Anti-Cancer Drugs 1999, **10**, pp. 693–698

Clinical review board

Results of an independent oncology review board of pivotal clinical trials of gemcitabine in non-small cell lung cancer

Stephen J Gwyther, Matti S Aapro,[1] Sarah R Hatty,[2] Pieter E Postmus[3] and Ian E Smith[4]
[1]East Surrey Hospital, Redhill, Surrey RH1 5RH, UK. [1]Clinique de Genolier, 1272 Geneva, Switzerland. [2]Sarah Hatty Associates, Esher, Surrey KT10 9UE, UK. [3]Department of Pulmonology, Free University Hospital, 1081 HV Amsterdam, The Netherlands. [4]Royal Marsden Hospital, Sutton, Surrey SM2 5PT, UK.

Response rates reported in early phase II clinical trials are often not reproduced in subsequent larger or phase III studies. Independent review of claimed partial or complete responders to gemcitabine was undertaken in four pivotal, open-label phase II studies of advanced, non-small cell lung cancer (NSCLC) to provide accurate, consistent, reproducible response rates. Patients were chemonaive and had stage III or IV NSCLC. In three trials, gemcitabine (800 and 1250 mg/m^2) was administered once-weekly for 3 weeks followed by a rest week. In the fourth, gemcitabine (90 mg/m^2) was given twice-weekly for 3 weeks in every 4 weeks. The primary endpoint was response rate. Of the 374 evaluable patients, 114 (30%) were claimed as responders. Independent review reduced this to 79 (21%). The response range was reduced from 25–35 to 20–23% after validation; 95% confidence intervals did not overlap. Consistent application of response criteria by an independent panel significantly reduced response rates but produced greater consistency and reproducibility. These results confirm that gemcitabine is active against NSCLC. Subsequent larger-scale studies have produced comparable response rates, vindicating the use of independent review. Independent review is recommended for all trials using response rate as a primary endpoint. [© 1999 Lippincott Williams & Wilkins.]

Key words: Gemcitabine, independent review, non-small cell lung cancer, response rates.

Support for these studies was provided by Eli Lilly and Co. SRH was an employee of Eli Lilly and Co.

Correspondence to SJ Gwyther, East Surrey Healthcare NHS Trust, East Surrey Hospital, Canada Avenue, Redhill, Surrey RH1 5RH, UK.
Tel: (+44) 1737 768511; Fax: (+44) 1737 780395;
E-mail: GwytherSJ@aol.com

Introduction

Historically, discrepancies have occurred in the evaluation of new chemotherapeutic agents in cancer trials. The reasons are varied but include insufficient care in the selection of the patient population, differences in prior therapy, difficulties in applying adequate staging systems for the disease and inadequately sized trials.[1,2] Additional factors which may affect the response rate are the protocol definition and choice of target lesions, response criteria definition, the quality, and hence reproducibility, of the clinical and radiological data, and tumor assessments and investigator bias. These and other factors have frequently called into question the validity of the reported treatment responses and have led to disappointing response rates when larger-scale trials have subsequently been performed. These issues have been apparent for many years, but have not yet been adequately or satisfactorily addressed. Consequently, in recent trials of new anti-cancer agents reported to show activity against a range of solid tumors, there has been a move toward the use of external review boards to reassess all patients who are claimed to show a response to treatment.[3–5] The use of review boards to independently assess data from clinical trials is generally welcomed by clinicians and the pharmaceutical industry, because independent verification of the data enhances the reliability and credibility of the data. Regulatory authorities also appreciate such data and, although not mandated, recent guidelines suggest such processes should be undertaken. For instance, the latest draft guidelines from the

European Community (EC) Commission of Proprietary Medicinal Products (CPMP) recommends that 'external independent review of the responses should be undertaken'.[6]

The Food and Drug Administration (FDA) are prepared to accept 'a surrogate endpoint which is reasonably likely to predict clinical benefit'[7] and therefore a significant objective response rate associated with tolerable treatment toxicity would be likely to support accelerated approval.

