

6-1984

## Clinical Trials: I. Design and Ethical Issues

Barbara Tilley

Follow this and additional works at: <https://scholarlycommons.henryford.com/hfhmedjournal>



Part of the [Life Sciences Commons](#), [Medical Specialties Commons](#), and the [Public Health Commons](#)

---

### Recommended Citation

Tilley, Barbara (1984) "Clinical Trials: I. Design and Ethical Issues," *Henry Ford Hospital Medical Journal* : Vol. 32 : No. 2 , 147-151.

Available at: <https://scholarlycommons.henryford.com/hfhmedjournal/vol32/iss2/13>

This Article is brought to you for free and open access by Henry Ford Health System Scholarly Commons. It has been accepted for inclusion in Henry Ford Hospital Medical Journal by an authorized editor of Henry Ford Health System Scholarly Commons.

## Special Series

### Clinical Trials: I. Design and Ethical Issues

Barbara Tilley, PhD\*

This is the first in a series of four articles focusing on clinical trials which will appear in this Journal in the next several issues. This article will present some ethical and design issues. The second article will provide information on sample size and randomization; the third, on stopping rules and analyses; and the final article, on management issues.

#### Definition and History

Clinical trials are defined, for this series of articles, as a "prospective study comparing the effect and value of interventions against a control in *human subjects*" (1). Clinical trials involving animals will not be discussed.

One of the earliest clinical trials was carried out by John Lind in 1747 among sailors who had scurvy. Twelve patients were given six treatments including vinegar and cider. Two of the 12 were given only oranges and lemons and were soon well enough to nurse the others (2). In the 1930s, Fisher developed methods for statistical inference based on random allocation which he applied to agriculture; these methods were the foundation of modern clinical trials. The first reported randomized clinical trial was carried out by Amberson who tested a therapy for tuberculosis; patients were allocated by a flip of a coin. Hill is credited with being the first to recognize not only the importance of randomization but its usefulness for valid statistical evaluation. Hill's first trial in 1944 was to test a remedy for the common cold, and the second was to test streptomycin as a cure for tuberculosis (3).

#### Rationale

Clinical trials are performed for many reasons. First, given the uncertainty about the course of a disease and the variations in biologic measures, it is usually impossible to say, on the basis of uncontrolled clinical observations, whether a new treatment makes a difference in outcome. Chalmers gives the following example from his observation of medical practice. A man with gastrointestinal hemorrhage dies without surgery. The second man with the same problem has early surgery and does well. The next also has early surgery and does

well. The fourth man has early surgery and dies, so the following patient has late surgery because of the experience with the previous patient (4). Randomized clinical trials allow decision making based on a comparison of past experiences rather than on one or two recent events in highly variable patients.

Clinical trials are also useful to determine the incidence of adverse events. Without clinical trials, drug toxicities, especially rare events, could go unnoticed. For example, 3,038 patients were studied in a trial of practolol, a beta blocker given to prevent recurrent myocardial infarction. Two patients were observed to have sclerosing peritonitis and three to have oculomucocutaneous syndrome. These symptoms were unusual enough to discourage the use of the drug, but it took a trial of this size to detect these problems.

Third, events which occur naturally in the general population could go unnoticed. In the Coronary Drug Project (1), a long-term, lipid-lowering study in men with coronary heart disease, cardiac arrhythmias were noted on the annual 12-lead electrocardiograms. This finding was not surprising, but the percentages differed by treatment group. In the clofibrate-treated group, 33.3% experienced arrhythmias, as compared to 32.7% in the niacin group, and 28.2% in the placebo group. These differences were statistically significant.

Finally, adverse events thought to be attributed to therapy may be shown to be independent. Again, in the Coronary Drug Project, clofibrate was thought to be associated with nausea, but when the placebo group was compared to the clofibrate group, there was no difference in nausea.

#### Design

Cox (5) outlines the requirements for a good experiment which can be applied to clinical trials. These re-

---

Submitted for publication: March 26, 1984

Accepted for publication: August 20, 1984

\* Department of Research, Division of Biostatistics, Henry Ford Hospital

Address reprint requests to Dr. Tilley, Division of Biostatistics, Henry Ford Hospital, 2799 W Grand Blvd, Detroit, MI 48202.



quirements include the absence of systematic error, precision, acceptable range of validity, simplicity, and calculation of uncertainty.

