

# Setting Standards

Stephen Bates and Kelie Luby at Perceptive Informatics review the common assessment criteria used in oncology clinical trials

RECIST, WHO, IWG, NCI-WG, PCWG2 – in the realm of oncology, these acronyms provide information on everything from the type of disease to the way the disease is assessed. Standardised oncology assessment criteria are essential to the validity of oncology clinical trial outcome data. Imaging as a surrogate endpoint has become a major focus in oncology clinical trials due to the ease of detection and quantification with conventional radiography techniques, further enhancing the validity of trial data. However, the use of imaging as an endpoint in clinical trials studies is hampered by a lack of a firm endorsement by regulatory agencies (principally in the US and EU) on specific quantification methods, leaving much of the actual protocol design, and thus application of standard oncology assessment criteria, in the hands of the individual sponsors. As a result, sponsors have focused on a relatively small number of academic articles for standardised assessment criteria and these criteria are favourably accepted by regulatory agencies if applied and modified appropriately. However, once data is generated based on the selected assessment criteria, sponsors will be subjected to scrutiny by regulators as to how the criteria was applied, particularly if applied in such a way as to bias the outcome of results. To this end, it is a worthwhile step to ensure that regulators are in agreement with the way a particular assessment criteria will be applied in a trial prior to study start up.

## THE BASIS FOR ONCOLOGY ASSESSMENT CRITERIA FOR IMAGING

The foundation of any oncology assessment criteria utilised to assess

imaging is to attempt to standardise and quantify lesion measurements into an objective overall patient assessment. At the baseline, imaging assessment lesions are generally categorised as measurable or non-measurable. Quantitative measurements are taken when a lesion is deemed measurable, otherwise qualitative assessments are made (for example present, absent or progressed). Lesions are defined as ‘new’ if they appear after the baseline imaging exam. Various changes in these three categories are tracked and assessed over subsequent imaging assessments, and then combined into an overall patient assessment. Often, critical details on lesions are ‘missing’ from the assessment criteria. Thus, appropriately modifying standardised assessment criteria to account for these overlooked scenarios is the first hurdle in validating clinical trial data.

The physical, and thus imaging manifestations of the therapeutic indication, influence the complexity of this assessment. For example, lung cancer tends to form numerous, sometimes hundreds, of small nodules in the lung; ovarian cancers tend to form a large, solitary, irregular lesion (known as omental cake) that entwines normal tissue; prostate cancers tend to form bone lesions that are difficult to measure because they are encased in bony tissue and typically do not change size in response to treatment. Additionally, it is not uncommon for the individual lesion assessments to conflict (for example, decreasing baseline disease burden, but appearance of new disease/lesions), which adds another layer of complexity when attempting to collate these individual responses into a single overall response.

## FIRST STEPS IN STANDARDISING ASSESSMENTS: WHO

In 1979, one of the first attempts to quantify endpoints in oncology was undertaken by the World Health Organization via the *WHO Handbook for Reporting Results of Cancer Treatment* (1). This handbook was the result of an international effort by The European Organisation for Research on Treatment of Cancer (EORTC), The National Cancer Institute of the United States (NCI), the International Union against Cancer (UICC), and other organisations. The goal of this publication was to provide an internationally acceptable common language for oncology studies, as well as similar quantification of results to allow accurate metadata analysis. Although WHO focused more on clinical endpoints, the response determinations that were suggested apply directly to imaging studies and were eagerly seized upon for guidance on imaging-related endpoints.

Many of the basic premises of tumour assessment and terminology introduced by WHO remain familiar today. For example, the ‘three categories’ of lesions (measurable, non-measurable and new) continue to be echoed in more recent criteria, and some definitions (such as that for complete response) have persisted. In recent years, WHO has been gradually supplanted by newer methods that target specific indications or mechanisms of action.

Most recently, in 2008, the WHO methodology underwent a modification by Hodi *et al* (2). The authors addressed a shortcoming of the original WHO article in that it made little allowances for immune



**Table 1: RECIST modifications to the WHO criteria**

WHO limitation	Concern	RECIST recommended modification
Unclear how to integrate response assessments from measurable and evaluable (non-measurable) lesions	No standardisation among local researchers  Particular confusion over conflicting responses in lesion types (for example, progressing measurable lesions, but responding evaluable lesions)	Established new lesion nomenclature consisting of 'target' (measurable) and 'non-target' (non-measurable) lesions as well as a methodology to compare assessments in both categories (in addition to 'new lesions') to provide a clear overall assessment
No criteria for minimum measurable lesion size	No standardisation among local researchers  Differing imaging techniques vary in precision	Minimum measurable lesion size is defined as double the slice interval for CT/MRI, but not smaller than 20mm with conventional techniques (such as conventional CT, X-ray) or 10mm with spiral CT
Ambiguous definition of progressive disease (PD) via tumour measurement	Some local researchers defined PD based on a single lesion, while others took into account the total lesion burden  Researchers varied in the number of lesions selected and measured	Progressive disease via tumour measurement is defined as a 20 per cent increase in the sum of total lesion measurements as compared to nadir  A maximum of 10 measurable lesions was established with no more than five in any single organ
Does not clearly address three dimensional imaging techniques	Advances in CT and MRI technology and access resulted in confusion with regard to three dimensional measuring techniques	Established all measurements will be one-dimensional and measured in the axial plane

