Pharm-Olam™
Helping Create A Healthier World

Response Evaluation in Oncology Trials

The Evolution Continues

An ebook from
Pharm-Olam International
drawn from its webinar series on
Key Clinical Research Topics.

A Growing Reliance on

Tumor-Centered Endpoints

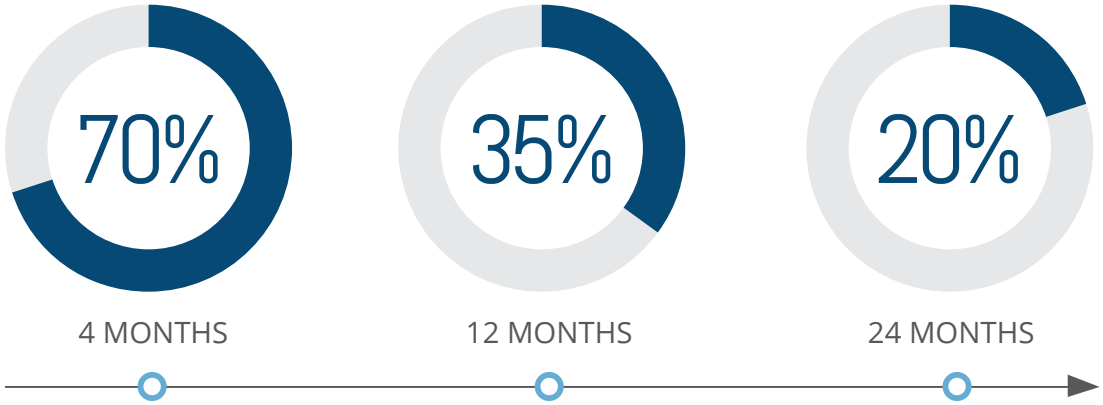
Overall Survival (OS) is the historical gold standard measurement in cancer clinical trials. However, the longer people with cancer live, the harder it is to observe survival differences over the course of a typical clinical trial.

Over the past 25 years, the use of Overall Survival as the primary endpoint in oncology clinical trials has steadily declined to be replaced by event free survival but response rate continues to be evaluated as primary end point specially in early phases of clinical trials.* Although OS remains the gold standard, it has been more difficult to measure due to the confounding influence of other therapies and the fact that patients are living longer, which extends the trial timeline.

Thus, the evidence goals of many oncology trials today hinge on having a definitive measure of treatment response that is a predictor of overall survival. The response criteria covered here are predominantly used in Phase II trials, but if they can be shown to correlate with overall survival, they can be used in Phase III trials. (The fact is, sometimes they correlate, and sometimes they don't!)

In this e-book, we review the history of response evaluation in cancer trials, detailing how measures have evolved along with technology; explain the principles of response evaluation; and share our view of its future. Our discussion is limited to solid tumors and so does not include response criteria for leukemia.

PROBABILITY OF OS STATISTICAL SIGNIFICANCE (%)



