Pergamon

European Journal of Cancer, Vol. 33, Suppl. 2, pp. S3–S7, 1997
© 1997 Elsevier Science Ltd. All rights reserved
Printed in Great Britain
0959–8049/97 $17.00+0.00

PII: S0959-8049(96)00446-7

A Call for Change in Anticancer Drug Evaluation

E. Cvitkovic

Service des Maladies Sanguines et Tumorales, Hôpital Paul Brousse, 12–14 avenue Paul Vaillant Couturier, 94800 Villejuif, France

In the evaluation of anticancer agents, the present emphasis on the use of measurements determining the drug's capacity for cytoxicity and tumour shrinking prevents oncologists from defining efficacy measures which may be more relevant to clinical reality. In addition, oncologists are not yet making adequate or appropriate use of the advanced technologies available, and the use of standardised response criteria does not ensure that results are clinically meaningful. There is a strong need for flexibility in the choice of the endpoint to be measured according to the type of cancer. It is important that faster, smaller-scale trial designs must be developed so that promising, new anticancer agents are made available as quickly as possible. © **1997 Elsevier Science Ltd. All rights reserved.**

Key words: therapeutic evaluation, response rate, methodology
Eur J Cancer, Vol. 33, Suppl. 2, pp. S3–S7, 1997

INTRODUCTION

UP UNTIL a few years ago, oncologists suffered a long period devoid of new candidates for anticancer drug development. More recently, many new therapeutic possibilities, including taxoids, topoisomerase I inhibitors and new antimetabolites have been discovered. Many other agents are also on the horizon. An overhaul of the current approval criteria for new agents is now essential, as anticancer agents occupy a unique position among medicines (Table 1). These drugs are developed and evolve in their therapeutic potential over many decades, most of that time being after registration. The real merits of such drugs are found only when the agents have been used to treat many different cancers, in many different combinations and in a wide variety of temporal interactions with other treatments. It takes time for the drug's dosage and administration schedule to be optimised, taking into account the toxicity in both acute and chronic use. Similarly, it is only with time that it may be established, for example, that a drug's usefulness is limited in the palliative, pretreated patient setting, yet becomes much greater in curative, neoadjuvant or adjuvant use.

With such considerations in mind, there is an urgent need for new anticancer agents to be recognised as active and to be made available as quickly as possible via an accelerated process of clinical research and registration. The need to improve survival in a multicentre, randomised trial in order to have access to a new, active anticancer agent should be questioned. If today's labyrinthine requirements had been in place in the past, it is highly probable that existing compounds, such as 5-fluorouracil (5-FU), would not be available today. 5-FU has a single agent bolus activity of $\leq$ 15% and no demonstrated advantage in terms of time to progression or survival in multicentre, large-eligibility, controlled trials in the advanced disease setting. This drug would, therefore, be unlikely to obtain present-day registration and marketing approval. Yet, its role in the clinic is both established and expanding; indeed, the modulation of 5-FU with other agents or altered administration schedules, and a practical application if its pharmaco-dynamic profile, is still ongoing today.

That the assessment methods which are currently used in clinical trials have rendered some useful service in the past remains without question, but it may be argued that these methods are now obsolete and impractical. For example, the rule that a new agent must prove superior in efficacy to previous ones is logical, but may potentially lead to sophistry. Poor treatment outcomes are still associated with many cancers. In this respect, current methods of assessing the efficacy of anticancer drugs have, indeed, failed, and a major rethink of drug development methodology is required.

An identifiable reason for the present impasse lies in the current oncological paradigm, which takes as read that demonstrable cytotoxicity and tumour shrinking are essential properties of any 'useful' anticancer agent. At the same time, there

Table 1. Special characteristics of anticancer agents

1. Their full development takes place after registration, often over decades (e.g. 5-fluorouracil, ifosfamide, anthracyclines).
2. Their full potential lies in their capacity for combination with other agents and other treatment modalities.
3. Their activity may vary at different periods of the disease's natural history (e.g. adjuvant versus neoadjuvant use, first-line palliation, salvage).
4. They all have moderate or severe, acute and/or chronic side-effects and toxicities.

Table 2. When is a cancer treatment worthwhile?

- What does it mean?
- Does it prolong life?
- Does it induce tumour regression?
- Does it slow the cancer's progression?
- If so, in what specific clinical context?
- Is it better tolerated than other treatments?
- Is it easier to use than other treatments?
- Is is less expensive than other treatments?
- Does its use make life worth living?

has been a failure to identify and define relevant clinical models that may help to detect significant antitumour activity. These aspects of the problem are now addressed in more detail.

