

Chapter 8

Type of Studies - IV. Intervention Studies

New therapeutic concepts arise from time to time. Drug discovery and development is the method by which an idea in the mind of a research worker is converted into a safe and effective therapeutic agent for the treatment of disease. Ideas for new treatments arise from a variety of sources. Some therapeutic hypotheses are suggested by the mechanisms of disease at the cellular or molecular level. Other ideas come out of clinical observations when patients being treated for one condition show improvement of symptoms of another. Some ideas arise out of epidemiologic studies. The road from the identification of a promising new drug or treatment to its becoming standard treatment is a long and arduous one. It goes through four stages viz. drug discovery; pre-clinical; clinical and post-registration development. In this chapter we would consider only the clinical stage.

Whatever the source of the new method for treating a disease, it becomes necessary to put it to test, because a variety of conditions like coincidence, faulty comparisons, or spontaneous changes in the course of disease can obscure the true relationship between treatment and its effects. It is almost always necessary to test therapeutic hypotheses by means of clinical research in which data are collected on the clinical course of patients who have actually experienced the treatment.

Clinical Trials

Clinical trials are a special kind of cohort study in which the conditions of study like selection of treatment groups, nature of interventions, and management during follow-up are specified by the investigator for the purpose of making unbiased comparisons. Clinical trials are the only study design that approximates laboratory experiments where all extraneous influences are tightly controlled by the researcher. They are used widely for assessing drug therapy, and also for evaluating methods of clinical management of particular conditions (e.g. child birth), for comparing medical versus surgical management (e.g. coronary artery disease), for evaluating health services, health administration, and so on. Whatever has been discussed about cohort studies applies also to studies of treatment.

Clinical trials are generally reserved for the final mature stage of investigating a disease. By this time the basic descriptive characteristics have been studied, and the Who? What? Where? When? aspects of the disease are reasonably established. A thorough work-up for establishing the etiology has been done, and the clinical as well as public health approaches for dealing with the disease are already in place. At this stage, clinical trials help to test the efficiency of treatment and provide greater insight into the underlying pathology. Well designed clinical trials have helped to establish the most efficient regimens for the treatment of tuberculosis, for the management of acute lymphocytic leukaemia, and are being conducted at present for the management of coronary heart disease. In clinical trials subjects are enrolled on the basis of exposure, but the distinguishing feature is that the investigator assigns the exposure. That is why clinical trials can provide data of high quality approaching those in controlled

experiments. If treatment has been allocated at random and if the sample is of adequate size a high degree of validity can be achieved, which is not possible with other observational designs (see figure 8.1).

Prior to conducting a trial one should carefully consider whether the intervention is well enough developed to permit evaluation; whether the preliminary evidence that the intervention is likely to be beneficial including some appreciation of the size of the treatment effect?

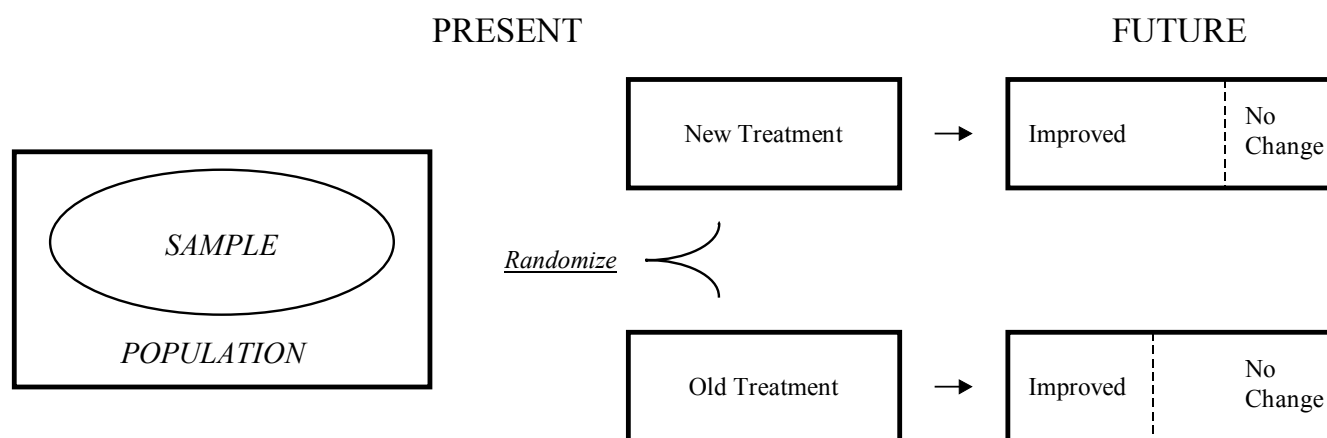


Figure 8.1: Basic Design of a Clinical Trial.

