Anti-Cancer Drugs 2000, **11,** pp. 433–437

Review paper

Radiologist review versus group peer review of claimed responses in a phase II study on high-dose ifosfamide in advanced soft tissue sarcomas of the adult: a study of the EORTC Soft Tissue and Bone Sarcoma Group

Stephen J Gwyther,[1] Ole S Nielsen,[2] Ian R Judson,[3] Martine van Glabbeke[4] and Jaap Verweij[5]

[1]East Surrey Hospital, Surrey and Sussex NHS Trust, Canada Avenue, Redhill, Surrey, RH1 5RH, UK. [2]Aarhus University Hospital, Norrebrogade 44, 800 Aarhus C, Denmark. [3]CRC Centre for Cancer Therapeutics, Institute of Cancer Research, E Block, 15 Cotswold Road, Sutton, Surrey SM2 5NG, UK. [4]EORTC Data Center, avenue E Mounier 83 Bte 11, 1200 Brussels, Belgium. [5]Rotterdam Cancer Institute (Daniel den Hoed Kliniek) and University Hospital, PO Box 5201, 3008 AE Rotterdam, The Netherlands.

The Soft Tissue and Bone Sarcoma Group (STBSG) of the EORTC ran a phase II study to assess the therapeutic activity of high-dose ifosfamide in patients with advanced soft tissue sarcomas by means of response rate (RR). Investigators claiming a response submitted the relevant chest radiographs (CXR) or scans to two other members of the STBSG for peer review. The reviewers completed a questionnaire indicating overall response or reasons for rejecting the claimed responses. An independent radiologist also reviewed the cases and he was blinded to the results of the peer review until the study was concluded. Twenty-two patients were reviewed by the radiologist and peer review, and the completed questionnaires were retrospectively reviewed. Two differences were noted, one partial responder (PR) was regarded as stable disease by the radiologist and one PR by peer review was determined a complete response by the radiologist. The radiologist found subsequent evidence of progressive disease in three patients who initially showed a PR, whilst the review group noted only one. This study suggests peer review in this tumor type is a satisfactory method of achieving an accurate, objective RR.

Key words: Claimed responses, independent radiological review, peer review, response rate.

Correspondence to SJ Gwyther, East Surrey Hospital, Surrey and Sussex NHS Trust, Canada Avenue, Redhill, Surrey RH1 5RH, UK.
Tel: (+44) 1737 231601; Fax: (+44) 1737 231783;
E-mail: GwytherSJ@aol.com

Introduction

A reliable, accurate and objective indicator reflecting clinical benefit and survival to new chemotherapeutic agents during an ongoing study is highly desirable.

Response rates (RR) using radiological techniques are frequently utilized as a major end point in phase II studies and act as a surrogate for clinical benefit, to help determine whether potential new agents show activity. If activity is demonstrated then further large-scale phase III studies can be undertaken using traditional end points such as survival to determine patient benefit. For optimal objectivity and to allow comparability between studies, its is important to apply a uniform set of response criteria. Independent review of claimed responses to confirm response increases objectivity and reduces potential bias, adding to the scientific credibility, and increases confidence in the process.[1] It may also reduce the variability between RR and survival statistics. The Soft Tissue and Bone Sarcoma Group (STBSG) frequently use RR as an end point in phase II studies and peer review of claimed responses by two other members of the STBSG is undertaken. To determine whether peer review is acceptable to accurately confirm claimed responders or whether serious flaws occurred, a radiologist independent from the study group also reviewed claimed responders, but was blinded to the results of the group peer review. We report the findings of this study.

Patients and methods

Eligibility criteria

Patients aged between 16 and 65 years were required to have histological proof of soft tissue sarcoma, evidence of progressive disease within 6 weeks prior to treatment and at least one clearly defined bidimensionally measurable lesion, which was required to be at least 2.0 cm diameter in the lungs and 2.5 cm diameter in the skin, s.c. tissue, lymph nodes, mediastinum, retroperitoneum and visceral organs. Lung metastases could be measured by chest X-ray (CXR) or computed tomography (CT), liver metastases by ultrasound or CT and all other lesions by CT. Magnetic resonance imaging (MRI) was also an acceptable imaging modality. Bone lesions, hepatomegaly, lymphodema, ascites and pleural effusions were considered non-measurable.

Other criteria included World Health Organization (WHO) performance status 0–2, adequate hepatic, renal and bone marrow function, no more than one prior combination chemotherapy or two single-agent regimens, and informed consent. Patients who had a second primary malignancy, those who had undergone radiotherapy to the sole measurable lesion and those for whom regular follow-up was impractical were excluded. Ethical approval for the study was obtained.

