MedicalBiostatistics.com

VALIDITY OF CLINICAL TRIALS

Validity is the ability to provide correct answers to the questions under investigation. In the context of clinical trials, besides randomisation and other features of design as discussed, tools to achieve validity are proper selection of subjects, blinding, multiple measurements, and adequate reporting of results.

SELECTION OF PARTICIPANTS

Validity of results of a clinical trial depends heavily on proper selection of participants. They should really represent the target group. When an inordinately large number of eligible subjects are available that pass the inclusion and exclusion criteria, they must be randomly selected for inclusion in the trial. This allows generalisability. One method could be systematic (e.g., every fifth), and the second method is to include consecutive eligible patients arriving in a clinic/ hospital within a specified period. Else random number can be used for selection.

Ethical considerations such as informed consent can preselect a biased group. Some patients or some clinicians may have strong preference for a particular therapy and they can refuse randomisation. Some eligible patients may refuse to participate when they are told that they could be randomised for placebo or the existing therapy. Some may refuse because it is a trial and not treatment per se. Considerable efforts may be needed to keep such refusals to minimum.

In addition, the groups should be such that there is a-priori uncertainty about the efficacy of the test therapy in them. This is called **patient equipoise** and helps to ensure that the patients are homogenous material. Another such term is **clinical equipoise** that is used for collective uncertainty among clinicians about the efficacy of the regimens under trial. (Published definitions of equipoise vary and often conflicting.) Lilford (2001) cites the example of amniocentesis and chorionic villus sampling for clinical equipoise although in his opinion the latter is twice as risky for miscarriage. Clinical equipoise is the condition under which patients should rationally accept randomisation. This also provides insurance against prejudiced assessment of the patients by the investigators.

CONTROL GROUP IN A CLINICAL TRIAL

Controls are needed to realistically assess the difference brought about by intervention. They provide a yardstick against which the gains are measured. Controls might match for baseline because of randomisation but it is also necessary that they be exposed to the same procedure and the same maneuvers as the cases.

There is always a question about using a placebo on patients who are known to have the disease because they need an active ingredient to cure their ailment. However placebo can be used in the following situations.

1. No standard treatment is available, i.e., the existing treatment modality has very doubtful results—perhaps no better than placebo.
2. New evidence has emerged regarding the doubtful efficacy of the existing therapy.
3. The existing regimen is too costly, or is rarely available to the population at large.
4. On patients who have already been given sexisting treatment and have not been benefited, and no second line of treatment is available for them.
5. The test regimen is add-on to the existing regimen. This means that all patients in the trial, including those on placebo, would receive the normally prescribed therapy any way.
6. Patients refuse to accept existing therapy, and are willing to be part of a trial where they know that they can receive placebo.

In situations where these conditions are not met, a group on existing therapy can serve as control.

An apparently simple approach could be comparison of the group on new regimen with a group previously treated with an alternative regimen. The latter are called **historical control**. They must be similar subjects. The flaw in this approach is that some factors may have changed over time. Use it only after assuring that no such change has occurred. If a change has occurred, it should be properly accounted for in the interpretation of results.

In some situations it is possible to give different treatment at the same time to known pairs such as two eyes or two limbs of the same persons. Twin studies also come under this category. Randomisation can be done within each pair to determine which one will receive test regimen and which control regimen. If the trial is on comparison of methods such as pulse oximeter and sphygmomanometer blood pressure readings at the same time in the two arms, many pairs would be easily available. If the trial is for treatment regimen, it would be extremely difficult to find matched pairs such as equal severity of glaucoma in the two eyes, or both limbs with same degree of paralysis.

RANDOMISATION

The difference in outcome can be legitimately ascribed to the intervention when the participants with and without intervention are equivalent to begin with. Randomisation is the process by which the participants are randomly allocated to receive one or the other treatment. It is a very potent tool to achieve equivalence and minimizes self-inflicted bias. Randomisation avoids the chances of biased allocation that can occur when the participants chose to be in a particular group. This works well for trials on a large number of participants but occasionally fails for small samples. If your sample size is small, identify pairs of participants matched for baseline characteristics and randomly allocate one of each pair to the test arm and the other to the control arm.

