EUROPEAN JOURNAL OF CANCER 47 (2011) 697–701

available at www.sciencedirect.com

ScienceDirect

journal homepage: www.ejconline.com

Recommendations for capture, validation and summarisation of data from studies using RECIST

Melanie Paules [a,*], *Michelle Casey* [a], *Grant Williams* [d], *R. Suzanne Swann* [a], *Philip S. Murphy* [c], *Vanessa M. Salazar* [a], *Diane Foose* [a], *Barbara Bailey* [a], *Patrick Therasse* [b]

[a] *GlaxoSmithKline, Collegeville, PA, USA*
[b] *GlaxoSmithKline Biological, Rixensart, Belgium*
[c] *GlaxoSmithKline, Middlesex, UK*
[d] *Williams Cancer Drug Consulting, LLC, Wayne, PA, USA*

ARTICLE INFO

Article history:
Received 1 October 2010
Received in revised form 16 December 2010
Accepted 22 December 2010

Keywords:
Response criteria
Solid tumours
Guidelines
Response rate
Data analysis
Censoring

ABSTRACT

Response Evaluation Criteria in Solid Tumors (RECIST) is commonly used in oncology clinical trials and provides a standard approach for the assessment of treatment. However, data capture, validation and summarisation are complex. This article focuses on managing solid tumour lesion and response assessment data from capture through summarisation. Conventions for capturing lesion data, as well as considerations for data validation and summarisation, are provided. Recommendations are based on a review of data capture forms (including data items and instructions), data validation practices and algorithms for data summarisation across solid tumour studies at a single company. The intent of the authors is to share our experiences at GSK in the spirit of harmonisation by transparently describing our decisions and methods.

1. Introduction

The Response Evaluation Criteria in Solid Tumors (RECIST), version 1.0 has been widely used in oncology clinical trials.[1] It is expected that the recently published version, RECIST 1.1 will become increasingly adopted for new studies.[2] While RECIST provides simple definitions for the assessment of treatment outcomes in solid tumour studies, processes for data capture, validation and summarisation are complex.

A project began within GSK to develop a harmonised approach for processing and summarising disease progression and response evaluation data according to RECIST. A small, multi-disciplinary group within the oncology organisation was formed comprising two clinicians, a radiologist, two programmers, a clinical scientist, a data manager and two statisticians. This expert team facilitated discussion and decision-making to provide direction to an internal oncology standards team to implement or modify standard response and lesion assessment data capture forms, validations, algorithms and statistical displays. This standards team is a cross-functional group that develops and maintains standards for data collection through reporting and analysis for phase 1–4 oncology

* *Corresponding author*: Address: GlaxoSmithKline, 1250 South Collegeville Road, P.O. Box 5089, Collegeville, PA 19426-0989, USA. Tel.: +1 610 917 5104; fax: +1 10 917 4701.
E-mail address: Melanie.a.paules@gsk.com (M. Paules).
0959-8049/$ - see front matter
doi:10.1016/j.ejca.2010.12.023

studies. The team provides a mechanism to ensure compliance with industry standards and uniformity across trials.

2. Purpose of this guidance

In this paper, we present recommendations for data capture, validation and summarisation based on a single company's experience implementing RECIST across multiple compounds. However, we recognise that there are limitations of this guidance, which is based on a consensus approach within the expert team. These recommendations do not cover all possible situations where another response assignment should be used. The intention behind publishing this document is to help those facing the recurrent problems of practical implementation of RECIST. We have found that these approaches emphasise the importance of clearly defined protocol requirements for scans, and provide consistent methodology to process and summarise RECIST data. Outside the primary scope of the project were trials using volumetric tumour measurements; prostate cancer trials with PSA endpoints; Inflammatory Breast Cancer (IBC), lymphoma, mesothelioma and GIST trials. Nevertheless, these trials can apply some of the recommendations from the project. For example, recommendations can be adapted for malignant lymphoma studies using Cheson criteria for response.