These criteria are specifically for new drugs in the treatment of certain serious or life-threatening illnesses including the refractory cancer setting, when no present therapies with meaningful benefit are currently available. The above statement demonstrates a genuine desire to expedite the widespread use of new, active agents in advanced cancer, for the benefit of the patient, based on sound scientific principles. This paper presents the findings of an external review board which reviewed the results of four early phase II clinical trials that were pivotal to the future development and registration of the anti-cancer agent gemcitabine (Gemzar®, Eli Lilly and Co, Indianapolis, IN), for the treatment of non-small cell lung cancer (NSCLC).[8-11]

It is of particular note that since these studies were concluded, the FDA's proposals for reform in 1996 included a section recommending increased use of the accelerated approval mechanism because 'commonly partial tumor shrinkages are induced and evidence has accumulated that such responses are often directly linked to longer or better patient survival. In fact for certain new agents, FDA has already begun to rely on a reasonably high rate of verifiable objective partial responses to the therapy as a basis for approval of agents to treat refractory malignancies without requiring evidence of improved survival and quality of life'.[12] The FDA clearly state that with refractory malignancy, or when there is no suitable alternative, objective evidence of antitumor activity is a reasonable basis for approving the drug, although post-approval studies will usually be required to confirm the effect on survival and quality of life.

The mandate of the review board was to provide an independent assessment of data from all patients for whom a partial or complete response to gemcitabine was claimed by the investigator, and to thus ensure the accuracy and reproducibility of the reported response rates in the four studies.

Patients and methods

Study design

A total of 417 patients with inoperable stage III or IV NSCLC were entered into four phase II open-label trials involving 21 different centers. Entry required all patients to be aged 18 years or older, to have a histologically proven diagnosis of inoperable stage III or IV NSCLC, with a WHO performance status (PS) of 0–2. Patients were required to present with at least one bidimensionally measurable lesion as evidenced by one or more radiological imaging modalities: chest radiograph (CXR), ultrasound (US), computed tomography (CT) or by physical examination (PE). In two of the studies (accounting for 227 evaluable patients), measurable disease was defined as bidimensionally measurable disease with margins clearly defined by diagnostic imaging studies with the largest dimension being 1×1 cm for CT and US, and greater than 2 cm for CXR. For PE, lesions had to be 2×2 cm. In the other two studies, bidimensionally measurable disease was required, with no minimum size stipulation. All patients were required to be chemonaive. Previous radiotherapy was acceptable in three of the four studies, but any measurable lesions were required to be outside the irradiated field. In three of the trials, gemcitabine was administered at doses between 800 and 1250 mg/m^2, once weekly for 3 weeks followed by a fourth rest week. In one of the trials, gemcitabine was administered at a starting dose of 90 mg/m^2, twice-weekly for 3 weeks in every 4 weeks. All four protocols were approved by the local Institutional Review Board and were performed in accordance with the Declaration of Helsinki, after relevant regulatory approval. All patients gave informed consent prior to study entry.

The primary endpoints for all the trials were response rate and duration of response. Patients in all four studies were qualified for response assessment if they met the following criteria: histological or cytological confirmation of NSCLC; at least three injections of gemcitabine; no prior chemotherapy; bidimensionally measurable disease at study entry and no concurrent anti-neoplastic therapy. The WHO guidelines for reporting cancer results were applied to measurable disease.[13] Response rates were defined as the number of patients who responded divided by the number of evaluable patients.

The independent review process

The mandate of the Oncology Review Board (ORB) was: (i) to verify claimed responses (a response to treatment included either partial or complete responses) in order to ensure the response rate had not been exaggerated by misinterpretation of radiological imaging factors; (ii) to review all radiological imaging in conjunction with the clinical summary; and (iii) to obtain a unanimous decision as to the appropriate response classification. All relevant documentation relating to each claimed responder was reviewed. The documentation relating to radiology (CXR, CT and US scans) was reviewed and assessed by an independent radiologist prior to the ORB meeting. The radiologist examined and re-measured the lesions, and any other abnormalities were identified and discussed by the ORB. All relevant documentation was reviewed by the ORB, including assessing severity of disease-related symptoms, and all relevant laboratory and clinical data. After thorough review of the data, a unanimous consensus was reached by the members of the ORB as to whether the response was valid. The investigator was informed if the ORB reached a different conclusion or if further information was required. The investigator could request the ORB to revisit the data if additional, clinically relevant data could be provided to the ORB. If agreement could not be reached between the ORB and the investigator, the case report forms (CRF) reflected the investigator-determined patient response according to the principles of Good Clinical Practice (GCP). However, the response determined by the ORB was used in all data analysis and submissions to the regulatory authorities.

Results

A total of 417 patients with stage III or IV NSCLC were enrolled into four open-label trials. Two of the four trials were conducted solely in Europe, one in Europe and Canada, and one in South Africa. A summary of the baseline characteristics of the patients enrolled is presented in Table 1.