Randomisation should be done with the help of random numbers. Methods such as alternation and even-odd date of birth are also applied but they can be misused. They are called **quasi-random allocations**. They may not actually bias the trial but the intention is suspect, and is not advisable in a blinded trial. Some trials use nonrandom allocation for which the appropriate term is controlled trial because a control group is still present. Nonrandomised controlled trials are valid only when the test and control groups match for baseline characteristics.

Random or nonrandom, effectiveness of allocation can be checked by post-hoc comparison of the participants in test and control groups. If there is appreciable difference, appropriate adjustments are made at the time of statistical analysis. A less realised importance of randomisation is that it provides a valid base for using statistical methods since these methods require random samples. Randomisation is not random sampling yet it helps in providing a base to use statistical methods.

Box 1: Random allocation in clinical trials

In the wake of unaccounted variations and uncertainties, the best insurance of initial equivalence among the groups by far is randomisation, though this is not a guarantee. By giving equal opportunity to the subjects to be assigned to one group or the other, it is fair to expect that the unaccounted factors such as age, gender, and grade of disease, will be distributed nearly equally, thereby helping to achieve baseline homogeneity across groups. No group is likely to have participants of a particular type that can favour or go against the test regimen. Beware though that random allocation is not random sampling. The former is a strategy to achieve initial equivalence of the groups so that the difference emerging after the intervention can be legitimately ascribed to the intervention (internal validity). The latter is for representativeness of the target population so that the results can be generalised (external validity). The participants should closely mirror the target population.

If the number of available eligible subjects is large (say 4000) and a few (say 30, 40 and 45, in three groups) are to be randomly assigned to the three groups respectively, pick 30+40+45 = 115 distinct random numbers. All these numbers should be less than or equal to 4000. First 30 subjects with these serials are assigned to group I, next 40 to group II and remaining 45 to group III. The website *randomization.com* would do all this easily. In this example the number of subjects in the groups is unequal but generally these numbers would be equal.

If the patients are consecutively attending a clinic, a systematic allocation beginning with assigning the first patient to a random group is easiest to implement. This works well when arrival of patients does not follow any specific pattern. Allocation of patients admitted on odd dates to one group and on even dates to the other group (or any other such mechanism) is fine so long as it is not altered mid-way for some subjects on some pretext. Our experience is that this is open to abuse and not as good as random allocation. Thus this is generally termed as quasi-random.

In large-scale trials, particularly in a community, groups of participants such as schools or clinics are randomised instead of subjects. This is called **cluster randomisation**.

Random allocation can be open so that concerned people know that they are randomised to which group. But the strategy of concealment of allocation is followed almost universally. See Box 2 for details.

Example 1: A randomised controlled trial on nutritional supplementation for GIT cancer

In Italy (Gianotti et al. 2002), a total of 305 patients with preoperative weight loss <10% and cancer of the gastrointestinal tract were **randomised** to receive either ① preoperative artificial nutrition supplementation, or ① + postoperative jejunal infusion (perioperative group), or no artificial nutrition (conventional group). There are three groups in this RCT including one **control** (conventional group).

Side note: The outcome variables were postoperative infections and length of hospital stay. Intention-to-treat analysis and differences between the groups showed that preoperative supplementation was as effective as perioperative administration, and both strategies are superior to the conventional approach.

BLINDING AND MASKING

An important method to minimise bias is blinding. When the patients do not know that they are receiving placebo or therapy then this is called single blinding. This eliminates the possibility of patients psychologically changing their response when they know that they are in the placebo group. They may feel discriminated against. Also, patients who know that they are receiving a new regimen may either exhibit increased anxiety or may have favourable expectations. Bias resulting from all these is called **Hawthorne effect**. In a broad sense, blinding is not merely concealment of allocation but also making arrangements for similar handling of two groups so that this does not break until the results are available.

