ELSEVIER

European Journal of Cancer 41 (2005) 1528–1532

European Journal of Cancer

www.ejconline.com

Assessment of the value of confirming responses in clinical trials in oncology

Jose Luis Perez-Gracia [a,*], Maria Muñoz [b], Grant Williams [c], Jun Wu [d], Eva Carrasco [b], Ignacio Garcia-Ribas [b], Ana Peiro [b], Jose Maria Lopez-Picazo [a], Alfonso Gurpide [a], Ana Chopitea [a], Salvador Martín-Algarra [a], Jesus García-Foncillas [a], Johannes Blatter [e]

[a] *Medical Oncology Department, Clínica Universitaria de Navarra, Avenida de Pío XII 36, 31008 Pamplona, Spain*
[b] *Clinical Research Department, Eli Lilly and Company, Madrid, Spain*
[c] *US Food and Drug Administration, USA*
[d] *Clinical Research Department, Eli Lilly and Company Indianapolis, USA*
[e] *Clinical Research Department, Eli Lilly and Company, Frankfort, Germany*

Received 6 December 2004; received in revised form 20 January 2005; accepted 27 January 2005
Available online 27 April 2005

Abstract

The requirement for a second assessment to confirm initial tumour response is required by all response guidelines. Its rationale, however, is not clear. We have conducted this study to compare validity of response rate assessment determined with and without secondary confirmation. Using specified criteria, nine trials of one single cytotoxic drug including 416 patients were selected from a pharmaceutical database. Objective response rates were determined by a single determination and by two separate determinations. 81 responses (19.5%, [15.8–23.6%]) were scored by the confirmation method and 97 responses (23.3% [19.3–27.7%]) by the no-confirmation method. The Kappa (κ) coefficient of 0.89 indicates good agreement between both methods. This is the first study that systematically compares response rates calculated with and without performing response confirmation. Results show good agreement between both methods. We suggest that assessing response without confirmation may be the preferred method. These results should be confirmed by additional studies in a variety of cancer settings.

Keywords: Response assessment; Response guidelines; Clinical trials

1. Introduction

Clinical response rate is the most widely used surrogate marker of anticancer drug activity. Several guidelines have been developed to ensure that response is assessed using uniform and consistent criteria [1–5]. All guidelines state that once a clinical response is observed, it must be confirmed by repeating the disease assessment at some minimum interval, usually 3–6 weeks from the initial response observation. Nonetheless, the rationale for requiring response confirmation is neither clearly explained in any of the guidelines for response assessment nor is it evident on careful analysis.

Common sense dictates that simple standards are preferred unless more complicated standards provide some added value. Even though the requirement for response confirmation, a more complicated standard, appears to be universally accepted, it is not clearly based on biological considerations or on experimental data. We performed this study to evaluate the added value of

* Corresponding author. Tel.: +34 658 92 02 12; fax: +34 948 255 500.
E-mail address: jlgracia@unav.es (J.L. Perez-Gracia).

0959-8049/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ejca.2005.01.023

requiring response confirmation compared to requiring only a single response observation.

2. Methods

2.1. Study selection criteria

The Eli Lilly database was the clinical trial source for this study. We selected available final response information from all published and unpublished phase II trials exploring the activity of one specific drug, as single agent or in combination, in a wide variety of tumour types. All studies used Southwestern Oncology Group (SWOG) criteria [4] for response assessment and confirmation was mandatory in all of them. Response rates determined by assessment of the best response for each patient were compared with response rates determined after confirmation was performed. In patients in whom initial response was not subsequently confirmed for any reason, the remaining response information was analysed.

All the studies were high-quality registration trials performed in experienced investigational sites and according to Good Clinical Practice guidelines. Consequently, all responses were evaluated by qualified specialists. All the trials were approved by the appropriate Ethical Review Board of the centre and by regulatory authorities.