3. Data capture recommendations

3.1. *General set-up of data capture forms*

As a standard practice visit level lesion and response data are collected, where 'visit' refers to any visit, scheduled or unscheduled, during treatment or follow-up and 'response' refers to RECIST responses (CR = complete response, PR = partial response, SD = stable disease, PD = progressive disease, Non-CR/Non-PD, NE = not evaluable). Underlying lesion data as well as assessments of response are captured using four separate forms: target lesions, non-target lesions, new lesion and response. It is understood that there may be other information available to the investigator, (e.g. symptomatic deterioration) which may impact the response provided. Therefore, in the case of PD, a text field is used to capture other evidence of progression although PD based on a radiological assessment is preferred.

The target lesion form contains: lesion index number (LIN), organ, a textual description of the lesion location, method, assessment date, lesion diameter and lesion status (see Table 1). LIN is derived by the system to ensure that the same LIN is associated with a given lesion across all visits. The lesion location description aids consistency of reporting by reducing ambiguity when following multiple lesions in a given organ. Lesion status is used in conjunction with the diameter to accurately describe the lesion. As an example, lesions that become too small to measure will have the diameter reported as '5 mm', which is included in the calculation of the sum of the diameters, and a lesion status of 'Too small to measure'.

In addition to the data fields noted above, the target lesion form provides RECIST calculations (including sum of the diameters and percentage changes from baseline and nadir) and captures the response for target lesions. The calculations are 'read-only' fields which aid in data capture and validation as they provide a mechanism for verifying that the target lesion response follows RECIST.

Table 1 – Lesion status.

Target lesion status	Absent Present Too small to measure Split or divided Merged or coalesced Not done Not assessable
Non-target lesion status	Absent Present but not unequivocal progression Unequivocal progression Not done Not assessable

The non-target lesion form contains: LIN, organ, location description, method, assessment date and lesion status (see Table 1). In addition, the form collects the overall assessment of non-target response, which is cross checked with the individual non-target lesion statuses.

The response form captures the investigator assessment of response at a visit and, in the case of PD, details on whether the PD is based on radiological or non-radiological assessment (e.g. symptomatic deterioration or other methods not used for RECIST calculations (e.g. lab values)). Target and non-target lesion responses are electronically mapped from the respective forms as 'read-only' fields to ensure consistency. In addition, a prompt for the presence or absence of new lesions is provided.

In the event that new lesions are present, a separate form is generated to capture LIN, organ, location description, method(s), assessment date(s) and lesion diameter (if available).

3.2. *Conventions for split lesions*

Split lesions can be either be documented as separate 'child' lesions or reported as a summary under the 'parent' lesion. For example if lesion '1' splits into two lesions, lesions 1a and 1b can be reported with diameters recorded for each 'child' lesion. However, it was decided that split lesions can be managed more efficiently if split lesions are reported as a summary under the parent lesion. Using this method the number of lesions followed remains consistent across the visits. The diameters of both 'child' lesions are measured, summed and recorded as the diameter for the originally recorded lesion with a lesion status of 'Lesion split or divided'.

3.3. *Conventions for merged or coalesced lesions*

Conventions for merged or coalesced lesions have also been developed. The method for entering data must ensure that duplicate measurements are not reported to ensure the sum of the diameter calculation is correct. The following approach was decided upon: if two existing lesions merge, the diameter

of the confluent mass is measured and recorded as the diameter for one of the lesions; 0 mm is recorded as the diameter for other lesion. A lesion status of 'Lesion merged or coalesced' is entered for both lesions.

3.4. *Lesion status ('Not done' vs. 'Not assessable')*

It is critical to understand the reason for missing data, therefore, the lesion status must distinguish between scans that were not performed (i.e. Not done) and scans which are performed but from which the radiologist could not provide an assessment (i.e. Not assessable). A text field captures the reason a lesion was not assessable (e.g. image quality, fibrosis).

