Methodological and Statistical Topics in Randomized Controlled Clinical Trials

Claudia Caminha Escosteguy

Rio de Janeiro, RJ - Brazil

Among research methods, randomized controlled clinical trials have constituted one of the main scientific advances during the 20th century. The randomized controlled clinical trial is a type of experimental study used as a reference standard for research methods in epidemiology and is considered the best source of available scientific evidence and the best source for determining the efficacy of an intervention.

The randomized controlled clinical trial is a prospective study that compares the effect and the value of an intervention (prophylactic or therapeutical) with controls in human beings. In this type of study, the investigator distributes, by chance, the factor of intervention to be analyzed through the technique of randomization; therefore, the experimental and control groups are formed through a chance distribution process to reduce or eliminate interference by variables other than those being studied. The intervention being studied can be drugs, techniques or procedures [1,2]. The term "efficacy" refers to the result of an intervention under ideal, controlled conditions, such as in the controlled clinical trial. The term "effectiveness" refers to the result of an intervention carried out in an average clinical environment, which includes the imperfections of implementation that characterize the daily world [2].

According to Feinstein [3], the idea of the distribution of a treatment through randomization was proposed by Fisher in 1923 and applied to agricultural research. In 1926, this idea was used for clinical studies for the first time by Amberson and coworkers, who tested the value of a gold compound in the treatment of tuberculosis. This was also the first blinded study reported, which means that the patients were not informed of the treatment administered. The controls received an injection of distilled water; the term "placebo", however, was used for the first time in 1938 in the study of Diehl on influenza virus vaccine [3].

In a generic way, the term "clinical trial" can be applied to any form of planned experiment involving patients and designed to elucidate the treatment most appropriate for future patients with a given medical condition. Some authors also use the term "non-controlled clinical trial" to describe a study in which all participants receive intervention. In reality, this would be only a descriptive study of the effects of an intervention in a group. The majority of the authors do not consider this type of study a clinical trial, calling it a non-controlled experiment. The more puristic authors reserve the term "clinical trial" only for the randomized controlled trials and do not accept its use for controlled trials that are not randomized [2].

The clinical trials with drugs are frequently classified into four main phases of experimentation [1]:

Phase I – These are trials of clinical pharmacology and toxicity in man, primarily related to safety and not to efficacy, and usually are carried out in healthy volunteers. The main objective is to determine an acceptable dose of a drug, the one that can be administered without causing severe side effects. This information is frequently obtained from experiments with fractional doses, in which a volunteer receives increasing doses of the drug, according to a predetermined schedule. Phase I also comprises studies on drug metabolism and bioavailability. After the studies on healthy volunteers, the initial trials with patients will also constitute a part of phase I. Typically, phase I studies may require 20 to 80 individuals and patients.

Phase II - These are initial trials of clinical investigation of the effects of treatment, still comprising a small investigation of drug effectiveness and safety, with careful monitoring of each patient. Sometimes, phase II trials can be carried out as a process of drugs, used to differentiate those with a true potential of effect among the several inactive or excessively toxic ones, so that the selected drugs can pass to phase III. Rarely, phase II requires more than 100-200 patients per drug.

Phase III – Large-scale assessment of the treatment. After the drug proves to be reasonably effective, it is necessary to compare it on a large scale with the standard treatment(s) available for the same medical condition, in a controlled clinical trial comprising a large enough number of patients. For some authors, the term "clinical trial" would be a synonym for these phase III trials, which constitute the most rigid form of clinical investigation of a new treatment.

Phase IV – Postmarketing surveillance phase. After

Hospital dos Servidores do Rio de Janeiro
Mailing address: Claudia Caminha Escosteguy - Av. Alexandre Ferreira, 361 - 22470-220 - Rio de Janeiro, RJ

the commercial approval of a research program, there are still some questions to be considered in regard to large-scale and long-term monitoring of side effects and additional studies on morbidity and mortality. Sometimes, the term "phase IV trial" has been used to describe exercises of promotion of a new drug directed at the medical public, which is not to be confused with the research of the clinical trial itself.

