APPENDIX IV RECIST criteria

RESPONSE EVALUATION CRITERIA IN SOLID TUMORS

(RECIST criteria)

Revised version of the WHO criteria published in the WHO Handbook for reporting results of cancer treatment (Geneva, 1979)

June 1999

1. Introduction

Definitions of objective tumor response were first developed, widely disseminated and adopted by clinical researchers in the mid-late 1970s 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 definitions published in the 1979 WHO Handbook and by Miller in 1981, have been the criteria most commonly used by investigators around the globe. Their criteria for complete and partial response have been validated by their ability to select agents for further development, some of which have been later shown to improve the important outcome of survival.

However, some problems have developed with the use of WHO criteria:

1. methods for integrating 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. definitions of progressive disease are related to change in a single lesion by some and to a change in the sum of lesion products by others,

4. the arrival of new technologies (CT, MRI) has led to some confusion about how to integrate tri-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 which, as far as was possible, addressed areas of conflict and inconsistency.

In so doing, a number of principles were identified:

(a) Despite the fact that "novel" therapies are being developed which 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 4 categories of CR, PR, SD, PD should be retained in any new revision.

b) Because of the need to retain some ability to compare favourable results of future therapies with those currently available, it was agreed that no major discrepancy in what might be considered a partial response by new or old measurement criteria should exist.

(c) In some institutions the technology now exists to determine changes in tumor volume which are very precise, or changes in tumor metabolism which 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 partial response in particular is an arbitrary convention ---- there is no inherent meaning for an individual patient of a 50% change in tumor area (~ 65% volume change). Thus in developing definitions for future use, it was not felt that increased precision was an important goal for its own sake. Rather, standardization and simplification of methodology was desirable. Nevertheless, the guidelines proposed in this document are not meant to discourage development and validation of new techniques which may provide more reliable surrogates endpoints than objective tumor response to predict a potential therapeutic benefit for future cancer patients.

(d) 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 to adopt criteria requiring a greater increase in size to consider a patient to have PD. These concerns have led to a similar change within these revised WHO criteria.

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

Screening measures other than objective tumor shrinkage that may appropriately allow selection of cytostatic agents in phase II trials

Definitions of serum marker response and recommended methodology for their validation

Specific tumors or anatomical sites presenting unique complexities.

Important note

Data from collaborative studies including more than 4000 patients assessed for tumor response support simplifying 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. This new approach is implemented in the present document. The results of the different retrospective analyses (comparing both approaches) performed using these different databases are described in appendix II, III and IV.

2. Various purposes of “Response Evaluation”

This paper is intended to explore the definitions, assumptions and purposes of tumor response criteria and offer guidelines that 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, non-investigational 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:

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 are typically phase II trials of investigational agents/regimens (see section 2.2 below) and it is for use in this precise context that the present document has been developed;
as a prospective endpoint in more definitive clinical trials designed to provide an estimate of benefit for a specific cohort of patients. These are often randomized comparative trials or single-arm comparisons of combinations of agents with historical controls. In this setting, objective tumor response is used as a surrogate endpoint for other measures of clinical benefit, including time to event (death, disease progression) and symptom control (see section 2.3 below).

to guide the clinician and patient/subject in decisions about continuing the current therapy. This purpose is applicable both to clinical trials and to routine practice (see section 2.1 below), but use in this context 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.

2.1 Response outcomes in daily clinical practice of oncology

The evaluation of tumor response in daily clinical practice of oncology may not be performed according to predefined criteria. It may be based on a subjective medical judgement 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.

2.2 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 complete responses are sometimes employed in a similar fashion. Utilization of response rate in this way is not encumbered by an implied assumption as to the therapeutic benefit of such responses but rather implies some degree of biological antitumor activity of the investigated agent.

For certain types of agents (notably 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 direct investigators to select anticancer agents for further study. Some agents selected this way eventually proved 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 proven to be of clinical value in later more definitive trials.

However, in most circumstances, a new agent achieving a prospectively determined response rate considered of sufficient interest to warrant further testing may not prove in subsequent randomized phase III trials to be an effective treatment for the studied disease. Random variables as well as selection biases (both known and unrecognized) can have an overwhelming effect in small uncontrolled trials. These trials are an efficient and economical 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 prior to 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.

