

Understanding Randomized Controlled Trials

Entendendo Ensaios Clínicos Randomizados

MARCO AURÉLIO PINHO DE OLIVEIRA¹; RAPHAEL CÂMARA MEDEIROS PARENTE²

¹ *Doctorate in Epidemiology from the Institute of Social Medicine of the University of the State of Rio de Janeiro (UERJ). Chief, Gynecology Division, UERJ;* ² *Doctorate in Gynecology (Human Reproduction) from the Federal University of the State of São Paulo (UNIFESP). Master's in Epidemiology from the Institute of Social Medicine of the University of the State of Rio de Janeiro (UERJ).*

Bras. J. Video-Sur, 2010, v. 3, n. 4: 171-175

Accepted after revision: July, 2010.

Clinical trials constitute a powerful tool for evaluating health interventions, be they medication interventions or not.¹ The first clinical trial, similar to the design we know today, was published in the late 1940s,² when the statistician Sir Austin Bradford Hill randomly assigned patients with pulmonary tuberculosis into two groups: those who would receive streptomycin and those who would not receive the medication. In this way Hill was able to evaluate, in an unbiased fashion, the efficacy of streptomycin. Despite the increasing publication of controlled clinical trials, some aspects of the design and the analysis are still misunderstood and misinterpreted.

Clinical trials are studies in which one group receives a treatment or exposure and is accompanied and compared with a control group. Unlike observational studies in which the researcher does not interfere in the exposure, in clinical trials the researcher plans and actively intervenes in the factors that affect the sample, thus minimizing the influence of confounding factors. The allocation or assignment of research subjects may be random or nonrandom.

Although randomized controlled trials (RCTs) are the gold standard for determining the effect of a therapy, for various reasons its use in surgery is not as widespread as in other medical specialties.³ Clinical trials are laborious and costly to carry out, and the development and execution of clinical trials of surgical interventions poses additional methodologic and practical challenges.

Randomized comparisons of surgical interventions have been carried out many years. Various surgical interventions used for long periods of time were found to be ineffective and were abandoned

after they were subjected to this kind of study.⁴ The ligation of the internal mammary artery was abandoned after two small clinical trials found no difference in relation to placebo. More surprising, the number of clinical trials in surgery has not accompanied that of other specialties.

A review found a marked increase in the number of randomized clinical trials from 1990 to 2000.⁵ And yet only 3.4% of all articles published by important surgical journals were clinical trials.⁶ The contrast with other clinical areas can be seen by comparisons of audits of databases for clinical practice in internal medicine and various surgical specialties. Whereas half of interventions in internal medicine are based on evidence from clinical trials, two surgical audits reported that fewer than one quarter of surgical interventions have comparable evidence.⁶⁻⁷

The main problem associated with clinical trials in surgery – and which impedes their design and execution – is the difficulty in blinding. Masking is the process of retention of information about the interventions received by each group is a key element in the design of RCTs. When done successfully, it is viewed as having an important role in preventing the introduction of an information bias. Blinding can be applied to the study subjects, to the researchers, and even to the statisticians calculating the results of the study. In clinical trials of surgical interventions, blinding poses a much greater challenge than in trials of pharmaceutical interventions. It is impossible to bar a surgeon from knowing what technique he will be using. Thus, it is impossible to blind the surgeon. But blinding generally does not occur in surgical interventions even for the study participants. This happens because of the routine practice of the patient returning to the

surgeon who operated post-operatively for follow-up. The blinding of patients is feasible when the same surgical access is used for different surgical procedures. Blinding patients is more difficult when the surgical approaches are different, for example, laparotomy compared with laparoscopy, and it is virtually impossible when one compares surgery with medical management. Although surgery as a placebo is theoretically possible – and there are examples in the literature⁸ – the conception of such studies is often impeded by ethical considerations.

Carefully conducted randomized clinical trials are the gold standard among studies to guide our day-to-day clinical practice. In developing a RCT one seeks a study that is large enough that an important clinical effect can achieve statistical significance, but not so big that minor outcomes are significant.

