The randomized clinical trial: Study design and execution

Lloyd R Sutherland, MDCM, FRCP, FACP

LR SUTHERLAND. The randomized clinical trial: Study design and execution. Can J Gastroenterol 1993;7(7):557-560. The clinical trial has been described as the indispensable ordeal and physicians base many clinical care decisions on information gained from such trials. As well, the methodology of a trial can be used just as effectively to look at various treatment strategies. The accomplishments of clinical trials include that they have modified practice patterns (eg, the prescribing habits of physicians in the '90s, compared with the '60s, are more often founded on reality as reflected by clinical trials) and have buttressed the concept that disease is quite heterogeneous (as the inclusion and exclusion criteria are defined within a clinical trial, it is possible to begin to appreciate the heterogeneity of the disease). This paper will consist of a brief review of the historical developments of the clinical trial and the various stages of a clinical trial from initiation to completion. Various design features, focusing on selection of the test and control treatments, establishing comparable study groups, and some suggestions about blinding and controlling bias will also be reviewed.

Key Words: Clinical trial, Methodology, Review

A FAMOUS RESEARCHER WAS ONCE asked to give a lecture on the design of clinical trials. He reported that his first thought was to stand up, say "Be careful", and sit down. The clinical trial has been described as the indispensable ordeal and we base many of our decisions in clinical care on information gained from such trials. Although the focus of this paper is the use of clinical trials as a technique of testing pharmacological therapy, the methodology of a trial can be used just as effectively to look at various treatment strategies, comparing, for example, colonoscopy versus air-contrast barium enema.

What have the clinical trials accomplished to date? There is some evidence that they have modified practice patterns. The prescribing habits of physicians in the '90s, compared with the '60s, are more often founded on reality as reflected by clinical trials. Patients have reason to be reassured that the therapies they are being offered are based on solid scientific evidence.

Another benefit from clinical trials is the concept that disease is quite heterogeneous; for example, not all peptic ulcers are the same, disease location may be important in inflammatory bowel disease. As the inclusion and exclusion criteria are defined within a clinical trial, it is possible to begin to appreciate the heterogeneity of the disease.

This paper will consist of a brief review of the historical developments of
the clinical trial and the various stages of a clinical trial from initiation to completion. Various design features, focusing on selection of the test and control treatments, establishing comparable study groups, and some suggestions about blinding and controlling bias will also be reviewed.

The first clinical trial was described in the Bible in the Book of Daniel. Daniel tells that while in Babylon, he and his companions were not thriving on the local Babylonian diet so he suggested to the Captain of the Guard that they be divided into two groups; one group would eat the traditional Israeli diet and the other would continue the Babylonian diet. Several weeks later, those on the Babylonian diet were feeling much better. Daniel tells that while in Babylon, he and his companions were not thriving on the local Babylonian diet so he suggested to the Captain of the Guard that they be divided into two groups; one group would eat the traditional Israeli diet and the other would continue the Babylonian diet. Several weeks later, those on the Babylonian diet were still feeling unwell but those on the Israeli diet were feeling much better.

In the modern era, Lind (in 1747) was the first to incorporate an untreated control group in his studies on scurvy. Sailors were divided into groups which received the regular diet or various dietary manipulations, one of which included citrus fruits. Haygarth, in 1799, was the first to use a sham procedure. At that time there was a growing interest in passing metal rods over individuals to cure their arthritis. He had new rods made that contained wood, not iron, and demonstrated that patients had the same responses to the wooden rods as to the iron rods. Gull, in the 19th century, was the first to use a placebo treatment. Fisher, in 1923, was the first to suggest that randomization was an indispensable component for experimentation. By the 1940s, the first multicentre trial had been organized.

Although clinical trials have become increasingly important, it is a bit disconcerting that in reviewing citations in the current medical literature, only about 20% of all the trials that are published annually are randomized. As well, less than 10% are double-blinded and fewer than 5% have placebo controls.

Controversy continues as to when a trial should be initiated. Various authorities feel that a trial should be initiated at the first whiff of a new idea or as soon as a new medication is synthesized. Ideally the trial should be initiated before the compound or the practice becomes part of general use. There is often a very narrow window between when a new intervention is being introduced and when it can be assessed within a trial. For example, until recently there have been very few trials looking at coronary artery bypass surgery compared with medical therapy. The medical community was not prepared to accept trials because they thought the efficacy of the therapy had already been accepted.

In terms of protocol development, the Canadian scene has changed over the past 10 or 15 years. In the past, clinical trial protocol offered to the investigator by the pharmaceutical industry was set in stone with little chance for embellishment or refinement. Today most protocols are developed in conjunction with the industry, usually through the convention of a steering committee with three or four physician advisors. After the protocol is finalized, it is sent to the Health Protection Branch (HPB) for review. The '60-day rule' states that HPB has 60 days to respond to the protocol and if no issues are raised within 60 days, patient recruitment can commence. The protocol then enters through the treatment and follow-up phases, followed by analysis and preparation of the final report.

