scanners vary in the images produced. Some of the factors involved include the magnet strength (high-field magnets require shorter scan times, typically 2-5 minutes), the coil design, and patient cooperation. Wherever possible, the same scanner should be used. For instance, the images provided by a 1.5-Tesla scanner will differ from those provided by a 0.5-Tesla scanner. Although comparisons can be made between images from different scanners,

such comparisons are not ideal. Moreover, many patients with advanced malignancy are in pain, so their ability to remain still for the duration of a scan sequence – on the order of 2-5 minutes – is limited. Any movement during the scan time leads to motion artifacts and degradation of image quality, so that the examination will probably be useless. For these reasons, CT is, at this point in time, the imaging modality of choice.

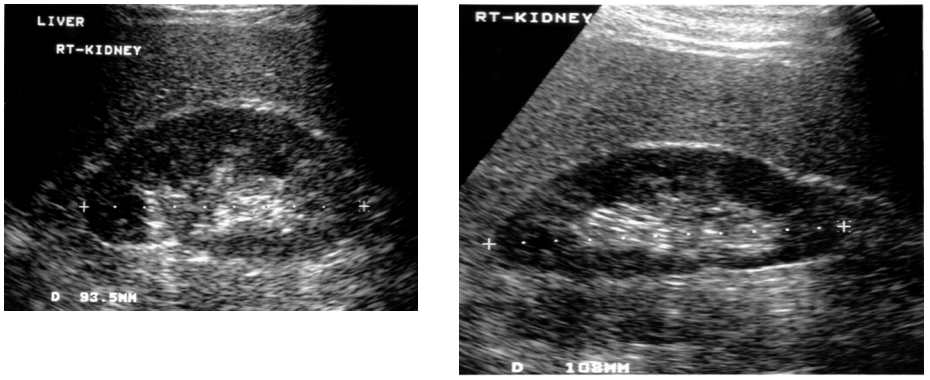

Fig 3. A) Ultrasound scan of a normal structure, the right kidney, which has been measured as 93 mm with the use of callipers. B) Ultrasound scan of the same kidney taken a few minutes later when it measures 108 mm. It appears to have increased in size by 16%. The difference is due to foreshortening of the kidney in panel A. The lack of anatomic landmarks makes accurate measurement in the same plane on subsequent examinations difficult. One has to hope that the measurements given on the hard copy film are a true and accurate reflection of events.

- Ultrasound examinations should not be used in clinical trials to measure tumor regression or progression of lesions that are not superficial because the examination is necessarily subjective. Entire examinations cannot be reproduced for independent review at a later date, and it must be assumed, whether or not it is the case, that the hard-copy films available represent a true and accurate reflection of events (Fig. 3). Furthermore, if, for example, the only measurable lesion is in the para-aortic region of the abdomen and if gas in the bowel overlies the lesion, the lesion will not be detected because the ultrasound beam cannot penetrate the gas. Accordingly, the disease staging (or restaging for treatment evaluation) for this patient will not be accurate.

The same imaging modality must be used throughout the study to measure disease. Different imaging techniques have differing sensitivities, so any given lesion may have different dimensions at any given time if measured with different modalities. It is, therefore, not acceptable to interchange different modalities throughout a trial and use these measurements. It must be the same technique throughout.

It is desirable to try to standardize the imaging modalities without adding undue constraints so that patients are not unnecessarily excluded from clinical trials.

APPENDIX II. RELATIONSHIP BETWEEN CHANGE IN DIAMETER, PRODUCT, AND VOLUME

Appendix II, Table 2. Relationship between change in diameter, product, and volume[a]

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

[a]Shaded areas represent the response evaluation criteria in solid tumors (diameter) and World Health Organization (product) criteria for change in tumor size to meet response and disease progression definitions.

APPENDIX III. RESPONSE EVALUATION CRITERIA IN SOLID TUMORS (RECIST) WORKING GROUP AND SPECIAL ACKNOWLEDGMENTS

RECIST Working Group

P. Therasse (Chair), J. Verweij, M. Van Glabbeke, A. T. van Oosterom, European Organization for Research and Treatment of Cancer (Brussels, Belgium); S. G. Arbuck, R. S. Kaplan, M. C. Christian, National Cancer Institute, United States (Bethesda, MD); E. Eisenhauer, National Cancer Institute of Canada Clinical Trials Group (Kingston); S. Gwyther, East Surrey Hospital (Redhill, U.K.); and J. Wanders, New Drug Development Office Oncology (Amsterdam, The Netherlands).

Retrospective Analyses

L. A. Rubinstein, National Cancer Institute, United States; B. K. James, A. Muldal, W. Walsh, National Cancer Institute of Canada Clinical Trials

Group; S. Green, Southwest Oncology Group (Seattle, WA); M. Terenziani, National Cancer Institute (Milan, Italy); D. Vena, Emmes Corporation (Rockville, MD); R. Canetta, J. Burroughs, Bristol- Myers Squibb (Wallingford, CT); A. Riva, M. Murawsky, Rhone- Poulenc Rorer Pharmaceuticals Inc. (Paris, France).

APPENDIX IV. PARTICIPANTS IN THE OCTOBER 1998 WORKSHOP TO DEVELOP THE FINAL RESPONSE EVALUATION CRITERIA IN SOLID TUMORS (RECIST) DOCUMENT AND FURTHER ACKNOWLEDGMENTS

Participants

S. C. S. Kao, Children's Cancer Study Group (Iowa City, IA); D. Grinblatt, Cancer and Leukemia Group B (CALGB) (Chicago, IL); B. Giantonio, Eastern Cooperative Oncology Group (ECOG) (Philadelphia, PA); F. B. Stehman, Gynecologic Oncology Group (GOG) (Indianapolis, IN); A. Trotti, Radiation Therapy Oncology Group (Tampa, FL); C. A. Coltman, Southwest Oncology Group (SWOG) (San Antonio, TX); R. E. Smith, National Surgical Adjuvant Breast and Bowel Project (Pittsburgh, PA); J. Zalcberg, Peter MacCallum Cancer Institute (Melbourne), Australia; N. Saijo, National Cancer Center Hospital (Tokyo, Japan); Y. Fujiwara, National Institute of Health Sciences (Tokyo); G. Schwartsmann, Hospital de Clinicas de Porto Alegre (Brazil); A. Klein, Health Canada, Bureau of Pharmaceutical Assessment (Ottawa, ON); B. Weinerman, National Cancer Institute of Canada Clinical Trials Group (Kingston, ON); D. Warr, Ontario Cancer Institute/Princess Margaret Hospital (Toronto); P. Liati, South Europe New Drugs Organization (Milan, Italy); S. Einstein, Bio-Imaging Technologies (West Trenton, NJ); S. Ne´grier, L. Ollivier, Fe´de´ration Nationale des Centres de Lutte contre le Cancer (Paris, France); M. Marty, International Cancer Cooperative Group/French Drug Agency (Paris); H. Anderson, A. R. Hanauske, European Organization for Research and Treatment of Cancer (EORTC) (Brussels, Belgium); M. R. Mirza, Odense University Hospital (Denmark); J. Ersboll, The European Agency for the Evaluation of Medicinal Products (Bronshoj, Denmark); C. Pagonis, Cancer Research Campaign (London, U.K.); S. Hatty, Eli Lilly and Co., (Surrey, U.K.); A. Riva, Rhone-Poulenc Rorer Pharmaceuticals Inc. (Paris); C. Royce, GlaxoWellcome (Middlesex, U.K.); G. Burke, Novartis Pharma AG (Basel, Switzerland); I. Horak, Janssen Research Foundation (Beerse, Belgium); G. Hoctin-Boes, Zeneca (Macclesfield Cheshire, U.K.); C. Weil, Bristol-Myers Squibb (Waterloo, Belgium); M. G. Zurlo, Pharmacia & Upjohn (Milan); S. Z. Fields, SmithKline Beecham Pharmaceuticals (Collegeville, PA); B. Osterwalder, Hoffmann- La Roche Inc. (Basel); Y. Shimamura, Taiho Pharmaceutical Co. Ltd. (Tokyo); and M. Okabe, Kyowa-Hakko-Kogyo Co. Ltd. (Tokyo).

Additional comments were received from the following:

A. Hamilton, R. De Wit, E. Van Cutsem, J. Wils, J.-L. Lefèbvre, I. Vergote, M. S. Aapro, J.-F. Bosset, M. Hernandez-Bronchud, D. Lacombe, H. J. Schmoll, E. Van Limbergen, P. Fumoleau, A. Bowman, U. Bruntsch, EORTC (Brussels); B. Escudier, P. Thiesse, N. Tournemaine, P. Troufleau, C. Lasset, F. Gomez, Fe´de´ration Nationale des Centres de Lutte contre le Cancer (Paris); G. Rustin, Mount Vernon Hospital (Northwood Middlesex, U.K.); S. B. Kaye, Western Infirmary (Glasgow, U.K.); A. Goldhirsch, F. Nole`, G. Zampino, F. De Braud, M. Colleoni, E. Munzone, T. De Pas, International Breast Cancer Study Group and Istituto Europeo di Oncologia (Milan); M. Castiglione, J. F. Delaloye, A. Roth, C. Sessa, D. Hess, B. Thürlimann, C. Böhme, T. Cerny, U. Hess, Schweizer Arbeitsgemeinschaft für Klinische Krebsforschung (Bern, Switzerland); H. J. Stewart, Scottish Cancer Therapy Network (Edinburgh, U.K.); A. Howell, J. F. R. Robertson, United Kingdom Coordinating Committee on Cancer Research (Nottingham); K. Noever, Bio-Imaging Technologies (Monheim, Germany); M. Kurihara, Toyosu Hospital, SHOWA University (Tokyo); L. Seymour, J. Pater, J. Rusthoven, F. Shepherd, J. Maroun, G. Cairncross, D. Stewart, K. Pritchard, National Cancer Institute of Canada Clinical Trials Group (Kingston); T. Uscinowicz, Health Canada, Bureau of Pharmaceutical Assessment (Ottawa); I. Tannock, Princess Margaret Hospital (Toronto); M. Azab, QLT Phototherapeutics (Vancouver, Canada); V. H. C. Bramwell, Canadian Sarcoma Group (London); P. O'Dwyer, ECOG (Philadelphia); A. Martin, S. Ellenberg, U.S. Food and Drug Administration (Rockville, MD); C. Chow, D. Sullivan, A. Murgo, A. Dwyer, J. Tatum, National Cancer Institute (Bethesda, MD); R. Schilsky, CALGB (Chicago, IL); J. Crowley, S. Green, SWOG (Seattle, WA); R. Park, GOG (Philadelphia, PA); V. Land, B. D. Fletcher, Pediatric Oncology Group (Chicago, IL); B. Hillman, University of Virginia (Charlottesville); F. Muggia, New York University Medical Center (New York); C. Erlichman, Mayo Clinic (Rochester, MN); L. H. Schwartz, Memorial Sloan-Kettering Cancer Center (New York, NY); S. P. Balcerzak, Ohio State University Health Sciences Center (Columbus); G. Fleming, CALGB (Chicago); G. Sorensen, Harvard University (Cambridge, MA); H. Levy, Thomas Jefferson University (Philadelphia); N. Patz, Duke University (Durham, NC); C. Visseren-Grul, Eli Lilly Nederland BV (Nieuwegein, The Netherlands)/J. Walling, Lilly Research Laboratories (Indianapolis); P. Hellemans, Janssen Research Foundation (Beerse, Belgium); L. Finke, Merck (Darmstadt, Germany); A. Man, N. Barbet, Novartis Pharma AG (Basel); G. Massimini, Pharmacia & Upjohn (Milan); J, Jimeno, Pharma Mar (Madrid, Spain); I. Hudson, SmithKline Beecham Pharmaceuticals (Essex, U.K.); and J. Krebs, R. A. Beckman, S. Lane, D. Fitts, SmithKline Beecham Pharmaceuticals (Collegeville).

APPENDIX V. RETROSPECTIVE COMPARISON OF RESPONSE/DISEASE PROGRESSION RATES OBTAINED WITH THE WORLD HEALTH ORGANIZATION (WHO)/SOUTHWEST ONCOLOGY GROUP CRITERIA AND THE NEW RESPONSE EVALUATION CRITERIA IN SOLID TUMORS (RECIST) CRITERIA

To evaluate the hypothesis by which unidimensional measurement of tumor lesions may substitute for the usual bidimensional approach, a number of retrospective analyses have been undertaken. The results of these analyses are given below in this section.

1. Comparison of Response and Disease Progression Rates by Use of WHO (or Modified WHO) or RECIST Methods

1.1. Trials Evaluated

No specific selection criteria were employed except that trial data had to include serial (repeated) records of tumor measurements. Several groups evaluated their own data on one or more such studies (National Institute of Canada Clinical Trials Group, Kingston, ON; U.S. National Cancer Institute, Bethesda, MD; and Rhone-Poulenc Rorer Pharmaceuticals Inc., Paris, France) or made data available for evaluation to the U.S. National Cancer Institute (Southwest Oncology Group and Bristol- Myers Squibb, Wallingford, CT)

1.2. Response Criteria Evaluated

Not all databases were assessed for all response outcomes. At the outset of this process, the most interest was in the assessment of complete plus partial response rate comparisons by both the WHO and new RECIST criteria. Once these data suggested no impact of using the new criteria on the response rate, several more databases were analyzed for the impact of the use of the new criteria not only on complete response plus partial response but also on stable disease and progressive disease rates (see Appendix V, Table 4) and on time to disease progression (see Appendix V, Table 5).

1.3. Methods of Comparison

For each patient in each study, baseline sums were calculated (sum of products of the two longest diameters in perpendicular dimensions for WHO and sum of longest diameters for RECIST). After each assessment, when new tumor measures were available, the sums were recalculated. Patients were assigned complete response, partial response, stable disease, and progressive disease as their “best” response on the basis of achieving the measurement criteria as indicated in Appendix V, Table 3. For both WHO and RECIST, a minimum interval of 4 weeks was required to consider complete response and partial response confirmed. Each patient could, therefore, be

assigned a best response according to each of the two criteria. The overall response and disease progression rates could be calculated for the population studied for each trial or dataset examined.

(Note: For WHO progressive disease, as is the convention in most groups, an increase in sums of products was required, not an increase in only one lesion.)

1.4. Results

2. Evaluation of Time to Disease Progression

Time to disease progression was evaluated, comparing WHO criteria with RECIST in a dataset provided by the Southwest Oncology Group (SWOG). Since SWOG criteria [5] for disease progression is a 50% increase in the sum of the products, or new disease, or an absolute increase of 10 cm² in the sum of the products, this dataset provided the means of assessing the impact of time to disease progression differences between a 25% increase in the sum of the products and a 20% increase in the sum of the longest diameters (equivalent to approximately a 44% increase in the product sum).

Appendix V, Table 3: Definition of best response according to WHO or RECIST criteria[a]

Best Response	WHO (change in sum of products)	RECIST (change in sums longest diameters)
CR	Disappearance Confirmed at 4 weeks[b]	Disappearance Confirmed at 4 weeks[b]
PR	50% decrease confirmed at 4 weeks[b]	30% decrease confirmed at 4 weeks[b]
SD	Neither PR nor PD criteria met	Neither PR nor PD criteria met
PD	25% increase No CR, PR, or SD documented before increased disease	20% increase No CR, PR, or SD documented before increased disease

[a]WHO = World Health Organization; RECIST = Response Evaluation Criteria in Solid Tumors; CR = complete response; PR = partial response; SD = stable disease; and PD = progressive disease.

[b]For the Bristol-Myers Squibb (Wallingford, CT) dataset, only unconfirmed CR and PR have been used to compare best response measured in one dimension (RECIST criteria) versus best response measured in two dimensions (WHO criteria). The computer flag identifying confirmed response in this dataset could not be used in the comparison for technical reasons.

Appendix V, Table 4. Comparison of RECIST (unidimensional) and WHO (bidimensional) criteria in the same patients recruited in 14 different trials[a]

Tumor	Criteria	No Pts Eval	Best response				RR	PD rate
			CR	PR	SD	PD		
Breast[b]	WHO	48	4	22			54%	
	RECIST	48	4	22			54%	
Breast[c]	WHO	172	4	36			23%	
	RECIST	172	4	40			26%	
Brain[b]	WHO	31	12	10			71%	
	RECIST	31	12	10			71%	
Melanoma[b]	WHO	190	9	37			24%	
	RECIST	190	9	34			23%	
Breast[d]	WHO	531	50	102			29%	
	RECIST	531	50	108			30%	
Colon[d]	WHO	1096	12	137			14%	
	RECIST	1096	12	133			13%	
Lung[d]	WHO	1197	60	317			32%	
	RECIST	1197	60	318			32%	
Ovary[d]	WHO	554	24	108			24%	
	RECIST	554	24	105			23%	
Lung[b]	WHO	24	0	4	16	4	17%	17%
	RECIST	24	0	4	19	1	17%	4%
Colon[b]	WHO	31	1	6	15	9	23%	29%
	RECIST	31	1	5	16	9	21%	29%
Sarcoma[b]	WHO	28	1	4	13	10	18%	36%
	RECIST	28	1	5	17	5	21%	18%
Ovary[b]	WHO	45	0	7	19	19	16%	42%
	RECIST	45	0	6	21	18	13%	40%
Breast[e]	WHO	306	18	114	117	57	43%	19%
	RECIST	306	18	108	124	56	41%	18%
Breast[e]	WHO	360	10	73	135	142	23%	39%
	RECIST	361	10	70	139	142	22%	39%
TOTAL (all studies where tumor response was evaluated)	**WHO**	**4613**	**205**	**977**			**25.6%**	
	RECIST	**4614**	**205**	**968**			**25.4%**	
TOTAL (all studies where PD as well as CR+PR were evaluated)	**WHO**	**794**			**315**	**241**		**30.3%**
	RECIST	**795**			**336**	**231**		**29%**

[a]WHO = World Health Organization (3); RECIST = Response Evaluation Criteria in Solid Tumors; CR = complete response; PR = partial response; SD = stable disease; PD = progressive disease; and RR = response rate.
[b]Data from the National Cancer Institute of Canada Clinical Trials Group phase II and III trials.
[c]Data from the National Cancer Institute, United States phase III trial.
[d]Data from Bristol-Myers Squibb (Wallingford, CT) phase II and III trials.
[e]Data from Rhone-Poulenc Rorer Pharmaceuticals Inc., (Paris, France) phase III trials (note: one patient in this database had unidimensional measured lesions only and could not be evaluated with the WHO criteria).

Appendix V, Table 5. Proportions of patients with disease progression by different assessment methods[a]

	No. Pts	%
Total No. of progressors	234	100
Progress by appearance of new lesions[b]	118	50
Progress by increase in pre-existing measurable disease	116	50
Same date of progression by WHO and RECIST Criteria	215	91.9
Different date of progression	19	8.1
Earlier PD with WHO criterion	17	7.3
Earlier PD with unidimensional criterion	2	0.9

[a]PD = progressive disease; WHO = World Health Organization; and RECIST = Response Evaluation Criteria in Solid Tumors.
[b]Also includes a few patients with PD because of marked increase of nonmeasurable disease.

Appendix V, Table 6. Magnitude of time to disease progression disagreements when differences existed[a]

	No. Pts	%
Number of progressors with differing progression dates	19	8.1
8 to 9 weeks difference	3	1.3
12 weeks difference	1	0.4
24-31 weeks difference[b]	2	0.9
Difference uncertain due to censoring of either WHO or RECIST progression time[c]	13	5.6

[a]WHO = World Health Organization; RECIST = Response Evaluation Criteria in Solid Tumors.
[b]For one patient, progression by RECIST (one-dimension) criteria preceded that by WHO criteria by 24 weeks due primarily to one-dimensional growth. For a second patient, with a colon tumor that increased in cross-section by 25%, then regressed completely, and then recurred, progression by WHO criteria preceded that by RECIST criteria by 31 weeks.
[c]As indicated in Appendix V, Table 6, 13 of the 19 patients had uncertain disease progression time differences when comparing RECIST and WHO criteria. In these patients, the RECIST progression criteria were not met by the time that disease progression by Southwest Oncology Group (SWOG) criteria (5) had occurred (50% increase or a 10 cm^2 increase in tumor cross-section). Notably, six of these patients had the same disease progression dates determined by use of WHO (25% bidimensional increase) and SWOG (50% bidimensional increase) criteria. Since 20% unidimensional increase (RECIST) is equivalent to approximately 44% bidimensional increase, it is likely, although not certain, that disease progression by RECIST unidimensional criteria would have occurred soon after disease progression by SWOG and WHO criteria. For three patients, the difference between the WHO and SWOG 50% bidimensional increase was 10–12 weeks. Again, it is likely, although it cannot be proven, that RECIST criteria would have been met soon after. The remaining four of the 13 patients where difference between WHO and RECIST progression times are uncertain were categorized as progressive disease following SWOG's criteria (5) because of an increase of the tumor surface of greater than or equal to 10 cm^2. For these patients, the magnitude of the difference is entirely uncertain.

2.1. Dataset Evaluated.

The dataset includes 234 patients with progressive disease as defined by the SWOG [5]. All patients had baseline measurable disease followed by the same technique(s) until disease progression. The tumor types included were melanoma, colorectal, lung, and breast cancers.

REFERENCES

1. Zubrod CG, Schneiderman SM, Frei E III, Brindley C, Gold GL, Schnider B, et al. Appraisal of methods for the study of chemotherapy of cancer in man: comparative therapeutic trial of nitrogen mustard and thio phosphoamide. J Chronic Dis 1960;11:7–33.
2. Gehan E, Schneidermann M. Historical and methodological developments in clinical trials at the National Cancer Institute. Stat Med 1990;9: 871–80.
3. WHO handbook for reporting results of cancer treatment. Geneva (Switzerland): World Health Organization Offset Publication No. 48; 1979.
4. Miller AB, Hogestraeten B, Staquet M, Winkler A. Reporting results of cancer treatment. Cancer 1981;47:207–14.
5. Green S, Weiss GR. Southwest Oncology Group standard response criteria, endpoint definitions and toxicity criteria. Invest New Drugs 1992;10:239– 53.
6. James K, Eisenhauer E, Christian M, Terenziani M, Vena D, Mudal A, et al. Measuring response in solid tumors: unidimensional versus bidimensional measurement. J Natl Cancer Inst 1999;91:523–8.