#### **Absence of systematic error**

One of the best ways to eliminate systematic error is to select the appropriate control group. Clinical trials also must have carefully defined protocols to which investigators strictly adhere. Where possible, the studies should be blinded; that is, neither the investigator nor the patient should know which treatment the patient will receive so that the patient or the investigator will not influence the trial outcome. Finally, follow-up efforts must be the same for both groups, to gather as much information as possible on the study endpoint without regard to the treatment allocation.

#### **Precision**

It is important to carefully define the trial endpoint in order to determine the precision of the estimate of the endpoint. An endpoint is the criterion by which patient benefit is measured. In cancer, the endpoint may be survival, tumor shrinkage, or duration of tumor disappearance. In heart disease, it may be survival or the occurrence of a new cardiac event. There is also increasing interest in quality of life, an endpoint which is more difficult to quantify. If the standard error of the estimate of the trial endpoint is too small, then resources have been expended uselessly, and the study could have been carried out with a smaller sample size. On the other hand, if the sample size is too small, resources have been wasted because no valid conclusions can be reached. The best way to guarantee precision is to estimate the appropriate sample size before beginning the trial.

#### **Range of validity**

The range of validity determines the generalizability of conclusions. This range of validity is usually defined by the eligibility criteria for the study. And here there is a tradeoff. If entry criteria are too broad, effects may be hidden in the variability introduced by including patients who are less likely to benefit. On the other hand, if criteria are too narrow, the subgroup may contribute little information about the population to which the study is directed. An example of entry criteria which are neither too narrow nor too broad comes from the recently completed Lipid Research Clinics Program (6). Men aged 39-59 with hyperlipidemia were randomized into two groups, one that received a lipid-lowering drug and one that did not. This group was selected because it was at high risk for a new cardiac event. Thus, the entry criteria were narrow enough to assure that a sufficient number of cardiac events would occur. The study showed that the group given the

cholesterol-lowering drug had fewer events than the placebo groups. Because the study concurred with previous studies and there was a dose-response relationship within the study, investigators were able to generalize their results to conclude that lowering cholesterol is useful in preventing coronary heart disease in the general population.

#### **Simplicity**

The best way to ensure a successful study is to clearly define the protocol. If the protocol is too complicated, useless data can be collected, or the investigators may not complete the protocol. Simplicity of study design also makes the study easier to describe, decreasing the possibility of misinterpreting the results when the study is published.

#### **Calculation of uncertainty**

This is the probability that the observed results could have occurred by chance alone. The best way to calculate uncertainty is to do a rigorous statistical analysis.

### **Controls**

Several types of controls are proposed in the literature: historical, crossover, and randomized. These are suggested because one type of control cannot be used in all settings. Randomized comparison groups are ideal, but there are situations where this is not possible. For example, randomized trials are unethical to test adverse effects of possibly noxious agents. Randomized trials are also difficult to perform when there are several alternatives, or when technologic advances produce improved agents so rapidly that long-term trials of a previous agent become obsolete before the trials are completed (7).

#### **Historical or nonrandomized**

Two types of nonrandomized controls have been proposed: controls from the literature and matched controls from previous or concurrent studies (8). In many places around the country, cooperative cancer groups commonly carry along the best treatment regimen from a previous study. The new treatment is then tested against the best previous treatment. These groups also use common protocols, so it has been proposed that investigators enter consecutive patients on the new treatment regimen, using patients on the previous treatment for a comparison group.

Simon (9) wrote that nonrandomized controls could be useful if they were from the same institution which is carrying out a new study with the same eligibility criteria, workup, etc. It is also important to have a static referral pattern, that is, no changes in the types of patients who are referred to an institution over time. For example, if an institution discovers a new treatment for



breast cancer, the publicity could lead to a sudden increase in numbers of patients and a change in the type of all cancer patients seen. Prognostic factors must be similar in the two groups, and Simon also insists that the prognosis *should be close to 100% predictable* before the advent of the new therapy (9). Tukey suggests that when a study is carried out without randomization, the results should be analyzed by insisting on a 30-50% improvement over the historical controls. If this analysis is indecisive, it will be necessary either to change to a randomized design or drop the trial (10).