2.3 Response outcomes in clinical trials as a surrogate for palliative effect

2.3.1 Usage in non-randomized clinical trials

The only circumstance in which objective responses in a non-randomized trial can permit a tentative assumption of a palliative effect is when there is an actual or implied comparison with historical series of similar patients. This assumption is strongest when the prospective statistical plan provides for matching

of relevant prognostic variables between cases and a defined series of controls. Otherwise, there must be at the very least a prospective statistical plan that provides 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 non-quantifiable biases inherent in non-randomized trials, proof of benefit still requires eventual confirmation in a prospectively randomized controlled trial of adequate size. The appropriate therapeutic benefit endpoints for such a trial are survival, progression free survival, or symptom control (including quality of life).

2.3.2 Usage 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 correlation with more definitive endpoints reflecting therapeutic benefit such as survival, progression free survival, reduction in symptoms or improvement (or maintenance) of quality of life.

3. Measurability of tumor lesions at baseline

3.1 Definitions

At baseline, tumor lesions will be categorized as:

* 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 3.2).
OR * non-measurable:	all other lesions, including small lesions (longest diameter < 20 mm with conventional techniques or < 10 mm with spiral CT scan) and truly non-measurable 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, using a ruler or calipers. All baseline evaluations should be performed as close as possible to the treatment start and never more than 4 weeks before the beginning of the treatment.

Lesions that are considered as truly non-measurable 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;

• cystic lesions;

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

3.2 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.

3.2.1 Clinical lesions

Clinical lesions will only be considered measurable when they are superficial (e.g. skin nodules, 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.

3.2.2 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 V.

3.2.3 CT, MRI

CT and MRI might be the best currently available and reproducible methods to measure target lesions selected for response assessment. Conventional CT and MRI should be performed with cuts of 10 mm or less in slice thickness contiguously. Spiral CT should be performed using a 5 mm contiguous reconstruction algorithm. This applies to the chest, abdomen and pelvis. Head & neck and extremities usually require specific protocols.

More details concerning the use of these methods of assessment for objective tumor response evaluation are provided in appendix V.

3.2.4 Ultrasound

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

3.2.5 Endoscopy, laparoscopy

The utilization of these techniques for objective tumor evaluation has not yet been fully and widely validated. Their uses in this specific context require sophisticated equipment and a high level of expertise that may only be available in some centers. Therefore, the utilization of such techniques for objective

tumor response should be restricted to validation purposes in reference centers. However, such techniques can be useful to confirm complete pathological response when biopsies are obtained.

3.2.6 Tumor markers

Tumor markers alone cannot be used to assess response. If markers are initially above the upper normal limit, they must normalize for a patient to be considered in complete clinical response. Specific additional criteria for standardized usage of PSA and CA-125 response in support of clinical trials are being developed.

3.2.7 Cytology, histology

These techniques can be used to differentiate between PR and CR in rare cases (for example, residual lesions in tumor types such as germ cell tumors, where known residual benign tumors can remain).

The cytological confirmation of the neoplastic origin of any effusion that appears or worsens during treatment when the measurable tumor has met criteria for response or stable disease is mandatory to differentiate between response or stable disease (an effusion may be a side effect of the treatment) and progressive disease.

Note: New techniques to better establish objective tumor response will be integrated into these criteria when they are fully validated with standard procedures and definitions to be used in the context of tumor response evaluation. In the meantime, the criteria provided in this document provide the standard to which newer techniques should be compared in investigational studies.The use of these standardized criteria for comparison to criteria based upon new technology is expected to facilitate and speed development of better approaches.

4. Tumor response evaluation

4.1 Assessment of overall tumor burden and measurable disease

4.1.1 **To assess objective response, it is necessary to estimate the overall tumor burden at baseline and use this as a comparator for subsequent measurements. 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 3.1). If the measurable disease is restricted to a solitary lesion, its neoplastic nature should be confirmed by cytology/histology.**

4.1.2 Baseline documentation of “Target” and “Non-Target” lesions

All measurable lesions up to a maximum of 10 lesions representative of all involved organs should be identified as ***target lesions*** and will be recorded and measured at baseline. Target lesions should be selected on the basis of their size (lesions with the longest diameter) and their suitability for accurate repetitive measurements (either by imaging techniques or clinically). A sum of the longest diameter (LD) for *all target lesions* will be calculated and reported as the baseline sum LD. The baseline sum LD will be used as reference to further characterize the objective tumor response of the measurable dimension of the disease.