2.2. Statistical methods

Response rates were calculated by dividing the total number of responses by the total number of evaluated patients, as defined in the respective protocols. Ninety five per cent confidence intervals (CI) of the rates were constructed using exact binomial method [6]. The difference in percentage between response rates with and without confirmation was estimated. The 95% CI of this difference was calculated from the response agreement rate of both methods and its 95% CI, by multiplying the reciprocals minus one of both limits of this interval by the confirmed response rate.

An assessment of the agreement between both methods was obtained based on the Cohen's Kappa (κ) coefficient [7]. The relationship between the κ coefficient and strength of between-method agreement may be interpreted in the following manner: 0.81–1.00, almost perfect; 0.61–0.80, substantial; 0.41–0.60, moderate; 0.21–0.40, fair; 0–0.20, slight and <0, poor [8].

All statistical analysis was done using SAS®.

3. Results

Nine phase II trials that met the inclusion criteria were selected. The studies included a total of 416 patients, in which 81 responses (19.5%) were observed when response confirmation was required compared with 97 responses (23.3%) relying on only the first response assessment. The average difference in response rates was 3.9% (2.1–6.6%). The κ coefficient for all studies combined was 0.89. All κ coefficients were greater than 0.81 (almost perfect agreement) except for one κ coefficient of 0.73 (substantial agreement). Confidence intervals for response rates calculated by both methods overlapped widely for each study. Global results are shown in Table 1 and results for each individual trial are shown in Table 2.

Responses were observed but not subsequently confirmed in 16 patients. Only one of these patients had disease progression in the next cycle after initial response documentation. In seven patients, no further evaluation of response status was performed. In the remaining eight patients, progression was observed, but in each case there was at least one visit where response might have been confirmed, yet no re-assessment was performed.

4. Discussion

Our results show that response rates determined by the first assessment of best response observed in each patient are reasonably similar to the results obtained when confirmation by a second evaluation is required. In each of the trials analysed, we observed a κ coefficient reflecting a high level of agreement between the two methods. On average, results in this solid tumour population show a slightly higher response rate when no confirmation is required. It seems unlikely that decisions taken based on these trials would have been changed if confirmation had not been performed.

Responses were observed in the first assessment, but were not confirmed in 16 patients. As presented above, only one of them had disease progression in the very next cycle after the first assessment. Therefore, we can hypothesise that at least some of the remaining 15 patients may have qualified responses if disease had been re-evaluated at the minimum interval allowed. This represents a potential source of heterogeneity in the results, which could depend on factors that are extrinsic to the true activity of the drug, such as the willingness and/or

Table 1
Response data from all patients included in the trials selected for the study, assessed with and without confirmation of first best response observed

	Confirmed response		
	Responders	Non responders	
1st Response evaluation			
Responders	81	16	97
Non responders	0	319	319
	81	335	416

Table 2
Response rates of individual trials selected for the study, assessed with and without confirmation of first best response observed

Tumour type	Number of evaluable patients	Response assessment	Complete responses	Partial responses	Overall response		κ
					Rate (%)	(95% CI)	
NSCLC	36	Confirmation	0	13	36.1	(20.8–53.8)	1.0
		No confirmation	0	13	36.1	(20.8–53.8)	
NSCLC	51	Confirmation	0	9	17.7	(8.4–30.9)	0.88
		No confirmation	0	11	21.6	(11.3–35.3)	
Bladder	28	Confirmation	0	9	32.1	(15.9–52.4)	0.92
		No confirmation	0	10	35.7	(18.6–55.9)	
Breast	78	Confirmation	0	7	9.0	(3.7–17.6)	0.86
		No confirmation	0	9	11.5	(5.4–20.8)	
Breast	72	Confirmation	3	12	20.8	(12.1–32.0)	0.88
		No confirmation	3	15	25	(15.5–36.6)	
Pancreas	41	Confirmation	0	5	12.2	(4.1–26.2)	0.73
		No confirmation	0	8	19.5	(8.8–34.9)	
Head and neck	34	Confirmation	0	9	26.5	(12.9–44.4)	0.86
		No confirmation	1	10	32.4	(17.4–50.5)	
Colorectal	40	Confirmation	1	5	15.0	(5.7–29.8)	0.83
		No confirmation	1	7	20.0	(9.1–35.7)	
Gastric	36	Confirmation	2	6	22.2	(10.1–39.2)	0.92
		No confirmation	2	7	25	(12.1–42.2)	
Overall	*416*	*Confirmation*	*6*	*75*	*19.5*	*(15.8–23.6)*	*0.89*
		No confirmation	*7*	*90*	*23.3*	*(19.3–27.7)*	