Nonrandomized controls present several problems whether they are simultaneous controls, controls from previous studies, or controls from the literature. Diagnostic techniques, staging procedures, or secondary treatments often vary in the different studies. Differential bias in selecting patients for a particular treatment could be present, especially when controls from the literature are used. There could also be differences in determining ineligibility for treatment and differences in the distribution of prognostic factors (11).

Some of these dangers have been illustrated by other authors. Byar, et al (12) reported on a large series of prostatic cancer patients who were selected by the same criteria. Both groups were given placebos, and yet the two groups had substantially different survival rates over time. Pocock (13) reported on 19 instances where a collaborative group carried over one treatment for two successive studies. For 4 out of 19 pairs of trials, there were differences in outcome ( $p < 0.02$ ) when the two trials of the same treatment were compared.

### Crossover designs

Another type of study which has been used widely in the past is a crossover design. Patients are started on either the treatment or placebo; after a certain period of time, the patients are switched to the opposite group. Thus, each patient is used as his/her own control. Crossover designs have been used in patients with chronic diseases such as diabetes, or in Phase II cancer studies where the evaluation of a new treatment is made on patients who have received some previous treatment. Crossover designs may increase the sensitivity of a study and reduce the sample size necessary to carry it out. However, a crossover design has some strong disadvantages. Patients may experience changes over time that are unrelated to the treatment. Also a treatment may be influenced by previous treatments or responses.

Crossover designs are discouraged by the Statistical Advisory Committee of the Food and Drug Administration (FDA). If the relative efficacy of treatment in a second period differs from that in the first or is conditioned by the first period of response, the patient can-

not be used as his/her own control, thus invalidating the trial. Also, to determine if this interaction exists, a crossover design requires almost as many patients as a noncrossover design, with little reduction in sample size. Finally, crossover designs are frequently analyzed incorrectly, ignoring the design (1,9).

### Randomized control group

In randomized trials, the patients are allocated by some random process to either the treatment or the control group. Randomization is the most ethical procedure when there is no knowledge about the relative efficacy of a treatment; it also decreases the bias that might occur if patients are allocated without randomization.

Chalmers says that randomization should begin with the first patient (14). An argument frequently used against randomization is that it may be unethical for physicians to randomize patients to treatments which they believe are inferior, but physicians who truly believe that one treatment is better than another should not participate in a randomized clinical trial.

To evaluate the use of randomized clinical trials in surgery, Gilbert, et al (15) used the Medical Literature Analysis and Retrieval System (MEDLARS) to identify all randomized surgical and anesthesia trials reported in English with at least ten patients in each group for the years 1964-1976. Forty-six studies were identified, and 49% of the innovations were successful when compared to the standard. That is, when assessed by randomized clinical trial, innovations in surgery and anesthesia were successful only about one half of the time. Since innovations brought to the stage of a randomized clinical trial are thought to be beneficial, the failure of 51% provides strong evidence for the value of a randomized trial in checking these innovations.

A danger of poorly controlled studies is that large numbers of such studies may create an illusion of strong evidence that causes investigators to believe in the efficacy of the new therapy. This accumulation of evidence is demonstrated in Table I which shows the early studies of the usefulness of diethylstilbestrol (DES) to prevent spontaneous abortion during pregnancy. These studies led to widespread use of DES. Yet the studies had either contrived controls, no controls, or historical controls. Contrived controls included reported rates of spontaneous abortion for the same hospital in previous years. In contrast (Table II), where alternate randomized controls (every other patient is put on the placebo) or simultaneous controls (the first 100 receive treatment, the next 100 receive placebo) were used, DES was shown to be no more effective than a placebo in preventing spontaneous abortions (16).



