

How to appraise a clinical trial critically

LLOYD R SUTHERLAND MD FRCPC

LR SUTHERLAND. How to appraise a clinical trial critically. Can J Gastroenterol 1995;9(7):401-404. Most clinicians wish to base their therapeutic decisions on scientific data but are often overwhelmed by the number of case reports, open series and other therapeutic trials published each year. It is essential to develop a personal screening plan that can alert the clinician to which reports deserve extra attention, as they may alter management, and to which reports can be ignored. The first step is to review the abstract. Decisions related to therapeutics should be based on 'randomized', 'controlled', 'double-blinded', 'clinical trials'. If these key words are not included in the abstract, it is unlikely that the report will change clinical practice. Each of these terms describes an essential element that attempts to assure that the results of the trial will be unbiased and generally applicable to clinical practice. The next step is to examine the 'Methods' section. Clinicians should be interested in the inclusion and exclusion criteria. Two questions arise. First, do the study participants resemble patients in your practice? Second, how many patients were assessed in order to enrol the study population? If only a few of the patients screened actually entered the trial the results may not be of use for the general population with the disease. Another important screening process is to review the 'Statistical' section. It is not necessary to be a statistician, but one should read the section to determine whether a sample size calculation was performed and whether confidence intervals have been calculated around the major end-points. By remembering these key concepts, clinicians can reduce the number of journal articles they read without compromising their ability to be informed of major breakthroughs in the management of disease. (*Pour le résumé, voir page 402*)

Key Words: Blinding, Clinical trials, Critical appraisal, Controls, Methodology, Randomization

Department of Medicine, University of Calgary, Calgary, Alberta

Correspondence and reprints: Dr LR Sutherland, Department of Medicine, University of Calgary, 3330 Hospital Drive NW, Room 1751, Calgary, Alberta T2N 4N1. Telephone 403-220-4500, fax 403-283-3028, e-mail lsutherl@acs.ucalgary.ca

This paper was presented at the Trends in Inflammatory Bowel Disease Therapy meeting, April 6 to 9, 1994, held in Victoria, British Columbia. This paper has also been published in Sutherland LR, et al, eds. Inflammatory Bowel Disease: Basic Research, Clinical Implications and Trends in Therapy. Boston, Dordrecht and London: Kluwer Academic Publishers, 1994

THE PRESCRIPTION OF MEDICATIONS is a daily activity for most physicians. The act of prescribing incorporates a number of factors including issues related to effectiveness, appropriate use, comparability with other medications, adverse events and cost. The answers to most of these issues can be found within the body of the available medical literature. But therein lies a dilemma: clinicians wish to base their therapeutic decisions on scientific data but are often overwhelmed by the number of case reports, open series and other therapeutic trials published each year.

For example, a search within MEDLINE for 1992 identified 3285 citations for the therapy of gastrointestinal disease. The majority of the publications, however, were isolated case reports, preliminary studies of groups of patients all receiving the same medication ('open series') or reviews. It is doubtful that any of these reports would have modified practice patterns. Such references should be skimmed and considered as perhaps previews of how therapy may change in the future. However, the history of medicine, including gastroenterology, is littered with reports of medications and interventions that appeared promising in small open series but failed the more

Comment faire la critique d'un essai clinique ?