The patients to be studied are selected from a larger number of patients with the condition of interest. This requires the definition of inclusion and exclusion criteria. Broad **inclusion** criteria make the task of assembling the study subjects relatively easy, and the findings are generalizable better. One also needs to take into account the rarity or otherwise of the outcome in selecting inclusion criteria. Well thought out **exclusion criteria** will help exclude those with extraneous conditions which could distort the outcome, or those who possess contraindications for the interventions, or are unlikely to comply with the treatment. Another matter to consider when deciding about recruitment of subjects is the sample size. One ought to have a reasonable sample size in order to come up with acceptable results.

The next step is to measure base-line variables. The purpose is to define and describe the characteristics of the study cohort; to ensure that the disease stages in study and comparison groups are similar; to decide on the outcome variables, and to measure predictors of the outcome. Measuring baseline variables at the outset helps in the statistical adjustment of the results to reduce the effects of chance maldistribution in baseline factors.

The subjects are then divided into two groups of comparable prognosis. One group, called the experimental group, is exposed to the new treatment. The other group, called control group, is treated the same in all ways except for the new treatment. Any difference in the outcome is attributed to the intervention.

The main reason for structuring clinical trials in this way is to avoid bias when comparing the value of two or more kinds of treatments. Physicians regularly apply therapies tested in groups of patients to individual subjects. The likelihood of success in an individual patient depends on the degree of certainty evident in the group, and the scientific strength of the method used. The validity of clinical trials depends on how well they result in an equal distribution of all determinants of outcome, other than that being tested in the treatment and control groups. Randomization eliminates bias caused by any existing confounding variables. It ensures that there are no systematic differences between intervention groups with regard to factors that may affect outcome. Blinding for intervention eliminates bias in the choice of subjects for interventions (e.g. patients being assigned unintentionally to new treatment because of age, sex, or severity of disease). Blinding for assessing the outcome eliminates bias in judging the outcome. Any pre-conceived views in subjects and clinicians cannot influence the assessment of outcome.

Efficacy, Effectiveness and Compliance

These factors matter in all clinical trials. What the researcher is asking at first is "Can the treatment work under ideal circumstances?" Then the next question is "Does it work in ordinary settings?" The terms efficacy and effectiveness are applied to these concepts. So efficacy is a measure of the success of a new intervention in the circumstances of a clinical trial performed in the artificial environment of a research institution. Effectiveness is a measure of the success of the same

intervention when implemented in the rough and tumble of a health service. Compliance is the extent to which patients follow the treatment.

Issues of Concern in Methodology

The researcher must address four areas of methodological concern. These are:

- entry criteria or how people get into the study;
- intervention, which must be reproducible, practical and free of bias;
- subject allocation to different types of intervention under study; and
- selection of comparison groups.

Entry Criteria

The kinds of patients that are included in a trial determine the extent to which conclusions can be generalised to other patients. Patients who are not included in a trial are those who do not meet the inclusion/exclusion criteria, refuse to participate, or are considered unlikely to co-operate once they have been enrolled. Naturally if subjects are carefully screened to exclude such people the evaluation would be much tidier. But reality is different. Many patients have several health problems, take a variety of medications or at times forget to take their pills. Trials which are high on exclusion and focus strongly on internal validity by optimizing conditions for testing interventions are called **explanatory trials**. They are intended for finding out whether the new treatment really works, and if so, how well. The focus of such trials is on efficacy. By contrast, **pragmatic trials** evaluate treatment under routine clinical situations. Their focus is on overall effectiveness (including placebo effect, if any). These two types of trials are sometimes referred to as efficacy trials and efficiency trials respectively. In explanatory trials intermediate outcomes are also measured to gain biological knowledge. In pragmatic trials analysis is always by intention to treat even if subjects drop out or change groups.

Patients who refuse to participate are systematically different (by social class, severity of disease, other problems) from those who agree to enter the trial. Similarly, those who participate are also likely to differ from non-participants in several ways which may affect the rate of development of the end points. Factors like age, sex, socio-economic status, education, health consciousness can all affect subsequent morbidity and mortality, and generalizability. For all these reasons, patients in a clinical trial generally end up as a highly selected biased sample. Hence in planning clinical trials careful thought needs to be given to generalizability when selecting the study subjects and the interventions.