**TABLE I**  
**Diethylstilbestrol in Prevention of Abortion and**  
**Other Accidents of Pregnancy (16)**

Enthusiastic Studies				
First Author	Date	No. of Patients	Controls	Blinding
Smith	1946-1954	1,300 +	Contrived	No
Gitman	1950	51	None	No
Davis	1950	42	Historical	No
Ross	1951-1953	200	None	No
Pena	1954	200	Historical	No
White*	1949-1953	642	Contrived	No
Plate	1954	29	None	No

\*Only diabetics studied

**TABLE II**  
**Diethylstilbestrol in Prevention of Abortion and**  
**Other Accidents of Pregnancy (16)**

Negative Studies				
First Author	Date	No. of Patients	Controls	Blinding
Crowder	1950	100	Alternate	No
Robinson	1952	93	Alternate	No
Ferguson	1953	393	Alternate	Yes
Dieckmann	1953	1,646	Alternate	Yes
Randall	1955	237	Simultaneous	No
Reid*	1955	148	Random	Yes

\*Only diabetics studied

## Ethical Issues

In designing a trial, it is impossible to avoid ethical issues. For example, an uncontrolled, poorly designed trial is clearly unethical. A primary issue, of course, is the timing of a clinical trial. There is only a small period of time when a trial can be carried out. Before this time, the medical community has so little evidence about efficacy that they would be unwilling to randomize any patient to the new treatment. After that time, the medical community is so convinced of the efficacy of the new treatment that they would be unwilling to randomize the patient to the control group.

It is important to evaluate the past evidence in determining whether a trial is necessary. In the 1950s, chloramphenicol was used in an uncontrolled study of patients with typhoid fever, with an overwhelming positive effect. Ten years later, in the 1960s, a double-blind, randomized trial was carried out in which 23% of the placebo group vs 8% of the treatment group died. It is difficult to justify the randomized clinical trial when the effect of treatment has already been clearly demonstrated (4). On the other hand, outside factors may hinder the introduction of an innovation, making a randomized clinical trial possible. For example, the previously mentioned Hill trial of streptomycin for tuberculosis was carried out because exchange regulations in Britain after World War II limited the importation of the drug. This made the double-blind study feasible. Only a limited amount of the drug was available, and all patients who could possibly have had the drug were randomized to receive it (10).

It is also unethical to continue a clinical trial to measure the outcome with better precision. For example, the Beta Blocker Heart Attack Trial was designed to test whether the use of the drug could prevent myocardial infarctions in those who had one cardiac event. While it would be useful to know how long the patient must take a beta blocker in order to be protected against recurrent myocardial infarctions, the successful results of the study led to an early termination of the trial. It became apparent that it was unethical to withhold this drug from the untreated group.

The cost of clinical trials also raises ethical issues. As a society we cannot afford to study all medical questions by clinical trials. Thus, the practitioner is often forced to make decisions on informed opinion alone. This is an important component of medical care, and advocates of clinical trials are obliged not to disturb this process too greatly.

Discussion continues about the allocation of funds for clinical trials. Is it ethical to spend large amounts of money on clinical trials to the detriment of more basic laboratory research? When this issue is evaluated with regard to a particular clinical trial, it is important to estimate the losses sustained in choosing the undesirable treatment, or continuing to administer it for years and to exclude treatment costs (10). An example is coronary bypass surgery. The cost of doing the surgery should not be considered as part of the cost of carrying out trials of this procedure.

There are also difficult questions relating to statistical analyses. First, when repeated tests of the hypothesis are planned at different points in time, the alpha level for rejecting the hypothesis of no difference in the two groups must be adjusted. The more the data are inspected, the wider the rejection region has to be, at least

early in the trial. If too many tests are planned, the trial may be unnecessarily prolonged.

Stopping the trial also provides some interesting ethical problems. If a trial is terminated too early, conclusions may not be accepted by the medical community. If it is terminated some time after results appear definitive, one study group may suffer needless harm. These decisions are made not only on the statistical criteria but also on the judgments of the advisory boards. A unique and less preferred stopping rule was reported in the *New York Times* (17). Anturane study researchers (18) were divided over stopping their trial. The researchers agreed to send an article on the trial to *The New England Journal of Medicine* with the understanding that if the journal accepted the article as statistically valid, the trial would stop and the patients would be informed of the results.

