Evaluation of response: new and standard criteria

P. Therasse

European Organisation for Research and Treatment of Cancer (EORTC) Data Center, Brussels, Belgium

Introduction

Most cancer patients suffering from some form of advanced or locally advanced cancer will be treated with anticancer agents to stop the natural evolution of the disease. Anticancer agents developed over the last four decades behave as 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 size of tumor lesions that were clinically or radiologically evaluable. This process of evaluation of response to treatment is nowadays integrated into the daily practice of every oncologist and is also used in the investigational setting (clinical trials) to define the antitumor 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.

Evolution of methodology

In the first clinical trial in solid tumors initiated in the 1950s, tumor response was already taken as an endpoint based on the subjective evaluation reported by the physician [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 tumors [3]. The principles of response evaluation, which are still valid today, can be summarized as follows:

• Overall cancer burden can be characterized by a quantitative evaluation of tumor lesions, which are measurable, and a qualitative 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 methods of investigation 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, which 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, insufficient shrinkage of measurable tumor lesions; progression, increase of measurable tumor lesions beyond a certain percentage or appearance of one or more new lesions.
• Complete and partial response should be confirmed with a second examination, which should take place within a certain time scale after the responses were originally observed.
• To avoid large variations due to interobserver variability, it is recommended that responses should be reviewed and confirmed by an independent panel of experts.

The specific aspects of the WHO/UICC criteria, amongst others, are as follows:

• Measurable lesions are quantified either by their surface (bidimensional lesions, product of longest perpendicular diameters) or by the longest diameter (unidimensional lesions) when only one dimension can be accurately measured.
• Bone lesions can be measured following specific criteria.
• Tumor load is evaluated for each organ independently and the overall response to treatment is calculated by combining the responses observed in each organ.
• Partial response is attributed when a decrease of 50% in the entire tumor burden (objectively for measurable lesions and subjectively for others) is recorded. Progression status is assigned when there is an increase of 25% in one specific tumor lesion or if the tumor load in one organ increases globally by 25% (based on the same principles as for partial response) or if a new lesion is discovered.

Since 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 detail in the WHO recommendations have stimulated the development of amended 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 1990s 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 these different versions (some of them published and others unpublished) of the original WHO criteria have rendered the accuracy of relative comparisons 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 as well as in terms of interobserver variability [6–9].

The evolution of cancer imaging, the importance given to the response rate endpoint and the increasing number of new anticancer agents to be tested (and therefore the number of centers to be involved in drug development) have been other factors demanding a coordinated effort to revisit, update and possibly improve the existing criteria.

This difficult exercise started in 1996 with representatives of three cancer research organizations: EORTC, US National Cancer Institute and the National Cancer Institute of Canada, Clinical Trials Group. A comprehensive revised version of the WHO criteria was published in February 2000 [10] under the acronym RECIST (Response Evaluation Criteria In Solid Tumors). These criteria have been rapidly 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 tumor lesions are based on the longest diameter only (unidimensional measurement).
- Cancer lesions are considered as measurable only when their longest diameter is \( \geq 2 \) cm when measured with conventional techniques or \( \geq 1 \) cm when measured by spiral computed tomography (CT) scan.
- Precisions are given as to which method of investigation can be used and how it can be used.
- The overall tumor load is represented by pre-selected measurable lesions which are set as target lesions whilst all other lesions (measurable and non-measurable) 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 \( \geq 30\% \).
- Response status should be confirmed after a minimum interval of 4 weeks. Stable disease is defined following two evaluations that are separated by an interval 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 in order to define the best overall response.
- CA125 can be used as an indicator that alone may determine progression of disease after first-line treatment in advanced ovarian cancer.

The use of one dimension only to measure tumor lesions has been based on the work published by James et al. [11]. Retrospective analysis using a cohort of patients from 14 different studies (>4000 patients) demonstrated that using two dimensions or one dimension for tumor lesion 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 tumor growth is required to define disease progression.

Although the RECIST criteria were launched in 2000, the harmonization process of the response criteria has continued through the implementation of a question and answers section on the internet (http://www.eortc.recist.be) which provides further clarification or additional information for specific aspects relating to response evaluation using these criteria. Also, proposals to modify the existing criteria or add new criteria are considered on a regular basis by the RECIST working group. Adaptations of the RECIST criteria are currently being studied for specific tumor types such as brain tumors, mesothelioma and pelvic tumors.

Beside the RECIST criteria developed to be applicable to most solid tumor types, specific criteria have been developed for evaluating response and progression in non-Hodgkin’s lymphoma [12] and prostate cancer [13].

**Summary**

Standard response criteria are necessary to quantify the degree of anticancer activity of new anticancer agents. These are used principally in phase II clinical trials that aim to define the response rate of a new agent or new combination of agents. They are also used to define the progression status of patients included in any clinical trial. The most commonly used set of criteria have been the WHO criteria [3]. In 2000, the RECIST criteria [10] based on unidimensional measurement of tumor lesions were published and are now commonly used in place of the WHO criteria.

**References**