Chapter 4

Measuring the Clinical Response. What does it mean?

Therasse P.

Eur J Cancer. 38(14):1817-23, 2002.

ABSTRACT

The clinical response to treatment is an important indicator of the therapeutic effect of anticancer agents. Its value and interpretation has to be carefully considered within the context that it is used. In daily practice, response assessment is combined with other indicators of the patient's condition to contribute to the decision-making process.

In clinical trials, it is widely used to identify and quantify the anti-tumour activity of new agents. In this context, response evaluation is conducted on the basis of strict predefined criteria such as the World Health Organization (WHO) or Response Evaluation Criteria In Solid Tumors (RECIST) criteria. The RECIST criteria have recently been proposed and offer a detailed guidance to perform a response evaluation. Clinical response is also used as an indicator of therapeutic efficacy in combination with other indicators. Its value as a surrogate indicator of a survival benefit remains unclear in most instances and can hardly be established within the framework of a single randomised trial.

With the development of new anticancer agents that behave differently to cytotoxics, clinical benefit will have to integrate concepts of disease stabilisation or time to progression. Over the next decade, oncologists will be able to assess the biological response before the clinical response, and a lot of work and energy will have to be dedicated to assess the predictive and, possibly, the prognostic value of the biological response with regard to the clinical response, as well as more definitive measures of clinical benefit.

1. INTRODUCTION

In cancer management, evaluating the therapeutic effect of anticancer therapy is a process undertaken every day by oncologists. In most instances, decisions to continue, change or stop systemic therapy are driven by the response to treatment recorded for each patient.
The same approach is also applied under predefined conditions to test new anticancer agents in clinical trials and quantify their level of antitumour activity. Sometimes, it also contributes with other indicators to define the real clinical benefit (the efficacy) provided to the patients with new therapeutic strategies.

Because the clinical response is the only indicator readily available to evaluate the therapeutic effect of anticancer treatment, many oncologists are tempted to use this indicator as a surrogate of long-term clinical benefit for the patients. Unfortunately, such correlation between response and long-term benefit has rarely been demonstrated.

The methodology used to evaluate the response to treatment has also substantially evolved over the past decades, starting from a complete subjective evaluation reported by the treating physician [1] to move to a complex set of objective criteria attempting to standardise the response evaluation process [2–6].

More recently, several new classes of anticancer agents have been discovered. These new drugs often operate through different mechanisms than those previously developed inducing massive cell kill. Accordingly, the methodology used to evaluate clinical response will not only require adaptation to use new tools and techniques to monitor response to treatment, but may also require a subtle different approach to monitor the therapeutic effect of these new classes of anticancer agents.

2. MEASURING RESPONSE TO TREATMENT IN DAILY PRACTICE

In the daily practice of oncologists, the clinical response reported after each patient's visit results from the combination of different indicators out of which the most important are the response to treatment of the anatomical indicators (tumour lesions), the biological indicators (tumour markers and biochemistry) and the patient's condition.

The clinical response so reported, directly contributes to the evaluation of the risk/benefit ratio procured by a certain treatment which also takes into account the subjective opinion of the physician and the patient's performance (Fig. 1).

In this setting, criteria used to determine the clinical response must be adapted to the real life, taking into account the socio-economical constraints (costs, insurance and resources) and the comfort of the patient. The global risk/benefit ratio of a certain standard treatment is supposed to be known (from previous clinical trials) and the role of the physician is to adjust the treatment to the specific conditions of each patient.

In this context, it is obvious that rigorous criteria to define the clinical response should not be applied systematically. It is important, however, to rely on robust and reliable assessments to support important decisions such as initiating or changing systemic therapy.

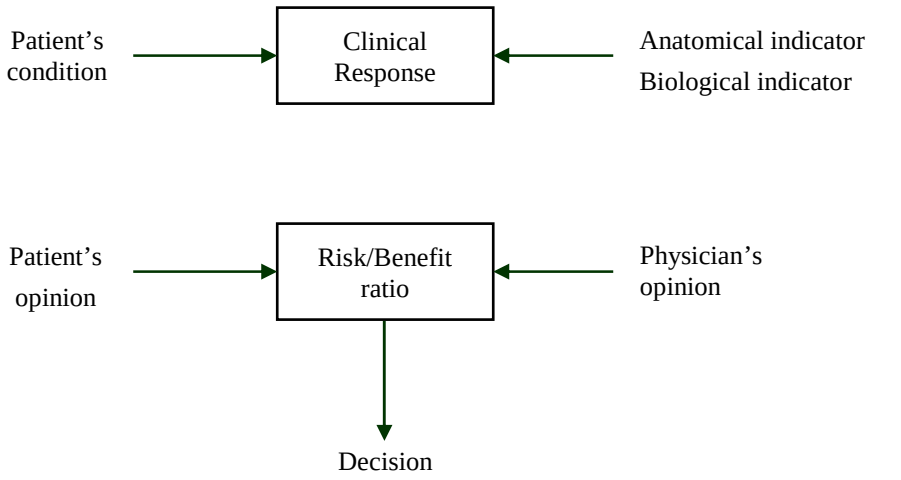

Fig. 1. The role of clinical response in daily practice.

3. MEASURING THE CLINICAL RESPONSE TO DETERMINE THE ANTITUMOUR ACTIVITY OF NEW ANTICANCER AGENTS

The evaluation of the response rate to determine the level of antitumour activity of new anticancer agents or new combination of existing agents is performed in clinical trials. These are usually phase II clinical trials with the determination of the response rate being the main endpoint. Several statistical standard designs (Gehan, Simon, Fleming) [7] are used to identify and also quantify the biological antitumour activity of anticancer agents. Such evaluation can be done qualitatively against the natural history of the disease or quantitatively against a known level of antitumour activity provided by standard therapeutic strategies.

Measuring the clinical response to determine the antitumour activity of new anticancer agents requires the application of rigorous criteria for several reasons mentioned below.

1. The investigation is conducted on a small cohort of patients which decreases the chance of treating a large cohort of patients with an inactive drug, but increase the probability of uncontrolled biases.
2. The therapeutic window of anticancer agents is usually narrow and the conditions under which a new treatment is tested should be optimised to avoid disregarding a potentially active drug.
3. The evaluation of the response rate in phase II clinical trials is a critical step in the development of new agents and just precedes the large clinical trials involving

thousands of patients and the submission of a registration file to regulatory authorities.

4. Often, several parallel identical investigations are conducted. It is expected that the results of these clinical trials will all show the same outcome. Reproducibility of the results will be assessed by regulatory authorities.
5. Finally, the methodology on which the evaluation of clinical response has been developed is fragile, complex and relatively imprecise with significant interobserver variability. Rigorous criteria should be used to decrease the probability of errors, misinterpretation and potential harm to future patients.

It is indeed true that the methodology used to evaluate the clinical response under this setting does not correspond to 'real life' conditions. Nevertheless, it is, so far, the best methodology that has been used successfully to screen anticancer agents and select those to be further investigated in a large cohort of patients to determine their efficacy. The weakness of the methodology as mentioned above can be illustrated by several examples.

Table 1 illustrates, for three different tumour types, the variation in response rate reported by the investigator compared with the response rate reported after response review. It is interesting to note that not only the response rate is systematically lower after response review, but also that the percentage of response downgraded is substantial. Moreover, Thiesse and colleagues have identified the causes leading to errors which could be attributed, in his series of 489 patients with renal cell carcinomas, to error in tumour measurement in 45% of the cases, error in the selection of target lesions in 45% of the cases, and in 10% of the cases errors that were relatively independent of the evaluator.

Table 1. Response rate reported by different authors before and after independent review

Author	Tumor Type	Patients (N)	RR before review (%)	RR after review (%)	% of responses downgraded
Gwyther (8)	NSCLC	374	30	21	NA
Biganzoli (9)	MBC	564	36	30	19
Thiesse (10)	RCC	489	17	13	23

RR = response rate; NSCLC = non-small cell lung cancer; MBC = metastatic breast cancer; RCC = renal cell carcinoma; NA = not available.

Table 2 illustrates the variation reported by Schrijvers and Vermorken [11] for phase II clinical trials investigating single agent or combination of agents in head and neck tumours. It is clear that these findings illustrate not only random variations that cannot be easily controlled for in phase II, trials but also variations as to how the response rate and the individual response for each patient have been evaluated.

Table 2. Reported ranges of response rate for single agent and combination chemotherapy in head and neck cancer

Single Agent		Combination	
Agent	**RR (%)**	**Agent**	**RR (%)**
Paclitaxel (Taxol)	20 – 43	Paclitaxel + Cisplatin	33 – 77
Docetaxel (Taxotere)	37 – 45	Paclitaxel + Carboplatin	23 – 62
Topotecan	0 – 14		
Vinorelbine (Navelbine)	11 – 25		

RR = response rate.

Table 3 illustrates the small cohort effect of phase II clinical trials. Although the examples mentioned are both related to breast cancer, the conclusion that response rates seen in phase II trials are systematically higher than those reported in phase III trials (under the same conditions) can be generalised to all tumour types.

Table 3. Response rate reported by different authors for phase II and phase III investigating the same regimens

Author (Ref.)	Development phase	Patients (N)	Response rate (%)
	Adriamycin + Paclitaxel in metastatic breast cancer		
Gianni (12)	Phase II	35	94
Biganzoli (13)	Phase III randomized	138 (1 arm)	58
	Epirubicin + Cyclophosphamide + G-CSF in locally advanced breast cancer		
Piccart (14)	Phase II	29	87
Therasse (15)	Phase III randomized	220 (1 arm)	57

G-CSF = Granulocyte-colony stimulating factor

These differences in response rates between phase II and phase III trials can be attributed to many different causes. Some of these causes can probably be controlled for such as the selection criteria (ensuring the homogeneity of the population being studied) and the compliance with the protocol (precision in measurement and follow-up). However, these factors alone can hardly explain the large differences observed between response rates.

3.1. The methodology to evaluate the antitumour activity

In the first clinical trial in solid tumours initiated in the 1950s, tumour response was already taken as an endpoint based on the subjective evaluation reported by the physicians [1].

By the end of the 1970s, a group of breast cancer specialists, under the auspices of the International Union against Cancer (UICC), set the principles under which response to treatment in breast cancer should be evaluated [2]. This work was subsequently adopted and integrated into the recommendations set by the World Health Organization (WHO) for the evaluation of cancer treatment in solid tumours [3]. The principles of response evaluation, which are still valid today, can be summarised as follows:

- The overall cancer burden can be characterised by a quantitative evaluation of tumour lesions, which are measurable, and a qualitative evaluation of tumour lesions, which are not measurable.
- The combination of the quantitative and qualitative evaluations provides an estimation of the treatment effect characterised by one of the following four categories: complete response, partial response, stable disease and progression.

The specificity of the WHO/UICC criteria, amongst others, is:

- Measurable lesions are characterised either by their surface (bidimensional lesions, product of longest perpendicular diameters) or by the longest diameter (unidimensional lesions) when only one dimension could be accurately measured.
- The tumour load is evaluated for each organ independently and the overall response to treatment results from the combination of the response observed in each organ.
- Partial response is attributed when a decrease of 50% of the entire tumour burden (objectively for measurable lesions and subjectively for others) is recorded. Progression status is assigned when there is an increase of 25% of the entire tumour burden (based on the same principles as for PR).

After 1981, many non-anticancer drugs have been developed, and many researchers have also started to investigate different combinations of treatments. The experience acquired over the years and the lack of details in the WHO recommendations has stimulated the development and the use of the amended version of the WHO criteria. For example, the South West Oncology Group (SWOG) published their version of the WHO criteria in 1990 [4], with a different cut-off point to define the progression status. In addition, the European Organization for Research and Treatment of Cancer (EORTC) published its version of the WHO criteria (5) defining minimum sizes for lesions from different organs to be considered as measurable.

Over the years, the use of the different versions (published and unpublished) of the original WHO criteria have rendered the accuracy of the comparison of results of identical investigations very unreliable. In addition, the evolution of cancer imaging, the importance given to the response rate endpoint and the increasing number of new anticancer agents to be tested, required a coordinated effort to review the existing criteria and attempt to 're-harmonise' the methodology throughout the entire oncology community.

This difficult exercise started in 1996 under the leadership of three research organisations: EORTC, National Cancer Institute of the United States (NCI US) and National Cancer Institute Canada—Clinical Trials Group (NCIC CTG). A comprehensive revised version of the WHO criteria was published by the group in February 2000 [6] under the acronym of RECIST (Response Evaluation Criteria In Solid Tumors). Within 2 years, these criteria have been adopted by most research groups, the pharmaceutical industry and the regulatory agencies.

The specificity of the RECIST criteria can be summarized as follows:

- All measurements of tumour lesions are based on the longest diameter only (unidimensional measurement).
- Cancer lesions are measurable when their longest diameter is >2 cm when measured with conventional techniques or >1 cm when measured by spiral computed tomography (CT) scan.
- Precisions are given as to which method of measurement can be used and how it can be used.
- The overall tumour burden is represented by selected target lesions and all other lesions are recorded, but not measured.
- Changes in the sum of the longest diameters of all target lesions will define the status of partial response and stable disease.
- Partial response status is defined when the sum of the longest diameters of all target lesions has decreased by 30% or more.
- Response status should be confirmed with a minimum interval of 4 weeks. Stable disease is defined following an interval between two measurements that is protocol-specific (depending on the disease being studied).
- Progression status is defined by an increase of 20% of the sum of the longest diameters of all target lesions or by a non-equivocal progression in non-target lesions or by the appearance of a new lesion.
- Precisions are provided as to how to combine the results of the evaluation of target and non-target lesions and define the overall response.

- Precisions are provided as to how to interpret successive evaluations to define the best overall response.
- CA125 can be used as an indicator that alone may determine a progression of the disease after first-line treatment in advanced ovarian cancer.

The use of one dimension only to measure tumour lesions has been based on the work published by James and colleagues [16]. Retrospective analysis using a cohort of patients from 14 different studies (>4000 patients) demonstrated that using two dimensions or one dimension for tumour lesions measurement did not change the response rate of each individual study. The rate of progression may be slightly different (lower with the RECIST criteria) since a larger difference in tumour growth is required to define disease progression.

Although the RECIST criteria have been launched in 2000, the harmonisation process of the response criteria has continued through the set-up of a Questions and Answers section on the web (eortc.recist.be). In addition, proposals to modify the existing criteria or add new criteria are considered regularly by the RECIST working group. Adaptation of the RECIST criteria are currently studied for specific tumour types such as brain tumours, mesothelioma and pelvic tumours.

Beside the RECIST criteria developed to be applicable to most solid tumours types, standard criteria have been developed for evaluating response and progression in specific settings such as non-Hodgkin's lymphoma [17] and prostate cancer [18].

4. CLINICAL RESPONSE AS AN INDICATOR OF TREATMENT EFFICACY IN PHASE III CLINICAL TRIALS

How can we demonstrate the efficacy of a new treatment in oncology? In other words, do we need to demonstrate an improvement in long-term survival? An improvement in time to progression? An improvement of quality of life? A better control of the symptoms of the patient or perhaps simply an improvement in the clinical response rate? All these endpoints (either primary or secondary objectives) of clinical trials are potential valid indicators of treatment efficacy when they are directly related to an improvement of the risk/benefit ratio for the patients.

One of these endpoints can be preferred to the others according to the context of the study (early disease versus advanced disease, symptomatic versus non-symptomatic disease and so on...). In the advanced setting, when patients usually present measurable lesions, time to progression and/or clinical response rate is often taken as primary endpoint of the trial. The latter, of course, provides a certain advantage over the other in as much as it can provide an answer relatively quickly to the question being investigated.

In this context, one may question the real meaning of clinical response in terms of improvement of the risk/ benefit ratio. This is not an easy question and it requires, a priori,

some clarification over the relationship between clinical response and long-term benefit or in other words to what extent is clinical response a surrogate indicator of another measure of treatment efficacy? It is well known that, in the long-term, those who usually respond to the treatment will do better than the others. This is what oncologists observed in their daily practice and this is also what subset analysis of clinical trials can demonstrate (comparing the survival of the responders with the non-responders). Indeed, responders do better, but does it make clinical response a valid surrogate of survival? The answer to that question is no. Making such a correlation between clinical response and survival is a well known pitfall [19]. In this particular case, selecting responders to analyse and compare their survival with the group of patients that does not respond to treatment is in fact equivalent to selecting a group of patients with specific characteristics including known and unknown prognostic factors that can influence the outcome, both in terms of response to treatment and in terms of survival for this subgroup of patients [20]. However, the effect produced by the treatment on the tumour can be mediated through different bio-molecular mechanisms for response and survival.

One specific outcome measure (such as clinical response) can be considered as a true surrogate indicator of another outcome measure (such as survival) only when the effect of the treatment on the surrogate can reliably predict the effect of the treatment on the final clinical outcome [21]. In other words, the treatment effect on the disease globally should be entirely mediated through the effect seen on the surrogate marker. Such a correlation has rarely been demonstrated in all disciplines of medicine. First of all, it requires very large data-sets of clinical data of patients treated under relatively identical conditions. Moreover, the statistical methodology deployed to prove such correlation and remove all potential confounding factors is extremely complex.

In oncology, several groups have attempted to demonstrate such a correlation between response and long-term outcome in breast [22], ovarian [23], nonsmall cell lung [24] and colorectal cancers [25]. Only in one study [25] could response to treatment be identified as an independent prognostic factor that predicts survival. That study by Buyse and colleagues also highlights a number of important points such as:

- The relationship between response and survival may depend on the drug, the schedule and the dose for the same disease.
- Large improvements in clinical response are needed to achieve a meaningful improvement in survival. In metastatic colorectal cancer, a twofold increase in response rate corresponded to a 1.12 increase in survival. In addition, the overall correlation between clinical response and survival may be substantially influenced by the rate of complete response (and even more by the rate of complete pathological response), but also by the efficacy of second-line treatments (the better it is, the more difficult it is to establish the relationship).

The main conclusion of this exercise was that, for individual trials, the response rate alone cannot realistically predict the benefit for survival. However, a good clinical response rate

should trigger phase III trials with more definitive outcome measures and should also encourage an extension of the drug development in the adjuvant setting, even without a survival benefit in the advanced setting.

5. CLINICAL RESPONSE AS A MEASURE OF EFFICACY

Using clinical response as a direct measure of treatment efficacy may be relevant under specific conditions [26]. The assumption that clinical response may indicate a certain clinical benefit will not only depend on the observed response rate and the degree of improvement over the existing standards of treatment, but should also take into consideration other characteristics of the responses observed and the drug studied.

The average duration of response, the rate of complete response (and in particular complete pathological response) and the localisation of sites responding to treatment are important characteristics. The pharmacological profile of the drug (and in particular the toxicity profile) together with the previous experience observed with the same class of drugs and in the same population are also important.

Finally, the reproducibility of the response rate in other clinical trials should confirm the overall trend observed with a particular drug. It is clear that in this setting no firm rules can be established and the opportunity to use clinical response alone or in association with other indicators of clinical benefit should be considered on a case by case basis.

6. MEASURING CLINICAL RESPONSE IN PHASE III CLINICAL TRIALS

In phase III clinical trials attempting to demonstrate a definitive efficacy advantage of the treatment being studied, the response rate is usually used as a secondary endpoint which may on the one hand support the primary endpoint and, on the other hand, may also be used to adjust the response rate reported from the phase II data under conditions which are usually closer to real life.

Under these conditions, the evaluation of response in phase III clinical trials may not require the same rigour as for phase II trials aiming at determining the degree of antitumour activity. More flexibility could be considered for specific requirements such as the necessity for confirmation, the number of selected target lesions, and the necessity for response review. However, when the response rate is used as a primary endpoint (which usually implies that sample size calculations are driven by a target difference in response rate between the two treatments) a rigorous methodology, as used for phase II trials should be used.

Trials with this main objective can indeed provide a rapid answer in comparison with those using time to event endpoints. However, they require much more resources and time from all of the involved parties and they are usually under-powered to assess more definitive objectives listed as secondary endpoints. The literature is overloaded with small

inconclusive phase III trials that were developed and conducted using the response rate as their main endpoint.

7. FUTURE DEVELOPMENT

Changes in the methodology developed to evaluate the clinical response will not only depend on progress in cancer imaging, but will also be affected by the 'new' classes of anticancer agents that are under development [27,28]. Amongst these new drugs, those having a biological antiproliferative effect inducing delays in tumour growth should be carefully evaluated. It is conceivable that these agents might not systematically generate rapid tumour regression (and therefore measurable response), but may simply result in stabilisation of the disease or even may decrease the rate of tumour growth. For these agents, more attention will have to be given to stable disease and time to progression than to a pure clinical response depending on tumour shrinkage. However, although the endpoints of clinical trials may have to be revised, shifting from response/survival towards progression/stabilisation, the methodology used to assess response remains valid for assessing stabilisation and/or progression.

In the future, molecular responses will first be considered based on functional imaging providing early indicators of antitumour activity. Techniques for documenting anticancer effects will be targeted to follow the mechanism of the anticancer agents tested such as: Positron Emission Tomography (PET) scanning to monitor the glucose metabolism and indirectly the proliferative activity; Magnetic Resonance Spectroscopy (MRS) to analyse the cellular energetics and the membrane turnover; Magnetic Resonance Imaging (MRI) to document tumour perfusion, vascularity and permeability and molecular imaging to follow the intracellular signalling pathways, as well as to monitor gene and drug delivery.

However, even the most developed of these techniques still requires validation with regards to the interpretation of the results in terms of 'response to treatment'. Large correlative studies with the current criteria and standard clinical outcome measures will be required. When all of these issues have been solved, cost issues will have to be addressed to confirm the cost-effectiveness of these new techniques before they can be implemented in clinical trials and routine care.

REFERENCES

1. Zubrod CG, Schneiderman SM, Frei E III, et al. Appraisal of methods for the study of chemotherapy of cancer in man: comparative therapeutic trial of nitrogen mustard and thio phosphamide. J Chronic Dis 1960, 11, 7–33.

2. Hayward JL, Carbone PP, Heuson JC, et al. Assessment of response to therapy in advanced breast cancer: a project of the Programme on Clinical Oncology of the International Union Against Cancer, Geneva, Switzerland. Cancer 1977, 39, 1289–1294.

