PhUSE EU Connect 2018

Paper AS04

Step by Step Guide to Efficacy Analysis in Solid Tumors Oncology Clinical Trials

Anastasiia Tiurdo, Intego Group, Kharkiv, Ukraine

ABSTRACT

Cancer is a scourge of the modern society, so more and more pharmaceutical companies work on discovering a treatment that will fight against the different types of this disease. Nowadays the majority of clinical trials evaluating cancer treatments for the objective response in solid tumors use a set of rules called RECIST. This abbreviation is commonly used in a clinical programmer's daily routine, but not everyone has a deep understanding of the RECIST requirements.

This paper provides a simple step by step description of the RECIST compliant efficacy analysis in solid tumors oncology studies starting from the data collection by an investigator up to the final statistical results based on this data.

INTRODUCTION

Assessment of the change in the tumor burden is an important part of the clinical evaluation of cancer therapeutics. Both the tumor shrinkage, known as Objective Response, and time to disease progression, known as Progression-Free Survival, are significant endpoints in cancer clinical trials. However, both of them can be useful only if based on the widely accepted and readily applied standard criteria.

The topic of this paper is the RECIST1.1 (Response Evaluation Criteria in Solid Tumors, version 1.1) – the criteria that are nowadays commonly used in the solid tumors oncology clinical trials where the objective response, assessment of the stable disease or time to progression analyses are included in the study endpoints list, since all of these outcomes are based on the assessment of the anatomical tumor burden and its change in the study follow-up – the main subject of the RECIST Guideline.

The paper should be helpful in understanding the main concepts, principles and algorithms of the RECIST1.1 criteria, comprising the features of the baseline tumors assessments and categorization, requirements of the disease follow-up process and reporting of the final results.

RECIST BACKGROUND/HISTORY

The first serious attempt to provide a criterion to evaluate the patient's response to the therapy used to treat the cancer disease was published in 1981 by the World Health Organization (WHO). As it often happens with pioneer work, WHO criteria were not ideal, as they raised many questions without answers and controversial points that could lead to very different conclusions about the efficacy of the same regimen, so it was clear that the criteria were not optimal.

In response to this problem, the European Organization Research and Treatment for Cancer (EORTC) in collaboration with the National Cancer Institute of the United States, and the National Cancer Institute of Canada Clinical Trials Group published the initial version of The Response Evaluation Criteria in Solid Tumors (RECIST) in February 2000. The key features of that document included definitions of minimum size of measurable lesions, instructions on how many lesions to follow up and the use of unidimensional, rather than bi-dimensional, measures for the overall evaluation of the tumor burden. Moreover, the important thing was that the regulatory authorities accepted RECIST as an appropriate guideline for the clinical trials where the primary endpoints are an objective response or progression.

Since then, many investigators, cooperative groups, industry and government authorities were using the RECIST criteria in the assessment of treatment outcomes, so in some years, based on the received experience, lots of new issues and comments, related to the initial version of the RECIST, were collected. Finally, this resulted in the development of a revised RECIST guideline (version 1.1) in 2009 - the main topic of this paper.

TUMOR MEASURABILITY AT BASELINE

All existing tumor lesions/lymph nodes should be examined at baseline and categorized as measurable or non-measurable. The study protocol may contain the list of organ sites for obligatory evaluation at baseline, which is most likely defined by the type of cancer investigated in a particular clinical trial.

If at least one measurable lesion is present, it is said that the patient has a Measurable Disease. In case the objective tumor response is a study primary endpoint, then only patients with a measurable disease can be involved in the efficacy analysis. However, if primary endpoints include the tumor progression only, the study protocol must specify whether the patients with a non-measurable disease only are also eligible.

Since the restriction to the measurable disease may slow down the recruitment to the study, trials often allow to entry both types of patients. The only thing that should be made clear in such studies is Progressive Disease definition, namely, the protocol must contain the detailed description of findings that would be classified as the disease progression for patients with a non-measurable disease only.

To ensure the most honest tumor response, the following RECIST1.1 requirements must be taken into account at baseline:

- All evaluations should be done as close to the study treatment start as possible and never more than 4 weeks before the beginning of the treatment.
- As many lesions as possible present at baseline should be documented per patient. Otherwise, a lesion, omitted at baseline, may be considered as a new one at further visits that may lead to incorrect final efficacy conclusion.
- Each lesion must be accurately measured in at least one direction and all measurements should be recorded in metric notation.

