Non-inferiority trials: Why, what and how?

Mara Therese P. Evangelista, MD and Nicholas Earl Huber

Why use non-inferiority trials?

Most clinical trials are designed to show that an experimental intervention, such as a drug or procedure, is superior to an existing intervention (or to no intervention, if none is currently available). These designs are called superiority trials. However, sometimes an experimental intervention is not expected to be clinically superior to an existing product, but has other advantages, such as being cheaper, easier to administer, or has a better safety profile. This latter concept is important in dermatology because some drugs that are considered the standard of care may cause unwanted side effects (e.g. steroids), or are perhaps too expensive or are non-sustainable for some our patients.

There has been a large interest in herbal medications in the Philippines because of the availability of raw materials and their lower cost compared to existing drugs. Studies on herbal medications are often done using the superiority design. However, we believe that this is not always the best design in this setting. It is very important to understand that if you do not find a significant difference between treatments in superiority trials, you cannot conclude that the treatments are similar (or equivalent), you can only conclude that they are not different. For example, if you try to compare a herbal medication A to a topical drug B using a superiority trial and your statistical analysis shows that there is no significant difference between the two treatments, you will not be able to conclude that the treatments are similar, i.e. that the effect of herbal lotion A is the same as topical drug B; you can only conclude that herbal medication A is not different from topical drug B. To understand this further, let us discuss the concept of hypothesis testing.

Hypothesis testing: A review:

A statistical hypothesis is an assumption about a population parameter (e.g. difference in proportions between patients cleared, difference in means in post-treatment SCORAD, risk ratios of developing adverse effects). This assumption may or may not be true. When conducting a hypothesis test, we consider two propositions: the null hypothesis, and the alternative. The alternative hypothesis is what we hope our data will prove. The null hypothesis, in contrast, is presumed to be true, until the data provide sufficient evidence that it is not. A similar idea underlies the U.S. criminal justice system ("innocent until proven guilty"): In the statistical world, the null hypothesis is taken to be true until the alternative is proven true. What is essential here is the null hypothesis is never proven true; you simply fail to reject it. Therefore, any hypothesis test has only two possible outcomes:

1. **Reject the null hypothesis** (if p-value < than alpha, usually set at 0.05) and conclude that the alternative hypothesis is true.
2. **Fail to reject the null hypothesis** (if p-value > alpha) and conclude that there is not enough evidence to suggest the null hypothesis is false.

In superiority trials, the null hypothesis is that there is no difference between treatments, while the alternative hypothesis is that the treatments are different (or that one treatment is superior). Therefore, your hypothesis test has one of two outcomes:

1. **Reject the null hypothesis of no difference** if the p-value is less than alpha (i.e. $p < 0.05$) and conclude that the treatments are significantly different.
2. **Fail to reject the null hypothesis of no difference** if the p-value is greater than alpha (i.e. $p > 0.05$) and conclude that the treatments are not significantly different.

As you can see, even if we fail to reject the null hypothesis, it does not mean the null hypothesis is true. In simpler terms, if **we have no evidence of a difference** between treatments, this should **not be considered the same as we have evidence of no difference between interventions** or, similarly, **we would not have evidence that they have the same effect**. You will need a different study design to investigate if they are indeed similar in terms of clinical effect. Going back to our studies on herbal medications: to begin with, it is already unlikely that a herbal medication is better than a drug, so a superiority study will probably not show a significant difference between treatments. However, the herbal medication may be "at least as good" as the drug, and this may be missed by a superiority study.

What is the goal of a non-inferiority trial?

In a non-inferiority study, the goal is to demonstrate that the experimental intervention has an effect sufficiently close to the effect of an active control (ideally the current gold standard). There is no placebo arm in the study, instead the gold standard plays the role of the control. The goal of the study is to show that the effect of the experimental intervention is not inferior to the effect of the active control by a specified amount, called the non-inferiority margin, or delta (discussed further below). The null and alternative hypotheses correspond to a null hypothesis of inferiority and an alternative hypothesis of non-inferiority, i.e. as follows:

- Null hypothesis: Experimental intervention is inferior to the active control by delta or more
- Alternative hypothesis: Experimental intervention is inferior to the active control by less than delta

Therefore, while a superiority study might not be able to reject the null hypothesis of no difference between treatments (and consequently might not put further interest in the experimental intervention), a non-inferiority study could also rigorously show that a herbal treatment is truly non-inferior, which could in turn impact clinical practice. It is worth noting that a non-inferiority study design can still show that the herbal medication is superior to the active control, even though it is called a non-inferiority trial.

