18

The Design of Clinical Trials

18.1. Introduction

The main emphasis of this book is the analysis of medical data. The quality of data available for analysis clearly depends on the design used for its collection. In a medical trial, investigators must balance considerations of ethics, simplicity and good statistical practice, and it is often difficult to give anything more than general advice about the characteristics of a well-designed study. However, there are a number of good resources on the design of clinical trials which can be consulted for more detailed discussion.

In this chapter, we shall briefly present a few of the issues which are frequently discussed and comment on trial organization. We also explore the role of randomized treatment assignment in clinical trials in somewhat greater detail. The use of randomized trials has been the subject of considerable debate, and we feel it deserves some discussion here. Sections on intention-to-treat analyses, factorial designs and repeated significance testing provide a short overview of these aspects of trial design. Finally, we conclude the chapter with a brief introduction to the important topic of the sequential analysis of a clinical trial.

18.2. General Considerations

Perhaps the primary requirement for a good clinical trial is that it should seek to answer an interesting question. The choice of treatments to be compared and the patients on whom they are to be compared largely determine the practical importance of discovering whether the treatments differ.

Strict entrance requirements which generate a very homogeneous patient population facilitate precise treatment comparisons with a small number of patients. However, the results of a larger study with a more heterogeneous population would probably be more convincing to a practising physician.

A trial with two highly divergent treatments is simple and is likely to produce a result more quickly than a trial with two similar treatments, or one involving more than two treatments. This observation is an important one since, for various reasons, it is often tempting to stop a trial before conclusive results have been obtained. On the other hand, sophisticated designs frequently allow more comprehensive inferences to be deduced. It is also important to ensure that the intended treatments are acceptable to the clinicians who must enroll their patients into the trial. Therefore, in selecting treatments, a balance must be struck among these various factors.

The design stage of a clinical trial should also specify data collection procedures. The information which will be collected concerning each patient at entry into the study should be identified. These baseline variables can be used in the analysis of the trial results to adjust for patient differences in the treatment arms. Therefore, information which is gathered at entry should be related to the chosen endpoints or to potential side effects of treatment; this latter aspect is sometimes overlooked. Since collecting data on a patient at the time of entry into a study is generally easier than attempting to recover relevant baseline information at a later time, it is advisable to record as much baseline data as is feasible.

Collecting data on only a few endpoints will make follow-up easier, and also reduces the chance of serious bias due to differential follow-up among patients. At the same time, as much information as possible should be recorded concerning each endpoint of the study. The time until a certain event is observed is more informative than a mere record of its occurrence. All patients who enter a trial should be followed, even if they abandon the treatment protocol, since exclusion of these patients can introduce bias. Similarly, the treatment groups which are compared in the primary analysis should be groups based on the treatments which were originally assigned (see also §18.5), because this comparison reflects how the treatments will perform in practice. Of course, it may be of scientific interest to restrict a comparison to those patients receiving and tolerating treatment regimens, for example, but the more general comparison, based on assigned treatments, is usually more valuable in the long run. Note that, in order to avoid bias, treatment assignment should only occur after informed consent procedures.

Another point of frequent discussion concerns the stratification of treatment assignment by prognostic factors. The statistical methods which have been developed, such as regression models, reduce the need for precisely com-

parable treatment groups. It seems reasonable, however, to consider stratifying a trial on a few factors of known prognostic significance and to attempt partial balance on other factors via randomization. The effectiveness of the randomization in achieving this balance should be examined. Peto et al. [36] argue in favor of no stratification, but it is more cautious, and perhaps more convincing, to balance on major prognostic factors rather than to rely solely on sophisticated statistical analyses to adjust for imbalances in the treatment arms. While it is generally agreed that excessive stratification is complicated, often unnecessary, and may even result in poor balance if only a few patients are entered in each stratum, an alternative to stratified randomization does exist. The advent of widely available computing resources allows the use of a technique called *minimization*. Minimization aims to provide an effective randomization scheme when there are more than two or three prognostic factors on which stratification might be appropriate, and therefore the risk that stratification on those factors will lead to poor balance is no longer negligible.

The goal of minimization is not to ensure balance within each of the potentially many strata that are defined by all possible combinations of the relevant prognostic factors. Instead, it only aims to ensure that, when each prognostic factor is examined individually, there is appropriate balance between treatment assignments. The balance that has been achieved prior to randomizing a newly enrolled subject is examined, and the probability of assigning that subject to each of the various treatments offered in the trial is then specified for the new subject in a way that is likely to reduce any imbalance that may be present. The algorithm to specify the appropriate randomization probabilities is relatively complex and thus requires access to computer resources. It is this complexity which provides protection against selection bias.

There is debate concerning certain issues raised by the use of minimization, but we are not able to address those issues here. While it represents a method of treatment assignment which is being used increasingly, some caution about its routine adoption may be wise. For many trials, a moderate level of stratification may be sufficient, and easier to implement.

In the early design stage, the inferences which are to be drawn from the study should be identified. For example, suppose that a clinical trial of two adjuvant therapy regimens following surgery for breast cancer is being planned. The response variables which are of interest are remission duration and survival. The study protocol should specifically mention that remission duration and survival will be used to compare the two treatment regimens. In addition, the statistical procedure that will be used to analyze the results of the trial should be specified.

When the study has ended and the data are analyzed, it may be determined that the treatment A arm of the study had fewer metastases to the ova-

ry than the treatment B arm, and this difference is statistically significant at the 5% level. Perhaps no other site of metastasis suggested there was a difference between the treatments, and the comparison of the two treatments on the basis of remission duration also indicated no difference. These results should not lead us to conclude that treatment A is better than B.