Median Post-Progression Survival

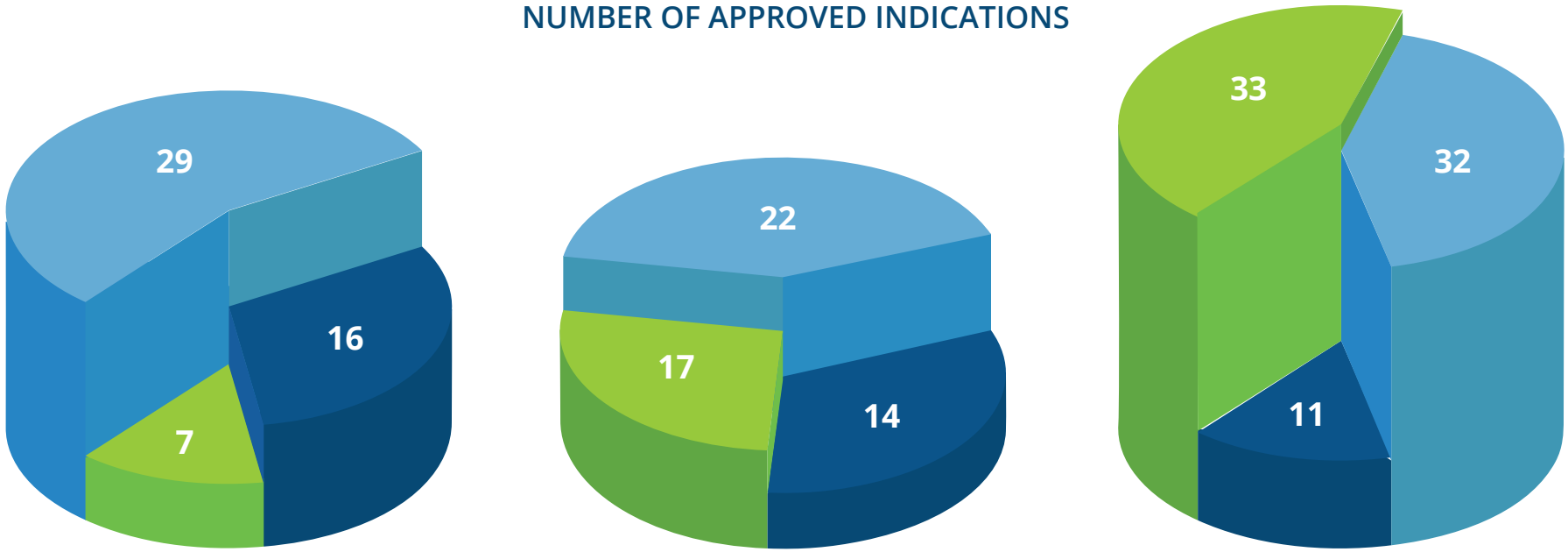
SOURCE: Adapted from Broglio KR and Berry DA; J Natl Cancer Inst. 2009:101:1642-1649.

* "*Oncology Endpoints in a Changing Landscape*," Genentech.

Change in the Use of Primary Endpoints for FDA-Approved Cancer Indications Over Time

Over the time span represented here, the percentage of trials with a time-to-event primary endpoint increased while the percentage of trials with survival as their endpoint decreased. The absolute number of trials with survival as their primary endpoint diminished slightly, however.

NUMBER OF APPROVED INDICATIONS



1990 - 1999
53 Indications*

2000 - 2005
53 Indications

2006 - 2011
76 Indications

Time To Event Response Rate Survival

* One trial in this period used symptoms palliation as a primary endpoint.
Adapted from Martell 2013

The Principles of
Response Evaluation

To be deemed valid, a response evaluation must be:

- A quantifiable assessment of measurable lesions;
- A qualitative assessment of non-measurable lesions (such as effusions); and
- Performed before treatment and during treatment at regular intervals using the same method of investigation.

Most response criteria can be classified into four categories:

- Complete response (CR)
- Partial response (PR)
- Stable disease (SD)
- Progressive disease (PD)

Before a complete or partial response can be definitively declared, the response must be confirmed radiologically.

The Earliest Measures of
Tumor Response

Decades ago, tumor size was evaluated by physical palpitation and the use of rulers and callipers. In 1976, Moertel and Hanley published the results of multiple trials in which they attempted to standardize response assessments using this method.

In the studies, experienced oncologists were instructed to measure spheres (laid out on a mattress and covered in foam rubber) that simulated tumor masses. The goal of the research was to determine the smallest size difference that could be measured reliably, given human error.

Moertel and Hanley concluded that:

- True tumor response should be defined as a >50% reduction

Response must be measured in two dimensions and must register as >25% reduction in the product of diameters.

Read the Full Report »

WHO Refinements

A few years later, in 1979, **The World Health Organization (WHO)** formed a committee to further standardize response criteria for solid tumors, based on measures taken by palpitation and X-rays. (Although CT technology existed, it was not yet widely available.) The group recommended that a determination of response be based on the sum of bi-dimensional measures (the greatest perpendicular dimensions) and offered the following definitions:

- CR = Complete disappearance of tumor for at least four weeks
- PR = >50% reduction from baseline, confirmed at four weeks
- No Change (There was no terminology describing stable disease)
- PD = >25% increase in tumor size, or the appearance of a new lesion

This work was a step forward, but still had some limitations, including:

- Because the measure was bi-dimensional, progressive disease would be declared based on an 11% increase in each dimension
- There was no explicit instructions on the number of tumor foci to be measured
- There was no clarity on the smallest size lesion that could be measured

Note that Moertel and Hanley's 50% reduction threshold carried over into the WHO criteria.