Moreover, when excessively large, RCTs can be prohibitively expensive.⁹ The sample size for any study is calculated based on the number of type I and II errors that the researcher is willing to tolerate. A strength of RCTs is that random assignment minimizes the impact of confounding factors by distributing them equally among the groups. However, there are major drawbacks of a single large RCT.

First, some RCTs are so methodologically rigorous that findings are only applicable to the population studied (problems of external validity or generalizability). In addition, patients participating in clinical trials may have a baseline risk for the outcome of interest that is much greater than the average of the patients of the population or the subjects may be from very different populations. For example, some studies that have investigated calcium intake for the prevention of pre-eclampsia have been carried out in countries with very different diets, which makes it difficult to extend conclusions to populations with other diets. This can limit the generalizability of findings. Another issue is that some clinical trials may not have a sufficient number of individuals in important subgroups.

A final concern with randomized controlled trials is the issue of publication bias. Publication bias occurs when the findings of studies that report positive results (i.e, statistically significant) are more likely to be published than studies with negative results. The studies that report significant results may result in a greater number of publications and will also be published in journals of greater impact. In this way publication biases may affect meta-analysis since these

are unable to consider trials with unfavorable findings that were never published.

A randomized clinical trial is a prospective study of human subjects comparing the effect and value of an intervention versus a control.¹⁰

1. Randomized controlled trials should be considered when:

- a. There is uncertainty about the effect of an exposure or treatment;
- b. The exposure can be changed in the study.

2. Potential limitations of randomized clinical trials include:

- a. Limited generalizability of the study population.
- b. Limited generalizability of the environment under study.
- c. Randomized clinical trials address a specific study question.

3. The usual measures of the magnitude of effect in randomized clinical trials are relative risk and risk difference.

In RCTs, a treatment or procedure may be compared to no treatment, treatment with a similar medication (e.g., same class of drugs in head-to-head trials), with a placebo, or a preexisting standard of care.¹⁰ RCTs are conducted under controlled conditions to ensure that the study hypothesis is tested in a reproducible manner. RCT protocols often cannot be repeated in clinical practice. The highly specialized conditions in which in a randomized trial is conducted may result in conclusions that are internally valid, but which cannot be generalized to patients with the same disease outside the study.¹⁰

4. Randomized clinical trials are designed to definitively answer a specific question, focusing strictly on a research hypothesis, isolating the effect of one or a small number of therapies. Randomized clinical trials are not designed to evaluate the mechanisms by which a therapy can produce benefits or harm. They usually have narrow clinical applicability. The conclusions derived from RCTs are limited to the situations in which the exposure of interest can be modified in a study environment, for example, using medications or lifestyle changes. There are innumerable exposures that are extremely worthy of investigation, but which cannot be easily modified, such

as genes, serum markers, socioeconomic conditions, etc.

Randomized controlled trials are limited to specific clinical situations in which the exposure of interest can be easily modified. Nor can they be used in situations where exposure is known to be deleterious, because it is not ethical. For example, we cannot get someone to smoke in order to be compared with a group who do not smoke because we already know hazards of smoking.

The investigation of new drugs by the pharmaceutical industry involves RCTs in different phases. They are carried out in basically four phases¹⁰:

Phase I Studies

The first step in developing a new drug is to understand whether the medication is well tolerated in a small number of people. Although not controlled clinical trials, these types of studies are referred to as Phase I studies. Participants in Phase I trials are either healthy adults or people with the specific disease that the drug is intended to treat. Occasionally, phase I studies cannot be performed in healthy adults because the drug has unacceptable adverse effects; this is the case with chemotherapeutic agents. Phase I trials seek to determine what maximum dose a drug can be administered before unacceptable toxicity occurs. These studies are begin with low doses in a limited number of people, with the dose increased gradually.

Phase II Studies

Phase II studies are designed to evaluate whether a drug has biological activity and to determine its safety and tolerability.

