

# “Points to Consider on Switching between Superiority and Non-inferiority” 及び “Guideline on the Choice of the Non-inferiority Margin” の紹介

Introduction of “Points to Consider on Switching between Superiority and Non-inferiority” and “Guideline on the Choice of the Non-inferiority Margin”

西山 智

Hiroshi Nishiyama

武田薬品工業株式会社 統計解析部

Biostatistics, Takeda Pharmaceutical Company Limited

e-mail: Nishiyama\_Hiroshi@takeda.co.jp

This paper introduces the contents of two documents from the European authority, “Points to Consider on Switching between Superiority and Non-inferiority” and “Guideline on the Choice of the Non-inferiority Margin”. Issues relevant to possible applications in Japanese clinical trials are also discussed.

*Key words:* Non-inferiority, Superiority, Switching, Delta, Choice of Margin.

## 1. はじめに

欧州の規制当局で纏められている “Points to Consider on Switching between Superiority and Non-inferiority” (CPMP, 2000) と “Guideline on the Choice of the Non-inferiority Margin” (CHMP, 2005) は、前者は臨床試験の治療群間比較の形式としての主目的が優越性である場合に非劣性の検討を実施すること、及び主目的が非劣性である場合に優越性の検討を実施することについて、後者は非劣性試験における非劣性の限界値 ( $\Delta$ ) の設定について議論している。以下では、前者をスイッチングガイドライン、後者をデルタガイドラインと呼ぶことにする。

デルタガイドラインは、ごく最近の 2005 年 7 月に ‘Points to Consider’ から ‘Guideline’ となり、2006 年 1 月から欧州で効力が発生する。‘Points to Consider’ から ‘Guideline’ とされる段階で、非劣性試験の実施に要求する妥当性に関して、ICH E10 より踏み込んだ表現に変更され、プラセボを同時対照に含むことが可能な場合は、3 アーム (被験薬, 対照薬, プラセボ) 試験の実施を常に推奨するとしている。

これら 2 つの議論はそれほど新しいものではない。ICH E9 (「臨床試験の統計的原則」(厚生省医薬品安全局, 1998)) ガイドラインと関連が深く、対照群の選択について特別に取り扱った ICH E10 (「臨床試験における対照群の選択とそれに関連する諸問題」(厚生労働省医薬局, 2001)) が Step 4 となったのが 2000 年 7 月 20 日であるのに対して、デルタガイドラインは 1999 年 12 月から 2004 年 1 月にかけて議論されている。デルタガイドラインは、非劣性試験において限界値の設

定は不可避であるから ICH E10 と同時期に議論されているのは自然である。これに対して、臨床試験の主目的を試験開始後に変更する可能性を含む、都合良すぎると批判が出そうにも思えるスイッチングガイドラインに関連して、1995 年に優越性と非劣性は仮説の包含関係から多重性の問題がないことが議論されている (Morikawa and Yoshida, 1995)。また、スイッチングガイドラインは、1999 年 2 月から ICH E10 と同時期に議論されており、それほど新しいアイデアではないことが分かる。

本稿は、スイッチングガイドライン及びデルタガイドラインの内容を紹介するとともに、その論点を提示し、日本の臨床試験への適用に関する議論の題材とすることを目的とする。

## 2. 優越性と非劣性

簡単のために、評価変数は正規分布に従い、正の方向に大きいほど効果大きいとする（両ガイドラインも同様の前提）。被験薬 ( $T$ ) と対照薬 ( $R$ ) の母平均の差を  $\delta(=\mu_T - \mu_R)$  とすると、優越性の検証を目的とする場合の帰無仮説と対立仮説はそれぞれ  $H_0^S: \delta \leq 0$ ,  $H_1^S: \delta > 0$  であり、非劣性の場合にはそれぞれ  $H_0^{NI}: \delta \leq -\Delta$ ,  $H_1^{NI}: \delta > -\Delta$  である。ただし、 $\mu_T$ ,  $\mu_R$  はそれぞれ被験薬 ( $T$ ) と対照薬 ( $R$ ) の母平均、 $\Delta > 0$  である。片側の仮説検定を考えているので、有意水準  $\alpha$  を 2.5% とする。

