MEDICAL PRACTICE

For Debate . . .

Do retrospective controls make clinical trials “inherently fallacious?”

LAWRENCE CRANBERG

British Medical Journal, 1979, **2**, 1265-1266

In 1957, a panel of the British Medical Association reported on a study, which employed retrospective controls, of using cortisone to treat polyarteritis nodosa. One conclusion was: “Its major importance is perhaps to emphasise that the assessment of therapeutic activity by the use of retrospective controls is an inherently fallacious method and no substitute for the strictly controlled clinical trial.”¹

The impact of this conclusion was probably increased by the accompanying citation of a paper on clinical trials by the distinguished medical statistician A B Hill.² The conclusion was later quoted without adverse comment by Hill (although in 1965 he spoke favourably of retrospective controls) in his important work *Principles of Medical Statistics*,³ and must have had considerable influence on the design of clinical trials. The conclusion is quoted again in the latest edition of Hill’s book.⁴

Surprising indictment

And yet, the indictment of retrospective controls as “inherently fallacious” is surprising not only because of lack of qualification, but because it results from a particular clinical trial rather than a general discussion of methodology. To prove a method “inherently fallacious” requires demonstration of a logical contradiction or a violation of a well-established scientific principle. But no such demonstration exists either in the report of the panel or in Hill’s cited work. On the contrary, the panel gives three convincing reasons for using retrospective controls

in its particular case: (1) the grave prognosis of the untreated disease in contrast to the “very favourable” reports on the proposed treatment already published; (2) the rarity of the disease; and (3) the pressure to reach a conclusion on the influence of the treatment on patient survival before the treatment became generally available. These circumstances are by no means uncommon and include common diseases with a wide range of manifestations—for example, cancer.

The methodological conclusion of the panel rests on the fact that the control group was not adequately matched to the experimental group. The two differed in the incidence of hypertension, a parameter that influences patient survival considerably. But the inference to be drawn from this need not be against retrospective controls. Quite the contrary: the fact that the medical records were good enough to show the difference made it possible to sort out the importance of hypertension and to assess the data more critically. The more refined classification of the data meant poorer statistics on each of the four subgroups: treated and untreated, with and without hypertension. But the poor statistics reflect only the paucity of the cases, and are irrelevant to the question of the inherent validity of retrospective controls versus concurrent or prospective controls. The use of retrospective controls did not produce a misleading result, but, as might have happened with concurrent controls, led to the discovery of an initially hidden variable.

To determine the relative merits of different methods, the results which would be obtained from them must be compared. The panel has not done this, but relative merit can be assessed from the data given.

Comparison of methods

With retrospective controls, the entire group of 17 living individuals could be treated, rather than only nine, if concurrent controls had been used. Thus all patients with severe constitutional illness received the benefits of cortisone treatment.

At the same time, data whose statistical weight was greater than if concurrent controls had been used, were obtained on side effects.

The panel recognised an obligation to provide to all the patients the known short-term benefits of cortisone. The obligation is an ethical one, and in the reference to Hill cited by the panel has been discussed as follows in a section entitled "Aims and ethics":²

"The first step . . . (in a clinical) trial is to decide precisely what it is hoped to prove, and secondly to consider whether these aims can be ethically fulfilled. It need hardly be said that the latter consideration is paramount and must never, on any scientific grounds whatever, be lost sight of. If a treatment cannot be ethically withheld then clearly no controlled trial can be instituted. All the more important is it, therefore, that a trial should be begun at the earliest opportunity, before there is inconclusive though suggestive evidence of the value of the treatment. Not infrequently, however, clinical workers publish favourable results of three or four cases and conclude their article by suggesting that this is the method of choice, or that what is now required is a trial on an adequate scale. They do not seem to realise that by their very publication they have vastly increased the difficulties of that trial or, indeed, made it impossible. For if it be a question of recovery or death, then the ethical situation produced by even inconclusive evidence may at times be such that the treatment cannot be withheld."

Faced with "very favourable" results from cortisone, the panel clearly followed these principles. And it prudently adopted retrospective controls, where the alternative was no controls. The panel cannot be faulted for its choice of retrospective controls or for the outcome of the trial. The panel opted for a method which was ethical and prudent in the circumstances. But serious, ethical questions could have been raised if the so-called standard on concurrent controls had been used.