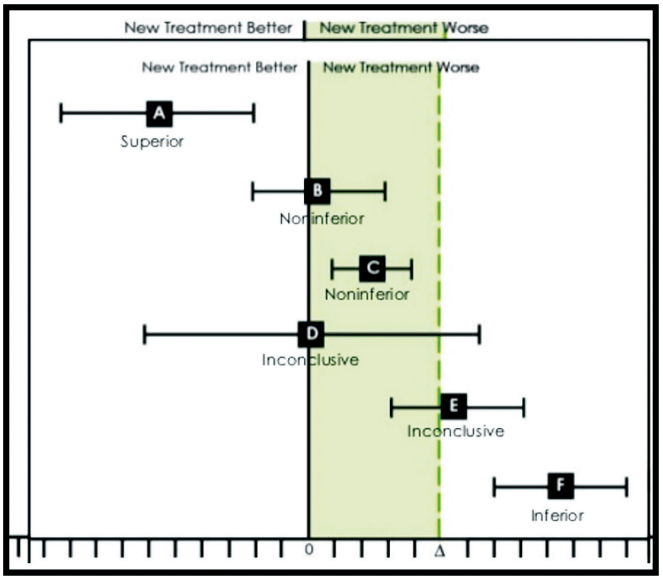
How do you establish non-inferiority?

When computing for the sample size of a clinical trial, you will have to specify the difference between the treatments that you are hoping to detect (e.g. you expect 90% of patients to have complete clearance with treatment A, while 60% will improve with treatment B: the treatment difference is 90 – 60 = 30). The larger the treatment difference you want to detect, the smaller the sample size. If you are comparing an experimental intervention to a placebo (which expectedly has no effect) in a superiority trial, you will assume that the treatment difference is large. However, if you are comparing two interventions you assume to have similar effects on the outcome, the treatment difference will expectedly be small (e.g. you expect 90% of patients to have complete clearance with treatment A, while 85% will improve with treatment B: the treatment difference is 90 – 85= 5), and this will need a much larger sample size.

Establishing non-inferiority means that the clinical performance of the experimental intervention is, at worst, only marginally inferior to that of the active control. 'Marginally' is defined in advance as a detriment of clinical unimportance. This margin is the most important decision in determining the sample size of a non-inferiority trial and has a central role in concluding non-inferiority. As previously mentioned, this difference needs to be set at a realistically small level (because we want to show that the interventions are similar in terms of effect i.e. a small treatment difference, as close to zero as possible), and because of this non-inferiority trials often need to be very large. So how do we set this "clinically unimportant margin?" We do so by pre-defining the smallest level of inferiority (delta) of the experimental treatment which if true is a clinically acceptable difference. A general rule of thumb is that delta must be considerably smaller (1/2 or 1/3) than the minimal clinical difference in a superiority trial. For example, you want to compare herbal medication M to antifungal drug O in patients with tinea corporis using a non-inferiority design. Previous studies comparing drug O to placebo in tinea corporis showed that 75% of patients had complete clearance with antifungal drug O, while 5% had resolution with placebo; treatment difference 75 – 5 = 70%). Therefore, we can assume delta to be 1/2 or 1/3 of the treatment difference of the referenced superiority trial, and the delta for your non-inferiority trial can be 35% or 23%.

The details of sample size calculation and statistical analysis for non-inferiority trials will not be discussed in this article. However, it is good to know that non-inferiority is established using the confidence interval approach. Once your measures of treatment effect have been established (e.g. mean difference, difference in proportions, relative risks, etc), your statistician will also give you confidence intervals for these treatment effects. Essentially, these are the guidelines for interpreting confidence intervals (CIs) for non-inferiority trials:

- When entire CI < 0 = superior
- When upper limit of CI < delta = treatment non-inferior
- When upper limit of CI is > delta = inconclusive
- When entire CI > delta = clearly inferior



Examples:

1) Delta: 10%
   Treatment difference: 0.3%
   Confidence interval: -4.6 – 5.2% (upper limit of CI [5.2] < delta)
   Non-inferior? YES

2) Delta: 10%
   Treatment difference: 22%
   Confidence interval: 13 – 25% (entire CI > delta)
   Non-inferior? NO, Inferior

3) Delta: 5%
   Treatment difference: -2.3%
   Confidence interval: -3.2 to – 1.2%
   Non-inferior? NO, Superior(entire CI < 0)

Non-inferiority trials also have particular characteristics which make emphasis on rigorous methods even more important than in trials designed to show superiority. The methodology of the non-inferiority trial must be as similar as possible to the referenced superiority trial (type of patients, regimen, outcome measures, follow-up period). In the previous example where we want to compare herbal medication M to antifungal drug O in patients with tinea corporis using a non-inferiority design, the methods should be as similar as possible to previous studies comparing drug O to placebo in tinea corporis.

