Annotations

Clinical trials, statistics, and dilemmas

A number of journals over the last decade have drawn attention to faults in the design and statistical analysis of published clinical trials.¹⁻⁵ Research workers may have splendid ideas for evaluating treatments but too often embark on studies with little knowledge of the discipline of good design and very little understanding of the concepts behind the mathematics of statistical analysis. In order to educate the profession in these matters the *Journal of Pediatrics* has introduced 'occasional critiques' of published papers to identify strengths and weaknesses of design and analysis.⁶ Reviews by Schwartz and colleagues,⁷ Gore and Altman,⁸ and Pocock⁹ are accessible to those with little knowledge of mathematics, and the *British Medical Journal's* statistical guidelines for contributors to medical journals¹⁰ should help the novice research worker plan his study to avoid, as far as possible, the many pitfalls he might otherwise encounter. Clinical trials whose results are inconclusive because of poor planning are wasteful of time and resources and their ethics are questionable.

Why are clinical trials important?

The proper evaluation of treatment should avoid anecdote as a way of promoting or condemning a certain treatment. Anecdote is anathema to critical appraisal because it does not remove bias. 'I have seen a number of patients whose symptoms have improved by treatment X. They will vouch for it.' This line of thinking is still widely prevalent and when such testimonials are presented, whether at academic meetings or in the media, they should be challenged. The statement would be better put 'I have formed the clinical impression that . . . Could it be important and worth validating? Is it real or is it artefactual? I must attempt to answer these questions objectively.'

'SUCCESS' STORIES

If, for example, it was publicised that a treatment for AIDS had produced a 'fantastic' improvement in three or four patients in whom it had been tried, it would be understandably very difficult to enrol patients for a clinical trial to evaluate such treatment. Some of the worst examples of this kind of

promotion are the world wide 'treatments' available for cerebral palsy. Families who have heard of their 'success' may pay huge sums of money to have their child treated. Even when a treatment that has been publicised to be of possible value has been subsequently shown to be worthless, it can still be very difficult to convince both patients and doctors that this is so. The Third World countries, for example, continue to pay huge sums for valueless anti-diarrhoeal preparations.¹¹

NON-VALIDATED TREATMENTS ALREADY IN PRACTICE

Once treatments have become part of standard practice without proper validation it can become extremely difficult to test them. What exactly is the place of physiotherapy in the management of cerebral palsy? For which patients is psychotherapy useful? If a treatment is 'safe' then why not use it just in case it may be valuable? Such examples of treatments have been chosen because at first sight both may be considered harmless and thereby accorded the 'benefit of the doubt.' Both, however, are time consuming and costly with potential effects on the rest of the family. Hopes may be raised without justification. As it would be very difficult to enrol patients for randomised controlled trials of these treatments versus no treatment, other methods, such as standard treatment against a less intensive variation of treatment, would have to be devised.

Patients, especially those with lethal or chronically disabling disorders, are a very vulnerable group. It could be argued that the offering of treatment which has not been validated or shown, in the widest context, to be harmless is a form of exploitation. This is of course a contentious issue and deserves continuing debate. It may be that for society as a whole just 'doing something', whether of proven value or not, is good enough reason to continue certain treatments that are particularly difficult to evaluate.

ASSOCIATION AND CAUSE

Some treatments are particularly difficult to validate. The management of patients with cystic fibrosis at specialist centres has been shown to be associated with improved mortality rates.¹² Associa-

tion is not cause but it is easy to argue that in this case it seems reasonable to conclude that the association is causal. It may indeed be reasonable but nevertheless untrue! It should be made very clear that when policy decisions are made on the basis of such associations they are based only on conjecture. Unwanted effects of treatment may also be difficult to evaluate. The nature of the relationship between oxygen treatment and the retinopathy of prematurity has not been, and may never be, defined.¹³ Yet the conviction, based on circumstantial evidence, of the causal nature of the association has influenced both medical and legal opinion for the last three decades.

Randomised controlled clinical trials

In 1773 when Dr Johnson considered why population numbers may have been exaggerated in the past he remarked: 'To count is a modern practice, the ancient method was to guess; and when numbers are guessed they are always magnified.'¹⁴ An analogy to this could be made in respect of clinical trials: controlled studies remove bias; uncontrolled studies exaggerate the value of a new treatment, often grossly. As accurate counts must have been difficult in Johnson's day, the conducting of a controlled clinical trial 'to remove bias' is far from easy in the present day.

