Non-inferiority clinical trials: concepts and issues

Estudos clínicos de não-inferioridade: fundamentos e controvérsias

Valdair Ferreira Pinto*

Abstract

Non-inferiority trials are experimental models designed to determine whether a new treatment or procedure is not less effective than an established one, which is considered as standard. They are especially important in the assessment of treatments in which the use of placebo is impracticable. They differ substantially from the classical superiority trials and require different approach, especially in the planning and the data analysis. This paper is a review of the key differences between non-inferiority and traditional clinical studies. There is a considerable amount of misunderstanding on the correct use of this experimental design, which certainly compromises the credibility of some clinical assessments.

Keywords: Clinical trial, Research Design, statistics & numerical data.

Resumo

Estudos de não-inferioridade são modelos experimentais desenvolvidos com o objetivo de determinar se um novo tratamento ou procedimento não é menos eficaz que outro já estabelecido e considerado como controle. São de grande importância no estudo de tratamentos em que o uso de placebo é inviável. Requerem metodologia diferente dos estudos clássicos, chamados de estudos de superioridade, especialmente no planejamento e análise estatística. O presente artigo é uma revisão dos principais elementos diferenciais entre estudos de não-inferioridade e estudos tradicionais. Existe ainda em nosso meio um considerável grau de desconhecimento sobre o uso correto dessa metodologia, o que certamente compromete a validade de alguns estudos clínicos.

Palavras-chave: Ensaio clínico, Projetos de Pesquisa, estatística & dados numéricos.

Introduction

The randomized comparative clinical trials are currently considered the best experimental design to assess issues related to treatment and prevention¹. Classically, they are defined as medical experiments projected to determine which of two or more interventions is the most effective facing the randomized allocation of patients to different study groups. In general, one of the groups is considered as the control one – which sometimes may refer to the absence of treatment, placebo or, more often, a treatment of recognized efficacy. Statistical resources are available to validate conclusions and to maximize the chances of identifying the best treatment. These models are called superiority trials, whose objective is to determine whether a treatment under investigation is superior to the comparative agent. Some non-randomized patient allocation alternatives were proposed to minimize the variance between groups and to increase the efficiency and sensitivity of the study; some of these methods have concrete statistical basis³; however, the block randomization procedures with or without stratification are still the most used methods. A common logical error made in the interpretation of traditional clinical trials is to admit the equivalence of treatments when it is not possible to show significant differences, that is, the non-significance in a traditional text does not support the conclusion of absence of difference (type II error). In conditions of low statistical power, there may be important differences between treatments, and the null hypothesis is not rejected due to an insufficient number of

Study carried out at Valdair Pinto Consultoria em Medicina Farmacêutica, São Paulo, SP.

^{*}Consultant of Pharmaceutical Medicine; Ex-professor of Biostatistics at Faculdade de Medicina da Universidade do Triângulo Mineiro, Uberaba, MG; Ex-international medical director at Pfizer Inc., Nova York, USA.

The present paper is part of a project of medical education in non-inferiority trials that was partially and unconditionally supported by Associação da Indústria Farmacêutica de Pesquisa (Interfarma).

Received on Apr 19, 2010. Accepted on Jun 23, 2010. J Vasc Bras. 2010;9(3):145-151.

patients, to the excessive variability or to some bad habits in designing or conducting the study. This inference error has been frequently emphasized by the expression "The absence of evidence is not an absent evidence"^{4,5}.

As it is fundamentally impossible to demonstrate that two treatments are equal, since the 1970s new methodological procedures have come up, allowing the development of evidence studies - destined to show the absence of significant differences between treatments and non-inferiority trials, which have been increasingly used and aim to show, based on certain criteria, that a new treatment is not less effective than another existing treatment. The equivalence studies have become essential to the regulatory approval of generic medications, while the non-inferiority trials are currently conducted in situations in which comparisons to placebo are unfeasible and active controls are necessary⁶. The essential difference between superiority, equivalence and noninferiority trials is the formulation of hypothesis to be tested. Table 1 depicts the algorithms of analysis for these three types of trials: T represents the measure of efficacy of the new treatment and C stands for the efficacy of the control treatment. Regarding the superiority trials, to reject the null hypothesis means that T is superior to C; for the non-inferiority trials, it means that the difference between C and T is smaller than a margin "M" and, for the equivalence trials, it shows that the difference between C and T is not smaller or bigger than a margin "M". Basically, the word "equivalent" means not inferior and not superior, and testing the equivalence refers to the analysis of the symmetrical region defined by [+M, -M], as shown in Figure 1.

