Guideline on Non-inferiority Design of Drug Clinical Trials

（For Trial Implementation）

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, 2020

Table of Contents

1. Overview 1

2. Conditions of Application 1

2.1 Historical Evidence of Active Comparator's Efficacy 2

2.2 Constancy Assumption 3

2.3 Good Trial Quality 3

3. Key Points in Trial Design 4

3.1 Statistical Hypothesis 4

3.2 Active Comparator 5

3.3 Analysis Population 6

4. Determination of Non-inferiority Margin and Statistical Inference 7

4.1 Fixed Margin Method 7

4.2 Synthesis Method 9

5. Other Considerations 10

5.1 Potential Benefits Relative to Loss of Efficacy 10

5.2 Switching between Non-inferiority and Superiority 11

5.3 Three-arm Non-inferiority Design 11

5.4 Communication with the Regulatory Agency 12

Appendix 1: Key Formulas 13

A1.1 Fixed Margin Method 13

A1.2 Synthesis Method 13

Appendix 2: Examples 14

A2.1 Fixed Margin Method 14

A2.2 Synthesis Method 15

Guideline on Non-inferiority Design of Drug Clinical Trials

1. Overview

When confirming the efficacy of a drug, superiority trials (superiority of the test drug over placebo or the active drug) are often the ideal option. 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 are 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 comparator.

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

2. Conditions of Application

Non-inferiority trials utilize active comparators as the control, in order 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 comparator. Because the superiority of the active comparator over placebo cannot be directly observed in a non-inferiority trial, assumptions on the definitive efficacy of the active comparator needs to be made. Non-inferiority trials need to ensure adequate assay sensitivity, i.e., the ability to distinguish an effective active comparator from those that are less effective or ineffective. Detailed discussion about assay sensitivity 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 is derived from the results from existing well-designed and conducted clinical trials. Based on these trials, and taking into account the degree of variability among them, a reliable estimate for the efficacy of active comparator over placebo may be established, which is a key parameter to determine the non-inferiority margin in a non-inferiority trial.

For certain symptomatic treatments and specific indications, such as psychiatric diseases, etc., it is often difficult to obtain a robust estimation of efficacy of the active comparator over placebo based on existing trials. Non-inferiority trials with such active comparators may not be used to confirm the efficacy of test drug. Therefore, in these disease areas, non-

inferiority trials should be used with caution, or alternatively 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. As a result, efforts should be made to ensure the efficacy of the active comparator in a non-inferiority trial remains consistent with that from historical trials, i.e., the constancy assumption is satisfied. The constancy assumption can be impacted by a number of factors, such as the trial participants, use of concomitant medications, definition and determination of efficacy endpoints, dose level and potential resistance of the active comparator, and statistical analysis methods. Over time, the definition of the treated disease, its diagnostic criteria, and treatment options may have changed, and hence the constancy assumption can be impacted, resulting in insufficient assay sensitivity of the non-inferiority trial and challenges in trial results interpretation. 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 deviation from protocol eligibility criteria, poor adherence, use of concomitant medications that

impacts the drug effect evaluation, measurement bias, randomization/grouping errors, and high dropout rate, etc., may all create bias in the estimation of efficacy difference between the test drug and active comparator. 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 trial quality during the design and conduct of non-inferiority trials.

3. Key Points in Trial Design

When designing a clinical trial, the trial objectives, study population, choice of control, evaluation variables, statistical assumptions, sample sizes, and method of data analysis and interpretation should all be considered. General considerations of clinical trial design as covered in other guidelines, such as those published by ICH and the *Biostatistical Guideline for Drug Clinical Trials* by China National Medical Products Administration, should be followed and hence are not described in detail in this guideline. Instead, this guideline focuses on key design elements specific to non-inferiority trials, including statistical hypotheses (where the non-inferiority margins are described in Section 4), and choice of active comparator and analysis populations.