PATIENT SELECTION AND INFORMED CONSENT

Dudley has recently drawn attention to a trial where difficulties in patient selection have cast doubt on the result of a time consuming and no doubt costly study.¹⁵ When patients are invited to participate in a trial of a new treatment where, at the toss of a coin, they may be allotted either the new treatment or an old or no treatment, some refuse to participate and select the treatment they prefer. This is one of the problems encountered when informed consent to enter a study is necessary. For example, if a comparison of home and hospital management of a behavioural problem were to be undertaken many families might not wish to take the chance of their child being randomised to a particular group. The behavioural patterns of the participants and the non-participants may well be different. How to deal with those who elect not to enter must be considered at the planning stage. The outcome of all patients who present with entry requirements must be included in the report.

TRIAL SIZE, CLINICAL SIGNIFICANCE, AND STATISTICAL SIGNIFICANCE

Another problem that must be considered at the design stage is the size of the trial—that is, how

many patients must be enrolled in order to detect reliably whether or not a new treatment is useful. To estimate this the smallest clinical change worth demonstrating should be decided. For example to demonstrate that a new treatment would reduce the admission rate for gastroenteritis by 25% might be worthwhile, whereas to demonstrate with statistical significance that it made a reduction of 5% might be considered not worthwhile. The number of patients required for a trial is inversely proportional to the square of the size of change considered clinically important to identify. Another question often asked by drug companies is whether a new drug is as good as one already marketed. If the new drug is cheaper or is thought to have fewer side effects, how much less effective would it have to be to make it an unacceptable alternative? The number of patients enrolled for the trial would have to be sufficient to demonstrate reliably whether or not this difference existed. If the clinical change is to be measured by a quantitative method, such as change in peak expiratory flow rate or the concentration of a serum constituent, then the variability of the measurement in the population—that is, its standard deviation—should also be known. Using this information together with figures for the risk of a false positive result—that is, the Type I error, usually 5% but preferably 1%—and for the risk of a false negative result—that is, the Type II error, usually between 5 and 20%—the number of patients to be enrolled for the study can be calculated.⁹ It is when insufficient attention has been paid to the exclusion of large Type II errors that the problem of 'negative' studies arises. Many studies which have reported 'no significant differences' between two treatments have simply not been powerful enough to show a difference, which may well have existed. Such studies have presented inconclusive results because of failure to enrol enough patients. An excellent commentary on this problem together with a worked example has recently been presented by O'Brian Smith.¹⁶

What size of change would be important to demonstrate may be difficult to decide. Another difficulty may be accruing sufficient numbers. The study of preventive measures and treatment for the retinopathy of prematurity, for example, would require the enrolment of patients from several centres. If the desired trial size is too large then it is better abandoned before it starts.

MODELS AND PILOT STUDIES

It is worth while to set out on paper during the planning stage possible series of results and then to analyse them as if for presentation including statistical analyses. This is a good way of developing a

'feel' for the variability in response and size of the trial and also a good exercise for understanding the concepts behind the statistics. Plans at this stage can be made for the timing of interim analysis, if this might be necessary.

If there is anxiety about accrual of patients, which may be particularly difficult when informed consent is necessary, a pilot study should be undertaken. This also gives an opportunity to examine the logistics of data collection, clarity of questionnaires, and so on. Apart from these concerns the period of the pilot study allows time for 'thought clarification', which may well have a good positive effect on a study protocol.

STATISTICAL CONSIDERATIONS

The *British Medical Journal* has led the way in this country in laying down guidelines for the analysis of clinical trials.¹⁰ The researcher should become familiar with presenting his results using confidence intervals when appropriate as these emphasise the clinical importance of results more than the statistical importance.¹⁷ (Observations which require non-parametric analysis do not, however, lend themselves easily to this.) Other useful concepts to understand are the proper collection and handling of ordinal data^{18 19} and the organisation of a crossover study.^{8 20}

Presentation of results

Good presentation is good communication. Apart from this, presenting the reader (and referee!) with all the data allows him to evaluate the analyses that were carried out for himself. Guidelines for the best ways of doing this diagrammatically are easily available.^{8 21}

Responsibilities

Clearly the investigator and his supervisor have responsibilities to familiarise themselves with the requirements of good study design. It is a waste of both investigators' and patients' time if every effort has not been made to ensure that a trial is likely to give a conclusive result. Ethical committees, among other considerations, must be aware of the importance of good study design and assure themselves of the scientific validity of any proposal.²² If no member of the committee is able to properly evaluate this, the committee should consider sending the proposal for consultation.