3. Miller AB, Hogestraeten B, Staquet M, Winkler A. Reporting results of cancer treatment. Cancer 1981, 47, 207–214.

4. Green S, Weiss GR. Southwest Oncology Group standard response criteria, endpoint definitions and toxicity criteria. Invest New Drugs 1992, 10, 239–253.

5. Tumor eligibility and response criteria for phase II and III studies. In EORTC Data Center Manuel. EORTC, Brussels, 1992 (monograph).

6. Warr D, McKinney S, Tannock I. Influence of measurement error on assessment of response to anticancer chemotherapy: proposal for new criteria of tumor response. J Clin Oncol 1984, 2, 1040–1046.

7. Kramar A, Potvin D, Hill C. Multistage designs for phase II clinical trials: statistical issues in cancer research. Br J Cancer 1996, 74, 1317–1320.

8. Gwyther SJ, Aapro MS, Hatty SR, Postmus PE, Smith IE. Results of an independent oncology review board of pivotal clinical trials of gemcitabine in non-small cell lung cancer. Anticancer Drugs 1999, 10, 693–698.

9. Biganzoli L, Lohrisch C, Paridaens R, Therasse P, Piccart M. Analysis of two EORTC trials in metastatic breast cancer (MBC) support the recommendation of external response review (ERR) in trials with response rate (RR) as primary endpoint. Eur J Cancer 36(Suppl. 5), S90 (abstr 227).

10. Thiesse P, Ollivier L, Di Stefano-Louineau D, et al. Response rate accuracy in oncology trials: reasons for interobserver variability. J Clin Oncol 1997, 15, 3507–3514.

11. Schrijvers D, Vermorken JB. Update on the taxoids and other new agents in head and neck cancer therapy. Curr Opin Oncol 1998, 10, 233–241.

12. Gianni L, Munzone E, Capri G, et al. Paclitaxel by 3-hour infusion in combination with bolus doxorubicin in women with untreated metastatic breast cancer: high antitumor efficacy and cardiac effects in a dose-finding and sequence-finding study. J Clin Oncol 1995, 13, 2688–2699.

13. Biganzoli L, Cufer T, Bruning P, et al. Doxorubicin (A)/Taxol (T) versus doxorubicin/Cyclophosphamide (C) as first line chemotherapy in metastatic breast cancer (MBC): a phase III study. Proceedings ASCO 2000 (abstr 282).

14. Piccart MJ, Bruning P, Wildiers J, et al. An EORTC pilot study of filgrastim (recombinant human granulocyte colony stimulating factor) as support to a high dose-intensive epiadriamycin-cyclophosphamide regimen in chemotherapy-naive patients with locally advanced or metastatic breast cancer. Ann Oncol 1995, 6, 673–677.

15. Therasse P, Mauriac L, Welnicka M, et al. Neo-adjuvant dose intensive chemotherapy in locally advanced breast cancer (LABC): an EORTC-NCIC-SAKK randomized phase III study comparing FEC (5FU, epirubicin, cyclophosphamide) vs high dose intensity EC+G-CSF (Filgrastim). Proceedings ASCO 1998 (abstr 472).

16. James K, Eisenhauer E, Christian M, et al. Measuring response in solid tumors: unidimensional versus bidimensional measurement. J Natl Cancer Inst 1999, 91, 523–528.

17. Cheson BD, Horning SJ, Coiffier B, et al. Report of an international workshop to standardise response criteria for non-Hodgkin's lymphomas. J Clin Oncol 1999, 17, 1244–1253.

18. Bubley GJ, Carducci M, Dahut W, et al. Eligibility and response guidelines for phase II clinical trials in androgen-independent prostate cancer: recommendations from the prostate-specific antigen working group. J Clin Oncol 1999, 17, 3461–3467.

19. Weiss GB, Bunce H III, Hokanson JA. Comparing survival of responders and nonresponders after treatment: a potential source of confusion in interpreting cancer clinical trials. Controlled Clin Trials 1983, 4, 43–52.

20. Fleming TR, DeMets DL. Surrogate endpoints in clinical trials: are we being misled? Ann Intern Med 1966, 125, 605–613.

21. Buyse M. On the relationship between response to treatment and survival time. Stat Med 1996, 15, 2797–2812.

22. A'Hern RP, Ebbs SR, Baum MB. Does chemotherapy improve survival in advanced breast cancer? A statistical overview. Br J Cancer 1988, 57, 615–618.

23. Torri V, Simon R, Russek-Cohen E, Midthune D, Friedman M. Statistical model to determine the relationship of response and survival in patients with advanced ovarian cancer treated with chemotherapy. J Natl Cancer Inst 1992, 84, 407–414.

24. Chen TT, Chute JP, Feigal E, Johnson BE, Simon R. A model to select chemotherapy regimens for phase III trials for extensive-stage small-cell lung cancer. J Natl Cancer Inst 2000, 92, 1601–1607.

25. Buyse M, Thirion P, Carlson RW, Burzykowski T, Molenberghs G, Piedbois P. Relation between tumor response to first-line chemotherapy and survival in advanced colorectal cancer: a meta-analysis. Lancet 2000, 356, 373–378.

26. Pazdur R. Response rates, survival, and chemotherapy trials. J Natl Cancer Inst 2000, 92, 1552–1553.

27. Buthiau D, Khayat D. New imaging techniques and clinical applications: visualizing pathology's course. Helix 2000, 3, 4–13.

28. Rowinsky EK. The pursuit of optimal outcomes in cancer therapy in a new age of rationally designed target-based anticancer agents. Drugs 2000, 60(Suppl. 1), 1–14.

Chapter 5

RECIST vs. WHO: Prospective comparison of response criteria in an EORTC phase II clinical trial investigating ET-743 in advanced soft tissue sarcoma

Therasse P, Le Cesne A, Van Glabbeke M, Verweij J, Judson I for the EORTC Soft Tissue and Bone Sarcoma Group.

Eur J Cancer. 2005 Jul;41(10):1426-30.

ABSTRACT

The present study was set up just after the publication of the response evaluation criteria in solid tumors (RECIST) as a prospective validation exercise in soft tissue sarcoma. Forty-nine patients were entered into a phase II clinical trial aiming at determining the activity and safety of ET-743 (Ecteinascidin) in second line advanced soft tissue sarcoma. Response to treatment and progression were monitored following the WHO criteria and RECIST.

Discordances between WHO and RECIST criteria for the best response were reported for two cases: one no-change (WHO) reported as partial response (RECIST) and one progression (WHO) reported as no-change (RECIST). In terms of date of progression, 3 patients progressed on WHO criteria while they were still stable with RECIST. Overall the results of the study would not have changed if RECIST had been used instead of WHO criteria.

In conclusion, response criteria as defined by RECIST are adequate to measure response and progression in non-GIST soft tissue sarcoma and can be used instead of the modified WHO criteria.

1. INTRODUCTION

Response evaluation criteria in solid tumors (RECIST) was introduced in February 2000 to facilitate and improve the evaluation and the reporting of responses in early clinical trials aiming at determining the level of anti-tumor activity of new anti-cancer agents [1]. The new criteria gave much more precision as to how tumor lesions should be assessed and how responses should be reported, also taking into account modern imaging techniques. RECIST uses a uni-dimensional measure (the longest diameter) to quantify measurable tumor lesions as opposed to the bi-dimensional method (cross-sectional longest diameters) usually employed with most other sets of response criteria [2–4]. On the basis of previous studies [3,5], RECIST defines measurable lesions as lesions with a minimum size depending on the method of investigation. Following a principle already implemented in the SWOG response criteria [3], the rules defining objective progression were voluntarily scaled down as compared to the WHO criteria so that the increase in measurable overall tumor burden should be greater with RECIST (20% in one dimension is equivalent to 44% in 2 dimensions) than with WHO criteria (25% in 2 dimensions) to qualify for progression. Following this last criterion, there was some concern that time to progression could be longer using RECIST as opposed to WHO criteria and this was identified up front as an issue requiring some attention in future trials before drawing definitive conclusions.

The objective measurement of tumor lesions has been used for decades in advanced soft tissue sarcoma to screen new agents or new regimens. The original WHO criteria have been adapted (modified WHO criteria) to improve the accuracy of response assessment in this tumor type [5,6]. The aim of the current study was to test RECIST in a prospective trial in parallel with WHO criteria and establish new references (using RECIST) in this tumor type for future trials if significant differences were identified compared with modified WHO criteria.

2. PATIENTS AND METHODS

The present study was conducted in the framework of a non-randomised phase II study investigating the anticancer activity and safety of Ecteinascidin (ET-743 – a novel tetrahydroisoquinoline compound isolated from the marine ascidian Ecteinascidia turbinata) in pre-treated advanced soft tissue sarcoma. The clinical trial was conducted by the EORTC Soft Tissue and Bone Sarcoma Group (STBSG). After the publication of the RECIST in February 2000, the original clinical trial protocol was officially amended to extend the sample size and collect information prospectively and in parallel about response and progression as assessed both by RECIST and WHO criteria. Patients eligible for entry in the study were required to have histologically proven measurable metastatic or unresectable loco-regional recurrent soft-tissue sarcoma. Mesothelioma, chondrosarcoma, neuroblastoma, osteosarcoma, Ewings sarcoma, embryonal rhabdomyosarcoma and dermatofibrosarcoma were excluded. Patients with gastro intestinal stromal tumors (GIST) were treated in a separate study.

All patients were to have a documented progressive disease at inclusion, with defined target lesions at physical examination, on X-rays and CT scan. For the purpose of this project, the eligibility criteria required the presence of at least one measurable lesion fulfilling the definition of both (modified) WHO criteria and RECIST. The protocol specified that maximum three target lesions per organ and maximum five target lesions overall were to be reported and used for assessing response. WHO criteria were used as reference criteria for therapeutic decisions (discontinuation of treatment). Other eligibility criteria were standard and have been outlined in detail in a previous paper together with the results of the therapeutic activity of ET-743 [7]. ET-743 was administered at a dose of 1.5 mg/m^2 intravenously as a 24 h continuous infusion every 3 weeks using a central venous line.

Response to treatment was evaluated every 2 cycles (every 6 weeks), with repeated clinical and relevant radiological assessments based on disease extension at presentation. For all responding patients, the hospital records and all available films were reviewed by two independent investigators. A response was accepted only if they reached consensus. In the absence of consensus the worst response category was assigned. Patients were considered evaluable for response if they had received a minimum of two cycles of treatment. In case of rapidly progressive disease after one course, the patient was removed from study and classified as treatment failure. If response had not been assessed, patients were included in the following categories: early death from toxicity in case of death occurring within 6 weeks due to signs of toxicity; early death from malignant disease if death occurred within 6 weeks after commencing chemotherapy due to soft tissue sarcoma and without signs of toxicity; a further classification was early death from other cause if death occurred in the same period of a cause not related to malignant disease. Patients who had stable disease or exhibited complete or partial responses remained on treatment until treatment completion (6 cycles), disease progression, unacceptable toxicity or patient refusal. Patients with evidence of drug related clinical benefit were allowed to continue on therapy after 6 cycles.

The Simon two stage design has been separately applied to each patient cohort (one cohort before and one cohort after the amendment) to allow determination of response rates and progression with RECIST. All analyses presented in this paper are exploratory and descriptive and have been produced using VISTA, the software developed by EORTC to handle clinical trial data.

3. RESULTS

Between March 2000 and November 2000, 49 patients were recruited by 7 participating centers. Two patients were initially declared ineligible by the study coordinator for the main efficacy analysis. One patient had a lung target lesion with a longest diameter of 17 mm on CT scan while the selection criteria required at least one target lesion >20 mm and another patient had only one target lesion that had been previously irradiated. However, considering an intent to treat analysis for all patients for whom we had data on both WHO and RECIST evaluations, these patients have been included in the present analysis.

The original localisation of the disease is described in Table 1.

Table 1. Primary sites of disease

	n=49 (%)
Head and Neck	2 (4)
Trunk	7 (14.3)
Visceral intra-abdominal	5 (10.2)
Retroperitoneum	6 (12.2)
Uterus	8 (16.3)
Girdle	5 (10.2)
Lower arm	16 (32.6)

Most of the patients had either one (21 patients/ 42.9%) or two (14 patients/28.6%) different anatomic sites involved (considering target and non-target lesions) and only 10 (20.4%) and 4 (8.2%) patients had 3 or 4 different sites involved, respectively. Twenty-nine patients had only one target lesion and 11 patients had 2 target lesions (Table 2). Target lesions were located in one organ only for 44 patients (Table 3) and the distribution of lesions per organ/system is described in Table 4. Following the modified WHO criteria used for decision making in this protocol 2 patients presented a partial remission (PR), 30 patients achieved no-change (NC) and in 17 patients progressive disease was recorded as best overall response. The comparison of response assessment between WHO criteria and RECIST is described in Table 5. Discordances between WHO criteria and RECIST for the best response were reported for two cases: one no change (NC) (WHO) reported as partial response (PR) (RECIST) and one progressive disease (PD) (WHO) reported as NC (RECIST).

Table 2. Number of target lesions per patients

WHO	**RECIST**					
	1	**2**	**3**	**4**	**5**	**Total**
1	29 (100%)					29 (59.2%)
2		11 (100%)				11 (22.4%)
3			7 (100%)			7 (14.3%)
4				1 (100%)		1 (2%)
5					1 (100%)	1 (2%)
Total	**29**	**11**	**7**	**1**	**1**	**49**

Table 3. Number of target lesions by organ per patient

Lesions	Organs			
	1	**2**	**3**	**Total**
1	29 (65.9%)			29 (59.2%)
2	7 (15.9%)	4 (100%)		11 (22.4%)
3	7 (15.9%)			7 (14.3%)
4			1 (100%)	1 (2%)
5	1 (2.3%)			1 (2%)
Total	**44**	**4**	**1**	**49**

Table 4. Organ/system involved

Involved sites	Any lesions (n=49)[a]	Target lesions (n=49)[a]
Primary	18	11
Lymph nodes	6	3
Lung	33	22
Liver	9	6
Skin	1	-
Other soft tissue sites	16	15
Bone	6	-

[a]Patients may have more than one site involved.

Table 5: Best response to therapy WHO vs. RECIST

WHO	RECIST			
	PR	**NC**	**PD**	**Total**
PR	2			2 (4.1%)
NC	1	29		30 (61.2%)
PD		1	16	17 (34.7%)
Total	**3 (6.1%)**	**30 (61.2%)**	**16 (32.6%)**	**49**

PR = partial response; NC = no change; PD = progressive disease.

The progression status evaluated according to WHO criteria or RECIST is presented in Table 6. In this analysis, 15 patients were not evaluable for the comparison RECIST/WHO. Two patients stopped treatment for toxicity reasons before progression and 13 patients progressed after the end of the planned treatment period and had no comparative measurements recorded at the time of progression. Among the remaining 34 patients, 3 patients were identified as PD following the WHO criteria while they were still stable (NC) following RECIST. For 2 of these patients, therapy was discontinued (as per protocol) at the time of WHO progression. One patient died rapidly and the other patient survived another year. The third patient was continued on therapy for another 6 months despite WHO progression (erroneously reported as NC (WHO) by the investigator but truly NC following RECIST) achieving a partial remission (WHO and RECIST) that remained stable for another year. In the present study, the decision rules set up for the further development of ET-743 would not have been affected if RECIST had been used instead of the modified WHO criteria.

Table 6. Timing of progression with RECIST and WHO criteria

	Progression *n*=49 (%)
Non evaluable	15 (30.6)
Same date of progression	31 (63.3)
- progression with new lesion(s)	18 (58)
- progression by increase of pre-existing of tumor burden	13 (42)
Progression by RECIST after progression by WHO	3 (6.1)

4. DISCUSSION

The present study is interesting for several reasons including that this is the first study prospectively testing both RECIST and WHO criteria in advanced soft tissue sarcoma. Using the response rate to decide whether or not to continue or stop further investigations with ET- 743 the same decision would have been taken whether WHO criteria or RECIST had been used. These decision rules were built into the protocol. However based upon the observed specific character of the anti-tumor activity generated by ET-743 (long lasting absence of progression), in further planning more attention was given to the time to progression and progression rate (or rate of progression arrest) to quantify the activity of ET- 743. As WHO criteria were initially designated as the criteria on which the therapeutic decisions should be taken, it has not been possible to assess and compare progression rates obtained with the two sets of response criteria especially for long lasting disease stabilisation after treatment completion. Should RECIST have been selected as the principal criteria for this study a long and difficult debate would have followed whether the very long time to progression was only due to the use of RECIST instead of WHO criteria, or due to the intrinsic anti-tumor activity of ET-743. This constitutes clearly one of the limitations of this study as is the case for all prospective validation studies published so far [8–15]. This

study, albeit relatively small, suggests that for screening types of trials such as the phase II study design, the simpler RECIST is as satisfactory as the more complex WHO criteria, particularly for development planning. This study does not enable us to assess whether if RECIST was used as principal selection criteria for response, if it would have cut down the number of eligible patients compared to WHO criteria (with no minimum size for tumor lesion) since the WHO modified criteria used in this study (and previous studies in the same tumor type) are even more strict in the selection of patients than RECIST. It is also important to note that, as in many other tumor types, progression is identified with the appearance of a new lesion in a majority of patients (58% in this study) as opposed to an objective increase in existing tumor burden. This confirms that although a relative precision is needed to measure the overall tumor burden, the true impact of measurement errors as well as the importance given to the magnitude of tumor burden increase (25% in 2 dimensions with WHO or 20% in one dimension with RECIST) on the correct estimation of the progression rate is relatively small. In the present study, only three patients (6.1%) were identified as progressing according to the modified WHO criteria while they were still considered as stable using RECIST. The natural history of these three patients (after being identified as progressing following the WHO criteria) supports the concept on which the progression rule in RECIST has been elaborated. That is to say that prolonging the time to progression by requiring a larger increase in tumor burden than with former WHO criteria may have almost no impact on patients truly progressing (the delay between WHO and RECIST progression will be very small). However, it may on the other hand help patients who might still benefit from further treatment and who are therefore less exposed to unfortunate therapeutic decisions based on measurement imprecision or errors.

Even though up to five target lesions could be reported as per protocol the large majority of patients had less than 3 target lesions reported and almost all of them were situated in the same organ/system. It is, however, difficult to interpret these data without knowing the real number of potential target lesions at baseline. One could indeed be victim of under-reporting of target lesions or on the contrary conclude that the problem of having to follow a lot of target lesions (up to five following this protocol) is not a true problem in this tumor type.

Of the eight currently published prospective validation studies comparing WHO criteria to RECIST, four involve primary lung cancer [8,11,13,14], one involves metastatic colorectal cancer [9], and one lung and liver lesions from breast cancer [10], and finally two involve mesothelioma [12,15]. Apart from the mesothelioma studies all indicate a similar outcome in terms of response rates regardless whether RECIST or WHO criteria are used. Because of the particular characteristics of mesothelioma, it can be expected that both RECIST and WHO are not adequate to measure the true tumor burden. Several solutions have been proposed but there is currently no consensus on a preferred system. Three of the studies performed provide information on WHO criteria and RECIST in terms of progression [9,10,13] but as indicated and importantly, none of these studies used RECIST as primary criteria and, therefore, the overall conclusion drawn from comparing RECIST and WHO criteria remains slightly biased. In two studies [9,10], as in the current one, few patients with PD using WHO criteria would still have been considered as stable with RECIST. In

one (small) study [13], there was no difference in time to progression when using either RECIST or WHO criteria. Apart from mesothelioma, all other studies performed confirmed that RECIST (and the uni-dimensional approach) is suitable to measure response and progression. In conclusion, our study confirms that RECIST can be used for decision making in screening studies in soft tissue sarcomas. Putting this study in perspective with other studies in more common tumor types supports the implementation of RECIST as standard criteria for response evaluation but also for monitoring progression.

CONFLICT OF INTEREST STATEMENT

None declared.

ACKNOWLEDGMENTS

The authors thank all contributors from the EORTC

Soft Tissue Sarcoma Group who have contributed to the study and accepted to duplicate their efforts in monitoring their patients to facilitate this study. This publication was supported by the Grant No. 5U10 CA11488-34 from the National Cancer Institute (Bethesda, Maryland, USA). Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the National Cancer Institute.

REFERENCES

1. 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.
2. Miller AB, Hogestraeten B, Staquet M, et al. Reporting results of cancer treatment. Cancer 1981, 47, 207–214.
3. Green S, Weiss GR. Southwest oncology group standard response criteria, endpoint definitions and toxicity criteria. Invest New Drugs 1992, 10, 239–253.
4. Tumor eligibility and response criteria for phase II and III studies. EORTC Data Center Manuel. Brussels, 1992, monograph.
5. Van Glabbeke M, van Oosterom AT, Steward W, et al. Selection of large and objectively measurable target lesions in EORTC phase II trials: impact on recruitment and response rate. EORTC soft tissue and bone sarcoma group (STBSG). Eur J Cancer 1993, 29, 1943–1947.
6. Van Glabbeke M, van Oosterom AT, Oosterhuis JW, et al. Prognostic factors for the outcome of chemotherapy in advanced soft tissue sarcoma: an analysis of 2185 patients treated with anthracycline-containing first-line regimens – a European organization for research and treatment of cancer soft tissue and bone sarcoma group study. J Clin Oncol 1999, 17, 150–157.
7. Le Cesne A, Blay JY, Judson I, et al. Phase II study of ET-743 in advanced soft tissue sarcomas: a European organisation for the research and treatment of cancer (EORTC) soft tissue and bone sarcoma group trial. J Clin Oncol 2005, 23, 576–584.

8. Werner-Wasik M, Xiao Y, Pequignot E, et al. Assessment of lung cancer response after nonoperative therapy: tumor diameter, bidimensional product, and volume. A serial CT scan-based study. Int J Radiat Oncol Biol Phys 2001, 51, 56–61.
9. Trillet-Lenoir V, Freyer G, Kaemmerlen P, et al. Assessment of tumour response to chemotherapy for metastatic colorectal cancer: accuracy of the RECIST criteria. Br J Radiol 2002, 75, 903–908.
10. 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.
11. Erasmus JJ, Gladish GW, Broemeling L, et al. Interobserver and intraobserver variability in measurement of non-small-cell carcinoma lung lesions: implications for assessment to tumor response. J Clin Oncol 2003, 21, 2574–2582.
12. Byrne MJ, Nowak AK. Modified RECIST criteria for assessment of response in malignant pleural mesothelioma. Ann Oncol 2004, 15, 257–260.
13. Konishi K, Kuriyama K, Chino S, et al. CT evaluation of response to chemotherapy and/or radiotherapy in primary lung cancer: comparison of response evaluation criteria in solid tumors (RECIST) and the WHO criteria, and comparison of both methods with the histological evaluation. Nippon Igaku Hoshasen Gakkai Zasshi 2004, 64(1), 41–45.
14. Grossi F, Belvedere O, Fasola G, et al. Tumor measurements on computed tomographic images of non-small cell lung cancer were similar among cancer professionals from different specialties. J Clin Epidemiol 2004, 57, 804–808.
15. van Klaveren RJ, Aerts JG, de Bruin H, et al. Inadequacy of the RECIST criteria for response evaluation in patients with malignant pleural mesothelioma. Lung Cancer 2004, 43, 63–69.

