Usage of Non-inferiority Design : Points to Consider from Regulatory Agency and Pharmaceutical Companies

Department of Applied Statistics

Yonsei University

Seung-Ho Kang

1. The Goal of Non-Inferiority Trials

The goal of the non-inferiority trial is to show that the effect of the test drug (T) is not inferior to the effect of the active control (C) by a amount of NI margin M (>0).

When a larger response indicates better outcome,

- $H_0 : \mu_T \leq \mu_C - M$ vs $H_1 : \mu_T > \mu_C - M$ . . . . (9.1)

A difference of less than M is considered clinically meaningless.

There are two different goals of non-inferiority trials.

(1) to show that the effect of the test drug (T) is not inferior to the effect of the active control (C).

- That is, the test drug is not worse than active control in terms of the primary endpoint (almost identical), but in other aspects, such as safety, test drugs have advantages over active controls.

  (a) purpose of medical research

  (b) phase IV clinical trials

. This goal is irrelevant to obtaining a marketing authorization for the test drug from the regulatory agencies.

. Many people misunderstand that this is the only goal of non-inferiority clinical trials.

. The word "non-inferiority" originally came from the first purpose.

(2) to get a marketing authorization for the test drug from the regulatory agency

- The criteria for efficacy of the test drug is, in principle, to prove superiority over placebo. Why?

- But, when there is an effective treatment that provides an important benefit (e.g., life-saving or preventing irreversible injury) available to patients for the condition to be studied in the trial, the use of placebo is not ethically acceptable.

Although we test the hypothesis (9.1) by comparing the test drug and active control in non-inferiority trials, the ultimate goal is still to show that the test drug is superior to placebo.

$$H_0 : \mu_T \leq \mu_{P,put} \qquad H_1 : \mu_T > \mu_{P,put}$$

where $\mu_{P,put}$: putative effect of placebo in non-inferiority trial.

Strictly speaking, this hypothesis is statistically impossible to test.

Therefore, we need additional assumptions and data such as historical trials, assay sensitivity, constancy assumption.

The word "non-inferiority" originally came from the first purpose.

But, "non-inferiority" is used to show that the test drug is superior to placebo in the second purpose.

This has confused many people.

If we show that the test drug is superior to active control rather than showing non-inferiority, of course, we can obtain marketing authorization for the test drug.

But, it is very difficult in most cases, because the efficacy of active control is quite good.

How can we show that the test drug is superior to placebo by using active control when there is no placebo arm?