response. Lesions undergoing immune response therapy, such as with monoclonal antibodies, will often inflame during the early stages of therapy. This inflammation can mimic lesion progression and as such, can cause premature treatment discontinuation if handled according to the common definitions which end treatment at the time of progression. These immune-related response criteria (irRC) seek to minimise false positive disease progression utilising a number of modifications to the original WHO criteria such as assessing only measurable lesions (greater than 10mm), as well as including the measurements of new lesions into the sum of all measurable lesions. A major result of the irRC criteria is that it allows for patients to continue therapy beyond what other criteria would conclude as progression events. The irRC has piqued interest in the industry with its new and novel approach and as a result there have been several workshops sponsored by the FDA and

American Society of Clinical Oncology to further discuss and develop this criteria.

#### NEW GUIDANCE: RECIST

In 2000, Therasse *et al* authored what may be considered the most influential publication to date regarding oncology imaging endpoints (3). In many ways, the RECIST article was written as a follow-on to WHO as it sought to address many of WHO's ambiguities (see Table 1). The RECIST article also established a number of other principles of oncology imaging endpoints that were subject to various interpretations and debate.

Despite its widespread use, the RECIST article continues to exclude the newer technologies, recommending that simple uni-dimensional measurements should be used for all measurable lesions. However, the article cites a series of studies that compared the WHO and RECIST measurement

methodologies using the same patients in retrospective fashion. After analysing 234 patients from 14 separate studies encompassing eight cancer types, an overall response concordance rate of 91.9 per cent was achieved between the two methodologies, suggesting the simplicity and reproducibility of simpler measurements was effective for the broad implementation required for late phase clinical trials.

#### MODIFICATIONS TO RECIST: PCWG2

While RECIST provided guidance in many areas where WHO did not, RECIST did not fully meet the need for study-by-study modifications in imaging-based studies. Prostate cancer, in particular, is very difficult to quantify and assess with imaging due to bone infiltration. Metastatic prostate cancer primarily presents itself as numerous and painful bone lesions. Not only are they difficult to detect with standard imaging techniques (CT scans and x-rays), but due to the bony nature of the tissue, it is virtually impossible to detect changes in lesion size. In 2009, Scher *et al* provided updated guidance for this particular indication (4) and made a few major modifications to the RECIST assessment including:

- Biomarker, prostate specific antigen (PSA), testing
- To prevent false positive calls of progression, a minimum of 12 weeks must elapse before the first assessment of either PSA or imaging (principally radionuclide bone scans)

- Detection of two new lesions on a bone scan is required for a determination of progressive disease

### THE NEW RECIST: RECIST 1.1

RECIST and its modifications remain widely used by sponsors since its original publication almost a decade ago. In that time, however, several concerns surfaced. In response, many of the original RECIST contributors regrouped and published a new revision in 2009. In an article by Therasse *et al*, they make additional refinements to the criteria in response to the ever-increasing amount of data available reflecting its actual use in trials (5). These revisions include decreasing the maximum number of measurable lesions from 10 to five and a subsequent reduction in target lesions per organ from five to two. Additionally, RECIST 1.1 utilises a bi-dimensional assessment method for lymph nodes only and redefines pathological lymph nodes as those greater than 10mm. Lastly, RECIST 1.1 requires a 20 per cent increase in the sum of total lesion diameters and a 5mm minimum absolute change for progression determinations. Qualitative use of FDG-PET is also allowable to support assessments of progressive disease.

### BEYOND RECIST & WHO: LIQUID TUMOUR ASSESSMENT CRITERIA

Chronic lymphocytic leukemia (CLL) and non-Hodgkin's lymphoma (NHL) are closely related diseases that, due to their 'liquid tumour' character, pose particular problems with standardised assessments. A series of articles, much like RECIST, have become the *de facto* standard for these indications.

