

RES MEDICA

Journal of the Royal Medical Society



The Use of Controls in the Assessment of Clinical Evidence

C. Vaughan Ruckley
M.B., CH.B.

Abstract

The prize-winning essay from the B.M.A. Essay Competition for Provisionally Registered General Practitioners. 1961. The essay is abridged.

The incorporation of standards or controls into clinical experimentation has become, over the past decade, a widely accepted practice. So much so that there is a danger that the medical profession in general may become uncritical of the practical and, more important, the ethical problems involved. The subject should be under constant review.

Are controls really necessary? How best can they be employed? How is the resulting evidence to be assessed? Under what circumstances can the use of controls be justified on ethical grounds?

Copyright Royal Medical Society. All rights reserved. The copyright is retained by the author and the Royal Medical Society, except where explicitly otherwise stated. Scans have been produced by the Digital Imaging Unit at Edinburgh University Library. Res Medica is supported by the University of Edinburgh's Journal Hosting Service: <http://journals.ed.ac.uk>

ISSN: 2051-7580 (Online) ISSN: 0482-3206 (Print)

Res Medica is published by the Royal Medical Society, 5/5 Bristo Square, Edinburgh, EH8 9AL

Res Medica, Autumn 1961, 3(1): 39-46

doi: [10.2218/resmedica.v3i1.376](https://doi.org/10.2218/resmedica.v3i1.376)

THE USE OF CONTROLS IN THE ASSESSMENT OF CLINICAL EVIDENCE

By C. VAUGHAN RUCKLEY, M.B., CH.B.

The prize-winning essay from the B.M.A. Essay Competition for Provisionally Registered General Practitioners, 1961. The essay is abridged.

The incorporation of standards or controls into clinical experimentation has become, over the past decade, a widely accepted practice. So much so that there is a danger that the medical profession in general may become uncritical of the practical and, more important, the ethical problems involved. The subject should be under constant review.

Are controls really necessary? How best can they be employed? How is the resulting evidence to be assessed? Under what circumstances can the use of controls be justified on ethical grounds?

THE NEED FOR CONTROLS

Progress in medicine depends upon experiment. Formerly the physician based his belief in the efficacy of his simples and mixtures upon either the dogma of his mentors, hallowed by time and seldom criticised, or his own observations upon individual patients. This has led to the perpetuation, even into this supposedly enlightened age, of many remedies whose true worth has never been accurately assessed.

Although the majority of the remedies inherited from the 19th century and beyond have been discarded with the revision of the pharmacopoea in 1955, the need for objective assessment of the drugs we prescribe has never been greater. The reasons for this are twofold.

The drug houses are pouring out every day a flood of new preparations, many of which are either pharmacologically identical or differ, one from the other, by only a small degree in action or in side effects. This is too much for the practitioner to sort out for himself. Even if he had the time and energy he might not by now have the inclination, for the extravagant claims made on behalf of so many products can breed a cynical indifference, and a reversion to well-tried medicines.

He needs therefore a guide for speedy reference. Though he may not be trained in statistics, if he knows that a drug has been assessed by a standardised scientific procedure, such as a controlled trial, his problems and his scepticism are much diminished.

The second reason is an economic one. That the best treatment should speedily be selected from a multiple choice is very relevant to the cost of the Health Service in this country. Not only should the best be propagated, but the less effective products should just as quickly be discarded.

In diseases where no effective treatment exists, especially when they are either so trivial as to inconvenience the patient only a little, such as the

common cold, or are highly malignant, such as acute leukaemia, there may be no necessity or indeed justification for the inclusion of a control group in a trial.

THE CONCEPT OF CONTROLS

The first clear definition of clinical controls has been attributed to Laplace in 1814. "To determine the best of several treatments it is sufficient to try each of them on the same number of patients, keeping all conditions constant . . . the superiority of the most beneficial treatment will manifest itself the more, the greater the number of cases."

This concept has gained widespread acceptance only during the past ten years. It continues to have its foes. Basically it involves the application of a statistical method to experimentation with human subjects in a manner previously applied only to animal studies. Planning the experiment is now recognised to be a major undertaking, necessitating from the start having a clear grasp of the questions it is hoped to answer, and proceeding to detailed specification of the kind of subjects to be included, the treatment to be compared and the measurements to be taken. The points at which different types of trial are particularly vulnerable to bias are more widely appreciated, as is the value of such precautions as randomisation, "blind" comparisons and the inclusion of placebos. The number of successfully completed trials is growing and they serve as models for future work.

ERROR

With the exceptions already mentioned, the inclusion of controls is essential for every kind of clinical trial. Where no control is exercised other than the subjective response of the patient and the clinical impression of the physician, the scope for error is great.