The sample size should be large enough to overcome the influence of any chance sources of error, and the randomization procedure should be carefully checked for any technical errors.

Another issue to consider is whether the diagnosis is accurate? How was the condition under treatment defined? Do the subjects really have the disease they are supposed to have? What were the

criteria for entry? Are the methods used for classifying subjects practical and reproducible in smaller health facilities? For all these reasons patients from specialty centres do pose a problem. They represent a selected slice of the problem rather than the full spectrum.

Intervention

The treatment regimen should be carefully defined, be reproducible in simpler settings, and should make practical sense. Compliance bias arises when there are dissimilarities in the treatments being compared and cause differences in patient adherence to treatment (e.g. a difficult diet; bad taste; and so on). When the overall compliance for all the treatments being compared is low (e.g. due to social stigma in the case of tuberculosis or leprosy), the beneficial effects of any one can be masked. Another problem is that of other competing interventions, with which the patient continues and they go unrecognized. They may be the cause of the good outcome. Or one of the groups may change their habits or life-style in a manner that reduces an undesirable outcome.

Subject allocation

Bias invariably occurs when the groups being compared are unequal with respect to risk factors or existing conditions which can influence outcome. If these are identified at the outset the subjects may be stratified according to prognosis.

Drop-outs from the study pose a major difficulty. Not only are the data lost but also attrition affects outcome if it is related to the treatment. Different rates of drop-outs between the intervention and comparison groups can lead to under or over estimation of effects of treatment.

The end points of the trial should be carefully defined. All the most important possible outcomes should be included. Sometimes a new treatment may turn out to be very effective. In such cases the trial needs to be terminated to make the results known to the scientific community. The same applies when the new treatment has serious side effects and has to be stopped.

Comparison groups

The value of a treatment can only be judged by comparing the results with those of some alternative form of treatment, which should be as appropriate as possible. Controls are ideally selected from the same population as the subjects.

When the outcome in a disease is predictable (e.g. subacute bacterial endocarditis), a separate control group is less important. However, most therapeutic decisions do not involve diseases with such predictable outcomes. If the clinical course is unpredictable, and also varies from one patient to another assessing treatment effects by clinical observation can be extremely unreliable e.g. systemic lupus erythematosus.

There may also be the so-called Hawthorne effect. (The tendency for people to change their behavior because they are receiving special interest and attention).

Which comparison treatment?

The choice of the comparison treatment depends on the objectives of the study. Clinical trials are usually designed as critical tests of the therapeutic alternatives being assessed. The research question must be clear with carefully defined measures of outcome, with realistic estimates of sample size and of event rates in the control group and a plausible reduction in the event rates in the treatment group. As we have seen, clinical trials are of two types: pragmatic and explanatory studies. Pragmatic studies are concerned with deciding about treatment policies. Patients are selected from a wide population; treatments reflect current clinical practices, and major outcome variables are measured. Explanatory studies, on the other hand, are primarily for evaluating new treatments and how drugs work. The sample of patients is homogeneous; treatments are not necessarily the standard ones and comparisons are made with an old treatment, the usual treatment, or placebo.

If the outcome is likely to vary because of age, sex, or stage of advancement of the disease it is necessary to ensure that treatment groups are balanced with regard to these characteristics. To do so the procedure called BLOCKING is done. Each participant is first classified with respect to these variables and then randomized within the subgroup or block for allocation of treatment. For blocking to be effective within-block variation should be much less than between-block variation. By comparing outcome of treatment within the same block the block effects are eliminated in the comparison. Randomization is applied to assignment of treatment to subjects within the blocks to further reduce the influence of unknown variables. Hence the dictum to follow is “Block what you can and randomize what you cannot”.

Depending on the objectives of the study comparison is made with one or more of the following:

- No intervention. Do patients receiving a new treatment end up better than before treatment?
- Placebo treatment
- Usual treatment

Randomization.

By randomization the potential for bias in allocation to study groups is removed. Moreover, the study groups will tend to be comparable with respect to all potentially confounding variables as well as other unsuspected confounders. However this is no guarantee that differences will not arise by chance between the two groups. Such a risk is greater when the number of patients is small.

Blinding (also called masking).

In a clinical trial there are three groups of performers: those who give the treatment (clinicians), those who receive it (patients), and those who assess the effects (researchers). Often the clinicians and the

investigators are the same people. All the three groups involved in the trial are prone to change their behaviour if they know which subjects are receiving what treatment.

Blinding refers to making the various participants unaware of what treatment is being given to whom in order that they do not change their behaviour.

