Non-inferiority trials

Priya Ranganathan, C. S. Pramesh[1], Rakesh Aggarwal[2]

Department of Anaesthesiology, Tata Memorial Centre, Homi Bhabha National Institute, [1]Division of Thoracic Surgery, Tata Memorial Hospital, Tata Memorial Centre, Homi Bhabha National Institute, Mumbai, Maharashtra, [2]Jawaharlal Institute of Postgraduate Medical Education and Research, Puducherry, India

Abstract

Studies sometimes aim to show that a new intervention is not substantially worse than the existing standard of care while offering some benefits, for example, lower cost, decreased toxicity, or easier administration. Such studies are called non-inferiority (NI) trials. In this article, we look at some aspects of NI trials.

Keywords: Randomized controlled trials as topic, research design, research methodology

Conventionally, clinical trials are carried out to show that a new (experimental) intervention has superior efficacy to the existing standard of care (the comparator). Sometimes, however, trials are conducted to show that a treatment is almost as good as (or not much worse than) the comparator. The rationale of such trials is that the new treatment might offer other benefits such as lower cost, greater ease of administration, or reduced toxicity. These are called non-inferiority (NI) trials. Three examples of NI trials are listed below:

a. Oral supplementation with 20 mg/day of zinc had been proven to decrease the duration and number of diarrheal episodes in children; however, this was associated with gastric irritation and vomiting. Dhingra *et al.* conducted a randomized study to look at the effect of lower dose zinc (5 and 10 mg rather than 20 mg) in children with acute diarrhea. They hypothesized that a lower dose would be almost as effective in controlling diarrhea, and be associated with less vomiting. Thus, lower dose zinc would be noninferior to standard dose zinc.[1]

b. The addition of trastuzumab to adjuvant chemotherapy results in improved survival in women with human epidermal growth factor receptor 2-positive, early, potentially curable breast cancer. The standard duration of therapy is 12 months of adjuvant treatment. However, this is expensive and has side effects including cardiac toxicity. The PERSEPHONE trial hypothesized that a shorter duration of trastuzumab therapy would be noninferior in terms of survival and offer benefits of being cheaper and having fewer adverse events.[2]

c. Goyal *et al.* showed that azithromycin is noninferior to amoxicillin-clavulanic acid for treatment of respiratory exacerbations in children with bronchiectasis with the convenience of once-a-day dosing.[3]

Each of these examples involves a trade-off between the risk of (marginally) decreased efficacy and another benefit, which makes the experimental treatment an acceptable alternative to the standard.

How to cite this article: Ranganathan P, Pramesh CS, Aggarwal R. Non-inferiority trials. Perspect Clin Res 2022;13:54-7.

Another use of the NI design can be when one product is available (making the use of placebo as a comparator ethically challenging), but more products are needed. For example, with COVID-19 vaccines – we have vaccines but need more vaccines and hence resort to showing that a newer vaccine is "noninferior" to an existing vaccine.

FORMULATING THE HYPOTHESES FOR NON-INFERIORITY STUDIES

Every research study starts with a baseline assumption (the null hypothesis) and a contradictory alternative hypothesis.[4] In a traditional superiority study, one starts with the null hypothesis that there is no difference between treatments and then tries to prove that there is a difference (by disproving the null hypothesis and accepting the alternate hypothesis). For example, the DREAMS study compared dexamethasone with standard treatment for postoperative nausea and vomiting after gastrointestinal surgery.[5] The primary outcome was the proportion of patients with vomiting within 24 h after surgery. The null hypothesis was that the proportion of patients with vomiting in the dexamethasone group would be equal to that in the placebo group. The alternative hypothesis was that the proportion of patients with vomiting in the dexamethasone group would be different from that the placebo group. The alternative hypothesis does not specify whether the experimental treatment is better or worse than the control – this is known as a two-sided hypothesis and is analyzed using two-tailed tests. The objective of the study is to reject the null hypothesis and accept the alternative hypothesis, i.e., prove that the treatments are dissimilar.

On the other hand, in a NI study, the null hypothesis is that the experimental treatment will be inferior to the standard by a margin greater than a predefined value (this margin is known as the margin of NI or delta and is explained later in this article). The alternative hypothesis states that the experimental treatment will, at worst, be only marginally inferior to standard treatment, i. e. by a margin not exceeding delta. The objective of the study is to reject the null hypothesis and accept the alternative hypothesis, and thus establish that the experimental treatment is noninferior to the standard. This is an example of a one-sided hypothesis, which means that we are only interested in testing for inferiority or its absence (and not in whether the experimental treatment is superior to the standard). In this case, the data are analyzed using a one-tailed test.