METHODS OF TUMOR ASSESSMENT

An important point is that the same technique of assessment should be used to characterize each particular lesion at baseline and during the follow-up in order to provide the most accurate and meaningful feedback on the tumor burden. The following methods are possible to be used:

- CT (computed tomography) and MRI (magnetic resonance imaging) are the best currently available and reproducible methods to measure lesions selected for the response assessment, so they should be used if possible.
- Clinical examination results should be used only if the lesion cannot be imaged. Otherwise, imaging evaluation has a priority, because it is more objective and may be reviewed at the end of the study.
- Lesions on a chest X-ray are acceptable as measurable lesions when they are clearly defined and surrounded by the aerated lung. However, CT is preferable.
- Ultrasound is not useful in the assessment of the lesion size and should not be used as a method of measurement. If new lesions are identified by ultrasound in the course of the study, confirmation by CT or MRI is advised.
- The utilization of endoscopy and laparoscopy for an objective tumor response should be avoided. However, such techniques can be useful in confirming a complete pathological response when biopsies are obtained.
- Tumor markers alone cannot be used to assess a response and are considered only as a supportive method, namely, if the markers are initially above the upper normal limit, they must normalize for a patient to be used for complete clinical response confirmation when all the lesions have disappeared.
- Cytology and histology can be used to differentiate between Partial and Complete Response in rare cases if that is defined in the study protocol.

MEASURABILITY OF TUMOR LESIONS (EXCLUDING MALIGNANT LYMPH NODES)

Longest Diameter should be used as a measurement characteristic of all the solid lesions types in the tumor burden analysis, except for lymph nodes measurements. A measurable tumor lesion is defined as a lesion with the minimum size of:

- 10 mm by CT scan (CT scan slice thickness no greater than 5 mm. When CT scans have the slice thickness greater than 5 mm, the minimum size for a measurable lesion should be twice the slice thickness).
- 10 mm caliper measurement by clinical examination (lesions, which cannot be accurately measured with calipers, should be recorded as non-measurable).
- 20 mm by chest X-ray.

All the other lesions are considered non-measurable. These are small lesions with longest diameter < 10 mm as well as truly non-measurable lesions. Truly non-measurable lesions include leptomeningeal disease, ascites, pleural or pericardial effusion, inflammatory breast disease, lymphangitic involvement of skin or lung, abdominal masses/abdominal organomegaly identified by a physical exam that is not measurable by reproducible imaging techniques.

MEASURABILITY OF MALIGNANT LYMPH NODES

Unlike other tumor types, a short axis is used as a measurement characteristic of lymph nodes in the tumor burden analysis (the nodal size is normally reported as two dimensions in the plane, in which the image is obtained, and the smaller of these measures is the short axis). Such decision has been made, since the lymph nodes are normal anatomical structures, which may be visible by imaging even if not involved by tumor.

RECIST 1.1 provides the next idea of lymph nodes classification:

- Malignant lymph nodes must be >= 15 mm in a short axis when assessed by CT scan (CT scan slice thickness being no greater than 5 mm) to be considered pathologically enlarged and measurable.
- All the other lesions with >= 10 to <15 mm short axis are considered non-measurable.
- Nodes that have a short axis < 10 mm are considered non-pathological and should not be recorded or followed.

SPECIAL CONSIDERATIONS REGARDING LESION MEASURABILITY

One of the important changes added to the RECIST1.1 guideline, comparing to the initial version, is the notes about measurability of some specific tumors that helped to answer commonly asked questions and made the process of tumor measurement more standard across all the oncology clinical trials using the RECIST criteria. The notes are as follows:

- Bone scan, PET scan or plain films techniques can be used to confirm the presence or disappearance of bone lesions, but are not allowed to be used to measure bone lesions.
- Lytic bone lesions or mixed lytic-blastic lesions, with identifiable soft tissue components, that can be evaluated by cross sectional imaging techniques such as CT or MRI can be considered as measurable lesions if the soft tissue component meets the definition of measurability described above.
- Blastic bone lesions are non-measurable.
- Cystic lesions that meet the criteria for radiographically defined simple cysts should not be considered as malignant lesions (neither measurable nor non-measurable).
- ‘Cystic lesions’ which are thought to represent cystic metastases can be considered as measurable lesions, if they meet the definition of measurability described above. However, if non-cystic lesions are present in the same patient, these are preferred for selection as the target lesions.
- Tumor lesions situated in a previously irradiated area, or in the area subjected to another therapy, are normally not considered measurable to exclude the effect of the previous treatment on the final objective response. But in case the lesion has demonstrated progression, it can be considered measurable if a corresponding condition is described in the study protocol.

TUMORS AT BASELINE: EVALUATION AND CATEGORIZATION

To assess tumors objective response or future progression in cancer clinical trials, it is necessary to estimate the overall tumour burden at baseline and use it as a comparator for subsequent measurements. The first step in this process is to divide all the tumors identified at baseline into two separate groups - target lesions and non-target lesions. This classification will be valid until the patient discontinues the study, meaning, a tumor defined as target at baseline can not become non-target during the follow-up and vise versa, even if the target lesion would reduce in size, so that it is not measurable any more, or if non-target one would icrease, so that it becomes measurable.