Chapter 6

Magnetic Resonance Imaging of the Axial Skeleton Enables Objective Measurement of Tumor Response on Prostate Cancer Bone Metastases

Tombal B, Rezazadeh A, Therasse P, Van Cangh P, Vande Berg B, Lecouvet F.

Prostate. 65(2):178-87, 2005.

ABSTRACT

Background: There is currently no technique to image quantitatively bone metastases. Here, we assessed the value of MRI of the axial skeleton (AS-MRI) as a single step technique to quantify bone metastases and measure tumor response.

Methods: AS-MRI was performed in 38 patients before receiving chemotherapy for metastatic HRPCa, in addition to PSA, computed tomography of the thorax, abdomen and pelvis [CT-TAP] and Tc-99m bone scintigraphy. A second AS-MRI was performed in 20 patients who completed six months of chemotherapy. Evaluation of tumor response was performed using RECIST.

Results: Only 11 patients (29%) had RECIST measurable metastases in soft-tissues or lymph nodes on baseline CT-TAP. AS-MRI identified a diffuse infiltration of the bone marrow in 8 patients and focal measurable metastatic lesions in 25 patients (65%), therefore doubling the proportion of patients with measurable lesions. Transposing RECIST on AS-MRI in 20 patients who completed six months of treatment, allows the accurate estimation of complete response (n=2), partial response (n=2), stable disease (n=5), or tumor progression (n=11), as it is done using CT-TAP in soft tissue solid metastases.

Conclusions: MRI of axial skeleton enables precise measurement and follow-up of bone metastases as it is for other soft-tissue metastasis.

Abbreviations: AS: axial skeleton; CT: computed tomography; HRPCa: hormone resistant prostate cancer; MRI: magnetic resonance imaging; PSA: Prostate Specific Antigen; PCa: Prostate Cancer; RECIST: Response Evaluation Criteria in Solid Tumors; Tc-99m: Technetium-99m.

INTRODUCTION

Prostate Cancer (PCa) is the most frequent cancer in men over the age of 55. PCa is characterized by an exquisite tropism for bone that results in a high incidence of bone metastases [1]. Bone metastases and their complications represent the major cause of pain and death from prostate cancer and account for the largest part of the cost of treatment [2]. Hopefully, a better understanding of the relationship between prostatic cells and the bone microenvironment has lead to the development of new specific therapy to interfere with the spread of bone metastases, *e.g.* as bisphosphonates, and antagonists of the endothelin's receptor [3-6]. Interestingly, while major therapeutic breakthroughs are rapidly concretizing, little progresses are made in the diagnostic and objective characterization of bone metastases, which to some extent might have blurred our appreciation of the activity of such new bone-specific agents. Classically indeed, the activity of a new agent is evaluated by measuring the extent of tumor shrinkage it induced in the tumor or its metastasis, using clinical assessment, CT or MRI imaging [7]. When it comes to bone metastases, which are diagnosed and followed using Tc-99m bone scintigraphy and standard X-Rays or CT scanner, there is no robust imaging technique to measure tumor size and response to therapy. Investigators then rely on the limited proportion of patients with soft tissues' lesions or on composite endpoint embedding objective and non objective measures. This might be sufficient in routine practice, but it becomes much more crucial for the development drugs targeted against bone metastasis [3,5,8]. Magnetic resonance imaging of axial skeleton (AS-MRI) is a reliable method to study the bone marrow of the axial skeleton and to detect primary and secondary bone tumors [9-11]. AS-MRI is routinely used for the work-up of suspected vertebral fractures or spinal cord compression in cancer patients [12,13]. The present study assesses the value of AS-MRI as an objective method to detect and measure bone metastases. The study suggests that AS-MRI is a suitable technique to perform evaluation of tumor response in bone.

MATERIALS AND METHODS

Patient characteristics

This prospective study has enrolled 38 consecutives patients with androgen independent hormone resistant PCa, according to Bubley et al. [14]. All patients were included in prospective phase II or III trials with one of the following agents: estramustine phosphate, irofulven, mitoxantrone, docetaxel. Tc-99m bone scans and CT scans of the thorax, abdomen and pelvis (CT-TAP) were performed prior to inclusion in compliance with the study protocols. Metastatic status was confirmed by a positive CT-TAP and/or a positive bone scan with targeted plain radiographs. In case of measurable lymph nodes or other soft-tissue metastases, tumor quantification was performed using RECIST [15].

Patients were followed-up according to the protocols with CT-TAP and bone scans performed every 12 weeks. Objective assessment of tumor response in soft tissue metastases was performed in patients who received at least six months of continuous therapy. Prostate

Specific Antigen (PSA) was measured every 4 weeks with the Tandem PSA assay® from Hybritech.

MRI of the axial skeleton

Addition of AS-MRI studies to the protocol was approved by our local ethic's committee and informed consent was obtained from all patients. Baseline AS-MRI was performed within two weeks before initiation of treatment. Eighteen patients discontinued treatment within six months; 11 for treatment related toxicity and 7 for death related to disease progression. A second AS-MRI study was obtained to assess tumor response in the remaining 20 patients who completed six months of therapy.

Imaging parameters

AS-MRI investigated the entire spine, pelvis and proximal femurs that contain the vast majority of red (hematopoietic) bone marrow containing areas. AS-MRI was performed on a 1.5 Tesla MR magnet (Philips Intera) using the following parameters:

Cervico-thoracic spine. Sagittal plane, synergy phase-array spine coil, 400 mm field of view (FOV), 448x512 matrix, 9x5 mm thick sections, 0.5 mm interslice gap, with median section on spinous processes. T1-Weighted spin-echo (SE) MR images (TR/TE: 412/20[msec]; acquisition time, 04.00 min) and T2-Weighted fast-spin echo (FSE) MR images (TR/TE: 2500/90; FSE factor: 12; acquisition time, 04.02 min) were systematically acquired.

Lumbo-sacral spine. Same parameters, but both T1-Weighted MR images and T2-Weighted FSE fat-saturated MR images were systematically obtained.

Pelvis and proximal femurs. Coronal plane, body coil, 400 mm FOV, 512x512 matrix. 16x6-mm-thick sections, 0.6 mm interslice gap, covering the pelvis from the anterior cortices of the pubis to the cortices of the posterior iliac crests in the AP plane and from the iliac crests to the medio-diaphyseal regions of the femurs. T1-Weighted MR images (TR/TE: 460/16; acquisition time, 03.49 min) and T2-Weighted MR images (TR/TE: 3560/120; TSE factor: 10; acquisition time, 02.58 min) were systematically obtained. Total imaging time varied between 24 and 30 min. All examinations were prospectively stored on optical disks for further review and measurements.

Image analysis and definition of MRI patterns of bone marrow involvement

MRI appearance of the spinal and pelvi-femoral bone marrow was categorized into one of the following well-defined categories [16,17].

Normal. Fig. 1 is characterized by an homogeneous high signal intensity on T1-Weighted MR images, higher than disk and muscle signal intensity, a homogeneous intermediate

signal intensity marrow on T2-Weighted MR images, and a homogeneous low signal intensity marrow on T2-Weighted fat-saturated MR images.

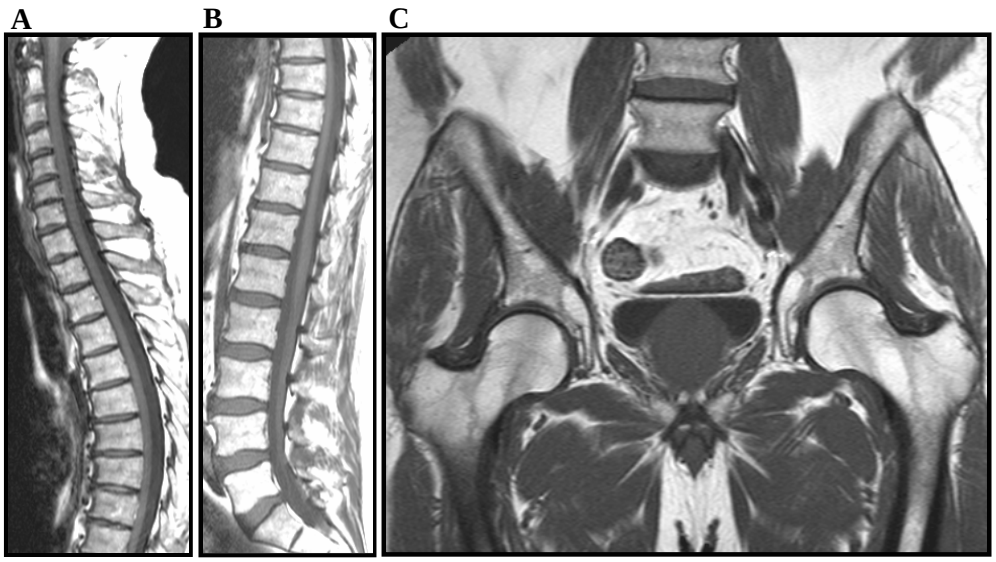

Fig. 1. Sagittal T1-weighted images of the thoracic (A) and lumbar spine (B), and coronal T1- weighted image of the pelvis (C) illustrate normal bone marrow appearance: homogeneous high (bright) signal intensity of the vertebral and pelvi-femoral bone marrow which is much higher (brighter) than the signal intensity of adjacent intervertebral disks and muscles.

Focal metastatic lesions (focal marrow replacement pattern). Fig. 2 has nodular area(s) with a low signal intensity on T1-weighted MR images, a low to intermediate signal intensity on T2-weighted MR images, an intermediate to high signal intensity on T2-weighted fat-saturated MR images. These focal lesions can be measured (cf. infra).

Diffuse marrow infiltration. Fig. 3 is characterized by a homogeneous low signal intensity spinal marrow on T1-weighted MR images, identical to or lower than disk and muscle signal intensity. It shows high signal intensity on fat-saturated T2-weighted MR images, which enables differentiation of this status from benign marrow hyperplasia which can be induced by the treatment and is characterized low signal intensity on both T1 and T2 weighted MR images. Within the pelvis, it is characterized by a low signal intensity marrow on T1-weighted MR images, identical to or lower than muscle signal intensity, with or without expansion of low signal intensity marrow within proximal femurs (which normally contain high signal intensity "yellow" marrow in an adult). Diffuse marrow infiltration pattern is not measurable.

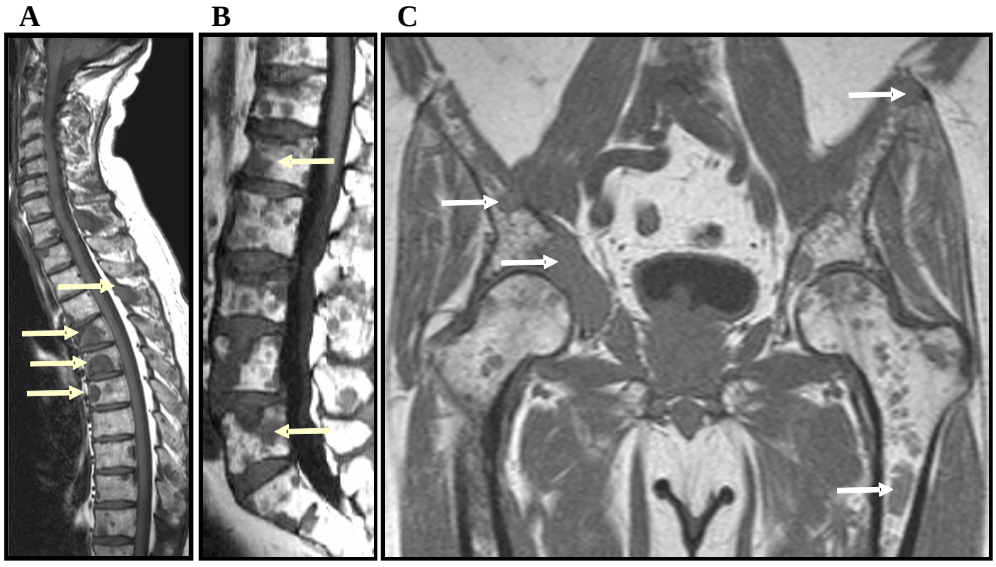

Fig. 2. Sagittal T1-weighted magnetic resonance images of the thoracic (A) and lumbar spine (B), and coronal T1-weighted MR image of the pelvis (C) show multiple focal areas of marrow replacement representing metastases (arrows show several examples); their signal intensity is much lower (darker) than that of adjacent normal bone marrow.

Measurement of bone metastases at baseline and after six months of therapy

Baseline measurement was performed on the MRI examination performed within two weeks before treatment initiation. Lesions were measured using the Philips Easy Vision Workstation software. As required by RECIST, a lesion was considered measurable (target lesions) when the largest maximal diameter was ≥ 10 mm for spinal lesions and ≥ 12 mm in the pelvi-femoral region (≥ 2 x slice thickness to avoid partial volume artifacts)(15). Lesions with smaller diameters were recorded but considered non measurable. If more than ten measurable lesions were found in a patient, only those with the largest diameters were recorded, according to RECIST recommendations. The total amount of lesions and maximal diameter of each lesion were recorded. The same method was used on follow-up MRI studies obtained in 20 patients six months after initiation of the treatment. Securing of the measuring plane between sequential MRI in a single patient is ensured by aligning the examination on the spinous processes of the vertebra for the cervico-thoracic and lumbo-sacral spine and on the anterior cortices of the pubis for the pelvis and proximal femurs. Objective response was assessed in these 20 patients using RECIST, resulting in 4 categories: complete response (CR), disappearance of all lesions; partial response (PR) at least a 30% reduction in the sum of the longest diameters of target lesions, between baseline and follow-up MRI studies; progression (PROG), at least a 20% increase in the sum of the longest diameters of target lesions, taking as reference the smallest sum longest diameter

since the treatment started or the appearance of one or more new lesions; stable disease (SD), neither sufficient decrease nor increase in lesion size to qualify for PR or PROG (15).

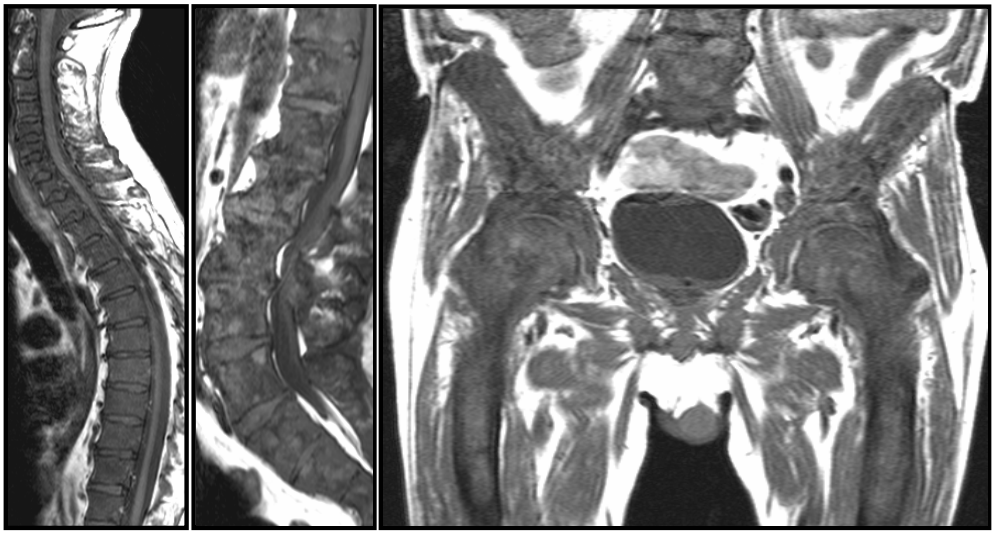

Fig. 3. Sagittal T1-weighted magnetic resonance (MR) images of the thoracic (A) and lumbar spine (B), and coronal T1-weighted MR image of the pelvis (C) show diffuse low (dark) signal intensity of the bone marrow representing diffuse metastatic infiltration.

Assessment of treatment response using PSA and soft tissue lesions

PSA response is defined according to the Bubley guidelines for phase II clinical trials in androgen-independent prostate cancer (14) (PSA normalization: value ≥ 0.2 ng/ml; PSA decrease: a PSA decline ≥ 50%, confirmed by a second PSA value 4 or more weeks later; PSA progression: a ≥ 25% increase over the baseline (and an increase in the absolute value PSA level by at least 5 ng/mL). RECIST were used to define tumor response in patients with measurable soft tissue lesions on CT-TAP. These responses defined on the basis of PSA-levels and of changes in soft-tissue lesions were compared to the response defined on AS-MRI studies.

RESULTS

Patient characteristics at entry

Patient characteristics at baseline are summarized in Table 1. Thirty patients (79%) had a positive Tc-99m bone scan with metastases confirmed by plain radiographs of the positive bone scan areas. All 30 patients had metastases located in the axial skeleton (spine, pelvis,

ribs and/or skull), 13 patients in the appendicular skeleton, and no patient had appendicular lesions only.

Table 1. Patient characteristics at entry

	n	%
Number of patients	38	
Positive bone scans	30	78,9
Positive CT-TAP	13	34,2
- Lymph nodes	13	34,2
- Soft tissue metastasis (liver)	1	2,6
RECIST measurable soft-tissue lesions	11	28,9

Pelvic lymph nodes were identified on CT-TAP in 13 patients (34.2%) and liver metastasis in 1 patient; 11 (28.9%) of 12 having measurable target lesions according to RECIST (*i.e.* ≥ 10 mm longest diameter on CT, given a CT slice thickness of 5 mm) (Table 2).

MRI appearance of the bone marrow on the baseline MRI study

The different patterns of AS-MRI are summarized in Table 2 and illustrated in Fig. 1-3. AS-MRI was considered normal in 5 patients (13 %). Focal metastatic lesions were identified in 22 patients (57.9%), diffuse marrow infiltration involving both the spine and pelvi-femoral areas in 8 (21.1%), and diffuse bone marrow infiltration within the spine but measurable focal lesions in the pelvifemoral area in 3 (7.9%).

Table 2. Bone marrow patterns on initial MRI studies

	n	**%**
No detectable lesions (normal pattern)	5	13.1
Focal lesions (spine ± pelvis)	22	57.9
Diffuse infiltration (spine ± pelvis)	8	21.1
Diffuse infiltration (spine) and focal lesions (pelvis)	3	7.9
Measurable lesions in bone after transposition of RECIST	25	65.7
< 10 lesions	20	52.6
> 10 lesions	5	13.1

Comparison of the proportion of patients with measurable target lesions in soft tissue and in bone.

After unidimensional measurements and transposition of RECIST criteria on AS-MR images, 25 (65.7%) patients had measurable "target" bone lesions including 20 with < 10 lesions and 5 with ≥ 10 lesions (Table 2 and Fig. 4 to 6). Transposition of RECIST to bone metastases using AS-MRI thus increased the proportion of patients with measurable metastatic disease from 28,9 % with measurable soft tissue lesions on CT-TAP to 65,7% with measurable lesions on AS-MRI.

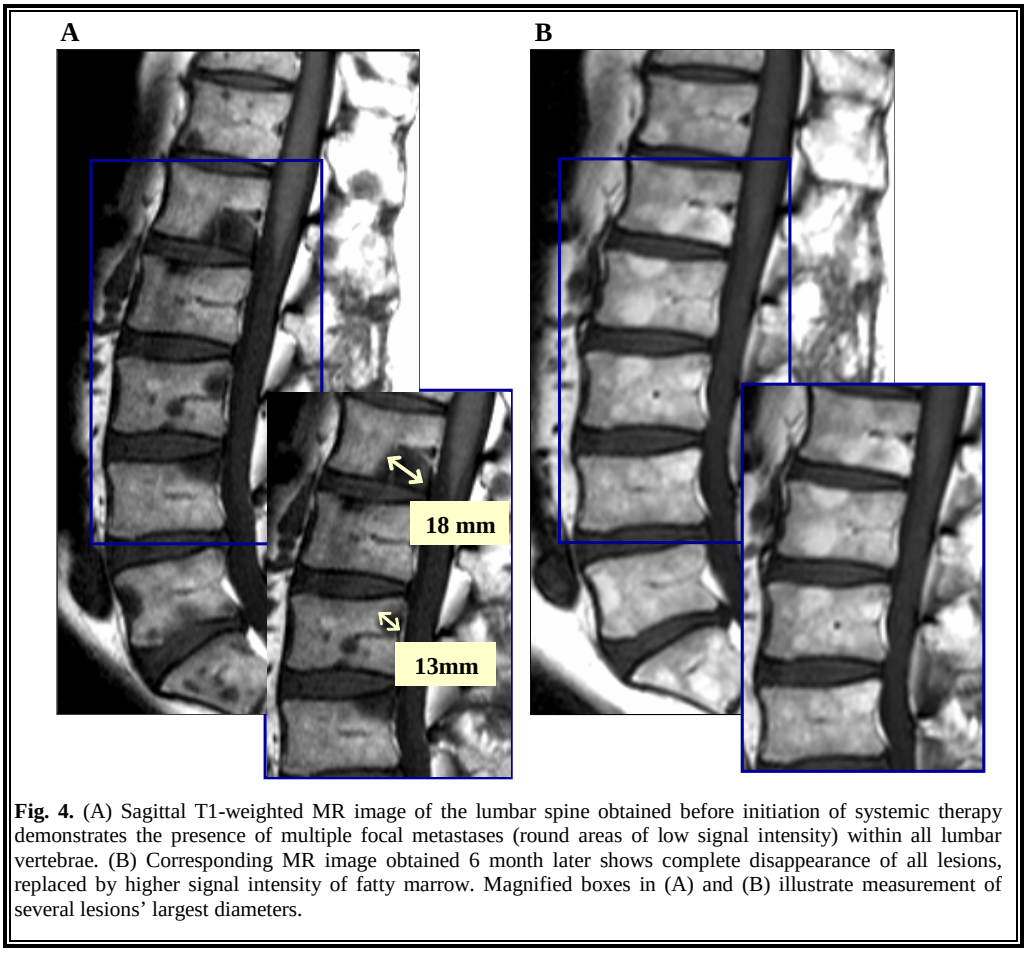

Fig. 4. (A) Sagittal T1-weighted MR image of the lumbar spine obtained before initiation of systemic therapy demonstrates the presence of multiple focal metastases (round areas of low signal intensity) within all lumbar vertebrae. (B) Corresponding MR image obtained 6 month later shows complete disappearance of all lesions, replaced by higher signal intensity of fatty marrow. Magnified boxes in (A) and (B) illustrate measurement of several lesions' largest diameters.