If ten sites of relapse were examined, then, because of the multiple comparisons problem which we discussed in chapter 16, it is not unlikely that one of the ten sites will, by accident, suggest there is a difference between treatments. If there was no reason, prior to the study, to suspect a treatment effect at a particular site of relapse, then the discovery of such an effect should be viewed with caution, especially if the designated principal comparison does not identify a treatment effect.

A major reason for specifying, in advance, the statistical procedure which will be used in the analysis is that it is possible to find perhaps ten different statistical tests which compare remission duration in two treatment groups. It might happen that one of these tests is just significant at the 5% level, while the rest suggest there is no significant difference. Such a 'search for significance' is entirely inappropriate; therefore, a reasonable test procedure should be specified before the study begins and used when the data are analyzed.

It may be that there is valid medical information in the unexpected results from a single relapse site, or that the statistical test indicating a treatment difference is particularly sensitive for the type of data produced by the study. If there is reason to suspect that this is the case, then the results should certainly be reported. How the results arose should also be reported, and it should be made clear that they need to be confirmed in other studies before being generally accepted. On the other hand, one can be much more confident about a result identified by a test which was specified prior to a detailed examination of the data.

18.3. Trial Organization

The previous section dealt with issues that arise in the design of clinical trials in a very general way. To implement an actual trial requires that attention be given to a myriad of details. In this section we comment briefly on the major aspects of trial organization. Our aim is simply to highlight key features, and we assume that interested readers who wish to undertake a clinical trial will both read more widely and hold discussions with researchers who have experience in running trials.

Specific considerations may arise if a trial is undertaken with an aim of gaining regulatory approval for a new treatment, as is common for new drugs

developed by pharmaceutical companies. We do not attempt to discuss details concerning such trials here but, of course, the basic structure of all trials should be similar.

The bedrock of any clinical trial is the trial protocol. In this document the details regarding justification for the trial question, the details of the trial treatments, eligibility of patients, assignment of treatments to patients, primary outcomes, primary analyses and the monitoring of trial progress are specified, along with the sample size calculations that justify the expected size of the trial. Recently, some medical journals have adopted the policy of not publishing trial results unless a trial protocol has been officially registered prior to the beginning of the trial. In any event, it is the protocol that drives the day-to-day activities associated with the trial.

The trial protocol is also central to the submissions that are made to ethics committees which approve the implementation of the trial in various clinical jurisdictions. Considerable resources can be required to attain the necessary ethical approvals, especially if a trial aims to recruit patients from more than one centre. Increasingly, there are national and international guidelines about the running and reporting of trials. Demonstrated adherence to these requirements also requires resources.

Most trials will require a specific source of funds. A research grant proposal to support a trial will usually provide information similar to that found in a trial protocol, but will generally devote more attention to the justification for the trial, including a summary of available information on the proposed treatments. The latter will sometimes include a meta-analysis of related studies; chapter 20 provides a brief introduction to that subject. A grant proposal will usually include less clinical detail than does the protocol; however, adequate financial details will be required to support the request for trial funding. The submission of a grant proposal may precede the completion of a draft trial protocol; nonetheless, the basic outline of the protocol must be in place in order to achieve funding for the trial.

The most common organizational structure for a trial is to have three primary committees. The first is often called the Trial Management Committee or Team. This group of individuals is usually headed by the principal investigator(s) for the trial, and is charged with the day-to-day running of the trial. The Trial Steering Committee, on the other hand, is chaired by an individual who is independent of the investigators who designed and are implementing the trial, in order to provide independent oversight of the study. The Steering Committee will have other independent representation, usually including statistical expertise, and often including lay individuals from either the general public or disease interest groups. Trial investigators may also sit on this committee. The Trial Steering Committee is vested with primary respon-

sibility for the ethical running of the trial. The final committee is usually the Data Monitoring Committee (DMC). This is a group of individuals who are entirely independent of the trial, and who are charged with monitoring the trial from an ethical perspective. The responsibilities of the DMC include ensuring that the trial is recruiting patients in a timely manner, being aware of adverse events associated with trial treatments and, when necessary, having access to the accumulating evidence concerning possible efficacy differences among trial treatments. The minimal requirement for such a committee is to have specific expertise in the clinical area of the trial, statistical expertise to interpret trial data, and general experience in the running of clinical trials. Specific ethical or legal expertise and lay participants may be included in such a committee, but there is wide variation from one trial to another. The DMC is advisory and reports to the Trial Steering Committee.

We have not done justice to the complexities of trial organization in this very brief discussion, but reiterate that there are many specialized sources of additional information for the interested reader.

18.4. Randomized versus Historical Controls

A randomized clinical trial is generally regarded as the 'gold standard' for a clinical investigation. Nevetheless, there can be questions concerning the use of such trials. One of the major concerns is often the ethical problem of allowing a random event to determine a patient's treatment.

We do not intend to summarize the various issues which have been discusssed with respect to randomized clinical trials. Arguments for and against their use have been advanced in the past, and interested readers can consult references 36–45 from the late 1970s and early 1980s. References 46–47 provide more recent discussions. In this section, we address only the question of whether there are alternative designs which are as informative as a randomized trial. The issue is fundamental to all discussions of randomized trials.

We will assume that the purpose of a medical trial is to make a comparative statement about the efficacy of two or more treatments. Therefore, the accuracy of this statement is important. It has been argued that this particular assumption regarding a medical trial is inappropriate. Freireich [43] has argued that a comparative trial which shows major differences between two treatments is a bad trial because half the patients have received inferior treatment. Although, in a sense, this is true, we feel that any wider perspective on clinical research will encompass a desire to know the true relative merits of different treatments.