Read WHO's Original Handbook »

World Health Organization

A Major Step Forward with RECIST

Twenty years later, in 2000, a group of oncologists working under an initiative of the National Cancer Institute (NCI), aimed to introduce consistent evaluation criteria that could be used internationally with the use of CT scans, which had become widely available. Their work resulted in the **Response Evaluation Criteria in Solid Tumors (RECIST)**, which specified:

- Up to 10 target lesions could be assessed;
- Transaxial imaging with CT was mandatory;
- The measure should be based only on the single longest dimension;
- The lesion must be at least 1 centimeter; and
- Target and non-target lesions should be regarded differently

COMPARISON

RECIST

- CR = Complete disappearance of tumor for at least four weeks
- PR = >30% decrease from baseline for at least four weeks
- SD = Neither PR nor PD
- PD = >20% increase from the nadir of the new lesion

WHO

- CR = Complete disappearance of tumor for at least four weeks
- PR = >50% reduction from baseline, confirmed at four weeks
- No Change (There was no terminology describing stable disease)
- PD = >25% increase in tumor size, or the appearance of a new lesion

Read more about the introduction of RECIST »

A Scientific Debate Rages

Over the few years, as new assessment technologies and drugs with new modes of action were introduced, the applicability of RECIST came under question.

The most burning questions were:

- How valid is anatomical unidimensional assessment? Is it worth moving to volumetric or functional assessments?
- How applicable is RECIST to non-cytoxic drug studies?
- How should RECIST be applied in Phase III trials that have progression, not response, as the primary endpoint?
- Is assessing fewer than 10 lesions adequate?
- How should FDG-PET scans and MRI technology be used?
- How well does tumor response correlate with OS?
- Is confirmation truly needed?
- How should FDG-PET scans and MRI technology be used?

A New & Improved RECIST

A working group of clinicians from academia, government, and industry, along with imaging specialists and statisticians published **RECIST 1.1** in 2009. The new recommendations included:

- Evaluation should be based on the longest dimension for tumors and the short axis for lymph nodes
- The minimum measurable lesion sizes for

CT and MRI are ≥10mm on the long axis and 2X the slice thickness; for chest X-ray it is ≥20mm on the long axis; and for lymph nodes it is ≥10mm on the short axis.

- Up to five target lesions in two organs are allowed
- FDG-PET scans may be used to determine PD and to confirm CR
- Confirmation of PR and CR is only necessary in non-randomized trials having Objective Response Rate (ORR) as the endpoint

Read about the development of RECIST 1.1 »

RECIST 1.1 also defined non-target lesions as: bone lesions, leptomeningeal disease, pleural/ pericardial effusion and ascites, inflammatory breast disease, lymphangitis and cystic lesions.

RECIST 1 vs. RECIST 1.1

	RECIST 1	RECIST 1.1
Tumour Burden	10 Targets (5 per organ)	5 Targets (2 per organ)
Lymph Nodes	Like any other lesion	Short axis, defined Nr size
PD Definition	20% Increase in SLD	20% Increase in SLD 5 mm absolute increase
Non-Measurable PD	Unequivocal	More details
Confirmation	Required for CR & PR	Required in non-randomized trials with RR as 1ry endpoint
New Lesions		Section on FDG-PET

RECIST 1.1 Further Clarified That:

- Patients with measurable disease should be included in protocols with Objective Response Rate (ORR)
- The selected lesions should be the largest that are reproducibly measurable
- Multiple non-target lesions of the same organ should be recorded as a single item
- Lesions with prior local treatment should not be used unless or until PD is documented
- Osteolytic lesions with a soft tissue component can be included
- Lesions that split or coalesce on treatment
- For PD, there must be a ≥ 5 mm absolute increase in the single largest diameter (SLD) and there must be an overall change in the non-target tumor burden