Assessment of response using MRI (Fig. 4-6)

The follow-up AS-MRI studies obtained in 20 patients after six months of therapy were compared to baseline AS-MRI studies to determine patient's individual responses (table 3),

using RECIST transposed to MRI studies. Complete response was observed in 2 patients (Fig. 4), partial response in 2 (Fig. 5), progression in 11 patients (Fig. 6) (9 with increase in size and number of focal metastases; 2 with evolution from focal to diffuse bone marrow infiltration at MRI,) and stable disease in 5 patients.

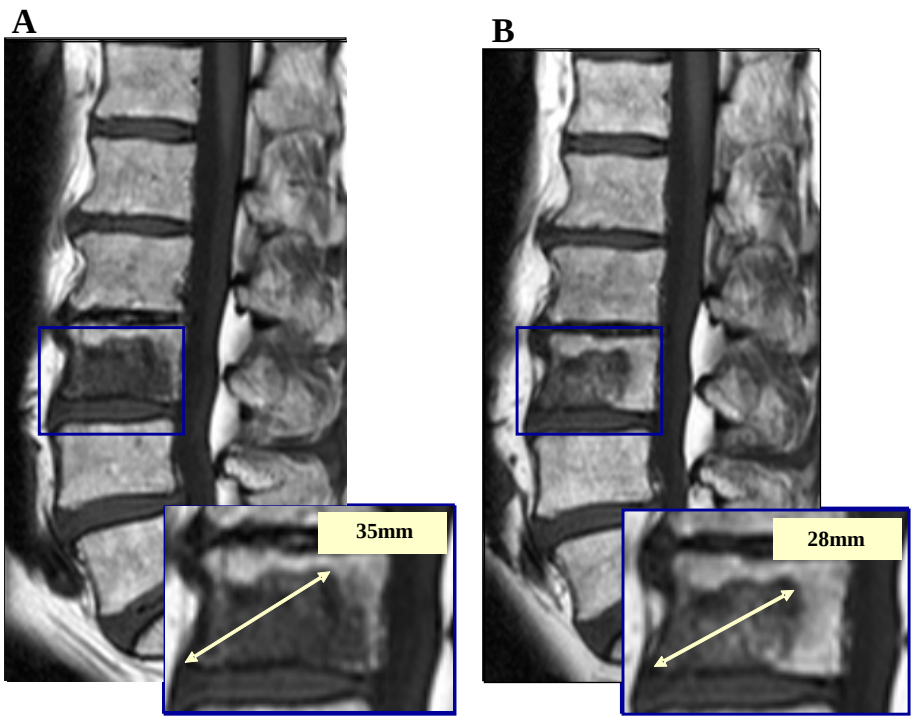

Fig. 5. (A) Sagittal T1-weighted MR image of the lumbar spine obtained before initiation of systemic therapy in a 72-year old patient demonstrates the presence of a focal lesion (round area of low spinal intensity) within the L3-vertebral body. (B) Corresponding MR image obtained 6 month later shows evident decrease in lesion size (partial response). Magnified boxes in (A) and (B) illustrate measurement of lesion's largest diameters.

Correlation with routine evaluation of treatment response (Table 3)

Only five (patients 3, 8, 12, 18 and 20) out of the 20 patients with a AS-MRI follow-up had RECIST measurable soft tissue target lesions identified on CT-TAP. In these five patients, response categorization using AS-MRI and pelvic CT-TAP studies were in perfect agreement (Table 3). Correlation of PSA and MRI evaluation of response showed agreement in 14 of the 20 (70%) patients**.**

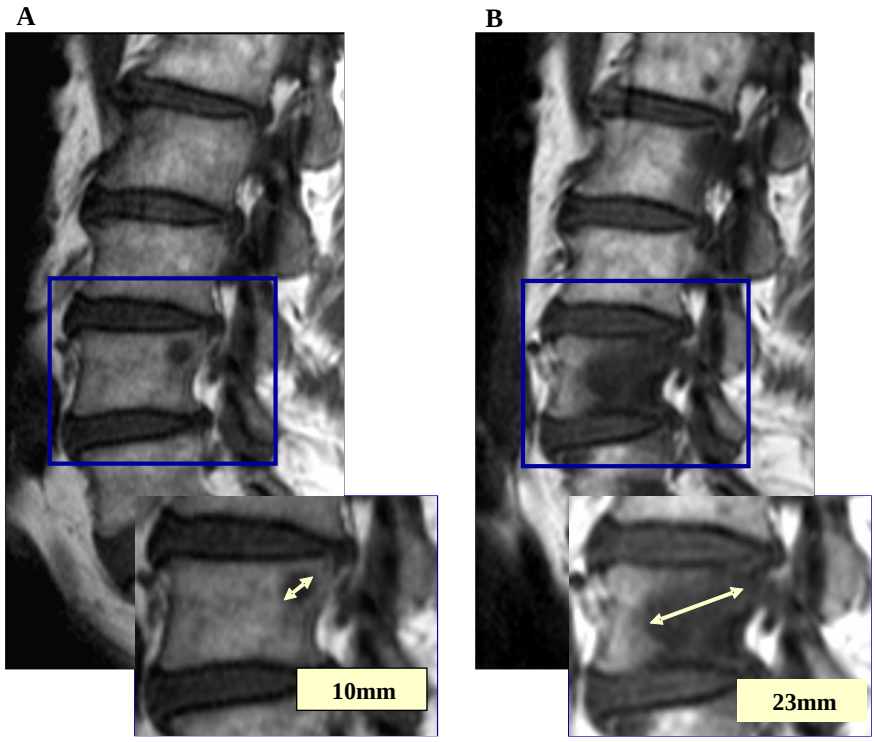

Fig. 6. (A) Sagittal T1-weighted image of the lumbar spine obtained before initiation of systemic therapy shows a focal metastasis in the L4-vertebral body (round area of low signal intensity). (B) Corresponding MR image obtained 6 month later shows evident increase in size of the L4-lesion, and appearance of multiple new focal lesions (progression). Magnified boxes in (A) and (B) illustrate measurement of the L4-lesion's largest diameters.

DISCUSSION

The initial assessment of antitumour activity of a new agent is usually performed by measuring the extent of tumor shrinkage induced by this agent in the primary tumor or in its metastases. The use of standardized guidelines secures the comparison of results between agents and investigators. In the present study, we have arbitrarily decided to transpose RECIST guidelines. Published in 2000, RECIST guidelines result from the joined effort of EORTC, NCI, and NCIC[1] and are based on the ability to perform unidimensional measurements of the largest tumor diameter in solid lesions clearly identifiable by clinical examination or preferentially by CT or MRI [15]. In the case of PCa, the application of

[1] European Organization for Research and Treatment of Cancer, National Cancer Institute of the United States, National Cancer Institute of Canada

Table 3. Evaluation of response to treatment in 20 patients

	Initial MRI		6 months follow-up MRI		Evaluation of response after 6 months of treatment		
					Bone lesions	Biological	Soft tissue metastases
P	**N° lesions**	**Total Size***	**N° lesions**	**Total size***	**MRI**	**PSA**	**Target Lesions**
1	7	139,5	0	0	CR	Normalization	NA
2	>10	252	0	0	CR	Normalization	NA
3	9	232	1	44	PR	Decrease ≥ 50 %	PR
4	2	25	1	15	PR	Normalization	NA
5	>10	300,2	>10	289,5	SD	PROG	NA
6	9	130,1	9	155,6	SD	PROG	NA
7	1	37,1	1	31,5	SD	Decrease ≥ 50 %	NA
8	>10	172	>10	141	SD	Decrease ≥ 50 %	SD
9	3	134,6	4	165,1	SD	Decrease ≥ 50 %	NA
10	7	168,8	7	213,8	PROG	PROG	NA
11	3	57	>10	355,9	PROG	PROG	NA
12	3	51,1	>10	280,2	PROG	Decrease ≥ 50 %	PROG
13	5	123	7	173	PROG	PROG	NA
14	9	141,6	9	240	PROG	PROG	NA
15	4	57,6	4	80	PROG	PROG	NA
16	6	121,7	6	235,4	PROG	PROG	NA
17	0		1	10.2	PROG	PROG	NA
18	0		>10	275	PROG	PROG	PROG
19	3	38,7	Diffuse pattern		PROG	PROG	NA
20	3	50	Diffuse pattern		PROG	PROG	PROG

* = Sum of maximal lesion diameters
NA = not available, CR = complete response; PR = partial response; SD = stable disease; PROG = disease progression

RECIST guidelines is unfortunately confounded by the high percentage of patients with bone metastatic disease, which in absence of a quantitative imaging technique to measure the size of bone metastases, are considered non measurable [15,18-20]. Bone metastases are detected in 65-75% of metastatic PCa, while measurable lesions in lymph nodes or soft

tissues are present in less than 40%, except in the very later stages of the disease [4,6]. In the current series, only 29.5% of the patients had RECIST target metastases, a proportion comparable to recently published trials with the same category of patient [4,6]. This has become even more crucial with the development of new drugs specifically tailored to bone metastases, such as bisphosphonates or inhibitors of the endothelin axis. Since it is not possible to measure in large series of patient their true efficacy on bone metastases, they are presently evaluated using composite endpoint incorporating Tc-99m bone scans progression, metabolic markers, symptoms, occurrence of new complications. Tc-99m bone scintigraphy, which has been used for more than 25 years, allows a quantitative estimation of the total number of so-called hot spots but lacks specificity, requiring additional radiographic or CT examination of suspect areas to differentiate neoplastic from degenerative or (micro) traumatic changes [21-25]. Except for the appearance of new lesions, Tc-99m does not allow precise measurement of tumor response, therefore identifying correctly disease progression [21]. The uptake of Tc-99m by individual hotspots metastases does only poorly reflect the size and activity of a metastasis, since the uptake is dependent on blood flow and osteoblastic activity and potentially biases by the induction a so-called false-flare phenomenon, in which bone healing cannot be distinguished from the scintigraphic appearance of new metastases or progressive disease [26,27].

Alternatives to Tc-99m bone scan are developed but lack practical applicability and are not ready to be implemented in clinical trials. Computed tomography of Tc-99 suspect areas produces high-quality skeletal images and can be used to measure response in selected regions. But Tc-99m bone scan are required to identify regions of interest, since acquiring serial images of more than a small part of the skeleton is impractical. In addition, the total accumulated radiation's dose required by whole spine CT preclude the systematic use of this technique for the determination of therapeutic response in clinical trials. Bone-specific biochemical markers, *e.g.* N-terminal and C-terminal byproducts from synthesis of procollagen type I, alkaline or osteocalcin are currently evaluated to diagnose and measure the extent of bone metastasis disease throughout the skeleton. Several studies have shown that systemic therapy can reduce circulating levels of these markers, but so far, there is still little evidence that they can be used as surrogate of tumor response [21].

MRI is a powerful single step and non invasive examination to image the axial skeleton [10]. MRI has been shown more cost-effective than standard procedures in several bone related disorder [28,29]. Physiological and pathological variations in the composition of red marrow and its distribution among normal subjects have been precisely described , so that trained radiologists can consistently differentiate metastases from red marrow abnormalities [30]. MRI is routinely used for the work-up of suspected vertebral fractures and spinal cord compression in symptomatic cancer patients [12,13]. The different MRI patterns of bone marrow involvement in neoplastic disease have been precisely described, so that discriminating metastatic lesions from benign marrow abnormalities such as marrow hyperplasia induced by chemotherapy can easily been made by trained radiologists [17]. The superiority of AS-MRI over the bone scan/X-Rays couple for the detection bone metastases has already been demonstrated [31-35].

Our results confirm this sensitivity, since metastatic lesions were identified in 33 out of 38 patients (86%), three of them having negative bone scans. Interestingly, AS-MRI has not been used prospectively yet to measure and/or monitor tumor response to drug therapy in bone. The current study addressed this issue in a homogeneous series of unselected patients with metastatic HRPCa. Its results are triple. Firstly, the study demonstrated that AS-MRI enabled a precise measurement of bone metastases. Focal metastatic lesions were found in 22 patients, diffuse (non measurable) metastatic marrow infiltration in 5, and a combination of both patterns in the spine and pelvis respectively in 3 patients. Second, AS-MRI increased the proportion of patients with measurable disease.

Transposition of RECIST guidelines to AS-MRI studies was easy and allowed the easy definition of measurable target lesions in almost two thirds of patients. Consolidation of the measurement's plane between sequential examinations was facilitated by the use of highly stable vertebral and pelvic bony landmarks. In the present series, the proportion of patients with measurable target lesions raised from approximately 30% to 66% when adding AS-MRI to CT-TAP. Third sequential AS-MRI enabled follow-up of bone metastatic lesions under treatment. In 20 patients treated during 6 months, the response was assessed in the axial skeleton by AS-RMI as easily as it is for soft tissue tumors using CT-TAP. In contrast to Tc-99 bone scans, objective measurement on sequential AS-MRI studies offered the opportunity to categorize tumor response as it is done in other tissues. Altogether, these results appeared very promising with regard to the proportion of patients who could be objectively evaluated in phase II/III trials of new anticancer agents and take advantage of therapy monitoring.

Several remarks deserve further attention. In contrast to bone scans, AS-MRI studies focus on the axial skeleton (spine and pelvi-femoral areas) and do not allow identification of metastases in the peripheral skeleton. This does not appear significant if used to assess the efficacy of a drug on bone metastasis, given the exquisite tropism of PCa metastases for the pelvis and lumbo-sacral spine [36]. As shown by Traill et al, isolated appendicular metastases occur very seldom and MRI survey limited to the spine and pelvis do not overlook significant peripheral metastases [31]. This is confirmed in our study since all patients with appendicular lesions also had measurable axial lesions. Additionally, the major trust of AS-MRI is not to provide an exhaustive identification of bone metastases but to allow precise tumor measurement in a subset of these lesions. In contrast to X-ray or CT scan, AS-MRI does not allow discriminating osteolytic from osteoblastic lesions, but on a therapeutic standpoint however, they both represent targets to control. It might equally be assumed that we did not obtain histopathological correlation to confirm the diagnosis of metastatic disease and to assess lesion evolution under treatment. In most clinical trials, such confirmation is not done for other primary or secondary tumors either, for example in routine follow-up of lung or liver cancer or metastases using CT. As in this situation, it seems reasonable to consider that the changes in size and number of lesions measured by AS-MRI during a treatment course reflect tumor response (*i.e.*, decrease in size and/or number: regression; increase: progression). Moreover, the agreement between the responses determined on the basis of AS-MRI and on the basis of CT-TAP of soft-tissues lesions when available strengthens this interpretation. In the present study, the only discrepancies in

response assessment were observed between AS-MRI and PSA levels evolution. The limitation of PSA as a surrogate for tumor response are well known [14,37]. Although recent data from the docetaxel trials indicate that PSA response might indicate better survival, it is noticeable that these conclusions were drawn without a standard of reference to correlate response in bone and PSA response [4,6].

The major limitation of this study remained the limited sample size and the absence of cross-validation by different investigators on different sites**.** Larger multicentric studies are mandatory to confirm the critical role of AS-MRI for lesion quantification and response evaluation, and to assess its potential superiority over PSA levels follow-up for response determination. Another potential prolongation of this study is the evaluation of AS-MRI for metastases measurement and follow-up in other cancers which frequently involve bones.

CONCLUSION

This study showed that in a large majority of metastatic PCa patient, AS-MRI is an objective straightforward quantitative tool to measure metastatic disease, to follow lesions under treatment, and to increase the proportion of patients who could take advantage of the monitoring of the efficacy of most recent anticancer agents.

REFERENCES

1. Sartor O. Prostate cancer and bone: a unique relationship with multiple opportunities for targeted therapy. Clin Prostate Cancer 2004;3(2):71-72.
2. Carlin BI, Andriole GL. The natural history, skeletal complications, and management of bone metastases in patients with prostate carcinoma. Cancer 2000;88(12 Suppl):2989-2994.
3. Jimeno A, Carducci M. Atrasentan: targeting the endothelin axis in prostate cancer. Expert Opin Investig Drugs 2004;13(12):1631-1640.
4. Petrylak DP, Tangen CM, Hussain MH, Lara PN, Jr., Jones JA, Taplin ME, Burch PA, Berry D, Moinpour C, Kohli M, Benson MC, Small EJ, Raghavan D, Crawford ED. Docetaxel and estramustine compared with mitoxantrone and prednisone for advanced refractory prostate cancer. N Engl J Med 2004;351(15):1513-1520.
5. Saad F, Gleason DM, Murray R, Tchekmedyian S, Venner P, Lacombe L, Chin JL, Vinholes JJ, Goas JA, Zheng M. Long-term efficacy of zoledronic acid for the prevention of skeletal complications in patients with metastatic hormonerefractory prostate cancer. J Natl Cancer Inst 2004;96(11):879-882.
6. Tannock IF, de Wit R, Berry WR, Horti J, Pluzanska A, Chi KN, Oudard S, Theodore C, James ND, Turesson I, Rosenthal MA, Eisenberger MA. Docetaxel plus prednisone or mitoxantrone plus prednisone for advanced prostate cancer. N Engl J Med 2004;351(15):1502-1512.
7. Therasse P. Measuring the clinical response. What does it mean? Eur J Cancer 2002;38(14):1817-1823.
8. Dearnaley DP, Sydes MR, Mason MD, Stott M, Powell CS, Robinson AC, Thompson PM, Moffat LE, Naylor SL, Parmar MK. A double-blind, placebocontrolled, randomized trial of oral sodium clodronate for metastatic prostate cancer (MRC PR05 Trial). J Natl Cancer Inst 2003;95(17):1300-1311.

9. Daffner RH, Lupetin AR, Dash N, Deeb ZL, Sefczek RJ, Schapiro RL. MRI in the detection of malignant infiltration of bone marrow. AJR Am J Roentgenol 1986;146(2):353-358.
10. Godersky JC, Smoker WR, Knutzon R. Use of magnetic resonance imaging in the evaluation of metastatic spinal disease. Neurosurgery 1987;21(5):676-680.
11. Vogler JB, 3rd, Murphy WA. Bone marrow imaging. Radiology 1988;168(3):679-693.
12. Carmody RF, Yang PJ, Seeley GW, Seeger JF, Unger EC, Johnson JE. Spinal cord compression due to metastatic disease: diagnosis with MR imaging versus myelography. Radiology 1989;173(1):225-229.
13. Khaw FM, Worthy SA, Gibson MJ, Gholkar A. The appearance on MRI of vertebrae in acute compression of the spinal cord due to metastases. J Bone Joint Surg Br 1999;81(5):830-834.
14. Bubley GJ, Carducci M, Dahut W, Dawson N, Daliani D, Eisenberger M, Figg WD, Freidlin B, Halabi S, Hudes G, Hussain M, Kaplan R, Myers C, Oh W, Petrylak DP, Reed E, Roth B, Sartor O, Scher H, Simons J, Sinibaldi V, Small EJ, Smith MR, Trump DL, Wilding G. Eligibility and response guidelines for phase II clinical trials in androgen-independent prostate cancer: recommendations from the Prostate-Specific Antigen Working Group. Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology 1999;17(11):3461-3467.
15. Therasse P, Arbuck SG, Eisenhauer EA, Wanders J, Kaplan RS, Rubinstein L, Verweij J, Van Glabbeke M, van Oosterom AT, Christian MC, Gwyther SG. 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(3):205-216.
16. Vande Berg BC, Malghem J, Lecouvet FE, Maldague B. Classification and detection of bone marrow lesions with magnetic resonance imaging. Skeletal Radiol 1998;27(10):529-545.
17. Vanel D, Dromain C, Tardivon A. MRI of bone marrow disorders. Eur Radiol 2000;10(2):224-229.
18. Dreicer R. Metastatic prostate cancer: assessment of response to systemic therapy. Seminars in Urologic Oncology 1997;15(1):28-32.
19. Papac RJ. Bone marrow metastases. A review. Cancer 1994;74(9):2403-2413.
20. Hanks C, Myers C, Scardino P. Cancer of the prostate. In: DeVita VTJ, Hellman S, Rosenberg SA, editors. Cancer: Principles and Practice of Oncology (5th edition): Lippincott Co Philadelphia, PA; 1993. p 1073-1113.
21. Clamp A, Danson S, Nguyen H, Cole D, Clemons M. Assessment of therapeutic response in patients with metastatic bone disease. Lancet Oncol 2004;5(10):607-616.
22. Hardy JG, Kulatilake AE, Wastie ML. An index for monitoring bone metastases from carcinoma of the prostate. Br J Radiol 1980;53(633):869-873.
23. Imbriaco M, Larson SM, Yeung HW, Mawlawi OR, Erdi Y, Venkatraman ES, Scher HI. A new parameter for measuring metastatic bone involvement by prostate cancer: the Bone Scan Index. Clin Cancer Res 1998;4(7):1765-1772.
24. Jorgensen T, Muller C, Kaalhus O, Danielsen HE, Tveter KJ. Extent of disease based on initial bone scan: important prognostic predictor for patients with metastatic prostatic cancer. Experience from the Scandinavian Prostatic Cancer Group Study No. 2 (SPCG-2). European Urology 1995;28(1):40-46.

