

Non-inferiority trials are unethical because they disregard patients' interests



Silvio Garattini, Vittorio Bertele*

Equivalence trials¹ have been widely used to assess new drugs, but have recently lost ground to a non-inferiority design. This type of trial is usually accepted by regulatory authorities for approval of new drugs or new indications, although the US Food and Drugs Administration has raised some concerns.² In this paper, we argue that the scientific community should ban non-inferiority and equivalence trials because they are unethical, whatever measures are taken to prevent their methodological pitfalls and inappropriate interpretation of results.³⁻⁸ Exceptions might exist, but we could not identify a situation in which patients can justifiably be entered into a trial that will not provide them with any advantage.

Pretext for looking for non-inferiority

Use of equivalence or non-inferiority rather than superiority designs implies the intention of not trying to prove any additional value of new drugs. However, the declared aim is to expand treatment options for patients with poor tolerance of, or no response to, available products. Drug producers argue that there is no reason to define the benefit-risk profile of new agents as better than those of existing drugs: it is enough to show that they are similar. One does not even need to know whether a new drug with some innovative peculiarities—for example, longer activity—is more effective. The added value rests on the probability of better compliance with, for instance, once-a-day administration compared with a more complex regimen. Similarly, the added value of a more convenient formulation arguably lies in its ease of use. Generally, when a new drug is claimed to have only minor advantages or no advantage over available products, a superiority test is not believed to be necessary; non-inferiority allows new products to compete with older ones on the basis of small differences made to seem to benefit patients.

Looking for non-inferiority or overlooking differences?

What is wrong with this approach? Problems arise from the definition of non-inferiority and the statistical criteria for its basis.³⁻⁸ Non-inferiority is a kind of similarity within a limit. The limit is the degree of tolerable inferiority of the new drug compared with the standard treatment. This arbitrary difference in efficacy, the non-inferiority margin or delta, is decided before doing the study. Non-inferiority is judged to have been established when the point estimate and 95% CI of the effect of the new drug do not fall outside the preset non-inferiority margin. A non-inferior test drug could actually be less effective or less safe than the comparator,

but not to the extent that it is recognised as such. For example, if the non-inferiority limit is set at 7.5%, an increase in the incidence of serious events or deaths—say 7% instead of the 5% currently established for the comparator—is not seen as large enough to mark a difference between the new and the control drug. The new drug will therefore be considered non-inferior to the old drug, even if in 1000 patients treated with the former, there could be 20 more deaths than with the latter.

These arguments also apply to equivalence trials, which aim to prove similarity of a new drug to the comparator, since true equivalence is theoretical and is difficult to demonstrate. Equivalence means that a new drug is not much worse than the comparator (as in non-inferiority trials), but also is not much better. Similarity is defined by limits that include a superiority margin as well as a non-inferiority margin. Since equivalence trials explore the differences between control and study treatments in both directions, they provide a more reliable estimate of the relative efficacy of two treatments than do non-inferiority trials. However, use of a non-inferiority limit exposes equivalence trials to the same ethics issues.

No limits to the non-inferiority limit

The wider the non-inferiority interval, the smaller the sample needed. The smaller the sample, the smaller the investment needed to do the trial, as well as the greater the chance of overlooking a difference and concluding non-inferiority. This situation has led to the adoption of extreme hypotheses, which are just arbitrary, yet approved by ethics committees and allowed in the scientific literature; for example, in the COMPASS study,⁹ the thrombolytic saruplase was judged equivalent to streptokinase for post-myocardial infarction, even though the saruplase group had 50% more deaths than the control group. Therefore, in absolute numbers, saruplase would be regarded as effective and safe as streptokinase even if there were 35 deaths per 1000 treated in addition to the 70 with streptokinase alone. The test of this questionable hypothesis only required about 3000 patients, in an era in which testing the superiority of tissue-type plasminogen activator over streptokinase involved about 90000 patients overall in three large clinical trials.¹⁰⁻¹²

The results of trials like COMPASS also arouse concern about the breadth of their confidence intervals. Sometimes these are so wide that what is judged non-inferior from a statistical point of view might actually be questionable from a clinical point of view.^{13,14}