優越性が目的の場合に、ポジティブな結果 ( $H_0^S$  を棄却) に対する表現、

- 差の両側 95% 信頼区間がゼロを含まない (完全にゼロを超える)。
- 差は有意水準 2.5% で有意である。

は同義であり、図 1 の “SUPERIORITY SHOWN” に対応する。

また、非劣性が目的の場合に、ポジティブな結果 ( $H_0^{NI}$  を棄却) に対する表現は、

- 差の両側 95% 信頼区間の下限が  $-\Delta$  を超える。

のように検定を用いずに信頼区間のみで表現することが一般的であり、図 1 の “NON-INFERIORITY SHOWN” に対応する。

非劣性と優越性に関する検定の帰無仮説には包含関係がある。すなわち、 $H_0^{NI}$  が真のとき  $H_0^S$  は真である。言い換えると、 $H_1^S$  が真のとき  $H_1^{NI}$  は真であり、図 1 において “SUPERIORITY SHOWN” であれば “NON-INFERIORITY SHOWN” であることに対応する。この包含関係によ

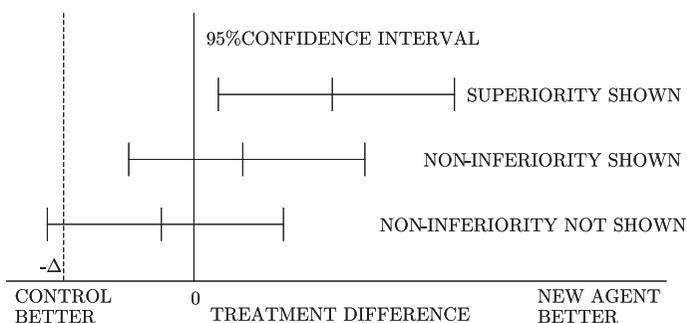


図 1. 優越性と非劣性の信頼区間による表示

S98 “Points to Consider on Switching between Superiority and Non-inferiority” 及び “Guideline on the Choice of the Non-inferiority Margin” の紹介 西山 智

り、非劣性の検定が有意であれば、優越性の検定を実施するという手順を採用することで多重性の問題が生じない (Morikawa and Yoshida, 1995) . これが, “Switching” というアイデアを議論する基礎であり, 動機となったのであろう .

### 3. Points to Consider on Switching between Superiority and Non-inferiority

本章では, スイッチングガイドラインの記述の抜粋と要約を示し, 論点を提示する .

#### 3.1 非劣性試験における優越性の解釈

スイッチングガイドラインは, 非劣性試験において, 優越性に目的を変更することは, 下記を満足すれば可能である述べられている . ニュアンスを誤って伝えないために, 原文と日本語訳を併記した . なお, 以降の日本語訳は著者の訳である .

- The trial has been properly designed and carried out in accordance with the strict requirements of a non-inferiority trial .  
( 試験は, 非劣性試験に対する厳密な要求に基づいて, 適切にデザインされ, 実施されていること . )
- Actual p-values for superiority are presented to allow independent assessment of the strength of the evidence.  
( 証拠の強さを独立に評価できるように, 優越性に関する実際の p 値を示すこと . )
- Analysis according to the intention-to-treat principle is given emphasis.  
( ITT 原則に従った解析を強調すること . )

スイッチングガイドラインを読む限り, 比較の形式を, 非劣性から優越性とする変更は, 受け入れやすいようである . すなわち, 非劣性試験を実施する際の多くのクリアすべき課題 ( 特に, 分析感度 ( assay sensitivity ) の存在 ) が, 優越性の形式に変更することで問題となくなることである . さらに, 優越性の証明そのものが分析感度の存在の証拠となるからである .

#### 3.2 優越性試験における非劣性の解釈

スイッチングガイドラインは, 優越性試験において, 非劣性に目的を変更した場合, 導かれる結論の信頼性は乏しいとしている .

