SPECIAL ARTICLE

New Guidelines to Evaluate the Response to Treatment in Solid Tumors

Patrick Therasse, Susan G. Arbuck, Elizabeth A. Eisenhauer, Jantien Wanders, Richard S. Kaplan, Larry Rubinstein, Jaap Verweij, Martine Van Glabbeke, Allan T. van Oosterom, Michaele C. Christian, Steve G. Gwyther

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. [J Natl Cancer Inst 2000; 92:205–16]

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 end points 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 end point 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 end point 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 end point 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 end points 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, end point. The trial should be large enough that differences in response rate can be validated by association with more definitive end points 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 target 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 end point 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 end point. 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 *(1).*

(*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.

(*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 end point 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.

Table 1. Overall responses for all possible combinations of tumor responses in target and nontarget lesions with or without the appearance of new lesions*

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

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

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 end point. 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 end points. 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 end point 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 end points. 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 mecha-

nism. It is clear though that, in an uncontrolled trial proposing to utilize progession-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 end point 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 end point, 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 end point 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 end point. 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.

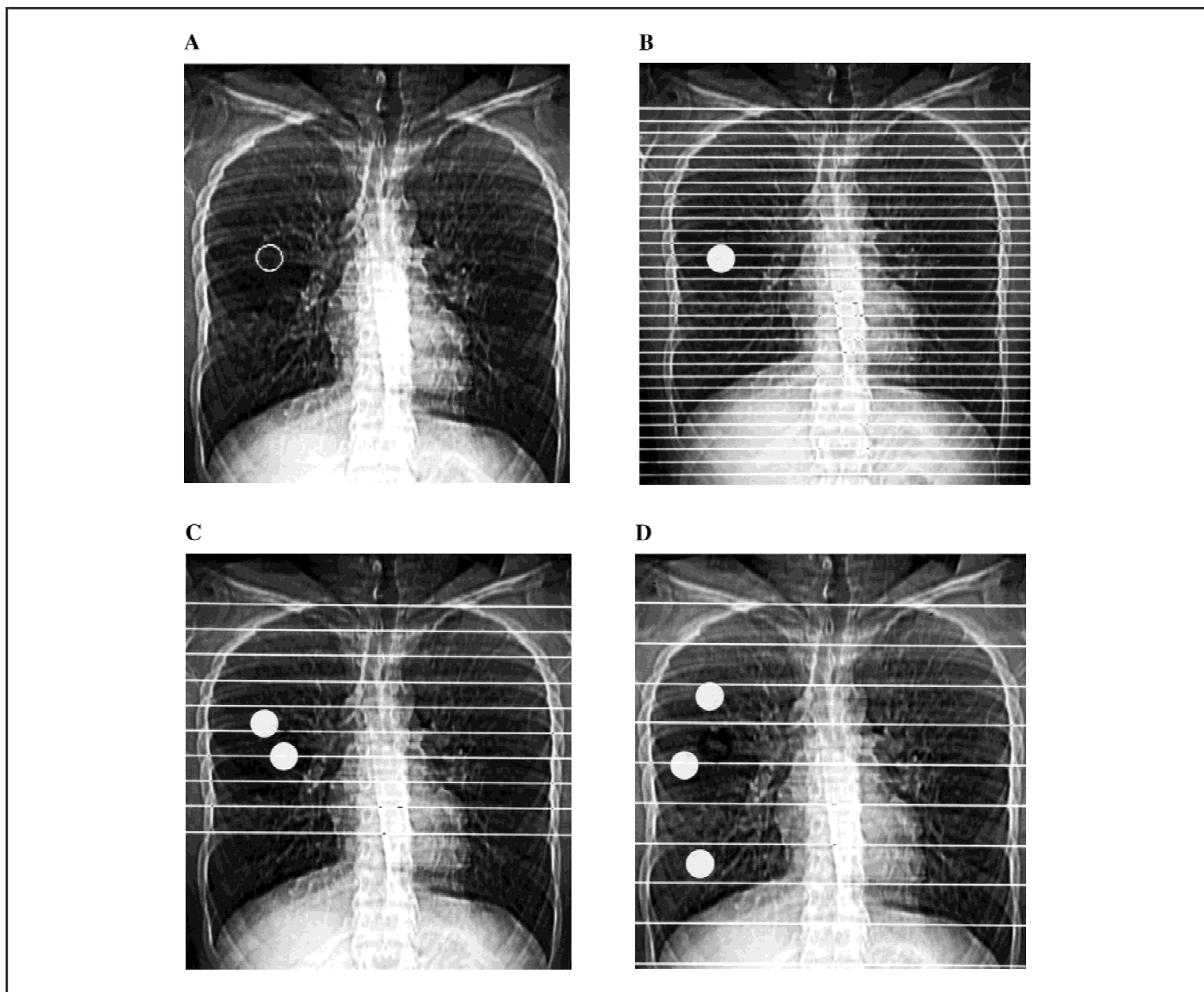Spiral CT is now the standard in most hospitals involved in cancer

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.