There are four levels of blinding. First, those allocating patients to treatment groups do not know who is assigned what treatment. Secondly, patients are unaware of what treatment they are receiving. Thirdly, physicians and the nursing staff should not know what treatment the patients under their care are receiving. Finally, those who assess the outcome should not know which patient received what treatment.

Blinding is especially necessary when the outcome is easily influenced by the knowledge of what treatment has been received e.g. pain, nausea, disability, and so on. Hard outcomes like death, recurrence etc. are not affected so much.

In blinding careful consideration should be given to the following:

Will it be ethical? The double blind procedure should not result in harm or undue risk to patients.

Will it be practical? For some of treatments (e.g. surgery vs. medical treatment) it may not be possible to arrange a double blind trial.

How well is bias avoided? The purpose of blinding is to avoid bias, and one needs to assess how serious the bias might be without blinding.

Is a compromise possible? For example, one may opt for partial blinding by using independent blinded evaluators.

Blinding is not always possible. Open (non-blind) trials or single blind trials (either only the investigator or the patient is blind to the allocation) are sometimes unavoidable. In trials of different types of patient management, or of surgical procedures, or alternative therapies full blinding is often impossible. In such cases the researcher can take comfort from the fact that blind assessment of outcome is often more important than blinding the intervention.

Now that we have discussed the general framework of clinical trials let us next consider some of the practical issues in their design.

Some of the **practical considerations** in planning clinical trials are as follows:

- 1). Eligible subjects should be recruited before assignment to groups is determined. This guards against bias in recruitment and differential refusal rates among study and control subjects.
- 2). The protocol designed should be such that it asks a relevant clinical question, is easily replicable, is capable of being "blinded", and is resistant to drift during the course of the study. Thought should be given to the fact that it may be necessary to deal with complications arising from the disease or the treatment, or to respond to new problems. Also allowance should be made as to how optimal blood level of the drug is to be achieved and monitored without the clinician coming to know what the treatment is.

3). In selecting alternative treatment or placebo, a cross-over design may be helpful to bring out the difference in response. Cross-over studies have the advantage of not needing large sample sizes compared to classical parallel studies. In cross-over designs the switching over should be determined by predetermined time intervals rather than by patient response to therapy or disease state. There may also be a need to have a washing out period during which response is not being assessed.

4). The sample size should be adequate. As we saw in Chapter .. (page ..) a clinical trial with a small number of patients carries a risk of failing to demonstrate a difference when one exists. A few minutes spent on calculating the sample size provides useful guide about how many subjects to recruit.

Drop-outs, Withdrawals and non-compliance

These can create several difficulties and particularly with regard to bias. The number of patients dropping out of each group, the reasons for their withdrawal and the potential bias resulting from their withdrawal should be carefully considered. A withdrawal rate in excess of 20% causes serious concern about the validity of the trial.

In the case of high withdrawal rate analysis of the data can take one of the following approaches:

1. Intention - to - treat analysis. This is analysis of outcome by groups as randomized (including the drop-outs). It is the preferred approach for pragmatic studies. This is the preferred method for dealing with the problem of withdrawals. The difficulty arises because the decision to withdraw from a trial is often related to the performance on treatment. For example, the subject develops side effects and decides to drop out of the trial. If side effects are related to only one of the treatments, then several subjects assigned to that treatment may drop out, and make comparison unreliable. To get over this difficulty, data from all subjects who were randomized to the treatment are included in the analysis. This may involve substituting, for example, the last value recorded of the outcome measure, or the drop-out's average value during the trial. Often both intention-to-treat and 'on-treatment' analyses are performed. If the results differ, it is usually because of a problem of unequal groups, and only the results of intention-to-treat analysis are taken as valid. The same approach applies for non-compliers.
2. Analysis by treatment received. It is the preferred approach for explanatory studies. Usually it is advisable to perform the intention-to-treat analysis in addition to this type of analysis. The reason is that a difference apparent from an analysis by treatment received is less convincing if not supported by a similar difference from an intention-to-treat analysis. The exclusion of any randomised patient from the analysis can lead to biased results. The reason is that in a subtle way it changes the research question. One is performing the clinical trial to monitor the actual effects of a treatment and so has randomized on the basis of offering a treatment. Hence theoretically only the entire groups allocated to treatments by

randomization are truly comparable. Once patients have been randomized to a treatment group their subsequent health experience must be assessed and analysed along with that group regardless of whether they comply with the regimens. The comparison that is optimal for assessing the benefits of an intervention is to analyse by the group patients were assigned to. The rule of thumb to follow is “Once randomized, always analyse”.

