

Guideline on Non-inferiority Clinical Trials

(Draft for Public Review)

English Translation by: Jianing Di and Xinxu Li

Disclaimer: The English is for information only and not an official translation and under any dispute the Chinese will prevail



Center for Drug Evaluation, NMPA

September, 2019

Table of Contents

1. Overview.....	1
2. Application Conditions	1
2.1 Historical Evidence of Active Comparator’s Efficacy	1
2.2 Constancy Assumption.....	2
2.3 Good Trial Quality	2
3. Key Points in Trial Design.....	2
3.1 Statistical Hypothesis.....	2
3.2 Active Control.....	3
3.3 Analysis Population	3
4. Determination of Non-inferiority Margin and the Corresponding Statistical Inference	3
4.1 Fixed Margin Method	4
4.2 Synthesis Method.....	5
5. Other Considerations	6
5.1 Potential Benefits Relative to Loss of Efficacy	6
5.2 Conversion between Non-inferiority and Superiority.....	6
5.3 Three-arm Non-inferiority Design.....	6
5.4 Communication with Regulatory Agencies	7
Appendix 1: Key Formulas	8
A1.1 Fixed Margin Method	8
A1.2 Synthesis Method.....	8
Appendix 2: Example	9
A2.1 Fixed Margin Method	9
A2.2 Synthesis Method.....	9

Guideline on Non-inferiority Clinical Trials

1. Overview

When confirming the efficacy of a drug, superiority trials (superiority of the test drug over placebo, the lower dose of the test drug, or the active drug) are commonly considered. Where superiority trials are not applicable, e.g., the use of placebo control might be considered unethical, consideration may be given to the use of non-inferiority trials. Non-inferiority trials were designed to confirm the clinical efficacy of the test drug, in the sense that the difference in treatment effect is within a clinically acceptable range even in case the test drug appears to be inferior to the active control.

The purpose of this guideline is to describe the application, design elements, non-inferiority margins, statistical inference, and other regulatory considerations in order to guide clinical trial stakeholders to understand, conduct, and evaluate non-inferiority trials. This guideline applies primarily to confirmatory clinical trials supporting the registration of drugs and biologics for marketing, but can also be used as a reference for exploratory clinical trials.

2. Application Conditions

Non-inferiority trials usually utilize active controls but sometimes are also supplemented with placebo (e.g., in a three-arm non-inferiority trial). Non-inferiority trials need to ensure adequate assay sensitivity, i.e., the ability to differentiate various levels of efficacy. Detailed discussion about assay sensitivities can be found in ICH E10, *Choice of Control Group and Related Issues in Clinical Trials*.

To ensure assay sensitivity of a non-inferiority trial, the following three aspects should be considered:

2.1 Historical Evidence of Active Comparator's Efficacy

In general, the efficacy of the active comparator relative to placebo are derived from existing well-designed and conducted clinical trial results. Based on the results of these trials, and taking into account the degree of variability among them, it is feasible to establish a reliable estimate for the efficacy of active control over placebo, which is a key parameter to determine the non-inferiority margin.

For some indications, such as certain symptomatic treatments, psychiatric indications, etc., it is often difficult to obtain a robust estimation of efficacy based on existing trials (e.g., even if the trial was well designed, it is sometimes difficult to obtain a robust conclusion that the active control is superior to placebo). As a result, non-inferiority trials with such kind active controls should be used with caution, and a three-arm non-

inferiority trial including placebo might be considered if allowed from an ethical perspective.

2.2 Constancy Assumption

The efficacy estimation of the active comparator over placebo mostly relies on historical clinical trials. In a non-inferiority trial, it is important that the efficacy of the active comparator appears to be consistent as it was in historical trials, i.e., the constancy assumption is satisfied. The constancy assumption can be impacted by a number of factors, such as the subject population, concomitant therapies, definition and determination of efficacy endpoints, dose level and potential resistance of the active comparator, and statistical analysis methods. Over time, if the definition of the treated disease, the diagnostic criteria, and the treatment methods have changed, the constancy assumption can be impacted, resulting in insufficient assay sensitivity of the non-inferiority trial and challenges in interpreting the trial results. Therefore, when the constancy assumption is difficult to verify, non-inferiority trials should be used with caution.