TARGET LESIONS

If more than one measurable lesion is available at baseline, all the measurable lesions up to a maximum of two lesions per organ and five lesions in total, that are, if possible, representatives of all cancer-affected organs, should be defined as target lesions, recorded and measured at baseline. The maximum number of target lesions is not accidental. In the first version of the RECIST Guideline ten lesions up to five per organ were allowed to be classified as target, but further data warehouse analysis showed no loss of information if the lesion number reduced from ten to five, so it does not make any sense to spend time and resources on evaluation of more than five target lesions. Furthermore, the protocol may reduce the maximum number of target lesions (from five to three, for example) if patients without a measurable disease are allowed to be involved in the efficacy analysis in order to make results of the trial more generalizable.

The main idea of the target tumors selection is that the selection of target lesions should be based on their size, i.e. the ones with the longest diameter are preferable, as well as their suitability for accurately repeated measurements either by imaging techniques or clinically. It may be the case that the largest lesions are not suitable for the follow-up

because of their unstable or non-legible configuration. In such cases, identification of the largest most reproducible lesions is advised. In addition, if both cystic and noncystic lesions are present for the same patient, non-cystic lesions are preferred for the selection as target lesions independently of the size.

As soon as the list of target lesions is prepared, a sum of the diameters (the longest diameter for non-nodal lesions and a short axis for nodal lesions, as mentioned before) of all target lesions should be calculated and reported as the baseline sum diameters (BSD). Then BSD will be used as a reference by which to characterize any objective tumor response in terms of the measurable disease on the post-baseline visits.

NON-TARGET LESIONS

All the other lesions that have not been defined as target, including pathological lymph nodes, should be identified as non-target lesions and also be recorded at baseline. The number of non-target lesions is not limited and it is allowed to report multiple non-target lesions, located in the same organ, as a single item on the CRF.
While some non-target lesions may actually be measurable, they are not required to be measured at baseline as well as during the follow-up, but should be assessed qualitatively: the fact of presence, absence or unequivocal progression of each non-target lesion should be collected.

TUMORS EVALUATION AT POST-BASELINE VISITS

The study protocol must specify how often post-baseline tumor evaluations should be repeated. Ideally, each target and non-target lesion should be examined at every scheduled visit and the schedule should not be affected by a study drug delay, omission, interruption or any other events that may lead to imbalance in a treatment arm in the timing of the disease assessment. In some cases, the protocol may allow assessing certain non-target organs less frequently, for example, only if either the complete response in the target disease or the progressive disease in this particular non-target organ is suspected.

FREQUENCY OF THE FOLLOW-UP RESPONSE ASSESSMENTS

Frequency of tumor re-evaluation is usually defined depending on the type and schedule of the study treatment. For instance, it is a common practice for phase II studies where the benefit effect of study therapy is not yet fully investigated to follow-up tumors every 6 – 8 weeks, although smaller or greater time intervals can be used to keep consistency with the length of the treatment cycles.

The necessity and frequency of tumor evaluation after the end of the study therapy depends on the primary efficacy endpoint. If the trial's main goal is a time to an event, then it is reasonable to continue the routine scheduled re-evaluation of the protocol specified disease sites after the treatment end. Otherwise, if the response rate is the main endpoint of the study, advisability of the after treatment tumor assessment should be described in the protocol, since post-treatment responses may be not recommended to use in the response rate calculation.

It should be noted that the frequency of the follow-up tumor evaluations influences such efficacy endpoints as duration of response, duration of stable disease and progression free survival, therefore, it should be carefully defined. This effect should be taken into account while comparing the results of several clinical trials.

TARGET LESIONS FOLLOW-UP

Similar to what has been done at baseline, the diameter of each target lesion, namely, the longest diameter for non-nodal lesions and a short axis for nodal lesion, should be measured and then added to the Sum of Diameters in order to assess a target lesions response at the given post-baseline visit. The response criteria according to the RECIST 1.1 are as follows:

Table 1: Target Lesions Response Criteria (RECIST 1.1)

Response	Criteria to be met
Complete Response (CR)	a) All target lesions disappeared; b) All lymph nodes must be reported as non-pathological (short axis <10 mm)*.
Partial Response (PR)	a) At least 30% decrease in the Sum of Diameters compared to the BSD.
Progressive disease (PD)	a) At least 20% increase in the Sum of Diameters, taking as reference the smallest Sum of Diameters (usually called NADIR as the lowest point) reported at the prior study visits. b) The Sum of Diameters must demonstrate at least 5 mm increase compared to NADIR.
Stable Disease (SD)	a) There has been no significant decrease for PR or increase for PD in the size of the target lesions.