THE CONFIDENCE INTERVAL APPROACH

In a previous article, we have discussed the concept of confidence intervals (CIs).[6] In brief, while a study gives us one observed value for a result, CIs provide an estimate of the possible range of values for that result in the population. In the DREAMS study, the incidence of postoperative vomiting in the first 24 h after surgery was 25.5% in the dexamethasone arm versus 33.2% in the standard care arm (risk ratio: 0.77).[5] Therefore, in this study, dexamethasone reduced the risk of vomiting by 0.23 (1.0 − 0.77) folds or by 23%. The 95% CI for this risk ratio ranged from 0.65 to 0.92; this means that we are 95% confident that dexamethasone is superior to standard care, though the real effect size in the population could vary from 8% benefit (1.0 − 0.92) to 35% benefit (1.0 − 0.65). Since this was planned as a superiority study (with two-sided alternative hypothesis), we try to find out both the minimum and the maximum possible effects of the experimental treatment, and the direction of the effect; therefore, we calculate two-sided CIs for the difference. The value of 95% for the CI arises from the type 1 error or alpha value set at the beginning of the study – allowing an error of 5%, we need to be 95% certain that any difference between treatments which we find at the end of the study is a true difference and has not occurred by chance.[4]

In a NI trial, the focus is on the worst possible outcome with the experimental treatment. With a type 1 error of 5%, we want to be 95% certain that even in the worst case, the experimental treatment does not differ from the standard by more than the predefined value of delta. Therefore, we calculate a one-sided 95% CI to determine the maximum difference that might be seen in the population. If the new treatment is to be considered noninferior, then the lower limit of the 95% CI should lie within the margin of NI. Here, we are not concerned about the least difference between the two treatments or about whether the new treatment is in fact superior to the standard.

It is not essential to use 95% CI and, as for other types of studies, one could use a different confidence level cutoff. In fact, the US FDA mandates that such studies use a 97.5% CI cutoff. This is in keeping with the fact that this is a one-sided CI and the traditional 5% error permitted with a two-sided hypothesis is likely to be equally distributed on the two sides.

Figure 1 shows the various possible results of a study and the interpretation.

ESTABLISHING THE MARGIN OF NON-INFERIORITY OR EQUIVALENCE

The validity of a NI trial hinges around the margin of NI (known as delta). The delta represents the largest loss
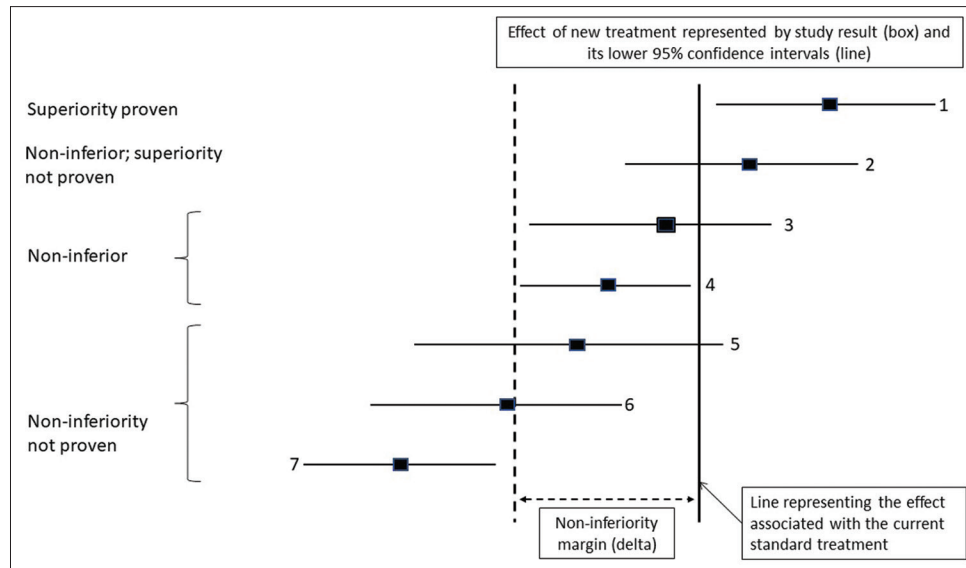
Figure 1: Interpretation of the results of a non-inferiority study. The vertical solid line represents the observed effect of the current standard, and the vertical dotted line is located at a distance of the predetermined inferiority margin (or delta) below it. For non-inferiority to be established, it is expected that the lower bound of the 95% confidence intervals (CIs) of the observed effect of the new treatment is not below the dotted line. Hence, for the lower three lines (lines 5–7), non-inferiority is not established. For the next three lines (lines 2–4), non-inferiority is established. The result represented by line 1 establishes not only non-inferiority but also superiority (since the lower bound of the 95% CIs of the effect of the new intervention exceeds the observed effect of the comparator or the current standard intervention)

of effect that would be considered acceptable in practice. There are no clear guidelines on how to choose delta, and it is largely a matter of clinical judgment. Typically, if the standard treatment has an effect size "x" over placebo, then the delta for a NI study has to be a small proportion of "x" so that the experimental treatment remains noninferior to the standard and is definitely better than placebo. Since sample size is inversely proportional to the delta, a very small delta will result in larger sample sizes; however, using a large delta to counter this defeats the assumption of NI, as the difference then becomes clinically important.