Forty-three patients (10.3%) were not qualified for efficacy. Twenty-eight patients received insufficient therapy, 11 had no measurable disease, two insufficient therapy and no measurable disease, one received concomitant anti-cancer therapy, and one had no histologic confirmation of NSCLC. The ORB only examined data from patients for whom the investigator claimed a response. Of the 374 evaluable patients, responses were claimed for 114 (31%). Of these, 35 (31%) were rejected by the ORB, leaving a total of 79 validated responders and reducing the overall response rate from 31 to 21%. Table 2 shows the distribution of claimed and ORB-verified responders across the four trials.

Table 1. Baseline characteristics of the patients enrolled in all four studies

	Study 1[11]	Study 2[8]	Study 3[10] (twice-weekly dosing regimen)	Study 4[9]
Total entered	161	82	90	84
Male/female	124/37	49/33	48/42	65/19
Mean age (range)	59 (35–75)	52 (23–71)	58 (38–72)	59 (35–75)
WHO PS 0, 1, 2	17, 134, 10	20, 45, 16[a]	3, 75, 10[a]	2, 81, 1
Stage IIIA/IIIB/IV	7, 50, 104	18, 24, 40	9, 21, 60	15, 34, 35

[a]One patient PS 3.

Table 2. Distribution of claimed and ORB-validated responders

Study	No. of patients enrolled	No. of evaluable patients	No. of claimed responders	No. of ORB-validated responders	No. rejected	Rejection rate (%)
Study 1	161	151	52 (34%)	33 (22%)	19	37
Study 2	82	71	22 (31%)	16 (23%)	6	27
Study 3	90	76	21 (28%)	15 (20%)	6	29
Study 4	84	76	19 (25%)	15 (20%)	4	27
Total	417	374	114 (31%)	79 (21%)	35	31

95% confidence interval of evaluable patients: claimed responders 26–35%; validated responders 17–25%.

An intent-to-treat analysis of all 417 patients resulted in a claimed response rate of 27% (range 21-32%) and a validated response rate of 19% (range 17-21%). In that analysis, patients with non-measurable disease were excluded as responders but included in the denominator. Of the 35 claimed responders not verified by the ORB, 25 were rejected because they had stable disease and not a partial response, three had progressive disease, one patient had a partial response lasting less than 4 weeks and six patients had either non-measurable disease or no convincing evidence of disease.

Table 3 shows the number of patients assessed and rejected for each response assessment modality. The most frequently used methods were CT (71%) and CXR (24%) with rejection rates of 31 and 19%, respectively. Only six (5%) patients were assessed by PE alone and thus could not be independently reassessed.

Discussion

Response rates are commonly used to assess the activity of new anti-cancer agents. Consequently, it is imperative that the criteria used to define a response are accurately and consistently applied throughout any given trial and between different trials.

Errors in tumor measurement are major confounding factors in response assessment. As far back as 1976, Moertel and Hanley studied the effect of measurement error by physical examination.[14] Experienced oncologists were required to measure the size of pairs of steel spheres covered by foam rubber and placed on a mattress to simulate lesions. A false positive response rate of 8% and a false positive progression rate of 19% were found. Others have confirmed these observations[15,16] and suggested that other factors, such as size and number of lesions being assessed, influence response assessment. Radiological response assessment has also been studied. For instance, assessment of lesions seen on CXRs show an error rate that is inversely proportional to the size and number of lesions, and intra-observer variability[16-18] is less than inter-observer variability. Similar results have been obtained where tumors have been assessed by CT.[19]

Table 3. Number of patients assessed and rejected by main assessment methods

	CT	CXR	PE	Total
Total assessed	81 (71%)	27 (24%)	6 (5%)	114
No. rejected	30 (37%)	5 (19%)	0[a]	35 (31%)

[a]Patients assessed by PE only could not be objectively reviewed.

Failure to accurately assess response rates in early pivotal clinical trials can lead to artificially high response rates and thus unwarranted development of inactive agents with large numbers of patients being exposed to these inactive and often toxic agents.

In this study, the ORB comprised a minimum of three oncologists and one radiologist, all of international reputation and not otherwise involved with the study. The outcome of this review was that the claimed response rate across the four trials was reduced from 25-35 to 20-23%. Although the study in which gemcitabine was administered twice-weekly resulted in a similar response rate to the weekly schedules, it was associated with significantly more toxicity in the form of fever and flu-like symptoms.[10]

The rejection rate varied between 27 and 37%. The principal reasons for rejection were failure to achieve or maintain a greater than 50% reduction in total tumor product for at least 4 weeks and misinterpretation of normal structures or non-malignant pathology for tumor on scans and CXR. The consistency of validated response rate clearly demonstrates the value of an independent review. By obtaining a second opinion, inconsistencies in the application of response criteria were highlighted and an objective conclusion about the outcome of treatment applied. Furthermore, the results of subsequent larger-scale, multicenter, randomized trials[20] have produced comparable response rates, a finding which lends further support for the use of independent ORBs.