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 end point, 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.

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.

A

B
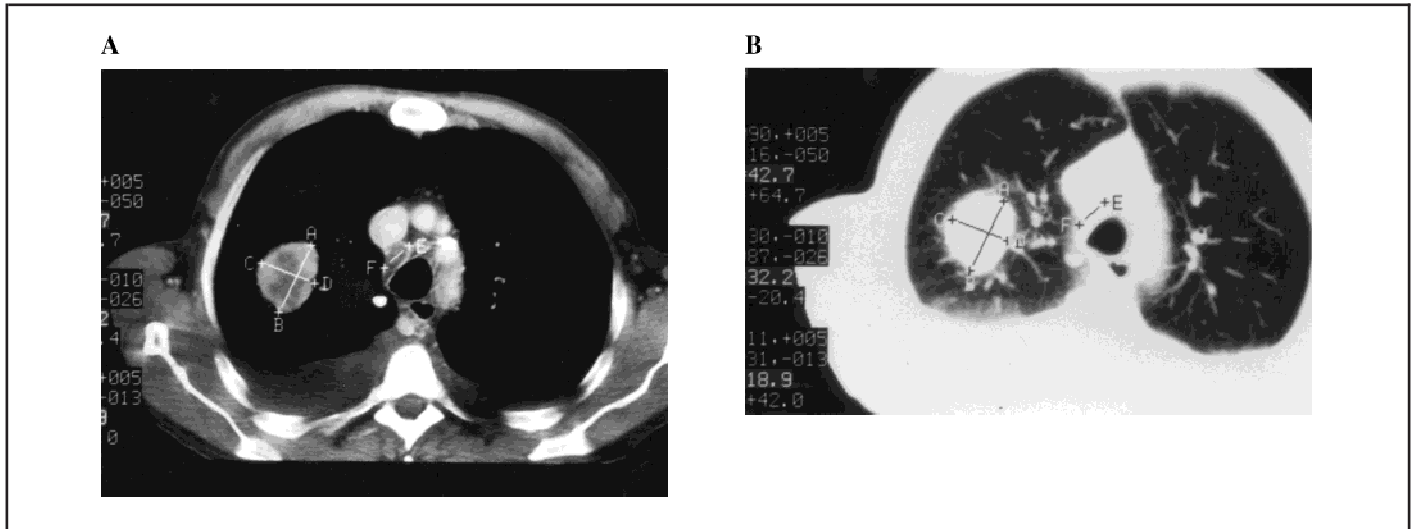
Fig 2. **A**) Computed tomography (CT) scan of the thorax at the level of the carina on "soft-tissue" windows. Two lesions have been measured with calipers. The intraparenchymal lesion has been measured bidimensionally, using the greatest diameter and the greatest perpendicular distance. Unidimensional 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.
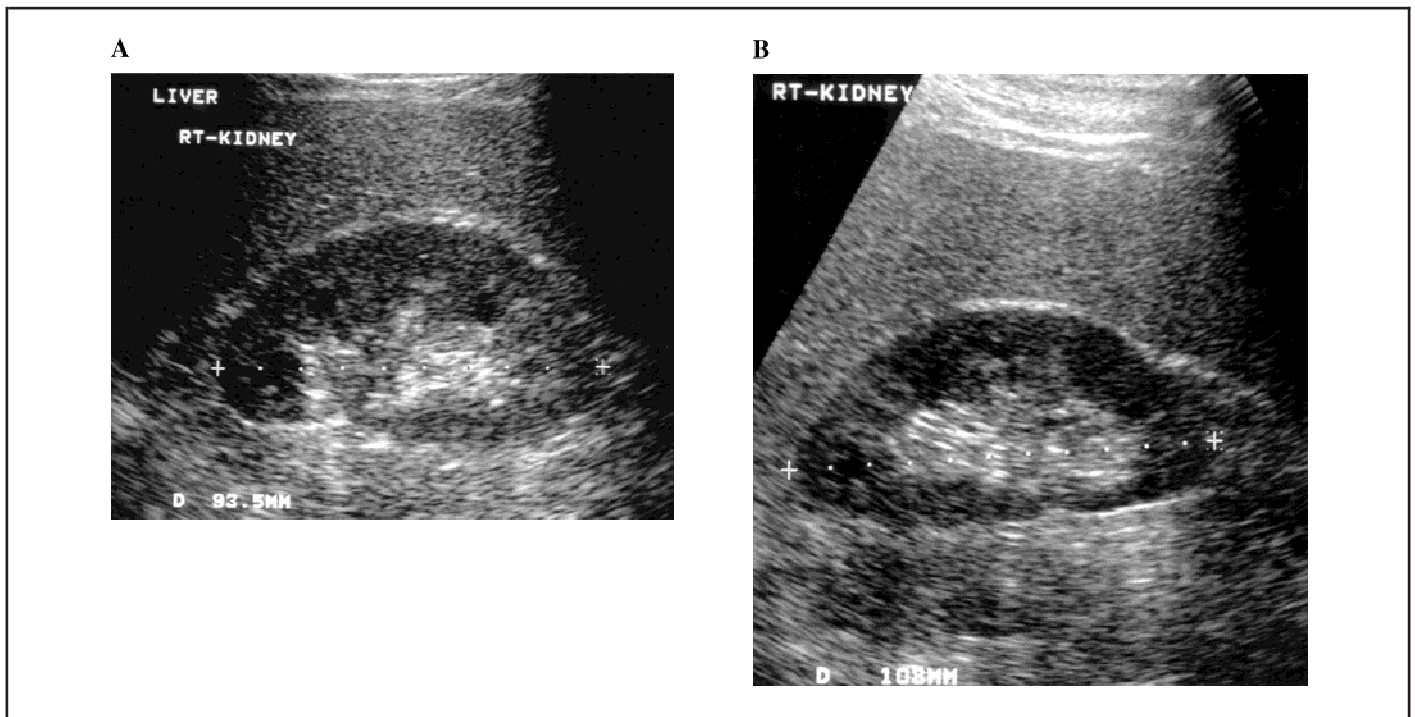