In a trial designed to show superiority, events such as protocol deviations (e.g. non-compliant patients) and patients being lost to follow-up tend to reduce the power of the study by making the treatments more similar. These trials often use intention-to-treat (ITT) analyses because this analysis is the more conservative approach (more likely to conclude the null hypothesis of no difference between treatments). However, the point of a non-inferiority trial is to show that the treatments are similar, so protocol deviations and patients lost to follow up will make it easier to conclude the experimental intervention is similar (not inferior to) to the active control, and an ITT analysis is not as desirable. Therefore, non-inferiority trials must employ and report both ITT and per protocol analysis (complete case analysis that excludes dropouts and protocol violtators), and see if they agree with each other. A conservative approach would be to base conclusions on the more pessimistic of the analyses.

Summary:

- Choice of control in a non-inferiority trial: must be current standard of care and benefit clearly established against placebo in previous superirity trials
- When to use non-inferiority trials:
  - Presence of established, efficacious treatment
  - New treatment matches efficacy (and unlikely to have higher efficacy), and may have other advantages:
    - Fewer side effects
    - Easier to administer
    - Cheaper
    - Profit driven (branded vs generic)
    - Greater adherence
- Disadvantages of a non-inferiority trial:
  - Sample size much larger than a superiority trial
  - Rigorous methods much more important than a superiority trial
- Characteristics of a good non-inferiority trial:
  - Standard trial procedures (randomization, blinding) with methods similar to the referenced superiority trial
  - Active comparator that is well-established and backed by at least one superiority trial
  - A pre-stated non-inferiority margin that is statistically AND clinically justified
  - Both ITT and per protocol analyses used and reported
  - Results reported as CIs
  - Conclusion drawn from comparison of CI with non-inferiority margin

Table 1. A summary of differences between superiority and non-inferiority trials.

	SUPERIORITY TRIAL	NON-INFERIORITY
Objective	Show that experimental intervention is superior	Show that one treatment is at least as good as the alternative in respect of patient response
Null hypothesis	No difference between the arms	Treatment is worse than control by more than the pre-set margin
Alternative hypothesis	There is a difference between the arms	Treatment is non-inferior to control based on a pre-set margin
Sample size	Smaller (you expect bigger treatment differences)	Larger (you expect smaller treatment differences)
Primary analysis	Intention to treat analysis	Both per protocol and intention to treat analyses

About the authors:

Dr. Mara Therese P. Evangelista graduated from the University of Santo Tomas Faculty of Medicine and Surgery, *cum laude*. She took her residency in Dermatology at the Jose R. Reyes Medical Center where she was Chief Resident in her senior year. She was the Board Topnotcher of the Philippine Dermatological Society's Board Examination last 2014. She is is multi-awarded both internationally and locally for her excellent research work in the field of Dermatology. She is the first Asian and Filipino to win the Everett Fox Award for Outstanding Clinical Research in the American Academy of Dermatology's Annual Meeting in Denver, Colorado. She has also won first place in the Advanced Residency Training and Education Research Poster Contest in Orlando, Florida. She has presented her research works at multiple international conferences, including the American Academy of Dermatology, World Congress of Dermatology, American Contact Dermatitis Society and International Society of Dermatopathology. Dr. Evangelista is currently on her second year of her Masters in Clinical Research in the University of London. She completed her 1st year with distinction, top of her class. She is also an internationally board-certified dermatopathologist (International Committee for Dermatopathology Board Certifying Examination, Frankfurt, Germany) and had her Dermatopathology Visiting Fellowship at the University of California San Francisco.

Mr. Nicholas Huber studied Economics and Statistics at Harvard College, where he graduated *summa cum laude*, and holds a master's degree in Engineering for Sustainable Development from Cambridge University, where his dissertation earned the highest marks possible and won the course prize for best scientific poster. Mr. Huber is currently a data analytics consultant to the executive team of PLDT, where he is bringing advanced data tools and data-driven thinking to the group of companies. Before that, he was a data scientist at Airbnb in San Francisco, where he helped develop and analyze automated pricing tools for hosts. He is passionate about data visualization and communication, and his blog posts on learning to program using free, online resources have been read by hundreds of thousands of people, all around the world. ■

* *You may contact the authors at:*
mevangelistamd@gmail.com
nicholas.e.huber@gmail.com