One should remember that, preceding the clinical trials, there should be a prior equally important program of pre-clinical research, including the synthesis of new drugs and studies in animals in regard to metabolism, efficacy, and, moreover, potential toxicity. In reality, this pre-clinical phase accounts for the majority of the estimated cost with the research on drugs. Currently, the majority of the clinical trials are related to the assessment of new drugs and is mainly funded by the pharmaceutical industry. It is estimated that, in the universe of new drugs synthesized in laboratories, only one out of 10 000 reaches the phase of clinical trials and only 20% of those are eventually marketed. A complete research program related to a drug lasts from 7 to 10 years, of which, almost half of the time is used in clinical trials, involving millions of dollars[1], emphasizing the role played by the pharmaceutical industry.

Randomized controlled clinical trials have the following main characteristics [1-5]: a) **they are experimental studies** and, therefore, involve important questions of ethics; b) **prospective architecture**: they have the architecture of a cohort study, meaning that they are prospective, with the particularity that the investigator uses a technique of random allocation (randomization) to form groups with similar characteristics, so that the individuals of a group receive a certain type of treatment while those of the other group remain as controls; c) **control**: it is necessary to compare the experience of a group of patients undergoing the new treatment with a group of similar patients who receive the conventional treatment. If there is no conventional treatment of real value, it can be appropriated to use a control group of non-treated patients. The most adequate technique for distributing the individuals in treated and control groups is randomization, which allows the allocation by chance; d) **randomization**: is a decision process that allows the study and control groups to be allocated by chance, being the best technique to avoid selection bias. In addition, it reduces the possibility of confusion bias. The beauty of randomization lies in the fact that it allows the distribution of known and unknown outcome determinants in a similar manner between the study and control groups, if the sample size is large enough[6].

There are several techniques of randomization [1,2]: **simple randomization** – it is the most frequently employed technique; the patients are directly assigned to study and control groups, with no intermediate stages. For example, by using a table of randomly selected numbers, where the odd numbers are assigned to the treatment group and the even ones to the control group; **block randomization** – it is characterized by the formation of equal size blocks with a fixed number of individuals, inside which the treatment in question is distributed, block by block, until the process of individual allocation in the research is finished. There is the advantage of providing equal number of participants in the study and control groups, even if the trial is interrupted before the expected end. It is also useful in studies with a reduced number of patients because the simple randomization performed with the aid of a table of randomly selected numbers only assures homogeneity between the groups when there is a large number of participants to be randomized; **paired randomization** – initially, pairs of participants are formed and allocation by chance is performed inside the pair, so that one individual receives the study treatment and the other the control treatment; **stratified randomization** – initially, strata are formed and the random allocation is performed inside each stratum; **randomization by minimization** – initially, the simple randomization is used but after the allocation of some individuals, the characteristics of the groups are analyzed and the calculation is reperformed as some new participants are recruited. These new participants will be allocated to one of the groups to reduce the detected differences or to maintain the already achieved balance. It is a new technique and computer technology allows several variables to be followed at the same time, so that a minimum of differences will be obtained between the groups.

In addition to the main characteristics already described, the following methodological questions should be considered when a randomized controlled clinical trial is performed:

Sample size – The trial should recruit a number of patients large enough to obtain a reasonably precise estimation of response to each treatment involved. Even though there are practical and ethical considerations in regard to sample size, a standard statistical approach refers to the estimations of the power of the study. There are five important questions in regard to sample size [1,2]: 1) *what is the main objective of the trial*? – for example, to verify if acetylsalicylic acid has any value in the prevention of post-infarction death, which is different from verifying if it prevents infarction or if it prevents death and re-infarction; 2) *what is the main outcome measurement*? – for example, death due to any cause within the first month of treatment, which is different from death due to a cardiovascular cause; 3) *how the data will be analyzed so that a difference in treatment can be detected*? – the most simple form is the comparison of percentages, for example, the percentage of deaths in the treated and placebo groups; a chi-square test will be used and a significance level of 5% will be accepted as showing evidence of a difference in treatment; 4) *what kind of outcomes can be anticipated with the standard treatment*? – for example, a 10% mortality is estimated in patients of the control group in the first month of treatment; 5) *what is the smallest difference in treatment considered important to be detected and with which degree of precision*? – it is important to stress that moderate reductions (for example, 20-25%) of the event of interest can require the randomization of thousands of patients [7].

