Practical Problems in Long Term Clinical Trials

K. K. ÜBERLA

Institut für Medizinische Informationsverarbeitung, 8000 München 70, Federal Republic of Germany

Summary: In this paper various practical problems of long term clinical trials are discussed according to the three phases of such trials – planning, execution and evaluation. Practical problems in planning include, for example, the development of a protocol, study organization, inclusion criteria, determination of the endpoints, number of subjects, problems with randomization and double blinding. Problems in execution covered concern: responsibility within the hospital, central monitoring, continuous data entry, regular reports, open information and interim evaluation. Practical problems in evaluation include selection, missing data, comparability of treatment groups, compliance, exploratory data analysis and conflicts of interest. Important general problems are concerned with asking sensible questions, balanced between methodology and medical content, coping with extraneous problems and risk-benefit evaluation. Throughout the paper the limitations of statistical techniques are stressed in comparison with the judgement and experience of the statistical investigator.

1. Introduction

For the purpose of this paper I consider to be a "long term trial" any trial with an observation time of more than 6 months. The trial must be prospective; I will not include retrospective or case-control studies. However, a long term study is not necessarily a randomized clinical trial: observational studies conducted on one group of patients followed up for more than 6 months may also be considered.

One cannot investigate chronic diseases without long term trials as they are highly complex in their causes and pathogenesis and response times to treatments are generally long. Serious side-effects may occur decades after treatment. It is likely that several variables will be confounded. Treatment regimens are diverse and may well be specific to an individual patient.

I would estimate that less than 10% of medical knowledge about chronic diseases is established in a statistically acceptable way. Less than 20% of the decisions in drug regulation might rely on methodologically sound studies. Methodology therefore seems to be not so important in practice. The world is rather complex and nearly everywhere there is a trap for the uninitiated investigator.

Treating this wide field I will cluster the practical problems around three phases of a long term trial: problems in planning, problems in execution and problems in evaluation of such a trial. A fourth group of problems is more general.

2. Practical Problems in Planning

Initiating a long term trial. This starts with sketching the first description, which should be basic and simple, as well as flexible and encouraging for sponsors and clinical investigators. It should detail the questions to be answered, the treatments to be used, some idea of the patients and hospitals to be included and specification of the endpoints. It is wise not to incorporate too many questions and to have only one coordinator during this phase. There should be a rough idea of the cost and duration of the trial. Possible motivations for performing a trial may be scientific or financial. It is vital to approach the right people and to get them interested in protocol development and participation. This is a rather critical phase and many long term trials do not survive it. Getting an idea about the possible benefits and risks, and about the chances of success in a trial is more of an art than a method; getting it initiated is often a matter of luck.

Development of the protocol. The development of a good protocol takes approximately one year. The state of the art in treating the disease has to be reviewed thoroughly. Experts and possible participants should meet several times; the protocol must be drafted

and revised frequently and external experts should be asked to review the procedures and forms. The protocol is the manual for the action of many people over several years and it should be very specific and detailed. Staff change; the protocol and its appendices should be one of the few constants in a long term trial. The time and effort required for getting a good protocol is usually considerably underestimated.

The clinical investigators should be involved extensively and the central organizing and statistical facilities should be developed and pre-tested. The way things are organized during this phase and the way people communicate and interact will determine the quality of the final protocol and consequently how good the trial will be. The weight of a protocol in kilograms is not necessarily an indicator of quality. Not every description has to start with Adam and Eve. Some procedures are in common use and physicians have had their education [1]. One has to find the right compromise as to the amount of detail necessary. In West Germany we prefer shorter protocols than in the United States, without omitting essential details.

Study organization. Responsibilities have to be clearly defined and the organizational structure has also to be clear. It may vary considerably, depending on the scope and the duration of the trial and on the participants. There are clear cultural differences between study organization in the U.S.A., Japan, Italy and West Germany. In the latter country the number of committees is minimized and the organization is kept as simple and as cheap as possible. However, every study organization has at least three levels, as indicated in Fig. 1.

The highest level contains the financing agency and the policy or advisory board. The latter can be used simultaneously as a protocol review committee, if the structure is to be simple. The second level contains the principal investigator and the study coordinator, who might be the same person. The three possible committees at this level – executive, steering and editorial committees – consist of members of the participating clinics and the coordinating, statistical and data processing centres. The third level consists of these centres, the participating hospitals or clinics and other central facilities. This is the level of day-to-day operation.