In the PERSEPHONE study, designed to assess NI of the experimental group (6 months of trastuzumab), the clinically acceptable NI was defined as the 4-year disease-free survival being not worse by an absolute value of 3% than that of the standard group (12 months of trastuzumab), which was estimated to be 80%.[2] This 3% NI margin was decided before the start of the trial based on consensus from the trial development group that included patient and public involvement groups.

INTENTION-TO-TREAT VERSUS PER-PROTOCOL ANALYSIS

In a previous article in this series, we have discussed the differences between intention-to-treat (ITT) and per-protocol (PP) analyses.[7] ITT analysis includes all patients irrespective of whether they received the treatment they were randomized to get; ITT provides an estimate

of the real-life effectiveness of the intervention. On the other hand, PP analysis includes only those patients who strictly adhered to the protocol and gives an estimate of the efficacy of the intervention in an artificial setting where all the participants adhere to and complete the allocated treatment, as planned.

For superiority trials, ITT analysis is the preferred method of analysis since PP analysis tends to overestimate the treatment effect, which may not reflect the effect likely to be seen in clinical practice. On the other hand, in NI trials, we are interested in determining the maximum possible difference between the experimental treatment and the comparator to rule out inferiority; here, if there is poor patient compliance to the experimental treatment, an ITT analysis could dilute the difference between treatments and make an inferior treatment appear to be noninferior. Thus, when analyzing a NI trial, PP analysis is the key analysis; thus, both ITT and PP analyses should be conducted and both approaches should show NI for a conclusive opinion.

SWITCHING BETWEEN SUPERIORITY AND NON-INFERIORITY

Often, if a superiority trial shows no significant difference, one is tempted to conclude that there is no difference between the two groups, and that they are similar. In a previous article, we have addressed the issue of how no evidence of effect is not evidence of no effect.[8] Since the rationale, hypothesis, and margin of difference in a NI trial

are completely different from a superiority trial, one cannot conclude NI based on a negative superiority trial.

On the other hand, if both the lower and upper limits of the CI of the result of a NI study lie above the line of no difference, then one can conclude superiority. Readers may refer to an article by Ganju for further details regarding this.[9]

REPORTING OF NON-INFERIORITY TRIALS

The Consolidated Standards of Reporting Trials (CONSORT) initiative was launched in 2001 to overcome problems arising from inadequate reporting of randomized controlled trials.[10] A separate extension specific to NI trials was added in 2006 and updated in 2012, to improve the quality of reporting of NI trials and to help readers and reviewers to assess the validity of trial results.[11] Researchers conducting NI trials or readers critically appraising NI trials are encouraged to go through the checklist for NI trials in the CONSORT extension statement.

Financial support and sponsorship
Nil.

Conflicts of interest
There are no conflicts of interest.

REFERENCES

1. Dhingra U, Kisenge R, Sudfeld CR, Dhingra P, Somji S, Dutta A, *et al*. Lower-dose zinc for childhood diarrhea – A randomized, multicenter trial. N Engl J Med 2020;383:1231-41.
2. Earl H, Hiller L, Vallier AL, Loi S, McAdam K, Hughes-Davies L, *et al*. Six versus 12 months' adjuvant trastuzumab in patients with HER2-positive early breast cancer: The PERSEPHONE non-inferiority RCT. Health Technol Assess 2020;24:1-190.
3. Goyal V, Grimwood K, Byrnes CA, Morris PS, Masters IB, Ware RS, *et al*. Amoxicillin-clavulanate versus azithromycin for respiratory exacerbations in children with bronchiectasis (BEST-2): A multicentre, double-blind, non-inferiority, randomised controlled trial. Lancet 2018;392:1197-206.
4. Ranganathan P, Cs P. An introduction to statistics: Understanding hypothesis testing and statistical errors. Indian J Crit Care Med 2019;23:S230-1.
5. DREAMS Trial Collaborators and West Midlands Research Collaborative. Dexamethasone versus standard treatment for postoperative nausea and vomiting in gastrointestinal surgery: Randomised controlled trial (DREAMS Trial). BMJ 2017;357:j1455.
6. Ranganathan P, Pramesh CS, Buyse M. Common pitfalls in statistical analysis: "P" values, statistical significance and confidence intervals. Perspect Clin Res 2015;6:116-7.
7. Ranganathan P, Pramesh CS, Aggarwal R. Common pitfalls in statistical analysis: Intention-to-treat versus per-protocol analysis. Perspect Clin Res 2016;7:144-6.
8. Ranganathan P, Pramesh CS, Buyse M. Common pitfalls in statistical analysis: "No evidence of effect" versus "evidence of no effect". Perspect Clin Res 2015;6:62-3.
9. Ganju J, Rom D. Non-inferiority versus superiority drug claims: The (not so) subtle distinction. Trials 2017;18:278.
10. Schulz KF, Altman DG, Moher D; CONSORT Group. CONSORT 2010 statement: Updated guidelines for reporting parallel group randomised trials. BMC Med 2010;8:18.
11. Piaggio G, Elbourne DR, Pocock SJ, Evans SJ, Altman DG; CONSORT Group. Reporting of non-inferiority and equivalence randomized trials: Extension of the CONSORT 2010 statement. JAMA 2012;308:2594-604.