2.3 Good Trial Quality

The trial quality is the basis for adequate assay sensitivity of non-inferiority trials. Various trial quality issues, including protocol violations, poor adherence, use of concomitant medications, measurement bias, randomization/grouping errors, and high dropout rate, may all create bias in the efficacy estimation. These potential trial quality issues are often not in favor of the superiority conclusions, but maybe conducive to non-inferiority conclusions. Therefore, it is particularly important to ensure quality during the design and conduct of non-inferiority trials.

3. Key Points in Trial Design

When designing clinical trials, the trial objectives, evaluation variables, statistical assumptions, controls, sample sizes, and analysis populations should be considered. General considerations of clinical trial design as covered by other guidelines, such as those published by ICH (e.g., E8, E9) and by China National Medical Products Administration (e.g., *Biostatistical Guideline for Drug Clinical Trials*), are not within the scope and focus of this guideline. Instead, this guideline focuses on design points specific to non-inferiority trials, including statistical hypotheses (where non-inferiority margins are described in Section 4), and choice of active comparator and analysis populations.

3.1 Statistical Hypothesis

For different measures and different types of variables, the null hypothesis (H_0) and alternative hypothesis (H_1) of a non-inferiority trial are expressed differently. In Table

1, Δ as represents the non-inferiority margin, the absolute measures include the difference in means and rates, etc, and the relative measures include the rate ratio, hazard ratio, odds ratio, etc. In addition, the response variables are divided into those where higher values are to be considered better (HVB) and those lower values are to be considered better (LVB).

Table 1 The null hypothesis and alternative hypothesis (H_1) of a non-inferiority trial (H_0)*

Type of measure	HVB variable	LVB variable
absolute measure	$H_0: T - C \leq -\Delta (\Delta > 0)$ $H_1: T - C > -\Delta (\Delta > 0)$	$H_0: T - C \geq \Delta (\Delta > 0)$ $H_1: T - C < \Delta (\Delta > 0)$
relative measure	$H_0: T / C \leq 1/\Delta (\Delta > 1)$ $H_1: T / C > 1/\Delta (\Delta > 1)$	$H_0: T / C \geq \Delta (\Delta > 1)$ $H_1: T / C < \Delta (\Delta > 1)$

* T represents the effect of the test group; C represents the effect of the active control group.

3.2 Active Control

The active control in non-inferiority trials must have sufficient evidence of superiority over placebo, including a reliable estimate of the treatment effect. Therapies currently used as the standard of care or with the best effect should be selected as the active comparator. If the selected active control does not have sufficient evidence of efficacy, then there exists a meaningful risk in using it to evaluate other test drugs.

3.3 Analysis Population

The statistical analyses of non-inferiority trials should usually be performed based on the intention-to-treat (ITT) principle. However, it should be noted that analyses based on the ITT principle may not necessarily be conservative in non-inferiority trials, therefore, conclusions from the ITT analyses should generally be supported by analyses based on the per-protocol set (PPS). When the conclusions of ITT and PPS analyses are inconsistent, further assessment are needed to explain the observed inconsistency.

4. Determination of Non-inferiority Margin and the Corresponding Statistical Inference

The non-inferiority margin is defined as the largest clinically acceptable loss of efficacy when comparing the test drug with the active comparator. In addition, in order to adequate assay sensitivity, the non-inferiority margin should not be greater than the clinical benefit of the active control compared with placebo. The determination of the non-inferiority margin relies on both statistical consideration and clinical judgement, and should be described in detail in the protocol.