3.1 Statistical Hypothesis

The statistical hypothesis of non-inferiority should be clearly stated in the

study protocol. For different measures and types of variables, the statistical hypothesis in a non-inferiority trial may be stated differently (see Table 1). Specifically, the null hypothesis (H_0) reflects inferiority and the alternative hypothesis (H_1) reflects non-inferiority; M indicates the non-inferiority margin; the absolute measures include the difference in means and rates, etc., whereas the relative measures include the rate ratio, hazard ratio, odds ratio, etc. In addition, the response variables are divided into those for which higher values represent better treatment effect (HVB) and those for which lower values represent better treatment effect (LVB).

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

Type of measure	HVB variable	LVB variable
Absolute measure	H_0: $T-C \le -M$ $(M>0)$ H_1: $T-C > -M$ $(M>0)$	H_0: $T-C \ge M$ $(M>0)$ H_1: $T-C < M$ $(M>0)$
Relative measure	H_0: $T/C \le 1/M$ $(M>1)$ H_1: $T/C > 1/M$ $(M>1)$	H_0: $T/C \ge M$ $(M>1)$ H_1: $T/C < M$ $(M>1)$

* T represents the effect of the test group; C represents the effect of the active comparator group; M represents the non-inferiority margin.

3.2 Active comparator

The active comparator in non-inferiority trials must have clear and sufficient evidence of superiority over placebo, including a reliable estimate of the treatment effect. Therapies currently used as the standard of care or as the optimal treatment option should be selected as the active comparator. If the selected active comparator 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

In a superiority trial, analyses performed based on the intention-to-treat (ITT) principle are often considered conservative. However, that is not necessarily true in non-inferiority trials. Certain trial quality issues, such as poor adherence, high dropout rate, incorrect classification of primary endpoint, etc., may conceal the treatment difference between test drug and control drug, resulting in an incorrect non-inferiority conclusion when the test drug is in fact inferior to the control drug.

On the other hand, trial participants' adherence to study protocol may be related to the actual drugs or treatment they received. Therefore, analyses based on the per-protocol set (PPS) may also introduce biases. For instance, to evaluate the treatment effect of subjects who can tolerate and continue receive the treatment, the PPS may not include comparable subjects from different treatment groups. Any analysis based on PPS should focus on the treatment effect in the clinically targeted population, and confirm that the observed treatment effect is due to the test drug instead of potential confounding factors (e.g., duration of observation and difference in subject characteristics).

For non-inferiority trials, to reduce the quality issues, attention to study quality should be paid starting from the design stage and the study quality should be continuously monitored during study conduct and data analysis. In case of an open-label non-inferiority trial, the attention to study

quality is of particular importance because it is often challenging to prove that no bias is introduced during trial enrollment, endpoint evaluation and other trial-related activities.

4. Determination of Non-inferiority Margin and 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. Therefore, in order to guarantee the superiority of test drug over placebo, the non-inferiority margin should not be greater than the clinical benefit of the active comparator compared with placebo. The determination of the non-inferiority margin relies on comprehensive statistical evaluation and clinical judgement, and these considerations should be described in detail in the study protocol.