An important element in protocol design is the definition of the question to be answered. In certain studies it is clear that too many questions have been built into the trial and the trial becomes so diffuse that the major question is obscured. Accordingly, it is important when designing the trial that the question be defined simply in a circumspect fashion.

There are certain principles to follow when determining the control treatment or mechanism. An underlying principle of any trial is the belief that the benefits of the study or therapy will outweigh whatever risks there are for patients who are participating in the study.

First of all, one should be able to distinguish one treatment from the other. The simplest way to ensure that the two treatments differ as much as possible is to use a placebo. However, one must be able to justify either therapy. This becomes an issue when placebos are used. North America trials still may use placebos, but they are much less popular in Europe and other parts of the world. The use of a placebo can, I believe, still be justified as it is the best way to determine in an efficient manner whether the new therapy has any effect. If a placebo is to be used, one must have clearly defined exit points so that the patient can be removed if his or her clinical situation deteriorates. Patients and their physicians always feel that the patient must be on placebo if they do not do well. Higher drop-out rates can be expected and larger number of patients may be required.

Inclusion of a placebo also gives a much better understanding of the side effects profile. A good example of the importance of placebos is the trial of the hepatitis B vaccine that was carried out in New York in the 1970s. The trial used a placebo group, but one might have argued, why is a placebo necessary for a vaccine? During the course of the trial, two patients developed Guillain-Barré syndrome. Both were on placebo. The possibility remains that if the researchers had not had the placebo group, the Guillain-Barré syndrome would have occurred during that vaccination trial and the sponsor might have withdrawn the vaccine because of a concern regarding neurological side effects.

Doubt should exist about the efficacy of either treatment. If one clearly is superior to the other, then the trial should not proceed, except perhaps in the setting of a placebo. The treatment and test control treatments have to be acceptable to patients. There should be an ethical basis for either treatment, and both need to be compatible with the health care needs of patients in society. It is not appropriate to use a control treatment that requires special intervention that physicians could not use in their own hospital or office. In summary, the test control treatment should be similar to the real world use as practical.

Turning to randomization, although
it generally is considered to be important from a statistical point of view, the major reason for using randomization is to provide a method for making bias-free treatment assignments. The most important randomization guideline is that treatment assignment cannot be determined in advance. This precludes using strategies such as allocation by date of birth, date of assessment, hospital identification number, etc. By knowing in advance which treatment will be assigned, the physician may be biased and only enter patients into the treatment module they believe to be most appropriate. Future assignments should not be predicted from past assignments. This means, for example, that patients should not be randomized in blocks of two. Often problems in allocation can be avoided by using a centralized allocation system by which one cannot predict from one assignment what will happen in the next.

The design should try to achieve the highest level of blinding possible. There are three levels of blinding: unblinded — both patient and physician know the assignment; single-blind — either the patient or the physician knows the assignment but the other does not; double-blind — neither the physician nor the patient knows the treatment allocation. It is possible to have a single-blind study or, in fact, an unblinded study and still get very objective information. What is required is that those who are doing the assessments are blinded to allocation. In this scenario the essential data should be collected by individuals who have no knowledge of the treatment. Particularly when subjective outcome measures are used — for example, a quality of life, Reichert or linear scale — the measurements should be made by individuals who are blind to the assignment.

The allocation scheme should be reproducible, documented and leave a clear audit trail. This effectively rules out, for example, flipping a coin to tell which treatment to allocate. It would be very hard, 20 years later, when the Food and Drug Administration comes to audit, to document the results of a coin toss. What is required is a log of the randomization which can then be audited. It should be possible from looking at the scheme to determine if departures from the sequence occurred. "A difference, to be a difference, must make a difference." This leads into the determination of the sample size required and the issue of clinical versus statistical significance. A variety of trials have detected rather minuscule differences between the treatment and the control treatment. If this is a difference in mortality rates, that may be acceptable. However, in dealing with Crohn's disease, a slight change in the Crohn's Disease Activity Index, although statistically significant, may not be clinically important. Sample size calculations are vital for the design of a clinical trial. Unfortunately, many trials do not report their calculations prior to initiation of the trial. A basic rule of thumb would be that the wider the difference in clinical outcome between the two treatments, the fewer patients required. For example, for a moderate difference in response (60% versus 30%) 60 patients might be required but for a small difference (60% versus 50%), 400 patients would be required. This is one of the reasons the placebo controlled studies are popular as one can anticipate wider differences in response.

Discussion about sample size leads to a review of alpha and beta levels as well as type I and type II errors. Table 1 compares and contrasts the principles behind these terms. The first assumption is that two therapies, treatment A and treatment B, are compared in a clinical trial. The issue is the probability that the conclusions based on the study in fact reflect the true state or reality.