Editors of journals and referees of papers submitted for publication and presentation at academic meetings bear a great responsibility for criticising work of poor design and statistical validity. The use of a check list may be worth while.²³ The statistical

considerations for a trial including calculations for the number of patients to be enrolled should be clearly laid out in the methodology section. Referees should encourage authors to present, in a suitable form, their raw data, as far as is possible.

Conclusion

One of the ethical considerations of clinical research is its scientific validity. This needs to be considered, and seen to have been considered, at the planning stage of any clinical trial. It is the business of investigators, supervisors, ethical committees, referees of papers and editors of journals to ensure that, as far as possible, trials are conducted which have the greatest chance of reaching a conclusive result. Time and resources are otherwise wasted and the only beneficiary is the investigator who may have learnt from his mistakes. Policy decisions may sometimes have to be made based on information from retrospective and uncontrolled studies and when this is done it should be made clear that they are based on conjecture only; the risk of hampering further research must be considered and acknowledged.

I am grateful to Professors Dudley and Simpson and to Dr Carol Devateux for helpful comments on a first draft.

References

- Freeman JA, Chalmers TC, Smith H, *et al*. The importance of beta, the type II error and sample size in the design and interpretation of the randomised controlled trial. *N Engl J Med* 1978;**299**:690-4.
- Pocock SJ. Current issues in the design and interpretation of clinical trials. *Br Med J* 1985;**290**:39-42.
- Prescott RJ. The small clinical trial: is there a better way? *Thorax* 1986;**41**:822-3.
- Hennekens C, Buring JE. Need for large sample sizes in randomised trials. *Pediatrics* 1987;**79**:569-71.
- Pocock SJ, Hughes MD, Lee RJ. Statistical problems in the reporting of clinical trials. *N Engl J Med* 1987;**317**:426-32.
- Editorial. Analyses of statistical analysis. *J Pediatr* 1986;**109**:827.
- Schwartz D, Flamant R, Lellouch J. *Clinical trials*. London: Academic Press, 1980.
- Gore SM, Altman DG. *Statistics in practice*. London: British Medical Association, 1982.
- Pocock SJ. *Clinical trials: a practical approach*. London: John Wiley, 1983.
- Altman DG, Gore SM, Gardner MJ, Pocock SJ. Statistical guidelines for contributors to medical journals. *Br Med J* 1983;**286**:1489-93.
- Chetley A, Gilbert D. *Problem drugs*. The Hague: Health Action International, Drukkerij Aktief, 1986.
- Nielsen OH, Schiøtz PO. Cystic fibrosis in Denmark in the period, 1945-1981. Evaluation of centralised treatment. *Acta Paediatr Scand [Suppl]* 1982;**301**:107-19.
- Silverman WA. Epoché in retinopathy of prematurity. *Arch Dis Child* 1986;**61**:522-5.
- Samuel Johnson. In: Chapman RW, ed. *Journey to the Western Islands of Scotland*. London: Oxford University Press, 1924.

- ¹⁵ Dudley HAF. Extracranial-intracranial bypass, one; clinical trials, nil. *Br Med J* 1987;**294**:1501–2.
- ¹⁶ O'Brian Smith E. Interpreting results of a 'negative' study. *J Pediatr* 1986;**109**:828–9.
- ¹⁷ Gardner MJ, Altman DG. Confidence intervals rather than P values: estimation rather than hypothesis testing. *Br Med J* 1986;**292**:746–50.
- ¹⁸ Moses LE, Emerson JD, Hosseini H. Analysing data from ordered categories. *N Engl J Med* 1984;**311**:442–8.
- ¹⁹ MacKenzie CR, Charlson ME. Standards for the use of ordinal scales in clinical trials. *Br Med J* 1986;**292**:40–3.
- ²⁰ Vere DW. Validity of cross-over trials. *Br J Clin Pharmacol* 1979;**8**:5–6.
- ²¹ Reynolds L, Simmonds D. *Presentation of data in science*. The Hague: Martinus Nijhoff, 1982.
- ²² Robinson RJ. Ethics committees and research in children. *Br Med J* 1987;**294**:1243–4.
- ²³ Gardner MJ, Machin D, Campbell MJ. Use of check lists in assessing the statistical content of medical studies. *Br Med J* 1986;**292**:810–2.

SHEILA MCKENZIE
Rush Green Hospital,
Romford,
Essex RM7 0YT