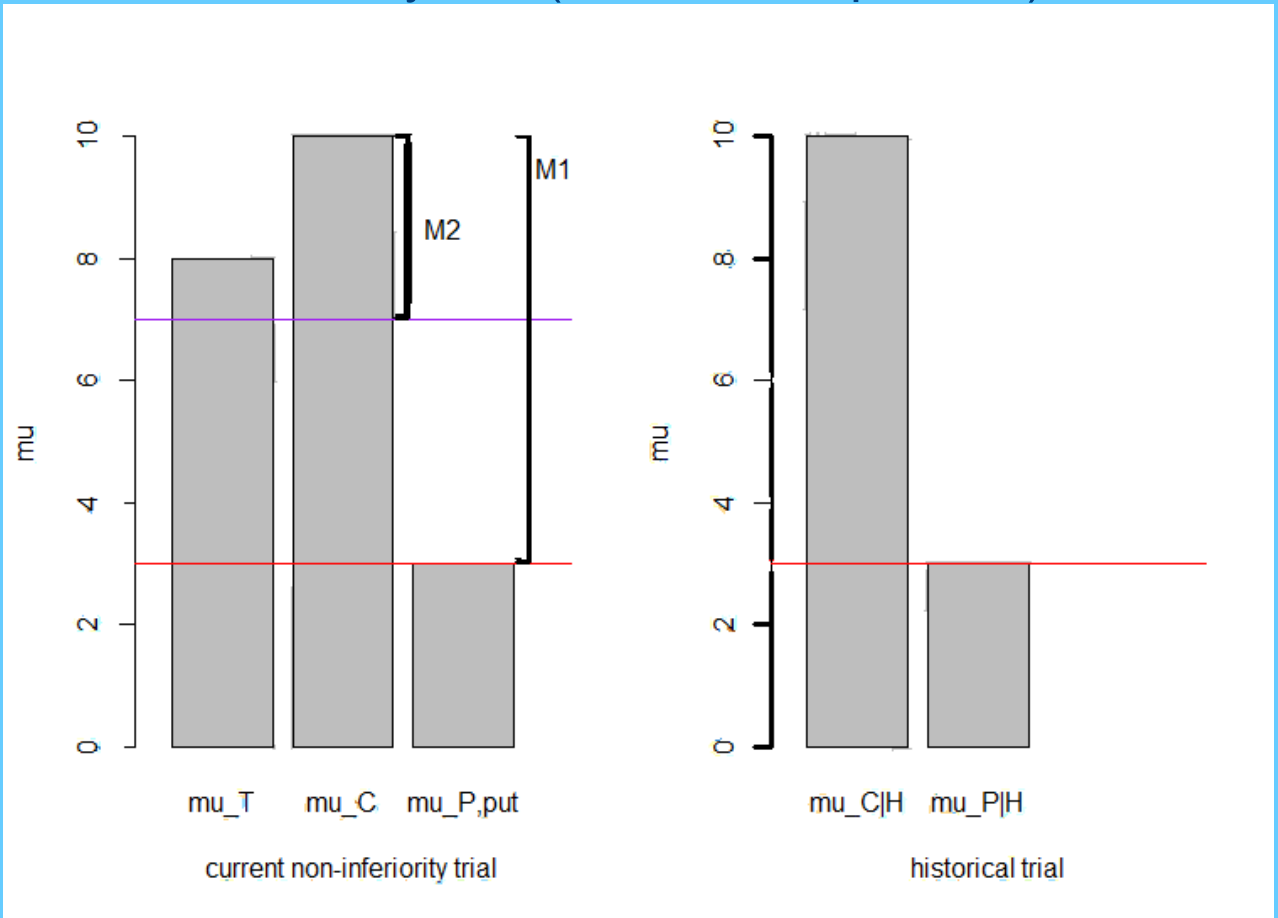
KEY IDEA

Active treatments have already been shown to be superior to placebo in historical clinical trials.

Therefore, in the current non-inferiority trial, the test drug may be claimed to be superior to placebo if the test drug is non-inferior (comparable to) the active comparator.

Indirect comparison!!

Key idea of non-inferiority trial (indirect comparison)



However, in order to establish the above method, we need the assumption that, if there is a placebo arm in the current non-inferiority trial, the active control would be still superior to placebo, because the active control was shown to be superior to placebo in historical trials.

In other words, the efficacy of the active control should be reproducible in current non-inferiority trials.

You may think that this is too obvious assumption.

But, it may not be….

2. The efficacy of the active control may NOT be reproducible in current non-inferiority trials, although it was proven in historical trials.

Example) For effective anti-depressant drugs which was shown to be superior to placebo in historical trials, approximately 50% of all placebo-controlled anti-depressant trials fail to show that those anti-depressant drugs are superior to placebo (US FDA 2016).

This is because current non-inferiority trials and historical clinical trials differ in the following aspects:

1) Demographic factors such as age and gender
2) Social economic level
3) Severity of disease
4) Dosage
5) Concomitant drugs
6) Compliance of protocol

- Clinical Trials are not mathematical proof!!

3. Assay Sensitivity

- If we try to demonstrate that the test drug is more effective than placebo through non-inferiority trial without placebo arm, the efficacy of active control which was proven in historical trial should be reproducible in current non-inferiority trial.

⟹ **Assay Sensitivity**

- Definition of Assay Sensitivity

  Assay sensitivity is a property of a clinical trial defined as the ability to distinguish an effective treatment from an ineffective treatment.

- Meaning of assay sensitivity in non-inferiority trial

- Active control has already shown a benefit of M1 compared to placebo in historical trials. If the current non-inferiority trial had included placebo arm (of course, in reality there is no placebo arm due to ethical reason), active control would have shown a benefit of compared to placebo.

- However, since there is no placebo arm in current non-inferiority trial, there is no strict way to confirm assay sensitivity.

- If the following three conditions are satisfied, assay sensitivity can be checked indirectly.

(1) HESDE (Historical evidence of sensitivity of drug effects)

HESDE means that prior studies, which were appropriately designed and conducted trials in the past, regularly showed active control to be superior to placebo.

This consistent finding in past studies allow for a reliable estimate of the effect size of active control compared to placebo.

HESDE cannot be determined for many symptomatic treatments (e.g., treatments for depression, anxiety, insomnia, angina, pain).

(2) Similarity between current non-inferiority trials and historical clinical trials.

. The characteristics of the patient population
. Important concomitant treatment
. Definitions of study endpoints
. Dose of active control
. Entry criteria
. Analytic approaches

(3) The conduct of current non-inferiority trial is of high quality.

