TRIALS AND TRIBULATIONS

By

A. A. BAKER, M.D., D.P.M.
Deputy Physician Superintendent

and

J. G. THORPE, Ph.D.
Research Psychologist
Banstead Hospital, Sutton, Surrey

WITH the advent of effective physical therapies, psychiatrists have become increasingly interested in the problems posed by clinical trials. Professor Bradford Hill, in the Preface to the Sixth Edition of his *Principles of Medical Statistics* (1) noted that ". . . passage of time has . . . brought clinical trials into prominence and fashion . . .". More recently Professor Pickering has described the clinical trial as "the most important new method introduced into clinical science in recent years" (2). Some interesting differences between controlled and uncontrolled trials are well brought out in the paper by Foulds (3) in which he found that therapeutic success was claimed in 85 per cent. of studies not using control groups. Among those studies using control groups, however, only 25 per cent. obtained a positive result. He concluded that "claims for the success of a treatment are closely associated with absence of the means whereby these claims can be scientifically substantiated".

This paper will discuss the use of clinical trials in psychiatry and the particular problems posed by this branch of medicine. We will first consider the clinical problems, next statistical methods, and lastly some suggestions for the future.

Clinical trials only become relevant when a new and potentially effective therapy is available. They can be extremely simple in design—i.e. the treatment of G.P.I. by malaria only required that a group of patients who undoubtedly had the disease received this treatment. In the usual course of events the illness is progressive to death. Any treatment which halted or averted this must be effective without any need for a control group, since all other patients form a control. Moreover, the attitudes of nurses and doctors may colour the clinical picture but do not alter the final outcome.

Quite a different problem is the treatment of schizophrenia. This disease too has acquired an evil reputation—partly because so many of the patients hoarded away in mental hospials during the last century were suffering from it. With the advent of physical methods of treatment "cures" of this condition were claimed for 60 per cent. or more of the patients treated. If schizophrenia is an incurable disease, these are very good results, and no control group is needed. Schizophrenia however is not incurable and has a considerable spontaneous remission rate, and unfortunately a high relapse rate too. Allowing for a spontaneous remission rate of 30 per cent. in patients admitted to hospital (the natural remission rate of all schizophrenic illnesses occurring in the community is quite unknown), the clinics using physical treatment appeared to have an advantage. But active clinics attract treatable cases and it has become

https://doi.org/10.1192/bjp.105.441.1082 Published online by Cambridge University Press

apparent that some of the physical treatments are only effective in cases with a good prognosis, even without treatment. It is obvious that any treatment which is only given to those patients likely to recover will appear to be an effective therapy.

It is of major importance to realize that psychiatric patients respond to their human environment. The results of the same treatment will be different from hospital to hospital, and even from ward to ward. This invalidates the majority of studies comparing control groups in other hospitals or wards or in the same hospital at different times.

Even when comparable conditions for both treated and control groups are available, our problems have only begun. Again, to use schizophrenia as an example (although most other functional psychiatric disorders could be used)—there is no precise entity to treat. Some experts consider it is one disease, others as six diseases, and yet others a multiplicity of reactions to various stresses. There are considerable variations in the symptoms from country to country, and even within one country: i.e. the simple peasant girl will show a different clinical picture to the saucy cockney. The results obtained in a hospital serving a rural area will be different from those in a hospital serving an urban area.

Two groups of patients may differ in their ability to respond to treatment. Under present conditions this difference may not only be due to the usual variables of age, sex, heredity, etc., but will also be dependent on their previous exposure to therapy. For example, nowadays few patients arrive in hospital without having had a trial of one or other of the tranquillizing drugs. We are always treating, in hospital, a group of patients who have already been selected by their partial or complete failure to respond to this therapy in their home environment.

The criteria for results of treatment are at present unsatisfactory. We have no way of measuring the common symptoms of psychiatric illnesses, nor even the presence of many, with certainty. Discharge from hospital may depend as much on hospital policy as on the patient's condition. At times the most topsy-turvy occurrences are noted: i.e. the paranoid patient may discharge himself from hospital at the height of his illness, but ask to be re-admitted when he has improved, acquired insight, and realizes his need for help. In countries with no National Health Service the duration of stay in hospital may be a reflection of the ability of the patient to pay rather than on the true duration of his need. The duration of stay can also be partly determined by legal requirements rather than medical.

Attempts to measure behaviour are only of value if we know the circumstances under which the behaviour occurs and the attitudes and expectations of the other persons concerned. It is striking that the reports of clinical trials usually reflect the expectations of those in charge at least as much as the effect of the therapy. We can recognize the personality of our colleagues from the reports of their clinical trials. These may vary from the enthusiast with marked success (but with no control group) to the cynic with everything carefully controlled who finds no positive results—"the nurses were warned not to allow the trial to influence their attitude to the patients"!