The non-inferiority trials have great applicability in Oncology, preventive Cardiology and in the assessment of anti-infectious agents. The regulatory authorities have also been requiring non-inferiority trials for the assessment of biosimilars (products obtained by biotechnological processes such as therapeutic proteins,

 Table 1 - Hypothesis formulation for superiority, non-inferiority and equivalence trials

Study type	Null hypothesis	Alternative hypothesis
Superiority	Ho: C - T ≥ 0	Ha: C – T < 0
Non-inferiority	Ho: $C - T \ge M$	Ha: C – T < M
Equivalence	Ho: $ C - T \ge M$	Ha: C – T < M

T and C are measures of efficacy for the new treatment and the control, respectively; M is the non-inferiority/equivalence margin

hemoderived products, monoclonal antibodies, etc). A product that has proved to be non-inferior in relation to an established treatment regarding an efficacy variable may, however, present important advantages such as better tolerability, use convenience, galenic advantages, different metabolic pathways, less interactions, among others.

An appropriate definition for the non-inferiority trials must indicate that they are destined to establish whether a new treatment is not less effective than a standard treatment even though a tolerance margin be previously established and named as non-inferiority mar $gin (M)^7$. In these studies, the null hypothesis is that the treatment under investigation is inferior to the control due to a difference higher than or equal to M, and the alternative hypothesis is that the difference between treatments is smaller than the margin. The method of choice for analysis of non-inferiority trials consists in the construction of confidence intervals, usually 95% (95%CI). The treatment is considered non-inferior if the inferior limit of the 95%CI of the difference between treatment and control does not include the value of the specified margin.

In the planning, analysis and interpretation of noninferiority trials, at least five factors must be carefully taken into account to assure the validity of the study:

- 1. determination of the non-inferiority margin;
- 2. the number of patients needed to conduct the study;
- 3. control of study sensitivity;
- 4. definition of the analyzed population;
- 5. ethical justification.



Figure 1 – T: treatment; C: control. T is superior to C if the confidence interval of the difference is entirely at the right of zero, non-inferior if entirely at the right of -M and equivalent if contained in the equivalence zone between -M and +M

A brief discussion on each of these critical determinants was the objective of the present paper. However, the focus of this article were the non-inferiority trials and much of the discussion is also applicable to the equivalence trials.

Non-inferiority margin

As indicated above, M quantifies the maximal loss of clinically acceptable efficacy for the studied treatment to be considered as non-inferior to the control. Consequently, it may not exceed the minimum clinically relevant difference that would be used in a superiority study and, to assure the maintenance of some efficacy, it may not be equal to or higher than the integral effect of the control treatment. Its specificity is a hard but essential task, as well as one of the necessary elements to the calculation of the sample size. Excessively high M values increase the probability that inferior treatments be considered as non-inferior while lower and more conservative values demand bigger samples, thus making the studies more expensive due to the larger number of patients, besides obvious ethical implications. The M value must be established based on clinical and statistical considerations, and must be defined prior to the study; it may be specified in absolute or relative terms, as may the differences between means or proportions or odds ratio logarithms, or even relations between the hazard rates.

In a simple manner, M may be determined as a percentage of the control effect estimated for the current study, usually between 10 and 20%. Its definition must, however, take into consideration the therapeutic field and the magnitude of the control group effect; for example, for anti-infectious agents, more conservative margins are recommended (e.g. 10%) when the expected effect is around 90%, and more ample margins (e.g. 20%) when the anticipated effect is inferior to 80%. The existence of other possible benefits must also be considered; a larger margin is accepted if there are clinical advantages such as an important reduction in adverse effects. In situations in which the study evaluates concrete outcomes such as mortality or irreversible morbidity, the value of M would be really hard or even impossible to define.