Valuable notes:

1) Note that the Baseline Sum of Diameters (BSD) should be taken as a reference to evaluate Partial Response of Target Lesions, while a criterion for Progressive Disease is based on NADIR - the smallest Sum of Diameters reported at the prior study visits. BSD may be used as NADIR, if the baseline result is the smallest in the study.
2) * Since a normal lymph node is defined as having a short axis of <10 mm, the Sum of the Diameters value does not have to be zero to meet the Complete Response criteria, if at least one lymph node is present as a target lesion. Therefore, lymph nodes identified as target lesions should always have the actual short axis measurement recorded, even if the short axis is small enough to consider the lymph node as non-pathological. In order to qualify for Complete Response, CRF may be designed to have target nodal lesions recorded in a separate section, to be able to confirm that each node achieved a short axis < 10 mm.
3) During the study all target lesions should be measured at each subsequent evaluation, even when being very small. However, sometimes lesions or lymph nodes become so faint on CT scan that radiologists may not be able to provide an exact measure. In such circumstances assessors are allowed to report these lesions as being 'too small to measure', but in addition they must provide one of two default values to use for the Sum of Diameters calculation: 0 mm, if the radiologist thinks that the lesion is likely to have disappeared, or 5 mm, if the lesion is believed to be present and is faintly seen but too small to measure. But if the radiologist is able to assign an actual measure, that should be done anyway, even if the diameter is below 5 mm.
4) It may occur that non-nodal target lesion has split up during the treatment. In this instance, the longest diameters of the fragmented portions should be added to the Sum of Diameters. In the opposite case, when the lesions coalesce, a plane between them may be maintained which would aid in obtaining maximal diameter measurements of each ndividual lesion. If the lesions have coalesced so that they are no longer separable, the longest diameter should be defined as the longest diameter for this 'coalesced lesion'.
5) The example of Target Lesions Response definition is shown below:

Example 1: Target Response Derivation

	Subject ID	Visit	Study Day	Tumor ID	Location	Parameter Name	Analysis Value	Analysis Value Unit
1	001	Screening	-1	Target Lesion 1	Lung	Longest Diameter	31	mm
2	001	Screening	-1	Target Lesion 2	Liver	Longest Diameter	59	mm
3	001	Screening	-1	Target Lesion 3	Lymph Node	Short Axis	20	mm
4	001	Screening	-1	All Target Lesions		Sum of Diameters	110	mm
5	001	Week 8	56	Target Lesion 1	Lung	Longest Diameter	28	mm
6	001	Week 8	56	Target Lesion 2	Liver	Longest Diameter	50	mm
7	001	Week 8	56	Target Lesion 3	Lymph Node	Short Axis	21	mm
8	001	Week 8	56	All Target Lesions		Sum of Diameters	99	mm
9	001	Week 16	112	Target Lesion 1	Lung	Longest Diameter	19	mm
10	001	Week 16	112	Target Lesion 2	Liver	Longest Diameter	33	mm
11	001	Week 16	112	Target Lesion 3	Lymph Node	Short Axis	16	mm
12	001	Week 16	112	All Target Lesions		Sum of Diameters	68	mm
13	001	Week 24	168	Target Lesion 1	Lung	Longest Diameter	19	mm
14	001	Week 24	168	Target Lesion 2	Liver	Longest Diameter	35	mm
15	001	Week 24	168	Target Lesion 3	Lymph Node	Short Axis	19	mm
16	001	Week 24	168	All Target Lesions		Sum of Diameters	73	mm
17	001	Week 32	224	Target Lesion 1	Lung	Longest Diameter	27	mm
18	001	Week 32	224	Target Lesion 2	Liver	Longest Diameter	43	mm
19	001	Week 32	224	Target Lesion 3	Lymph Node	Short Axis	19	mm
20	001	Week 32	224	All Target Lesions		Sum of Diameters	89	mm

Target Disease Response for each follow-up visit

BSD =110 mm

NADIR = 110 mm
→ 11 % Decrease from BSD
→ Target Response is **SD**

NADIR = 99 mm
→ 38 % Decrease from BSD
→ Target Response is **PR**

NADIR = 68 mm
→ 34 % Decrease from BSD
→ Target Response is **PR**

NADIR = 68 mm
→ 31 % Increase from NADIR
→ Target Response is **PD**

NON-TARGET LESIONS FOLLOW-UP

As mentioned previously, non-target lesions are not measured, but assessed qualitatively, that makes the criteria of their response not as transparent as the one for the target lesions. The RECIST 1.1 suggests the following algorithm to be used:

Table 2: Non-target Lesions Response Criteria (RECIST 1.1)

Response	Criteria to be met
Complete Response (CR)	a) Disappearance of all non-nodal non-target lesions; b) All lymph nodes must be < 10 mm in a short axis to be considered normal sized; c) Normalization of the tumor-marker level.
Non-CR/Non-PD	At least one condition is true: a) Persistence of one or more non-target lesions; b) Maintenance of the tumor marker level above the normal limits.
Progressive Disease (PD)	a) Unequivocal progression* of non-target lesions

Valuable notes:

* Unequivocal Progression of a non-target disease is a specific concept that needs additional comments. The problem is that the very nature of the non-target disease makes it impossible to have the objective criteria for the disease progression, since by definition all lesions (or at least most of them) are truly non-measurable. That is why to consider Unequivocal Progression of non-target lesions, the increase must be really substantial.