25. Soloway MS, Hardeman SW, Hickey D, Raymond J, Todd B, Soloway S, Moinuddin M. Stratification of patients with metastatic prostate cancer based on extent of disease on initial bone scan. Cancer 1988;61(1):195-202.
26. Hayward JL, Carbone PP, Heusen JC, Kumaoka S, Segaloff A, Rubens RD. Assessment of response to therapy in advanced breast cancer. Br J Cancer 1977;35(3):292-298.
27. Lemieux J, Guimond J, Laberge F, St-Pierre C, Cormier Y. The bone scan flare phenomenon in non-small-cell lung cancer. Clin Nucl Med 2002;27(7):486-489.
28. Annertz M, Wingstrand H, Stromqvist B, Holtas S. MR imaging as the primary modality for neuroradiologic evaluation of the lumbar spine. Effects on cost and number of examinations. Acta Radiol 1996;37(3 Pt 1):373-380.
29. Rubin SJ, Marquardt JD, Gottlieb RH, Meyers SP, Totterman SM, O'Mara RE. Magnetic resonance imaging: a cost-effective alternative to bone scintigraphy in the evaluation of patients with suspected hip fractures. Skeletal Radiol 1998;27(4):199-204.
30. Vande Berg BC, Malghem J, Lecouvet FE, Maldague B. Magnetic resonance imaging of normal bone marrow. Eur Radiol 1998;8(8):1327-1334.
31. Traill ZC, Talbot D, Golding S, Gleeson FV. Magnetic resonance imaging versus radionuclide scintigraphy in screening for bone metastases. Clin Radiol 1999;54(7):448-451.
32. Freedman GM, Negendank WG, Hudes GR, Shaer AH, Hanks GE. Preliminary results of a bone marrow magnetic resonance imaging protocol for patients with high-risk prostate cancer. Urology 1999;54(1):118-123.
33. Chan Y, Chan K, Lam W, Metreweli C. Comparison of whole body MRI and radioisotope bone scintigram for skeletal metastases detection. Chin Med J (Engl) 1997;110(6):485-489.
34. Heshiki A. Bone marrow MRI in prostate cancer. Adv Exp Med Biol 1992;324:209-215.
35. Turner JW, Hawes DR, Williams RD. Magnetic resonance imaging for detection of prostate cancer metastatic to bone. J Urol 1993;149(6):1482-1484.
36. Suzuki T, Shimizu T, Kurokawa K, Jimbo H, Sato J, Yamanaka H. Pattern of prostate cancer metastasis to the vertebral column. Prostate 1994;25(3):141-146.
37. Carducci MA, DeWeese TL, Nelson JB. Prostate-specific antigen and other markers of therapeutic response. The Urologic Clinics of North America 1999;26(2):291-302, viii.

Chapter 7

RECIST revisited. A review of validation studies on tumor assessment

Therasse P, Eisenhauer EA, Verweij J
Eur J Cancer. 2006 Apr 6; [Epub ahead of print]

ABSTRACT

The Response Evaluation Criteria In Solid Tumors (RECIST) were developed in the late 1990's to replace the WHO criteria for response evaluation. The new criteria included important changes such as unidimensional tumor measurement, selection of target lesions with a minimum size, details concerning imaging modalities and a new threshold for assignment of objective progression.

RECIST was published in February 2000 and very quickly came into operation first in clinical trials performed under the auspices of EORTC, US NCI or NCI Canada Clinical Trials Group but was adopted quickly thereafter by the entire cancer clinical research community. Because several key features of RECIST were based on analysis of retrospective clinical data it was felt important to carefully monitor the implementation of the guidelines and stimulate prospective validation studies. This paper reviews the literature that has been published on RECIST from 2000 up to November 2005. In total 60 papers and ASCO abstracts directly refer to research studies or reviews related to RECIST and its implementation. Amongst the 60 references identified for this review, 11 papers refer to validation studies (7 prospective and 4 retrospective), 6 papers refer to the comparison of unidimensional measurements versus bi or tri-dimensional measurements, 12 papers address issues raised with the implementation of RECIST in Mesothelioma and Gastro-Intestinal Stromal Tumors and 4 papers report on an adaptation of RECIST for specific tumor types.

In general, RECIST has been well received by the scientific community and most validation studies fully support the implementation of the new criteria. As expected, however, some issues have been identified. In keeping with the mathematical differences in definition of progression, RECIST delays the identification of progression as compared to WHO criteria in some instances. RECIST criteria are not easily applicable in some types of trials such as those in pediatric tumors and in mesothelioma. Furthermore, anatomical changes in the tumor as described by RECIST may be detected later than functional changes in some circumstances, as for example in Gastro-Intestinal Stromal Tumors treated with Imatinib. However, there is no other universal method of tumor assessment as yet and functional imaging methods won't be validated and widely available before some time. The findings of this review together with experience acquired thus far and the results of some ongoing research projects pave the way for RECIST 2.0 hopefully later this year.

INTRODUCTION

Response Evaluation Criteria In Solid Tumors (RECIST) were introduced by a small international working group in February 2000 to facilitate, improve and standardize the evaluation and the reporting of objective tumor outcomes in early clinical trials investigating new anti-cancer agents [1]. In comparison to earlier response assessment systems, the new criteria gave much more detailed recommendations on how to assess tumor lesions, how to report responses, and also took into account recent developments in medical imaging techniques. RECIST uses a uni-dimensional measure (the longest diameter) to quantify measurable tumor lesions as opposed to the bi-dimensional product (longest diameter multiplied by its perpendicular) which was commonly employed by earlier iterations of response criteria [2,3,4]. Building on the work of others [3,5], RECIST defines measurable lesions as those with a minimum size depending on the method of investigation. Following a principle already implemented in the SWOG response criteria [3], the threshold for defining objective progression was arbitrarily increased as compared to the WHO criteria, i.e the increase in measurable overall tumor burden required for progression was greater in RECIST (20% in one dimension being approximately equivalent to a 44 % increase in bidimensional product) than in the WHO criteria (25% increase in product).

Following the publication of RECIST, standard case report forms (CRFs) and protocol sections were created by the working group and made available on the web. A special email address was created to receive and answer questions related to the implementation of the criteria. A website was created to host the Questions and Answers to facilitate the implementation of the criteria (www.eortc.be\recist). Although the last comment on the website was posted in 2003, the RECIST working group continues (every week!) to answer questions and provide support for the interpretation of the criteria in specific situations.

After the publication of RECIST, some investigators raised concerns about the interest, the pertinence and the applicability of the new criteria. The main purpose of this paper is to review the work performed and published by other colleagues on the usefulness of the criteria in general and their validation in specific tumor types when available.

REVIEW METHODOLOGY

The search strategy was simple and made through PUBMED using the word RECIST as keyword to identify titles and abstracts published between February 2000 and November 2005. This search strategy identified 99 referenced papers. Only those manuscripts reporting on original work focused on the methodology of response evaluation and RECIST were retained for detailed review. Also excluded were editorial comments and non-English literature. Ultimately 43 papers satisfied these criteria. A second search was undertaken of abstracts published in the American Society of Clinical Oncology (ASCO) annual conference proceedings between 2001 and 2005. This identified a further 9 abstracts (and related data in oral presentations or posters) that have not yet been followed by a full paper.

Finally, examination of the reference lists in the 43 full papers yielded another, 8 additional papers which met the review criteria. Thus in total, 60 studies (51 papers and 9 ASCO abstracts) were identified for inclusion in this review.

RESULTS

The studies included focused either on general principles related to the *implementation* of RECIST (or tumor evaluation) or on a prospective or retrospective attempt to *validate* the utility of RECIST in certain tumor types. Accordingly the results of this review have been divided into general considerations and tumor specific considerations.

General considerations

One of the first papers to refer to RECIST was a commentary of Padhani and Husband [6]. The authors outline the problems inherent to the morphological assessments of tumors independently of the number of dimensions being measured and briefly explore the development of functional imaging as a tool of the future. However their conclusion is crystal clear: “current criteria should remain unchallenged until better functional parameters emerge”. One year later the same first author [7] analysed RECIST and its impact on radiology departments highlighting the possibility that the implementation of RECIST could translate into an increased workload. The paper concluded that, while the issue of workload requires careful monitoring, this factor alone should not be an argument to be less stringent in response assessment in the performance of clinical trials. Institutions that can not provide this service should be considered incapable of performing studies where response assessment is crucial. In 2004, the International Cancer Imaging Society (ICIS) published a consensus statement about the evaluation of the response to treatment of solid tumors [8], including a number of issues related to the implementation of RECIST (Table 1). Another paper [9] published almost simultaneously but in another journal identified very similar issues. It is interesting to note that on one hand these authors cite concern about the potential increase in workload created with the application of RECIST (specifically the requirement to measure up to 10 lesions if multiple measurable lesions are identified) while on the other hand advise consideration for the use of 3 dimensional measurements which, to date, has not been shown to be more useful than 1 dimensional measurement (for the purpose of response evaluation) but is certainly much more complex and time consuming.

The general concordance between RECIST and WHO criteria was tested retrospectively in a cohort of 130 patients with different tumor types and entered into different protocols [10]. In line with the larger increase in lesion size required for definition of PD found in RECIST, it was shown that about 1/3 of patients normally identified as PD with WHO criteria would still be classified as SD with RECIST. The authors also used this dataset to create multiple simulations to artificially change tumor shape to demonstrate that increasing the irregularity of lesions may decrease the concordance rate of PR and SD categories between the two methods.

Table 1. Main concerns expressed by the International Cancer Imaging Society concerning RECIST

Topics	Concerns
General	RECIST is a first step in the good direction but requires well trained radiologists who should also be involved in the planning of trials
WHO vs RECIST	Several studies have shown a good concordance between RECIST and WHO for response but less good concordance for time to progression. This should be taken into account for planning of future trials
1 dimension vs 3 dimensions	3 dimensional imaging is available in many centers and should be considered in the guideline
Specific issues	▪ Lymph-node should be measured in the short axis ▪ Changes in tumor consistency (calcification, necrosis..) should also be reported for accurate evaluation of tumors ▪ Cystic lesion shouldn't be systematically excluded ▪ Why 10 target lesions? No scientific rationale ▪ Appendix concerning imaging technology should be updated ▪ MRI could be used to measure bone disease (breast, prostate) ▪ More information should be given on the use of contrast for imaging

In an analysis of 32 North Central Cancer Treatment Group (NCCTG) trials including 2374 patients [11] it was suggested that 2 lesions were sufficient to provide a reliable tumor assessment. However only 23 % of the patients studied had more than 2 lesions at baseline. Schwartz et al [12] used the data of 36 patients with multiple target lesions to simulate hundreds of possible groupings between target lesions and suggested that when 6 lesions were measured bidimensionally and 4 lesions were measured unidimensionally, the average variance was decreased by 90%. The same authors subsequently developed a mathematical model [13] which could in theory enable physicians to calculate the optimal number of lesions to follow to decrease the variance to an acceptable level for patients with a large number of lesions at baseline. Although this concept is certainly interesting it could increase the workload for response assessment so it would be strengthened if it underwent further testing on a large database generated on prospective studies.

Perez-Gracia and colleagues [14] analysed 9 phase II trials (same drug but different tumor types) including 416 patients amongst which 97 responses were first recorded and 81 were later confirmed. Most unconfirmed responses were due to patients lost to follow-up.

Because of the high correlation coefficient between the rate of confirmed and non confirmed responses, the authors suggest that confirmation of response may not be necessary and should be studied in a larger setting. The relevance of response confirmation is part of an ongoing debate. The need for confirmation of response should be discussed as an approach for further simplifying the RECIST.

Three papers from the same group of authors addressed the problems of the implementation of RECIST in paediatric oncology. Two papers [15,16] focussed on technical problems while the third paper [17] provided a short illustrative description of 10 cases and the issues raised by attempting to measure the tumor response according to RECIST. A summary of the most important points raised is presented in Table 2. There are very few studies with a focus on the methodology of response assessment in paediatric solid tumors and the problems reported by the authors seem to relate more to the application of RECIST in clinical practice than to the use of RECIST for screening potential new anticancer agents in clinical trials. It is not uncommon that individual patients (adult and pediatric) in oncology practice cannot have response assessed according to RECIST if they do not have measurable disease: these cases would not normally be eligible for inclusion in phase II trials where response is the endpoint.

In summary, the major issues identified in these reports were related to the need for measurement of 10 lesions, the fact that some patients with progression by WHO were considered to have stable disease by RECIST, proposals for different approaches to lesion measurement, and the minimal impact in terms of overall response rate of the requirement for response confirmation.

Table2. Main concerns raised by pediatricians regarding the implementation of RECIST in pediatric studies.

Topics	Concerns
General	▪ Problems related to disseminated disease with diffuse infiltration
	▪ The minimum size of target lesions should be smaller than 10 mm with current available techniques (multislice CT)
Imaging	▪ The need for repetitive exposure to radiation burden when multiple CT should be performed hence ultrasonography avoids unnecessary radiation exposure and ensure much better compliance for children
	▪ Bone lesions should be acceptable as they can be evaluated by MRI
	▪ All radiological plans should be considered to measure tumor lesions and not the axial plan only
	▪ Functional imaging should also be considered as a possible modality for tumor evaluation in some instances

Tumor specific considerations

Lung Cancer

Three papers compared RECIST and WHO criteria for the assessment of response in Non Small Cell Lung Cancer [18-20]. The results of the three studies are summarized in Table 3. They report a good correlation between unidimensional measurements and sum of bidimensional products in keeping with the data supplied in the RECIST paper. The application of RECIST translated into an ineligibility rate of 5% because of the requirement for a minimum size of the target lesion [20], a factor on which WHO is silent. Three other papers [21-23] analyzed the intra and inter-observer variability in tumor response evaluation using either RECIST or RECIST and WHO criteria but did not make a comparison in terms of response and progression rates between the two set of criteria. In all three papers, the unidimensional approach reduced the inter-observer variability of the measurements and the misclassification of some patients. This is further supported by the observation that the inter-observer variability improves when minimum lesion size criteria are used [20]. Interestingly the papers were almost unanimous in recommending RECIST for tumor evaluation in future trials in NSCLC.

Mesothelioma

Tumor evaluation in mesothelioma has always been a difficult problem. The original WHO criteria were already modified in 1997 [24] to enable the use of unidimensional measurements for response assessment. Despite this adaptation, the modified WHO criteria do still not seem optimal given the frequent discordance between the evaluation of bidimensionally measured lesions and unidimensionally measured lesions [25]. One might have expected that RECIST, focusing only at unidimensional measurement, would be less confusing [26]. Yet, in 2 small studies (34 patients and 4 patients respectively) [27, 28] there was considerable discordance between the evaluation performed according either RECIST or WHO criteria, both for objective response and progression. Byrne [25] proposed modified RECIST criteria for mesothelioma (using the longest perpendicular diameter to chest wall or mediastinum measured at 2 sites at 3 different levels on CT scan). He tested these modified RECIST criteria against the more complex WHO modified criteria retrospectively with data from two trials and found a good correlation between them. Responses reported with the modified criteria were associated with longer survival and improved lung function. Similar outcomes had been reported by others [29]. Further testing in reasonably sized prospective trials is encouraged for the definitive validation of the modified RECIST for mesothelioma.

Breast and Colorectal Cancer

Four studies [30-33] have compared the response rates observed with WHO criteria and RECIST in metastatic breast or colorectal cancer (Table 3). Interestingly, the study by

Table 3: Prospective and retrospective studies comparing WHO criteria with RECIST

Author	Tumor type	P/R[a]	Sample size	RR (%) RECIST	RR (%) WHO	PD (%) RECIST	PD (%) WHO
Werner (18)	Locally Adv Lung	P	22	87	87	NA	NA
Cortes (19)	Metastatic NSCLC[b]	R	164	52	52	26	26
Watanabe (20)	Metastatic NSCLC	R	120	19.3	20	13	17.5
Trillet Lenoir (30)	Metastatic Colorectal	P	91	25	20	41	43
Prasas (31)	Metastatic Breast	P	86	50	50	NA	NA
Choi (32)	Metastatic Colorectal	P	41	36	32	NA	NA
Muro (33)	Metastatic Esophageal	P	52	20	24	39	43
Therasse (42)	Metastatic STS[b]	P	49	6.1	4.1	32.6	34.7
Negrier (51)	Metastatic RCC[b]	P	61	14.7	11.4	44.3	57.3
Schwartz (49)	Metastatic RCC	R	53	5.6	3.7	11.3	17
Park (60)	Mix of tumors	R	79	30.4	31.6	30.4	38

[a]P = prospective; R = retrospective
[b]NSCLC = non small cell lung cancer; RCC = renal cell cancer; STS = soft tissue sarcoma

Trillet-Lenoir [30] has investigated the original WHO criteria (as published) as well as a modified version (no progression declared when only one lesion is progressing and no more than 5 lesions measured per organ) which correlated very well with RECIST both for response and progression. In a second study [31], although the overall response rates were identical between RECIST and WHO criteria, 4 patients were recorded with a worse response and 4 with a better response with RECIST. These were all patients with irregularly shaped tumors. Tran [34] compared the change in lung metastases measurements in 15 patients using WHO criteria (2D), RECIST (1D) and a 3 dimensional measurement system (3D). The author reported that the 1D and 3D measurements were concordant in 29 of 30 classifications, the 2D and 3D measurements were concordant in 23 of 30 classifications and 1D and 2D were concordant in 24 of 30 classifications. Despite the good concordance between 1D and 3D assessments, the level of agreement (measured with Kappa statistic) did not reach significance and the overall correlation between the various methods was considered fair to poor. The Author acknowledged that many tumors were irregularly shaped and thus presented challenges in correctly calculating the 3D measurements which

might in part explain the results. Kimura et al [35] retrospectively assessed 50 breast cancer patients comparing RECIST with the standard criteria used by the Japanese Cancer Society for breast cancer evaluation. Because the Japanese criteria are different from the WHO criteria, the meaning of the comparison with RECIST is not particularly clear. However, the author noted that 32 % of patients would have been ineligible for studies had RECIST been used because of the minimum size for target lesions and the exclusion of patients with bone metastases only.

Gastrointestinal Stromal Tumor (GIST) and Soft Tissue Sarcoma

GIST is a relatively rare disease as compared to other cancers and until recently there has been no effective systemic therapy available. Imatinib, an oral tyrosine kinase inhibitor, has proven to be very effective to treat GIST. Early metabolic responses can be observed with ^{18}Fluorodeoxyglucose (FDG) positron emission tomography (PET), a technique that assess metabolic activity in tissues. The FDG PET responses usually precede objective response by several weeks [36]. Many investigators have reported problems with the evaluation of GIST based on changes in tumor size [37-41]. GIST's can sometimes increase in size as a result of the metabolic response (intratumoral haemorrhage or mixoid degeneration) or intratumoral nodules can be reactivated showing an increased metabolic activity (translating early resistance to treatment) while the size of the tumor remains globally stable. FDG-PET has some limitations for evaluating tumor response to treatment, such as specificity, access, costs and quantitative measurements. Stroszczynski [40] proposes to associate MRI examinations with changes in tumor size and Choi [41] proposes a new set of criteria based both on changes in tumor size and changes in tumor density measured on CT images. The criteria proposed by Choi are currently being evaluated in a multicenter study.

Therasse [42] studied the outcome of 49 non-GIST soft tissue sarcoma patients treated for metastatic disease with ET-743. The response rates reported using RECIST and WHO based criteria were 6% and 4% respectively for partial response, 61% for no change (stable disease) using both methods and 33% and 35% respectively for progression. The outcome of the study would have been the same regardless of the criteria used. RECIST may correlate much less favourably with histological response in locally advanced high-grade soft tissue sarcoma treated with neoadjuvant doxorubicin and ifosfamide. In a total of 41 patients 11 had a good histologic response while only 1 patient had a response by RECIST [43]. Similar lack of correlation has also been observed with WHO criteria, so it may be that the neoadjuvant approach in soft tissue sarcomas is not suited for any size-related assessment of response.

Prostate Cancer

Prostate cancer metastases are well known to be difficult to measure by imaging. Tombal [44] reported that only 11 (29%) of 38 consecutive metastatic hormone refractory prostate cancer patients had measurable disease according to RECIST. Twenty-five patients had focal metastatic bone lesions identified and potentially measurable on axial-skeleton MRI

(see below). In another study [45] two cohorts of patients with hormone refractory (HRPC) (31 patients) and hormone sensitive (HSPC) (124 patients) prostate cancer respectively, were analysed for eligibility to enter clinical trials in which RECIST criteria were to be used. 39% of HRPC patients and 51% of HSPC would have qualified on the basis of having at least one measurable lesion, while 13 % and 44% respectively had bone-only disease (which is non-measurable). In this prostate study, most of the visceral measurable lesions were in lymph-nodes for which the evaluation by RECIST may also be difficult. Clearly, the evaluation of prostate cancer in screening phase II studies may require different tools such with outcome measures for clinical benefit defined according to stage of the disease and/or composite endpoints that include the various clinical dimensions of the disease (measurable disease, QOL, PSA, Bone, pathology…)[45].

Brain tumors

While in the past brain tumors were thought to be difficult to assess for response, this currently seems less of an issue. Four separate studies [46-49] on a total of 204 patients showed a high concordance between 1D (RECIST), 2D (WHO) and 3D (volumetric) measurements in detecting responses both in childhood and adult brain tumors. For disease progression, the results were a little less uniform. One study suggested progression was detected later with the 1D measurement [46]. Although 3 of these studies have as yet only been reported in abstract form, the data suggest that RECIST is a useful tool for brain tumor measurement.

Renal Cell Carcinoma

In two studies on metastatic renal cancer [50, 51] outcome according to RECIST and WHO criteria correlated extremely well (Table 3).

Other Issues

The issue of functional imaging has already been mentioned. Gopinath [52] assessed functional volume variations against RECIST in 22 patients with neuroendocrine tumors of carcinoid type in the liver and treated with chemotherapy. He found out that tumor functional volume assessed by single-photon emission computed tomography (SPECT) predicted clinical outcome (as measured by a reduction in pain, flushing or abdominal symptoms) for 59% of patients and RECIST only for 36% of patients.

Another problem is the use of a primary tumor for response assessment, if the tumor is localized in a hollow organ. Indeed the nature of the cancer growing for instance within the wall thickness of the oesophagus makes measurements based on RECIST difficult. Response to treatment is even more difficult to image because of post-treatment fibrous stenosis [53]. Tahara proposed to incorporate in RECIST a set of simple criteria based on endoscopic evaluation and histology to validate the complete response. The application of the proposed criteria correlates well with survival in a cohort of 139 patients reviewed

retrospectively. RECIST was assessed in advanced gastric cancer by Yoshida [54]. He compared the response rate of 161 patients with measurable disease included in a phase III trial and concluded that response rates by the old and new criteria were nearly equal.

