The controlled trial

SIR RICHARD DOLL
F.R.S., M.D., F.R.C.P.

Emeritus Professor of Medicine, University of Oxford and Honorary Consultant, Imperial Cancer Research Fund’s Cancer Epidemiology and Clinical Trials Unit, Radcliffe Infirmary, Oxford

Introduction

In 1945 I was looking for a job, after having spent a short post-war convalescence as a junior assistant in the Medical Unit at St. Thomas’s Hospital, when I had the good fortune to be introduced to Dr Avery Jones. He had just then had the idea that clues to the causes of gastric and duodenal ulcers might be obtained if differences could be found in the frequency with which they occurred in men and women in different occupations, particularly if the ulcers proved to be unusually common in transport workers, who, it was commonly believed, tended to be afflicted because of the irregularity of their meals. His clinical responsibilities, however, prevented him from conducting a survey himself and he obtained a grant from the Medical Research Council to employ a medical assistant to conduct one under his supervision. I was offered an appointment to work on the project and secured as a result, a life long friend and teacher with whom I collaborated actively in research for 24 years.

Our work on the occupational distribution of gastric and duodenal ulcers did not produce any major surprises, but it did establish the prevalence of each type of ulcer by sex and age and showed that factors of importance in the aetiology of the two types of ulcer were different. Duodenal ulcers were found to occur more or less evenly throughout the population, with some excess in men holding responsible positions at work, while gastric ulcers were found to occur most often in the poorest socioeconomic groups (Doll, Avery Jones and Buckatzsh, 1951).

Two years later, when the collection of data was complete, I obtained a post in the Medical Research Council’s Statistical Research Unit, but continued to work with Avery Jones at the Central Middlesex Hospital in an honorary capacity. At that time new methods of treatment for gastric and duodenal ulcers were being proposed with monotonous regularity and it was not difficult to compile a list of recommended treatments commencing with every letter of the alphabet. What was difficult was to assess whether any of them had any effect at all. So much so that some sceptical physicians even began to enquire whether the standard dietary treatment with milk and eggs had any advantage over normal food, and one carried his doubts so far as to treat 20 chronic gastric ulcer patients with daily injections of distilled water (Gill, 1947). The results showed that the patients’ pain disappeared and their ulcers healed as quickly on this treatment as in a control series treated previously on orthodox lines.

Controlled trials with random allocation

By this time, Avery Jones had come under the influence of Bradford Hill, who was now the honorary director of the MRC’s Statistical Research Unit and had been a member of the Council’s Committee which had supervised the conduct of our survey, and Avery Jones was convinced that the only reliable way to reach conclusions about the value of an ulcer treatment was to follow Hill’s prescription for a clinical trial, using concurrent controls and allocating the method of treatment to each patient at random. This type of trial first began to be used in 1946, when a committee of the Medical Research Council planned a trial to assess the efficacy of streptomycin in the treatment of pulmonary tuberculosis. Its characteristics were described succinctly in the Committee’s report (Medical Research Council, Streptomycin in Tuberculosis Trials Committee, 1948) and were elaborated as an article in the British Medical Bulletin by Bradford Hill (1951) who had been primarily responsible for the design of the Committee’s trial.

That controls were needed when introducing a new treatment had, of course, been appreciated for some years, although there were still some doctors who persisted in believing that they could assess the value of a treatment by observing the patients’ response and comparing it, in their own minds, with the effect of the vis mediatrix naturae with which their educa-
tion and experience had (as they thought) made them totally familiar. There was, however, no consensus about the type of controls to be used. Many sought them in patients who had been treated in other ways by other doctors in other places, in the patients they had themselves treated previously, or in concurrent patients who did not ‘volunteer’ for the new treatment. Others had come to realise that trials that relied on such comparisons were liable to give rise to false conclusions, because of differences in the character of the patients constituting each group and in the ancillary treatments they were likely to have received. They realised, therefore, that it was essential to compare groups of patients whose specific type of treatment had been determined uninfluenced by knowledge of the details of their particular state, other than that they fell within the general category of patients suitable for the specific treatment, and whose other treatment was managed as similarly as possible; and they had sought to achieve this by giving the treatment under investigation to alternate patients or to patients admitted on (say) certain dates in the month.