There are two possible cases:

I. When a Measurable disease is present:
It should be noted that just a minor size increase of one or more non-target lesions is usually not sufficient to report unequivocal progression status. To report non-target disease unequivocal progression, there must be an overall substantial worsening in the non-target disease such as, even providing SD or PR response in the target disease, the overall tumor burden has increased sufficiently. Since the measurable disease usually gives the most significant contribution to the tumor burden, it is clear that the overall progression based on non-target progression is extremely rare to occur.

II. When the patient has a Non-Measurable disease only:
Basically, the same concept as above should be applied, however, in this case there is no measurable disease available to estimate the contribution of non-measurable lesions to the tumor burden increase. Since worsening in the non-target disease cannot be easily quantified, a special test is required to be used to assess unequivocal progression. The idea is to compare the increase that would be required to report PD for the measurable disease with the increase in the overall disease burden based on the change in the non-measurable disease only. If the increase in the overall tumor burden is at least 73% in volume (that is equivalent to 20% increase in the sum of diameters for measurable lesions), then unequivocal progression should be documented.

The examples of Unequivocal Progression of non-target lesions are: an increase in a pleural effusion from ‘trace’ to ‘large’, an increase in the lymphangitic disease from localised to widespread, or may be described in protocols as ‘sufficient to require a change in therapy’.

NEW LESIONS IDENTIFICATION

The appearance of a new malignant lesion is always considered to be a disease progression, but the finding of a new lesion should be unequivocal. This means that it should not be related to differences in the scanning technique, imaging the modality change and should represent nothing other than the malignant tumor. A lesion identified in study follow up in a location that was not examined at baseline should be considered a new lesion and indicate the disease progression too.

In case a new lesion is equivocal (for example, its size is too small or necrosis in the existing lesions), an additional evaluation is recommended to confirm an appearance of the truly new disease. If repeat scans acknowledge there is definitely a new lesion, then the progression should be declared using the date of the initial scan.

TIME POINT RESPONSE

A response to the study treatment that describes the status of the patient’s oncology disease in general is called Overall Response and should be reported at each follow-up time point scheduled in the protocol. It is based on the three main factors described above: the target disease response (if available), the non-target disease response and the appearance of new lesion(s).

Tables 3 and 4 below provide a summary of overall response status derivation for two types of patients: patients who have a measurable disease at baseline and patients with a non-measurable disease only respectively.

Table 3: Overall Response Criteria (RECIST 1.1), Measurable Disease available

Target Response	Non-target Response	New Lesion Identified	Overall Response
Complete Response (CR)	Complete Response (CR)	No	Complete Response (CR)
Complete Response (CR)	Non-CR/non-PD	No	Partial Response (PR)
Complete Response (CR)	Not evaluated	No	Partial Response (PR)
Partial Response (PR)	Non-PD or not all evaluated	No	Partial Response (PR)
Stable Disease (SD)	Non-PD or not all evaluated	No	Stable Disease (SD)
Not All Evaluated	Non-PD	No	Unevaluable (NE)
Progressive Disease (PD)	Any	Yes or No	Progressive Disease (PD)
Any	Progressive Disease (PD)	Yes or No	Progressive Disease (PD)
Any	Any	Yes	Progressive Disease (PD)

Table 4: Overall Response Criteria (RECIST 1.1), Measurable Disease not available

Non-target Response	New Lesion Identified	Overall Response
Complete Response (CR)	No	Complete Response (CR)
Non-CR/non-PD	No	Non-CR/Non-PD
Not All Evaluated	No	Unevaluable (NE)
Unequivocal Progression (PD)	Yes or No	Progressive Disease (PD)
Any	Yes	Progressive Disease (PD)

In case the tumor measurement was not performed at a particular time point, despite the fact it was required according to the protocol, the Overall response should be set to NE, meaning that the patient was not evaluable at that time point. If only a subset of lesions was investigated, but not all, in most situations the Overall Response should be reported as NE as well, unless it is clear that the contribution of non-assessed lesions would not influence the response. This is theoretically possible in case of PD. For instance, at baseline the patient had three target lesions reported with the sum of diameters of 40 mm, and at one of the follow-up visits only two out of three target lesions were examined, but SLD for those two was 80 mm – that is already enough to document Progressive Disease for the Target Response (30% increase in SLD) regardless of the results for the third missing lesion.

It sometimes befalls that discontinuation of treatment might be required for patient because of the global deterioration of health status, while no objective evidence of the disease progression is present. In such case the reason for the study treatment withdrawal should be reported as ‘symptomatic deterioration’, but under no circumstances it should be considered as objective response of PD without evaluation of the target and non-target disease and the new lesions. Every effort should be made to document the objective progression even after the treatment discontinuation.

In some situations it may be difficult for the investigator to distinguish a residual disease from a normal tissue. And if the evaluation of the complete response depends on this determination, additional investigations are recommended to avoid a false-positive CR report.

