

Tools & Techniques - Statistics: Non-inferiority trials

Timothy John Collier

Department of Medical Statistics, London School of Hygiene and Tropical Medicine, London, United Kingdom

The classic randomised placebo-controlled parallel group trial is designed to demonstrate the superiority of a new treatment over an inactive placebo. However, where a proven effective treatment for the condition being investigated exists, randomising patients to a placebo can be ethically unacceptable, particularly when denying the patient an effective treatment could have serious consequences on morbidity or mortality. According to the Declaration of Helsinki¹, the effectiveness of a new intervention must be tested against those of the best current proven interventions. The use of a placebo is only advocated where no proven intervention currently exists, or if the patient will not be subject to any risk of serious or irreversible harm. Similarly, the International Conference on Harmonization guidance on the choice of control group² states that where an available treatment is known to prevent serious harm, such as death or irreversible morbidity in the study population, it is generally inappropriate to use a placebo.

In such situations a randomised controlled trial (RCT) may still be designed to demonstrate the superiority of a new treatment over an active control. However, advances in treatments, improved outcomes, and the diminishing incremental benefits of new treatments mean that demonstrating superiority becomes far more difficult, requiring increasingly large trials.

Consequently, non-inferiority (NI) clinical trials have become increasingly common over the past 20 years. These are trials in which the goal is to demonstrate that a new treatment is not less effective or worse than an active control by more than a pre-specified margin. Hence, NI trials are one-sided in nature since the concern is only to rule out inferiority beyond a certain amount without considering whether the new treatment might be more effective.

The choice of the active control treatment is crucial. This should be the best available proven treatment - typically the "gold standard" if one exists. Selecting a treatment that is not the best available or the efficacy of which is unproven or uncertain is both ethically and scientifically inappropriate. Ideally, there should be a good body of reliable and consistent placebo-controlled trial evidence demonstrating superiority so that the efficacy and safety profile of the control treatment is clearly understood. In order to be able to draw sound conclusions about the efficacy of the new treatment compared to placebo, the NI trial should closely follow the design (particularly in terms of the primary endpoint and patient characteristics) of those trials in which the superiority of the control treatment versus placebo has been demonstrated.

When seeking to demonstrate non-inferiority it is usually assumed that the new treatment carries some other advantage over the active control. For example, the new treatment may be cheaper, less invasive, e.g., laparoscopic versus open surgery, or be easier to administer, e.g., once weekly versus daily injections. The new treatment might also be more tolerable or have fewer or less serious side effects, which could lead to improved adherence or greater efficacy on secondary endpoints. If these benefits are substantial a limited loss of efficacy may be considered a price worth paying. NI trials can also involve development of so-called "me too" treatments which simply add to the number of available treatments for a condition. Although such treatments may not carry any substantial additional benefits in terms of efficacy, cost, or safety, they do increase the range of therapeutic options from which individual patients might benefit.

NI trials are statistically challenging in terms of design and analysis. It is important to understand that demonstrating non-inferiority is not equivalent to failing to demonstrate superiority in a standard RCT. In a standard superiority trial the null hypothesis is that the treatments are equal in effect with the alternative hypothesis that the two treatments are not equal. A statistical test is carried out and a p-value calculated - the smaller the p-value the greater the evidence against the null hypothesis. However, large p-values (large typically meaning values greater than 0.05) should not be interpreted as proving the null hypothesis. If this was the approach, the goal of demonstrating non-inferiority could easily be achieved by carrying out a small (underpowered) trial, or through poor standards of trial conduct which tends to dilute any treatment differences.

A different statistical approach is required to demonstrate NI. Since it is impossible to prove the equality of effect of two treatments, the approach taken in NI trials is to rule out the possibility that the new treatment is less effective than the comparator by more than a pre-specified amount - referred to as the margin of non-inferiority (MNI). The sample size is chosen so that if the new treatment truly is non-inferior the upper limit of the 95% confidence interval (CI) for the difference in treatment effect (Control - New) will be less than MNI. At the end of the study the treatment difference is calculated and, if the upper limit of the 95% CI is less than MNI, non-inferiority is claimed.

Figure 1 shows estimates and 95% CIs for three hypothetical NI trials. Note that since the goal of NI trials is to rule out inferiority we are simply interested in the upper limit of the CI. In scenarios 1 and 2, NI can be claimed since the upper limit does exceed MNI whereas non-inferiority cannot be claimed in scenario 3, since

*Corresponding author: Department of Medical Statistics, London School of Hygiene and Tropical Medicine, London, WC1E 7HT, United Kingdom. E-mail: Timothy.Collier@lshtm.ac.uk