3.5. *Method of assessment*

Each lesion should be assessed by the same method throughout the study. However, there are instances when a new modality is used, and the impact depends on the response to be assigned and whether the method was used for the evaluation of a target or non-target lesion. When any target lesion is assessed by a method other than that used at baseline, the response must be reported as 'not evaluable' since accurate calculations of the changes from baseline and nadir cannot be satisfactorily made. In addition to the requirement for target lesions to be assessed by the same method as used at baseline, a CR requires non-target lesions to be assessed by the same method used at baseline. However, a PR, SD or PD may be assigned even if the non-target lesions are assessed by a method that differs from baseline; these cases can be treated as a protocol deviation in the analysis if appropriate.

3.6. *Specificity of non-target lesion data collection*

The data capture of non-target lesions can vary. In some cases each non-target lesion is recorded as a separate entry and an assessment is made for each non-target lesion while in other cases non-target lesions are grouped by organ and an overall assessment is made for the organ. Following non-target lesions individually and processing them algorithmically may result in a single non-target lesion driving the response. As per RECIST 1.1 it is sufficient to capture and follow non-target sites of disease by organ.[2]

3.7. *Collection of 'Additional' scans*

Target and non-target lesion forms capture 'Method' of assessment. Supplementary forms capture any further scans performed to assess disease status, e.g. modalities other than the primary imaging modality used to rule out progression (i.e. based on new symptoms). The rationale for capturing 'Additional' scans is threefold. First, it provides a complete record of all scanning conducted, including negative scans. Second, it supports database reconciliation to ensure that when centralised review of images is required, all known imaging investigations are collected. Third, it helps to clearly differentiate correlative imaging from the primary imaging modality.

4. Considerations for data validation and summarisation

The visit response data are often used to programmatically derive best overall response (BOR) and date of progression,

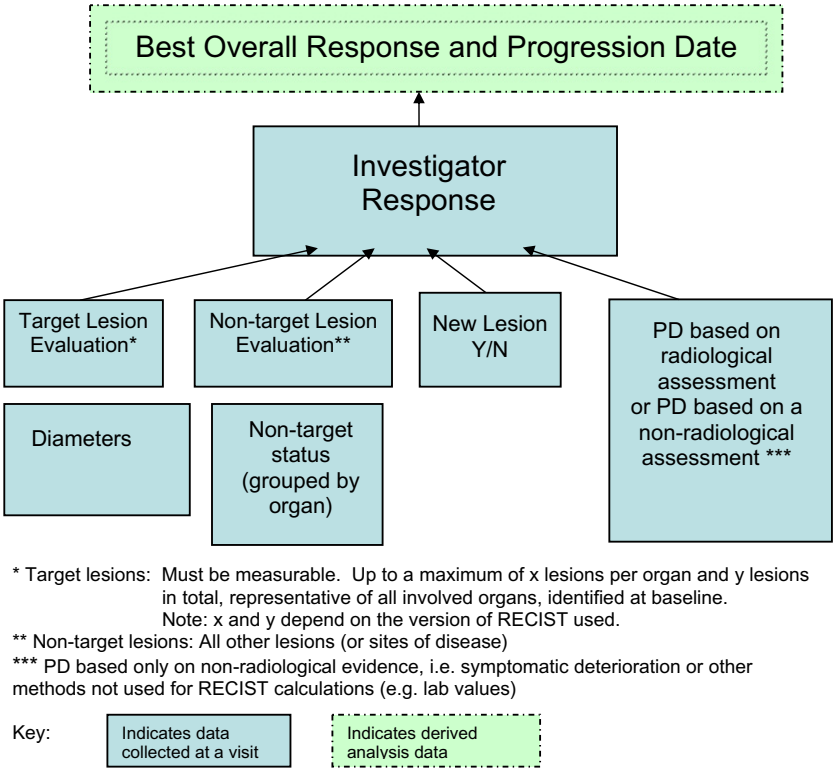

Fig. 1 – Relationships between Response Data Elements.

which are used in BOR and time-to-event (TTE) analyses. Because analyses depend on the underlying data, considerations for data validation and summarisation must follow the same principles. If an organisation collects BOR the same principles apply and validations of the collected BOR will also be needed. The diagram provides the relationships between data elements (see Fig. 1).