As it has already been mentioned, for equivocal progression findings the study treatment may be continued up to the next scheduled assessment. However, if at the next scheduled assessment some progression is confirmed, the date of progression should be the earliest date when the progression was suspected.

BEST OVERALL RESPONSE

The best overall response (BOR) is the best response reported for the patient during the study treatment period. The responses documented after the end of the study treatment may be considered during the best overall response determination, but this should be clearly defined in the study protocol, as well as how any additional therapy treated before the progression will affect the best overall response definition. The best overall response is determined once all the data for the patient is known and, depending on the study requirement, BOR may need a confirmation.

In the simplest case, when the confirmation is not required, best overall response is defined as the best response across all the time points. The BOR of SD needs a separate comment, as when stable disease is assumed to be a BOR it must meet the criteria of minimum time from baseline defined in the protocol (generally 6 - 8 weeks). If this time minimum has not been achieved, the response should be defined based on the subsequent assessment, so that, for instance, if SD was reported during the fourth treatment week and PD – during the eighth week, while the protocol requires minimum time from baseline of 6 weeks, PD should be claimed as BOR. Based on the same logic, BOR for the

patient lost to follow-up after the SD response that did not satisfy the minimum time from baseline requirement should be considered unevaluable.

Another more complex algorithm should be used in case BOR needs confirmation. The main goal of the response confirmation is to avoid overestimation of the response rate observed and to demonstrate that the response identified initially is not the result of a measurement error.

At the first glance, it may seem that the confirmed response gives the results that are more objective. It is obvious that a response rate rises if no confirmation is necessary, and, in fact, in the initial version of the criterion (RECIST 1.0) confirmation of the response was obligatory for all the studies. However, for randomized studies or studies where time to progression is the target goal, the confirmation of response does not give any value to interpretation of the efficacy result. That is why RECIST 1.1 suggests the response confirmation requirement only for clinical trials where there is no concurrent comparative control or where an objective response rate is the primary efficacy endpoint. Although if the unconfirmed response is collected, especially for non-blinded studies, it makes sense to provide central review of the responses to be safe from the investigator's bias.

The main principle of the best overall response confirmation is that the conclusion being made should not only be based on the overall response from the given visit, but also considering a subsequent time point assessment result, i.e. the changes in the tumor measurements must be confirmed by the repeat schedule assessments to report a partial or complete response as BOR. Generally, subsequent assessments sufficient for BOR confirmation should be performed no less than 4 weeks in between (the sufficient time period is defined on the strength of the schedule of tumor assessments). In the case of a stable disease, the follow-up measurements must meet the stable disease criteria at least once after the study entry at a minimum interval (usually 6 – 8 weeks) similar to the unconfirmed BOR case.

The general scheme of the confirmed best overall response definition according to RECIST1.1 guideline is available in Table 5:

Table 5: Best Overall Response Confirmation Algorithm (RECIST 1.1)

Overall Response at the First Time Point	Overall Response at the Subsequent Time Point	Best Overall Response
Complete Response (CR)	Complete Response (CR)	Complete Response (CR)
Complete Response (CR)	Partial Response (PR)	Stable Disease (SD), Progressive Disease (PD) or Partial Response (PR)*
Complete Response (CR)	Stable Disease (SD)	Stable Disease (SD) given that minimum criteria for SD duration met, otherwise, Progressive Disease (PD)
Complete Response (CR)	Progressive Disease (PD)	Stable Disease (SD) given that minimum criteria for SD duration met, otherwise, Progressive Disease (PD)
Complete Response (CR)	Unevaluable (NE)	Stable Disease (SD) given that minimum criteria for SD duration met, otherwise, NE
Partial Response (PR)	Complete Response (CR)	Partial Response (PR)
Partial Response (PR)	Partial Response (PR)	Partial Response (PR)
Partial Response (PR)	Stable Disease (SD)	Stable Disease (SD)
Partial Response (PR)	Progressive Disease (PD)	Stable Disease (SD) given that minimum criteria for SD duration met, otherwise, Progressive Disease (PD)
Partial Response (PR)	Unevaluable (NE)	Stable Disease (SD) given that minimum criteria for SD duration met, otherwise, Unevaluable (NE)
Unevaluable (NE)	Unevaluable (NE)	Unevaluable (NE)

Valuable notes:

1) The case when Complete Response had been reported at the first time point, but was not confirmed on a subsequent visit, needs an additional explanation, since even Partial Response met at subsequent time point would mean Disease Progression comparing to the previous result. In such situation the best response would depend on whether the minimum duration for Stable Disease was met: if so – Stable Disease would be reported, if not – Progressive Disease would be claimed. Nevertheless, sometimes the subsequent scan may clarify that small lesions are still in place, although are non-measurable anymore, so this would mean that at the first time point the patient experienced a Partial, but not Complete Response. If so, the original Complete Response should be changed to the Partial Response and the best confirmed response set to the Partial Response.