Following tumor markers is another possible approach to assessing change in tumor burden. Successive CA-125 measures were compared with sequential tumor measurements as prognostic factor for survival in 131 patients receiving second line chemotherapy for advanced ovarian cancer [55]. The study was retrospective and based on patients from a single institution treated with similar regimen in first and second line. In this setting, CA-125 response was 2.6 times better than clinical response (assessed by RECIST) in predicting survival. However, in this example, the authors have examined both endpoints in relationship to their predictive value in survival in individual patients. This is a somewhat different application than use of the endpoint to screen new drugs for activity. The authors also confirm the difficulties in using radiographic techniques for assessing recurrent ovarian carcinoma.

Metastatic site related considerations

Bone metastases represent a frequent problem in breast and prostate cancer. Approximately 20-30% of these patients will present with bone metastases as the only metastatic site. Since RECIST considers bone metastases as non measurable, several authors have tried to assess this problem [7, 9, 35, 44, 45]. While they confirmed the problem, they unfortunately could not provide alternatives. In contrast, Hamaoka [56] proposed a set of criteria requiring the combination of different imaging modalities (skeletal scintigraphy, plain radiography, computed tomography and/or magnetic resonance imaging) depending on the characteristics of the lesions. Tombal [44] investigated the possibility of using the axial skeleton MRI (AS-MRI) for quantitative imaging of bone metastases of prostate cancer. He prospectively performed AS-MRI in a cohort of 30 patients with HRPCa and a positive Tc-99m bone scan before and 6 months after starting chemotherapy, and suggested that the proportion of patients eligible to enter trials based on RECIST increased from 28.9% to 65.7% with measurable lesions on AS-MRI. The feasibility of this approach should be tested prospectively taking into account technical issues, cost and time constraints.

Lymph-nodes and nodal masses can be manifestations of many cancers, but RECIST has yet only been reported compared to WHO in primary lymphomas. Sohaib [57] compared CT assessment in 1D, 2D and 3D for tumor response in 16 patients with either lymphoma or germ cell tumors. He concluded that whichever method is used there is limited influence on the classification of treatment response. Assouline [58] pooled the data from 3 phase II lymphoma trials (115 patients) and compared RECIST with the International Working Criteria (IWC) [59]. RECIST was slightly adapted to make the criteria relevant for response in lymphoma and the overall response rates were 42% and 46% for IWC and RECIST respectively with identical progression rates.

DISCUSSION

RECIST has become the most frequently used response criteria for clinical trials investigating new treatments for solid tumors. The criteria are used to define response rate, progression rate and/or time to progression irrespective of the stage of development of new cancer therapeutics. Some features of the criteria have also been rapidly implemented in day to day practice of oncologists for standard patient care.

Overall, many authors agree that the development of RECIST with rigorous evaluation of many underlying aspects of response assessment has been very valuable. However, RECIST is not the universal panacea that one would like to have to precisely measure tumor response and progression in all possible situations and with all type of cancer therapeutics. Interestingly, the implementation of RECIST has also revealed a number of otherwise uncovered problems related to response evaluation in specific situations [8,9,17,25,29,35,45]. Although RECIST may not have provided an answer to all problems it has the merit of having stimulated the discussion and therefore improved awareness and harmonization!

The first objective of this review was to look into all prospective and retrospective studies attempting to validate RECIST against the WHO criteria. An overview of all studies directly comparing the two methods is presented in Table 3. In keeping with the retrospective data compiled in the RECIST manuscript, none of these studies found major differences in response rate between the two methods, while some found a slightly longer time to progression or lower progression rate for RECIST. The latter was expected and confirms that the changes in the definition of progression with RECIST translate into fewer patients being classified as having progression at a certain time as compared with WHO criteria. This should be taken into account when time to progression or progression free survival serve as a primary endpoint in non randomized phase II trials where the primary hypothesis is constructed with reference to historical controls based on WHO criteria evaluations. This is of particular importance now that we move towards the development of drugs which may not induce rapid tumor shrinkage. Reference matrices for this purpose can be derived from recent large phase III which have used RECIST for response and progression assessment. EORTC is currently developing such references for several tumor types.

Apart from directly comparing RECIST to WHO criteria several groups have investigated the value of volumetric measurement (3D) as opposed to bi-dimensional (2D) (WHO criteria) or uni-dimensional (1D) (RECIST) approaches for assessing tumor response. There was a belief, particularly for lung and brain tumors, that volumetric assessment would be better. In addition, radiologists [8,9] and paediatricians [15,16,17] have raised concerns on the deficits of non-volumetric measurements. However, in all reported studies [18,34,46,47,49,57] the correlation between 1D and 3D dimension was quite good, with one possible exception for classifying progression in childhood cancers which would have been detected earlier with a 3D assessment [46]. However criteria for declaring progression were

not proportionally identical. Apparently the use of more laborious 3D measurements does not add great value to the much simpler 1D assessment of RECIST for the purpose of response determination in clinical trials.

RECIST has been found not optimal for assessment of response in mesothelioma and gastro-intestinal stromal tumors (GIST). Mesothelioma, because of its specific growth characteristics deserves specific criteria, and some adaptations of RECIST as proposed by Byrne [25] seem an appropriate compromise to keep criteria as simple as possible yet keeping a good correlation with clinical outcome. For GIST the problem is less well elucidated The implementation of RECIST came in parallel with the clinical development of Imatinib in GIST and early treatment monitoring of this targeted therapy with FDG-PET identified a metabolic response much before any morphological response (tumor shrinkage) could be reported according to RECIST [36,37,38,39]. However, this is only a chronology issue and does not yet disqualify RECIST. And for many different reasons [40,41], not least the limited availability of the tool, FDG-PET is at present not yet a universal method to assess tumor response in GIST. Yet, also in GIST, size changes might be opposite to clinical benefit observed, and early changes in tumor density without significant size changes may still harbour progression or response. This renders RECIST (as presently defined) not totally appropriate for use in GIST, and other criteria are currently being tested [41].

The development of Imatinib shows that we have to be careful with the evaluation of new cancer therapeutics using standard tools. Evaluation solely based on morphological changes can be misleading and an accurate early determination of response may require functional and molecular techniques that assess metabolism, growth kinetics, angiogenesis growth factors, tumor cell markers and in vivo genetic alterations and gene expression. The US National Cancer Institute therefore has designed a very large research program to develop imaging in oncologic drug development [61, 62].

Other potential refinements and facilitations of the process of response evaluation [11,12,13,14] are as important. In this respect, the projects of Tombal [44], Assouline [58] and Tahara [54] are worthwhile following. Further large scale validation of some of these pilot projects is planned and the results should hopefully be available for RECIST 2.0!

By virtue of the search system used it has to be assumed that this review may be incomplete. Other useful commentary may have been included in the discussion sections of phase II and phase III trial reports, which we would have missed by the methods used in searching the literature. However we believe we have captured many important assessments of RECIST in this review.

RECIST was implemented five years ago and since has been used at a level far beyond expectation. This indicates it serves a purpose. The large majority of the validation studies reported in this review support the use of RECIST as a tool for tumor evaluation in most common situations. The review also identified areas requiring specific criteria or attention.

Imaging in oncology, and in drug development in particular, is rapidly developing and requires continuous research and validation. A revised version of RECIST will take into account the experience accumulated so far and should provide guidance informed by the experience of many in the field to aid in assessment of novel agents.

REFERENCES

1. 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.
2. Miller AB, Hogestraeten B, Staquet M, Winkler A. Reporting results of cancer treatment. Cancer 1981, 47, 207-214.
3. Green S, Weiss GR. Southwest Oncology Group standard response criteria, endpoint definitions and toxicity criteria. Invest New Drugs 1992, 10, 239-253.
4. Tumor eligibility and response criteria for phase II and III studies. EORTC Data Center Manuel. Brussels, 1992, monograph.
5. Van Glabbeke M, van Oosterom AT, Steward W, Verweij J, Mouridsen H. Selection of large and objectively measurable target lesions in EORTC phase II trials: impact on recruitment and response rate. EORTC Soft Tissue and Bone Sarcoma Group (STBSG). Eur J Cancer 1993, 29, 1943-1947.
6. Padhani AR, Husband JE. Are current tumour response criteria relevant for the 21st century? Br J Radiol 2000, 73, 1031-1033.
7. Padhani AR, Ollivier L. The RECIST (Response Evaluation Criteria in Solid Tumors) criteria: implications for diagnostic radiologists. Br J Radiol 2001, 74, 983-986.
8. Husband JE, Schwartz LH, Spencer J, et al. Evaluation of the response to treatment of solid tumours – a consensus statement of the International Cancer Imaging Society. Br J Cancer 2004, 90, 2256-2260.
9. Bellomi M, Preda L. Evaluation of the response to therapy of neoplastic lesions. Radiol Med 2004, 107, 450-458.
10. 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.
11. Hillman SL, Sargent DJ, An MW, et al. Evaluation of RECIST criteria in determining the response to treatment in solid tumours: A North Central Cancer Treatment Group (NCCTG) investigation. Proc Am Soc Clin Oncol 2003, 22, 521 (abstr 2095).
12. Schwartz LH, Mazumdar M, Brown M, Smith A, Panicek DM. Variability in response assessment in solid tumours: effect of number of lesions chosen for measurements. Clin Cancer Res 2003, 9, 4318-4323.
13. Mazumdar M, Smith A, Debroy PP, Schwartz LH. A theoretical approach to choosing the minimum number of multiple tumors required for assessing treatment response. J Clin Epidemiol 2005, 58, 150-153.
14. Perez-Garci JL, Muñoz M, Willimans G, et al. Assessment of the value of confirming responses in clinical trials in oncology. Eur J Cancer 2005, 41, 1528-1532.
15. McHugh K, Kao S. Response evaluation criteria in solid tumours (RECIST): problems and need for modifications in paediatric oncology? Br J Radiol 2003, 76, 433-436.
16. Moon L, McHugh K. Advances in paediatric tumour imaging. Arch Dis Child 2005, 90, 608-611.

17. Barnacle AM, McHugh K. Limitations with the response evaluation criteria in solid tumours (RECIST) guidance in disseminated pediatric malignancy. Pediatr Blood Cancer 2005, Oct 17 (Epub ahead of print)
18. Werner-Wasik M, Xiao Y, Pequignot E, Currant WJ, Hauck W. Assessment of lung cancer response after nonoperative therapy: tumor diameter, bidimensional product, and volume. A serial CT scan-based study. Int J Radiat Oncol Biol Phys 2001, 51, 56-61.
19. Cortes J, Rodriguez J, Diaz-Gonzalez JA et al. Comparison of unidimensional and bidimensional measurements in metastatic non-small cell lung cancer. Br J Cancer 2002, 15, 158-160.
20. Watanabe H, Yamamoto S, Kunitoh H et al. Tumor response to chemotherapy: the validity and reproducibility of RECIST guidelines in NSCLC patients. Cancer Sci 2003, 94, 1015-1020.
21. Watanabe H, Leki R, Mori K, et al. The impact of an independent response evaluation committee (REC) using RECIST guidelines in a Four-Arm Cooperative Study (FACS) for advanced non-small cell lung cancer (NSCLC) in Japan . Proc Am Soc Clin Oncol 2004, 23, 629 (abstr 7065).
22. Erasmus JJ, Gladish GW, Broemeling L et al. Interobserver and intraobserver variability in measurements of non-small-cell carcinoma lung lesions: implications for assessment of tumor response. J Clin Oncol 2003, 21, 2574-2582.
23. Grossi F, Belvedere O, Fasola G et al. Tumor measurements on computed tomographic images of non-small cell lung cancer were similar among cancer progressionals from different specialties. J Clin Epidemiol 2004, 57, 804-808.
24. Byrne MJ, Davidson JA, Musk AW, et al. Cisplatin and gemcitabine treatment for malignant mesothelioma, a phase II study. J Clin Oncol 1999, 17, 25-30.
25. Byrne MJ, Nowak AK. Modified RECIST criteria for assessment of response in malignant pleural mesothelioma. Ann Oncol 2004, 15, 257-260.
26. Nowak AK. CT, RECIST, and malignant pleural mesothelioma. Lung Cancer 2005, 49, S37-40.
27. 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.
28. Monetti F, Casanova S, Grasso A, Cafferata MA, Ardizzoni A, Neumaie CE. Inadequacy of the new Response Evaluation Criteria in Solid Tumors (RECIST) in patients with malignant pleural mesothelioma: report of four cases. Lung Cancer 2004, 43, 71-74.
29. Heelan R. Staging and response to therapy of malignant pleural mesothelioma. Lung Cancer 2004, 45, S59-S61.
30. Trillet-Lenoir V, Freyer G, Kaemmerlen P, et al. Assessment of tumour response to chemotherapy for metastatic colorectal cancer: accuracy of the RECIST criteria. Br J Radiol 2002, 75, 903-908.
31. Prasad SR, Saini S, Summer JE, Hahn PF, Sahani D, Boland GW. Radiological measurement of breast cancer metastases to lung and liver: comparison between WHO (bidimensional) and RECIST (unidimensional). J Comput Assist Tomogr 2003, 27, 380-384.
32. Choi JH, Ahn MJ, Rhim H, et al. Comparion of RECIST and WHO response criteria in metastatic colorectal carcinoma. Proc Am Soc Clin Oncol 2005, 23, 301 (abstr 3721).
33. Muro K, Hamaguchi T, Ohtsy A, et al. A phase II study of single-agent docetaxel in patients with metastatic esophageal cancer. Ann Oncol 2004, 15, 955-959.

34. Tran LN, Brown MS, Goldin JG, et al. Comparison of treatment response classifications between unidimensional, bidimensional, and volumetric measurements of metastatic lung lesions on chest computed tomography. Acad Radiol 2004, 11, 1355-1360.
35. Kimura M, Tominaga T. Outstanding problems with response evaluation criteria in solid tumors (RECIST) in breast cancer. Breast Cancer 2002, 9, 153-159.
36. 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.
37. Reichardt P, Schneider U, Stroszczynski C, Pink D, Hohenberger P. Molecular response of gastrointestinal stromal tumour after treatment with tyrosine kinase inhibitor imatinib mesylate. J Clin Pathol 2004, 57, 215-217.
38. Choi H, Charnsangavej C, de Castro Faria S, et al. CT evaluation of the response of gastrointestinal stromal tumors after imatinib mesylate treatment: a quantitative analysis correlated with FDG PET findings. Am J Roentgenol 2004, 183, 1619-1628.
39. Goldstein D, Tan BS, Rossleigh M, Haindl W, Walker B, Dixon J. Gastrointestinal stromal tumours: correlation of F-FDG gamma camera-based coincidence positron emission tomography with CT for the assessment of treatment response – an AGITG study. Oncology 2005, 69, 326-332.
40. Stroszczynski C, Jost D, Reichardt P, et al. Follow-up of gastro-intestinal stromal tumours (GIST) during treatment with imatinib mesylate by abdominal MRI. Eur Radiol 2005, 15, 2448-2456.
41. Choi H. Critical issues in response evaluation on computed tomography: lessons from the gastrointestinal stromal tumor model. Curr Oncol Rep 2005, 7, 307-311.
42. Therasse P, Le Cesne A, Van Glabbeke M, Verweij J, Judson I; for the EORTC Soft Tissue and Bone Sarcoma Group. RECIST vs WHO: Prospective comparison of response criteria in an EORTC phase II clinical trial investigating ET-743 in advanced soft tissue sarcoma. Eur J Cancer 2005, 41, 1426-1430.
43. Schuetze SM, Eary JF, Griffith KA, et al. FDG PET but not RECIST agrees with histologic response of soft tissue sarcoma to neoadjuvant chemotherapy. Proc Am Soc Clin Oncol 2005, 23, 817s (abstr 9005).
44. Tombal B, Rezazadeh A, Therasse P, Van Cangh PH, Vande Berg B, Lecouvet FE. Magnetic resonance imaging of the axial skeleton enables objective measurement of tumor response on prostate cancer bone metastases. Prostate 2005, 65, 178-187.
45. Scher HI, Michael JM, Kelly WK, Schwartz LH. Prostate cancer clinical trial end points: "RECIST"ing a step backwards. Clin Cancer Res 2005, 11, 5223-5232.
46. Warren KE, Patronas N, Aikin AA, Albert PS, Balis FM. Comparison of one-, two- and three-dimensional measurements of childhood brain tumors. J Nat Cancer Inst 2001, 93, 1401-1405.
47. Sze G, Mehta MP, Schutlz CJ, et al. Radiologic response evaluation of brain metastases: unidimensional World Health Organization (WHO) response evaluation criteria in solid tumors (RECIST) vs bidimensional or 3-dimensional criteria. Proc Am Soc Clin Oncol 2001, 20, 59 (abstr 234).
48. Galanis E, Maurer MJ, Ballman KV, et al. Validation of neuroradiologic response assessment in gliomas: RECIST (1D) vs 2D measurements vs computer-assisted tumor area vs volume. Proc Am Soc Clin Oncol 2003, 22, 106 (abstr 423).
49. Shah G, Kesari S, Xu R, et al. Comparison of 1D, 2D, 3D and volumetric parameters in measuring tumor response in high-grade gliomas in adults. Proc Am Soc Clin Oncol 2004, 23, 112, (abstr 1523).

50. Schwartz LH, Mazumdar M, Wang L, et al. Response assessment classification in patients with advanced renal cell carcinoma treated on clinical trials. Cancer 2003, 98, 1611-1619.
51. Negrier S, Thiesse C, Vincent C, et al. Different measurement criteria give different tumor response classifications in patients with metastatic renal cancer (MRCC). Proc Am Soc Clin Oncol 2004, 23, 520, (abstr 6013).
52. Gopinath G, Ahmed A, Buscombe JR, Dickson JC, Caplin ME, Hilson AJ. Prediction of clinical outcome in treated neuroendocrine tumours of carcinoid type using functional volumes on 111In-pentetreotide SPECT imaging. Nucl Med Commun 2004, 25, 253-257.
53. Tahara M, Ohtsu A, Hironaka S, et al. Clinical impact of criteria for complete response (CR) of primary site to treatment of esophageal cancer. Jpn J Clin Oncol 2005, 35, 316-323.
54. 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 anticancer effects of advanced gastric cancer. Gastric Cancer 2000, 3, 128-133.
55. Gronlund B, Høgdall C, Hilden J, Engelholm SA, Høgdall EVS, Hansen HH. Should CA-125 response criteria be preferred to response evaluation criteria in solid tumors (RECIST) for prognostication during second-line chemotherapy of ovarian carcinoma? J Clin Oncol 2004, 22, 4051-4058.
56. Hamaoka T, Madewell JE, Podoloff DA, Hortobagyi GN, Ueno NT. Bone imaging in metastatic breast cancer. J Clin Oncol 2004, 22, 2942-2953
57. Sohaib SA, Turner B, Hanson JA, Farquharson M, Oliver RTD, Reznek RH. CT assessment of tumour response to treatment: comparison of linear, cross-sectional and volumetric measures of tumour size. Br J Radiol 2000, 73, 1178-1184.
58. Assouline S, Eisenhauer E. RECIST (Response evaluation criteria in solid tumors) applied to response in lymphoma. Proc Am Soc Clin Oncol 2004, 23, 582 (abstr 6606).
59. Cheson BD, Horning SJ, Coiffier B, et al. Report of an international workshop to standardize response criteria for non-Hodgkin's lymphomas. NCI sponsored international working group. J Clin Oncol 1999, 17, 1244.
60. Park JO, Lee SI, Song SY, et al. Measuring response in solid tumors: comparison of RECIST and WHO. Jpn J Clin Oncol 2003, 33, 533-537.
61. Kelloff GJ, Krohn KA, Larson SM, et al. The progress and promise of molecular imaging probes in oncologic drug development. Clin Cancer Res. 2005, 22, 7967-7985
62. Kelloff GJ, Hoffman JM, Johnson B, et al. Progress and promise of FDG-PET imaging for cancer patient management and oncologic drug development. Clin Cancer Res. 2005, 8, 2785-2808

Chapter 8

Summary and Conclusions

Samenvatting

The work reported in this thesis relates to one of the most critical aspects of cancer clinical trials that is the evaluation of treatment both in terms of responses and progression. There isn't any clinical trial published nowadays which does not report response and/or progression rates both of which contribute to the efficacy assessment on new treatments. Hence relying on a robust and validated methodology to assess these events is extremely important. Also important is the feasibility of the methodology proposed which should be as simple as possible and widely applicable across institutions involved in cancer clinical trials.

The standard World Health Organization criterion for partial response is a 50% or more decrease in the sum of the products of two measurements (the maximum diameter of a tumor and the largest diameter perpendicular to this maximum diameter) of individual tumors. However, 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. It has been hypothesized that the calculation of bidimensional products is unnecessary, and a 30% decrease in the sum of maximum diameters of individual tumors (assuming spherical shape and equivalence to a 50% reduction in the sum of the bidimensional products) was proposed as a new criterion. This is the scope of **chapter 2** in which we describe the results of a study comparing the standard response and the new response criteria to the same data to determine whether the same number of responses in the same patients would result. Data from 569 patients included in eight studies of a variety of cancers were reanalyzed. The two response criteria were separately applied, and the results were compared using the kappa statistic. The importance of confirmatory measurements and the frequency of nonspherical tumors were also examined. In addition, for a subset of 128 patients, a unidimensional criterion for disease progression (30% increase in the sum of maximum diameters) was applied and compared with the standard definition of a 25% increase in the sum of the bidimensional products. The overall agreement between the unidimensional and bidimensional criteria was generally found to be good. The kappa statistic for concordance for overall response was 0.95 and the conclusion was that one dimensional measurement of tumor maximum diameter may be sufficient to assess change in solid tumors.

In parallel to the work reported above several research organizations grouped their forces to tackle the review of the WHO criteria on the basis of the experience and knowledge acquired since the publication of the criteria in the early 1980's. After several years of intensive discussions with all partners involved in cancer clinical research, a new set of guidelines called RECIST (Response Evaluation Criteria In Solid Tumors) was prepared to replace the former WHO criteria. The RECIST guidelines are described in **chapter 3**.

