Eur J Cancer Clin Oncol, Vol. 21, No. 8, pp. 889–891, 1985.
Pergamon Press Ltd. Printed in Great Britain.

Editorial

RECENTLY the Editorial Board of this journal approved a proposition from the editor to publish guidelines for reports of clinical trials. We heard from the Editorial Board of Cancer Treatment Reports *that they had the same intention and shortly thereafter their text appeared in their journal. These guidelines conform well to what we wanted to recommend to our contributors. Thereafter it was decided to publish guidelines in the* European Journal of Cancer & Clinical Oncology. *The authors, Richard Simon and Robert E. Wittes, gave permission to publish their text and we thank them very much. Parts of the following text will, after a period of reflection, be incorporated in the recommendations to authors. We are inviting comments on these guidelines from our readers and contributors.*

METHODOLOGIC GUIDELINES FOR REPORTS OF CLINICAL TRIALS*

Effective communication of clinical trials results requires informative and properly analyzed reports. To help ensure the quality of publications in *Cancer Treatment Reports,* the Editorial Board has adopted a set of methodologic guidelines for manuscripts. The guidelines appear below and will appear in future issues as part of Instructions to Authors.

The guidelines, though not all-inclusive, address basic issues in the design, conduct and analysis of clinical trials. They are not details which are inessential, abstruse or difficult to implement. Institutions serving as research bases for clinical trials should have the mechanisms at hand for dealing with the implications of each of the guidelines. The principles they represent are general and sensible ones; accordingly, the Editorial Board asks both authors and reviewers to use the guidelines routinely in the preparation and review of manuscripts. As with any set of guidelines, exceptions can be made where circumstances justify them. After an initial trial period, the Editorial Board will assess the feasibility of making compliance with the guidelines a precondition for publication.

1. Authors should discuss briefly the quality control methods used to ensure that the data are complete and accurate. A reliable procedure should be cited for ensuring that all patients entered on study are actually reported upon. If no such procedures are in place, their absence should be noted. Any procedures employed to ensure that assessment of major endpoints is reliable should be mentioned (e.g. second-party review of responses) or their absence noted.

Comment: The intent here is that a report should make clear the extent to which the major data of the study rest on a firm and verifiable foundation. To ensure that all patients entered on a study are in fact included in the final report, the research base should have a formal registration mechanism for study entry. We are well aware that many (perhaps most) single institutions do not currently have such mechanisms. We also recognise that formal second-party review of responses, though highly desirable, is not a widespread practice. Quality control of response assessment requires much greater attention than it usually receives. Currently, numerous response criteria are employed, and the inter-observer reliability of these is almost totally unknown. In any case, where such procedures are in place, they should be explicitly cited in the Methods section of the manuscript.

We are hopeful that the increased attention in institutions to essential features of trials methodology, as emphasized by the NCI-supported site-visit monitoring program, will result in the more general implementation of formal registration mechanisms and data verification procedures similar to those already in place in the clinical cooperative groups.

---

2. All patients registered on study should be accounted for. The report should specify for each treatment the number of patients who were not eligible, who died or withdrew before treatment began. The distribution of follow-up times should be described for each treatment, and the number of patients lost to follow-up should be given.

*This text appeared in *Cancer Treatment Reports*, Vol. 69, No. 1, January 1985, and is reproduced with the kind permission of the authors Richard Simon and Robert E. Wittes.

Comment: Differences in policies for excluding patients from analysis are a source of variation in results among similar studies. Regardless of how response rates are calculated, all patients must be accounted for. This will permit the reader to recalculate rates as he or she wishes.

---

3. The study should not have an inevaluability rate for major endpoints of greater than 15%. Not more than 15% of eligible patients should be lost to follow-up or considered inevaluable for response due to early death, protocol violation, missing information, etc.

Comment: The 15% figure is obviously somewhat arbitrary, but inevaluability rates of $\geq$20% usually reflect inappropriate patient selection. For phase III studies, disqualifications are a source of potential bias; when the disqualification rate approaches the magnitude of the difference in outcomes being tested, the results are not sufficiently reliable.

---

4. In randomized studies, the report should include a comparison of survival and/or other major endpoints for all eligible patients as randomized, that is, with no exclusions other than those not meeting eligibility criteria.

Comment: Comparisons of outcomes in randomized studies that exclude eligible randomized patients are subject to potential bias. Patients who refuse further treatment, for example, may be prognostically favorable or unfavorable. This has been clearly demonstrated for placebo patients in major cardiovascular trials. Consequently, the analysis of randomized trials should contain comparisons of all eligible randomized patients. The report may also contain other comparisons.

---

5. The sample size should be sufficient to either establish or conclusively rule out the existence of effects of clinically meaningful magnitude. For "negative" results in therapeutic comparisons, the adequacy of sample size should be demonstrated by either presenting confidence limits for true treatment differences or calculating statistical power for detecting differences. For uncontrolled phase II studies, a procedure should be in place to prevent the accrual of an inappropriately large number of patients, when the study has shown the agent to be inactive.

Comment: The point here is basic but frequently not recognized. Small studies that find no statistically significant differences between treatments are generally indeterminate, not negative [1]. Unfortunately, such studies are usually erroneously interpreted as negative. The problem is that the statistical power of small studies (i.e. the probability of obtaining a statistically significant difference if the two treatments are truly different) is low. Reporting confidence limits in addition to or instead of significance levels clarifies the distinction between indeterminate and negative results. For example, suppose the response rate for treatment A is 10/20 (50%) and for treatment B is 8/20 (40%). This difference is not significant ($P = 0.75$). But approximate 95% confidence limits for the true differences in response rates are -20.7% to +40.7%. So the data are consistent with both a moderate difference favoring treatment B and a tremendous difference favoring treatment A. The trial is not negative but rather indeterminate; the P value is misleading, and the number of patients is inadequate.