4.1. *Target lesions reported as 'Not assessable' or 'Not done'*

Target lesions that disappear will be recorded with a lesion measurement of 0 which is used in the calculation of the sum of the diameters. If any measurement is not provided, the sum of the diameters cannot be calculated for purposes of assigning response or for use as the nadir for future assessments. However, the sum of the non-missing diameters should be calculated to evaluate PD. If an assessment of PD cannot be made, the response should be 'not evaluable' for the target lesions.

4.2. *Calculating nadir*

All previous data, including baseline but excluding the current assessment, are used to identify the nadir.

4.3. *Investigator-reported visit response*

There are instances when the investigator-reported response at a visit does not coincide with the supporting lesion data. In order to respectfully query the investigator-reported responses to ensure that data capture errors have not occurred, the follow hierarchy is used.

- First, the target response is compared to the RECIST calculations.
- Second, the non-target response is compared to the underlying assessment of each organ containing non-target disease.
- Third, the investigator-reported response is compared to ensure alignment with each of the following components: target response, non-target response, the presence of new lesions and the results of the question on whether PD is determined based on radiological or non-radiological assessment.

As the clinical management of the patient resides with the treating physician, the response as reported by the investigator is used in analyses of BOR, regardless of whether the data query resolution matches RECIST. However, in the cases where PD is documented based on non-radiological evidence, the BOR is derived based on the radiological data. As per RECIST, symptomatic deterioration is not a descriptor of objective response. Symptomatic progression can be used in progression-free survival and time to progression analyses as appropriate.

4.4. *Non-target lesion assessment schedule*

We note that the assessment schedules for certain sites of non-target disease can differ from the assessment schedule of other lesions, which may lead to differential evaluation of non-target lesions in different patient populations. As one example, bone scintigraphy is performed at baseline for all subjects and every 24 weeks thereafter for subjects with baseline bone disease. In a different example, scans used to evaluate CNS disease are required for all subjects at baseline and every 3 months thereafter regardless of the presence of CNS disease at baseline. Obviously, in the first example, the population with bone lesions at baseline has a higher chance of detecting new bone lesions during follow-up, whereas in the latter design, new brain lesions are equally detected in all subjects. Determination of the definition of 'adequate assessment' in these cases is a complex issue, affecting both BOR as well as TTE analyses.

The rules for censoring TTE endpoints are not specifically discussed here; however, the determination of visit responses directly affects these censoring rules. In particular, TTE endpoints are censored at the date of an adequate assessment. Adequate assessment for the purpose of censoring progression-free survival, time to progression and response duration is defined as any assessment where a response determination of CR, PR or SD can be made. A response of 'not evaluable' does not constitute an adequate assessment.

In protocols where separate scans of specific non-target lesions are required on a schedule that does not require them to be performed at each disease assessment, the response determination of CR, PR or SD will depend on the scans performed at each assessment (to determine the target lesion measurements and the assessment of non-target lesions covered by the same scans) and the adherence to the protocol specified schedule for specific sites of non-target disease which require separate scan(s). Scans of sites where disease was not present at baseline are not necessary to document a CR, PR or SD. The rationale is that RECIST responses are based on the evaluation of disease present at baseline as well as new lesions. It is assumed that the possible presence of new lesions will be evaluated as per clinical judgment.

Two approaches are recommended for assigning BOR when separate scans of specific non-target lesions are required on a protocol specified schedule that does not require them to be performed at each disease assessment.

- The standard approach applies when the protocol does not contain additional requirements for protocol specified sites of non-target disease beyond the expectations within RECIST. In the standard approach, at least one post-baseline assessment of non-target lesions is required within a protocol specified window only to verify disappearance of disease to establish a BOR of CR. Furthermore, a BOR of PR can be established without a post-baseline assessment of non-target lesions under the presumption that the non-target lesions are not worsening based on lack of any other findings suggesting progression.
- The rigorous approach requires at least one post-baseline assessment of specific non-target lesions to establish a BOR of both CR and PR within a protocol specified window. The rigorous approach builds on the standard approach by including requirements for establishing PR in addition to establishing CR.