Currently, there is a tendency to define the value of M predominantly by means of statistical considerations, thus putting the clinical judgment on a less decisive position. In this manner, by using historical data of studies that compared the control treatment to placebo, it is possible to derive indirect comparisons to assure that, even with the inexistence of a placebo group, the studied treatment is superior to it. It is known as comparison to putative placebo⁸. If the magnitude of the control effect is designed by C, and P is the effect of the placebo based on historical data, the non-inferiority margin "M" will always be a fraction of C-P, that is: M = x% de (C-P). In this manner, it is also possible to control how much of the effect of C should be preserved in T. Usually, 50 to 75% is accepted as the fraction of the estimated control effect to be preserved in relation to the placebo. In this case, M would be 25 or 50% of the liquid effect of C. A numeric example: suppose that the historical data have indicated an 80% efficacy of the control group and 60% of the placebo; the liquid effect of C will be 20%. To preserve 50%, "M" must be equal to 10% and, to preserve 75%, M must be equal to 5%. At this point, clinical considerations may help in the decision-making and define a final value.

The non-inferiority margin may also be determined by the so-called "50%-rule", endorsed by the Food and Drug Administration (FDA)⁹, which advocates that the value of M must be inferior (preferentially 50%) to the inferior limit of the 95%CI obtained from historical data that compare control treatment and placebo. In the numeric example above, supposing that the difference of 20% between proportions has been obtained from a sample of 200 patients per group, the 95%CI is 11.1-28.2%. Taking the half of the inferior limit (11.1%), the value suggested to M is 5.5%. This method is considered conservative, for it provides a double discount in the margin calculation, thus decreasing the power of the study in demonstrating non-inferiority.

Sample size

In the calculation of the number of patients necessary to a clinical trial, one must bear in mind the alpha probability (type I error or false positive) and the beta probability (type II error or false negative). In the context of the non-inferiority trials, somehow, these errors have a reversal interpretation. Alpha quantifies the risk of declaring false non-inferiority and patients may use an inferior medication in the future; beta quantifies the risk of false conclusion of inferiority, which means that further patients may not benefit from the new medication. Besides that, the size of the sample will also depend on the value stipulated for the non-inferiority margin and on the variability of data¹⁰.

Generically and simply, the equation for the calculation of the number of patients in a comparative clinical trial between parallel groups has the following view (Equation 1):

$$n = 2 \times z^2 \frac{s^2}{d^2}$$
 Equation 1

in which:

n is the number of patients in each group of treatment; z is a constant that depends on the choices for alpha and beta, and their values are obtained from probability tables that are cumulative of the normal distribution;

 s^2 is a measure of the data variability;

d is the minimal clinically relevant difference; in the classical case of alpha = 0.05 and beta = 0.2, the value of z^2 will be 7.9.

In this manner, the equation is reduced to Equation 2:

$$n = 16 \frac{s^2}{d^2}$$
 Equation 2

It corresponds to the elegant mnemonic suggestion of Lehr: "sixteen s-squared over d-squared"¹¹.

For non-inferiority trials, it is usual to stipulate alpha = 0.025 in unilateral tests, and beta = 0.1 (in this case, the constant z^2 will be equal to 10.5), as *d* is replaced by the non-inferiority margin "M" (Equation 3).

$$n = 21 \frac{s^2}{M^2}$$
 Equation 3

The methods to estimate the variance s^2 depend on the studied variable's nature (percentages, means or risk rates), a matter that is out of the scope of this paper.

Starting from the equation above, one may conclude that n increases rapidly when M decreases; for instance, in a situation in which the efficacy is estimated in 70%, the number of patients required in each group will be 83 for M = 20%, 147 for M = 15%, 330 for M = 10% and 1,319 for M = 5%. When M tends to zero, n tends to infinite, which indicates that it is impossible to prove that two treatments are identical. Once the statistical properties are maintained and M is always smaller than d (the clinically relevant difference of the non-inferiority trials), the number of patients needed for non-inferiority trials will always be larger than the corresponding number of classical superiority trials. For M equals half d, the number will be four times superior. In the case of prevalence trials, the number is still slightly bigger, because two hypotheses are tested for the symmetrical region [+M e -M], that is, non-inferiority and non-superiority.