If the assessing physician also does not know that the patient belongs to test group or control group then this is called double-blinding. This removes possible bias of the physician in patient assessment – at least mitigates any subconscious influence of the assessor on the outcome. Such precaution is an important criterion for validity of the results of a trial. Double-blind RCT is considered a "gold standard" to assess the efficacy of a new regimen. Some times the results are statistically evaluated without breaking the code for case and control group to eliminate statistician's bias. Then the trial is called triple-blind. The codes are broken after the analysis is over.

Although morality issues are attached to blinding because some information is withheld from the participants but it has distinct scientific advantages. It not only reduces possible bias in the responses and assessments but in fact can improve compliance and retention of the subjects by clearly demonstrating that all are being treated alike. Merely stating in your protocol that blinding would be done is not enough. Give full details how the blinding is to be implemented including how the two groups would be assessed and handled similarly for medical maneuvers. The difference between the two should be only the active regimen, and nothing else that can possibly alter the outcome. Also state about masking of the regimen and of the procedure to make them look alike, and administered in an undifferentiated fashion. If one regimen is once-a-day (OD) and the other is twice-a-day (BD), the OD group can be given placebo second dose to give an identical look. If such details are not fully clarified, the readers

remain sceptical about bias reduction. They must be convinced that blinding was in effect until all opportunities for bias have passed.

Blinding is easily said than done. Some methods of blinding are in Box 2. There are situations where blinding is not possible. For assessing the outcomes such as quality of life, readmissions, and falls after hip surgery, blinding is just not possible if one maneuvre is keeping the patients in hospital for a specified number of days, and the other is early discharge and home rehabilitation. In most surgical interventions, control has to be another kind of surgery, and not a 'placebo'. A sham surgery may be unethical because it exposes a patient to surgical risks. In either case, it is extremely difficult to enforce blinding in a surgical trial. The patient can be kept blind after proper consent but the surgeon definitely knows. However, mechanism can possibly be developed that all assessments subsequent to the operations are done by another surgeon who does not know that the patient belongs to a test surgery or a control surgery.

Box 2: Methods of masking and blinding

Masking is the term used for apparent similarity of the regimens under trial and of the procedures followed during the trial. Top ingredient of masking is that the placebo or the control should have exactly same physical properties – packaging, labeling, handling, colour, size, shape, smell, and possibly taste – so that the patients or nurses are not able to distinguish, nor the physician who is assessing the outcome. This provides an insurance against prejudiced assessment of the patients by the investigators. The control subjects must pass through the same medical rigmarole in terms of physical and laboratory assessments, diet, change of wards or beds so that there is no scope: one, of deciphering the group to which the patient belongs; and two, of biased response due to differential procedures. Masking is making arrangements that the identity of the groups does not break till the trial is over.

Blindness is the term used for patients and the assessors. It is the concealment of allocation. To implement blinding faithfully, it is necessary to have a referee who keeps the code and assigns subjects to test or control group according to a predevised plan such as random allocation. See text for the meaning of single blinding, double blinding, and triple blinding. Many times the term blinding is used to include masking.

COMPLIANCE

Bias can still occur in subtle or unknown ways in an RCT despite random allocation and blinding. A major source of bias is loss to follow-up. If the follow-up requires recalling or revisiting the patients, some may not turn up or refuse to cooperate, some may be untraceable, and some could die from unrelated causes. Even if the outcome assessment is within the hospital stay, some can leave against medical advice (LAMA). Another factor that could affect a clinical trial is the need to change the treatment modality mid-way if a patient develops a serious illness. Then there could be patients who did not follow the full regimen. This is called the partial compliance. Take preemptive steps to minimise such losses and plan to adjust the results if needed.