Importance of retrospective controls

The panel was understandably dissatisfied with the comparison of the experimental group with the retrospective group, but the fault lay not with the concept of retrospective control but with the limited size of the historical sample available. As medical record keeping improves, and, in particular, as computers are used to collect, classify, and recover data, the use of retrospective controls is likely to increase and produce substantial benefits.

This is not to say that the use of retrospective controls requires such techniques. The fact is—and here we come perhaps to the nub of the matter—retrospective controls are used informally by every investigator in the early stages of assessing a new treatment. For only by comparing the new treatment with the conventional can the investigator gauge whether the new has any promise at all.

Thus, the strictures against retrospective controls do not eliminate their use but only discourage their deliberate (and optimal) use. Instead of searching out comparable cases from past records, the experimenter falls back on a subjective assessment of his own experience. More ironically, he is discouraged from consulting published reports on the accomplishments of conventional treatment for that too represents the use of retrospective controls. Perhaps the *reductio ad absurdum* of the contention that the use of retrospective controls is "inherently fallacious" is that it disqualifies the use of published reports for evaluating current experimental data—surely one of the prime functions of those reports.

Because retrospective controls are used in an impromptu way rather than formally, preliminary experiments are assessed less reliably than they might be. And experimenters' subjective judgments of their own innovations are likely to be more favourable than judgments reached through controlled retrospective analysis, and unpromising lines of investigations may be undeservedly prolonged.

Example

I want now to illustrate my argument by considering a practical case, the clinical trial of fast-neutron treatment.^{3 6} This trial is typical of randomised trials with concurrent controls; there is an experimental group and a control group of comparable size. The experimental group is given a standard course of fast-neutron treatment, and the control group is treated with x-rays, the conventional treatment, which is given by co-operating doctors.

So far, there are fewer than 100 patients in each group, and the patients have disease at different sites and at different stages. Thus each group is really many sub-groups, and, indeed, the number in each sub-group is not much larger than in the groups of the study of using cortisone to treat polyarteritis nodosa. Despite their smallness, each sub-group has been divided further into experimental and control groups of about equal size. In this trial the controls have been given conventional treatment, whereas in the cortisone trial the controls had no special treatment at all. This difference is important because it raises the question of how representative is the conventional treatment given to the controls of the treatment given generally. This question can be answered only by mounting another trial in which the control group is compared with a much larger group culled from many centres. But presumably the same practical considerations that led to small control groups will militate against collecting a large group to control the controls.

An alternative method of validating the results obtained with the control group is to compare them with retrospective controls, of which there are many. But if the reasonableness of this method is accepted then it is difficult to see why there is any need for a concurrent control group in the first place as the experimental group is effectively being compared with the retrospective controls.

Thus the arguments given for using retrospective controls in the cortisone trial also apply to the fast-neutron trial, and are reinforced by the additional need in the fast-neutron trial of validating the conventional treatment given to the control group. This conclusion implies that previous work with fast-neutrons that was not controlled may be evaluated by comparing it with retrospective controls.

Conclusions

I conclude that the use of retrospective controls is essential; and that neglect of their formal use is unwarranted and may lead to inappropriate optimism about the promise of a new treatment, increase the risk of later disappointment, and increase the time and cost of making definitive evaluations of new treatments.

Hill has pointed out that every method of control used in clinical trials has its defects.^{3 4} No one method is best in every case, and the choice of method rests on consideration of the options and their consequences. Those designing a trial should not succumb to fashion in experimental style and assume that a "strictly controlled" clinical trial is necessarily a trial that uses concurrent controls.

References

- ¹ Panel on Collagen Diseases and Hypersensitivity, *British Medical Journal*, 1957, **1**, 611.
- ² Hill, A B, *British Medical Bulletin*, 1951, **7**, 278.
- ³ Hill, A B, *Principles of Medical Statistics*. London, Oxford University Press, 1967.
- ⁴ Hill, A B, *A Short Textbook of Medical Statistics*. Philadelphia, Lippincott Co, 1977.
- ⁵ Catterall, M, Bewley, D K, and Sutherland, I, *British Medical Journal*, 1977, **1**, 1642.
- ⁶ Committee on Radiation Oncology Studies and its Particle Subcommittee, *Cancer Clinical Trials*, 1978, **1**, 153.

(Accepted 31 August 1979)