The first step is to state the null hypothesis (H₀) — that there is no difference in effect between the two treatments. If the H₀ is rejected then the alternate hypothesis (H₁) can be accepted and a difference in effect between the two treatments expected. Setting the alpha and beta probability levels will determine the degree by which the investigator wishes to exclude the possibility that the rejection or acceptance of the H₀ was based on chance. Generally, the lower the alpha and beta levels are set, the more patients that are required. The alpha level reflects the degree of concern the investigator has regarding the possibility of a type I error. This occurs if the H₀ is rejected and it is concluded that treatment A is better than B but in reality the H₀ is true (treatment A and B are the same). The alpha level is generally set at 0.05, i.e., 1 in 20 probability that a type I error has occurred.

Similarly the beta level is linked with concerns related to type II errors. A type II error occurs when the H₀ is confirmed and it is concluded that treatment A is the same as treatment B but in reality the treatments differ. The beta level is generally set at 0.20 to 0.10, i.e., 1 in 5 to 10 probability. The reason the alpha level is set at a more rigorous level is that more damage can be assumed to be done to society if a new treatment is characterized as being superior when, in fact, it is not.

In conclusion, when designing a study, remember the 10 commandments of the randomized clinical trial.

**Thou shall randomize:** This is required to eliminate bias. There are a variety of reviewed various schemes that one can use, but the best way of randomization is a central, computerized registry.

**Thou shall double-blind:** This concept includes not only attention to randomization, but also efforts to manufacture placebos identical in both appearance.

---

**TABLE 1**

| True state of nature | H₀ true (no difference between A and B) | H₁ true (there is a difference between A and B) |
|----------------------|---------------------------------------|-----------------------------------------------|
| **H₀ true** (according to trial) | Correct decision | Type II error (beta level) |
| **H₁ true** (according to trial) | Type I error (alpha level) | Correct decision |

* Null hypothesis; † Alternate hypothesis

---

**Discussion** about sample size leads to a review of alpha and beta levels as well as type I and type II errors. Table 1 compares and contrasts the principles behind these terms. The first assumption is that two therapies, treatment A and treatment B, are compared in a clinical trial. The issue is the probability that the conclusions based on the study in fact reflect the true state or reality.

The first step is to state the null hypothesis (H₀) — that there is no difference in effect between the two treatments. If the H₀ is rejected then the alternate hypothesis (H₁) can be accepted and a difference in effect between the two treatments expected. Setting the alpha and beta probability levels will determine the degree by which the investigator wishes to exclude the possibility that the rejection or acceptance of the H₀ was based on chance. Generally, the lower the alpha and beta levels are set, the more patients that are required. The alpha level reflects the degree of concern the investigator has regarding the possibility of a type I error. This occurs if the H₀ is rejected and it is concluded that treatment A is better than B but in reality the H₀ is true (treatment A and B are the same). The alpha level is generally set at 0.05, i.e., 1 in 20 probability that a type I error has occurred.

Similarly the beta level is linked with concerns related to type II errors. A type II error occurs when the H₀ is confirmed and it is concluded that treatment A is the same as treatment B but in reality the treatments differ. The beta level is generally set at 0.20 to 0.10, i.e., 1 in 5 to 10 probability. The reason the alpha level is set at a more rigorous level is that more damage can be assumed to be done to society if a new treatment is characterized as being superior when, in fact, it is not.

In conclusion, when designing a study, remember the 10 commandments of the randomized clinical trial.

**Thou shall randomize:** This is required to eliminate bias. There are a variety of reviewed various schemes that one can use, but the best way of randomization is a central, computerized registry.

**Thou shall double-blind:** This concept includes not only attention to randomization, but also efforts to manufacture placebos identical in both appearance.
and taste. Patients will go to amazing lengths to try to find out which regimen they receive. It may be of interest to ask the patient and the physician at the end of the study whether they believe the patient received active therapy.

Thou should not do unto a patient that which thou would not have done to thyself: The treatment should always be justifiable.

Thou shall enter consecutive patients: It is helpful to keep a log of patients that, in fact, have been assessed but who did not enter the trial to determine if there are differences between patients who entered the study and those who did not. Remember that in a clinical trial, one is looking at a very homogeneous group and then trying to extrapolate from that group of patients to the general population. An example of a study which took a log seriously would be the Dutch study of steroids and sulfasalazine. They looked at 175 patients and enrolled 60 into the study. Such a report might indicate how generalizable the trial patient sample is to the general population.

Thou shall not deviate from the protocol: Protocol deviations are the bane of any clinical project.

Thou shall complete all forms: Completing all the forms, of course, for most of us is the happy task that often is assigned to the study nurse. It requires an obsessive nature but it is required in these times of medical audit.

Thou shall consult a statistician: Statistical evaluation is an art to itself. Statisticians like to be involved in the development process rather than be handed a computer disk with data after the trial has been completed. Many physicians over-rely on the P value. It is important to understand that there may be a variety of statistical techniques to analyze data and that each technique may arrive at a different P value.

Thou shall only seek answers to a few questions: This has to be balanced the fact that large multicentre trials offer an opportunity to gain additional information about the disease in question. Other questions can be added to the project as secondary questions, not necessarily as part of the assessment of the therapy.

Thou shall not adulterate the data.

Thou shall not take the HPB name in vain!
