Factors influencing response rates for advanced colorectal cancer chemotherapy

R. Labianca, G. Pancera & G. Luporini
Division of Medical Oncology, S. Carlo Borromeo Hospital, Milan, Italy

Summary

Background: The activity of various chemotherapy regimens used in the treatment of advanced colorectal cancer is assessed by different groups of investigators and in various trials by what appear to be common criteria. However, there may be substantial inter-trial variation in the interpretation and application of these criteria which contributes to differences in response rates reported for the same regimen.

Materials and methods: This paper reviews the most prominent studies in this field and examines the factors which may influence the assessment of activity in clinical trials such as patient selection, the definition and application of response criteria, the methods of assessment of time to progression and duration of response, factors related to the therapeutic regimen and statistical methods. Each factor is critically discussed.

Results: The analysis confirms that there is a large variability among the different studies and that an inter-trial comparison is often impossible, with subsequent difficulties for clinicians in determining the true impact of therapies.

Discussion: After briefly commenting on the various issues, this review makes recommendations about how to achieve consistency among trials, for instance by using standard criteria, by extending the use of randomization even in phase II trials and by evaluating high quality, well conducted clinical trials in a meta-analysis, thereby making possible comparison across trials. The conclusions, although specific to colorectal cancer, are also applicable to other advanced malignancies.

Key words: advanced colorectal cancer, chemotherapy, evaluation criteria, tumor response

Introduction

An accurate evaluation of objective response has traditionally been considered of primary importance in advanced cancer chemotherapy. A Working Group established by the American Society of Clinical Oncology (ASCO) recently re-defined by consensus the outcomes of cancer treatment [1] and emphasized the importance of improvement in overall survival and quality of life, both considered as 'patient outcomes', while tumor response was classified as a 'cancer outcome' and its value was restricted to its ability to predict the patient outcomes and to influence decisions about treatment. Of particular importance is the achievement of complete response, because it often predicts the potential of a regimen to improve survival. While response rate is an appropriate endpoint for phase II trials, in the phase III setting survival and quality of life should receive the greater amount of attention.

Therefore, the results of clinical trials in which the response rate is reported (not only phase II but also most phase III studies) significantly influence the opinions of medical oncologists, both in clinical practice and in the designing of innovative or confirmatory trials.

In the absence of a single standardized classification system, several criteria for reporting objective response are available, of which the most frequently used are the ones developed by the WHO [2] and UICC [3]. These two classification systems are similar and only patients who achieve a complete or partial response are considered to have responded objectively to treatment.

In advanced colorectal cancer nearly all clinical reports concerning chemotherapy include a declaration confirming the use of WHO or UICC criteria and some studies, chiefly those conducted in North America, also include a clinical evaluation of metastatic hepatomegaly according to the criteria established by Moertel [4]. In contrast, reduction of tumor markers, for example, carcinoembryonic antigen (CEA) and gastrointestinal cancer antigen (GICA) alone, even if considerable, is not used to assess a positive response to treatment.

Despite this apparent methodological consistency, it is a common observation that different papers report a
widely different, response rates for the same therapeutic regimen, with the most notable differences reported for the biochemical modulation of 5-fluorouracil (5-FU) by folinic acid (FA) or alpha-interferon (α-IFN). For instance, in phase III studies, biochemical modulation of 5-FU by FA has achieved response rates varying from 15% to 48% and a subsequent meta-analysis [5] reported an overall response rate of 23%. The optimal administration schedule of 5-FU plus FA is still not known. Combination regimens utilising low doses of FA have attracted particular interest following the report of response rates ranging from 31% to 43% by the North Central Cancer Treatment Group (NCCTG) [4, 6, 7]. However, these results were not reproducible in very recent trials comparing a daily regimen of 5-FU plus low-dose FA for 5 consecutive days with raltitrexed, a direct and specific thymidylate synthase inhibitor [8], high-dose levorotatory FA plus 5-FU [9] and a high dose-intensity regimen of FA plus 5-FU [10]; in all of these trials the response rates were between 10% and 17%. The final results of a large Southwest Oncology Group (SWOG) trial were recently published [11]. This study evaluated seven different 5-FU (± biochemical modulation) regimens, and using the SWOG criteria [12] the authors observed that a substantial percentage of patients (19%) were not fully assessable for response. The remission rates ranged from 13% to 24% (confirmed responses only, in which the initial assessment was confirmed by a second test at >/- 4 weeks) or from 15% to 29% (all responses), again indicating that responses are limited following bolus administration of 5-FU even if modulation with FA is employed.