The appropriate calculation of the sample size is important not only to validate the analysis and the interpretation, but also to the planning of resources; the budget of a trial depends greatly on the number of patients to be recruited. As to more complex experimental designs and more sophisticated analysis strategies, the methods herein described are not enough and, in many cases, simulations may be necessary.

Trial sensitivity

Trial sensitivity is the capability of determining an effective placebo treatment or, on the whole, of detecting differences between treatments when they actually exist¹². It depends on factors such as the magnitude of the treatment effect and the quality of study conduction, besides adherence, patient selection criteria and excess of data variability. It is, therefore, a more enclosing concept than the statistical power of the study.

Ideally, a non-inferiority trial should include a placebo group to assure that, in this precise study, the control group is effective. As this situation is almost always unviable, a non-inferiority comparison that indicates an efficacy difference between the control group and the studied treatment which is smaller than the non-inferiority margin possibly means that both treatments were effective or that both were not. A non-inferiority trial without a placebo group can only be validated if it is possible to assume that the control group is effective⁶. The presupposition of the control group efficacy must be transitively assumed or deduced based on historical data. The historical evidence of regular and consistent efficacy is known by the acronym HESDE (historical evidence of sensitivity to drug effects), which depends mainly on the studied pathology and on the therapeutic field. HESDE may not be easily assumed with regard to drugs that do not show to be regularly superior to placebo such as antidepressant drugs, products destined to dementia, gastroesophageal reflux, allergic rhinitis, among others. On the contrary, HESDE may be more easily adopted for anticoagulants in deep venous thrombosis, antibiotics in urinary tract infections or beta-agonists in bronchospasm. Besides the historical evidence, there must also be the presupposition of constancy, thus assuring that the efficacy evidence has not changed throughout time¹³.

More complex methods, referred to as synthetic methods¹⁴, intend to combine errors from two sources (historical and current data), thus building a kind of meta-analysis type I error, which allows the comparison between the new treatment and the placebo projected effect (putative placebo). Aiming at not compromising the validity of the presented conclusions, the sensitivity analysis must always be conducted and monitored based on external information.

Another important factor to assure the superiority concerning the placebo is the choice of the comparator. If a treatment that showed to be non-inferior to a recognized pattern is used as control in future studies, there might be the risk of demeaning the efficacy and ending up with treatments that are not superior to placebo. Slightly inferior treatments placed within the margin that became control in a new generation of non-inferiority trials result in progressive loss of reliability. This phenomenon of comparator degeneration is called "biocreep"¹⁵ and emphasizes the necessity of a careful control selection. The biocreep may occur, for example, when a generic drug is used as control of a new product, even if it has been considered non-inferior to the original product. It seems obvious that when A is non-inferior to B and C non-inferior to A it does not mean that C is not inferior to B. In a recent publication¹⁶, the diverse factors that could lead to the undesired biocreep were examined by means of simulations - as well as the choice of the control, and violations of constancy presuppositions were considered to be potential sources of the phenomenon.

Analysis population

The results of clinical trials may be assessed by considering two possible data groups. The first one includes the entire randomized population regardless of withdrawals, losses or lack of treatment adhesion, and it is defined by the intention-to-treat, referred to as ITT. The second group includes only the patients that completed the treatment and did not violate the protocol; it is a subgroup of the latter, referred to as per-protocol (PP). Sometimes, an intermediate group known as modified intention-to-treat (MITT) is used, including all patients, except those who did not receive any dose of the treatment.

For the superiority trials there is a consensus that the ITT group must be preferred. The justification is that this strategy prevents attrition bias, preserves the initial randomization and, more importantly, represents protection against suspicion of conscious or unconscious exclusion of undesired data. It is also known that the analysis according to ITT provides a more conservative result, reducing the differences between treatments and, for the same reason, introducing an opposite bias to the equivalence or non-inferiority trials. The analysis according to the PP principle is more effective in discriminating treatments because it is done in a population that is more receptive to it and, for that reason and according to many opinions, is the strategy adopted for the analysis in non-inferiority trials. However, this recommendation is not controversy free because it inflates the type II error. The pragmatic solution found by the main sanitarian authorities is to recommend the analysis by the two methods and require that they be robust and consistent. The analysis planning according to PP influences the calculation of the sample size and must allow an adequate provision for an estimated rate of losses.