To calculate of the sample size, one should consider the **alpha-level of significance** desired to detect a difference in the treatment and the **power of the study**, i. e., the degree of certainty that the difference between the treatments will be detected, if it really exists. An alpha-error or type I error is the probability of detecting a difference that in reality does not exist, i. e., the probability of a false-positive outcome; alpha is usually stipulated as 0.05. A beta-error or type II error is the probability of not detecting a difference when it really exists. The power of the study is 1-beta, and it is usually stipulated as 0.90.

When the necessary size of the sample is too large, the trial can be carried out in several centers, constituting the multicenter trial, which evidently requires special measures of organization and monitoring.

Trial organization and planning – It is of fundamental importance to precisely define: 1) which patients are eligible for the study, through well-defined inclusion and exclusion criteria; 2) which treatment is being assessed; 3) which outcomes or endpoints are of interest to be analyzed; 4) how the response of each patient will be verified.

Monitoring of the trial process – It is necessary to monitor protocol adherence, adverse effects, data processing, and the temporary analyses of the comparison among treatments. The possible protocol violations and deviations should be carefully analyzed, such as non-adherence to the treatment, participant dropout, incomplete assessment, and crossing between the study and control groups after randomization. This last deviation, for example, occurred in the nitrate versus control arm of the study GISSI-3, in which 57% of the patients randomized for control received nitrate, reducing the power of the study to detect a possible difference between the two groups [8].

Types of analysis – Analysis can be performed using two main formats [1,2]: 1) between those who have really completed the treatment in each group; 2) according to the intention to treat, in which all who were randomized to form groups are included, independently from having completed or not completed the treatment. This latter has been preferred, because it assures the maintenance of the random groups and assesses the treatment in the real world, with its imperfections. However, it is necessary to know what happened to those who have not completed the treatment, as well as if there was crossing between the groups. The dimension of these facts should also be known because if it is very large, this can represent bias.

Subgroup analysis – The fundamental result of a clinical trial is the description of the main outcome of interest in each of the main groups undergoing treatment. Although it may seem tempting to analyze the results in specific subgroups of patients, there are great risks inherent to this analysis. The first one is the inadequate number of patients, if the referred analysis was not part of the initial sample. The second is the risk of bias, since the subgroups selected according to characteristics considered after allocation to treatment may not be comparable, even though they were selected from the groups initially randomized. Third, when a great number of subgroups is examined, there is an increased chance that some of them will show a spurious statistically significant difference. A classical example of this possibility of spurious association was the analysis of the effect of zodiac signs in the ISIS-2 study, suggesting that acetylsalicylic acid was beneficial to all signs except Libra and Gemini to which there was an apparent damage [9,10].

Potential bias – The potential sources of bias are the following: the process of selecting the groups, the allocation to treatment, the achievement of the intervention in the proposed form, and the assessment of the results. Randomization controls the first two steps.

A disruption in the follow-up and non-adherence of the participants can introduce bias, mainly if they are differently distributed between the study and control groups, and should always be mentioned.

Bias of assessment (also called of information, of observation or of measurement) results from systematic differences in the way data about the event of interest are obtained from the several groups being studied. They are minimized when the double-blind technique is used with placebos; however, it is not always possible, even using this technique, to hide from observers and from the ones observed, the groups to which these latter belong.