Further inconsistencies occurred in the evaluation of 5-FU plus α-IFN. The high response rate (>60%) originally reported by Wadler et al. [13] was not confirmed by phase II studies performed in other institutions [14, 15], even though a confirmatory ECOG trial [16] obtained results similar to those of the original study. Subsequent phase III studies failed to demonstrate a significant difference between the combination of the two drugs and 5-FU alone [17] and the activity related to the addition of α-IFN was much lower than that reported in the first studies. The same trend was observed when 5-FU plus α-IFN ± FA was compared with 5-FU plus FA [18, 19]. The long-term continuous infusion of 5-FU was also reported by Lokich et al. to evince a response rate of approximately 30% in phase II and then III trials [20], but demonstrated more limited activity when included in a cooperative evaluation [11]. Earlier studies comparing the efficacy of 5-FU alone or in combination with nitrosoureas also had encountered this problem. Ten years ago Tonkin [21] reviewed a large series of papers published in major oncology journals of which twelve pertained to colorectal cancer. It was concluded that no article included all of the information needed to precisely define the criteria used for the assessment of tumor response and that this inadequacy, together with the variability in response criteria, strongly contributed to the large differences observed in response rate.

The purpose of this article is to review the possible reasons for this discrepancy and to make recommendations about how to achieve consistency among trials, and thereby permit meaningful inter-trial comparisons. For the reasons explained above, the discussion will concern both phase II and III trials.

Factors with a potential impact on response assessment

1. Patient selection

Studies which appear similar because they employ identical treatment regimens can actually have significant differences with respect to patient characteristics, many of which can have an impact on prognosis or influence drug tolerance.

Age

There is an increasing trend not to impose a strict age limit on clinical trials in advanced colorectal cancer and this tends to result in the inclusion of a large proportion of older patients (>70 years) in trials not specifically designed for this age group. This can necessitate a reduction in drug dosage which could have a negative impact on the clinical activity of the treatment. It is indeed a common observation that the modifications of drug pharmacokinetics and pharmacodynamics induced by aging itself and by the presence of concomitant diseases can reduce the potential benefits and increase the complications of antitumor chemotherapy [22].

Performance status and other characteristics

PS is widely recognized as an important prognostic factor and, as a consequence, the distribution of patients according to performance status can influence response to treatment and even survival [23]. However, most studies include only patients with a performance status of ECOG 0 to 2 and the percentage of patients with a score of 2 is usually limited. Because patients with a seriously impaired performance status have a poor prognosis, they should not be entered into comparative studies due to the lack of activity of standard chemotherapy regimens in this setting. Instead novel treatment regimens or schedules could be developed for this patient group.

It has been recognized [23] that PS alone may not be sufficient to give a prognostic characterization of patients with advanced colorectal cancer: haemoglobin levels, disease-free interval and proportion of symptomatic patients may also affect the results of a clinical trial and should be carefully considered.
Presence of measurable disease

The presence of measurable disease is mandatory for inclusion in phase II and in most phase III trials. However, patients without measurable disease are also included in some phase III studies, and symptomatic relief, time to progression and survival are employed as indicators of therapeutic activity in them. Some prominent US institutions and cooperative groups, for example, SWOG and NCCTG, use this method with particular consistency and include about 40% to 50% of patients without measurable disease in randomized studies [4, 6, 7]. Consequently, even though the evaluation of objective response in measurable patients is obviously unaffected, a possible survival advantage could be attributable to patients with less extensive disease who were included in the study. In our opinion, all phase III trials in advanced colorectal cancer should include only patients with measurable disease because: a) evaluation of objective response in a large number of cases can yield more precise information about the activity of a given therapeutic regimen; b) most patients with metastatic colorectal cancer actually present measurable lesions (e.g., liver, lung and lymph node metastases) provided that imaging techniques are adopted.