. Imprecise or poorly implemented entry criteria
. Poor compliance
. Use of concomitant treatments whose effects may overlap with the test drug
. Inadequate measurement techniques
. Errors in delivering assigned treatments
. High attrition

These sloppiness may reduce the observed difference between test drug and active control.

It may potentially lead to a false positive conclusion of non-inferiority.

In summary,

(1) HESDE (Historical evidence of sensitivity of drug effects)

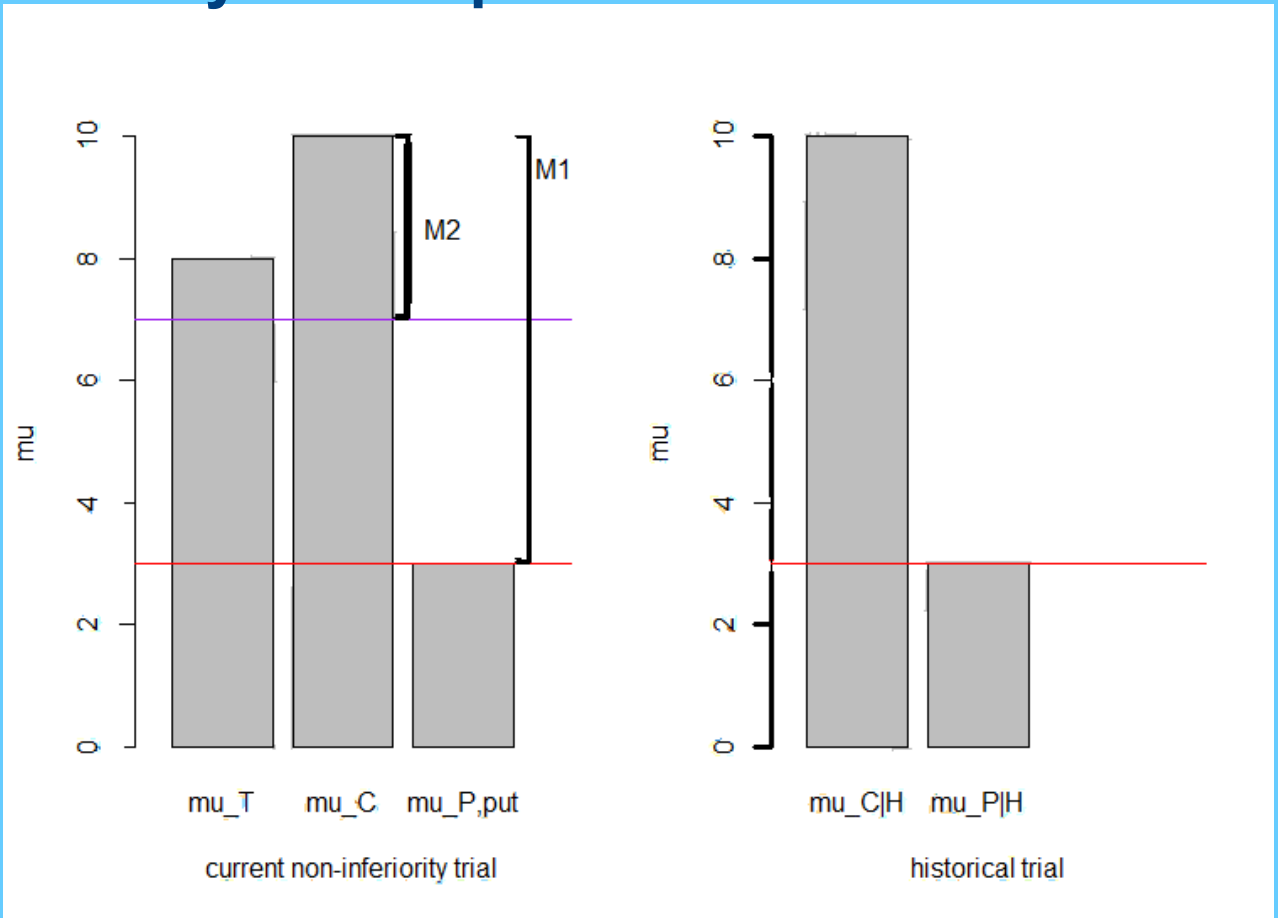
(2) Similarity between current non-inferiority trials and historical clinical trials

(3) The conduct of current non-inferiority trial is of high quality.

If these three conditions are satisfied, we can check assay sensitivity indirectly.

Subjective decision…

4. Constancy Assumption



The following notations are used to denote the effect (population mean) of primary endpoint.