CI, confidence interval; NSCLC, non-small cell lung cancer.

the ability of the physician/patient to perform or not the re-assessment in the minimum interval allowed.

The rationale for performing response confirmation has not been clearly explained in any of the guidelines that require its use. In one of the first guidelines developed to assess responses, specifically in breast cancer patients, the rationale provided for confirming responses was that some patients achieving a response may have an additional period where the response changed [2], but these changes also happen frequently even after response is confirmed. The original criteria from the World Health Organisation just noted that four weeks should be the minimum duration of reported response, even though it was recognised that in some trials a shorter duration of response may be useful [3]. The SWOG criteria requires a minimum period of 3–6 weeks to confirm response but gives no background to justify this procedure [4]. Finally, the most recent Response Evaluation Criteria in Solid Tumours (RECIST) guidelines state that the goal of confirming responses is to avoid overestimating response rates [5]. Nonetheless, even though overestimation of response rates is a real problem in oncology trials, it has never been proven that confirmation is an effective method to control it.

While the benefit of confirming responses is unproven, the potential difficulties seem straightforward:

- Response confirmation is a time consuming procedure for everyone involved in a clinical trial, from the patient to the clinician, including data management staff and radiologists. Furthermore, a small but certain risk of complications is associated to many radiological procedures, especially when they involve the use of intravenous contrasts.
- Response confirmation increases the expense of clinical trials and uses scarce resources. The introduction of new, more complex imaging techniques that will be used to assess response in the near future will make this problem even greater.
- Response confirmation is a potential source of heterogeneity for interpreting results from clinical trials. In first place, because assessing responses twice is more complicated than doing it once and the process is more prone to error [9]. In second place, because most guidelines state that response should be confirmed in a minimum period ranging from 3 to 6 weeks while a maximum period is not defined. Consequently, studies in which response is confirmed in the minimum period allowed may show better response rates than trials in which responses are confirmed at longer intervals. Such variations in the timing for confirming responses may be due to differences in the resources allotted to the trials, differences in the duration of a treatment cycle or other reasons, but seem in any case artificial and extrinsic to the activity of the drug tested. Finally, some trial response rates are likely reported without actually being confirmed. This seems likely in reports that are not peer-reviewed, such as abstracts or reports presented at meetings. This could artificially increase response rates in comparison with trials in which responses are confirmed by strict criteria. If response

confirmation were not performed, all these potential sources of heterogeneity would no longer be a matter of concern.

The requirement for response confirmation may have been intended to increase the precision of the objective response assessment (by correcting initial measurement errors) or to increase the clinical significance of objective response assessment (by "raising the bar" to require a minimal duration). Neither of these reasons for requiring confirmation seems relevant when assessing response rate either in a single-arm study or in a comparative study. In a single arm study, differences in response rates between different study populations are likely to be significantly larger than differences in response rates using these two methods, as shown in Table 2. The only meaningful information that a phase II trial can contribute regarding antitumour activity is a gross approximation of the response rate and a measure of response duration. For this purpose, a slightly lower response rate obtained using confirmation seems to be of no more value that the more simple method. In the setting of a randomised controlled trial, the single assessment method would seem to have a theoretical advantage. In randomised studies, the relative difference in response rates between study arms is more relevant than the absolute response rate. The confirmation method would seem to be more susceptible to bias, e.g., response rates may vary if the rate of obtaining confirmation scans is higher on one study arm than the other. In either case, in a single arm study or in a comparative study, there appears to be no clear or proven advantage to requiring confirmation. Our results suggest that there would be little cost or risk to assessing the response rate with the first observation and evaluating the quality of responses using response duration.