Phase III / IV Studies

Phase III studies are randomized clinical trials designed to assess the efficacy and safety of an intervention. The results of phase III studies are clinical endpoints such as death or tumor-free survival. Assessments of safety occur over a longer period than with phase II studies. Phase IV studies occur after regulatory approval and measure outcomes associated with a drug or intervention in routine clinical use in the general population.

In the critical evaluation of a randomized clinical trial, we should ask several important questions, using some information already outlined above.

The first question one should ask is “Are the patients selected for this study similar to patients that I treat?”

The study should state clearly the inclusion and exclusion criteria (e.g. age, gender, prior illnesses, etc.) The conclusions of studies with very restrictive inclusion criteria are limited in terms of their generalizability (but are more specific).

The study report should also mention how patients were selected from the general population. Several examples that might introduce biases: Were study enrollees all patients who presented to a tertiary hospital? Were they paid volunteers, and if so, were they patients drawn from a poor community near the hospital?

Remember that we can only generalize to populations similar to the one in the study. It is up to the reader to decide whether differences between the population selected in the study and the population that interests them are important or not.

Another key aspect is the clinical endpoint of interest. Many studies that are well structured methodologically, but evaluate variables that are not very useful clinically, or do not represent what they should. For example, in the treatment of endometriosis with an antiestrogen drug, the researcher finds only a statistically significant decrease in the American Fertility Society (AFS) score – which quantifies the extent of disease – and concludes that treatment is effective. But what endometriosis causes is basically infertility and pelvic pain, and the study in question did not assess what really matters (the clinical endpoint of interest), that is, if the patient’s pain improved or that she was able to get pregnant (the patient is not interested only in improving their AFS score). Another example would be a study to evaluate the effectiveness of a new chemotherapeutic agent, using as parameter only the reduction of tumor mass. The decrease of the tumor, however, may not correlate with survival, which is the outcome variable of interest for the patient.

Another important issue is how individuals were allocated between the treatment group and control (or placebo) group. Studies that do not include a control or placebo group run the risk of showing that a treatment is effective when in fact it is not. A “placebo effect” can even produce improvement in symptoms in more than 50% of subjects. The placebo may even be more effective than the medication being studied! In other words, a patient’s improvement might

occur not because of the treatment in question, but for other reasons: the natural history of the disease process, psychological aspects, etc. Sometimes, as for example, in the treatment of acute pelvic inflammatory disease, it is inappropriate to use of a placebo or have no treatment as the control. In these cases, the investigator can study a new drug using as a control group the treatment considered the standard of care.

The next step is to verify how the randomization was done between the treatment and control groups. Randomization implies the chance assignment of patients so that each case has the same probability of being allocated to the intervention or control group. Non-randomized controlled studies are viewed as not have the same weight of scientific evidence as RCTs, because in most cases the intervention and control groups are not equivalent (i.e., there is some bias). For example, in a study comparing Laparoscopic Burch Urethropexy and laparotomy, one could imagine that patients who were thinner and who had milder complaints of stress urinary incontinence might be selected for laparoscopy. If this selection bias were to occur, one could not compare these two distinct groups of patients, as it would put laparotomy at a disadvantage.

Randomization ends up balancing known factors such as age, weight, stage of illness, as well as unknown (or immeasurable) factors (e.g., some genetic factor that cannot readily be detected that might influence the clinical outcome). True randomization should be done with random numbers generated by computer (there are also tables for this purpose) that are placed in numbered sealed envelopes that are opened immediately before treatment is administered. Inappropriate methods of randomization – sometimes called “quasi-randomized” – include, for example, doing laparoscopic surgery on Mondays and Wednesdays and laparotomies on Tuesday and Thursday, or alternating laparoscopic surgery and laparotomy. These methods permit failures in the random assignment, even if unconsciously. For example, the surgeon (or the physician who referred the patient for surgery and is familiar with the design of the study) may consider a patient too obese for laparoscopy and subconsciously (or intentionally) postpone the surgery for a day, in order to coincide with day or turn for a laparotomy. Appropriate randomization will generally prevent this type of bias.