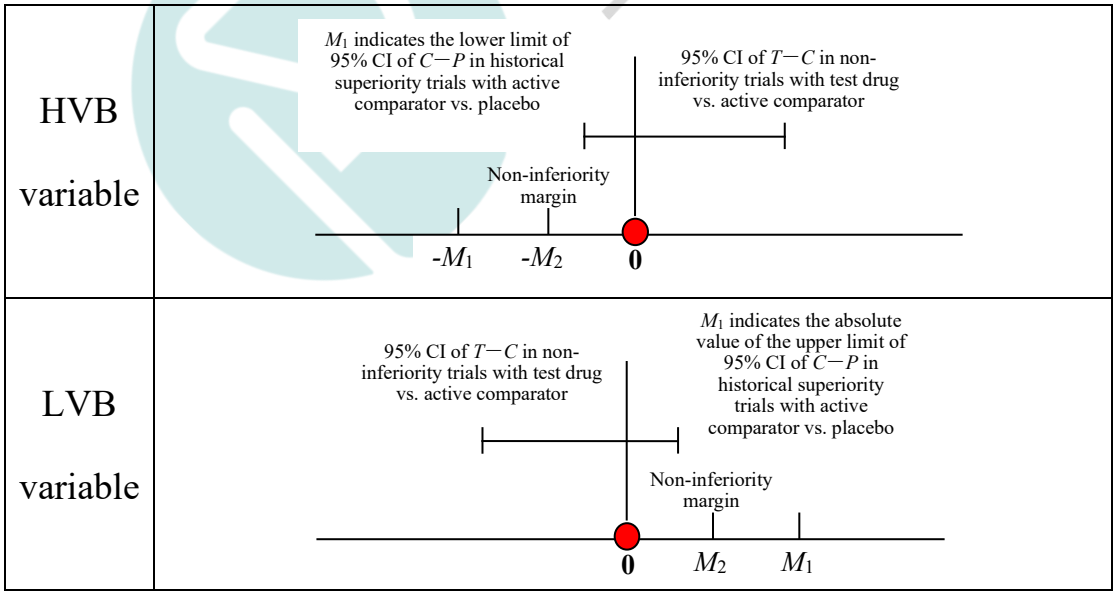
The determination of the non-inferiority margin and corresponding statistical inference are often performed based on the fixed margin method or the synthesis method. In usual cases, the fixed margin method can provide a more intuitive illustration of the efficacy of the test drug.

4.1 Fixed Margin Method

Let M_1 denote the treatment effect of active comparator over placebo. The estimation of M_1 usually relies on a meta-analysis of historical superiority studies, resulting in a 1-sided 97.5% (or 2-sided 95%) confidence interval (CI) for the treatment effect of the active comparator vs. placebo. The determination of M_1 is further illustrated in Figures 1 and 2. If concerns

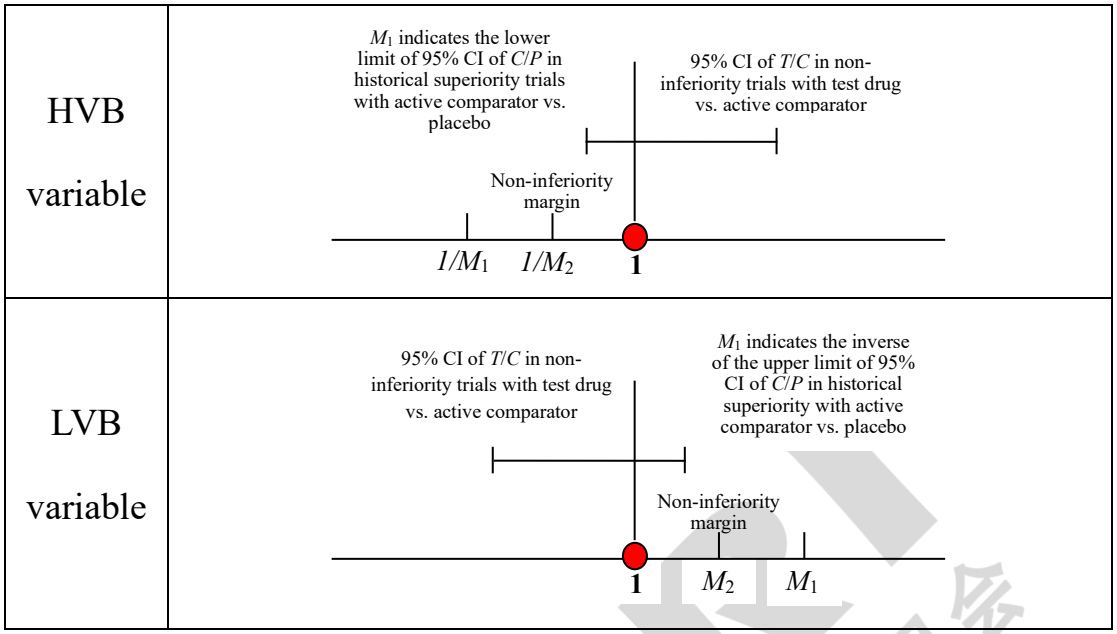
exist regarding the variability in historical evidence and the constancy assumption, a "discount" strategy can be used to determine M_1, i.e., further reducing M_1 (e.g., by half) to establish a more conservative value.