試験の分析感度を確保するために, 下記のこと特に注意を払う必要があると記載されている .

- finding evidence, direct or indirect, that the control treatment is showing its usual efficacy.  
( 対照薬が通常の効果を示す証拠を, 直接的または間接的に, 見出すこと . )
- comparing the trial with earlier trials which demonstrated the efficacy of the control agent.  
( 対照薬の有効性が示された以前の試験と比較すること . )
- in particular demonstrating levels of non-compliance and of loss of patients and data that are at least comparable to those in the earlier trials.  
( 特に, 遵守不良及び脱落の度合いについて, 以前の試験と同等以上の質であることを示すこと . )
- showing similar results form the full analysis set and PP analysis set.  
( FAS 集団と PP 集団で類似した結果を示すこと . )

これらは、限界値の設定に特化したものだけでなく非劣性試験の満たすべき一般的条件も含んでいる。優越性を主目的として計画された試験において非劣性を検討する場合には、非劣性の一般的条件が満たされることを実施計画書に記載しておくことが必要であろう。また、優越性が認められなかったために、計画していなかった非劣性を事後的に検討するような場合には、結果を見てから都合の良い限界値を設定していないこと、分析感度の存在、事前に計画していない目的を検討するに至ったことを説得力をもって説明する必要があると考えられ、より注意が必要であろう。

また、ガイドラインの‘Conclusion’として、これらが満足されれば変更は認められるかもしれない(“may be feasible”)と記載されている内容の抜粋を以下に示す。

- The non-inferiority margin with respect to the control treatment was pre-defined or can be justified. (The latter is likely to prove difficult and to be limited to rare cases where there is a widely accepted value of  $\Delta$ .)

(非劣性の限界値が事前に定められていること、あるいは(事後的でも)正当化できること。(事後的に正当化することは、難しいことがわかるであろうし、またそれは $\Delta$ の値が広く認められているという稀なケースに限定されるであろう。)

- Analysis according to the intention-to-treat principle and PP analysis, showing confidence intervals and p-values for the null hypothesis of inferiority, give similar findings.

(ITT原則とPPに従った解析対象の解析で、信頼区間とp値が類似した結果を示すこと。)

- The trial was properly designed and carried out in accordance with the strict requirements of a non-inferiority trial (see ICH E9 and E10).

(非劣性試験に対する厳密な要求に基づいて、適切にデザインされ、実施された試験であること。)

- The sensitivity of the trial is high enough to ensure that it is capable of detecting relevant differences if they exist.

(意味のある差が存在する場合には、それを検出できることを保証しうるだけの、十分高い試験の分析感度を有すること。)

- There is direct or indirect evidence that the control treatment is showing its usual level of efficacy.

(対照薬が通常の効果を示しているという直接的または間接的な証拠があること。)

スイッチングガイドラインでは、非劣性試験の満たすべき一般的条件が繰り返し強調されており、優越性を主目的として計画した場合に非劣性を検討することは推奨していないようである。著者の考えでは、優越性試験の中で非劣性の検討を実施する場合には、非劣性試験の一般的条件が実施計画書に記載されるべきであり、当初計画されていない非劣性を事後的に検討する場合は、その妥当性の説明はより説得力が伴う必要があるように思う。

### 3.3 スwitchingガイドラインの論点

- 主要な解析対象集団として、full analysis set と PP analysis set のどちらを選択するかについて、ガイドラインは、優越性試験は full analysis set を選択し、非劣性試験では両者とも同

様に重要で、類似した結果が示されることが重要であるとしている。著者の考えを補足すると、優越性、非劣性のどちらも、full analysis set の質が良くない状況では、得られた結果に不確かさが伴うと考えられ、類似した結果を示すのみでなく、試験の質についても検討する必要があると思う。試験の質について、説得力をもって示す最良の方法は、ガイドラインが推奨するプラセボ対照を含む 3 アーム試験ではないであろうか。また、非劣性の検討において、一般には full analysis set が PP analysis set より保守的な結果を与えると考えられているが、評価変数が 2 値データで、途中脱落が応答に依存する場合には、群間差の推定値は応答（割合）に依存して真値の前後で変化する状況もあり、注意が必要である。

