

Non-inferiority trials

Shengping Yang PhD, Gilbert Berdine MD

The current treatment of a childhood cancer requires radiation therapy. We are planning to test an experimental treatment protocol that omits radiation therapy and expect the experimental protocol to have similar efficacy. What is the best design for conducting such a study?

With continuous advances in healthcare, substantial improvements in clinical outcome are commonly seen in many diseases. For example, childhood acute lymphoblastic leukemia, which was considered one of the most fatal childhood cancers 50 years ago, has a cure rate approaching 90% in many developed countries, including the US. With such a high cure rate, the benefit of a newly developed treatment is expected to be only marginal. As a result, it might not be feasible to conduct a superiority trial for an experimental treatment because the size of such a study might be unreasonable.

A non-inferiority trial, on the other hand, tests whether an experimental treatment is 'at least as effective as' or 'at worst not much less effective than' an active control treatment. Very often, it is anticipated that the experimental treatment offers ancillary benefits, such as improved safety, lower cost, better tolerability, or more convenience in administration. As we know, radiation therapy might cause long term side effects, especially for children, and thus, the non-radiation therapy treatment protocol provides a great ancillary benefit by completely omitting radiation therapy. Therefore, to evaluate such an experimental treatment, a non-inferiority trial becomes an ideal choice.

On other occasions, non-inferiority trials can be used to show efficacy of an experimental treatment when it is unethical to include a placebo control in the trial. Specifically, the experimental treatment is compared with an active control, which is known from past superiority trials to be effective, and if the difference between the two treatments is sufficiently small, then under certain assumptions, the trial can support the conclusion that the experimental treatment is effective.

On the other hand, although non-inferiority trials are becoming increasingly popular, they have serious issues, mostly weaknesses compared with superiority trials that need to be addressed.

THE NON-INFERIORITY HYPOTHESIS

The null hypothesis of a non-inferiority trial is that the experimental treatment is inferior to the active control in terms of the outcome; the alternative hypothesis is that the experimental treatment is not inferior. Although meaningful, these hypotheses are associated with problems. If a non-inferiority trial is poorly executed – serious protocol violations, excessive attrition, inadequate sample size – then it is very likely that the difference between treatments diminishes, and there will not be enough evidence to reject the null hypothesis. As a result, the experimental treatment can be declared as non-inferior. However, such a conclusion can be a reflection of poor quality of trial design and/or implementation, rather than a true non-inferior experimental treatment.

Also associated with the null and alternative hypothesis is the protection from bias in a trial. Blinding is the most used technique to avoid bias. Blinding is less effective in a non-inferiority trial because it is more difficult to prevent a conscious bias toward

Corresponding author: Shengping Yang
Contact Information: Shengping.yang@ttuhsc.edu
DOI: 10.12746/swrccc.v5i21.424

equivalence than nonequivalence, especially for subjective end points.

NON-INFERIORITY MARGIN

As a component of the non-inferiority hypothesis, the non-inferiority margin is defined as the difference that is clinically acceptable to conclude that there is no difference in practice between treatments. It directly affects sample size/power calculation, assay sensitivity (see below), and eventually the success of a non-inferiority trial. However, there is no consensus on what is the best method for specifying a non-inferiority margin.

One approach for specifying a non-inferiority margin is based on the minimal difference in terms of clinical significance. However, this approach is arbitrary, and clinical significance is disease and outcome specific.

The alternative is to choose such a margin on the basis of the effect of the active control in historical placebo-controlled trials. In general, the margin should not be greater than the smallest response that the active control would be reliably expected in the planned trial compared to a placebo. Otherwise, we bear the risk of declaring that the experimental treatment is non-inferior to an active control, even if it has no effect at all compared to placebo. Specifically, supposing that the effect of the active control is C and the effect of the experimental treatment is T , then the null and alternative hypotheses are:

$$H_{null}: C - T \geq M$$

$$H_{alternative}: C - T < M$$

where M is non-inferiority margin.