Stopping Rules

It is helpful to have an independent monitoring group separate from the investigators. If a new treatment is found to be particularly successful or obviously harmful then early termination of the trial and publication of the results is mandatory.

Significance testing or Confidence Interval

The purpose of statistical significance tests is to assess whether observed differences between treatment groups could possibly have arisen by chance, or whether they indicate real effects. This is fine for the day-to-day practice. But it detracts from the real purpose of most research which is to determine the magnitude of the difference between groups. Confidence intervals give a range which includes the real treatment differences. A wide confidence interval indicates lack of precision which means inadequate sample size. Calculating the confidence interval is always important, and particularly so when the statistical tests indicate non-significant differences.

It should be obvious by now that the design and conduct of clinical trials requires a great deal of time and effort. There are situations where the benefits are not commensurate with the effort. For example they are less suited in the following situations:

- 1).When multiple regimens are being compared (because of the large number of subjects who need to be recruited)
- 2).For small changes in the therapeutic plan (the effort needed to do the study may outweigh the benefits)
- 3).Therapies are being changed in the course of the study.
- 4).Treatment with rare unexpected outcomes, or where outcomes become noticeable in distant future.

Ethical considerations may also limit the usefulness of clinical trials.

The main problem is the dilemma of balancing the internal purity of the study with external validity. Trying to achieve extreme methodological purity may result in findings that are difficult to apply in the usual clinical setting. On the other hand loosely constructed studies are very likely to be challenged.

Common Variations in the designs of Clinical Trials

Depending on the research question and the situation five types of common variations are usually found in the designs of clinical trials. These are as follows:

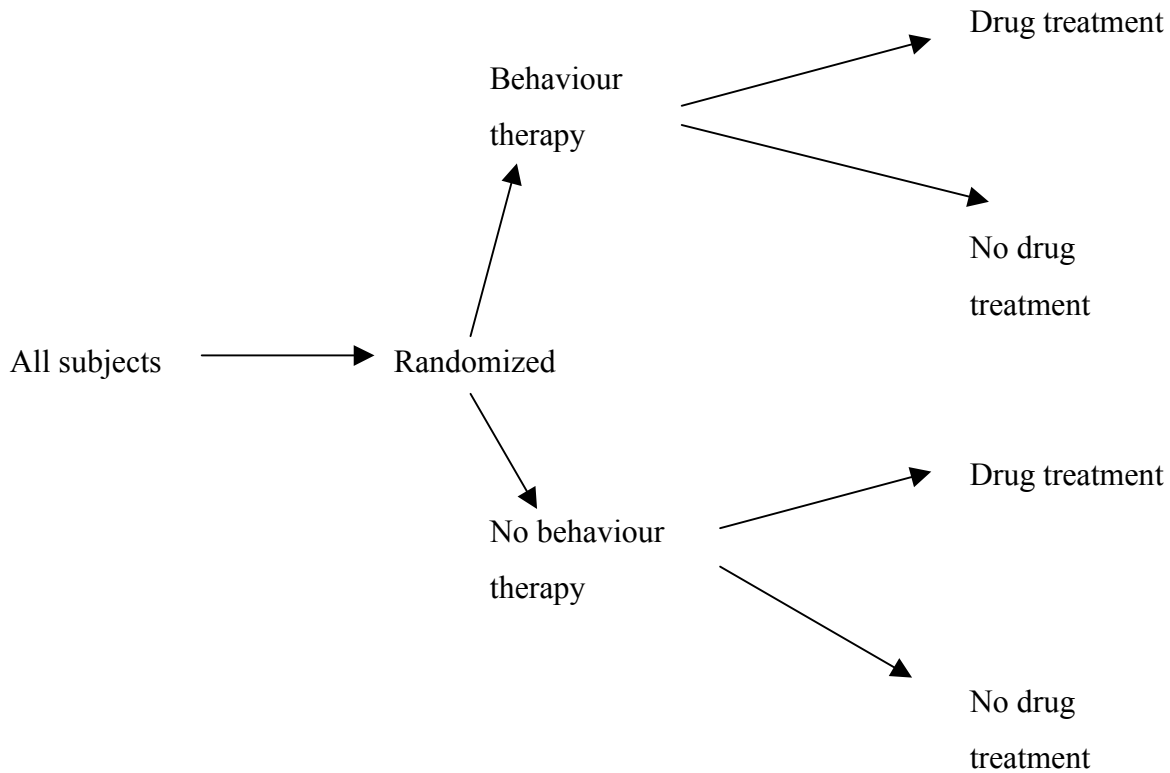
Run-in Design.