Many investigations into observer variation have emphasised the need to substitute the experimental for the observational approach.

Professor Hill has pointed out, however, that "within the framework of a clinical trial designed to contrast one group with another there is nothing whatever to inhibit the highly gifted clinical observer from observing . . ." The observational and experimental techniques are not mutually incompatible; indeed I believe one advantage of the development of controlled trials in Great Britain has been the need for, and the attempt to improve, observations of disease and disease processes.

"The emphasis placed on objective measurements," writes Oswald Savage, "has already resulted in more careful and accurate studies in these chronic diseases and has already produced new observations on the natural history of conditions such as arthritis."

The principle of therapeutic control is to provide for the group of patients who are to undergo a new and yet untried treatment a parallel group of cases similar in all respects, that is as regards all possible contributing causes except the one factor of treatment; or if the behaviour of the disease in a particular patient is to be observed before and after treatment, to make conditions the same in both periods except again for the factor of treatment.

Some sources of error may be unsuspected until the control method brings them to light. They represent all conditions which may be important for the origin as well as for the further development of the disease. Sources of error, or "concurrent causes," are divided by Herdan into 'internal' and 'external' causes.

Among the internal causes are firstly, the characteristics of the patient.

These include sex, age, nutrition and genetic constitution—all of which may influence profoundly his response to disease and its therapy.

Secondly, there are the characteristics of the disease itself. At different periods of time and in different subjects, particularly in chronic diseases, or in trials covering several months or years, the disease may vary in its stages, in its degree of severity, and in the virulence of an infecting organism. Delay in treatment will also allow evolution in the disease process, and will alter its response to treatment.

Among external causes are firstly environmental factors such as financial conditions, family affairs, conditions of employment and various physical disturbances which may substantially affect the progress of the disease.

Causes may be introduced through the treatment itself. These include the effects of transition from home to hospital — the very fact that a patient is moved into a hospital environment can evoke new mental and physical responses, as also can the attitudes of doctors and nurses. For example, in an organic disease such as ulcerative colitis where psychical factors are known to play an important part, even a dummy tablet, if recommended with conviction and enthusiasm by the physician, may evoke a perceptible clinical improvement. Thus in certain trials neither the doctor nor the patient should know whether he belongs to the trial or the control group, nor even, in the case of the patient, that such a dichotomy exists; hence the use of the 'double blind' method and dummy tablets.

It is not correct to assume that by introducing clinical controls into a trial we automatically eliminate error. However large a trial, and however similar control and trial groups may be, owing to the operation of concurrent causes it may lack generality. That is to say that the results, although true for a particular time, place and trial, may not be universally true.

The science of comparative therapeutics involves the study of just such larger problems. However, comparative therapeutics apart, it is probably true to say that the larger and more widespread a trial, the more reliable are its conclusions. Examples of two such experiments are the M.R.C. trial of ACTH, cortisone and aspirin in acute rheumatic fever, 1955, which involved the co-operation of many centres in the United Kingdom and the United States, and secondly the trial of the Salk polio vaccines which involved many thousands of patients.

THE CONSTRUCTION OF CONTROLLED TRIALS

According to the type of disease we are studying, so a particular type of trial is adopted and a particular statistical method applied to the analysis of the results. There are two broad categories: "within-patient" trials, where the patient is his own control, and "between-patient" trials, where control and trial are different patients.

It seems agreed by most authorities that in clinical trials, as in experimental pharmacology, when feasible, comparisons of treatments should be made within subjects rather than between subjects. If each subject receives all the treatments, variations in level of response from subject to subject cancel out when treatment averages are compared. The gain in precision is often striking (5).

A within-subject trial is a simple example of grouping. The general principle is to divide the patients into groups such that differences between groups represent important sources of variation that may inflate the experimental errors—that is, they are stratified according to known concurrent causes—then, if the experiment can be conducted so that each treatment is represented

equally often in every group, differences between groups are automatically eliminated from the comparisons of the treatment averages.

There are various designs by which grouping may be applied. A basic form is the simple cross-over pattern where the patients, and the duration of the trial, are divided into two for the comparison of two drugs. For example, 100 patients are included for the testing of 2 drugs over a period of 6 months. For the first 3 months 50 patients, who may either be randomly selected or subgrouped according to some common characteristic, are given drug A and 50 patients drug B. For the second 3 months the treatments are reversed. The patients are their own controls.