- ガイドラインでは、優越性が主目的のときに非劣性を検討する場合、及び非劣性が主目的のときに優越性を検討する場合のいずれの場合も実施計画書に事前に記載しておくべき（‘Pre-Definition’）と述べ、解析方法として非劣性から優越性への段階的手順を提示している。

単純に検定の多重性のみを考えると、優越性と非劣性に関する検定の仮説の階層関係において、主目的をどちらの方向で変更しても問題ない。実際には、優越性試験において非劣性を検討する場合においても、非劣性試験で ‘Pre-Definition’ が求められる一般的条件（分析感度、限界値の設定等）が同様に必要であろうから、試験計画時の主目的が優越性であっても ‘Pre-Definition’ が必須だろう。

これに対して、非劣性試験において優越性を検討する場合はどうであろうか。非劣性試験として計画されているから必要な一般的条件は実施計画書に ‘Pre-Definition’ されており、結果として非劣性試験に要求される質の高い試験が実施された上で、試験実施前に計画していなかった優越性が示された場合に、「優越性が検証された」との解釈に疑問はないだろうか。

- スイッチングガイドラインには、受け入れられる具体的な治療領域の例示がない。著者の感想として、スイッチングのアイデアを適用する場面が容易にイメージできるように記載されていないのは残念である。

#### 4. Guideline on Choice of the Non-inferiority Margin

デルタガイドラインでは、設定する「非劣性の限界値」について、統計学的な合理性と臨床的な正当性の両面からの説明を求めており、また、対象とする治療領域にあわせて説明することを求めている。治療領域横断的に一般化できる限界値の設定方法はなく、原則のみを示すとしており、ガイドラインの議論は ICH E9, ICH E10 で要求されている非劣性試験としての妥当性が確保されていることが前提であると強調している。

さらに、ガイドラインは、非劣性試験を実施することの妥当性、及び設定する限界値の正当性を完全に示すことは難しいと述べ、プラセボを対照として含めることが可能な場合には、全ての非劣性試験に 3 アーム試験を推奨している。‘CONCLUSIONS’ の章に、“A three-armed trial with test, reference and placebo allows some within-trial validation of the choice of non-inferiority margin and is therefore associated with fewer difficulties. This is the recommended design and should be used wherever possible.” と記載されている。

また、非劣性試験の目的には「demonstrating efficacy（プラセボに対する有効性の検証）」と

「establishing acceptable efficacy relative to the active comparator (対照薬に対する非劣性の検証)」があり、両者は限界値の設定方法が異なるとし、実施計画書に目的を明示することを強調している。

本章の次節以降に、上記の目的別に限界値の設定に求められることについて、デルタガイドラインから抜粋と要約を示し、論点を提示する。

#### 4.1 プラセボに対する有効性

デルタガイドラインは、非劣性試験の結果を解釈するためには、もしプラセボ対照試験が実施されたなら、プラセボに対する優越性は示されるだろうという最低限の前提が要求されるとしている。優越性試験では、統計的に有意であることに加えて、プラセボに対する差が臨床的観点から有用であることが求められるとしている。

##### 4.1.1 Three arm trials: test, reference and placebo

デルタガイドラインは、3 アーム試験は、被験薬と対照薬のプラセボに対する優越性を直接評価することができ、このとき、限界値の設定は必要ないとしているが、解釈には下記を考慮する必要があると記載されている。

- 被験薬と対照薬ともにプラセボに対する優越性が示されない場合、それは試験の分析感度が低いことを示唆していること。
- 優越性試験として、プラセボとの差の臨床的な妥当性を評価すること。
- 対照が効果の確立している承認済みの薬剤であるなら、対照薬のプラセボに対する差は、被験薬のプラセボに対する差の臨床的意義の評価に役立つこと。
- 対照薬のプラセボに対する優越性が示されない場合、または期待する差と大きく異なる場合、分析感度、限界値の設定など全てが疑わしく、どのようなポジティブな結果も、この試験単独からは信頼できないこと。