Lancet 2007; 370: 1875-77

Published Online

October 23, 2007

DOI:10.1016/S0140-

6736(07)61604-3

Mario Negri Institute for Pharmacological Research, Milan, Italy (Prof S Garattini MD, V Bertele* MD)

Correspondence to:

Dr Vittorio Bertele*, Mario Negri

Institute for Pharmacological

Research, Milan, 20156, Italy

bertele@marionegri.it

Unreliable messages from questionable methods

As with a superiority design in a placebo-controlled trial, evidence of non-inferiority to active comparators might allow drugs onto the market that are in fact less acceptable than those in current clinical use. Worse, if the difference between the standard treatment and placebo is small, dependent on the non-inferiority limit, the effect of the supposedly non-inferior drug might actually be close to that of placebo. In any case, the loss in efficacy might be greater than it appears, since the effect of the standard treatment includes that of placebo. For example, if the standard treatment prevents 30% of expected events and the non-inferiority margin allows the new drug to prevent only 20%, the allowable loss in efficacy appears to be a third—but if the placebo effect accounts for 10% of the overall action, half the efficacy could actually be lost. Thus, non-inferiority trials expose patients to clinical experiments without any assurance that the experimental drug is not worse than the standard treatment, and without really exploring whether it is better.

Commercial aims, not patients' interests

Are there specific reasons for allowing a non-inferiority approach? One reason cited is that for patients who do not respond to existing treatments, products with similar activity could offer a useful alternative. The aim is reasonable, but the approach is not. If the target is non-responders to current treatments, why not test the new agents' superiority in this subset, rather than its non-inferiority in the overall population? This approach would meet patients' needs best, but restricts the market that can be targeted by the drug companies.

Another suggested reason is that non-inferior drugs might be better tolerated or easier to use than existing treatments. However, these features are unlikely to be confirmed in non-inferiority trials, since any advantage should translate into better compliance and result in a superior rather than a non-inferior outcome.

Superiority trials are also said to generally take much longer and require many more patients than do non-inferiority trials, delaying the availability of potentially useful drugs. However, non-inferiority trials do not necessarily need a smaller sample size, which can be the result of selecting a large inferiority margin or of other questionable methodological choices.¹ Moreover, it is our view that a delay in the availability of proven effective drugs is preferable to early availability of potentially advantageous drugs whose real efficacy has not been formally established. Actual efficacy testing might never be done, particularly if patients no longer agree to be randomly assigned to older drugs.

A more convincing approach might be to test non-inferior efficacy for the sake of improved safety. This strategy is reasonable if the outcome events used to measure efficacy have clinical importance similar to those for safety as, for instance, deaths and haemor-

rhagic strokes after thrombolysis in acute myocardial infarction. In these circumstances, however, a superiority trial would be a preferable way to compare the effectiveness of two treatments in terms of survival without stroke by cumulatively measuring efficacy and safety events.

These examples are intended to show that any question of practical relevance to patients requires a test of superiority. The superiority approach, whether the hypothesis is verified or not, provides information about new drugs in the context of available treatments, whereas the non-inferiority trial does not. From a commercial point of view, to prove non-inferiority of new products is less risky than aiming to establish their superiority. Failure to prove superiority can tarnish the product's commercial image, although it could provide more information for doctors and patients. The non-inferiority approach is likely to overlook differences that might stop the product getting onto the market. A demonstration of non-inferiority leaves the product in a kind of limbo: its place in therapy is not established, although its place on the market is assured.