The same principle can be applied in various more complex designs. An example of a trial may illustrate the method. The relative values of Phenmetrazine and Dexamphetamine in the management of obesity have recently been compared (6). A 'double blind' procedure was used. Three series of tablets of similar appearance were prepared:— Phenmetrazine, Dexamphetamine and a control placebo, containing principally Lactose. A crossover technique was employed on the latin-square pattern.

1. A.B.C. 2. B.C.A. 3. C.A.B.

The patients were allotted to a particular treatment by a random sequence using a table of random numbers (Bradford Hill, 1955). Thus the two treatments under review were compared one with the other, and in addition, each patient was his own control. With the use of such control grouping far more information can be gained than simply the quantitative success of one treatment over another.

All patients throughout the trial were on a similar low caloric diet. As is always the case with the initiation of a reducing diet, regardless of the drugs given, the weight loss in all groups was greater during the first 6 weeks, due to initial dehydration. Since the precise effect of a treatment depended upon two things, namely the kind of tablet taken and whether it was administered in the first, second or third period, a more accurate result could be obtained if these two influences were disentangled. Grouping made this possible.

Thus an important concurrent cause could be eliminated, namely the evolution of the disease process at a particular time in the trial—in this case the regression of weight loss—and appropriate statistical correction terms could be applied. The conclusions drawn at the end of such a trial are therapeutically and statistically reliable.

In general, it is seldom possible to apply these within-subject methods to acute or rapidly progressive diseases. In fact, in clinical practice, as Doll has pointed out, the opportunities are few for using the patient as his own control because "when one drug has been tested the patient's condition is likely to be so altered, whether by the drug or by nature is immaterial, that nothing is to be gained by repeating the treatment with another drug." (7).

The majority of trials therefore have to be conducted by giving trial and control drugs to different patients, that is between-subject trials. The groups of patients should be similar in all respects except the treatment they receive.

In the past, controls were often retrospective, that is to say, attempts were made to compare the effects of a new drug on a group of patients over a period of time with the effects of a previous treatment and a different group of patients. This, as Armitage (8) and others have pointed out, is a practice which should be avoided because other factors may be operating.

For example, patients treated with anticoagulants for coronary thrombosis have been compared with 'control' groups treated prior to the introduction of anticoagulants in 1948. It has been shown that these two groups can not be fairly compared. The very fact that a specific treatment for thrombosis was

available meant that after 1948 a larger number of coronary thromboses of only moderate severity were admitted to hospital than previously. This alone was sufficient significantly to improve the survival rate of cases treated in hospital.

To avoid these and other difficulties in between-patient trials a variety of methods of random selection have been devised. Randomisation precludes the possibility of conscious or unconscious bias on the part of the clinician who is allotting patients to the trial.

Randomisation does not ensure that all groups are exactly equal; nothing can do this. It does ensure that they differ by an extent that is predictable and can be allowed for in the statistical analysis.

It may be thought preferable to take into account all the concurrent causes over which we have some control by methods of sub-grouping or stratification. This is particularly valuable where the numbers involved are small and where it is less likely that 'by chance' the two groups will be homogeneous. This sub-grouping can be incorporated into the system of alternates—known as compensating alternates—or can be combined with the random numbers method, especially if more than two sub-groups or strata are contemplated.

Two or more groups which at the commencement of a trial appear to be quite comparable may show, on reassessment at the end, to differ in important respects. In the M.R.C. trial in 1955, some 500 children were divided into three groups to compare the value of ACTH, cortisone and aspirin in the treatment of acute rheumatism. It was found that most of the concurrent causes were strictly comparable, but a marked difference happened to be present between groups in the numbers presenting with chorea (ACTH, 5.6%: cortisone, 11.4%: aspirin, 15.5%) and with congestive heart failure (ACTH, 14.2%: cortisone, 9%: aspirin, 6%). This difference is statistically significant and could therefore influence results, and it illustrates the value of sub-grouping.⁽⁹⁾

Doll points out, however, that it is seldom necessary to have more than a few subgroups, and that if it seems necessary to have a large number it suggests that the treatment is being tried on too heterogeneous a group of patients.

A statistical method which has been increasingly used in recent years is Sequential Analysis. The usual practice in trials is to postpone conclusions until final measurements from all subjects have been gathered. In a sequential trial on the other hand, a continuous statistical analysis is made as the data from each subject comes in. The trial is stopped as soon as the analysis indicates a clear-cut verdict of statistical significance.

This method may allow a reduction in amount of experimentation by 10 - 50% as compared with a fixed-size trial of the same discriminating power. For its suitable application there are certain requirements:—

- (i) Patients enter the trial in sequence over a period of time.
- (ii) Results should be quickly available.
- (iii) There should be potent reason for wishing to stop the trial as soon as possible.
- (iv) The primary object should be to perform a test of significance between new and standard treatments rather than a quantitative assessment (10).