The non-inferiority margin, M_2 (denoted as M in Table 1), is defined as the largest clinically acceptable loss of efficacy and can be defined as a certain proportion of M_1. Let f ($0<f<1$) be the lowest proportion of efficacy retention in M_1, hence 1-f represents the largest proportion of acceptable loss. With that, the formula that determines M_2 are described in Appendix 2, while the relationship between M_1 and M_2 is illustrated in Figures 1 and 2. The determination of f depends on clinical judgement. When there is great efficacy of the active comparator over 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 comparator; 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 comparator; *P* refers to placebo; *CI* refers to confidence interval.

Figure 2 Determination of non-inferiority margins for relative measures

Let the test level (α) be set at one-sided 0.025 (or two-sided 0.05). For an HVB variable with an absolute measure, non-inferiority can be concluded if the lower limit of the one-sided 97.5% (or two-sided 95%) CI of treatment effect (test drug vs. active comparator) is greater than $-M_2$ (or $1/M_2$ for a relative measure). For an LVB variable, non-inferiority can be concluded if the upper limit of the one-sided 97.5% (or two-sided 95%) CI of treatment effect (test drug vs. active comparator) is smaller than M_2, regardless of absolute or relative measure.

4.2 Synthesis Method

The synthesis method does not require the pre-specification of M_1, but constructs a test statistic *Z*, by combining data from historical superiority

trials of the active comparator vs. placebo and the current non-inferiority trials of the test drug vs. the active comparator, to assess if the test drug can retain a portion of the active comparator's treatment effect. The calculation formula for Z is provided in Appendix 2. Let $Z_{1-\alpha/2}$ denote the 100 (1-α/2)% percentile of the standard normal distribution. The non-inferiority conclusion can be made if Z is greater than $Z_{1-\alpha/2}$ for an HVB variable or if Z is smaller than $Z_{1-\alpha/2}$ for an LVB variable.

When the constancy assumption holds, comparing with the use of the fixed margin method, the use of the synthesis method may improve study efficiency (by reducing sample size or obtaining greater power without changing sample size). The synthesis method does not require pre-specification of M_1, but f in the study protocol based on clinical judgement.

5. Other Considerations

5.1 Potential Benefits Relative to Loss of Efficacy

Non-inferiority trials allow certain loss of efficacy in the test drug relative to the active comparator. Correspondingly, necessary compensation for such loss of efficacy, in terms of potential benefits in other aspects, should be considered. For example, as compared with the active comparator, other potential benefits may include shorter treatment duration, easier administration, fewer adverse reactions, and better adherence. The evaluation of potential benefits should consider the objectives of the non-

inferiority trial and the clinical question of interest.

5.2 Switching between Non-inferiority and Superiority

In the protocol of a non-inferiority trial, the switch between non-inferiority and superiority tests can be defined in advance. Specifically, the non-inferiority test can be conducted first, and the superiority test can be further carried out if the non-inferiority conclusion is established. In such cases, the superiority conclusion is established if the test is positive, and otherwise the study should conclude with non-inferiority only. If in the first step the non-inferiority conclusion was not established, then the study conclusion does not support non-inferiority and the superiority test should not be further performed.

In a superiority study with an active comparator, if the non-inferiority study is to be performed in case superiority is not established, such switch needs to be pre-specified in the study protocol. This includes the definitions of the non-inferiority hypothesis, non-inferiority margin, and strategy for multiplicity adjustment, etc.

