

REVIEW ARTICLE

Non-inferiority randomized trials, an issue between science and ethics: The case of the SYNTAX study

VITTORIO MANTOVANI¹, VINCENZO LEPORE², ANTONIETTA MIRA³ & EVA BERGLIN²

¹Department of Cardiac Surgery, University of Insubria and Ospedale di Circolo di Varese, Varese, Italy,

²Department of Cardio-Thoracic Surgery, Sahlgrenska University Hospital, Göteborg, Sweden and

³Department of Economics, University of Insubria, Varese, Italy

Abstract

Non-inferiority trials are questionable when death and serious complications are included among outcomes. The term itself “non-inferiority” is misleading, since such a study would not demonstrate that a new treatment is non-inferior to a control treatment, but simply that the inferiority would not reach a pre-specified level, deemed as acceptable by the designers of the trial. Group cross-over, assay-sensitivity and the need of a placebo arm are major issues for the reliability of non-inferiority trials.

The SYNTAX trial for severe coronary artery disease was designed on a non-inferiority margin of 6.6%. In this paper we show that the SYNTAX designers were ready to accept up to 30% higher rate of death and major adverse events to claim the non-inferiority of percutaneous coronary intervention versus coronary artery bypass grafting. Eventually the SYNTAX study failed because percutaneous patients sustained an even higher rate of adverse events. We propose major caution in performing non-inferiority randomized trials.

Key words: *Non-inferiority trial, percutaneous coronary intervention (PCI), coronary artery bypass graft (CABG), Syntax-study*

Randomized controlled trials (RCTs) are generally considered the best available source of evidence when comparing different medical treatments. Three different statistical concepts can be identified in RCTs: superiority, equivalence and non-inferiority trials (1).

In superiority trials one treatment is declared better than the control treatment if the outcome of the former outperforms the latter. An equivalence trial aims at demonstrating that the outcome of one treatment falls within a certain pre-set range above or below the outcome of the control treatment (1,2). Non-inferiority may be confused with equivalence, but this should be avoided (3–7). Non-inferiority trials do not demonstrate that the investigated treatment is actually non-inferior to the reference treatment in the sense that the difference in effect be null. Instead, the design aims to exclude that the inferiority would reach a critical value, which the study

designers consider the maximum tolerable loss of effect in order to accept the new treatment.

The aim of this paper is to discuss the concept of non-inferiority used in the SYNTAX-trial (8). The SYNTAX trial compared percutaneous coronary interventions (PCI) using drug eluting stents, versus coronary artery bypass grafting (CABG) in patients with three-vessel coronary artery disease and/or left main stem stenosis.

Non-inferiority trials

Theoretically, new treatments should be approved only if they are better than the currently accepted therapies, hence with superiority trials. Superiority trials commonly require a lower number of patients than other statistical models (2), and may thus be less expensive. Nevertheless, non-inferiority trials are used to test new treatments with increasing frequency (9).

The aim of a non-inferiority trial is to declare a new treatment acceptable, not to demonstrate improvement in clinical outcome (9). This approach was mainly used by pharmaceutical companies to obtain the permission from control agencies to market new drugs.

The European Medicines Agency (EMA) published in 2005 guidelines on the choice of the non-inferiority margin in non-inferiority trials (10). Basically, it might be appropriate to approve a new treatment with a marginally lower efficacy than the currently accepted reference treatment, on the condition that the new treatment offers other relevant benefits. This guideline states that “it may be useful to specify co-primary endpoints, one to demonstrate superiority in terms of the safety endpoint, the other non-inferiority on the efficacy endpoint”.

Several issues make the use of non-inferiority trials questionable:

- a) *“Bio-creep”*. If a new treatment is approved as non-inferior to the current golden standard, it might become itself the future control for a third treatment, which would then be a double step down compared to the original golden standard (6,11,12).
- b) *Cross-over bias*. Cross-over between the assignment groups generally confounds the results of any trial. This phenomenon is unfavorable for the new treatment in superiority trials, as it may narrow the advantage over the control. In non-inferiority trials, cross-over might be favorable for the new treatment, as it dampens the differences versus the control treatment and makes it easier to declare the non-inferiority. Intention-to-treat (ITT) analysis protects against cross-over bias in superiority trials (13); but might favor the new treatment in non-inferiority trials (2,14), making both ITT and per-protocol analysis advisable (1,15).
- c) *Assay-sensitivity*. A poor efficiency of researchers in detecting adverse events in a blinded randomized non-inferiority trial would favor the new treatment against the control treatment (12). Paradoxically, if there were no events at all in either group, the new treatment would automatically be non-inferior to the control group. This risk does not exist in superiority trials, where the sponsors are motivated in finding all adverse events, in the hope that they would be more frequent in the control group. If superiority is demonstrated, questions about assay sensitivity do not arise.
- d) *Placebo*. In a superiority trial, if superiority of the new treatment is proved over the active control arm, one does not need to control how superior the reference was over placebo, since a step forward in outcome is made anyway. In non-inferiority trials it is important to know exactly the margin of superiority of the control treatment over placebo, since we presume that the new treatment will place itself somewhere between the control and the placebo. A properly designed non-inferiority trial should include three arms: the new treatment, the control treatment and a placebo arm (10). The inclusion of placebo in the trials is a strong guarantee against a low assay-sensitivity and against choosing a non-inferiority margin too far from the active control and too close to the placebo. If a placebo arm is not possible or ethical, then the non-inferiority model is not appropriate. EMA published in 2010 guidelines on medical products for the treatment of Alzheimer’s disease and other dementias (16). This guideline states that “due to concerns over assay sensitivity, the use of a non-inferiority design without a placebo arm will not be accepted as proof of efficacy”.
- e) *Non-inferiority margin*. The choice of the non-inferiority margin itself is critical. A large non-inferiority margin allows the study designers to enroll a smaller number of patients. The choice of the non-inferiority margin should, according to EMA, be based “upon a combination of statistical reasoning and clinical judgement” (10). The non-inferiority margin should always be sufficiently far from placebo level and “independent of considerations of power” (10).
- f) *Endpoints*. According to EMA’s guidelines (10) is it “very difficult to justify a non-inferiority margin of any size in a study where the treatment under consideration is used for the prevention of death or irreversible morbidity and there is no second chance for treatment. Discussion of the number of extra deaths that are acceptable is ethically very difficult”.