It is not possible to conduct a clinical trial in such a way that the nurses concerned do not know something of the treatment and *also* something of the attitudes of those in charge. All the current physical therapies have such obvious effects and side reactions that they cannot be concealed. The results of treatment are always end results of interaction between patient, therapy and nursing staff. The response of the latter to an enthusiastic programme is usually enthusiastic

https://doi.org/10.1192/bjp.105.441.1082 Published online by Cambridge University Press

—to an ambiguous programme such as is presented by "double-blind" trials, it will be ambiguous—and this is likely to be reflected in the results. Those who insist that the nurse must not know which treatment is being given will tend to eliminate enthusiasm and produce an atmosphere in which nothing particular is expected—and such expectations are often fulfilled.

Reports in the Journals of new methods of treatment may be misleading, for the simple reason that in general only successful results are recorded. Many trials which show little benefit are either abandoned or not considered to justify reporting. Psychiatrists, as well as patients, live in and respond to a human environment. Who is there who wishes to be the first to report, or print a report, on a failure?

Some may feel that these problems are solved by applying a statistical analysis to the results of the trial. We will therefore examine the application of statistical methods and the logic of statistical analysis.

If we can regard the drug trial as an example of a scientific technique, an attempt is being made to establish a causal relation between drug and subsequent effect. This requires that two *identical* patients be chosen. One is treated with the drug, the other in an identical way except that the drug is absent. If the first patient improves, and the second does not, we can conclude that the drug caused the improvement, *without the application of a statistical analysis.* The essence of this method is that the patients should be identical, and also that the *only* difference between their treatments should be the presence or absence of the drug in question. This is impossible to achieve in practice and hence our technique has to be modified in the following way.

We choose not two identical patients, but two groups of patients. This is usually carried out by dividing a larger group into two by a random selection procedure. One of the groups is then treated with the drug, and the other is given the same treatment, but in this case the drug is absent. Next we apply statistical analysis to the difference between the final status of the two groups.

It is important to realize the function of the statistical analysis in a trial of this type, and this can best be done by noting the factors which could conceivably have contributed to the difference in status at the end of the experiment. There are three factors. The first, of course, is the presence or absence of the drug. The second is the possibility that the two groups of patients differed in their ability to respond to the drug, while the third is the possibility that the two groups of patients received different treatment over and above the presence or absence of the drug. The second and third factors are obviously error factors and are referred to by Lindquist (4) as Type I and Type II errors respectively. Type II errors include such factors as doctors' and nurses' expectations, the influence of relatives, and so on. Differences in final status can therefore be attributed to the drug, or to error, and it is most important to be able to determine the total amount of this error before we can tell whether there is any difference left between the two groups which we could attribute to the drug itself.

The statistical analysis which is applied to trials of this type will enable us to estimate the Type I error *only*. It will *not* tell us how much Type II error has contributed to the difference in final status. To take a practical example, suppose there is a 20 per cent. difference in relapse rate between two groups of patients and we wish to know whether this difference is statistically acceptable. The statistical analysis may indicate that a 10 per cent. difference in relapse rate could be expected from errors of the Type I variety. We could therefore conclude that, as the actual difference is twice as large as this, it is statistically

https://doi.org/10.1192/bjp.105.441.1082 Published online by Cambridge University Press

acceptable at about the 5 per cent. level of significance, and would therefore interpret the results as indicating that the drug was beneficial. We have entirely ignored the fact, however, that a further 10 per cent. difference in relapse rates might be attributable to Type II errors which, if added to the 10 per cent. difference due to Type I error, would account *entirely* for our observed 20 per cent. difference. In this case the conclusion which follows is that the difference in relapse rates between the two groups is *not* statistically acceptable.

Of course, the obvious precaution to be taken is to eliminate the Type II errors entirely in designing our experiment, and if we can be satisfied that they are completely eliminated, then our statistical analysis may be of benefit. It has been our experience, however, that it is difficult, if not impossible, to eliminate these errors for reasons we pointed out earlier.

In any case, even if we could eliminate them we would still be left with a fundamental problem. A point which is frequently ignored is that the very first step in setting up a controlled trial should be to collect two samples of patients in such a way that they are both representative of the population about which generalizations are to be established. Groups of patients for clinical trials are rarely selected in this way, and the more usual procedure is to divide a group of patients in a particular hospital at a particular time into two groups by a random selection procedure. This is *not* the same as selecting two groups at random from a definite population. In all statistical analyses, conclusions refer to the population from which the sample was drawn. In other words from the behaviour of the sample we generalize to the behaviour of the total population.

In some cases the fact that we do not work with strictly random samples may not be an important drawback, but in psychiatry, where there are no hard and fast rules for diagnosis, the drawback is a serious one. It means for instance, that a clinical trial of a new drug for schizophrenia which is carried out in a single hospital can tell us very little about the effectiveness of the drug in another hospital where the diagnosis of schizophrenia may have a different basis. In addition, even if there were complete agreement over diagnosis, the patients remaining in a hospital with an active discharge policy would differ from those in a hospital with a custodial care policy. Worse still, as clinical trials of this sort are usually carried out in a single ward from which two groups of patients are formed on a random basis, we are not even able to generalize over the population of patients in the same hospital.