##### 4.1.2 Two arm trials: test and reference

デルタガイドラインは、プラセボを含まない被験薬と対照薬の 2 アーム試験では、対照薬のプラセボに対する優越性が示された過去の試験に基づいて、被験薬のプラセボに対する間接的な比較による限界値の設定を強いられるとしている。対象とする患者集団における対照薬とプラセボの差の推定ができない、または過去の試験結果に一貫性がない場合、試験の分析感度には疑問があり、対照薬に対する優越性だけが解釈可能であるとしている。

ガイドラインには、対照薬の効果が推定できる質の良い過去のデータがある場合の限界値の設定法の例示として、“The ‘historical’ confidence interval compares the reference product with placebo ( $r - p$ ). The planned trial comparing the test and reference products will also produce a confidence interval (for  $t - r$ ). If these intervals are combined, an indirect confidence interval comparing the test product and placebo can be obtained ( $t - p$ ). Delta can be defined as the lower bound of  $t - r$  that ensures that the lower bound of the indirect confidence interval of  $t - p$  will be above zero. As the comparison is indirect it might be wise to be conservative and select some value smaller than that suggested by this indirect calculation.” という記載がある。要約すると、「対照薬のプラセボに対する差 ( $r - p$ )、及び被験薬と対照薬の差 ( $t - r$ ) の信

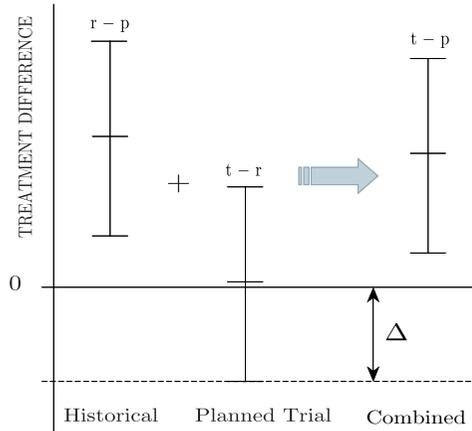


図 2. 非劣性の限界値の設定

信頼区間について、両者を統合することで間接的な被験薬のプラセボに対する差 ( $t-p$ ) の信頼区間が得られる。すなわち、非劣性の限界値を ( $t-r$ ) の信頼区間の下限として設定することにより ( $t-p$ ) の信頼区間がゼロを超えることを保証することができる。間接的な比較であるから、保守的かつ間接的な計算から得られる値より小さくすることが賢明だろう。」であり、これを図示すると図 2 となる。

#### 4.2 対照薬に対する非劣性

ガイドラインは、実際の治療において被験薬を使用するならば、対照薬を使用するのに比べて、有効性に関して重要な損失が無いことを検証することが、通常非劣性試験の目的であると述べており、プラセボと比較した対照薬の過去の試験からだけでは、この目的に適切な非劣性の限界値は設定できないとしている。

また、ガイドラインは、「対照薬のプラセボに対する効果の大きさの割合に基づいて限界値を設定するというアイデアは、プラセボより優れていることの保証を定式化したに過ぎない」としている。対照薬に対する非劣性の検証を目的とする場合には、対象とする治療領域で重要でないと考えられる差を根拠として適切な限界値を設定することと述べている。また、プラセボ対照を含まない試験では、被験薬のプラセボに対する効果は分からないことを指摘し、限界値は十分厳しく小さく設定することが、プラセボに対する効果及び対照薬に実質的に劣らないことの証明となり得るとしている。

質の良い過去のデータがない場合には、その治療領域における他の薬剤のデータの統合や、専門家が重要でないとする差を調査するアイデアも記載されている。

#### 4.3 特殊な状況

一方、非劣性の限界値を設定できない特殊な状況では（たとえば、死亡が評価変数）、優越性の検定において有意水準を緩める（たとえば、5%から 15%）というアイデアがあり得るとしている。すなわち、特殊な状況では、false positive と false negative の価値が異なり得るということである。他方で、両者の違いの最適なバランスを定めるのは難しいとも述べている。