The SYNTAX trial

The SYNTAX trial was a prospective randomized trial of percutaneous coronary intervention (PCI) with drug eluting stents versus coronary-artery bypass grafting (CABG) in coronary artery disease (8). Even though the SYNTAX was not a comparison between two drugs, the EMA guidelines offer the possibility for a discussion.

It was performed by randomizing 1 800 patients to either treatment. PCI would be declared non-inferior to CABG “if the one-sided 95% upper confidence limit for the difference was less than the pre-specified delta value (6.6%)”, the endpoint being a combination of death and major adverse events (myocardial infarction, stroke and repeat revascularization). The number of patients to enroll was calculated on the basis of an expected rate of primary end-point of 13.2% for CABG and 14.0% for PCI, based on literature analysis (8) (on-line supplement to the paper).

According to the published results from the SYNTAX trial, the primary outcome after one year occurred in 105 of 849 patients in the CABG group (12.4%) and in 159 of 891 in the PCI group (17.8%). Non-inferiority could not be declared because the difference between the two groups was 5.5% point with an upper one-tail 95% confidence interval of 8.3% (8).

Given the observed outcome rate of 105/849 for CABG, it can be calculated that the authors would have declared PCI non-inferior with up to 144 major events among 891 patients, i.e. 16.2% (17). This figure would give a 3.8% difference with an upper one-tail 95% confidence limit of 6.6%.

The ethics of this design can be questioned. By using a non-inferiority margin of 6.6%, the SYNTAX designers considered acceptable an approximately 30% higher rate of death and other major adverse events. If the outcome rate for CABG had been the expected 13.2%, i.e. 119 of 900 prospective patients, then PCI would have been declared non-inferior with up to 153 events in the 900 patients (17%) enrolled to PCI. This figure would result into a 3.8% absolute difference with an upper one-tail 95% confidence limit of 6.6%.

The justification for a non-inferiority study is that some loss of efficacy might be accepted in exchange for other benefits, first safety. If PCI is proposed as an alternative to CABG for three-vessel and main-stem disease, this should be done in the hope that survival and major complications would be at least equivalent and not up to 30% worse. In our opinion, the SYNTAX study disregarded the safety of the patients, by including death and major complications in the primary end-point of a non-inferiority study. The PCI patients, compared to CABG group, suffered 54 more major adverse events, including nine deaths.

Other methodological issues for discussion are:

- a) The EMEA guideline (10) states that if the actual performance of the reference treatment in the trial is different from “what was assumed when defining the non-inferiority

margin, then the chosen margin may no longer be appropriate.” In the SYNTAX trial the primary outcome for patients randomized to CABG was 12.4%, i.e. 6% lower than the expected value of 13.2%. Furthermore, in the CABG registry of patients enrolled but not randomized, the primary outcome rate was 8.8% (Table 7 of the online supplement to the paper) (8).