M can be set equal to the difference (or the lower bound of the 95% CI to account for uncertainty) between the active control and the placebo M_{AC} with respect to the outcome. However, due to differences in a study objective, M is often set so that only a fraction (f) of the efficacy of the active control M_{AC} is preserved, i.e., $M = (1 - f)M_{AC}$. It is obvious that the greater the efficacy fraction, the smaller the margin, and the larger the required sample size of a trial. According to the FDA guideline, a 50% preserved efficacy would be a meaningful choice.

Defining an appropriate non-inferiority margin is difficult and is subject to the objective of the trial (whether to provide evidence of efficacy of the experimental treatment, or to make comparative effectiveness evaluation), the nature of a disease, the seriousness of clinical outcome, the magnitude of active control efficacy, and the safety and cost of active control and experimental treatments.

THE CONSTANCY ASSUMPTION

As mentioned above, to specify the non-inferiority margin M , the efficacy of the active control M_{AC} has to be determined. However, due to ethical reasons, non-inferiority trials in general do not have placebo controls. Thus the effect of active control cannot be directly evaluated in the trial but has to be assumed from past trials external to the current trial. A constancy assumption thus needs to be made that the effect of the active control is assumed to be the same in the current trial as in the past superiority trials. This inevitably requires that many aspects of the current trial be the same as the past trials, and it is important to adhere to the treatment protocol more strictly in a non-inferiority trial than in a superiority trial.

ASSAY SENSITIVITY

Assay sensitivity is defined as the ability of a clinical trial to distinguish an effective treatment from a less effective or ineffective treatment. Without assay sensitivity, a trial is not capable of comparing the efficacy of two treatments, and thus non-inferiority is virtually automatically established, which is completely undesirable.

Demonstration of assay sensitivity is straightforward in a superiority trial. If a superiority trial shows a difference in efficacy, it automatically demonstrates assay sensitivity by definition. However, a non-inferiority trial is designed to rule out that there is any difference between treatments by a margin. Therefore, even it shows non-inferiority of the experimental treatment, it would not distinguish between whether the experimental treatment is truly non-inferior to the active control, or the trial

lacks assay sensitivity to detect a difference. As a result, without demonstrating assay sensitivity, a non-inferiority trial might lead to an erroneous conclusion of efficacy.

In order to have assay sensitivity, a non-inferiority trial must be designed the same (or very similar to) as the past trials which demonstrated the active control efficacy to ensure the constancy assumption is valid. In addition, assay sensitivity also depends on non-inferiority margin, and very often an investigator tends to specify a larger margin to reduce the sample size of a trial, which may reduce or eliminate assay sensitivity. (Power/sample size calculation is specific to the type of outcome of interest, and will not be discussed in this article.)

In summary, a non-inferiority trial is typically less reliable than a superiority trial, and has to be conducted with caution due to its inherent weaknesses, including arbitrariness in specifying non-inferiority margin, potentially invalid assumptions, and difficulties in controlling biases.

From: The Department of Pathology (Yang) and Internal Medicine (Berdine) at Texas Tech University Health Sciences Center, Lubbock, TX

Submitted: 10/4/2017

Conflicts of interest: none

REFERENCES

1. Pui CH, Evans WE. A 50-year journey to cure childhood acute lymphoblastic leukemia. *Semin Hematol* 2013 Jul; 50(3):185–196.
2. Food and Drug Administration. Non-inferiority clinical trials to establish effectiveness: guidance for industry 2016. Accessed Oct. 6, 2017.
3. Kaul S, Diamond GA. Good enough: A primer on the analysis and interpretation of noninferiority trials. *Annals Int Med* 2006;145:62–69.
4. Hahn S. Understanding noninferiority trials. *Korean J Pediatr* 2012;55(11):403–7.
5. Snapinn SM. Noninferiority trials. *Curr Control Trials Cardiovasc Med* 2000;1(1):19–21.