MISTAKES IN APPLYING **RECISIT 1.1**

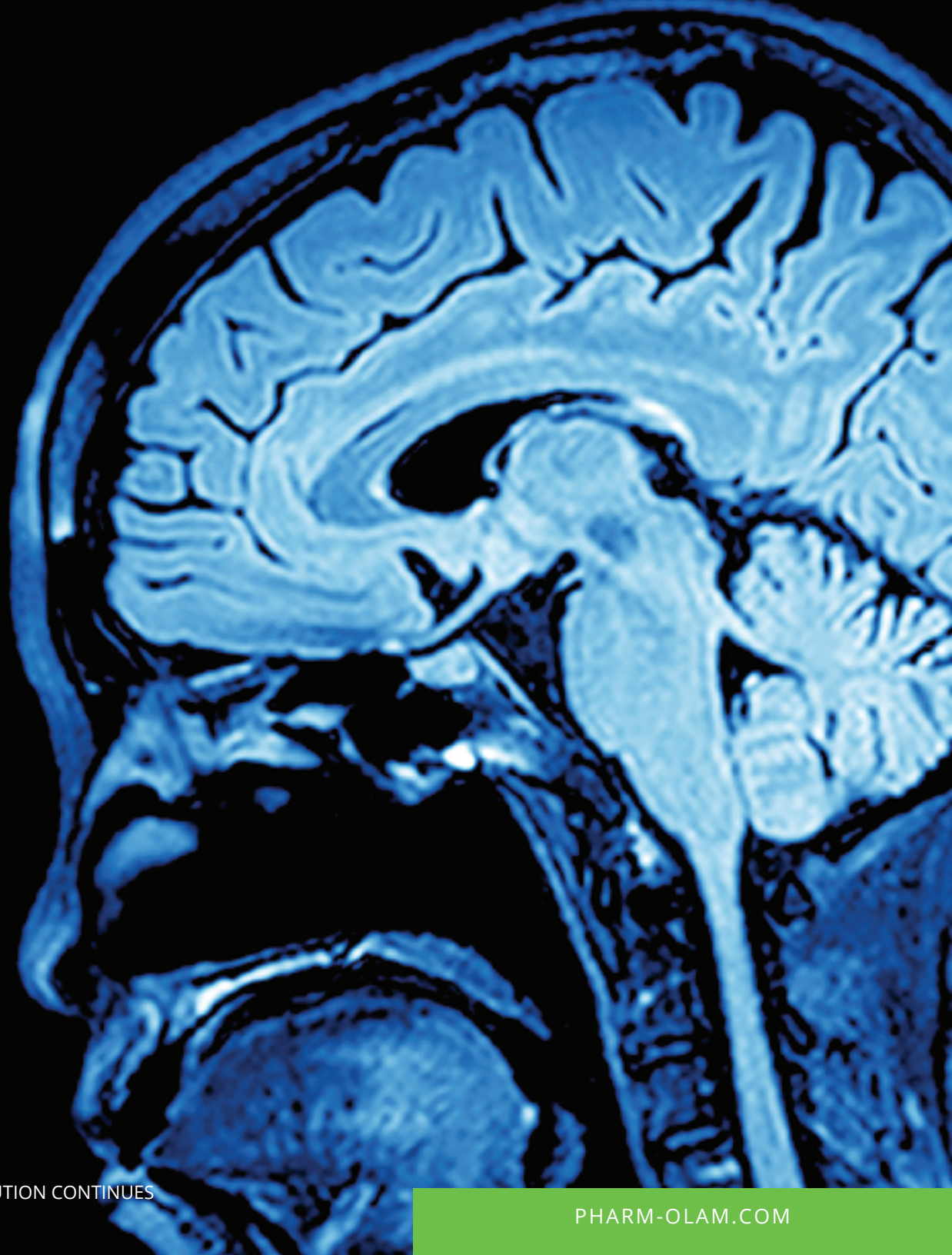
Despite the clarity of the RECIST 1.1 guidance, sponsors commonly make a number of mistakes in applying the criteria during clinical trials. These errors include:

- Assessing non-qualifying lesions for their number and size
- Measuring irregular, non-reproducible target lesions
- Including bone metastases
- Evaluating changes in cystic/necrotic tumors
- Considering reappearance of lesions as progressive disease
- Changing the baseline or nadir as a reference point in PD
- Following up inconsistently

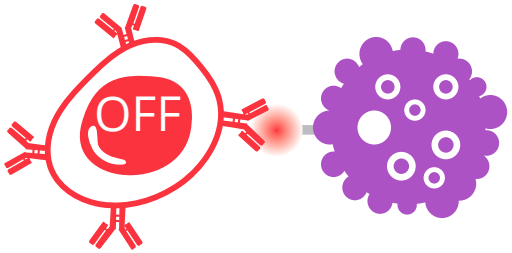
Limitations of Anatomical Response Criteria

Anatomical Response Criteria, as measured through CT scans or MRI may not be appropriate for:

- Cytostatic drugs used on longstanding, stable disease
- Predicting OS benefit for the limited anti-tumor size seen in hepatomas treated with sorafenib



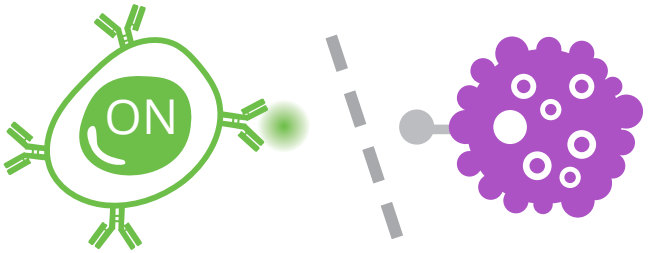
Tumor cells bind to T-cells to deactivate them



T-Cell Tumor

How Immunotherapy Works?

Immunotherapy drugs can block tumor cells from deactivating T-cells



T-Cell Drug Tumor

Enter Immuno-Oncology Drugs

During the trials for ipilimumab, the first monoclonal antibody, investigators observed something very unique: some patients who had initially achieved a stable disease state then had a decrease in tumor mass. For others, the disease progressed (either through an increase in the size of lesions or the appearance of new lesions) before showing response or stable disease.

We now know that with immune-oncology drugs, it may take longer to actually see the response, and RECIST may not do justice to development of these types of drugs. It is recommended that therapy not be discontinued until PD can be confirmed, that there be an allowance for "clinically insignificant" PD, and that researchers recognize that durable, stable disease may, in fact, represent anti-tumor activity.

Immune-Related Response Criteria (irRC)

With the development of many other immuno-oncology therapies came the need for specific, immune-related response criteria. These now specify:

- Assessment should be treated as a continuous variable and even new lesions that appear after baseline do not signal PD
- Lesions are measured bi-dimensionally
- The total tumor burden is calculated by adding the sum of perpendicular diameters at baseline to new lesions
- Response categories are the same as in WHO handbook
- The presence of a new lesion alone does not quality as PD if it does not add more than 25% to the tumor burden
- If there is a new lesion, but the overall tumor burden decreases, it would quality as PR or SD

The Evolution at a Glance

	1979	2000	2009	2009	2009
Criteria	WHO	RECIST 1	RECIST 1.1	PERCIST	irRC
Technology	X-Ray/CT	CT MRI	CT MRI	PET	PET CT MRI
Drugs	Cytotoxic	Cytotoxic	Targeted Therapy	Targeted Therapy	Immuno

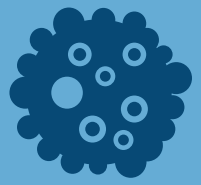
EASL, mRECIST, RECICL

Different organizations proposed different criteria to measure heptatocellular carcinoma when it became clear that RECIST was not satisfactory for a trial on sorafenib. The European Association for the Study of the Liver (EASL) criteria, a modified RECIST, and the Response Evaluation Criteria in Cancer of the Liver RECICL) (from a Japanese group) were all put forward.

Tumors are measured in two dimensions, and the dense accumulation of lipiodol is regarded as necrosis. For the first time, tumor markers such as alpha-fetoprotein, AFP-L3, and des-gamma-carboxy protein (DCP) were added.