WHAT DO WE MEAN BY 'USEFUL' ANTICANCER ACTIVITY?

We should not be restrictive in our definition of potentially 'useful' anticancer activity. It is short-sighted to limit the definition to include only the effects of a drug on the tumour and ignore those on the patient. The definition of an effective anticancer agent, first proposed in 1949, is given below [1]:

1. 'An agent which does not produce an appreciable objective clinical improvement . . . cannot be expected to prolong life . . .
2. An active agent may be expected to show a range of therapeutic efficacy . . .
3. An effective chemotherapeutic agent should induce consistent and predictable improvement . . .
4. The toxicity and hazards of administration . . . must be taken into consideration . . .
5. . . . since different agents are partially effective . . . their combination should enhance the effect.'

Much of this definition is still valid today, particularly the point that an anticancer agent must produce positive change in the clinical status of the patient if it is to be expected to prolong life. More recent analysis in the same tradition [2, 3] has helped to identify a series of questions which are useful as a blueprint for determining the overall usefulness of cancer treatments, in general, not just of anticancer agents (Table 2).

THE NEED FOR RELEVANT CLINICAL MODELS

Historically, there have been three distinct epochs in the methodology of anticancer drug evaluation.

- From 1948 to 1960, the dominant methodology used was that published by Karnofsky [1].
- From 1960 to 1977, evaluation techniques were modelled after Zubrod [4].
- In recent years, the established 'rules' for measuring the effectiveness of anticancer agents were those derived originally from the International Union Against Cancer (UICC) response criteria for breast cancer of 1977 [5]. These guidelines were 17 years old in 1994 and, since 1981, have formed much of the basis for the standardised approaches to the recording of patient data and the results of therapy recommended by the World Health Organisation (WHO) [6]. These same WHO/UICC guidelines are almost unchanged today.

In this context, today's oncologists are not making adequate or appropriate use of the new technologies and evaluation methods that are becoming available. When the WHO guidelines were developed, clinical examination, scintigraphy, simple echography and standard X-rays were the only measurement tools available. Today's methods include computed tomography (CT), nuclear magnetic resonance (NMR) imaging, intractivity sonography and, in the near future, positron emission tomography (PET), together with reliable tumour markers, such as prostate-specific antigen and other serum markers (e.g. CA 12.5, CA 15.3, CA 19.9, etc.). Yet, these methods have not been used over the years to re-think the evaluation criteria, only to define the existing response criteria even more restrictively. We now count in millimetres, not centimetres, but we still measure the surface area of planar cuts in irregular masses by the product of the diameters.

Another long-established problem is that posed by the differences found in both inter-observer and intra-observer variation, found even when 'objective' measurement techniques are used [7]. For example, it has previously been shown that there is only a 50% chance of an observer identifying a 50% reduction of tumour area as a partial response, using the standard WHO/UICC criteria [8]. Even worse, there is a strong chance that a lesion that is actually stable will be identified as progressive, with the result that treatment may be changed inappropriately. For the same investigator examining the same patient, there is a 43% chance that a stable lesion will be labelled progressive by the fourth follow-up examination. If a different investigator examines the patient, this chance increases to 45% by the second examination and 68% by the fourth [8].

The 50% reduction rule has no real significance, even when it is established by such controversial methods as the product of diameters of measurable lesions or the multidimensional assessment of evaluable lesions. The likelihood of response is multifactorial, but the chance of obtaining a 50% reduction is volume dependent. In addition, the percentage of reduction occurs in a continuum. A recent paper by Sobrero and associates illustrates the range of percentage reductions in measurable masses (Figure 1) [9]. It is evident that there is really no difference between, for example, a 43% reduction or a 52% reduction, yet the former is classified as a non-response and the latter, a response.

As a further complication, not even the systematic standardisation of response criteria ensures that results are clinically meaningful. Rigid interpretation of the WHO/UICC criteria demands the classification of a breast cancer patient with liver and bone metastes that are unchanged or of unknown status, but who has clinically minor shrinking of her skin and lymph node lesions, as a 'partial response' (Figure 2). We readily acknowledge the limited relevance of such change to the outcome for the patient, yet we continue to use such assignations in reporting our results. In this respect, it is worth remembering that response criteria were originally defined as a common language reference, and were by no means intended to become dogmas. Indeed, the original publication from which the WHO guidelines were derived stresses natural history parameters and clinical context as being more important than measurements [6]; such points were also included in the initial Karnofsky tenets.