Metastatic sites

Usually, the response rate for patients with liver metastasis does not differ from that observed in the treated patient population overall [24]. However, it is recognized that localizations of the lung are less aggressive than those of the liver, with a median survival of 12 versus 8 months [25], even though this site of disease is less responsive to medical treatment, and locally advanced or relapsed disease is poorly sensitive to chemotherapy. Therefore, the distribution of patients with respect to this variable can be of great importance when comparing the results of apparently homogeneous studies.

Size of liver metastases or extent of liver involvement

A patient with extensive liver involvement has a shorter survival and is probably less likely to respond to treatment than a patient with more limited involvement. A universally accepted classification system for hepatic metastases is not yet available, but some good staging criteria are available [26, 27] which should be adopted in clinical studies. Alternatively, the median percentage of liver involvement, as evaluated by CT scan, should be reported in the demographic characteristics of the study [27, 28]. Notably, some centres include only patients with extensive liver involvement in trials of systemic chemotherapy, as other treatments such as surgery or intra-arterial chemotherapy are considered more suitable for patients with less extensive disease. It is conceivable that the response rates observed by these centres will therefore be lower than those of institutions which include all patients in trials of systemic chemotherapy.

Biochemical parameters

As expected, patients with organ function impairment can tolerate only less intensive doses of chemotherapy which in turn means a reduction in the intensity of treatment and a possible negative effect on drug activity. In colorectal cancer, impaired liver function is common and therefore variations in eligibility criteria and/or in the composition of the case-list regarding biochemical parameters may also explain the occurrence of differing results using the same treatment schedules.

How to avoid the risks of an arbitrary patient selection? It is important that patients included in clinical studies evaluating the same treatment regimen present with comparable characteristics. This aim can be more easily achieved if all consecutive patients fulfilling the eligibility criteria are included in the trial and are not selected arbitrarily or entered into other concurrent studies conducted in the same centre. This prerequisite is always stated in the methodology section of clinical studies but is sometimes not respected in practice. A formal list of all potentially eligible patients and the reasons for not including any patient in the study could permit better control of trial quality. As discussed below, a simpler way to avoid an arbitrary selection of patients could be the adoption of randomization also in phase II studies.

2. Definition and application of response criteria

As mentioned previously, nearly all studies use the WHO or UICC criteria for classification of clinical response. Although this should permit good comparability between trials, differences can arise if clinical or instrumental techniques are used in order to measure, for example, a nodular lesion of the abdominal wall or a superficial lymph node. In our opinion, it would be better to always use an imaging technique, for example, CT scan, NMR or ultrasonography, which permits a deeper and more reproducible measurement of tumor lesions. Several years ago, a note of caution was raised against the frequent possibility of inter-observer variability (at least 30% of cases) in the clinical evaluation of an objective response to treatment [29, 30]. Although this risk also exists with some imaging methods, chiefly ultrasonography, it is probably lower than the risk associated with clinical evaluation.

Another potential source of confusion is the adoption of particular criteria for the clinical measurement of liver metastases [4], according to which there must be a reduction in malignant hepatomegaly (defined by extension of the liver edge at least 5 cm below the costal margin on quiet respiration) greater than 30% to qualify for a partial response. We believe it would be better not to use this method now that much more sophisticated techniques are available.