#### CLL

The 1988 National Cancer Institute-sponsored Working Group (NCI-WG) for chronic lymphocytic leukemia (CLL), published guidelines for the assessment of patients with CLL in clinical trials with the goal of providing standardised assessment criteria for comparisons between clinical trials and a basis for evaluation of future scientific studies. In 1996, Cheson *et al* revised the criteria to include both clinical trial and general practice recommendations and incorporate advances made since the 1988 NCI-WG criteria (6). Most recently in 2008, Hallek *et al* updated the NCI-WG recommendations for CLL due to

advances in technology and methods of evaluation (7).

#### NHL

In 1999's IWG criteria, Cheson *et al* published new criteria for standardised assessment of Non-Hodgkin's lymphoma which sought to address the wide variance in assessment of response in lymphoma such as normal lymph node size, requirements for progression, assessment frequency, and methods used for response assessment (8).

In 2007, Cheson *et al* revisited the criteria, suggesting advances since the original publication of the criteria, misinterpretations of the original criteria, and adoption of FDG-PET required revisions to be made (9). The revised guidelines include mechanisms which seek to address residual masses more effectively and included PET imaging in the assessment of response. Additionally, the often misused Complete Remission Unconfirmed (CRu) assessment category was eliminated.

### NEW CHALLENGES

These criteria (WHO, RECIST, IWG, NCI-WG and IWCLL) and their related modifications have been well accepted, but focus almost entirely on physical characteristics of cancerous lesions that can be detected with conventional imaging techniques, and even the human eye (for example skin lesions). Assessments of this type are advantageous as they can be easily assessed with standard imaging, are relatively easily understood by all physicians, and require little computational equipment. However, the criteria are very poor at determining metabolic functions of cancer cells and instead can only imply these functions by growth or regression of the tumour size. Functional imaging (such as PET, SPECT and DCE-MRI) offer mechanisms to overcome these limitations by assessing both the actual amount of living tumour tissue as well as its metabolic condition. The benefits of these techniques are obvious but there are roadblocks to wide-scale implementation. The limited number of scanners and facilities trained to use them, as well as the difficulty to validate and reproduce algorithms needed to assess lesions, have restricted the use of these methods.

### Choi Criteria

Despite the limitations, standardised criteria around functional imaging are starting to gain more attention and are becoming more common. In 2004, Choi *et al* published data from a unique study that directly compared lesions assessed with both CT (RECIST) and PET imaging (10). In this study, 173 gastrointestinal stromal tumours from 36 patients where morphologically assessed via RECIST guidelines and also metabolically assessed by measuring SUV via FDG PET. The results of this comparison suggested that RECIST tends to underestimate response in these patients as the overall size of lesions remains steady or decreases only marginally as compared to more radical changes in metabolic activity and density. However the authors point out that this technique is virtually useless for lesions that are FDG negative (approximately 21 per cent in this case) at baseline.

Choi *et al* again published a revised criteria in 2007 that fully marries RECIST and PET (11). This article claims to validate the original Choi methodology by demonstrating prospectively that the Choi criteria correlates to both time to tumour progression and disease specific survival far better than standard RECIST. Choi criteria requires only a 10 per cent decrease in size (as opposed to RECIST's 30 per cent) and a 15 per cent decrease in volume (which is not assessed at all in RECIST) in order to determine response. Although interesting, the application of this criteria is quite limited (GIST patients are a small fraction of cancer patients). What it does suggest is that further research may validate its application in other disease indications.

In 2006, the National Cancer Institute published one of the first comparisons of the multiple PET methods (12). This publication discusses the advantages and disadvantages of the three most common FDG-PET assessment methods: visual, standardised uptake value (SUV) and kinetic. While each method has benefits and drawbacks, the authors acknowledge that there is no single best method but do recommend kinetic assessments only for early phase imaging studies.

Dynamic contrast enhanced magnetic resonance imaging (DCE-MRI) is another

novel imaging technique that is attempting to fill in the gaps left by morphologic imaging techniques. DCE-MRI offers the ability to assess both the amount of angiogenesis as well as the ‘integrity’ of tumour vasculature in order to assess the effects of treatment on this mechanism. In 2007, Yankeelov and Gore produced a nearly encyclopaedic reference for this technique including discussions around the basic physics, acquisition techniques and data analysis algorithms (13).

## CONCLUSION

The assessment of imaging in oncology clinical trials is well supported by standard assessment criteria. However, modifications are often needed in order to correctly quantify changes observed on imaging. Morphologic criteria (for example WHO, RECIST, IWG, IWCLL, and so on) lead the way in terms of simplicity, reproducibility and general regulatory agency acceptance. However, limitations in these morphologic techniques and advances in functional imaging (notably FDG PET and DCE MRI) are culminating with a radical rethinking of the existing criteria. Contemporary revisions of the morphologic criteria are increasingly incorporating functional imaging into disease assessment. However, lack of widespread use, cost, ease of standardisation and complexity are limiting large scale adoption of functional techniques. Advances in pharmaceutical development suggest continued adoption of the newer imaging techniques that in many cases are the best methods available to assess early efficacy. This will no doubt lead to the development of new criteria as the need for faster and better efficacy assessments continues unabated.