Thus, there is a need for flexibility in evaluation measures and criteria according to the type of cancer. For example, in relatively slow-growing tumours, such as colorectal cancer or hormone-sensitive breast cancer, it is now being shown that stabilisation of the disease (SD) is equivalent, in terms of

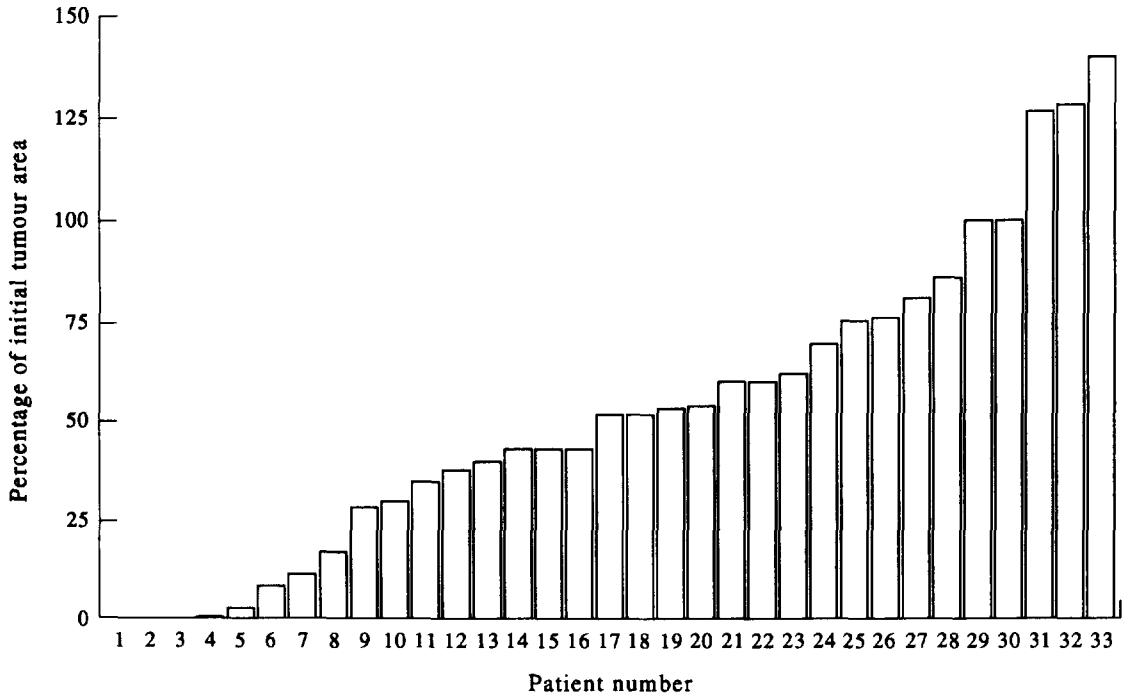


Figure 1. Tumour shrinkage as a continuum. The histogram shows the maximum percentage decrease in tumour area for each patient. Patients 1–3 showed complete regression, patients 4–16 had over 50% regression and patients 17–25 had 25–50% tumour regression. Reprinted by permission of the American Association for Cancer Research, Inc., from Sobrero AF *et al.*, *Clin Cancer Res* 1995, Vol. 1, pp. 955–960.

survival, to a partial response (PR) (Figure 3) [10]. In the context of first-line metastatic colon cancer, therefore, the stabilisation rate may be a clinically worthwhile measurement [10].

In contrast, the situation in rapidly growing disease, such as ovarian cancer, is different. For example, in a recent study in patients undergoing salvage therapy with paclitaxel, the regression rate, as measured by the marker serum CA 125, was the same for patients with stable disease as it was for those showing partial response, yet the actuarial survival of patients with stable disease more closely resembled that of patients whose disease progressed (Figure 4) [11]. In advanced ovarian cancer, therefore, the clinical relevance of disease stabilisation is, at best, limited.

Such findings may also serve as a warning about the value of markers, in general, particularly in measuring end-stage