Moreover, it would be suitable to standardize the radiological procedure necessary for an accurate evaluation of liver metastases: for instance the Gastrointes-
tinal Tract Cancer Cooperative Group of the EORTC (European Organization for the Research and Treatment of Cancer) recently produced a detailed proposal [31] which probably deserves validation in large cooperative clinical trials. According to this proposal, a standard procedure for liver measurement by ultrasound only could be adopted, while CT scan would not be mandatory in every case, but in doubtful ones only.

3. Assessment of time to progression or duration of response

Definition of the exact moment when progression of the disease occurs following a significant objective response or disease stabilization is of utmost importance in the assessment of duration of response or time to progression. It has been reported [21] that there is a trend toward higher response rates and longer durations of remission in those trials which evaluate patients less frequently and stipulate a shorter minimum duration as acceptable for definition of a positive response.

Another important issue concerns the method of evaluation. As progression of disease is defined as an increase of at least 25% in the sum of the lesion areas, it is obvious that an inter-operator difference can lead to a different evaluation when a small lesion is considered. In contrast, when a computer-assisted programme is used it will detect more limited changes, leading to a more frequent definition of progressive disease. The same problem can also occur when discriminating between two different categories of response; this is of particular importance when deciding if a reduction in tumor bulk may be classified as a partial response or a no change. Because a clear reduction in tumor mass of 80% to 90% is difficult to achieve in colorectal cancer with the regimens currently available, discrimination between a 49% or 51% shrinkage is theoretically important for classification of the objective response but is largely dependent upon the skill of the individual observer. Instead of quoting the response as a simple percentage and as a 'yes or no' outcome, a more suitable way of presenting the results would be to illustrate each case using a histogram; with this method the percentage reduction in tumor shrinkage is usually represented by a continuous spectrum with several cases falling within the central part of the histogram. The first examples of this simple but strict method of reporting [32] are already available and should be welcomed.

4. Independent review of response

The importance of independent review is becoming increasingly evident as more clinical trials are being conducted in Europe according to the rules of Good Clinical Practice (GCP). Even before the introduction of these guidelines some cooperative groups followed the policy of collectively reviewing positive responses which reduced the risk of false positive results. Furthermore, a review of all cases by a panel of external observers who are unaware of the patients' identity or treatment and the responses assigned by the clinical investigator can also be useful in reducing the possibility of false negative results, which are probably more frequent than expected. A 'certified' assessment of response is more reliable and convincing for the scientific community and should have a stronger impact on clinical practice. Every clinical study should include this type of control in the future.

5. Factors related to the therapeutic regimen

Dose intensity

In several malignancies achieving an optimal dose intensity is widely known to influence response rate. In colon cancer, data are available for the different 5-FU schedules with higher activity observed for continuous infusion [20]. However, within a given study a difference in response between patients receiving a dose intensity very near to the planned one and those who do not is rarely reported. Recently, a study [33] was carried out to further investigate the concept of 5-FU dose-intensity, and a clear relationship between drug plasma levels, toxicity and activity was demonstrated.

Modality of administration

The activity and toxicity of 5-FU can be altered by different administration schedules. The duration of administration appears to be of particular importance, with a marked increase in the therapeutic index resulting from long-term continuous infusion compared to bolus administration, as mentioned previously. Recently, particular attention has focused on the actual duration of administration of bolus 5-FU when modulated by FA [34]; a retrospective analysis [35] showed that patients treated with a 2- to 3-minute bolus injection achieved a higher response rate and experienced greater toxicity than patients receiving the drug more slowly over 10 to 15 minutes. Even though the results of randomized prospective studies evaluating the effect of these slight differences on the therapeutic profile of biomodulated 5-FU are not yet available, a more careful consideration of this issue now seems warranted [36].