It follows from this argument that the statistical approach as it is usually applied has added little to the scientific approach outlined above. By randomizing the patients of one ward into two groups we have probably achieved some measure of success in equating the two groups on their ability to profit by the treatment under trial, though the degree of success is usually unknown. We have no justification in regarding them as random samples of a population and unless this is so we have no justification in applying statistical analysis at all. Even if we had, the statistical analysis would never give us an estimate of the Type II errors contributing to the final difference.

If the results of clinical trials in psychiatry are to be applicable to all hospitals and all patients in the country, then it will be necessary to organize a national research project. If the problem is to evaluate a new treatment for patients with schizophrenia, the experiment should be carried out using the total schizophrenic population of hospitals selected on a random basis. The patients from each hospital should then be divided into two groups on a random or matching basis. One group should be given the new therapy and the other

https://doi.org/10.1192/bjp.105.441.1082 Published online by Cambridge University Press

a form of treatment as nearly matching the new in its exterior presentation as possible. The same doctor and nursing team should be concerned with both groups, and each should receive identical amounts of doctor and nursing time.

If the total groups in the hospitals concerned were too large, as is probable with most psychiatric illnesses, a random selection or representative sample from each should be obtained. It would be an advantage if the group was of such size that the treatment project could be carried out on one ward.

The application of analysis of variance to the results of such a national trial would enable generalizations to be made to cover all the hospitals in the country. Furthermore, it would allow an examination of the statistical significances of the differences in response to treatment exhibited by the various hospitals in the sample. It would provide clues to the types of patient likely to respond and the type of environment which would enable patients to improve. Hospitals not in the sample would be able to consider which of the hospitals in the experiment they most closely resembled, in order to decide how their own patients would probably respond to the new treatment.

It could be argued that such an elaborate trial is quite unnecessary, as the various manufacturers presenting new products already obtain the co-operation of a large number of different hospitals. There is an obvious fallacy in assuming that the results from hospitals interested in trials of new treatments will be the same as those with less interest, nor are they a random sample, which invalidates a statistical assessment.

There are very serious difficulties in organizing a national trial. Full approval of the proposed new treatment would have to be obtained in advance from all the hospitals and doctors. This would seem to be most unlikely under present conditions. Even if such a national trial became probable, we must point out that it has one serious defect, which would limit its value. This is inherent in the nature of psychiatric practice. Each hospital develops its own unique human atmosphere, determined by the medical and nursing staff. Patients react to this atmosphere and the extent of the success or failure of new treatment regimes will be determined by it. Because of this, each hospital must conduct its own research programme, quite apart from the results of any national trial (which might give useful guidance), or the present haphazard trials.

Most trials using placebos or variations on the double-blind technique are trying to eradicate the influence of staff attitudes. Although, in theory, this should give more information about the effects of the treatment alone, in practice no psychiatric treatment is used in such a manner. We have already noted that so-called "blind" trials fail to deceive the nurses. Apart from this, these trials contain a major fallacy insofar as they are intended to be applied to ordinary clinical practice. We do not want to know the effect of a new treatment used in an atmosphere of uncertainty, but its effects in a normal atmosphere of therapeutic hope.

We therefore suggest that the nursing and medical staff should have full knowledge of the trial in advance. A record should be made of their instructions and expectations before, during and after the trial. We should consider using our ingenuity in devising methods of recording human expectations and hopes rather than in developing techniques for deception. Apart from other assessments we should always record patients' illnesses in terms of their relationship to each other and to the nursing staff. As the trial progressed we would record changes in patients' attitudes to the nurses and vice versa. At the end of the trial the mean scores would be compared and experiments should be repeated on another ward with another nursing team. If this confirmed the findings on

https://doi.org/10.1192/bjp.105.441.1082 Published online by Cambridge University Press

the first ward, it would be highly probable that the results were reliable. If there were differences in the results, the experiment had not been accurately duplicated. The differences may of course be in the patients themselves, nursing attitudes or other variables. It should be noted that statistical analysis is not applied to this type of trial.

In conclusion we must point out that psychiatric patients present many special problems for those conducting clinical trials. In psychiatric illness, the human environment is of vital importance. No surgeon would consider comparing the results of operations, some of which were conducted under modern aseptic conditions, and others in a country cottage. Similarly, psychiatrists cannot compare results of clinical trials in differing human environments.

Reports on clinical trials are of limited value unless further information is given not only about the design of the trial but also of the total physical and human environment in which it takes place.

Lastly, the statistician should not be expected to apply his special techniques to a trial in which the design and method do not justify such action.

References

1. Bradford Hill, A., *Principles of Medical Statistics*. 6th Edition, 1957. London, The *Lancet* Ltd.
2. Pickering, Sir George, "Medicine's Challenge to the Educator", *Brit. Med. J.*, 1958, pp. 1121–1124.
3. Foulds, G. A., "Clinical Research in Psychiatry", *J. Ment. Sci.*, 1958, **104**, 259–265.
4. Lindquist, E. F., *Statistical Analysis in Educational Research*, 1940. Chicago: Houghton Mifflin Company.

https://doi.org/10.1192/bjp.105.441.1082 Published online by Cambridge University Press