Cross-sectional imaging techniques should be employed and CT is the modality of choice, because it allows quantitative reappraisal by a third party at a later date. In these studies, CT was the most common imaging modality used and its use allowed a thorough review of the data. However, even with CT, difficulties can be encountered when defining lesions, unless standardized protocols are employed. Factors such as the inability to determine the outline of a lesion from an adjacent blood vessel when an i.v. contrast medium has not been administered can also affect the accuracy of tumor measurement or render potentially measurable lesions non-measurable. The use of an i.v. contrast agent could have defined the margins of the great vessels from paratracheal lymph nodes and so some lesions that could not be measured would have been measurable, resulting in the possibility of a higher overall response rate for these studies. The difficulties interpreting and therefore validating CT scans of this nature would be reduced if the administration of an i.v. contrast agent were to be standard procedure

in all oncology trials unless contraindicated for medical reasons.

In two of the studies, minimum sizes for measurable lesions were stipulated as 1×1 cm for CT or 2×2 cm for CXR. There were 14 patients in all studies with lesions smaller than these minimum sizes, although only two patients were in the studies where a minimum size was stipulated. Only four of these 14 patients had claimed responses, two of whom were classified as non-responders because they did not meet the protocol-stipulated requirements for measurable disease.

CXRs are adequate for lesions which are clearly delineated. However, in areas such as the paratracheal and subcarinal regions, the hila and those areas obscured by the diaphragm and cardiac contour, CXRs simply do not demonstrate lesions clearly, if at all, whereas cross-sectional imaging in the form of CT does so and hence is preferred. Standardization of procedures is necessary, e.g. a postero-anterior chest radiograph must be performed in full inspiration at a constant tube to film distance to prevent differences in the 'apparent' size of the lesion due to magnification factors. Should this occur, a meaningful comparison is not possible.

Finally, the review board was mandated to review only those patients in whom a response was claimed by the study investigator. It may be useful in future to review patients with stable disease to ensure that no partial responders are overlooked. Furthermore, six claimed responders were found on review to have non-measurable disease, and so it may be prudent to review all patients at baseline to ensure they have measurable disease, thus enabling consistency between claimed responders and non-responders for measurability.

Response rates therefore can and do play a vital role in the further development of new active oncological agents, and by the same token can reject inactive agents at an early stage. One study showed only 'fair' agreement between readers as to whether a response had occurred on CXR after treatment of small cell lung cancer (SCLC), but no mention was made as to whether radiologists or oncologists viewed the scans or how experienced the readers were.[21] However, this suggests the application of uniform response criteria—a situation that has yet to be attained. Ideally, the same response criteria should be applied to all studies to allow a fair comparison.[22] Response criteria were initially defined by several cooperative groups as far back as the mid-1970s and the response criteria were updated in 1981.[23] Different cooperative groups have further amended these criteria up to the present day, although none have been universally accepted. Davis *et al.*[24] compared the evaluation criteria of four cooperative groups, where the criteria for complete response and partial response were standard but subtle differences existed in lesion assessment. The calculated response rates for the same data using the different criteria ranged from 38 to 52%.

Response criteria are by no means uniform. The WHO, Southwestern Oncology Group (SWOG) and Eastern Cooperative Oncology Group (ECOG) are probably the most commonly applied criteria, but all are different. Study protocols often try to 'improve' these criteria by modifying them. In order to allow for true comparison of results between different agents, and in particular for studies where new combination therapy is being investigated, a universally agreed set of response criteria should be developed and implemented. Taken together with independent review to ensure that the criteria are consistently applied, this could result in robust, reproducible response rates, and, therefore, accelerated development of active new anti-cancer agents and early rejection of inactive agents.

Conclusion

Independent review provided convincing data supporting the activity of gemcitabine against NSCLC. By use of an independent ORB, response criteria were applied uniformly, with greater consistency, and between-study reproducibility was achieved. These findings commend the use of independent ORBs in the drug development process and support the EC CPMP guideline advocating the use of independent review boards in the development of new anti-cancer agents.