6. Statistical considerations

Particular attention should be paid to the statistical methods section of any paper reporting the results of a clinical study in cancer chemotherapy. Some important points of this field need to be discussed:

- first, the 'publication bias': promising results are published faster and in journals with a higher impact factor than are negative results. Although this phenomenon concerns all malignancies, it can be particularly important in such a chemoresistant disease as colorectal cancer.
- the number of patients included in the study: al-
though the more recent trials generally follow the stringent criteria established for the sample size, great caution should be used when interpreting historical results or very promising early reports, particularly when presented as abstracts. Generally, there is a trend toward higher response rates in smaller studies due to the random variation. To reduce the risk of overestimating promising early results the report of the 95% confidence intervals (the width of which is inversely related to the 'true' response rate) is of utmost importance.

- only a few randomized phase II clinical trials are available in this field: randomization can be a simple mechanism for controlling the results of a study by 'weighing' the activity of an innovative therapy against that of a standard regimen.
- both in phase II and phase III trials, most important is the definition of 'evaluable' patients: several authors improperly exclude patients who do not fulfill all the entry criteria despite having received treatment, those who have not received the number of courses specified by the protocol, ones who are lost to follow-up or die early as a result of disease progression or adverse effects. In a disease such as advanced colorectal cancer, in which the absolute number of responses is generally limited, any modification in the total number of patients considered to be evaluable can lead to important differences in the results of the study and their subsequent interpretation and impact. The solution to this bias is the inclusion of all patients entered in the study in the final evaluation (intent to treat analysis). This practice is now widely accepted and applied in phase III trials but should also be used in phase II studies.

Discussion

The difficulty in obtaining concordant results in trials evaluating the same regimen is evident when surveying the extensive literature concerning advanced colorectal cancer. It appears that lower activity is generally observed when the regimen is evaluated outside the institution which conducted the original evaluation and when the number of treated patients is increased. This 'fall effect' has a detrimental effect on the reliability of promising results and can produce a pessimistic attitude in the oncology community.

Should we therefore conclude that every positive result in this field is only an illusion, even when the studies are performed by qualified investigators in renowned institutions? Certainly not. On the contrary, we have to realize that several factors, reported previously, can influence the activity of what is, or appears to be, the same therapeutic regimen. A critical analysis of the literature is therefore mandatory when interpreting different results reported for the same treatment regimen, even though a well-defined model capable of predicting a statistically significant relationship between the response rate and the stringency of the implemented criteria is as yet unavailable.

How can we overcome this problem? Possible solutions could be, firstly, the strict application of standard criteria to assess response and developing homogeneous trials. Agreements should be reached regarding, for example, an optimal case list composition, the practical method of drug administration, the use of imaging techniques to measure the response and the statistical methodology. As far as this last point is concerned, the above suggestions (publication of both positive and negative results, inclusion of an adequate number of patients and report of 95% confidence intervals, adoption of randomization also in phase II trials, intent to treat analysis in all studies) should be followed and their implementation could be more realistically expected, in most cases, than that of other recommendations. Usually, the more rigorous the criteria, the lower the response rates are likely to be, even though the difficulty in discriminating between a significant and a minor response could also lead, as discussed previously, to an underestimation of the clinical activity of a treatment regimen. With these guarantees several trials could be compared and provide reliable information regarding the activity and tolerability of the regimen under evaluation.

A second solution could be to combine all trials pertaining to a particular issue in a meta-analysis. The advantages and limitations of this instrument have been extensively discussed elsewhere [37, 38] and, with appropriate use and methodology [39], it is possible to overcome the lack of homogeneity of single trials in which discordant results have been reported. However, the bases of a good meta-analysis are high-quality, well-conducted clinical trials.

Even though complete homogenization of the criteria for response assessment is important for the entire field of medical oncology, advanced colorectal cancer could represent a priority due to the specific characteristics of the disease and the nature of the antiproliferative drugs employed in its treatment. Furthermore, every step toward improved methodological consistency will also have a positive impact on other cancers.

References

5. The Advanced Colorectal Cancer Meta-Analysis Project.