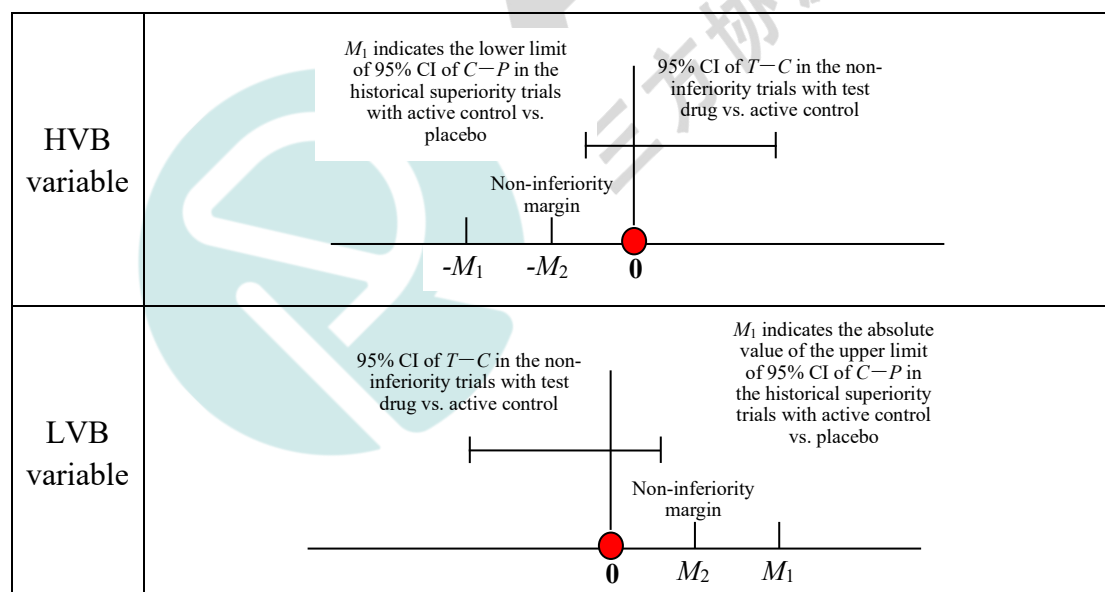
The method of determining the non-inferiority margin mainly includes the fixed margin method and the synthesis method. In usual cases, it is easier for the fixed margin method

to demonstrate the efficacy of the test drug.

4.1 Fixed Margin Method

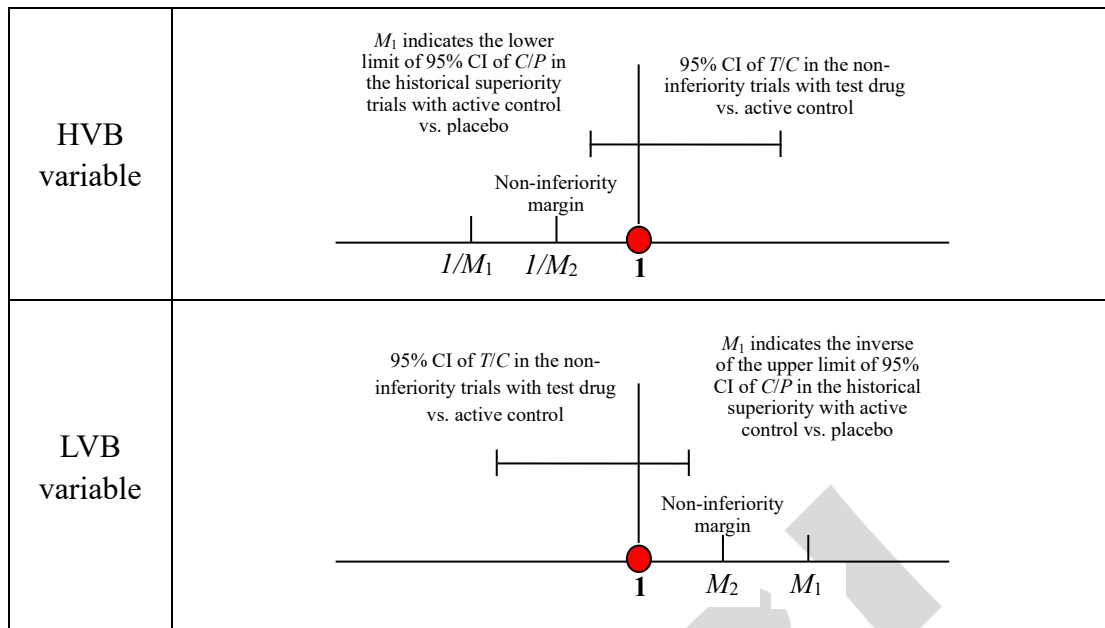
Let M_1 denotes the efficacy of active control over placebo. The estimation of M_1 usually relies on a meta-analysis of historical data, resulting in a 1-sided 97.5% (or 2-sided 95%) confidence interval (CI) for the treatment effect of the active control vs. placebo. The determination of M_1 is further illustrated in Figs 1 and 2. If concerns exist about the variability in historical evidence and the constancy assumption, a "discount" strategy can be used to determine M_1 , i.e., further reducing M_1 (e.g., by half) to ensure a more conservative estimation.