Enrolling patients in non-inferiority trials betrays their trust

We believe that non-inferiority studies have no ethical justification, since they do not offer any possible advantage to present and future patients, and they disregard patients' interests in favour of commercial ones. This situation betrays the agreement between patients and researchers set out in any fair informed consent form that presents randomised trials as the only ethical way to address clinical uncertainty. Non-inferiority trials claim minor advantages for the test drugs, but do not prove their efficacy compared with older products. Few patients would agree to participate if this message were clear in the informed consent form: as we said before, why should patients accept a treatment that, at best, is not worse, but could actually be less effective or less safe than available treatments?¹⁵

In conclusion, we believe that non-inferiority trials fail to meet the commitments of good clinical research: "Ask an important question, and answer it reliably".¹⁶ Although a non-inferiority study reduces research and development costs and commercial risks thereafter, it asks no relevant clinical questions. Randomisation should not even be allowed in such trials, since it is unethical to leave to chance whether patients receive a treatment that is anticipated to provide no extra benefit, but could be less safe and less effective than existing treatment options.

With regard to the reliability of the methods and consequently of the results, the uncertainty surrounding alleged non-inferiority is hard to accept; however small an increase in relative risk, the increase in risk unavoidably implies an absolute excess of adverse events in the population. Sometimes the risk turns out to be significantly greater in the test treatment group, without

this difference necessarily disproving non-inferiority.¹⁷ To expose patients to such risks in the trial and in real life, with no benefit in exchange, is clearly unethical. We hope that these arguments will foster a debate on the issue.

Conflict of interest statement

We declare that we have no conflict of interest.

Acknowledgments

We thank Iain Chalmers, Brian Godman, and Valter Torri for reviewing the manuscript and providing useful suggestions, and Judith Baggott for editorial assistance.

References

- Bertele' V, Torri, W, Garattini S. Equivocal equivalence. <http://www.marionegri.it/page.asp?idp=891441724> (accessed Oct 17, 2006).
- US FDA's non-inferiority stance could slow sinusitis approvals. SCRIIP—World Pharmaceutical News, Sept 18, 2006: 27.
- Siegel JP. Equivalence and noninferiority trials. *Am Heart J* 2000; **139**: S166–70.
- 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.
- Hung HM, Wang SJ, O'Neill R. A regulatory perspective on choice of margin and statistical inference issue in non-inferiority trials. *Biom J* 2005; **47**: 28–36.
- 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.
- Snapinn SM. Alternatives for discounting in the analysis of noninferiority trials. *J Biopharm Stat* 2004; **14**: 263–73.
- Kaul S, Diamond GA. Good enough: a primer on the analysis and interpretation of noninferiority trials. *Ann Intern Med* 2006; **145**: 62–69.
- Tebbe U, Michels R, Adgey J, et al. Randomized, double-blind study comparing saruplase with streptokinase therapy in acute myocardial infarction: the COMPASS equivalence trial. *J Am Coll Cardiol* 1998; **31**: 487–93.
- Gruppo Italiano per lo Studio della Sopravvivenza nell'Infarto miocardico. GISSI-2: a factorial randomised trial of alteplase versus streptokinase and heparin versus no heparin among 12 490 patients with acute myocardial infarction. *Lancet* 1990; **336**: 65–71.
- ISIS-3 (Third International Study of Infarct Survival) Collaborative Group. ISIS-3: a randomised comparison of streptokinase vs tissue plasminogen activator vs anistreplase and of aspirin plus heparin vs aspirin alone among 41 299 cases of suspected acute myocardial infarction. *Lancet* 1992; **339**: 753–70.
- The GUSTO investigators. An international randomized trial comparing four thrombolytic strategies for acute myocardial infarction. *N Engl J Med* 1993; **329**: 673–82.
- Bertele' V, Torri V, Garattini S. Inconclusive messages from equivalence trials in thrombolysis. *Heart* 1999; **81**: 675–76.
- Barbui C, Violante A, Garattini S. Does placebo help establish equivalence in trials of new antidepressants? *Eur Psychiatry* 2000; **15**: 1–6.
- Garattini S, Bertele' V, Li Bassi L. How can research ethics committees protect patients better? *BMJ* 2003; **326**: 1199–201.
- Yusuf S, Collins R, Peto R. Why do we need some large, simple randomized trials? *Stat Med* 1984; **3**: 409–22.
- Splawinski J, Kuzniar J. Clinical trials: active control vs placebo—what is ethical? *Sci Engineer Ethics* 2004; **10**: 73–79.