The controls in sequential trials are the patients on the standard treatment.

Sequential Analysis is the method used in the Stilboestrol-Oestriol trial in metastatic mammary carcinoma, which is currently being conducted in Edinburgh. It provides a good illustration of the technique.

It is designed to show whether there is any significant difference between therapy with a synthetic oestrogen, stilboestrol, and a naturally occurring one,

oestriol. The former is the standard control. Patients are allotted to one or other group by a 'double blind' method. The tablets, similar in appearance, are distributed by the sealed envelope system using a table of randomised numbers and applied by a statistician.

In order that progress can be assessed with some accuracy the criteria of admission to the trial are clearly defined, principally definite skin recurrence and/or metastases clearly seen on X-ray.

Each patient is seen at monthly intervals. The nosographic criteria of progress are based on X-ray changes, measurements of skin lesions, including photographic records and biochemical tests. They are assessed not by one but by a panel of doctors. As soon as it is clear that a patient is deteriorating despite the hormone therapy she is withdrawn from the trial.

THE STATISTICAL EVALUATION OF RESULTS

The construction of therapeutic trials implies a twofold control. The two control concepts may be distinguished by the names of 'clinical controls' and 'chance controls'.

Having arrived at the end of a trial with a quantitative difference in results between trial and control groups, the question then arises as to whether such differences can be regarded as significant. The alternative is that the differences observed are caused by fluctuations of the many causes which comprise 'chance'.

Hence just as important as the clinical control is the control of the observed differences, and this is affected by application of a calculated standard error or standard deviation.

There are three main types of fluctuations or variations of numerical quantities which obey the laws of chance:—

1. Random sampling fluctuations of numbers and of relative frequencies (percentages).
2. Biological variations of counts and measurements.
3. Experimental and observational errors.

Once a difference has been observed between the findings in the control group and in the trial group, a statistical significance test must be applied. The appropriate test depends upon the particular nosographic criterion by which the effect of treatment is being assessed. For example, differences in relative frequencies or of percentages of outcome are assessed by the standard error of a percentage and by the chi-square test, and differences between average duration of a disease by the standard error of the mean test.

For most practical purposes, having established a distribution curve from one control group we can draw an arbitrary line at a distance of two standard deviations from the mean. If an observed value from the trial group lies outwith this line it has a 95% chance of being statistically significant. If it lies beyond thrice the standard deviation it has a 99.7% likelihood of being significant.

ETHICAL PROBLEMS

The foregoing discussion has centred upon some of the problems which arise with the use of controls, their application and assessment. Considerations remain of far greater importance. They arise in clinical experimentation because of our belief, with Kant, that every human being, irrespective of his mental or social stature, should be regarded as equally important.

It has been said that many of the great men of history were great because they were bigots; that is to say they could only see one side of a problem, their

own side. In the same way, anyone, whatever his sphere of activity, dedicated in the pursuit of a policy or line of investigation, may become blind to points of view differing from his own.

Since this may be a quite subconscious process, such a person may bitterly resent the imputation that he is not fit to assess objectively the ethical implications of his own work.

Since the use of controls has assumed an integral part of clinical experiments it is very relevant, in discussing their use, to consider the ethics upon which human experimentation as a whole is based.

Can we justify experiments on humans at all? I have discussed earlier some of the material reasons why controlled trials are necessary in the propagation of new therapy. These trials cannot be confined to experimental animals. Findings in other species may have general or specific validity for man, but the ultimate establishment of such validity must rest in each instance upon direct observations upon man.

Shimkin has advanced important considerations for the justification of such experimentation. He points out that an unwillingness to experiment carries with it as much moral responsibility as active experimentation. He says, "... to do nothing, or to prevent others from doing anything, is itself a type of experiment..." and goes on:

"As much knowledge and as weighty reasons are required for one course of action as for the other, and it should be demonstrated that the proposed experiment is more dangerous or more painful than the known results of inaction" (11).

The Nuremberg Medical Trials resulted in the formulation in 1947 of the oft-quoted rules which serve as a guide, though not necessarily as an infallible credo, for the experimenter. Shimkin has reduced these rules to two primary principles:—

1. "Investigators must be thoroughly trained in the scientific disciplines of the problem, must understand and appreciate the ethics involved, and must then be competent to undertake and carry out the experiment.
2. The human experimental subjects must agree to the procedure and must not be selected upon any basis such as race, religion, level of education or economic status. In other words, the investigators and their subjects are human beings with entirely equal, inalienable rights that supersede any considerations of science or general public welfare."