#### 4.4 デルタガイドラインの論点

- プラセボを含む 3 アーム試験を原則推奨としているが、倫理的な観点から実施可能であろうか。ガイドラインでは倫理面についての記載は無い。
- 臨床的観点から、重要な差ではない範囲が設定できる治療領域とはどのようなものか。一般に、臨床的観点からの設定は可能か。
- 2005 年 9 月 21 日～22 日に行われた計量生物セミナーの中で、東京理科大学の吉村功教授は、許容限界値  $\Delta$  についてパラメータ空間とサンプル空間を区別して議論する必要があると指摘された。その内容について、著者の解釈を以下に示す。

ガイドラインでは、臨床的観点から正当な「許容限界値」を定めることを強調している。しかしながら、ガイドラインの言う臨床的な許容範囲とは、パラメータ空間上の仮説の採択・棄却の領域なのか、サンプル空間上の棄却限界値なのか、どちらを意図しているのか明記されていない。

すなわち、パラメータ空間上の議論とは、非劣性と認めて良い母平均の差  $\delta$  (被験薬の母平均-対照薬の母平均) を定めることであり、サンプル空間上の議論とは、得られたデータから非劣性と判断する母平均の差の信頼区間の下限値 ( $-\Delta$ ) を定めることに対応する。

著者は、「臨床的な許容範囲」という表現は、非劣性と認めて良い範囲と解釈する人が多いのではかと思う。すなわち、真の母平均の差  $\delta =$  「臨床的な許容範囲」のときには、高い確率で非劣性と判定したいと考えられる。しかし、ガイドラインの表現をそのまま用いて、帰無仮説を「 $H_0^{NI}: \delta \leq -\Delta$ 」とすると、真の母平均の差が  $-\Delta$  のときに、サンプルサイズに関係なく非劣性が示される確率は第 1 種の過誤 (2.5%) となる。したがって、この場合に十分な検出力で非劣性を示すためには、臨床的な許容範囲と棄却限界値としての  $\Delta$  は区別して定める必要がある。

#### 5. おわりに

ICH E9, スイッチングガイドライン, デルタガイドラインを並べて読むと、一般論としての非劣性試験に求められる妥当性の説明が最も難しい課題との印象を受ける。そうであれば、その課題が試験計画時にクリアできれば、主目的とする比較形式が優越性と非劣性のどちらであれ、他方の比較形式との 'Switching' 自体は大きな問題ではないように思う。実際には、優越性試験の中で非劣性を検討する場合には、そのハードルがより高くないか注意すべきだろうし、検証とは何かという観点から異論があるだろう。

#### 謝 辞

本論文の完成には査読者からの有益な示唆・助言によるところが多い。ここに記して感謝の意を表する。

S 104 “Points to Consider on Switching between Superiority and Non-inferiority” 及び “Guideline on the Choice of the Non-inferiority Margin” の紹介 西山 智

参考文献

- European Medicines Agency, Committee for Medicinal Products for Human Use (CHMP). *Guideline on the Choice of the non-inferiority Margin*. EMEA: London 2005,  
<http://www.emea.eu.int/htms/human/ewp/ewpfn.htm>.
- European Agency for the Evaluation of Medicinal Products, Committee for Proprietary Medicinal Products (CPMP). *Points to Consider on Switching between Superiority and Non-inferiority*. EMEA: London 2000,  
<http://www.emea.eu.int/htms/human/ewp/ewpptc.htm>.
- Koyama, T., Sampson, AR. and Gleser LJ. (2005). A framework for two-stage adaptive procedures to simultaneously test non-inferiority and superiority. *Statistics in Medicine* **24**, 2439-2456.
- Morikawa, T. and Yoshida, M. (1995). A useful testing strategy in phase III trials: Combined test of superiority and test of equivalence. *J. Biopharmaceutical Statistics* **5**, 297-306.
- 厚生省医薬品安全局 (1998). 「臨床試験における統計的原則」.
- 厚生労働省医薬局 (2001). 「臨床試験における対照群の選択とそれに関連する諸問題」.