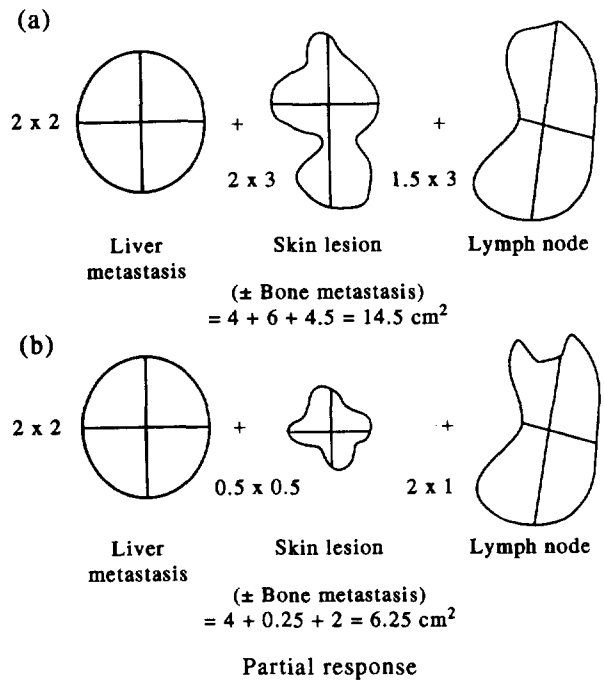


Figure 2. Objective response assessments of a patient with breast cancer (a) before treatment and (b) after treatment. The response was measured as the difference in summed cross-sectional lesion areas.

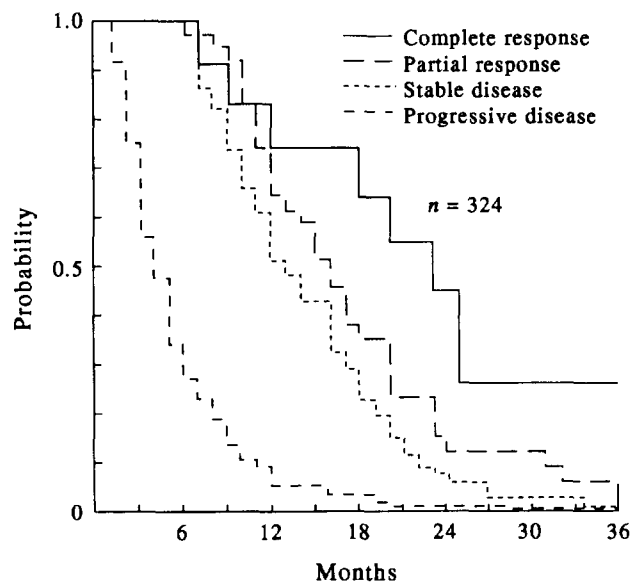
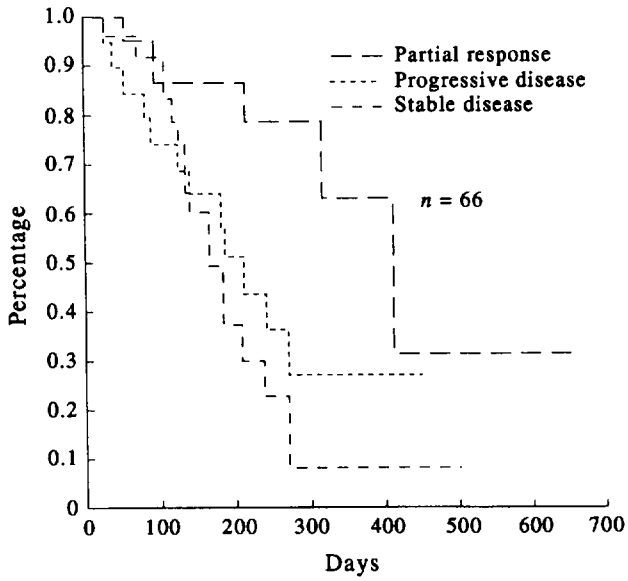


Figure 3. Probability of survival in colorectal cancer patients treated with either 5-FU alone or sequential methotrexate/5-FU/leucovorin according to response. Reprinted by permission of Macmillan Press Ltd, from Graf W *et al.*, *Br J Cancer* 1994, Vol. 70, pp. 559–563.

Table 3. Clinically relevant surrogate endpoints for therapeutic activity of a test agent used second-line or subsequently

To demonstrate activity in patients progressing while receiving standard best-choice agent combinations.	• **Response rate*** • Time to progression • Tumour markers
To demonstrate activity in the setting where no agents are very active and where the prognosis is short and ominous (e.g. liver or brain metastases).	• **Response rate** • **Time to progression**
To demonstrate activity in smouldering, difficult-to-evaluate disease.	• **Time to progression** • Response rate • Quality of life • Need for analgesics
To demonstrate durable activity in treatment-refractory chemosensitive tumours (e.g. choriocarcinoma, lymphoma, germ cell tumours).	• **Response rate** • **Duration of response**

* Response criteria to be defined *a priori* and *ad hoc*, not necessarily those used by the WHO-UICC. Main surrogate endpoints shown in bold.