- b) When a placebo arm cannot be included in a non-inferiority trial, historical data should be used as “putative placebo” in order to establish a reasonable non-inferiority margin and control for assay sensitivity. If no reliable historical data about a placebo effect are available, the reference treatment should be used as baseline and the researchers should try to demonstrate direct superiority of the new treatment over the reference, rather than non-inferiority (10). In the case of severe coronary artery disease a placebo group is clearly unethical; we suggest that the closest figure to placebo would be medical therapy. Historical survival data for coronary patients treated with CABG vs. medical therapy before PCI development are out-of-date. In the MASS-II trial, the PCI outcome was closer to medical therapy than to surgical treatment at five years (18), while after one year the survival free of cardiac mortality and myocardial infarction was even better for medical therapy than PCI (19). These data make the 6.6% non-inferiority margin of SYNTAX questionable not only from the ethical point of view, but even from a statistical perspective.
- c) The use of a composite end-point (MACCE, major adverse cardiac and cerebrovascular events: death, stroke, myocardial infarction and repeat revascularization) as main outcome has an intrinsic weakness, putting together events with very different impact on patient’s life. Its main justification is that analyzing the single adverse events per se would require a too large sample. The Authors of SYNTAX correctly concluded that PCI is inferior to CABG in this setting, but then compared the adverse events one by one using χ^2 , specifically Table 3 of their publication (8). We are seriously concerned that Table 3 would be misunderstood. It might support among physicians the idea that the inferiority of PCI versus CABG lies in a “minor component” of the composite outcome, i.e. the need of repeat revascularization, while the “substantial components” of

the outcome, i.e. death and myocardial infarction are similar, and stroke even better with PCI. Since the authors chose a non-inferiority model, which is not based on traditional χ^2 , they should be consistent and avoid using a statistical test that would require 30-times more patients, rather than just stating that “subgroup analyses can only be considered as hypothesis-generating”. Given the expected main outcome rates of 13.2% for CABG and 14.0% for PCI respectively, if the SYNTAX designers wished to analyze their data with χ^2 test, with $p < 0.05$ and power 90%, at least 53 750 patients should have been enrolled, and this for the combined outcome only (20).

Conclusions

Non-inferiority is a questionable concept bearing important shortcomings. In the SYNTAX-trial, the fact that PCI failed to demonstrate non-inferiority versus CABG, makes this result even more negative for PCI than one would believe at first glance.

Declaration of interest: The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

References

- Christensen E. Methodology of superiority vs. equivalence trials and non-inferiority trials. *J Hepatol.* 2007;46: 947–54.
- Jones B, Jarvis P, Lewis JA, Ebbutt AF. Trials to assess equivalence: The importance of rigorous methods. *BMJ (Clin Res).* 1996;313(7048):36–9.
- Greene WL, Concato J, Feinstein AR. Claims of equivalence in medical research: Are they supported by the evidence? *Ann Intern Med.* 2000;132:715–22.
- Siegel JP. Equivalence and noninferiority trials. *Am Heart J.* 2000;139:S166–70.
- Snapinn SM. Noninferiority trials. *Curr Control Trial Cardiovasc Med.* 2000;1:19–21.
- 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. Good enough: A primer on the analysis and interpretation of noninferiority trials. *Ann Intern Med.* 2006;145:62–9.
- Serruys PW, Morice MC, Kappetein AP, Colombo A, Holmes DR, Mack MJ, et al. Percutaneous coronary intervention versus coronary-artery bypass grafting for severe coronary artery disease. *New Engl J Med.* 2009;360:961–72.
- Garattini S, Bertele V. Non-inferiority trials are unethical because they disregard patients’ interests. *Lancet.* 2007; 370(9602):1875–7.
- Committee for Medicinal Products for Human Use (CHMP) guideline on the choice of the non-inferiority margin. *Stat Med.* 2006;25:1628–38.
- 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.
- Fleming TR. Current issues in non-inferiority trials. *Stat Med.* 2008;27:317–32.
- Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D, et al. The revised CONSORT statement for reporting randomized trials: Explanation and elaboration. *Ann Intern Med.* 2001;134:663–94.
- Piaggio G, Elbourne DR, Altman DG, Pocock SJ, Evans SJ. Reporting of noninferiority and equivalence randomized trials: An extension of the CONSORT statement. *JAMA.* 2006; 295:1152–60.
- Le Henaff A, Giraudeau B, Baron G, Ravaud P. Quality of reporting of noninferiority and equivalence randomized trials. *JAMA.* 2006;295:1147–51.
- EMA/CHMP Guideline on medical products for the treatment of Alzheimer’s disease and other dementias, doc. ref. CPMP/EWP/553/95 Rev.1. updated 2008 Nov 25, cited 2010 Feb 12. Available from: www.ema.europa.eu/pdfs/human/ewp/055395en.pdf.
- Dann RS, Koch GG. Methods for one-sided testing of the difference between proportions and sample size considerations related to non-inferiority clinical trials. *Pharm Stat.* 2008;7:130–41.
- Hueb W, Lopes NH, Gersh BJ, Soares P, Machado LA, Jatene FB, et al. Five-year follow-up of the Medicine, Angioplasty, or Surgery Study (MASS II): A randomized controlled clinical trial of 3 therapeutic strategies for multivessel coronary artery disease. *Circulation.* 2007;115:1082–9.
- Hueb W, Soares PR, Gersh BJ, Cesar LA, Luz PL, Puig LB, et al. The medicine, angioplasty, or surgery study (MASS-II): A randomized, controlled clinical trial of three therapeutic strategies for multivessel coronary artery disease: One-year results. *J Am Coll Cardiol.* 2004;43:1743–51.
- Faul F, Erdfelder E, Lang AG, Buchner A. G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav Res Methods.* 2007;39:175–91.