Study design

An open, non-randomized, multicenter phase II study using high-dose ifosfamide, given with mesna at a dose of 12 mg/m^2 as a continuous infusion over 3 days every 4 weeks, in advanced soft tissue sarcomas of the adult, both as first- and second-line chemotherapy, with appropriate stratification.[2]

The main aim of the phase II study was to assess the therapeutic activity of high-dose ifosfamide by RR using the WHO response criteria.

Peer review by two members of the group, other than the investigator claiming the response, was undertaken during one of the bi-annual meetings and the scans or CXR re-assessed. The reviewers completed an EORTC response assessment form, indicating date of review, overall response, site(s) of response, date of first response, date of first evidence of progression or reasons why the patient was not considered to have responded and the names of the reviewers. These forms were returned to the EORTC data center for entry into the data base and subsequent analysis.

In order to be validated as a response, it was necessary for the investigator and the two peer reviewers to unanimously agree, otherwise the worst response determined by peer review was assigned.

The aim of this particular facet of the study was to determine correlation between the radiologist and peer review group, to ensure the group review process provided a similar RR to the radiologist or whether a radiologist should be included in the review process in soft tissue sarcomas. The films were made available to the radiologist, so he could assess the films and record his results separately, but he was blinded from the results of the peer review. Similarly the investigators and group were blinded to the response evaluation determined by the radiologist.

Results

The radiologist reviewed scans and radiographs on a total of 22 patients claimed as responders in this study. The results are summarized in Table 1.

Of the 22 claimed responses, the radiologist and the peer review group rejected six (27%) and five (23%), respectively, because of either failing to meet the criteria of partial response (PR) or being considered non-evaluable.

The peer review group determined 17 responses and the radiologist 16. Two discrepancies were seen, one patient who was validated a PR by peer review was considered a complete response (CR) by the radiologist. The other, a PR by peer review was assigned stable disease (SD) by the radiologist.

The first patient had disease in the neck which almost disappeared with treatment, but some residual tissue remained. Peer review regarded this as representing residual active disease, whereas the radiologist considered this to represent normal tissue (Figures 1–3).

The other case was rejected by the radiologist because although there was a reduction in disease during treatment, the sum of the three individual lesions did not decrease by more than 50% even though the tumor product of two of the three lesions did so.

The radiologist noted three patients who responded subsequently developed progressive disease (PD),

Table 1.

	CR	PR	SD	PD	NE[a]	Total
Radiologist	1	15	4	0	2	22
Peer review	0	17	3	0	2	22

[a]Not evaluable.

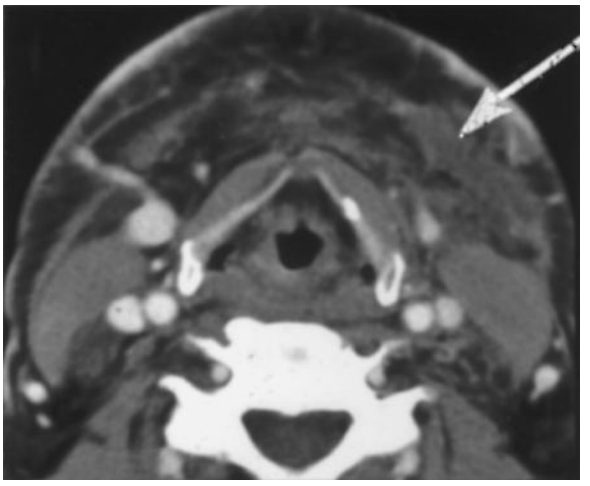

Figure 1. Contrast enhanced axial CT scan of the neck at the level of the thyrohyoid cartilage, performed prior to treatment. There is a bi-dimensionally measurable mass arising from the left platysma muscle (arrow).