**Crucial characteristics**

The MRC’s trial had broken new ground and established a model to which (with minor modifications) the great majority of useful trials have subsequently tended to conform. Its chief features were four. First, it introduced a system of allocating treatment by a random method, the treatment being decided only after the patient had been admitted to the trial, thus avoiding any possibility that knowledge of the type of treatment the patient was going to receive could influence the decision to include him (as it could when treatments were allocated alternately or on specific days of the week or month). This was not, of course, a new idea as it had been used freely in agriculture and to a less extent in laboratory research; it was first proposed by Fisher (1926), but it was the first time that it had been used in the clinical study of patients.

Secondly, care was taken to define precisely the type of patient suitable for admission to the trial, thereby diminishing the likelihood of obtaining conflicting results in the future, facilitating repetition using other treatments, and setting limits to the extent to which the results could legitimately be generalized.

Thirdly, arrangements were made to ensure that the events that would be used to assess the results were recorded unbiased by knowledge of the treatment that the patient had received. This, it was recognized, could be achieved either by counting as end results precise events like death or survival, or by making the trial ‘double-blind’ so that neither the patient nor the investigator knew at the time of treatment which of the types under investigation the patient was receiving, or by ensuring that results that involved an element of subjective judgment, such as the extent of an improvement in X-ray shadows, were determined independently by individuals ignorant of the treatment schedule.

Fourthly, attention was paid overtly and in detail to ethical considerations, asking not only whether it was ethical to conduct the trial in the way proposed, but also whether it might not be unethical ‘not to have seized the opportunity to design a strictly controlled trial which would speedily and effectively reveal the value of the treatment’ (Hill, 1963).

Two of the questions posed have to be asked afresh in every trial and will be answered differently in different circumstances: namely, whether it is ethical to withhold the trial treatment from a control series, given that it is ethical to use it at all, and whether the control patients should be given placebos so as to enable the trial to be conducted ‘double-blind’. A third question, whether the physician in charge could modify the schedule if required in the patient’s interest, was answered categorically for all trials: namely, that the doctor in charge must always do for his patient whatever he really believes to be essential for that patient to return him to health.

**Trials inspired by Sir Francis**

The trials that were conducted at the Central Middlesex and St Mark’s Hospitals by some of Avery Jones’ pupils are listed in an appendix. They consistently followed the pattern laid down by Bradford Hill and led to notable advances in demonstrating the value of many treatments and the relative uselessness of many others, some of which had been used almost routinely for many years. Among the features of these trials, three are, perhaps, particularly relevant in 1984.

**Factorial design**

One was an innovation that was introduced early on and enabled more information to be obtained more quickly than would otherwise have been the case: namely, the use of a factorial design that allows two or more treatments to be assessed on the same group of patients. The technique was pioneered before the days of random allocation by Wilson, Pollock and Harris (1946), who used it to test two treatments for infectious hepatitis. In their study, alternate patients admitted to hospital were prescribed 5 g of cysteine a day and alternate patients within each of the cystine and control groups were given diets containing 200 g or 70 g of fat a day made up to isocaloric by manipulating the amount of carbohydrate. One hundred and seven patients were admitted to the trial, all of whom contributed to the
assessment of both the value of adding cysteine and the value of reducing fat. No difference in the rate of recovery was seen between patients on the different fat content diets, but there did appear to be more rapid recovery in the patients given cysteine, due principally to a lower relapse rate.

When combined with random allocation this technique requires each combination of each treatment and its corresponding control (four combinations if two treatments are to be treated, eight if three are to be tested*) to be allocated at random to successive patients. It is convenient, however, for ease of analysis and to secure maximum economy if the randomization is restricted to ensure that the numbers of patients given each combination of treatments are equalized periodically, and this was always the practice in the Central Middlesex Hospital.

In the series of trials of treatments for gastric ulcer in Avery Jones’ department between 1948 and 1972, 19 treatments were tested two or three at a time on a total of 440 patients, as is shown in Table 1. Had a factorial design not been used, each treatment would have been tested, on average, on only 23 patients; with the factorial design the average was 55. With this number (small though it is) it was possible to show that three treatments were beneficial (4 weeks’ bed rest in hospital against ambulant treatment with return to work as soon as the patient felt fit enough, stopping smoking, and the prescription of carbenoxolone), that one dose of carbenoxolone (100 mg t.d.s.) was better than another (50 mg t.d.s.), and that one diuretic (spironolactone) given to prevent the water-retaining effect of carbenoxolone counteracted the effect of carbenoxolone while another (thiazide) did not. The other 14 comparisons gave non-significant results. The treatments tested in these comparisons were not necessarily ineffective, but they did not appear likely to have large enough benefits in the treatment of a relatively benign disease to justify further study, or else they suggested that the special treatments were harmful, and they were abandoned.