Another important aspect of compliance is participants intentionally flushing the drug (or placebo) in the toilet. If this is done for actual or fear of side-effects, it certainly adds to the bias. The patient will hardly ever confess of doing so. Thus this bias may never surface. Another instance of bias can be when a patient takes two doses at one time because he missed the previous one. A side effect may occur that otherwise would not occur if the prescription is adhered to. Though difficult, in this case the patient may admit doing so. Perhaps the only alternative in this case is to exclude this patient and count him under dropouts.

Suppose two out of 500 randomised to receive placebo died of liver failure. You subsequently discover that these patients actually received the test drug due to an administrative error. This lapse can throw the trial off-guard and all precautions must be taken that this does not happen. If such a lapse is found, the analysis will have to be geared to the new realities. Another 'lack of compliance' occurs when patients are switched from one group to the other as per their wishes. All these situations are difficult to handle at the time of analysis although analysis such as 'intention-to-treat' can be done to address some of these biases.

Realise that trials are generally done in ideal conditions that do not exist in practice. Thus the actual performance may differ. **Efficacy** of a treatment is what is achieved in a trial that simulates ideal conditions, and **effectiveness** is what is achieved in practical conditions when the treatment is actually prescribed. For clarity, the latter is sometimes called use-effectiveness. Effectiveness could be lower than efficacy because of lack of compliance of the regimen, inadequate care, nonavailability of drugs, etc. Experience suggests that nearly three-fourths of the patients do not exactly adhere to or persist with prescriptions. Thus patients and maneuvers adopted during a trial do not lend the results to be generalised for patients at large. Generally, such external validity of the trials is not high. But trials do establish the potency of a regimen to effect a change.

Effectiveness under practical conditions has brought **pragmatic trials** into focus. The patients recruited for this kind of trial are not homogeneous as in a regular clinical trial but reflect variations that occur in real clinical practice. Strategies such as randomisation and control are also not used. Because of a large number of intervening factors in this setup, the interpretation could be difficult. Statistically, the standard deviation could be relatively large. For details of pragmatic trials, see Roland and Torgerson (1998).

SIZE OF THE TRIAL

The number of subjects should be reasonably large in each group so that full clinical spectrum is represented and a trend, if present, can clearly emerge. This also ensures reliability of the results. It should have adequate power to detect a minimum medically relevant difference.

Completely new treatment strategies that have large benefit are rarely discovered. Many trials are on variation of the existing modalities in the hope that some improvement in specific type of cases can be achieved. Additional benefit from such minorly different regimen is also likely to be small because the comparison is with the existing regimen and not placebo. Statistical power considerations tell that the size of the trial must be large to be able to detect a small difference. Trials involving thousands of patients are increasingly becoming norm. For comparing rt-PA (alteplase) and streptokinase for cardiovascular disorders, a trial on more than 40,000 patients was planned (GUSTO 1993)

However the size of the trial can not continue to increase indefinitely. Large trials are expensive, difficult to manage and supervise, and run the risk of lacking uniformity. Sustained focus and keenness is also casualty. Insufficient number of patients in one centre may encourage you to conduct multicentric trial. This is even more difficult to manage as centres may like to retain their freedom to adopt modification as per their wisdom consistent with local conditions. On the other hand, time and resource constraints for an endeavour such as postgraduate thesis necessarily limit the size.

Adapted from:

Basic Fundamentals of Medical Research: For emerging researchers, 2012 by A. Indrayan

Details at http://www.medicalbiostatistics.com/MedicalResearchBook.pdf

---

REFERENCES

Gianotti L, Braga M, Nespoli L, Radaelli G, Beneduce A, Di Carlo V. A randomized controlled trial of preoperative oral supplementation with a specialized diet in patients with gastrointestinal cancer. Gastroenterology 2002;122:1763-1770.

GUSTO Investigators. An international randomized trial comparing four thrombolytic strategies for acute myocardial infarction. N Engl J Med 1993;329: 673–682.

Lilford RJ. Equipoise is not synonymous with uncertainty (Letters). BMJ 2001;322:574.

Roland M, Torgerson DJ. Understanding controlled trials: what are pragmatic trials? BMJ 1998;316:285.