Eliminating the requirement to confirm responses should not imply that a patient does not need to undergo subsequent evaluations of disease. Repeated evaluations are required to determine whether a partial response improves to a complete response, to evaluate response duration and progression-free survival and to determine when to discontinue therapy. In addition, response confirmation should be differentiated from the response review that is sometimes performed by independent external reviewers and which has proved to be an effective method to avoid overestimation of drug activity [10–12].

The main limitation of our study is that it has been performed using a limited number of phase II trials and with a single cytotoxic drug. Other studies should compare these methods in a variety of tumour types and with a variety of cancer drugs. If differences between the methods are reasonably predictable preference should be given to the method that is simpler and less expensive, laborious and prone to heterogeneity. In addition, if future studies demonstrate a significant difference between assessing responses with and without performing confirmation, they should also consider what is the advantage of using the more complex procedure.

In summary, to our knowledge this is the first study that has analysed the value of confirming responses in clinical trials in oncology. According to our data, response confirmation does not seem to add any value to response assessment using the first best response observed. Moreover, confirmation may arbitrarily reduce the true response rate of the tested drugs. It appears, that confirmation may be unnecessary, since it is expensive and time consuming and it may increase heterogeneity in results. Since response confirmation constitutes a widespread practice, additional studies are required to confirm our results before recommending definitive changes.

Conflict of interest statement

None declared.

Acknowledgement

We are indebted to James Symanowsky for his comments to the statistical design.

References

1. Hayward JL, Carbone PP, Heuson JC, *et al*. Assessment of response to therapy in advanced breast cancer: a project of the Programme on Clinical Oncology of the International Union Against Cancer, Geneva, Switzerland. *Cancer* 1977, **39**, 1289–1294.
2. Hayward JL, Rubens RD, Carbone PP, *et al*. Assessment of response to therapy in advanced breast cancer. A project of the programme on clinical oncology of the International Union against Cancer, Geneva, Switzerland. *Eur J Cancer* 1978, **14**, 1291–1292.
3. Miller AB, Hoogstraten B, Staquet M, *et al*. Reporting results of cancer treatment. *Cancer* 1981, **47**, 207–214.
4. Green S, Weiss GR. Southwest Oncology Group standard response criteria, endpoint definitions and toxicity criteria. *Invest New Drugs* 1992, **10**, 239–253.
5. Therasse P, Arbuck SG, Eisenhauer EA, *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* 2000, **92**, 205–216.
6. Leemis LM, Trivedi KS. A comparison of approximate interval estimators for the Bernoulli parameter. *The American Statistician* 1996, **50**, 63–68.
7. Cohen J. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* 1960, **20**, 37–46.
8. Landis J, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977, **33**, 159–174.

9. RECIST Working Party. How to combine the results of two subsequent evaluations to define the best overall response. June 2000. Available from URL: http://www.eortc.bc/.
10. Thiesse P, Ollivier L, Di Stefano-Louineau D, *et al*. Response rate accuracy in oncology trials: reasons for interobserver variability, Groupe Francais d'Immunotherapie of the Federation Nationale des Centres de Lutte Contre le Cancer. *J Clin Oncol* 1997, **15**, 3507–3514.
11. Biganzoli L, Lohrisch C, Paridacns R, *et al*. Analysis of two EORTC trials in metastatic breast cancer support the recommendation for external response review in trials with response rate as primary endpoint. *Eur J Cancer* 2000, **36**(suppl 5), S90. (abstract 227).
12. Gwyther SJ, Aapro MS, Hatty SR, *et al*. Results of an independent oncology review board of pivotal clinical trials of gemcitabine in non-small cell lung cancer. *Anticancer Drugs* 1999, **10**, 693–698.