How can the use of patients as controls be reconciled with such principles? Guttentag has drawn a distinction between the 'physician-friend' and the 'physician-experimenter'. The former has a personal relationship with his patient, sharing his distress and wanting to assist him. "Objective experimentation to confirm or disprove some doubtful or suggested biological generalisation is foreign to this relationship. This would involve taking advantage of the patient's cry for help and of his insecurity" (12).

There is no doubt that among medical research workers in this country there is a tendency to regard patients as experimental objects, rather than human beings, and since this attitude is principally to be observed in the teaching hospitals it may well be spreading.

The responsibility for care of the patient, and for intensive investigation or experimental procedures on that patient, should be in the same hands, not necessarily the hands of an individual but preferably of a group. The more extensive and potentially dangerous the experiment, the more widely should the responsibility be divided. The nature of the doctor-patient relationship is such that nothing ought ever to be done to the patient except to his direct advantage, unless he gives his consent. As Fox has put it, "it is this consent

that changes his status to that of volunteer, in which he becomes the legitimate object of experiment." (13)

Greiner has listed the circumstances under which he considers that a compromise may fairly be reached.

"In the first place, the risk of drug toxicity must not exceed that run by thousands of patients under conventional treatment. In the U.S.A. this means drugs and doses that have already been accepted by Government agencies.

"Secondly, each patient-participant shall have a medical condition which, in the ordinary course of events, would be treated by the test agent or another drug with similar action.

"Thirdly, the experimental design shall not permit deterioration of the medical condition.

"Finally, it is apparent that a number of trials would be impracticable if the patient understood exactly the nature of the experiment, or even that he were participating in an experiment at all. A rigid ethical ideal may be too restricting when facing real-life problems." As Greiner succinctly puts it, "a workable standard appeals to me even more than lip-service to a remote principle, however perfect." (14)

The very fact of an experiment being necessary implies that one treatment is inferior to the other, and therefore that one group of patients will derive more benefit than the other. As long as it is only known for certain to be true in retrospect, then the experiment can be entirely justified on ethical grounds.

While the trial group receive the new treatment the control group are given the best orthodox treatment. Crofton has said that since the control group receives the best treatment previously available, it is the safest group from the patient's point of view to belong to. "He may be denied the still improved advantage of the new treatment but he avoids its possible side effects. And, of course, the new treatment may prove to be inferior to the old."

Finally, if we reject controlled clinical trials as ethically inadvisable, what are we to propose as the alternative? Without controls, trials are unreliable; without trials the distribution of drugs may be haphazard and potentially dangerous. To quote from Hill, "in many settings the carefully designed controlled trial is far more ethical than the uncontrolled experimentation with unproven products to which patients are frequently exposed." (15)

It is not sufficient simply to take a passive role nor to obstruct those who are active in the development of new experimental methods. We have an obligation not merely to reap what has been sown, but to plough-under a little more untilled land for the next crop.

REFERENCES

1. 6. W. G. COCHRAN. "Newer Statistical Methods." Symposium on Quantitative Methods in Human Pharmacology and Therapeutics, Ed. Lawrence. Pergamon, 1959, 119.
2. J. HAMPSON, J. LORAIN, J. A. STRONG, *Lancet*, June, 1960, 1265.
3. 5. R. DOLL. "Practical Problems of Drug Trials in Clinical Practice." Symposium on "Quantitative Methods in Human Pharmacology and Therapeutics." Ed. Lawrence. Pergamon, 1959, 213.
4. P. ARMITAGE. "The Construction of Comparable Groups." Conference on "Controlled Clinical Trials." Chairman, A. Bradford Hill. Blackwell, 1960, 14.
7. M. B. SHIMKIN. "The Research Worker's Point of View." Symposium on "The Problem of Experimentation on Human Beings." *Science*, 1953, 117, 205.
8. O. E. GUTTENTAG. "The Physician's Point of View." Symposium on "The Problem of Experimentation on Human Beings." *Science*, 1953, 117, 207.
9. T. L. FOX. "The Ethics of Clinical Trials." Symposium on "Quantitative Methods in Human Pharmacology and Therapeutics." Ed. D. R. Lawrence. Pergamon, 1959, 222.
10. T. GREINER. Symposium on "Quantitative Methods in Human Pharmacology and Therapeutics." Ed. D. R. Lawrence. Pergamon 1959, 231.
11. A. BRADFORD HILL. "Aims and Ethics." Conference on "Controlled Clinical Trials." Chairman, A. Bradford Hill. Blackwell, 1960, 7.