Disease response	Alive	Dead	Total	Median
Progression	7	12	19	202
Stable	8	17	25	161
Partial response	16	6	22	410

Figure 4. Actuarial survival of paclitaxel-treated patients with ovarian cancer. Reproduced by permission of the Academic Press, from Pearl ML *et al.*, *Gynecol Oncol* 1994, Vol. 53, pp. 339–343.

therapy. Nonetheless, it remains important that markers are validated, and already they are viewed differently according to their context. It is, for example, readily acceptable that a fifth-therapeutic-line patient with a male germ cell tumour has received clinical benefit from an antitumoural effect on rapidly growing disease, when the proof of this is a 45-day duration of response and a 50% decrease in serum markers together with a 20% decrease in the product of diameter addition on two separate CT scans. It is less acceptable if the tumour is an ovarian cancer and not acceptable if it is prostrate cancer. I think we are right to be sceptical in these latter instances.

A more important matter is the lack of valid criteria for measuring dynamic changes in tumour progression. If a new agent could stop or slow down human tumour growth, or make the growth 'benign', the only currently available parameter to assess such an effect is the ill-defined time to tumour progression.

We accept biological response modifiers and hormones in such a context in controlled prospective trials, but there is little available in that respect in a phase II context. Therefore, there is a substantial risk of discarding 'inactive' drugs that may yet be used clinically.

WHAT PARAMETERS SHOULD WE MEASURE?

The volume and character of evidence required before it is accepted that an agent is useful therapy for a particular type of cancer depends on the situation being investigated. Table 3 summarises tentative proposals regarding clinically relevant surrogate endpoints that provide demonstrable evidence of the usefulness of a given test agent in a variety of clinical situations. In contrast, those measures I do not regard as acceptable include:

- Prognostic multivariate analyses that have not been validated in another patient cohort;
- outcomes scored on the basis of probable prognosis;
- surrogate endpoints that have not been validated externally;
- surrogate endpoints of valid surrogate endpoints.

CONCLUSION

Oncologists should strongly urge for the more rapid evaluation of new drugs using faster, small-scale trial designs. I contend, for example, that the liver metastasis specific activity of a promising new agent in breast cancer could usefully be assessed using only a small group of patients if, instead of 75-year-old women with metastatic bone disease, the patients enrolled were premenopausal, hormone receptor-negative, liver-metastatic breast cancer patients whose disease was anthracycline-refractory and rapidly progressive, or following failure of adjuvant therapy less than 6 months previously. If the results of such a trial were positive, this would demonstrate the usefulness of the drug in this respect using only 40 patients instead of the 600 that would otherwise be needed, with the added risk of diluting the beneficial effect of the drug in the heterogeneity of the large eligible population. Thus, I suggest that the only fixed rules in the evaluation of new anticancer drugs should be to always review and amend the existing fixed rules, and to avoid creating new ones that are not needed.

1. Karnofsky DA, Burchenal JH. The clinical evaluation of chemotherapeutic agents in cancer. In MacLeod CM, ed. *Chemother-*

apy of Cancer. New York, Columbia University Press, 1949, 191–205.
2. Tannock IF. Treating the patient, not just the cancer. *N Engl J Med* 1987, **317**, 1534–1535.
3. Tannock IF, Boyer, M. When is a cancer treatment worthwhile? *N Engl J Med* 1990, **323**, 989–990.
4. Zubrod CG, Schneiderman M, Frei E, *et al*. Appraisal of methods for the study of chemotherapy of cancer in man: comparative therapeutic trial of nitrogen mustard and triethylene thiophosphoramide. *J Chron Dis* 1960, **11**, 7–31.
5. Hayward JL, Rubens RD, Carbone PP, Heuson J-C, Kumaoka S, Segaloff A. Assessment of response to therapy in advanced breast cancer. *Br J Cancer* 1977, **35**, 292–298.
6. Miller AB, Hoogstraten B, Staquet M, Winkler A. Reporting results of cancer treatment. *Cancer* 1981, **47**, 207–214.
7. Gurland J, Johnson RO. How reliable are tumor measurements? *JAMA* 1965, **194**, 973–978.
8. Lavin P, Flowerdew D. Studies in variation associated with the measurement of solid tumors. *Cancer* 1980, **46**, 1286–1290.
9. Sobrero AF, Aschele C, Guglielmi AP, *et al*. Schedule-selective biochemical modulation of 5-fluorouracil: a phase II study in advanced colorectal cancer. *Clin Cancer Res* 1995, **1**, 955–960.
10. Graf W, Påhlman L, Bergström R, Glimelius B. The relationship between an objective response to chemotherapy and survival in advanced colorectal cancer. *Br J Cancer* 1994, **70**, 559–563.
11. Pearl ML, Yashar CM, Johnston CM, Reynolds RK, Roberts JA. Exponential regression of CA 125 during salvage treatment of ovarian cancer with taxol. *Gynecol Oncol* 1994, **53**, 339–343.