## References

1. World Health Organization, *WHO Handbook for reporting results of cancer treatment*, Geneva 1979
2. Hodi F, Hoos A, Ibrahim K, Chin H, Pehamberger H *et al*, Novel efficacy criteria for antitumor activity to immunotherapy using the example of ipilimumab, an anti-ctla-4 monoclonal antibody, *Journal of Clinical Oncology* 26(155): pp2,008-3,008, 2008
3. Therasse P, Arbuck S, Eisenhauer E *et al*, New guidelines to evaluate the

response to treatment in solid tumors, European Organization for Research and Treatment of Cancer, National Cancer Institute of the United States, National Cancer Institute of Canada, *J Natl Cancer Inst*, 2; 92(3): pp205-216, 2000

4. Scher H, Halabi S, Tannock I *et al*, Design and endpoint of clinical trials for patients with progressive prostate cancer and castrate levels of testosterone: recommendations of the prostate cancer clinical trials working group, *Journal of Clinical Oncology* 26(7): pp1,148-1,159, 2009
5. Eisenhauer E, Therasse P, Bogaerts J *et al*, New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1), *European Journal of Cancer* 45: pp228-247, 2009
6. Cheson B, Bennett J, Grever M *et al* National cancer institute-sponsored working group guidelines for chronic lymphocytic leukemia: revised guidelines for diagnosis and treatment, *Blood* 87(12): pp4,990-4,997, 1996
7. Hallek M, Cheson B, Catovsky D *et al*, Guidelines for the diagnosis and treatment of chronic lymphocytic leukemia: a report from the International Workshop on Chronic Lymphocytic Leukemia updating the National Cancer Institute-Working Group 1996 guidelines, *Blood* 111(12): pp5,446-5,456, 2008
8. Cheson B, Horning S, Coiffier B *et al*, Report of an international workshop to standardise response criteria for non-hodgkin's lymphoma, *Journal of Clinical Oncology* 17(4): pp1,244-1,253, 1999
9. Cheson B, Pfistner B, Juweid M *et al*, Revised response criteria for malignant lymphoma, *Journal of Clinical Oncology* 25(5): pp579-586, 2007
10. Choi H, Charnsangavej C and De Castro Faria S, CT evaluation of the response of gastrointestinal stromal tumors after imatinib mesylate treatment, *American Journal of Roentgenology* 183: pp1,619-1,628, 2004
11. Benjamin R, Choi H, Macapinlac A *et al*, We should desist using RECIST, at least in GIST, *Journal of Clinical Oncology* 25(13): pp1,760-1,764, 2007
12. Shankar LK, Hoffman JM, Bacharach S, Graham MM, Karp J, Lammertsma AA, Larson S, Mankoff DA, Siegel BA, Van den Abbeele A, Yap J and Sullivan D, Consensus recommendations for the use of 18F-FDG PET as an indicator of therapeutic response in patients in National Cancer Institute Trials, *J Nucl Med*, 47(6): pp1,059-1,066, 2006
13. Yankeelov T and Gore J, Dynamic contrast enhanced magnetic resonance imaging in oncology: theory, data acquisition, analysis, and examples, *Current Medical Imaging Reviews* 1(3): pp91-107, 2007

## About the authors



Stephen W Bates joined Perceptive Informatics Inc, a PAREXEL company, in 2001 as Program Director of the Oncology/Cardiology Medical Imaging Group. During the past eight years, he has had extensive experience in a wide range of clinical trials and imaging techniques, including the performance of both oncology (RECIST/Cheson) and CNS volume measurement. Stephen has a BSc in

Cell/Molecular Biology and an MS in Technology Management, both from Bridgewater State College. Email: [stephen.bates@perceptive.com](mailto:stephen.bates@perceptive.com)



Kelie Luby is a Senior Medical Writer at Perceptive Informatics. Kelie's responsibilities include developing trial-specific independent review design as defined by the Independent Review Charters and Reviewer Manual to support regulatory submissions and requirements. Kelie has a BSc in Chemistry and an MS in Organometallic Chemistry from the University of North Carolina, US. Kelie also holds a MSC in

Technical and Professional Writing with a concentration in medical, scientific and regulatory writing from Northeastern University where she also taught undergraduate technical writing. Kelie is a member of the American Medical Writers Association.

Email: [kelie.luby@perceptive.com](mailto:kelie.luby@perceptive.com)