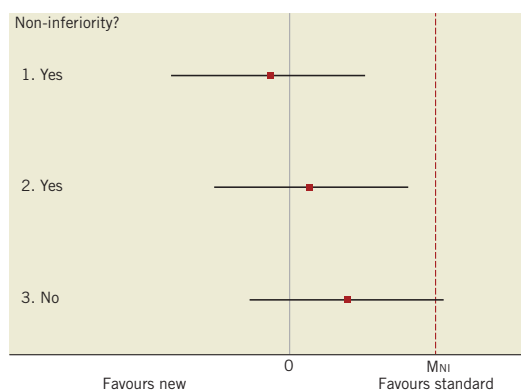


Figure 1. Three possible scenarios for results from a non-inferiority trial.

the upper limit extends beyond MNI. In the DUTCH PEERS trial³, which compared two third-generation drug-eluting stents, the margin of non-inferiority was set at 3.6%. The absolute risk difference was 0.88% (slightly favouring the active control), with the upper limit of the one-sided 95% CI being 2.69%. This corresponds to scenario 2 in **Figure 1**.

The decision on the value of MNI is perhaps the most critical issue in an NI trial and should involve careful clinical and statistical judgement taking into account a number of important factors. Choosing too narrow a margin can result in unfeasibly large trials and truly non-inferior treatments being missed, whereas a wide margin can allow treatments that are no better than a placebo to enter clinical practice. Critically, there needs to be a reliable and precise measure of the efficacy of the active control versus placebo (ECP). This could be estimated by means of a meta-analysis of the results from historical RCTs involving the control. The value of MNI involves judgement as to the maximum clinically acceptable loss of ECP; certainly MNI should be considerably smaller than ECP so that demonstrating non-inferiority infers a minimum efficacy of the new treatment compared to placebo. Consideration should also be given to the nature and frequency of the primary endpoint. The more serious and more frequent the endpoint, the more stringent MNI should be. The nature and scale of any potential benefits of the new treatment should also be considered. Where a new treatment carries major benefits, e.g., reduced cost, less invasive, lower risk, fewer or less serious side effects, then there can be greater flexibility in the choice of MNI. If the new treatment is simply a “me too” intervention then MNI should typically be small compared to ECP.

As the choice of MNI will usually be smaller than the clinically relevant difference used in a placebo-controlled superiority trial, NI trials tend to require larger sample sizes. For a given event rate, the smaller the value of MNI the greater the sample size required. It is important that the choice of the margin should be clearly reported as part of the sample size calculations.

Maintaining a high quality of study design and conduct is of great importance in ensuring the scientific validity of a NI trial. Factors such as poor blinding, low adherence, losses to follow-up and misclassification of endpoints tend to make two treatments appear

more similar. In a superiority trial such things generally bias results towards the null hypothesis and are therefore conservative – the result may be that effective treatments are missed. However, the opposite is true for NI trials. Poor quality produces bias towards the alternative hypothesis, and therefore increases the probability of demonstrating non-inferiority – the result may be that ineffective or even harmful treatments enter into clinical practice. Therefore, there must be an even greater emphasis on rigorous methods in NI trials.

Finally, it is important that the methods and results of NI trials are clearly reported in order to allow readers to draw reliable conclusions. In particular, the reason for using an NI design, the choice of active control and decision on the non-inferiority margin should be clearly explained with ethical, clinical and statistical justification. The CONSORT (Consolidated Standards of Reporting Trials) Statement has been extended for the reporting of NI trials⁴ and provides a useful author’s checklist of items.

Conclusion

NI trials have become increasingly common over the past 20 years, due in part to the decreasing incremental benefits of new treatments and the necessity of having an active control when randomising to an inactive placebo would be ethically unacceptable. However, NI trials carry particular ethical, statistical and organisational challenges which differ from those of standard superiority trials. In order to produce scientifically sound reliable results they need to be carefully designed, rigorously conducted and appropriately analysed and reported.

Conflict of interest statement

The author has no conflicts of interest to declare.

References

1. The World Medical Association, Declaration of Helsinki - Ethical Principles for Medical Research Involving Humans, 59th General Assembly, Seoul, Korea, October 2008.
2. International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH): Choice of Control Group and Related Design and Conduct Issues in Clinical Trials (ICH E-10) Food and Drug Administration, DHHS, July 2000. <http://www.fda.gov/ohrms/dockets/98fr/051401b.htm>
3. von Birgelen C, Sen H, Lam MK, Danse PW, Jessurun GA, Hautvast RW, van Houwelingen GK, Schramm AR, Gin RM, Louwerenburg JW, de Man FH, Stoel MG, Löwik MM, Linssen GC, Saïd SA, Nienhuis MB, Verhorst PM, Basalus MW, Doggen CJ, Tandjung K. Third-generation zotarolimus-eluting and everolimus-eluting stents in all-comer patients requiring a percutaneous coronary intervention (DUTCH PEERS): a randomised, single-blind, multicentre, non-inferiority trial. *Lancet*. 2014;383:413-23.
4. 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.