Another interesting bias is related to trial publishing and not to its development; this is the publishing bias, which is the tendency to publish the studies with positive results.
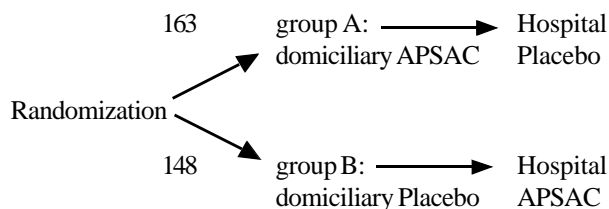
Factorial design – In this design, the effects of several factors are verified in one single trial. For example, in the study of drugs A and B, a factorial design will evaluate four groups of treatment: one using drug A, another using drug B, another using drugs A and B, and a control group with none of the drugs. One example is ISIS-2 [9], where the effects of acetylsalicylic acid, streptokinase, both, and none of them in patients suspected of having acute myocardial infarction (AMI) were assessed.

Cross-over type trial – Usually the trials make comparisons between patients, and each patient receives only one type of treatment. Sometimes it may be advisable to make sequential comparisons in the same patient, i. e., each patient of the study will receive more than one treatment. A major problem with the conventional parallel groups is the fact that patients vary a lot in regard to their initial stage of disease and their response to treatment. A great number of patients in each group is often necessary to estimate in a reliable manner the magnitude of any difference in effect [1,2].

One should not mistake the cross-over design for the "before and after" studies, in which all patients receive the

same treatment and their conditions are assessed before treatment onset and in many stages after it, and are, in effect, non-controlled studies [1,2].

An example of the cross-over type trial is the study of the GREAT Group on safety and efficacy of the domiciliary thrombolysis [11]:

```
       163      group A: ────────▶  Hospital
                domiciliary APSAC    Placebo
                 ▲
Randomization ───┤
                 ▼
       148      group B: ────────▶  Hospital
                domiciliary Placebo  APSAC
```

Blind assessment – Also called blinding: The justification for this technique lies in the potential of bias occurring when all individuals involved in the trial know which treatment the patient is receiving. In regard to the blind condition, there are three participants to be considered: the patient, the group of professionals applying the treatment, and the evaluator [1].

The Hawthorn effect refers to the tendency of individuals to change their behavior because they are targets of special interest and attention, despite the specific nature of the intervention they are receiving. A way of controlling this effect is through blinding and placebo use [4].

The patients' knowledge about receiving a new treatment can have a beneficial psychological effect on them and, in contrast, their knowledge of receiving a conventional treatment or no treatment at all can have an unfavorable effect. It is obvious that the impact depends on the type of disease and nature of treatment, but this possibility should not be underestimated even in non-psychiatric disorders.

In regard to the group of people applying the treatment, decisions related to changes in doses, particularities of the patient's examination, continuation of the trial treatment, and need for additional treatments are usually the responsibility of the assistant doctor, who can influence the course of treatment in several ways. These decisions can be influenced depending on the knowledge of to which trial group the patient belongs. The excitement about a new treatment can also be transferred to the patient and cause a change in his or her attitude, increasing the patient's adherence to treatment, for example.

In regard to the investigators who evaluate the results, if they are aware of each patient's treatment, there is a potential risk, for example, of registering more favorable responses for the treatment they consider superior. Not knowing the trial groups helps to avoid gauging bias, which is also minimized when the final event of interest is defined in the most objective form possible. A gauging bias might occur when the evaluation of response to treatment requires clinical judgement. Even in apparently well-defined events, such as AMI, clinical judgement is many times needed in borderline cases. In such cases, if the treatment state were known, there could be a tendency from the evaluator to direct the final diagnosis in favor of AMI or against it.

The term "double-blind" refers to those trials in which neither the patients nor the people responsible for their assistance and evaluation know the treatment being received. In reality, in these cases, the three types of participants are blind in regard to the treatment condition; however, as the same clinicians who work with therapeutics are often the ones who evaluate the patient, the term "double-blind" is adequate (it is not common to refer to a trial as triple-blind; usually the term double-blind is used).

Placebo use – A placebo is a substance of appearance, form and administration similar to that of the treatment being evaluated, but without the active principle. The main reason for introducing controls with placebo is to make the attitudes of the patients in the study and control groups of the trial uniform. The placebo effect is a response to a medical intervention that, despite being a definite result of it, has no relation to the specific mechanism of action [4]. A basic principle to be considered is that patients can not be ethically assigned to receive placebo if there is an alternative standard treatment of established efficacy.