Another condition to be considered is the quality of study conduction. A bad adherence, inaccurate measures and demeaned processes increase the variability and hide the differences between treatments. What may be devastating in a superiority trial produces inverse effects in noninferiority trials. By diluting the differences, the probability of false positive results increases, thus defining non-inferior treatments that would be differently assessed in other occasion. The absence of this "quality incentive" makes the noninferiority trials much more complex.

Ethical justification

The principle of clinical equipoise provides an ethical justification for randomized clinical trials. Clinical equipoise is the state of genuine uncertainty regarding which among two or more treatments is the most effective and safe - a clinical trials is conducted to reduce this uncertainty status. There is a considerable disagreement about the use of placebo. In some situations, its use is perfectly justifiable, and in others its employment is not considered to be ethical. The last update of the Declaration of Helsingue (2008)¹⁷ introduced more flexibility in the use of placebo in clinical trials, but the controversy persists. In Brazil, there is a resolution on the National Health Council indicating that "the benefits, risks, difficulties and effectiveness of a new method must be tested in comparison to the best current methods"18. In this manner, the noninferiority trials represent a methodological contribution to reduce the exposition of patients to placebo. However,

the proper definition of these studies implies that the analyzed treatment may be, in less than a margin, inferior to the control, and this is enough to raise ethical issues. A radical ethical objection to the non-inferiority trials was recently published¹⁹, and even suggested their banishment. Assertively, this does not seem to be the opinion of the whole scientific community.

Some criteria may help the ethical scrutiny of the equivalence and non-inferiority trials: the condition must not represent risk of a serious or irreversible damage, the existence of undeniable marginal benefits of the studied treatment (adverse effects, cost, the facility of administration etc.), the informed consent must be well explained and, of course, there must be a control of quality of the study planning and execution (definition of margin, trial sensitivity, adequate control definition etc.).

There are many problems and challenges concerning the conduction of non-inferiority trials. Besides a great deal of recent publications in specialized journals, guidance such as the ICH-E9²⁰, ICH-E10²¹, EMEA^{22,23} may be consulted. Till recently, the FDA recommendations were either generic, based on the ICH recommendations, or too specifically directed to certain products, indications or determined therapeutic fields. However, in March 2010, the FDS published a new and enclosing guidance²⁴ for the planning and analysis of non-inferiority trials. Although this is not its definitive version, the document provides important orientations. To prepare or update reports of non-inferiority trials, some checklists are available. Two of these are specially useful: one prepared by the CONSORT group²⁵ and another by the Canadian Partnership Against Cancer of the McMaster University²⁶. In Brazil, there is still a considerable degree of ignorance concerning this subject, as the use of inappropriate methodologies for non-inferiority trials is frequent. Many treatments classified as non-inferior would not have this classification had they been correctly conducted. Bringing this theme up on discussion and stimulating the development of local guidance seem to be opportunities to be taken.

References

- Fletcher RH. Evaluation of interventions. J Clin Epidemiol. 2002;55:1183-90.
- Altman DG, Bland JM. Treatment allocation by minimisation. BMJ. 2005;330:843.