The factorial design can, of course, be used only when the treatments under investigation are likely to affect the course of a disease in different ways. The combination of two drugs with similar pharmacological effects would not be appropriate because the combination would be the equivalent of increasing the dose of one and might cause undesirable toxic effects. Nor would combinations be appropriate if one treatment was ineffective in the presence of the other, or if two treatments interacted with each other so as to produce less effect than one treatment on its own. Such interactions are, however, rare and I have not come across any in practice, when care has been


<table>
<thead>
<tr>
<th>Treatments tested concurrently*</th>
<th>No. of</th>
<th>patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bed-rest v. ambulant, phenobarbitone, ascorbic acid</td>
<td></td>
<td>64</td>
</tr>
<tr>
<td>Milk drip, cabbage-juice, robaden</td>
<td></td>
<td>48</td>
</tr>
<tr>
<td>Milk drip, belladonna, robaden</td>
<td></td>
<td>32</td>
</tr>
<tr>
<td>Advice to stop smoking, bland v. normal diet, milk drip + alkali</td>
<td></td>
<td>32</td>
</tr>
<tr>
<td>Carbenoxolone, sulphated polysaccharide</td>
<td></td>
<td>64</td>
</tr>
<tr>
<td>Carbenoxolone, oestrogen</td>
<td></td>
<td>50</td>
</tr>
<tr>
<td>Carbenoxolone high dose v. low, oestrogen, spironolactone v. thiazide</td>
<td></td>
<td>46</td>
</tr>
<tr>
<td>Carbenoxolone analogue v. carbenoxolone, zinc sulphate, vitamin A</td>
<td></td>
<td>32</td>
</tr>
<tr>
<td>All combinations</td>
<td></td>
<td>46</td>
</tr>
</tbody>
</table>

*All patients given what was standard treatment at the time except when modification was requested by the special treatment. Comparisons made with standard treatment only unless otherwise specified.

taken to avoid combining treatments that are likely to act by the same mechanism—as, for example, two aneplatelet drugs may do when given to patients who have recently had a myocardial infarction.

The majority of new treatments under investigation, in fact, show little or no benefit in comparison with the best of the old, and one would indeed be fortunate if all the treatments in a factorially designed trial were beneficial. If they were, some quantitative interaction between them may well be expected and it is then a positive advantage to know what it is. A qualitative interaction that would completely prevent or reverse the effect of either treatment is, however, most unlikely and has yet to be encountered.

The overwhelming advantage of using a factorial design is the economy in the treatment of patients and the rapidity with which results can be obtained. Many controlled trials may take several years to complete, so that the advantage of being able to assess two or three treatments in parallel rather than seriatim is great.

Size of trial

Another feature of these gastroenterological trials was their size, which was generally small by modern standards. These standards often call for the treatment of many hundreds or thousands of patients and this is, indeed, essential if a disease is common, the end results lethal or seriously disabling, and the treatments under test are expected to yield, at best, only moderate effects (Peto et al., 1977; Peto, 1980).
Sir Richard Doll

To reduce by a fifth the fatality of myocardial infarction in patients admitted to hospital, from say 15% to 12%, would save several thousand lives a year if applied nationally and would make a bigger contribution to the expectation of life in the country as a whole than the complete elimination of many individual diseases. So would a reduction in the fatality of breast cancer by a similar proportion from (say) 50% to 40%. In neither case, however, could one expect to obtain any useful information from trials in 100 patients nor even (in the first example) from trials in 1,000. Small trials of therapies for these conditions are, in fact, more likely to mislead than to help. For, on the one hand, treatment that produces only a small benefit is unlikely to lead to a statistically significant result and, with bad luck, may even appear to be harmful. The treatment will, therefore, tend to be dismissed as ineffective, even though its application on a large scale might make a substantial contribution to national welfare. When, for example, Freiman et al. (1978) reviewed the results of 71 trials that were reputedly negative, they found that they had been carried out on such a small scale that real reductions in risk of 25 to 50% could easily have been missed. Treatments, on the other hand, that are totally ineffective will occasionally appear to produce a statistically significant benefit and these results are differentially liable to be published, because the benefit that is apparently produced in a small number of patients will inevitably be large.