GCIG

For ovarian cancer, the Gynecologic Cancer Intergroup (GCIG) developed criteria in which PD is based on either RECIST or the Cancer Antigen (CA) 125 blood test. The doubling of CA 125 from the upper limit of normal reliably predicts objective progression.



Cancer-Specific Criteria

While the general criteria were evolving, some groups were simultaneously developing specific criteria for specific cancer types. Criteria definitions are affected not only by the type of therapy, but also by the duration of therapy.

Choi

RESIST significantly underestimated the initial tumor response to imatinib in GIST, as significant changes in tumor density, enhancing antitumor nodules and tumor vessels were noted. Choi criteria use a combination of the values of tumor size and tumor density on CT scans.

CR is the disappearance of all lesions and no new lesions. PR is a decrease in size of 10% or a decrease in tumor density of 15% on a CT scan. PD is an increase in tumor size of 10%, the appearance of new lesions or new intra-tumoral nodules, or the increase in the size of the existing intra-tumoral nodules.

Choi is also used in response evaluation of metastatic renal cell carcinoma; high-grade, soft tissue sarcoma; solitary fibrous tumors, and hepatocellular carcinoma

MDA

In 2004, Hamaoka proposed using X-Rays, CT scans, MRIs, skeletal scintigraphy and other supportive technologies to quantify bone metastasis. The MD Anderson (MDA) criteria specify that follow up imaging is recommended every two to six months. Complete response is achieved when there has been complete fill-in or sclerosis of lytic lesions or normalization of osteoblastic lesions. PR is achieved when:

1. The sclerotic rim is seen along the lytic lesions
2. There is sclerosis of a previously undetected lesion
3. There is partial fill-in or sclerosis of lytic lesions
4. There is regression of measurable lesions or
5. There is a decrease in blastic lesions.

Every lesion need not regress, but no lesion should have progressed.

RANO

Within temozolomide radiotherapy trials, in 20% of cases, observed progression according to the standard scale was, in fact, pseudo progression.

In 2011, Response Assessment in Neuro-Oncology (RANO) criteria were published, permitting treatment to continue with PD, pending follow-up imaging.

PCWG2 Criteria

The Prostate Cancer Working Group (PCWG) has proposed separate criteria for prostate cancer because of the disease's unique profile. Bone scans can show a flare-up of the disease when the soft tissue mass is decreasing. The group wanted to incorporate measures of Prostate Specific Antigen (PSA), as this has long been part of clinical practice. The criteria state:

- Progression is an increase of >25% or increase of >2 ng/ml from the nadir
- A decline in PSA should be confirm three or more weeks later
- PSA progression alone is not an indication to stop treatment
- Progression based on bone scans = two new lesions on the first on-treatment scan, followed by two additional lesions on the next scan OR two new lesions on any scan after the first on-treatment scan (must be confirmed on a subsequent scan)

One area of continued discussion is whether it is best to rely on continuous, as opposed to discrete, sets of responses. Some suggest that it should be possible to take all the data points from continuous evaluation and, with the help of statistical tools, better predict overall survival.

Another area of current exploration is the suitability of response criteria in an era of individualized medicine. How should the criteria evolve to correctly predict overall survival when the pathology of a cancer is different from person to person? To begin with, the tumor response criteria must be specific to the treatment and type of tumor.

THE FUTURE OF **Response Evaluation**

There is growing interest in using FDG-PET scans (which are very sensitive) to validate functional biomarkers and show their correlation to overall survival.

Clearly, response evaluation has changed dramatically from 1976 to the present, and no doubt it will continue to evolve. Today, there are multiple types of responses that are being used both in clinical trials and clinical practice. Unfortunately, the science has not yet reached a point where evaluating tumor response always translates into a direct benefit to the patient. But research efforts will evolve towards translating tumor response to benefit patients.

See our Webinar: Response Evaluation in Oncology Trials »

For more information about Pharm-Olam's experience and its approach to evaluating Oncology clinical trials, please contact us at info@pharm-olam.com or visit www.pharm-olam.com.