All the patients are at first put on a placebo. After a few weeks those who have complied are randomized. This allows the researcher to check who is likely to comply and who will not. Another advantage is that it allows for any previous medication to get washed out of the system before base-line measurements are made.

A variant of the method is the use of an active drug instead of placebo during the run-in period. Those who do not show side effects to the active drug are randomized for treatment with a related new drug.

Factorial Design.

In this design two types of treatment are being compared simultaneously e.g. behavioral therapy and treatment with drugs. The subjects would be first randomized to say behaviour therapy, and then within each group to drug treatment. Factorial designs enable the researcher to investigate the separate and combined effects of both treatments. When all possible combinations of the levels (values) of the independent variables are included the design is said to be fully crossed. If one or more are left out it is called incomplete.



Matched-pair randomization

In this design subjects are selected in pairs matched on factors like age, sex, severity of disease and then randomly assigned to interventions. This design helps to balance out baseline confounding variables.

Group randomization

Instead of randomizing individuals groups are randomized e.g. villages or factories, or patients from separate general practices. The study thus takes into account naturally occurring clusters.

Time Series Design

In a group of subjects serial measurements are performed sequentially during treatment and control periods. Each subject thus serves as his own control. The advantage is that factors like age, sex, residence, social class and similar others get eliminated as confounders. The sample size doubles because each subject is now providing two sets of observations viz. intervention and control. The disadvantage is that of *carryover effect* in the form of residual influence of the intervention during the control period after it has been stopped.

Time Series Designs sometimes use repeated measure strategy. Intervention is repeatedly started and stopped. If the start and stop periods show matching patterns of outcome then the influence of confounding factors can be ruled out.

Cross-Over Design

Half the participants are randomly assigned to start with placebo treatment and then switched to active treatment while the other half do the opposite. The cross-over design permits between group *and* within group analyses. It has the advantage of control over confounding variables and effective doubling of the sample size. The disadvantage is that the duration of the study also doubles, besides the analysis becoming slightly more complex. The cross-over design is a good choice when subjects are few and difficult to recruit, and when carry-over effect is not likely to be a problem.

The cross-over design helps to overcome a major problem with conventional parallel group randomized trials. Patients vary so much in their disease state and response to treatment. This means that the investigator should observe substantial numbers of subjects on each treatment in order to assess reliably the difference in outcome. The cross-over design helps to overcome this difficulty. Moreover, it offers two separate ways of assessing the effect of treatment. First, comparison of one group against another; and secondly, each subject's response to one treatment against another.

n-of-1 trials

Randomized controlled trials are the best evidence of drug efficacy. There still remain a number of clinical situations in which treatment decisions on an individual patient cannot be based on such evidence due to a variety of reasons. Either no controlled trials have been carried out, or if carried out, the patient does not meet the eligibility criteria. In such a case the clinician may wish to examine the effect of treatment in individual patients whilst still maintaining the methodological safeguards provided by the randomized controlled trials. As we have seen such safeguards are needed because the clinician's judgment about response to treatment can be influenced by a number of factors like natural fluctuation in the disease; the placebo effect; and the patients' reticence to tell the doctor that the treatment did not work. Randomized controlled trials in individual patients (N of 1 RCT) have been

developed to deal with such situations. In the N of 1 RCT the patient undergoes pairs of treatment periods (one period with the active treatment and one with placebo or old treatment), assigned at random. Both the patient and the clinician are "blind" to allocation. The treatment effects are monitored, by designing quantitative measures of the patient's symptoms by means of diaries or questionnaires. Pairs of treatment periods are continued until effectiveness is proved or refuted. There should be at least three replications of alternative treatments before one can give a confirmed opinion about the efficacy or otherwise of the treatment.

N of 1 RCTs are best suited to chronic stable conditions. Ideally the therapy being assessed should have a rapid onset of action, and cease to act soon after it is discontinued. With such methodology the N of 1 trial becomes a more accurate test of the efficacy of treatment compared to open uncontrolled trials.

Field Work in Mother and Child Health

MCH activities are different from clinical management of individual patients. They are intended for things to happen in groups of families and communities e.g. promotion of immunization, oral rehydration therapy, health education, and so on. Interventions have two distinguishing characteristics. The interventions are programmes or policies, and the outcomes are the presumed effects of these activities. In order to assess the effectiveness of various inputs a number of research designs are employed depending upon available resources. These are as follows:

One Group - after only Design

The subjects are exposed to the intervention and the effects documented later e.g. a mass media programme for the promotion of family planning. Such a design allows only one value for the outcome variable viz. subjects exposed to the programme. Since the base line value of couples practicing family planning is not known it is not possible to measure if the proportion changed after the programme was introduced. Hence, neither the evidence of any links between the programme and its effects nor the evidence that the proportion of couples practicing family planning changed because of the programme can be provided.