A sample size that is insufficient to answer the question originally posed by the trial is a serious and complex issue. Clearly, however, oncologists and cancer patients are not well served by the publication of results which are inconclusive because of avoidable flaws in trial execution. The trial which does not accrue an adequate number of patients is a failed experiment; unless the reason for the poor accrual is itself illuminating, the field is no wiser after the trial than before.

Overaccrual to a trial of an agent which proves to be inactive is equally bad methodology and also is ethically questionable. A negative phase II trial should cease accrual as soon as it has shown that the activity of the agent is lower than the level that is medically important. Once this question has been answered, the addition of more patients is counterproductive scientifically and very difficult to defend ethically.

---

6. Authors should state whether there was an initial target sample size and, if so, what it was. They should specify how frequently interim analyses were performed and how the decisions to stop accrual and report results were arrived at.

Comment: This refers to the sequential analysis of data as they are accumulating. It is not appropriate to interpret significance levels and confidence intervals at face value if one repeatedly analyzes accumulating data. That is, stopping accrual and publishing results as soon as a P value falls below 0.05 is a procedure with a high

probability of producing erroneous conclusions [2]. Generally it is necessary to perform interim evaluations of results. But premature termination and reporting of the study should be based upon P values much smaller than 0.05 if unreliable results are to be avoided. Sequential analysis of results is a technical issue for which a statistician is usually required. For purposes of these guidelines, we feel that the manuscript should describe the initial target sample size, the history of interim analysis, and the circumstances concerning the decision to report results.

---

7. All claims of therapeutic efficacy should be based upon explicit comparisons with a specific control group, except in special circumstances where each patient is his own control. If nonrandomized controls are used, the characteristics of the patients should be presented in detail and compared to those of the experimental group. Potential sources of bias should be adequately discussed. Comparison of survival between responders and nonresponders does not establish efficacy and should not generally be included. Reports of phase II trials which draw conclusions about antitumor activity but not therapeutic efficacy generally do not require a control group.

Comment: Controls are generally not required for phase II trials because no claims of therapeutic efficacy are (or should be) made. Phase II trials attempt to evaluate only antitumor activity. Phase III trials, however, require controls. The recommendations state that nonrandomized studies should be performed in the cleanest possible manner using explicit controls for which comparability can be thoroughly evaluated on a patient-by-patient basis [3]. Comparison of survival between responders and nonresponders is not a valid way of establishing therapeutic efficacy [4–6]. This comparison can be biased in several ways. First, patients who die quickly are by definition nonresponders. Hence, there is a time bias. Second, responders may have more favorable prognoses regardless of treatment. They may have less disease, less prior treatment, and better performance status. They may also be more favorable with regard to unknown prognostic factors. A similar issue occurs in the evaluation of heart transplantation in nonrandomized studies. Patients who live long enough for a donor to be found may do very well without a transplant. For cancer studies, to evaluate the impact of a treatment on survival or disease-free survival, outcomes for all of the treated patients should be compared to those for an appropriate control group.

---

8. The patients studied should be adequately described. Applicability of conclusions to other patients should be carefully dealt with. Claims of subset-specific treatment differences must be carefully documented statistically as more than the random results of multiple-subset analyses.

Comment: Care should be employed in extrapolating results to the general population of patients. Only a small fraction of patients enter clinical trials, and they are not a random sample. Proper statistical methodology is necessary to distinguish true subset-specific treatment differences from the random results expected from multiple-subset analyses [7]. It is not generally recognized that, by chance alone, there is a 40% probability of finding at least one statistically significant false-positive treatment difference in the evaluation of ten disjoint subsets.

9. The methods of statistical analysis should be described in detail sufficient that a knowledgeable reader could reproduce the analysis if the data were available.

Comment: This stipulation is self-explanatory.

REFERENCES

1. Freiman JA, Chalmers TC, Smith H Jr *et al.* The importance of beta, the type II error, and sample size in the design and interpretation of the randomized clinical trial. Survey of 72 "negative" trials. *N Engl J Med* **299**, 690–694, 1978.
2. McPherson K. The problem of examining accumulating data more than once. *N Engl J Med* **290**, 501–502, 1974.
3. Gehan EA, Freireich EJ. Non-randomized controls in cancer clinical trials. *N Engl J Med* **290**, 198–203, 1974.
4. Mantel N. An uncontrolled clinical trial — treatment response or spontaneous improvement? *Controlled Clin Trials* **3**, 364–370, 1982.
5. Weiss GB, Bunce H, Hokanson JA. Comparing survival of responders and non-responders after treatment: a potential source of confusion in interpreting cancer clinical trials. *Controlled Clin Trials* **4**, 43–52, 1983.
6. Anderson JR, Cain KC, Gelber RD. Analysis of survival by tumor response. *J Clin Oncol* **1**, 710–719, 1983.
7. Simon R. Patient subsets and variation in therapeutic efficacy. *Br J Clin Pharmacol* **14**, 473–482, 1982.

ADDITIONAL REFERENCE

Buyse ME, Staquet MJ, Sylvester RJ, eds. *Cancer Clinical Trials: Methods and Practice.* Oxford, Oxford University Press, 1984.