The small trials that were carried out in Avery Jones' department were not, however, concerned with such crucial end-points as death, but with lesions or symptoms that could be graded and measured quantitatively, such as the size of an ulcer on radiological examination or the duration of remission in mild ulcerative colitis. Small changes in such quantifiable measures were clinically unimportant, while large changes could be expected from any treatment that materially affected the disease process. It was, therefore, reasonable to assess the value of treatment on small numbers of patients which varied in practice from 32 to 150. In some circumstances too, which are, however, uncommon in gastroenterological practice, it may be possible to use even smaller numbers effectively; as, for example, when the endpoint of interest is the suppression of symptoms that recur frequently and the patient can be used as his own control, receiving each treatment for periods of time in random order (Prichard et al., 1963).

When even moderate changes in outcome were likely to be important, as in the treatment of chronic ulcerative colitis of such severity that treatment was normally expected to require at least six weeks in hospital, Avery Jones and his colleagues appreciated that larger numbers of patients would be needed and collaborated in the conduct of a large scale trial under the auspices of the Medical Research Council, like those that had been so successful in determining the best combinations of drugs for the treatment of pulmonary tuberculosis. The results of this trial of 210 patients showed that cortisone increased the remission rate at the end of six weeks from 16% to 41% (Truelove and Witts, 1954 and 1955), a conclusion that has been amply confirmed many times since. It may be noted, however, that had it been necessary to judge the short-term results by death or survival, the number of patients would not have been large enough to produce a statistically significant result (five deaths in the cortisone groups against 11 in the controls, \( P=0.09 \)) while if the trial had been stopped at an early stage, as might have been recommended by an inexperienced ethical committee when the first four deaths all occurred in the treated group, the conclusion might have been the opposite of the truth.

Ethical issues

The attention paid to ethical considerations was one of the characteristic features of the controlled trial with random allocation when it was first conceived. This was not because random allocation created any special ethical problems in itself. The real ones with which the initiation of the modern trial were concerned were, in fact, already latent in all trials using concurrent controls who had not been self-selected by refusing to have the new treatment. The novelty of the procedure, however, added to the concern about the expansion of clinical research in general that had been fuelled by the atrocious experiments on prisoners that had been carried out in Nazi Germany and Japan during the war.

Avery Jones' pupils, who carried out so many controlled trials, worked on the principles set out by the Medical Research Council's Committee and subsequently summarized by Bradford Hill (1963) in his Marc Daniels lecture to the Royal College of Physicians. Many of these trials were, however, conducted in a way that would be at variance with the requirement to seek the patient's 'informed consent' to be included in the trial, that is now common. Such a requirement makes no sense unless it applies equally as a prelude to all medical treatment and it is difficult to believe that it is genuinely in the best interest of the patient always to explain the doubts that beset any medical advice about the efficacy of treatment and the complications of disease (Brewin, 1982). Both the World Medical Assembly (1975) and the Medical Research Council (1964), which have laid down that informed consent to inclusion in a trial should be obtained, provide exempt clauses which make it unnecessary in circumstances in which it is judged to be against the
patient's best interest to obtain it. Both, in these circumstances, require the decision not to seek informed consent to be approved by a supervisory group or an independent committee. Some circumstances in which approval will almost certainly be granted have been described elsewhere (Doll, 1983), but it is open to doubt whether ethical committees would agree with its dispensation if some of the trials listed in the Appendix in which it was not sought were to be repeated again.

What is not open to doubt is that several of the trials provided strong evidence in favour of the view that it is not so much the ethics of randomization that are in question, but those of prescribing treatment before randomized trials have been effectively carried out (Brewin, 1982).

References


Appendix

Reports of controlled clinical trials undertaken at the Central Middlesex and St. Mark's Hospital with the support of Sir Francis Avery Jones.

†1. DOLL, R. & PYGOTT, F. (1952) Factors influencing the rate of healing of gastric ulcers: admission to hospital, phenobarbitone, and ascorbic acid. Lancet, i, 171.


† These articles contributed data to those summarized in Table 1.
The controlled trial.

R. Doll

Postgrad Med J 1984 60: 719-724
doi: 10.1136/pgmj.60.709.719

Updated information and services can be found at:
http://pmj.bmj.com/content/60/709/719.citation

Email alerting service

Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

Notes

To request permissions go to:
http://group.bmj.com/group/rights-licensing/permissions

To order reprints go to:
http://journals.bmj.com/cgi/reprintform

To subscribe to BMJ go to:
http://group.bmj.com/subscribe/