Informed consent is another well-known ethical issue. Two recently published studies evaluated subjects' comprehension of studies (19,20). In a sample of

patients from the Beta Blocker Heart Attack Trial, most were well informed about study design and risks. However, 8% (5 patients) believed they were participating in a therapeutic program rather than a research project. A study of patients with psychiatric problems by Appelbaum, et al (20) revealed the same misconception. While the subjects understood the benefits and risks of treatment, many believed that study assignments were based on therapeutic considerations rather than randomization. These results stress the need for better explanations. Subjects may have to be told explicitly that scientific goals will have priority over therapeutic goals.

Finally, given a well-designed study which has been correctly analyzed and shown to benefit or harm, it is unethical to continue as if the trial had not been done. A prime example was Linn's study of scurvy. Forty years passed before the British navy provided lemon and orange juice to its sailors.

## References

1. Friedman LM, Furberg CD, DeMets D. *Fundamentals of clinical trials*. Boston: John Wright, PSG Inc 1982.
2. Bull JP. The historical development of clinical therapeutic trials. *J Chronic Dis* 1959;10:218-48.
3. Office of Technology Assessment. *The impact of randomized clinical trials on health policy OTA-BP-H22, and medical practice*. Washington DC: US Congress, 1983:5-7.
4. Chalmers TC. The randomized clinical trial as a basis for therapeutic decisions. In: Tygstrup N, Lachin JM, Juhl E, eds. *The randomized clinical trial and therapeutic decisions. Statistics: Textbooks and monographs*. New York: Marcel Dekker Inc 1982;43:14,18.
5. Cox DR. *Planning of experiments*. New York: John and Sons Inc 1958:1-13.
6. Lipid Research Clinics Program. The lipid research clinics coronary primary prevention trial results. I. Reduction in incidence of coronary heart disease. *JAMA* 1984;251:351-64.
7. Feinstein AR. An additional basic science for clinical medicine: III. The challenges of comparison and measurement. *Ann Intern Med* 1983;99:705-12.
8. Gehan EA, Feireich EJ. Non-randomized controls in cancer clinical trials. *N Engl J Med* 1974;290:198-203.
9. Simon RM. Design and conduct of clinical trials. In: DeVita VT, Hellman S, Rosenberg SA, eds. *Cancer — principles & practice of oncology*. Philadelphia: J.B. Lippincott Co, 1982;198-225.
10. Tukey JW. Some thoughts on clinical trials, especially problems of multiplicity. *Science* 1977;198:679-84.
11. Peto R, Pike MC, Armitage P, et al. Design and analysis of randomized clinical trials requiring prolonged observation of each patient. I. Introduction and design. *Cancer* 1976;34:585-612.
12. Byar DP, Simon RM, Friedewald WT, et al. Randomized clinical trials: Perspectives on some new ideas. *N Engl J Med* 1976;295:74.
13. Pocock SL. Randomized clinical trials. Letter. *Brit Med J* 1977;1:1161.
14. Chalmers TC. When should randomization begin? Letter. *Lancet* 1968;1:858.
15. Gilbert JP, McPeck R, Mosteller F. Statistics and ethics in surgery and anesthesia. *Science* 1977;198:684-9.
16. Chalmers TC. The impact of controlled trials on the practice of medicine. *Mount Sinai J Med* 1974;41:753-9.
17. Winfrey C. Again the ethics of medical research. *The New York Times*. 1978 Mar 5:(Sec E)19-20.
18. The Anturane Reinfarction Trial Research Group. Sulfapyrazone in the prevention of cardiac death after myocardial infarction: The anturane reinfarction trial. *N Engl J Med* 1978;298:291-5.
19. Howard JM, DeMets, BHAT Research Group. How informed is informed consent? The BHAT Experience. *Controlled Clin Trials* 1981;2:287-303.
20. Appelbaum PS, Roth LH, Lidz C. The therapeutic misconception: Informed consent in psychiatric research. *Int J Law Psychiatry* 1982;5:319-29.