The non-inferiority margin, M_2 (denoted as Δ in the statistical hypotheses), on the other hand, is defined as clinically acceptable loss of efficacy in M_1 . Let f ($0 < f < 1$) be the lowest proportion of efficacy retention in M_1 , hence $1-f$ represents the largest acceptable loss. With that, the formula that determines M_2 are described in Appendix 1, while the relationship between M_1 and M_2 is illustrated in Figures 1 and 2. The determination of f depends on clinical assessment. When the efficacy of the active control is very different from that of the placebo, or when the endpoint relates to irreversible morbidity or mortality, the selection of f should be carried out with caution.



Note: T refers to the test drug; C refers to active control; P refers to placebo; CI refers to confidence interval

Figure 1 Determination of non-inferiority margins for absolute measures



Note: T refers to the test drug; C refers to active control; P refers to placebo; CI refers to confidence interval

Figure 2 Determination of non-inferiority margins for relative measures

If the test level (α) is set at one-sided 0.025 (or two-sided 0.05), for the HVB variables, non-inferiority can be concluded in the following two scenarios:

- 1) For an absolute measure, the lower limit of the one-sided 97.5% (or two-sided 95%) CI of treatment difference (test drug vs. active control) is greater than $-M_2$.
- 2) For a relative measure, the lower limit of the one-sided 97.5% (or two-sided 95%) CI of treatment difference (test drug vs. active control) is greater than $1/M_2$.

Similarly, for the LVB variables, non-inferiority can be concluded if the upper limit of the one-sided 97.5% (or two-sided 95%) CI of treatment difference (test drug vs. active control) is smaller than M_2 , regardless of absolute or relative measure.

4.2 Synthesis Method

The synthesis method does not require the specific margins (e.g., M_1 and M_2), but constructs a test statistic Z by combining data from historical superiority trials of the active comparator with placebo and the current non-inferiority trials of the test drug and the active comparator. The statistic Z is used to assess if the test drug can retain at least a portion of the active comparator's treatment effect, and its calculation formula is provided in Appendix 1. For the HVB variables, if Z is greater than $Z_{\alpha/2}$ ($Z_{\alpha/2}=1.96$ when $\alpha=0.05$), the non-inferiority of test drug vs. the active comparator can be concluded; for the LVB variables, if Z is smaller than $-Z_{\alpha/2}$, the non-inferiority can be concluded.

When the constancy assumption holds, the use of the synthesis method may improve study efficiency (by reducing sample size or obtaining greater power with no change in sample size) over the use of the fixed margin method. The synthesis method does not require pre-specification of M_1 and M_2 , but f .

5. Other Considerations

5.1 Potential Benefits Relative to Loss of Efficacy

Non-inferiority trials allow certain loss of efficacy for the test drug, but necessary compensation for such loss of efficacy should be considered, i.e., the test drug needs to provide other potential benefits over the active comparator, such as shorter treatment duration, easier administration, fewer adverse reactions, and better compliance. The use of non-inferiority trials is meaningful only if the test drug provides one or more of these potential benefits.

5.2 Conversion between Non-inferiority and Superiority

In the non-inferiority trial protocol, the conversion between non-inferiority and superiority tests can be defined in advance. Specifically, the non-inferiority test can be conducted first. If the non-inferiority conclusion is established, the superiority test can be further performed. In such cases, the superiority conclusion is established if the test is positive, and otherwise the original conclusion of non-inferiority will reflect the final decision of the study. If the step-1 non-inferiority conclusion is not valid, the superiority test is no longer conducted, and the study conclusion does not support non-inferiority. No multiplicity adjustments are required for this process.