The new concept using one dimension to quantify tumor lesions was validated in a population of more than 4000 patients included in 14 different trials and integrated in the new criteria. The average overall response rate using bidimensional or unidimensional criteria were 30% and 29% respectively. The outcome of each individual trial was identical whether using WHO criteria or RECIST. The criteria to define progression were also modified in RECIST proposing an increase of 20% in the sum of the longest diameters as cut of for progression (corresponding to an increase of 44% in two dimensions). RECIST

also provides some philosophical background to clarify the various purposes of response evaluation. It proposed a model whereby the entire tumor burden is characterized by a limited number of target lesions (to be measured) and non-target lesions (not to be measured) extrapolating an overall response to treatment from a combined assessment of all existing lesions. Methods of assessing tumor lesions have been better classified, briefly in the guidelines and in details for various imaging techniques in a specific appendix. All other aspects of response evaluation have been reviewed and amended to increase clarity and/or precisions whenever appropriate. EORTC, NCI-US and NC-Canada immediately adopted these guidelines for all clinical trials.

The modifications of the response criteria with RECIST also put a strong emphasis on clinical response to treatment as an important indicator of the therapeutic effect of anticancer agents. The value and interpretation of clinical response has to be carefully considered within the context of its use and this is discussed in detail in **chapter 4**. In daily practice, response assessment is combined with other indicators of the patient's condition to contribute to the decision making process.

In clinical trials, it is widely used to identify and quantify the anti-tumor activity of new agents. This is used in early clinical trials to screen new anticancer agents an detect those which are worth further developing in large phase III efficacy studies with more robust endpoints. In this context, response evaluation should be performed on the basis of strict pre-defined criteria such as WHO criteria or RECIST.

Clinical response is also used as an indicator of therapeutic efficacy in combination with other indicators such as for example duration of response, quality of responses, and localization of responses. On that basis response rates in clinical trials have been widely used by the pharmaceutical industry and the regulators to speed up access to new drugs (with early marketing authorization) for patients when there is an unmet medical need. However, the value of clinical response as a surrogate indicator of longer time to progression or survival benefit remains unclear in most instances and can hardly be established in the frame of a single randomized trial. Accordingly measuring clinical response in phase III trials should not systematically follow the same rigorous principles as those applied in phase II. Measuring response in this context is still informative and should be done to support the main conclusions of the trial based either on progression free survival or overall survival. With the development of new anticancer agents behaving differently than cytotoxics (not inducing rapid tumor shrinkage), early clinical benefit also encompasses new concepts such as long disease stabilization and/or increased rates of patients free from progression at pre-defined fixed time points.

The RECIST criteria have been retrospectively validated on a large data base of phase II and phase III trials but because of the nature and the importance of the modifications proposed in RECIST, prospective validation has been awaited by the scientific community.

The study reported in **chapter 5** is one of those validation studies that was set up just after the publication of RECIST as a prospective validation exercise in soft tissue sarcoma. Forty-nine patients were entered into a phase II clinical trial aiming at determining the activity (and the safety) of ET-743 (Ecteinascidin) in second line treatment of advanced soft tissue sarcoma. Response to treatment and progression were monitored according to the WHO criteria and also RECIST, using the WHO criteria as reference for therapeutic decisions. Patients with gastro-intestinal stromal tumors were excluded from the analysis. All patients with at least one measurable lesion at entry satisfying both sets of response criteria were included in the study. Discordances between WHO and RECIST criteria for the best response were reported for two cases: one no-change (WHO) reported as partial response (RECIST) and one progression (WHO) reported as no-change (RECIST). In terms of date of progression, 3 patients progressed on WHO criteria while they were still stable with RECIST. Overall the results of the study would not have changed if RECIST had been used instead of WHO criteria. The main conclusion of the study was that response criteria as defined by RECIST are adequate to measure response and progression in non-GIST soft tissue sarcoma and can be used instead of the modified WHO criteria.

RECIST also points at specific problems with some disease sites being considered as non measurable because they cannot be assessed quantitatively such as for example ascites, pleural effusion, lymphangitis or bone disease. The last problem is particularly important since many patients with breast and prostate cancer only present with bone disease at time of treatment. **Chapter 6** reports the results of a study assessing the value of Magnetic Resonance Imaging of the axial skeleton (AS-MRI) as a single step technique to quantify bone metastases and measure tumor response in prostate cancer.

AS-MRI was performed in 38 patients before receiving chemotherapy for metastatic HRPCa, in addition to PSA assessment, computed tomography of the thorax, abdomen and pelvis [CT-TAP] and Tc-99m bone scintigraphy. A second AS-MRI was performed in 20 patients who completed six months of chemotherapy. Evaluation of tumor response was performed using RECIST.

Only 11 patients (29%) had RECIST measurable metastases in soft-tissues or lymph nodes on baseline CT-TAP. AS-MRI identified a diffuse infiltration of the bone marrow in 8 patients and focal measurable metastatic lesions in 25 patients (65%), therefore doubling the proportion of patients with measurable lesions. Transposing RECIST on AS-MRI in 20 patients who completed six months of treatment, allows the accurate estimation of complete response (n=2), partial response (n=2), stable disease (n=5), or tumor progression (n=11), as it is done using CT-TAP in soft tissue solid metastases. In conclusion, it is possible to measure bone metastases of the axial skeleton with MRI and assess tumor response objectively. The feasibility of this approach at a larger scale requires further investigations.

Chapter 7 of the thesis is dedicated to the review of the literature that has been published about RECIST from 2000 up to November 2005. In total 60 papers and ASCO abstracts directly referred to research studies or reviews related to RECIST and its implementation.

Amongst the 60 references identified for this review, 11 papers referred to validation studies (7 prospective and 4 retrospective), 6 papers referred to the comparison of unidimensional measurements versus bi or tri-dimensional measurements, 12 papers addressed issues raised with the implementation of RECIST in mesothelioma and gastro-intestinal stromal tumors and 4 papers reported on an adaptation of RECIST for specific tumor types.

In general, RECIST has been well received by the scientific community and most validation studies fully support the implementation of the new criteria. As expected, however, some issues have been identified. In keeping with the mathematical differences in definition of progression, RECIST delays the identification of progression as compared to WHO criteria in some instances. RECIST criteria are not easily applicable in some types of trials such as those in pediatric tumors and in mesothelioma. Furthermore, anatomical changes in the tumor as described by RECIST may be detected later than functional changes in some circumstances, as for example in gastro-intestinal stromal tumors treated with Imatinib. However, there is no other universal method of tumor assessment as yet and functional imaging methods won't be validated and widely available before some time.

FUTURE PERSPECTIVE

The findings of this review together with experience acquired thus far and the results of some ongoing research projects pave the way for a revised version of the criteria. This new version will clarify a number of issues of the previous version which have been addressed by the RECIST working group over the last five years through questions received from those applying RECIST in real situations and answers that were posted on the RECIST website. More data on the use of tumor markers (and references to other published criteria) will be integrated as well as specific criteria for particular tumor types such as mesothelioma. In collaboration with our colleagues' radiologists more attention will also be given to recommendations for the use of sophisticated imaging techniques and the utilization of contrast products. Finally, issues such as the need for confirmation of response and the minimum number of target lesions needed to make a correct tumor assessment are being investigated at EORTC in large data sets from recently completed large trials using RECIST. This is done in collaboration with our colleagues from Industry and the results of these investigations should be available by the time the revised guideline will be drafted. As expressed above, the need to quantify tumor lesions with clinical or images based measurements will remain for the next five to ten years whether this is used to measure response or assess tumor progression. It is therefore of paramount importance that clinical investigations use a similar and up to date set of response criteria to ensure uniform assessment of activity and efficacy of new cancer treatments.

Samenvatting

Dit proefschrift bespreekt een van de belangrijkste aspecten van klinisch oncologisch onderzoek, namelijk het evalueren van de behandeling zowel met betrekking tot respons als tot progressie. Heden ten dage wordt er geen enkele klinische studie gepubliceerd die niet respons en/of progressie percentages rapporteert. Beide eindpunten dragen bij tot de beoordeling van de effectiviteit van nieuwe behandelingen. Zodoende is het uiterst belangrijk om te kunnen vertrouwen op een robuuste en gevalideerde methodologie om deze eindpunten te bepalen. Eveneens belangrijk is de uitvoerbaarheid van voorgestelde methodologie. Deze moet zo eenvoudig mogelijk zijn en een brede toepassing mogelijk maken in alle instituten die bij klinische oncologische studies betrokken zijn.

Het standaard World Health Organisation (WHO) criterium voor partiële respons is een vermindering van minstens 50% van de som der producten van twee metingen (de maximale diameter van de tumor en de grootste diameter loodrecht op deze eerste) van de afzonderlijke tumoren. Theoretisch staat de eenvoudige som van de maximale diameters van de afzonderlijke tumoren echter in een betere lineaire verhouding ten opzichte van cel sterfte dan de som van de tweedimensionale producten. We hypothetiseerden dat de berekening van tweedimensionale producten overbodig is, en stelden een 30% vermindering van de som der maximale diameters van de afzonderlijke tumoren (die onder aanname van een bol vorm gelijkwaardig is aan een 50% vermindering van de som der tweedimensionale producten) voor als nieuw criterium. Dit is het onderwerp van **hoofdstuk 2** waarin we de resultaten beschrijven van een studie die het standaard en het nieuwe respons criterium vergeleek. Gegevens van 569 patiënten, uit acht studies bij verschillende kanker soorten werden opnieuw geanalyseerd. De twee respons criteria werden afzonderlijk toegepast en de resultaten werden vergeleken door middel van de kappa statistiek. Het belang van bevestigende metingen en de frekwentie van niet-bolvormige tumoren werden ook onderzocht. Tevens werd voor een deelgroep van 128 patiënten een eendimensionaal criterium voor progressie (30% toename in de som van de maximale diameters) toegepast en vergeleken met de standaard definitie van een 25% toename in de som van de tweedimensionale producten. De overeenkomst tussen de eendimensionale en tweedimensionale criteria was over het algemeen goed. We konden derhalve veronderstellen dat de simpele eendimensionale meting van de maximum tumor diameter voldoende zou kunnen zijn om verandering in de grootte van solide tumoren te beoordelen.

Parallel aan, en volgend op het boven beschreven onderzoek bundelden verscheidene onderzoeksorganisaties hun krachten om de aanpassing van de WHO criteria aan te pakken, uitgaande van de ervaring en kennis opgedaan sinds de publicatie van die criteria in de vroege jaren ’80. Na verscheidene jaren intensieve discussie met alle partners betrokken bij klinisch oncologisch onderzoek, werd een nieuwe reeks richtlijnen met de naam RECIST (Response Evaluation Criteria In Solid Tumors) voorbereid teneinde de vroegere WHO criteria te vervangen. De RECIST richtlijnen worden beschreven in **hoofdstuk 3**.

Het nieuwe concept dat één dimensie gebruikt om tumor letsels te kwantificeren werd gevalideerd in een populatie van meer dan 4000 patiënten uit 14 verschillende studies en vorm gegeven in de nieuwe criteria. De gemiddelde globale respons percentages voor de tweedimensionale en de eendimensionale criteria waren respectievelijk 30% en 29%. De eindresultaten van elke klinische studie afzonderlijk waren identiek voor de WHO en de RECIST criteria. De criteria om progressie te definiëren werden ook aangepast in RECIST: een toename van 20% in de som van de langste diameters werd voorgesteld als limietwaarde voor progressie (hetgeen overeenkomt met een toename van 44% in de tweedimensionale methode). RECIST verleent ook een filosofische achtergrond om de verschillende doelstellingen van respons evaluatie te verduidelijken. Het verschafte een model waarin de volledige tumor omvang wordt gekarakteriseerd door een beperkt aantal doel lesies (die exact gemeten moeten worden) en niet-doel lesies (die niet exact gemeten hoeven te worden), met extrapolatie van een globaal respons op de behandeling uit de gecombineerde beoordeling van genoemde lesies. Zo zijn de methodes om tumor lesies te beoordelen beter geklassifieerd, in het kort in de richtlijnen en in detail voor verschillende beeldtechnieken in een specifieke appendix. Alle andere aspecten van respons evaluatie werden bekeken en aangepast om de duidelijkheid en nauwkeurigheid te verbeteren waar noodzakelijk. EORTC, NCI-US en NC-Canada hebben onmiddellijk deze richtlijnen voor al hun klinische studies overgenomen.

De aanpassingen van de respons criteria door RECIST leggen ook sterk de nadruk op klinische respons op de behandeling als een belangrijke indicator van het therapeutische effekt van antikanker middelen. De waarde en interpretatie van klinische respons moet zorgvuldig beoordeeld worden in de context van zijn toepassing en dit wordt in detail besproken in **hoofdstuk 4**. In de dagelijkse praktijk wordt respons beoordeling gekombineerd met andere parameters van de toestand van de patiënt om bij te dragen aan de besluitvorming over de behandeling. In klinische studies wordt respons algemeen gebruikt om antitumor activiteit van nieuwe middelen te identificeren en te kwantificeren. Het wordt gebruikt in vroeg-klinische studies om nieuwe antikanker middelen te screenen en zodoende diegene te ontdekken die het waard zijn verder ontwikkeld te worden in grote fase III studies met meer robuuste eindpunten. In deze context zou respons evaluatie op basis van strikt gedefinieerde criteria zoals de WHO criteria of RECIST moeten worden verricht.

Klinische respons wordt ook gebruikt als een parameter van therapeutische effectiviteit in combinatie met andere parameters zoals bijvoorbeeld duur, kwaliteit en lokalisatie van de responsen. Zodoende worden respons percentages in klinische studies algemeen gebruikt door de farmaceutische industrie en de overheden om toegang tot nieuwe geneesmiddelen te versnellen (met vervroegde toelating tot de markt) voor ziekten waar geen goede behandeling voor beschikbaar is. Nochtans blijft de waarde van klinische respons als een surrogaat parameter van een langere tijd tot progressie of een overlevings verbetering vaak onduidelijk en kan die waarde nauwelijks vastgesteld worden in het kader van een enkele klinische studie. Dienovereenkomstig hoeft het meten van klinische respons in fase III studies niet systematisch dezelfde rigoureuze principes te volgen als die gebruikt voor fase II studies. Het meten van respons in deze context blijft nog steeds informatief en zou moeten worden gebruikt om de hoofd conclusies van de studie, gebaseerd op progressie

vrije overleving of algemene overleving, te ondersteunen. Voor de ontwikkeling van nieuwe antikanker middelen met een ander werkings mechanisme en doel (niet het induceren van snelle tumor verkleining) dan cytotoxica, omvat zogenaamd "vroeg klinisch voordeel (clinical benefit)" ook nieuwe concepten zoals langdurige stabilisatie van de ziekte en/of hogere percentages van patiënten die vrij van progressie zijn op van te voren bepaalde tijdspunten.

De RECIST criteria werden retrospectief gevalideerd in een groot databestand van fase II en fase III studies, maar gezien de aard en het belang van de in RECIST voorgestelde aanpassingen, verwachtte de wetenschappelijke gemeenschap ook een prospectieve validatie. De studie beschreven in **hoofdstuk 5** is een van de validatie studies die direct na de publicatie van RECIST werden opgezet als een prospectieve validatie, in dit geval in weke delen sarcoom. Negenenveertig patiënten werden in een fase II klinische studie opgenomen teneinde de activiteit (en de veiligheid) van ET-743 (Ecteinascidin) voor de tweede lijns behandeling van vergevorderd weke delen sarcoom. Respons op de behandeling, en progressie werden zowel beoordeeld volgens de WHO criteria als RECIST, waarbij de WHO criteria als referentie werden gebruikt voor therapeutische beslissingen. Patiënten met gastro-intestinale stroma tumoren werden uitgesloten van deze analyse. Alle patiënten met minstens 1 meetbare lesie bij aanvang van de behandeling, en die voldeden aan de voorwaarden voor het gebruik van beide sets respons criteria, werden opgenomen in de studie. Slechts in twee gevallen kwamen de WHO en RECIST criteria voor de beste respons niet overeen. Qua datum van de progressie, hadden 3 patiënten een progressie volgens de WHO criteria terwijl ze nog stabiel waren volgens RECIST. De totaalresultaten van de studie zouden niet anders zijn geweest als RECIST was gebruikt in plaats van de WHO criteria. De hoofdconclusie van onze studie was dat respons criteria zoals gedefinieerd door RECIST bruikbaar zijn om respons en progressie te meten in niet-GIST weke delen sarcomen en gebruikt kunnen worden in plaats van de aangepaste WHO criteria.

RECIST verwijst ook naar specifieke problemen in sommige ziekte lokalisaties die als niet meetbaar worden beschouwd omdat ze niet kwantitatief beoordeeld kunnen worden: bijvoorbeeld ascites, pleura vocht, lymphangitis of bot uitzaaiingen. Het laatste voorbeeld is bij uitstek belangrijk omdat vele patiënten met borst- en prostaatkanker enkel bot metastasen hebben ten tijde van de behandeling. Hoofdstuk 6 geeft de resultaten van een studie ter beoordeling van "Magnetic Resonance Imaging of the axial skeleton" (AS-MRI) als een eenstapstechniek om bot metastasen te kwantificeren en tumor respons bij hormoon refractair prostaatkanker (HRPCa) te meten.

AS-MRI werd bij 38 patiënten verricht vóór chemotherapie voor gemetastaseerde HRPCa, naast PSA beoordeling, CT-scan van de thorax, het abdomen en bekken [CT-TAP] en Tc-99m bot scintigrafie. Een tweede AS-MRI werd verricht bij 20 patiënten die zes maanden chemotherapie hadden gekregen. De evaluatie van tumor respons werd verricht volgens RECIST.

Slechts 11 patiënten (29%) hadden op de aanvangs CT-TAP meetbare metastasen volgens RECIST in weke delen localisaties of lymfeklieren. AS-MRI diagnosticeerde een diffuse infiltratie van het beenmerg bij 8 patiënten en focale meetbare gemetastaseerde lesies bij 25 patiënten (65%). Daardoor verdubbelde het percentage patiënten met meetbase lesies. De toepassing van AS-MRI analoog aan RECIST bij 20 patiënten die zes maanden waren behandeld maakte een nauwkeurige beoordeling van complete respons (n=2), partiële respons (n=2), stabiele ziekte (n=5), of tumor progressie (n=11) mogelijk, op voorwaarde dat CT-TAP wordt gebruikt voor weke delen metastasen. Hiermee is het mogelijk bot metastasen van het axiale skelet te meten met MRI en tumor repons objectief te beoordelen. De haalbaarheid van deze benadering op grotere schaal vereist verder onderzoek.

Hoofdstuk 7 van het proefschrift beschrijft de literatuur die over RECIST is gepubliceerd van 2000 tot November 2005. Een totaal van 60 artikelen en ASCO abstracts beschreven studies of reviews in verband met RECIST en de implementatie van RECIST. Van de 60 referenties die opgenomen werden in deze review, betroffen er 11 validatie studies (7 prospectieve en 4 retrospectieve), 6 betroffen de vergelijking van eendimensionale metingen met twee- of driedimensionale metingen, 12 artikelen behandelden kwesties die opdoken bij de implementatie van RECIST bij mesotheliomen en gastro-intestinale stroma tumoren en 4 papers beschreven aanpassingen van RECIST voor specifieke tumor types.

In het algemeen is RECIST zeer goed ontvangen door de wetenschappelijke gemeenschap en de meeste validatie studies ondersteunen de implementatie van de nieuwe criteria volledig. Nochtans werden, zoals verwacht, enkele problemen herkend. Overeenkomstig de wiskundige verschillen in de definitie van progressie, valt bij RECIST de progressie in sommige later, in vergelijking met de WHO criteria. RECIST zijn niet gemakkelijk aan te passen voor sommige specifieke studies zoals bij pediatrische tumoren mesotheliomen. Bovendien is het in sommige omstandigheden mogelijk dat anatomische veranderingen in de tumor zoals beschreven door RECIST later ontdekt worden dan funktionele veranderingen, bijvoorbeeld bij gastro-intestinale stroma tumoren behandeld met Imatinib. Er bestaat echter nog geen andere universele methode van tumor beoordeling, en funktionele beeldmethoden zullen voor de eerstkomende tijd niet gevalideerd en algemeen beschikbaar zijn.

TOEKOMST PERSPECTIEF

De bevindingen van de laatst beschreven review, tesamen met de ervaring die tot nu toe is opgedaan en de resultaten van enkele lopende onderzoeksprojekten bereiden de weg voor een herziene versie van de criteria. Deze nieuwe versie zal een aantal kwesties van de vorige versie die in de laatste vijf jaar door de RECIST werk groep zijn aangepakt verduidelijken aan de hand van vragen die werden ontvangen van diegenen die RECIST in de dagelijkse praktijk toepassen en de antwoorden die op de RECIST website werden aangegeven.

Meer gegevens over het gebruik van tumor markers (en referenties naar andere gepubliceerde criteria) zullen opgenomen worden, alsook specifieke criteria voor bijzondere tumor types zoals mesothelioom. In samenwerking met onze collegae radiologen zal ook meer aandacht gaan naar aanbevelingen voor het gebruik van geavanceerde beeldtechnieken en contrast middelen. Tenslotte worden kwesties zoals de noodzaak van bevestiging van respons en het gebruik van een minimum aantal lesies dat nodig is om een correcte tumor beoordeling te verkrijgen, onderzocht door EORTC in grote databanken van recent beëindigde grote studies die RECIST toepasten. Dit gebeurt in samenwerking met de farmaceutische industrie, en de resultaten van de onderzoeken zouden beschikbaar moeten zijn tegen de tijd dat de huidige richtlijn zal worden herzien. Zoals hierboven aangegeven zal voor de volgende vijf tot tien jaar de noodzaak blijven bestaan om letsels te kwantificeren met metingen op basis van klinische waarneming of beelden, of deze nu gebruikt worden om respons te meten of tumor progressie te beoordelen. Daarom is het van het hoogste belang dat klinische onderzoeken een gelijke en hedentendaagse reeks respons criteria gebruiken teneinde een uniforme beoordeling van de activiteit en effectiviteit van nieuwe kankerbehandelingen te garanderen.

Acknowledgements

I would to thank all of those who have contributed to this work and in particular Prof J. Verweij who gently looked after me to accomplish this task and Ms Z. Tayah who did most of the formatting work of the thesis.
I dedicate this work to my parents who both suffered from cancer.

Patrick Therasse, 2006