These rules apply to subjects with measurable disease in addition to sites of non-target disease. Specific rules would need to be adapted for subjects who have only non-measurable disease to ensure that adequate assessments are performed to document response.

In studies where the protocol states that specific scans are to be performed even when disease was not present at baseline, the recommendation is to not consider these additional protocol requirements when assigning the visit response and BOR. However, the statistical analysis plan can specify sensitivity analysis where more stringent requirements are needed to consider an assessment adequate for the purpose of censoring of progression-free survival, time to progression, and response duration analyses.

4.5. *Dates of response at each assessment*

For any one response assessment, scans to evaluate all sites of disease can be performed on multiple dates. Nonetheless, a single date is required for the associated response. The following rules are used to assign dates to responses:

PD: The date of progressive disease is the earliest date that PD has been documented, i.e. the earliest of:

- Dates of target lesion assessments when the target lesion response is PD
- Dates of non-target assessments where the lesion status is unequivocal progression
- Date(s) documenting new lesion(s)
- If specified in the protocol and/or analysis plan, the date of PD determined based on non-radiological assessment

If none of the above dates are available, the earliest scan or visit date for the assessment at which PD is reported is used.

SD: The date of stable disease is the earliest date of lesion assessment at a visit. The earliest date provides the most conservative method of determining SD because all scans used to determine SD are dated the same as or later than the protocol specified cut-off for SD.

CR/PR: The date of complete or partial response is the latest date of a lesion assessment at the visit. If any assessment dates are missing or not reported, then a date of response cannot be assigned. Use of the latest date aligns with the requirement to assess all target lesions when assigning response.

4.6. *Confirming response*

The team specified how to confirm CR or PR responses based on the response dates (e.g. CR or PR can be confirmed no earlier than 4 weeks after the first determination of response). Because different scans can be done on different dates, our approach is to (1) calculate the time between initial and confirmatory assessments for all target and non-target lesions separately; (2) select the minimum time; and (3) then to determine if the confirmation criterion were met.

4.7. *Non-measurable only disease*

As per RECIST subjects with only non-measurable disease can achieve a CR, Non-CR/Non-PD (classified as SD in RECIST 1.0 studies), or PD but not a PR. For studies with response rate as a primary endpoint, which inadvertently enrol subjects with only non-measurable disease, it was decided that separate summaries are necessary for subjects with measurable disease at baseline and subjects with only non-measurable disease at baseline.

For studies with response rate as a secondary endpoint, which enrol subjects with non-measurable only disease, response rates may be reported using an analysis including all subjects in the denominator or an analysis where only the subset of subjects with measurable disease at baseline are included. A consideration in deciding how to report response rate is whether consistency should be maintained with other studies across phases.

5. Summary

A comprehensive review and refinement of the practices for processing and summarising lesion assessment and response was conducted. The internal data standards team implemented these decisions and now a full set of standard components for lesion assessment and response data are available to study teams for data capture and validations, algorithms and standard displays for reporting. The relationship between the protocol specified requirements for scans and the impact on the analyses was clarified. This harmonised approach to RECIST across studies has resulted in greater efficiency and comparability of trial results.

Conflict of interest statement

None declared.

Acknowledgements

The members of the RECIST Standard Reporting Project would like to thank GlaxoSmithKline Oncology R&D for the support of this effort. We would also like to thank the following individuals in providing support to this effort and input during the development of the Guidelines for Response Assessment Data: Will Bushnell, Ohad Amit and Lauren McCann, GlaxoSmithKline USA.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ejca.2010.12.023.

REFERENCES

1. Therasse P, Arbuck SG, Eisenhauer EA, et al. New guidelines to evaluate the response to treatment in solid tumours (RECIST guidelines). *J Natl Cancer Inst* 2000;**92**:205–16.
2. Eisenhauer EA, Therasse P, Bogaerts J, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer* 2009;**45**:228–47.