References

1. Zelen M. Guidelines for publishing papers on cancer clinical trials: responsibilities of editors and authors. *J Clin Oncol* 1983; **1**: 164-9.
2. Simon R, Wittes RE. Methodologic guidelines for reports of clinical trials. *Cancer Treat Rep* 1985; **69**: 1-3.
3. Fumoleau P, Chevalier B, Kerbrat P, *et al.* Current status of Taxotere® (docetaxel) as a new treatment in breast cancer. *Breast Cancer Res Treat* 1995; **33**: 39-46.
4. Trudeau ME, Eisenhauer EA, Higgins BP, *et al.* Docetaxel in patients with metastatic breast cancer: a phase II study of the National Cancer Institute of Canada—Clinical Trials Group. *J Clin Oncol* 1996; **14**: 422-8.
5. Gwyther S, Bolis G, Gore M, *et al.* Experience with independent radiological review during a topotecan trial in ovarian cancer. *Ann Oncol* 1997; **8**: 463-8.
6. Notice for guidance on evaluation of anticancer medicinal products in man. CPMP/EWP/205/95; adopted December 1996.

7. Food and Drug Administration. New drug, antibiotics and biological drug product regulations: accelerated approval final rule. *Federal Register* 11 December 1992; **57**: 58942-60.
8. Anderson H, Lund F, Bach N, Thatcher N, Walling J, Hansen HH. Single-agent activity of weekly gemcitabine in advanced non-small cell lung cancer: a phase II study. *J Clin Oncol* 1994; **12**: 1821-6.
9. Abratt RP, Bezwoda WR, Falkson G, Goedhals L, Hacking D, Rugg TA. Efficacy and safety profile of gemcitabine in non-small cell lung cancer: a phase II study. *J Clin Oncol* 1994; **12**: 1535-40.
10. Lund B, Ryberg LM, Meidahl-Petersen P, Anderson H, Thatcher N, Dombernowsky P. Phase II study of gemcitabine (2′,2′-difluorodeoxycytidine) given as a twice weekly schedule to previously untreated patients with non-small cell lung cancer. *Ann Oncol* 1994; **5**: 852-3.
11. Gatzemeier U, Shepherd FA, Le Chevalier T, *et al.* Activity of gemcitabine in patients with non-small cell lung cancer: a multicenter, extended phase II study. *Eur J Cancer* 1996; **32A**: 243-8.
12. FDA Proposals for Reform. *Reinventing the regulation of cancer drugs*. March 1996.
13. *WHO handbook for reporting results of cancer treatment*. Geneva: WHO 1979.
14. Moertel CG, Hanley JA. The effect of measuring error on the results of therapeutic trials in advanced cancer. *Cancer* 1976; **38**: 388-94.
15. Lavin PT, Flowerdew G. Studies in variation associated with the measurement of solid tumors. *Cancer* 1980; **46**: 1286-90.
16. Warr D, McKinney S, Tannock I. Influence of measurement error on assessment of response to anticancer chemotherapy: proposal for new criteria of tumor response. *J Clin Oncol* 1984; **2**: 1040-6.
17. Herschorn S, Hanley JA, Wolkove N, *et al.* Measurability of non-small-cell lung cancer on chest radiographs. *J Clin Oncol* 1986; **4**: 1184-90.
18. Gurland J, Johnson RO. How reliable are tumor measurements? *J Am Med Ass* 1965; **194**: 125-30.
19. Hopper KD, Kasales CJ, Van Slyke MA, Schwartz TA, TenHave TR, Jozefiak JA. Analysis of interobserver and intraobserver variability in CT tumor measurements. *Am J Roentgenol* 1996; **167**: 851-4.
20. Eli Lilly and Co. Data on file.
21. Quoix E, Wolkove N, Hanley J, Kreisman H. Problems in radiographic estimation of response to chemotherapy and radiotherapy in small cell lung cancer. *Cancer* 1988; **62**: 489-93.
22. Grossman SA, Burch PA. Quantitation of tumor response to antineoplastic therapy. *Semin Oncol* 1988; **15**: 441-54.
23. Miller AB, Hoogstraten B, Staquet M, Winkler A. Reporting results of cancer treatment. *Cancer* 1981; **47**: 207-14.
24. Davis HL, Multhauf P, Klotz J. Comparisons of cooperative group evaluation criteria for multiple-drug therapy for breast cancer. *Cancer Treat Rep* 1980; **64**: 507-17.

(Received 26 June 1999; accepted 2 July 1999)