2) A special case takes place when NE time point assessments occurred. They may complicate BOR determination, so the protocol must specify how the missing response assessment should be covered. For example, it makes sense to consider the following chain of responses PR – NE – PR as a confirmed Partial Response.

3) The examples of Confirmed/Unconfirmed Best Overall Response definition are available below. Let us assume that the study protocol requirements are as follows:
   - follow-up tumor evaluations should be repeated each 8 weeks;
   - a minimum criterion for SD duration is 7 weeks (49 days);
   - subsequent assessments sufficient for BOR confirmation should be performed no less than 4 weeks (28 days) in between.

Best Overall Responses: Example 1

	Subject ID	Visit	Study Day	Parameter Name	Analysis Value, Character
1	001	Week 8	56	Overall Response	SD
2	001	Week 16	112	Overall Response	PR
3	001	Week 24	168	Overall Response	PR
4	001	Week 32	224	Overall Response	PR
5	001	Week 40	280	Overall Response	CR

1. **Unconfirmed BOR is CR.**
2. **Confirmed BOR is PR**, because there are two consequent PR observation with >= 28 days in between and there is no record available to confirm CR.

Example 2

	Subject ID	Visit	Study Day	Parameter Name	Analysis Value, Character
1	002	Week 8	56	Overall Response	PR
2	002	Week 16	112	Overall Response	NE
3	002	Week 24	168	Overall Response	PD

1. **Unconfirmed BOR is PR**.
2. **Confirmed BOR is SD**, because first PR assessment has met min criteria for SD duration.

Example 3

	Subject ID	Visit	Study Day	Parameter Name	Analysis Value, Character
1	003	Unscheduled Visit	30	Overall Response	CR
2	003	Week 8	56	Overall Response	PR

1. **Unconfirmed BOR is CR.**
2. **Confirmed BOR is SD**, because only two responses are available with 26 days in between, but the second PR assessment has met min criteria for SD duration.

OBJECTIVE RESPONSE RATE

Objective Response Rate (ORR) is a proportion of patients with the Best Overall Response equal to Complete or Partial Response and a denominator in this proportion could be various.

If the objective response rate is a primary efficacy endpoint, so the disease measurability is a study inclusion criterion, then the total number of eligible patients should be used, even if among them there are patients with major protocol treatment violations or if BOR is not evaluable. Some protocols may restrict the denominator to the treated patients only.

Otherwise, if ORR is a secondary endpoint, so the patient is not required to have a measurable disease to entry a trial, the method of reporting ORR must be specified in the protocol: the denominator may be the total number of all the randomized/enrolled patients, as well as a subset of patients with a measurable disease only.

Ideally, the two-sided 95% confidence interval should be calculated for the objective response rate.

CLINICAL BENEFIT RATE

Clinical Benefit Rate (CBR) is a proportion of patients with the Best Overall Response equal to Complete Response, Partial Response or Stable Disease. Stable Disease must meet the criterion of the minimum time defined in the protocol.

All other aspects are similar to the Objective Response Rate.

TIME TO EVENT ANALYSIS

Time to Event Analysis (TTE) is a set of methods for analyzing data where the outcome variable is the time up to the occurrence of an event of interest (can be measured in days, weeks, months or years as defined in statistical analysis plan). The typical events of interest in clinical trials efficacy analysis are death, disease progression, occurrence of new particular disease, etc.

Time to event analysis has two components that must be clearly defined: the beginning point and the endpoint that is reached either when the event occurs or when the follow-up time has ended. In the second case the concept of censoring arises - a condition in which the value of a measurement is only partially known.

To be more precise, in the simplest TTE analysis two types of observations can be present:

- The event occurred, and we are able to measure when it occurred – Event.
- The event did not occur during the observed time, and we only know the total number of days in which it did not occur – Censoring.

In case of censoring the time to event is calculated as the time up to the last date known that no event of interest had taken place up to this date.

The commonly used time to event parameters, that are based on the responses received using RECIST criteria, are provided below.

DURATION OF RESPONSE

Duration of Response is defined as the time from the first Partial or Complete Response criteria met, whichever was reported earlier, until the first date of documented overall disease progression. If the duration of the complete response is requested, then only Complete Response should be considered. Usually death is counted as a progression event, but there could be a case, where the cause of death is taken into account, so, for example, if a patient died in a car accident, it is definitely not a real progression event.

In case the patient is a responder, meaning, the best overall response of Complete or Partial Response was documented, but later on no progression event occurred, they should be censored at the date of the last available non-PD tumor assessment, i.e. the duration of the response will be defined as the time between the first response date and the date of the last tumor assessment.

DURATION OF STABLE DISEASE

Duration of Stable Disease is measured from the start of the treatment (the first treatment date for non-randomized studies, the randomization date for randomized studies) until the date when the criteria for Overall Disease Progression are first met. In case no progression occurred, the patient should be censored at the date of the last tumor assessment.