Two Groups - after only Design

This design is an advance over the previous one. Addition of a group which did not receive the programme provides more than one value for the input variable. Now we are able to determine whether input and outcome variables are associated.

However, the research design does not include the pre-programme measure to determine whether the dependent variables varied after the independent variable. It is therefore inadequate for providing the evidenceneeded. Besides there are many sources of spurious association in the design like, for example, differential selection, self selection, differential attrition and so on.

One group - before and after Design

This is a popular design. It allows the researcher to determine association by demonstrating a difference in the outcome variable before and after the programme. This design is an advance on both the previous designs since it allows the researcher to establish the time sequence. But there still is the problem of spurious association, and caution in drawing conclusions is needed. Factors other than the programme may be responsible for a change if change was at all noticed. There are five causes of spurious association:

- 1). Historical factors. These are influences external to the programme or policies e.g. change in employment, change in family structure because of some members migrating and so on.
- 2). Maturation. Changes occur in people independently of the programme e.g. people getting older, or families getting larger, and so on.
- 3). Testing. If evaluation required some form of testing, then previous experience of testing may change subsequent performance.
- 4). Instrumentation. If evaluation required the use of some appliance or instrument, then changes in the measuring process may be responsible for the difference observed.
- 5). Regression towards the mean. Extreme values at the time of first measurement normally tend to move towards the mean at subsequent measurements. This is often overlooked.

Two Groups - before and after Design

Because there are values of the outcome variable before and after the intervention such a design allows us to determine an association between programme activities and outcome. The research design also allows the investigator to show that change happened after the introduction of the programme activities. (If there was no change demonstrable then there may be other influences working, and these should be searched for before concluding that the programme was ineffective). Similarly, even if there was evidence of association, it is worth bearing in mind that comparison groups do not remain the same on characteristics (variables) other than the study variables.

This design is a good choice. If comparison groups were the same in all characteristics before the programme and remained so throughout the study except for the input variables then this design is as good as any. Its weakness lies in the fact that one can never be too sure whether the comparison groups are similar in all characteristics.

Some of the difficulty may be avoided by matching. In community studies it may not be possible to match on all possible attributes, and commonly age, sex, education, and socio-economic status are employed to achieve comparability. Random allocation is another way of achieving comparability.

Two Groups - after only with random allocation Design

The major limitation of this design is that randomization does not guarantee identical experimental and control groups on all variables except the input variable. Moreover, since input and outcome variables are not measured before the programme is commenced the researcher cannot be certain that the two groups are comparable on these variables before the experiment.

Two Groups - before and after with random allocation Design.

In this design measurement of the outcome and other variables before commencing the programme allows the researcher to determine how well random allocation worked. It is a robust design, but the contaminating effects of the factors causing spurious association should be borne in mind. One also must be cautious about generalization. The programme in combination with something else might be responsible for the variation in the outcome variable, and so the results of the same programme in other communities will not necessarily turn out to be identical.

Four Groups - 2 before and 4 after with random allocation Design

This design is actually two intervention designs in one. It has two groups in which before and after measurement is done, and two additional groups in whom only after measurement is done. Naturally such a design is more expensive and also time consuming.

Problems and Difficulties in designing Clinical Trials

In clinical trials problems arise because of bias, drop-outs, and sample size.

Control of bias

It is important to eliminate bias in the allocation of subjects to treatment groups, and in the assessment of response to treatment. This has been referred to in previous pages under randomization and blinding. It is also necessary to point out that various types of designs have been evolved in order to marry the need for controlling bias with the practical reality presented by different types of illnesses.

Sample size

In clinical trials selecting a sample of adequate size is particularly important so that one can be confident of detecting a clinically important difference should it exist, and also to be confident of having excluded the possibility of such a difference should none be detected by the study. The important considerations in calculating sample size for clinical trials besides those described in Chapter 3 are the following:

1. What features are to be used to assess response to treatment?
2. What is expected to happen in the control group? If the outcome variable is a continuous one, then mean and standard deviation are needed for calculating sample size. These can be obtained from a pilot study or from the literature. In the case of a discrete variable one needs a similar estimate of the event rate.
3. How small a difference one considers important and clinically significant?
4. How certain does one want to be of detecting a difference; that is power of the study. Generally, a power of 80 to 90% is regarded as adequate.