- 3. Chow SC, Chang M. Adaptive design methods in clinical trials a review. Orphanet J Rare Dis. 2008;3:11.
- Altman DG, Bland JM. Absence of evidence is not evidence of absence. BMJ 1995;311:485.
- Alderson P. Absence of evidence is not evidence of absence. BMJ. 2004;328:476-7.
- 6. Evans SR. Noninferiority clinical trials. Chance. 2009;22:53-8.
- 7. Pocock SJ. The pros and cons of noninferiority trials. Fundam Clin Pharmacol. 2003;17:483-90.
- D'Agostino RB Sr, Massaro JM, Sullivan LM. Non-inferiority trials: design concepts and issues – the encounters of academic consultants in statistics. Stat Med. 2003;22:169-86.
- Kaul S, Diamond GA, Weintraub WS. Trials and tribulations of non-inferiority: the ximelagatran experience. J Am Coll Cardiol. 2005;46:1986-95.
- Chow SC, Shao J, Wang H. Sample size calculations in clinical research. 2nd Ed. Boca Raton: Chapman & Hall/CRC Biostatistical series; 2008.
- 11. Lehr R. Sixteen S-squared over D-squared: a relation for crude sample size estimates. Stat Med. 1992;11:1099-102.
- 12. Pater C. Equivalence and noninferiority trials are they viable alternatives for registration of new drugs? (III). Curr Control Trials Cardiovasc Med. 2004;17:8.
- **13.** Kaul S, Diamond GA. Good enough: a primer on the analysis and interpretation of noninferiority trials. Ann Intern Med. 2006;145:62-9.
- 14. James Hung, HM, Wang SJ, Tsong Y, Lawrence J, O'Neil RT. Some fundamental issues with non-inferiority testing in active controlled trials. Stat Med. 2003;22:213-25.
- **15.** Fueglistaler P, Adamina M, Guller U. Non-inferiority trials in surgical oncology. Ann Surg Oncol. 2007;14:1532-9.
- Everson-Stuart SP, Emerson SS. Bio-creep in non-inferiority clinical trials [Internet]. Biostatistics Working Paper Series, University of Washington, Seatle; 2010 [cited Aug 13, 2010]. http://www.bepress.com/uwbiostat/paper359
- World Medical Association. Ethical principles for medical research involving human subjects, 59th WMA General Assembly, Seoul [Internet]. 2008. [cited Aug 13, 2010]. http://www.wma.net/ en/30publications/10policies/b3/index.html
- Brasil. Ministério da Saúde. Conselho Nacional de Saúde. Resolução 404/2008 de 1 de Agosto de 2008. Brasília: Diário Oficial da União – DOU 186 de 25 de Setembro de 2008; 2008. p. 45 ISSN 1677-7042.
- **19.** Garattini S, Bertele V. Non-inferiority trials are unethical because they disregard patients' interests. Lancet. 2007;370:1875-7.
- International Conference of Harmonization (ICH). Guideline E9 Statistical principles for clinical trials [Internet]. 1998. [cited Aug 16, 2010]. http://www.ich.org/cache/compo/475-272-1.html#E9
- International Conference of Harmonization (ICH). Guideline E10 Choice of control groups and related design and conduct issues in clinical trials [Internet]. 2000 [citado 2010 ago 16]. http://www.ich. org/cache/compo/475-272-1.html#E10

- 22. Committee for Proprietary Medicinal Products. Points to consider on switching between superiority and non-inferiority. Br J Clin Pharmacol. 2001;52:223-8.
- Committee for Proprietary Medicinal Products for Human Use. Guideline on the choice of the non-inferiority margin [Internet]. 2005. [cited Aug 16, 2010]. http://www.ema.europa. eu/docs/en_GB/document_library/Scientific_guideline/2009/09/ WC500003636.pdf
- 24. FDA Guidance for Industry. Non-inferiority clinical trials: draft guidance. US Dept of Health and Human Service FDA/CDER/ CBER [Internet]. 2010. http://www.fda.gov/downloads/Drugs/ GuidanceComplianceRegulatoryInformation/Guidances/ UCM202140.pdf
- **25.** Piaggio G, Elbourne DR, Altman DG, Pocock SJ, Evans SJ; CONSORT Group. Reporting of noninferiority and equivalence randomized trials: an extension of the CONSORT statement. JAMA. 2006;295:1152-60.
- 26. Canadian Partnership Against Cancer. Capacity Enhancement Program: non-inferiority and equivalence trials checklist [Internet]. McMaster University; 2009. [cited Aug 13, 2010]. http://fhs.mcmaster.ca/cep/documents/CEPNon-inferiorityandEquivalenceTria IChecklist.pdf

Correspondence: Valdair Ferreira Pinto Alameda dos Aicás, 229 CEP 04086-000 – São Paulo, SP E-mail: valdairpinto@hotmail.com