If things are made as simple as possible, there are – in addition

Study Organization

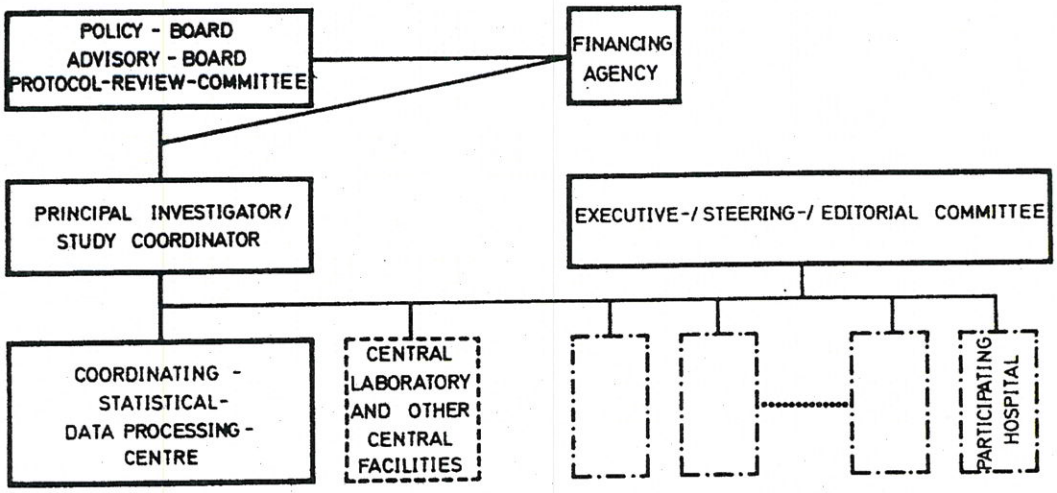

Fig. 1. Diagram to show three levels of study organization.

to the hospitals – five independent units in this structure, as indicated by the boxes in Fig. 1. Separating the tasks somewhat gives up to 13 independent organizational units, as indicated by the individual phrases in the boxes of Fig. 1. The number of persons and the number of committees involved have no empirical relationship to the scientific results. European studies are usually a little less structured in comparison with studies in the U.S.A.

Inclusion criteria and patient log-book. In defining the inclusion criteria, a practical problem is how far one should go into detailed description of symptoms, signs and measuring procedures. It is better to have too much detail than too little. A more severe practical problem is often that of getting an accurate patient log-book, containing all patients with the disease in a given hospital, as well as those *not* included in the trial for various reasons. Such a log-book is necessary for assessing the selectivity of different hospitals. It provides an epidemiological background for the study.

Determining the endpoints. In choosing the endpoints, one navigates between Scylla and Charybdis: on the one side are clear-cut events such as death, which are rather infrequent; on the other are less

clearly defined but more frequent events, such as certain side effects, clinical improvement or changes in laboratory values. The endpoint affects the number of subjects required.

Number of subjects. Determining the number of subjects envisaged is not only a statistical problem, to be solved by a set of defined equations. The hospitals, patients, money and time available, the selection criteria and the possible endpoint, the size of the expected effects, the drop-out rate and other factors all have to be taken into consideration. Usually one examines different combinations of the items mentioned and makes an educated guess. Determining the number of subjects is a practical and empirical problem. The clear-cut theoretical or statistical solutions only give hints to the investigator responsible for making decisions.

The more subjects used, the more information will be acquired. The less subjects used, the better for ethical reasons. Statistical formulae provide no clear-cut escape from this dilemma. One has to decide with responsibility, relying on judgment and experience. If one has 100 to 300 patients per group, well defined inclusion criteria, reasonably frequent endpoints and a drug effect ratio between groups of around two, one can hope to start a worthwhile study. Studies with several thousand patients per group might be necessary to answer certain questions. They should not, however, become standard practice in long term clinical trials, because of the various practical and theoretical problems that they pose. The number of patients actually recruited is often lower than the number estimated to be necessary.

Multi-centre trials. The practical problems in multi-centre trials are legion. During the planning phase, it is particularly time-consuming to develop the protocol, to secure the involvement and consent of many different institutions and to implement the compromises necessary to get one coordinated trial instead of several separate ones. One has to allow for different numbers of patients in the hospitals and for the drop-out of some hospitals during a study.