RÉSUMÉ : La plupart des cliniciens souhaitent baser leurs décisions thérapeutiques sur des données scientifiques, mais ils croulent souvent sous l'abondance de rapports de cas, de séries ouvertes et autres essais thérapeutiques publiés chaque année. Il convient donc d'adopter un plan d'analyse personnel pour signaler au médecin quels rapports méritent qu'on leur consacre plus d'attention parce qu'ils peuvent modifier le traitement, et lesquels sont à ignorer. La première étape est de lire le résumé. Les décisions quant aux traitements doivent se baser sur des essais cliniques randomisés contrôlés à double insu. Si ces mots clés ne sont pas inclus dans le résumé, il est peu probable que le rapport change la pratique clinique de quelque façon que ce soit. Chacun de ces termes décrit un élément essentiel qui assure l'impartialité des résultats de l'essai et en général, leur applicabilité à la pratique clinique. La prochaine étape est d'examiner la section « Méthodologie ». Les médecins doivent porter attention aux critères d'inclusion et d'exclusion. Deux questions viennent à l'esprit. D'abord, les participants ressemblent-ils à ceux de votre pratique ? Ensuite, combien de patients ont été évalués en vue de former la population étudiée ? Si seulement quelques patients ont été évalués et inclus dans l'étude, les résultats risquent d'être peu pertinents pour la population en cause dans la pratique. Un autre important processus d'analyse est de consulter la section « Statistiques ». Il n'est pas nécessaire d'être statisticien, mais il faut lire cette section pour déterminer si le calcul de la taille de l'échantillon a été effectué et si l'intervalle de confiance a été calculé relativement au principal paramètre. En se rappelant de ces notions fondamentales, les cliniciens peuvent réduire le nombre d'articles scientifiques à lire, sans compromettre leur capacité de rester informés au sujet des principaux progrès thérapeutiques.

rigorous methodology of the clinical trial.

Returning to the 3285 citations, simply restricting the search to randomized trials reduced the number of papers to 227. That is still a large number of papers to review but a task that can be managed and enhanced by the application of a few simple rules (1).

This review outlines strategies that should allow clinicians to develop a personal screening plan to reduce the number of papers that have to be carefully reviewed. A second goal is to assist in identifying those publications that can alter the pattern of prescribing medications.

THE ABSTRACT

Is the trial double-blind, randomized and controlled?: Although the natural history of the development of an indication for a particular medication includes an 'open label' study or 'open series' in which all patients receive the medication, the impact of such a study on clinical practice should be minimal. Such studies may be categorized as

"honest efforts" (personal communication). They should alert the clinician to the possibility that drug X may be of benefit for a particular disease. However, the results are likely to be greatly biased towards efficacy of the drug, and the sample size is generally small. Following the publication of such a case series, several additional clinical trials should be performed before the medication can be recommended for a particular disease or indication.

The first step to critical appraisal is to scan the title and abstract for these three key words: 'randomized', 'controlled' and 'double-blind'. The inclusion of these words or phrases in the abstract or key words section should alert the clinician to the possibility that this report may alter the way in which he or she practises medicine. Blinding of both patient and investigator, the use of a control substance (either placebo or other active medication) and randomization to prevent entry bias should produce a trial of reasonable quality. These techniques are fundamental to reducing the inherent bias, which is the clinical assump-

tion that the new medication will be superior to the conventional therapy, present in any trial. If the abstract confirms the inclusion of these key phrases, the reader should move on to the 'Methods' section.

THE METHODS SECTION

The 'Methods' section is one of the most important sections of the paper but is often skipped by the reader. This practice is often encouraged by the publisher, who in a few journals prints this section in the smallest type!

Randomization: As Meinert (2) has suggested, randomization provides part of the basis for statistical analysis, but from a practical point of view its major effect is to increase the chances of bias-free treatment assignment.

Randomization implies that each patient has an equal chance to receive either treatment. Physicians may often have a bias towards one therapy and may influence the trial results by assigning their patients to one treatment preferentially. The ideal strategy is for patients to be assigned by sealed envelopes based on random number generation. In the case of a multicentre trial, the system may be centralized. Schemes that rely on birth dates, hospital identification numbers or day of the month are susceptible to bias. Using a sealed envelope is not foolproof. There are several stories of investigators pulling envelopes until they found the therapeutic assignment they were looking for! A log that reports patient assignment by entry number and entry date provides a check against such interference.

The control: The control is the medication or intervention with which the newer therapy is being compared. Studies that do not include a control intervention do not have the same credibility as those that do. Within the context of the pharmaceutical sphere the control will be either a placebo or another active medication. The control should be identical in appearance to the active medication and, if possible, should taste the same. Study participants have been known to compare notes with each other or to take their medications to

pharmacists, chemists, etc, for analysis to break the code.