A

B

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.

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.

• 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.

• 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*

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

*Shaded areas represent the response evaluation criteria in solid 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

Retrospective Analyses

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

Participants

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, Fédération 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. Nolè, 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 analysis 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

Appendix V, Table 3. Definition of best response according to WHO or RECIST criteria*

Best response	WHO change in sum of products	RECIST change in sums longest diameters
CR	Disappearance; confirmed at 4 wks†	Disappearance; confirmed at 4 wks†
PR	50% decrease; confirmed at 4 wks†	30% decrease; confirmed at 4 wks†
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

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

†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*

Tumor site/type	Criteria	No. of patients evaluated	Best response				RR	PD rate
			CR	PR	SD	PD		
Breast†	WHO	48	4	22			54%	
	RECIST	48	4	22			54%	
Breast‡	WHO	172	4	36			23%	
	RECIST	172	4	40			26%	
Brain†	WHO	31	12	10			71%	
	RECIST	31	12	10			71%	
Melanoma†	WHO	190	9	37			24%	
	RECIST	190	9	34			23%	
Breast§	WHO	531	50	102			29%	
	RECIST	531	50	108			30%	
Colon§	WHO	1096	12	137			14%	
	RECIST	1096	12	133			13%	
Lung§	WHO	1197	60	317			32%	
	RECIST	1197	60	318			32%	
Ovary§	WHO	554	24	108			24%	
	RECIST	554	24	105			23%	
Lung†	WHO	24	0	4	16	4	17%	17%
	RECIST	24	0	4	19	1	17%	4%
Colon†	WHO	31	1	6	15	9	23%	29%
	RECIST	31	1	5	16	9	21%	29%
Sarcoma†	WHO	28	1	4	13	10	18%	36%
	RECIST	28	1	5	17	5	21%	18%
Ovary†	WHO	45	0	7	19	19	16%	42%
	RECIST	45	0	6	21	18	13%	40%
Breast‖	WHO	306	18	114	117	57	43%	19%
	RECIST	306	18	108	124	56	41%	18%
Breast‖	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%

*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.

†Data from the National Cancer Institute of Canada Clinical Trials Group phase II and III trials.

‡Data from the National Cancer Institute, United States phase III trial.

§Data from Bristol-Myers Squibb (Wallingford, CT) phase II and III trials.

‖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*

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

*PD = progressive disease; WHO = World Health Organizaiton; and RECIST = Response Evaluation Criteria in Solid Tumors.

†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*

	No. of patients	% (of 234, *see* above)
No. of progressors with differing progression dates	19	8.1
8–9 wks' difference	3	1.3
12 wks' difference	1	0.4
24–31 wks' difference†	2	0.9
Difference uncertain due to censoring of either WHO or RECIST progression time‡	13	5.6

*WHO = World Health Organization; RECIST = Response Evaluation Criteria in Solid Tumors.

†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.

‡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 cm² 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 cm². For these patients, the magnitude of the difference is entirely uncertain.

(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).

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 and 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.

NOTE