Ethical questions – Maybe the great catastrophe of congenital anomalies induced by thalidomide in the '60s has been a landmark for discussion and implementation of medical and public polices that take into consideration the ethical aspects related to the introduction of new treatments. Since 1926, in the USA, it has been required by law that an efficacy test be performed before new drugs are approved for marketing [2].

The basic international document for ethical discussion of clinical trials is the Helsinki Declaration from 1964, revised in Tokyo in 1975 [2]. Among the national relevant documents are the Medical Ethics Code [12] and the Research Rules Involving Human Beings from the National Board of Health [13]. Even when the investigation is thoroughly justified, some questions deserve consideration: one of the main points is the deprivation of the control group from a new treatment to which there is clear evidence of superiority in relation to the conventional treatment. The non-administration of an effective treatment to patients is only ethically acceptable if there are doubts in regard to treatment efficacy; the small sample size, informed by calculations, sufficient to answer the question being investigated should be used. The study should be immediately interrupted if, during its conduction, there is definitive evidence of benefit or absence of benefit of the treatment in question. The informed consent of the patient should always be present.

All questions that have been discussed so far are related to the internal validity of the trial. The dissemination of the randomized trials and their use as a standard to demonstrate therapeutic efficacy of drugs have caused good-quality scientific evidence to be available before the introduction of new therapeutic agents into clinical

practice. Another fundamental aspect to be discussed, however, is the possibility of generalization of the trial results. The external validity of a study implies the possibility of generalizing the results of the studied sample to other samples, beyond the target population studied. It also involves patient and ethnocultural variations, severity factors, in addition to considerations on the cost: benefit ratio, risk, infrastructure, and so forth. These considerations are justifiable only after the establishment of the internal validity of the study.

References

1. Pocock SJ. Clinical Trials. A Practical Approach. Brisbane: John Wiley & Sons, 1989.
2. Pereira MG. Epidemiologia, Teoria e Prática. Rio de Janeiro: Guanabara Koogan, 1995.
3. Feinstein AR. Clinical Epidemiology. The Architecture of Clinical Research. Philadelphia: WB Saunders, 1985.
4. Fletcher RH, Fletcher SW, Wagner EH. Epidemiologia Clínica: Bases Científicas da Conduta Médica. Porto Alegre: Artes Médicas, 1989.
5. Sackett DL, Haynes RB, Guyatt GH, Tugwell. Clinical Epidemiology: A Basic Science for Clinical Medicine. 2nd ed. Boston: Little Brown, 1991.
6. The Evidence Based Medicine Working Group. How to use an article about therapy or prevention, 1998.
7. Yusuf S, Wittes J, Friedman L. Overview of results of randomized clinical trials in heart disease. JAMA 1988; 260: 2088-93.
8. GISSI-3 – Effects of lisinopril and transdermal glyceril trinitrate singly and together on 6-week mortality and ventricular function after acute myocardial infarction. Lancet 1994; 343: 1115-221.
9. ISIS-2 Collaborative Group – Randomized trial of intravenous streptokinase, oral aspirin, both or neither among 17187 cases of suspected acute myocardial infarction. Lancet 1988; I: 349-60.
10. Yusuf S, Wittes J, Probstfield J, Tyroler H. Analysis and interpretation of treatment effects in subgroups of patients in randomized clinical trials. JAMA 1991; 266: 93-8.
11. GREAT Group – Feasibility, safety, and efficacy of domiciliary thrombolysis by general practitioners: the Grampian Region Early Anistreplase Trial. Br Med J 1992; 305: 548-53.
12. Conselho Regional de Medicina do Estado do Rio de Janeiro – Código de Ética Médica. Legislação dos Conselhos de Medicina, 1988.
13. Conselho Nacional de Saúde – Normas de Pesquisa envolvendo Seres Humanos. Res. CNS 196/96. Brasília: Ministério da Saúde, 1996.