Randomization problems. Randomization must be extensively planned in all practical details. Sealed envelopes containing the code for each patient must be numbered and sent back to the central organization after the trial. The randomization plan must be unknown

to the physician and there should be no possibility whatever of predicting from one patient the drug assignment of the next. It is intellectually attractive for physicians and patients to find out what treatment will be administered, so one has to take the necessary precautions. The statistical method of randomization is not so important as the practical problems involved.

Problems with double blinding. In planning a double blind study one has to go into a lot of detail. Taste, smell, colour and weight of the preparations have to be identical. If the drug changes some laboratory values, double blinding is usually not achievable. It is not sufficient to say that a double blind study was planned; one has to give detailed information on how this should be achieved and on the extent to which it actually was achieved. One can for instance ask the physicians which treatment they think the patients received. These individual estimates can then be compared with the actual treatments [2].

Extent of variables covered. In planning a long term trial one has to decide on the variables to be covered and documented. How often should patients be seen? To what extent should prognostic factors be included? It is tempting to argue that once the patient has been recruited, more information can be obtained without a significant increase in cost. However, attempting too much will make the data acquisition time-consuming, less practicable and inaccurate. A minimum of relevant data is better than a lot of missing or even wrong values.

Patient consent. When obtaining the patients' consent it is important to conform to legal requirements, which vary from country to country. In West Germany there is a minimum requirement to give information verbally on all relevant aspects of the trial. This must be done in the presence of at least one third party, for example a nurse. Written consent is advisable, even if it decreases the number of patients participating. In long term trials one must expect that patients change their minds after some time and withdraw.

Blind advisory group. In most trials it is worth having an advisory group to decide on the exclusion terms of individual patients without knowing their treatments. Such a group can be composed of external experts or of clinicians participating in the study. In the latter case

one has to make sure that there is real blindness; the participating clinician should not judge the patients from his own hospital.

3. Practical Problems in Execution

Execution of a long term trial over many years requires an efficient organization. Success with a long term trial is at least as much an organizational achievement as an intellectual or scientific one.

Responsibility within each hospital. Every clinic or hospital is responsible for the integrity of its data. There should be one scientist in each hospital who trains the co-workers and who continuously monitors adherence to the protocol. He can also be responsible for the organizational problems which arise within the clinic. Substitute personnel covering every function should be available, in order to make sure that the trial goes on during holidays or illness. As long as the medical director of the clinic is interested in the trial and as long as he is supporting it, this will encourage interest at other levels.

Central monitoring of a trial. A long term multi-centre trial must be centrally monitored, using telephone, telex, letters or personal visits as appropriate. There will be a continuous flow of data between the hospitals and the coordinating centre. The channels of communication should be specified in advance.

Continuous data entry and correction. After the central admission of a patient and his randomization, information will come in at specified intervals. Data sheets are checked for completeness of entries, admissibility of values and for plausibility and logical consistencies. Part of this continuous data entry and correction has to be done manually, part of it can be carried out by computer. The important task of continuous data entry and correction takes a lot of time and detailed work which are often underestimated.

Regular reports to hospitals. The performance of every clinic can be enhanced by regular reports. They should detail the numbers of recruited and excluded patients, updated information and data correction, other relevant information and problems which might have arisen. These reports can be sent monthly or at longer intervals.

It is stimulating for the clinicians to see the numbers at their own hospitals and at others, in order to compare the development of the trial at different places. A policy of open information is a good way of ensuring cooperation over a long period of time. New results from other studies and information from meetings outside the trial should be given to the clinicians quickly.

Interim evaluations and early termination of a trial. There are several statistical approaches to the problem, which may lead to different conclusions from the same data. There is no general agreement on the procedure for stopping a trial, prior to the end of the plan. As in other fields, statistical methods can be helpful but should never be used to provide an automated decision rule. Decisions by ethical committees might introduce a bias in favour of stopping a trial too early; democratic decisions seldom yield scientific truths. In my opinion, responsibility for stopping a trial should lie with an experienced investigator answerable to his peers. A practical way out of the dilemma might be to set up a decision rule at the very beginning, but not to follow it slavishly. Evaluations could be performed at reasonable intervals, not more often than three times, so that evidence has a chance to accumulate. The trial should then be stopped if the evidence is convincing, but not merely as dictated by statistical tests. In my opinion one has to come back to the individual judgement of the experienced scientist, comparing all available evidence and using sophisticated models only for what they are: tools which can sometimes be useful when the results are coherent with other information. As long as there is no general agreement on conclusions to be drawn from given results, one can only follow such an empirical course.