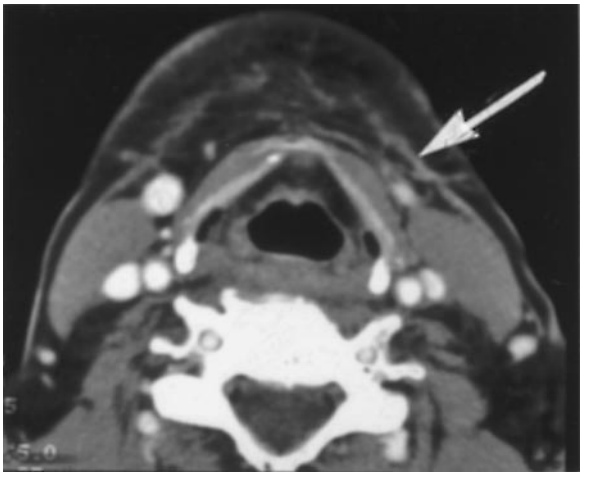

Figure 2. Corresponding CT scan performed after two courses of ifosfamide. The mass arising from the left platysma muscle is now very much smaller (arrow).

although the review group determined only one. In the latter case, the radiologist assigned PD at the examination prior to that assigned by the review group. In one patient where the radiologist recorded evidence of new disease, the new metastasis was seen abutting the pleura, although the single measurable lesion had increased in size significantly. The other patient had three measurable lesions, all of which progressively decreased in size, but on the final examination, a new metastasis was apparent in the right lung.

The peer review group validated a total of 18 responses of the 114 eligible patients, giving a RR of 16% in the overall study. The scans from one patient,

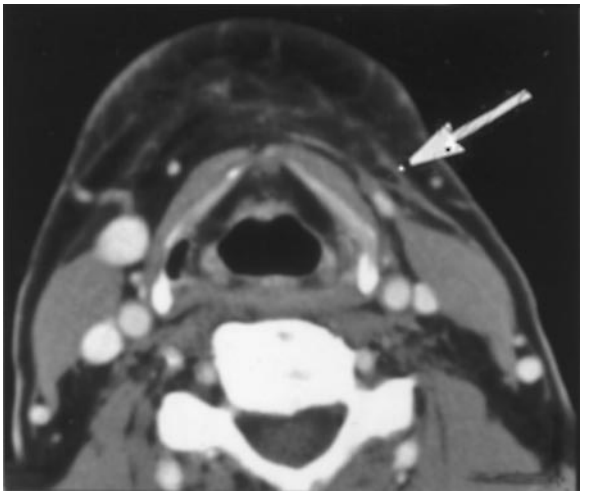

Figure 3. Corresponding CT scan performed after four courses of ifosfamide. The mass has now completely disappeared, the only remaining tissue is the platysma muscle (arrow).

previously assigned a PR by the review group, were not available for the radiologist to review, so this patient has been ommitted from this facet of the study. However, there was very good correlation of responses determined by the review group and the radiologist, 17 by the review group and 16 by the radiologist.

The number of patients involved in this study was too small to enable a meaningful statistical analysis to be undertaken.

Discussion

RR are frequently used as a measure of activity of a drug in new drug development in oncological phase II studies. A direct comparison with historical studies is difficult for several reasons. Firstly, advances in technology and new techniques over the last 25 years preclude a direct comparison, e.g. the CXR cannot be compared to a CT scan of the thorax.[1] Secondly, earlier studies compared response of the agent under study with historical controls using either existing therapy or best supportive care, where no beneficial treatment was available. Matters were compounded when the patient population studied was not clearly defined, and factors such as response to prior first-line therapy was not stated and claimed responders were not independently validated.[3–5]

The active treatment of advanced disease by chemotherapeutic agents and the large number of potential new agents available for study demands the use of a surrogate marker that can be reliably used to

indicate activity of potential new agents, thereby allowing the further development of active, beneficial agents whilst inactive and toxic agents which confer no patient benefit can be discarded at an earlier stage. This does not replace survival as the ultimate end point, but as yet there is no consistently reliable, accurate surrogate marker.[1] Quality of life studies are important but are necessarily subjective. The patient is likely to underestimate the severity of symptoms if this will help keep him in a study[1,6] when no alternative treatment is available. Some tumor types are easier to measure than others. Sarcomas tend to produce large, well-circumscribed metastases, which are usually easily visible. The use of i.v. contrast agents increases the tissue coefficient of vascular structures, so the borders of the matastases are better defined when lesions such as para-aortic lymph nodes abut adjacent normal vascular structures. Similarly, the use of i.v. contrast agents enable metastases in the liver and spleen to be better defined. Oral contrast agents are also used to delineate the bowel and avoid normal loops of bowel being misinterpreted as representing tumor. However, despite our best efforts, some disease is truly non-measurable using current imaging techniques.[1] Independently validated data adds credibility and reliability to the study because it shows reproducibility, although it decreases the claimed response rate by up to 30%.[7–13] The reduction in claimed RR is mainly due to misinterpretation of normal structures for active tumor and errors measuring lesion size. Many of these errors could be reduced by the introduction of uniform imaging protocols, both within and between trials.[1] Even so, some tumor types such as epithelial ovarian carcinoma are often difficult to assess radiologically, even when it is clinically apparent the patient is relapsing.