Appendix 8.1

A checklist of possible bias in clinical trials.

- What types of patients are considered for the trial? Are they to be similar to those commonly seen in practice or different?
- Are the patients for inclusion in the study an unbiased and randomized sample of all the patients with the disease.
- Are patients to be allocated to treatment groups without bias?
- With what treatment is the experimental intervention to be compared?
- What is the experimental intervention going to be?
- How will compliance with the treatment be measured / ensured?
- Are patients to cross over between treatment groups after randomization ?
- How is drop-out from the study to be avoided?
- Are there any unknown interventions other than those being studied?
- Are all clinically relevant outcomes to be reported?
- Will the outcomes be sought with equal vigour in both groups?
- Is the outcome well defined?
- Could the observed results occur by chance?

Appendix 8.2 Calculating sample size for clinical trials

The size of the sample in a clinical trial depends on a number of key questions:

1). What is the main purpose of the trial?

The answer determines the end points to be measured.

2). How will the data be analysed to detect the difference in outcomes between treatments?

The simplest forms of analysis are the differences between means and the χ^2 test. A 5% level of significance as showing evidence of treatment difference is a convenient starting point.

3). What type of results does one anticipate with placebo (or the standard treatment if it is to be used for comparison)?

4). How small a treatment difference is it important to detect and with what degree of certainty?

Very large treatment differences can be shown with small sample size. It is crucial to be able to identify the smallest difference that will be of clinical value, and it would be undesirable to fail to detect it. (It may be argued that any treatment benefit is relevant, but this is unrealistic and calls for a very large sample.

With regard to the degree of certainty with which treatment difference can be detected it is convenient to start with 80%

Calculating for qualitative outcome

The most common approach is to focus on a dichotomous outcome i.e. whether the outcome in each patient can be considered as “success” or “failure”.

The next step is to choose the following four items:

P_1 = percentage of success expected on the standard treatment.

P_2 = percentage of success expected on the trial treatment.

α = level of χ^2 significance test for detecting a treatment difference (usually set at 0.05)

$1 - \beta$ = degree of certainty that the difference $P_1 - P_2$, if at all present, would be detected (usually set at 0.8 (80%) or 0.9 (90%)). $1 - \beta$ is referred to as the power of the study to detect the difference between P_1 and P_2 .

α is referred to as Type 1 error, or the Risk of False Positive, and β is referred to as Type II error, or the Risk of False Negative.

The required number of subjects on each treatment is given by the formula:

$$N = P_1 \times (100 - P_1) + P_2 \times (100 - P_2) \div (P_2 - P_1)^2 \times f(\alpha, \beta)$$

The value of $f(\alpha, \beta)$ is obtained from the following table:

	β (Type II error)			
α (Type I error)	0.05(95%)	0.1 (90%)	0.2 (80%)	0.5(50%)
0.1	10.8	8.6	6.2	2.7
0.05	13.0	10.5	7.9	3.8
0.02	15.8	13.0	10.0	5.4
0.01	17.8	14.9	11.7	6.6

For $\alpha = 0.05$ and $\beta = 0.2$ (i.e. Power of 80%) the value of $f(\alpha, \beta)$ will be 7.9 and is highlighted in the table.

The method is a guide to sample size. However, it does serve the useful purpose of warning the investigator that if fewer subjects are recruited than indicated by the calculation the power of finding statistically significant difference between the two treatments is decreased and therefore the risk of a false negative conclusion is increased.

Calculating for a quantitative outcome

In many trials the outcome measure is numerical e.g. blood glucose level in a study on early feeding of newborns. The approach is similar to that described for qualitative outcome.

First one needs to specify the mean response μ_1 and its standard deviation sd_1 for the conventional method of feeding.

Next one decides what difference in the mean value achieved with the new feeding method would be important to detect. In other words what value of the difference $\mu_2 - \mu_1$ is important.

The statistical test in such a study would be a two sample 't' test, using a significance level of α at $P < 0.05$. One would also need to decide on the value of $1 - \beta$ i.e. the degree of certainty with which the difference would be detected if it exists.

The formula for calculating the required number of subjects on each treatment is

$$N = 2 \times sd_1^2 \div (\mu_2 - \mu_1)^2 \times f(\alpha, \beta)$$

The value of f is obtained from the table above in exactly the same manner as before.

One difficulty is in choosing the appropriate values of μ_1 and μ_2 and sd. The values may be obtained by doing a pilot run or from similar studies reported in the literature.