In many cases, particularly when two active medications are being compared, it is not possible to construct identical capsules. The appropriate adjustment in this case is to conduct a 'double-dummy' trial in which placebo, identical to both active medications, are prepared and the patients are assigned to either active medication and the placebo of the alternate therapy. The failure to use such a technique will cast a shadow over a study. For example, in one of the few comparison studies of two 5-aminosalicylic acid (5-ASA) preparations for maintenance of remission in ulcerative colitis, a double-dummy technique was not used (3).

In a recent report of the use of 5-ASA for the prevention of postoperative recurrence of Crohn's disease, although patients were randomized they were randomized either to treatment or no medication (4) rather than a placebo. Although the authors dealt with the issue in their discussion and felt that they had demonstrated no ill effect from the lack of a placebo, the study would have been enhanced by the use of one.

Blinding: Blinding is another essential element required in order to have confidence in the study results. The term usually implies that both the patient and the physician are blinded to the treatment assignment. The term 'triple-blind' refers to blinding of the patient, the physician and the statistician doing the analysis to the drug assignment.

Blinding is easiest to perform if two medications are being compared (see the section above). It is more difficult if two different interventions are being compared. For example, how would one blind a study comparing total parenteral nutrition with corticosteroid therapy? A variety of strategies can be used. One possibility is to define the outcome by objective criteria (change in hemoglobin, albumin, etc), to be analyzed by individuals unaware of the assignments. More subjective evaluations can be performed by individuals blinded to the assignment who may, for example, evaluate the participants using a telephone interview.

TABLE 1
Simple checklist for reviewing clinical trials

Method of randomization	Pharmacy or random numbers
Blinded	Placebo identical in appearance and taste, double-dummy technique if not possible to give both medications in identical capsules
Entry criteria	Clearly stated List of patients screened provided
Patient population	Results account for each patient, ie, completed, eliminated, adverse reactions
Statistics	Intent to treat analysis Sample size calculation or reference to consideration of an alpha or beta error Confidence intervals Life-table analysis
Compliance	Tested by pill count or biological equivalent

Other issues: How do the patients in this study compare with the patients in your own practice? A study of high quality will inform the reader of the number of patients that had to be assessed in order to arrive at the study population. Minor points to note in the 'Methods' section include acquisition dates (starting and stopping) and the use of a reject log. It is common to screen two to three times the number of patients required to arrive at the patient population. Was a log of screening patients kept? How do the excluded patients differ from those who were included? This information will assist the reader in relating the study population to his or her own practice.

How many patients dropped out of the study? This will give an impression of the side effect profile. If the medication is poorly tolerated, a greater proportion of patients will drop out. This rule does not apply as well in placebo controlled trials where a higher rate of drop out can be expected. Patients who are doing poorly are assumed to be on placebo and are withdrawn.

Awareness of these points will help the reader to determine how relevant the study population is to his or her practice. For example, if the study took several years to recruit a small number of patients or most of the patients approached declined, then the study population may not be representative of the total patient population.

Compliance should be assessed preferably by biological tests (serum or urine analysis) but pill counts are better than nothing.

STATISTICAL METHODS

Most clinicians skip over the 'Statistical Methods' section and regard it as a lot of mumbo jumbo or a black box. To a certain extent, the clinician has to rely on the abilities of journal reviewers and editors to ensure that the appropriate statistical techniques were used. There are, however, some questions that physicians should ask each time they review a clinical trial.

First, was an intention to treat compared with a per protocol analysis used? In the intention to treat analysis all patients who took even one tablet are included in the analysis. The per protocol analysis is confined to patients who have complied with all aspects of the protocol. An intention to treat analysis is more conservative, tends to give results that are less biased towards showing efficacy and may reflect what happens in the real world. In contrast, the per protocol analysis provides information related to what to expect in patients who tolerate the medication and who are compliant.