All other lesions (or sites of disease) should be identified as ***non-target lesions*** and should also be recorded at baseline. Measurements are not required and these lesions should be followed as “present” or “absent”.

4.2 Response Criteria

4.2.1 Evaluation of target lesions

* Complete Response (CR):	disappearance of all target lesions.
* Partial Response (PR):	at least a 30% decrease in the sum of LD of target lesions taking as reference the baseline sum LD.
* Progression (PD):	at least a 20% increase in the sum of LD of target lesions taking as references the smallest sum LD recorded since the treatment started or the appearance of one or more new lesions.
* Stable Disease (SD):	neither sufficient shrinkage to qualify for PR nor sufficient increase to qualify for PD taking as references the smallest sum LD since the treatment started.

4.2.2 Evaluation of non target lesions

* Complete Response (CR):	disappearance of all non-target lesions and normalization of tumor marker level.
* Non-Complete Response (non-CR)/ Non–Progression (non-PD):	persistence of one or more non-target lesion or/and maintenance of tumor marker level above the normal limits.
* Progression (PD):	appearance of one or more new lesions. Unequivocal progression of existing non-target lesions (1).

Although a clear progression of "non target" lesions only is exceptional, in such circumstances, the opinion of the treating physician should prevail and the progression status should be confirmed later on by the review panel (or study chair).

4.2.3 Evaluation of best overall response

The best overall response is the best response recorded from the start of the 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 4.3.1).

Target lesions	Non-Target lesions	New Lesions	Overall response
CR	CR	No	CR
CR	Non-CR/Non-PD 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

Note:

Patients with a global deterioration of health status requiring discontinuation of treatment without objective evidence of disease progression at that time should be reported as "symptomatic deterioration". Every effort should be made to document the objective 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, treatment periodicity).

In some circumstances it may be difficult to distinguish residual disease from normal tissue. When the evaluation of complete response depends upon this determination, it is recommended that the residual lesion be investigated (fine needle aspirate/biopsy) before confirming the complete response status.

4.3 Confirmatory measurement / Duration of response

4.3.1 Confirmation

The main goal of confirmation of objective response is to minimize the risk of overestimation of the response rate. This aspect of response evaluation is particularly important in non-randomized trials where response is the primary endpoint. In this setting, to be assigned a status of PR or CR, changes in tumor measurements must be confirmed by repeat studies 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 SD, follow-up measurements must have met the SD 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 4.3.3).

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 should be made clear when reporting the outcome of such studies.

4.3.2 Duration of overall response

The duration of overall response is measured from the time measurement criteria are met for CR/PR (whichever is first recorded) 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 CR until the first date that recurrent disease is objectively documented.

4.3.3 Duration of stable disease

Stable disease is measured from the start of the treatment until the criteria for progression are 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 between trials are to be made.**

4.4 Progression Free Survival

This document is focused primarily on the use of objective response endpoints. In some circumstances (e.g. brain tumors, investigation of non-cytoreductive anticancer agents), “response evaluation” may not be the optimal method to assess the potential anticancer activity of new agents/regimen. In such cases “progression free survival” (PFS) can be considered a valuable alternative to provide an initial estimate of biologic effect of new agents that may work by a non-cytotoxic mechanism. It is clear though that in an uncontrolled trial proposing to utilize PFS, it will be necessary to document with care the basis for estimating what PFS 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, for example with regard to selection and ascertainment. Uncontrolled trials using PFS 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.

5. Response review

For trials where the response rate is the primary endpoint it is strongly recommended that all responses be reviewed by an expert(s) independent of the study. Simultaneous review of the patients’ files and radiological images is the best approach.

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

6. Reporting of results

Every report should contain all patients included in the study. For the response calculation the report should contain at least a section with all eligible patients. Another section of the report may

detail the response rate for evaluable patients only. However, a response rate analysis based on a subset of patients must explain which patients were excluded and for which reasons. It is preferred that 95% confidence limits are given.

Response evaluation in Randomized Phase III trials Some of the guidelines recommended in this document 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. 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. Response evaluation in phase III trials may be an indicator of the relative antitumor activity of the treatments evaluated but may not solely predict the real therapeutic benefit for the population studied.**