	Current non-inferiority trial	Historical trial
Test drug T	μ_T	
Active control C	μ_C	$\mu_{C\|H}$
placebo P	$\mu_{P,put}$	$\mu_{P\|H}$

※ $\mu_{P,put}$ = putative placebo effect if there is placebo arm in current non-inferiority trial

Key Idea of Analysis of Non-Inferiority Trial

$$\mu_P = \mu_{P,put} = \mu_{P|H}$$

$$\mu_C = \mu_{C|H}$$

} ssumed,

We can simply combine the results of the current non-inferiority trial and historical trial. Indirect comparison!

$$(\mu_T - \mu_C) + (\mu_C - \mu_P) = \mu_T - \mu_P > 0$$

Constancy Assumption can be expressed mathematically as follows.

$$\mu_C - \mu_{P,put} = \mu_{C|H} - \mu_{P|H} \quad \ldots. \quad (9.3)$$

Constancy Assumption is that the efficacy of an active control against placebo in historical trial is the same as that of current non-inferiority trial.

But, Constancy Assumption cannot be checked directly in real practice because there is no placebo arm in current non-inferiority trial.

In general, where there has been substantial evolution over time in disease definition and treatment, or where the methodology used in the historical trials has become outdated, constancy assumption may not be supported.

Subjective decision…

Suppose that we choose non-inferiority margin M with the following condition.

$$\mu_{C|H} - \mu_{P|H} \geq M \qquad \ldots \qquad (9.4)$$

Then we can derive the following conclusion based on (9.1), (9.3) and (9.4).

$$\mu_T - \mu_C > -M \geq -(\mu_{C|H} - \mu_{P|H}) \quad \ldots \text{ from (9.1) and (9.4)}$$

$$\Rightarrow \quad (\mu_T - \mu_C) + (\mu_{C|H} - \mu_{P|H}) > 0$$

$$\Leftrightarrow \quad (\mu_T - \mu_C) + (\mu_C - \mu_{P,put}) > 0 \quad \ldots \text{ from (9.3)}$$

$$\Leftrightarrow \quad \mu_T - \mu_{P,put} > 0$$

5. How to determine the non-inferiority margin

Sponsors prefer wide margins.

VS

Regulatory agencies prefer narrow margins.

FDA guideline recommends the use of the statistical margin M1 and the clinical margin M2.

The Statistical Margin M1 should satisfy

$$\mu_{C|H} - \mu_{P|H} \geq M_1$$

If so, based on constancy assumption, accepting the alternative hypothesis $H_1 : \mu_T - \mu_C > -M$

implies $H_1 : \mu_T > \mu_{P,put}$

How can we determine M1 such that

$$\mu_{C|H} - \mu_{P|H} \geq M_1 \quad ?$$

Because we cannot confirm assay sensitivity and constancy assumption directly, we use conservative approach.

We take M1 as a lower confidence limit of 95% confidence interval of $\mu_{C|H} - \mu_{P|H}$ based on historical trials. (meta-analysis)

Why do we need the clinical margin $(M_2 > 0)$?

Although it is statistically acceptable that the test treatment is inferior to the active control by the amount of M_1 , it may not be clinically acceptable.

The decision of the clinical margin is made with pure clinical judgement that does not involve with statistical considerations. (It could be subjective….)

The non-inferiority margin is taken as the minimum of M1 and M2.

Assay sensitivity and constancy assumption are not verifiable directly.

The clinical margin could be subjective.

Therefore, the determination of the non-inferiority margin can be seen as negotiation between sponsors and regulatory agencies.

After the non-inferiority margin M is determined, we can test the hypothesis (9.1)

$$H_0 : \mu_T \leq \mu_C - M \qquad H_1 : \mu_T > \mu_C - M \qquad (9.1)$$

If the upper limit of 95% confidence interval of $\mu_C - \mu_T$ is smaller than M, we conclude

that the test treatment is non-inferior to the active control

and the test treatment is effective (that is, the efficacy of the test treatment is superior to that of placebo).

6. When non-inferiority trials are infeasible

Although the use of placebo is unethical, there may be some cases where non-inferiority trials are not feasible.

For example,

1.The treatment effect is so small that the sample size required to conduct a non-inferiority trial may not be feasible.

2. There is large study-to-study variability in the treatment effect.

In this case, the treatment effect may not be sufficiently reproducible to allow for the determination of a sufficiently reliable estimate of M1.

3. There is no historical evidence to determine a non-inferiority margin.

4. Medical practice has changed so much that the effect of the active control in historical studies is not clearly relevant to the current study.