The clinical relevance of the Stable Disease duration varies for different tumor types and grades. Therefore, it is highly recommended that the protocol specify the minimal time interval required between the two measurements for determination of Stable Disease. This time interval should take into account an expected clinical benefit that such a status may bring to the population under the study.

PROGRESSION FREE SURVIVAL

Progression Free Survival (PFS) is one more way to measure a potential anticancer activity. It is defined as the time from the start of the treatment (the first treatment date for non-randomized studies, the randomization date for randomized studies) until the disease progression or death from any cause. In case Progressive Disease/death did not take place, the patient should be censored at the date of the last reported tumor assessment.

PFS is increasingly used as the primary efficacy endpoint in the Phase III clinical trials, while Objective Response Rate (ORR) is usually chosen as the main goal for Phase II studies.

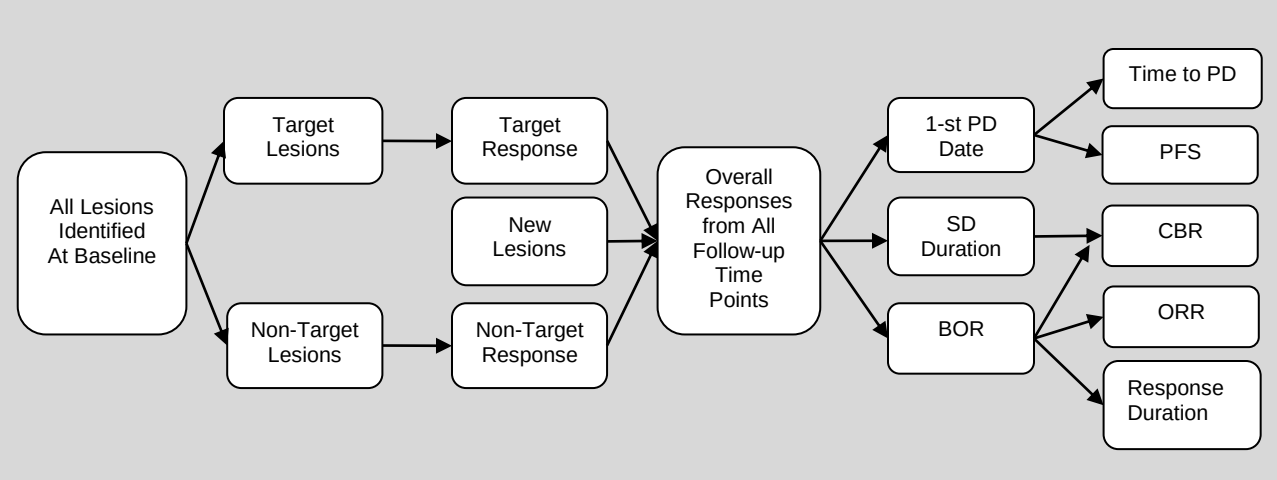
Ideally, trials with PFS as the primary end-point should be randomized and double-blinded, since otherwise promising PFS results may relate to the lucky patient selection and not the study treatment effect.

CONCLUSION

The paper provides a general step by step vision of the efficacy analysis in solid tumors oncology clinical trials, based on fundamental principles and algorithms of the Response Evaluation Criteria in Solid Tumors, version 1.1 (RECIST 1.1).

It also includes specific notes that should be considered in the tumor response analysis to receive the most accurate and honest statistical feedback.

The whole picture of the efficacy analysis described in the paper looks as follows:



After reading the paper a reader should be able to understand and implement in the analysis routine:

- the whole process of solid tumor evaluation at baseline, as well as at follow-up visits, including methods and frequency of assessments;
- the idea of tumors classification and the specificity of the analysis of each class;
- the definition and derivation algorithm for the common points of interest in oncology clinical trials, such as Best Overall Response, Time to Disease Progression, Progression Free Survival and so on;
- features of the oncology efficacy analysis depending on the study phase and type.

REFERENCES

1) New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1), 2009
URL: ctep.cancer.gov/protocolDevelopment/docs/recist_guideline.pdf

2) New Guidelines to Evaluate the Response to Treatment in Solid Tumors, 2000
URL: ctep.cancer.gov/protocolDevelopment/docs/therasserecistjnci.pdf

3) Common Pitfalls of RECIST 1.1 Application in Clinical Trials, 2015
URL:www.researchgate.net/publication/280575735_Common_Pitfalls_of_RECIST_11_Application_in_Clinical_Trials

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Please contact the author at:

Anastasiia Tiurdo
Senior Statistical Programming Analyst
Intego Group, LLC
43/2 Gagarin Avenue
Kharkiv, Ukraine/ 61001
Work Phone: +38 044 500 7020 ext. 2414
Fax: +38 057 728 3035
Email: anastasiia.tiurdo@intego-group.com