The conversion between non-inferiority and superiority tests should be clearly defined in the clinical trial protocol. If superiority tests are not defined in the protocol for non-inferiority trials, even though the post-hoc superiority tests also turn out to be positive, the conclusion of the study will stay at non-inferiority. On the other hand, in a superiority trial with active control, if the non-inferiority test is not pre-specified in the protocol, a positive post-hoc non-inferiority results after a negative superiority test will not be accepted.

5.3 Three-arm Non-inferiority Design

Subject to ethical conditions, a three-arm non-inferiority design consisting of a test drug group, an active control group, and a placebo group may also be considered. The three-arm non-inferiority design can also examine whether the active control is superior to placebo while testing the non-inferiority of the test drug to the active control, thereby establishing high assay sensitivity. Therefore, allowed by the ethical considerations, the three-arm non-inferiority design is often considered ideal for confirming the non-inferiority of the test drug to the active control.

5.4 Communication with Regulatory Agencies

Prompt communication with regulatory authorities is encouraged when the applicant plans to use a non-inferiority trial. Topics of communication include but are not limited to the choice of active comparator, the determination of non-inferiority margin, the conversion of non-inferiority to superiority tests, and alternative design considerations. Before the communication, the applicant should provide the relevant information of the trial protocol, statistical analysis plan, etc. to the regulatory agency. For example, when discussing a non-inferiority margin, the applicant should provide a detailed illustration of the determination of the non-inferiority margin, including the literature and meta-analysis results used.



ARI
三方协调委员会

Appendix 1: Key Formulas

A1.1 Fixed Margin Method

If M_1 is for an absolute measure, then $M_2 = (1 - f)M_1$.

If M_1 is for a relative measure, then $M_2 = e^{(1-f)\ln(M_1)}$.

A1.2 Synthesis Method

For the efficacy evaluation of an absolute measure,

$$Z = \frac{(\widehat{T} - \widehat{C}_n) + f(\widehat{C}_h - P)}{\sqrt{SE_{\widehat{T}-\widehat{C}_n}^2 + f^2 SE_{\widehat{C}_h-P}^2}}$$

For the efficacy evaluation of a relative measure,

$$Z = \frac{\ln(\widehat{T}/\widehat{C}_n) + f\ln(\widehat{C}_h/P)}{\sqrt{SE_{\ln(\widehat{T}/\widehat{C}_n)}^2 + f^2 SE_{\ln(\widehat{C}_h/P)}^2}}$$

Here, C_h and P represent the effect of the active comparator and placebo in historical superiority trials, respectively; T and C_n represent the effect of the test drug and the active comparator in the current non-inferiority trial, respectively; f is the pre-specified efficacy as a retained proportion of that of C_h relative to P ; SE is the standard error, whereas the SE from historical superiority trials need to be estimated through meta-analyses.

Appendix 2: Example

A2.1 Fixed Margin Method

Consider a non-inferiority trial that evaluates the efficacy of a novel anticoagulant ximelagatran using an active comparator warfarin. Warfarin is a highly effective orally active anticoagulant that has been approved for the treatment of patients with non-valvular atrial fibrillation and with the risk of thromboembolic complications. From 1989 to 1993, a total of six placebo-controlled trials of warfarin were published for the treatment of patients with non-valvular atrial fibrillation. The main trial results are summarized in Schedule 1, providing the basis for the determination of non-inferiority margin in the non-inferiority trial assessing ximelagatran against warfarin.