7. In a situation where a placebo-controlled trial would be considered unethical, but a non-inferiority trial cannot be performed, what are the options?

(1) When the new drug and established treatment are pharmacologically distinct, an add-on study where the new drug and placebo are each added to the established treatment.

T+A vs P+A

(2) A placebo-controlled trial for non-responders.

(3) A study in patients who cannot tolerate the established effective therapy

(4) A study of a population in which the effect of available therapy is not established

(5) For a drug with dose-related side effects, and where a dose lower than the usual dose would be considered ethical, a dose-response study may be possible.

(6) Early escape design, randomized withdrawal design

7. Other Issues

(1) If there are several active controls, which active control to choose?

Since the result of a non-inferiority trial depends on the active control, the choice of active control is very important.

- ICH E9 says …

Active comparators should be chosen with care.

An example of a suitable active comparator would be a widely used therapy whose efficacy in the relevant indication has been clearly established and quantified in well-designed and well-documented superiority trial(s) and that can be reliably expected to exhibit similar efficacy in the contemplated active control trial.

- ICH E9 says …

To this end, the new trial should have the same important design features (primary variables, the dose of the active comparator, eligibility criteria, and so on) as the previously conducted superiority trials in which the active comparator clearly demonstrated clinically relevant efficacy, taking into account advances in medical or statistical practice relevant to the new trial.

(2) Study Quality of a Non-Inferiority Trial

Poor quality of a non-inferiority trial can sometimes lead to an apparent finding of non-inferiority trial that is incorrect.

There is a critical need for particular attention to study quality and conduct when planning and executing a non-inferiority trial.

(For example) Suppose that the test treatment is actually inferior to the active control.

- However, if there are many dropouts, power decreases.
- We cannot detect the difference, although the difference exists.
- We conclude that the test treatment is non-inferior to the active control, although the test treatment is inferior to the active control (Type I error!!)

(3) ITT and PP

- In superiority trial, ICH E9 states clearly that ITT should be the primary analysis, because it is conservative.
- However, in non-inferiority trials, there is no clear guideline.
- The problem with the per protocol set is that the process of extracting only a small number of subjects who are well compliant with the clinical trial protocol is not random. It may cause bias.

- In non-inferiority trials, ITT includes non-adherence, dropouts, measurement problems, so ITT can bias toward no treatment difference (success of non-inferiority trial) when treatment difference exists. (Increase of type I error!!)

- PP may bias toward either no treatment difference or treatment difference, depending on situations. (Either directions are possible)

(4) How will we handle historical data?

Should historical data be treated as a constant, or treated as a random variable?

There are two different approaches. (Kang 2016).

(a) The within-trial type I error rate (Fixed Margin Method)

The calculation of the within-trial type I error rate treats only the data in the current non-inferiority trial as random variables.

No matter how a non-inferiority margin is determined from historical data such as a lower confidence limit of the difference between the active control and placebo, the non-inferiority margin that is finally chosen is treated as a constant, not a random variable.

The within-trial type I error rate is the type I error rate for approval of new treatment in real practice.

$$P(\text{reject } H_{01} | H_{01})$$

$$= P\left(\frac{\bar{X}_T - \bar{X}_C + \delta}{\hat{\sigma}_{TC}} > 1.96 | H_{01}\right)$$

$$n_1 = kn_2 \qquad n_2 = \frac{\left(z_{\alpha/2} + z_\beta\right)^2 \sigma^2 (1 + 1/k)}{\left(|\mu_T - \mu_C| - \delta\right)^2}$$

(b) The (unconditional) across-trial type I error rate.

(Fixed Margin Method, Synthesis Method)

Both the data in the current non-inferiority trial and the historical trial data are treated as random variables.

The unconditional across-trial type I error rate is calculated by incorporating all possible values of historical trial data, although historical trial has already finished and historical data have already been observed.

$$T = \frac{(\bar{X}_T - \bar{X}_C) - (\lambda - 1)(\bar{X}_{C|H} - \bar{X}_{P|H})}{\sqrt{\hat{\sigma}^2_{TC} + (\lambda - 1)^2 \hat{\sigma}^2_{PC|H}}}$$

$$P(T \geq 1.96|H_{04}) = \int P(T \geq 1.96|T_H = t_H, H_{04}) w_{T_H}(t_H) dt_H$$

How to handle historical data applies to the development of biosimilar product as well.

Reference

Kang (2016). Avoiding ambiguity with the type I error rate in non-inferiority trials. Journal of Biopharmaceutical Statistics. 26(3). 452-465.

Thank you