Second, it is worthwhile to check whether a sample size calculation is mentioned. The sample size calculation is particularly important if a 'negative' trial is reported. In certain cases, the conclusion that drug A is equivalent to drug B or placebo may be a function of low statistical power to detect a difference, which is related to an insufficient number of study subjects. Many early trials of Crohn's disease therapy that did not demonstrate efficacy are flawed because of small sample size (5). It was not until the 1980s and trials in-

volving hundreds of patients that the possible beneficial effects of 5-ASA could be identified. Broadly speaking, studies that use small patient groups can detect only very large differences in outcome. The larger the study sample, the smaller the difference in efficacy that can be detected. If there is no mention of such issues in either the 'Methods' or 'Discussion' section, this may indicate a poorly designed study.

Third, do the investigators report their major results with 95% confidence intervals? Confidence intervals provide the clinician with the range of possible results if the experiment was repeated, using the same number of patients, 95 out of 100 times. The values are sensitive to the number of patients enrolled in the trial. If the intervals include unity or a range of positive and negative values, the results are not statistically significant. Although most clinicians focus on the P value to help evaluate the results of a study, consideration of the confidence intervals is often more helpful and may explain inconsistent results.

Finally, presentation of the data in the form of a life-table will often clarify response rates to various medications.

RESULTS

How was the randomization assessed?: There should be an attempt to determine whether the treatment groups are comparable in terms of basic demographic characteristics and disease vari-

ables. An acceptable 'Results' section includes the comparison of the treatment groups to determine whether the randomization was successful.

Readers should not be distracted by concerns related to statistical significance compared with clinical significance. Consider a modest difference in effect between two medications that is statistically significant. The effect may be clinically important if the disease is serious or lethal, but may not be clinically important if the disease is not life-threatening. Whether the P value is 0.05, 0.01 or 0.001 does not determine clinical significance.

Subgroup analysis, if not part of the original study design, should be considered as pointing out the directions for future research rather than as definitive findings. Adverse events should be detailed, and the timing of significant end-points should be readily apparent to the reader.

CONCLUSIONS

Given the incredible growth of the medical literature and the responsibility that we as physicians have to remain current, strategies are necessary to focus attention on reports that have the potential to alter clinical practice.

Critically appraising the literature should assist in mastering the continually growing information available on which decisions for therapeutics are based. A simple checklist (Table 1) may assist.

Clinicians and health care agencies

wish to base their decisions regarding therapeutics on objective evidence gained through the execution of clinical trials. Just as with most things in life, not all trials are equal in terms of quality or reliability. Decisions regarding therapeutics should only be based on properly executed trials of high quality. It is not necessary to be a biostatistician or clinical trials expert to appraise critically the trial literature. A few simple concepts will assist in reviewing any report of a therapeutic intervention.

REFERENCES

1. Department of Clinical Epidemiology and Biostatistics McMaster University Health Sciences Centre. How to read medical journals: V. To distinguish useful from useless or even harmful therapy. *Can Med Assoc J* 1981;124:1156-62.
2. Meinert CL. *Clinical Trials Design Conduct and Analysis*. Oxford: Oxford University Press, 1986.
3. Courtney MG, Nunes DP, Bergin CF, et al. Randomised comparison of olsalazine and mesalazine in prevention of relapses in ulcerative colitis. *Lancet* 1992;339:1279-81.
4. Caprilli R, Andreoli A, Capurso L, et al. Oral mesalazine (5-aminosalicylic acid: Asacol) for the prevention of post-operative recurrence of Crohn's disease. *Aliment Pharmacol Ther* 1994;8:35-43.
5. Sutherland LR. 5-aminosalicylates for prevention of recurrence in patients with Crohn's disease: Time for a reappraisal? *J Clin Gastroenterol* 1991;13:5-7. (Edit)



Hindawi
Submit your manuscripts at
<http://www.hindawi.com>