A likely explanation why small, single center studies often report higher response rates than multi-center studies is the lack of a review process, where the data have not been scrutinized objectively by experienced investigators.

In this study, there was very good agreement between radiological and peer review. Interestingly, there was a slightly greater discrepancy between the radiologist and peer review in determining PD in those who had already responded. Although the numbers are small, the radiologist noted three cases and the review group one. In two cases, the disease was noted on the most recent examination and no further examinations were included. In the one case where both groups determined PD, the radiologist assigned PD at the examination prior to that by the review group. This may be significant in studies where time to progression rather than RR is the primary end point. This does not necessarily imply radiological detection reliably demonstrates PD before clinical examination, because the time interval between examinations is also important. To prove this hypothesis, large studies using standardized radiological protocols and a constant time interval between radiological examinations would be necessary.

Ideally, review should include both radiologists and oncologists with a broad knowledge of both the clinical and radiological features of that particular tumor type.

Conclusion

Peer review by oncologists and a radiologist experienced in the clinical and radiological presentations of soft tissue sarcoma demonstrated a similar RR. The radiologist did, however, identify PD in established responders before the review group, suggesting that a radiologist should be involved when assessing time to progression, although not necessarily when assessing RR as the primary end point.

The need for uniform radiological protocols is important to make meaningful comparisons both within and between studies.

References

1. Gwyther SJ. Response assessment using radiological methods. *Crit Rev Oncol/Hematol* 1999; **30**: 45–62.
2. Nielsen OS, Judson I, van Hoesel Q, *et al*. Effect of high dose ifosfamide in advanced soft tissue sarcomas. A multicenter phase II study of the EORTC soft tissue and bone sarcoma group. *Eur J Cancer*, in press.
3. Markman M, Hakes T, Reichman B, *et al*. Ifosfamide and mesna in previously treated advanced epithelial ovarian cancer; activity in platinum-resistant disease. *J Clin Oncol* 1992; **10**: 243–8.
4. Hoskins PJ, Swenerton KD. Oral etoposide is active against platinum resistant epithelial ovarian cancer. *J Clin Oncol* 1994; **12**: 60–3.
5. Seymour MT, Mansi JL, Gallagher CJ, *et al*. Protracted oral etoposide in epithelial ovarian cancer: a phase II study in patients with relapsed or platinum resistant disease. *Br J Cancer* 1994; **69**: 191–5.
6. Slevin ML, Stubbs L, Plant HL, *et al*. Attitudes to chemotherapy: comparing views of patients with cancer with those of doctors, nurses and general public. *Br Med J* 1990; **300**: 1458–60.
7. Gwyther SJ, Bolis G, Gore ME, ten Bokkel Huinink WW, *et al*. Experience with independent radiological review during a topotecan trial in ovarian cancer. *Ann Oncol* 1997; **8**: 463–8.
8. Gwyther SJ. Imaging in lung cancer. *Lung Cancer Ther* 1995; **10**: 301–5.

9. 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.
10. Carmichael J, Possinger K, Phillip P, *et al*. Advanced breast cancer: a phase II trial with gemcitabine. *J Clin Oncol* 1995; **13**: 2731–6.
11. Carmichael J, Fink U, Russell RCG, *et al*. A phase II study of gemcitabine in patients with advanced pancreatic cancer. *Br J Cancer* 1996; **73**: 101–5.
12. Shepherd FA. Phase II trials of single-agent activity of gemcitabine in patients with advanced non-small cell lung cancer: an overview. *Anti-Cancer Drugs* 1995; **6** (suppl 6): 19–25.
13. Gwyther SJ, Aapro MS, Hatty SR, *et al*. Results of an independent oncology review board of pivital clinical trials of gemcitabine in non-small cell lung cancer. *Anti-Cancer Drugs* 1999; **10**: 693–8.
14. Trimble EL, Adams JD, Vena D, *et al*. Paclitaxel for platinum refactory ovarian cancer: Results from the first 1000 patients registered to National Cancer Institute Treatment Referral Center 9103. *J Clin Oncol* 1993; **11**: 2405–10.
15. Eisenhauer EA, ten Bokkel Huinink WW, Swenerton KD, *et al*. European–Canadian randomized trial of paclitaxel in relapsed ovarian cancer: High dose versus low-dose and long versus short infusion. *J Clin Oncol* 1994; **12**: 2654–66.

(Received 26 March 2000; accepted 6 April 2000)