Nevertheless, well executed randomization does not guarantee the equivalence of the groups, especially with small samples. Before analyzing the data, it is good idea to check if there really was a balanced assignment of patients, particularly regarding those characteristics which directly influence the outcome, such as disease stage, age, weight, etc.). Unbalanced groups may bias the results.

One should also be concerned if the method used in measuring the clinical endpoint of interest was double-blinded. This means that neither the patient nor the professional making the assessment (for example, assessing the degree of postoperative pain in patients who may or may not have received the anesthetic marcaine in a surgical wound) may know whether the treatment was applied. The physician may unconsciously not fully appreciate the complaints of patients who belong to the group that received the investigational treatment, distorting the results. The patient may want to believe in a particular investigational or novel treatment and thus downplay or underreport symptoms. The double-blind method attempts to eliminate this potential type of subliminal bias. As with comparisons of clinical and surgical treatments, a double-blind approach is not always feasible. But you should not neglect to use a double blind approach whenever possible.

Another important aspect is the duration of follow-up. First make sure that the duration of follow-up was adequate to assess the clinical outcome of interest. For example, in assessing the use of laparoscopic transection of the utero-sacral ligaments for relief of dysmenorrhea, the duration of follow-up should be at least one year, as the high number of relapses after this period is well established. Another relevant issue is whether the percentage of patients who were lost to follow-up was similar among the groups. Follow-up rates below 70% make it virtually impossible to analyze the data, as the abandonment may have occurred because of side effects or even deaths stemming from the treatment.

The aforementioned errors are systematic, as opposed to random errors that can occur when working with any sample. The appropriate statistical analysis aims to identify whether the differences in results were obtained due to random errors or if a difference really does exist between the intervention and the control. The basic concepts of statistics will be addressed in the next issue of this journal.

ADDITIONAL READING

1. Medical Research Council. Streptomycin treatment of pulmonary tuberculosis. A Medical Research Council Investigation. *BMJ*.1948; 2:769-82.
2. Coutinho ESF, da Cunha GM. Conceitos básicos de epidemiologia e estatística para a leitura de ensaios clínicos controlados. *Rev Bras Psiquiatr*. 2005; 27(2):146-51.
3. Cook J. The challenges faced in the design, conduct and analysis of surgical RCTs. *Trials* 2009, 10:9.
4. Ko CY, Sack J, Chang JT, Fink A: Reporting randomised, controlled trials: where quality of reporting may be improved. *Dis Colon Rectum* 2002, 45:443-447.
5. Wente MN, Seiler CM, Uhl W, Büchler MW: Perspectives of evidence-based surgery. *Dig Surg* 2003, 20:263-269.
6. Ellis J, Mulligan I, Rowe J, Sackett DL: Inpatient general medicine is evidence based. *Lancet* 1995, 364:407-410.
7. Howes N, Chagla L, Thorpe M, McCulloch P: Surgical practice is evidence based. *Br J Surg* 1997, 84:1220-1223.
8. Kenny SE, Shankar KR, Rintula R, Lamont GL, Lloyd DA: Evidence-based surgery: interventions in a regional paediatric surgical unit. *Arch Dis Child* 1997, 76:50-53.
9. Scifres CM, Iams JD, Klebanoff M, Macones GA. Metaanalysis vs large clinical trials: which should guide our management? *Am J Obstet Gynecol* 2009; 200:484.e1-.e5.
10. Kestenbaum B. *Epidemiology and Biostatistics: An Introduction to Clinical Research*, DOI 10.1007/978-0-387-88433-2_7, © Springer Science Business Media, LLC 2009.

Correspondence Address:

MARCO AURÉLIO PINHO DE OLIVEIRA

Rua Coelho Neto, 55 / 201

Tel.: (21) 9987-5843

E-mail: maurelio@infolink.com.br