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 comparator group, and a placebo group may also be considered. The three-arm non-inferiority design can examine whether the active comparator is superior to placebo while testing the non-inferiority of the test drug to the active comparator, thereby establishing

clear assay sensitivity within the clinical trial. Therefore, when ethically appropriate, the three-arm non-inferiority design is often considered ideal for confirming the non-inferiority of the test drug to the active comparator.

5.4 Communication with the Regulatory Agency

Timely communications with the regulatory agency are 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 switch between non-inferiority and superiority tests, and considerations of alternative designs. Prior to the communication, the applicant should provide to the regulatory agency relevant information such as trial protocol that includes considerations of statistical analyses. 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.

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,

$$Z = \frac{(\widehat{T - C_n}) + (1-f)(\widehat{C_h - P})}{\sqrt{SE^2_{\widehat{T-C_n}} + (1-f)^2 SE^2_{\widehat{C_h-P}}}}$$

For the efficacy evaluation of a relative measure,

$$Z = \frac{\ln(\widehat{T/C_n}) + (1-f)\ln(\widehat{C_h/P})}{\sqrt{SE^2_{\ln(\widehat{T/C_n})} + (1-f)^2 SE^2_{\ln(\widehat{C_h/P})}}}$$

In these formulas, 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 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. Here, the relative measure is illustrated by a simple ratio (e.g., relative risk). Some other relative measures (e.g. the hazard ratio based on a proportional hazard model) cannot be written as a simple ratio in most cases. However, the same conclusion can be derived.

Appendix 2: Example

A2.1 Fixed Margin Method

Consider a non-inferiority trial that evaluates a novel anticoagulant ximelagatran against the 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		Relative risk (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 fixed effect meta-analysis of these six trials, the relative risk (RR) of warfarin versus placebo is estimated to be 0.361 with a 95% CI of (0.267, 0.489). Since this primary endpoint is an LVB variable, M_1 is the inverse of the upper limit of the 95% CI, i.e., M_1=1/0.489=2.04.

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 1.43.

In the non-inferiority trials of ximelagatran and warfarin, the estimated RR was 1.39 with a 95% CI of (0.91, 2.12). The upper bound 2.12 is greater than M_2. Therefore, based on the results of this trial, we 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. The synthesis method compares the efficacy of ximelagatran in the current non-inferiority trial to placebo in historical superiority trials of warfarin versus placebo. This is an indirect comparison without including a placebo arm in the current trial. The synthesis method combines the data from historical superiority trials (warfarin vs. placebo) with the data from the current non-inferiority trials of ximelagatran and warfarin to conduct a hypothesis test, demonstrating that a certain proportion of warfarin's efficacy over placebo is retained in the non-inferiority trial.

The key point of differentiation between the synthesis method and the

fixed margin method is that the efficacy of warfarin versus placebo (M_1) does not need to be pre-determined prior to the current non-inferiority trial. Although warfarin is not directly compared with placebo in the current non-inferiority trial, the assumption is that the efficacy of warfarin over placebo, if any, in the current non-inferiority trial is the same as that observed in the historical superiority trials that compared warfarin and placebo.

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

$$\mathrm{Ln}(RR\ \text{ximelagatran vs. warfarin}) \geq -0.5\ln(RR\ \text{warfarin vs. placebo})$$

In the non-inferiority trial, the RR for ximelagatran versus warfarin was 1.39, and the 95% CI was (0.91, 2.12). For the purpose of easy interpretation, based on the meta-analysis using the fixed margin method, the *RR* for warfarin versus placebo was 0.361 with a 95% CI of (0.267, 0.489). Based on this, the estimated HR on a logarithmic scale for ximelagatran vs. warfarin is $\ln(1.39) = 0.329$ with a standard error of 0.216. On the other hand, the RR estimate for warfarin relative to placebo was $\ln(0.361)$ = -1.02 with a standard error of 0.154. According to the

formula for synthesis methods, we have

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

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