Schedule 1. Placebo-controlled trials of warfarin for the treatment of non-valvular atrial fibrillation

Trial	Description	Events per person-year		Hazard ratio (95% CI) for warfarin versus placebo
		Warfarin	Placebo	
AFASAK	Open label, 1.2 years follow-up	9/413 = 2.18%	21/398 = 5.28%	0.41 (0.19, 0.89)
BAATAF	Open label, 2.2 years follow-up	3/487 = 0.62%	13/435 = 2.99%	0.21 (0.06, 0.72)
EAFT	Open label, 2.3 years follow-up	21/507 = 4.14%	54/405 = 13.3%	0.31 (0.19, 0.51)
CAFA	Double-blind, 1.3 years follow-up	7/237 = 2.95%	11/241 = 4.56%	0.65 (0.26, 1.64)
SPAF I	Open label, 1.3 years follow-up	8/260 = 3.08%	20/244 = 8.20%	0.38 (0.17, 0.84)
SPINAF	Double-blind, 1.7 years follow-up	9/489 = 1.84%	24/483 = 4.97%	0.37 (0.17, 0.79)

Based on a meta-analysis of the six trials, the hazard ratio (HR) of warfarin versus placebo is estimated to be 0.361 with a 95% CI of (0.248, 0.527). Since this endpoint is a LVB variable, M_1 is the inverse of the upper limit of the 95% CI, i.e., $M_1=1/0.527=1.90$.

The primary objective of this non-inferiority trial was to demonstrate that ximelagatran retains a substantial portion of efficacy of warfarin and therefore f was set at least 50%. As a result, the largest acceptable level of non-inferiority at the logarithmic scale is $(1 - 50\%)\ln(M_1)$, suggesting an M_2 of 1.38.

In the non-inferiority trials of ximelagatran and warfarin, the estimated HR was 1.39 with a 95% CI of (0.91, 2.12). The upper bound 2.12 is greater than M_2 . Therefore, the results of this trial cannot conclude that the effect of ximelagatran, in terms of risk reduction, is non-inferior to that of warfarin.

A2.2 Synthesis Method

Consider the same example where ximelagatran was compared to warfarin for the assessment of non-inferiority. The synthesis method compares the efficacy of ximelagatran in the current non-inferiority trial to placebo in historical superiority trials of warfarin versus placebo, hence an indirect comparison that is not based on including a placebo arm in the current trial. The synthesis method combines the data from historical superiority trials (warfarin vs. placebo) with the data

37 from the current non-inferiority trials of ximelagatran and warfarin to conduct a hypothesis test,
38 demonstrating that a certain proportion of warfarin's efficacy over placebo is retained in the non-
39 inferiority trial.

40

41 The key point of differentiation between the synthesis method and the fixed margin method is that
42 the efficacy of warfarin versus placebo (M_1) does not need to be pre-determined prior to the current
43 non-inferiority trial. Although warfarin is not directly compared with placebo in the current non-
44 inferiority trial, the assumption is that the efficacy over placebo, if any, of warfarin in the current
45 non-inferiority trial is the same as that observed in the historical superiority trials that compared
46 warfarin and placebo.

47

48 As such, the synthesis method statistically tests the null hypothesis that the inferiority of
49 ximelagatran compared with warfarin is less than half (50%) the risk reduction of warfarin
50 compared with placebo. This is a question that cannot be directly addressed by the fixed margin
51 method, as the placebo exists only in historical trials. To test on a logarithmic (log) risk scale, the
52 null hypothesis H_0 is:

53

$$54 \quad \ln(\text{HR ximelagatran vs. warfarin}) \geq -0.5\ln(\text{HR warfarin vs. placebo})$$

55

56 In the non-inferiority trial, the HR for ximelagatran versus warfarin was 1.39, and the 95% CI was
57 (0.91, 2.12). For the purpose of easy interpretation, based on the meta-analysis using the fixed
58 margin method, the HR for warfarin versus placebo was 0.361 with a 95% CI of (0.248, 0.527).
59 Based on this, the estimated HR on a logarithmic scale for ximelagatran vs. warfarin is $\ln(1.39) =$
60 0.329 with a standard error of 0.216. On the other hand, the HR estimate for warfarin relative to
61 placebo was $\ln(0.361) = -1.02$ with a standard error of 0.154. According to the formula for
62 synthesis methods, we have

63

$$64 \quad Z = \frac{0.329 + 0.5(-1.02)}{\sqrt{0.216^2 + [0.5(0.154)]^2}} = -0.789$$

65

66 Since $Z > -1.96$, the non-inferiority of ximelagatran as compared with warfarin cannot be
67 concluded.