4. Practical Problems in Evaluation

Selection bias. For most long term studies there is a lack of data about patients not admitted to the trial. We usually know very little about the criteria of selection and about the reference population. Nevertheless, a trial may be valid in respect of the assessment of drug effects. Effects so established are extrapolated from the sample observed to the whole population by analogy, not by statistical reasoning. Determining the selection bias is an unsolved practical problem in most clinical trials.

Missing data. In every long term trial there will be missing data. It remains a matter of judgement when to estimate such data by various means or when to ignore the patient records with missing values. There are no strict rules. Of course, every patient taken into the trial has to be counted in the final evaluation. Some basic data should be available from all patients; one can use such data to estimate the selection effect of the group with missing values.

Baseline data and comparability of treatment groups. The comparability of treatment groups should always be evaluated. Sometimes groups will not be comparable, possibly owing to a failure of the randomization procedure. This can be corrected by using stratification or other statistical adjustments for baseline differences.

Compliance. Careful scrutiny of a clinical trial almost always reveals deficiencies in patient compliance. The problem is particularly serious for long term trials, where motivation tends to decrease with time. In every long term trial there should be an evaluation of compliance. One practical question is whether patients not complying should be admonished during the trial and encouraged to conform to the treatment. Another question is whether to inform the physician about the compliance of his patients.

Generating hypotheses and using exploratory procedures. After the primary endpoints have been tested and evaluated, one need not stop the statistical evaluation of a long term trial. Such trials are so expensive and contain so much valuable information that one is stimulated to look at more than one question. Long term studies generate systematic information on the natural history of diseases and on the diagnostic and therapeutic procedures in use at particular times. A variety of statistical procedures are available for exploratory analyses and hypothesis generation. They should be used more extensively. Assessing the effect of risk factors and comparing it with the drug effect, examining the responders and non-responders to treatment and observing the stability of conclusions resulting from different methods of statistical analysis all provide valuable information. One should not shy away from trying various methods of analysis. Long term trials are at least partly observational studies in which one can use the corresponding techniques.

Conflicts of interest in evaluation and interpretation. Using different methods of analysis on the same data base, one may end up with different results. Manipulation by statistical analysis is therefore possible, for instance by suppressing certain information in the publication. It is good practice for scientific journals to require great detail on all aspects of a long term trial in order to rule out such possibilities. It is however indicative that only very few long term trials have so far opened their data bases completely to another evaluating group, with a different or neutral attitude to the study in question. Such exchange of information is highly desirable, even if it involves much painstaking work. Conflicts of interest and differences of interpretation often arise between the participants in a study. This may delay publication and valuable information may be lost through compromise. The statistician has to be the good conscience of the group, not allowing conflicts of interest to result in biased interpretation. This idea emphasizes that statistics is not merely a set of rules but a profession with a defined responsibility requiring personal integrity and judgement.

5. General Problems

Methodology and medical content. Sound methodology and medical content should be in harmony. Biostatisticians need to be well informed in the various medical fields, in order to apply appropriate methodology. It is not sufficient to be a good statistician: with a long term trial one has to become expert in the respective scientific and clinical fields. It remains a practical problem how to educate such people more efficiently [3].

Management problems. Conducting a long term study in medicine is a management problem involving special skills not necessarily possessed by the statistician or the clinical investigator.

Coping with problems from outside the trial. In conducting a long term trial one is not isolated from extraneous factors. Papers may be published which change the relevant scientific knowledge; the law may also change, as may the attitudes of patients, doctors and participants. Sometimes the impact from such extraneous factors is very strong and may lead to the termination of the study.

Risk–benefit evaluation. A risk–benefit analysis of a long term trial is highly desirable. There are however no satisfactory procedures or statistical techniques for such an assessment. The methodology is insufficient and needs further development.

Conflicting results from different trials. There remains the practical problem of coping with conflicting results from different studies. Usually a careful analysis of inclusion criteria, the drug regimen, the endpoints and other details will provide an explanation for these different results. There is a deficiency in the methodology available for this problem.

6. Conclusion

It is clear from the above catalogue of practical problems that there are serious difficulties in the conduct of long term clinical trials. However, they will continue to be needed as the only effective means of systematically accumulating knowledge of chronic diseases, their diagnoses and treatment.