

3 群間での非劣性検証に関する臨床試験 デザインについて

広島大学病院・総合医療研究推進センター * 飛田 英祐

Eisuke Hida

Hiroshima University Hospital

Center for Integrated Medical Research

§1. はじめに

医薬品等の開発では既存の治療法 (標準治療) と比べて, 新しい治療法 (試験治療) での有害事象や副作用の発現割合が低い, 投与回数の減少や患者への負担が軽減される投与経路などのコンプライアンスの向上が期待できる, 生存時間の延長や検査値を正常範囲に到達させるなどの直接的な治療効果の他に QOL (Quality of Life) の改善などのメリットがあれば, 治療効果 (主要評価項目) が臨床的に許容できる値 (非劣性マージン Δ) 以上に劣らなければ, 試験治療は有用性であると判断されることがある. この試験治療が標準治療の比べてある一定の値以上は劣らないことを証明する非劣性試験 (ICH E9 [1], ICH E10 [2], EMEA ガイドライン [3]) は多くの臨床試験において実施されており, 2010 年には米国 FDA (Food and Drug Administration) が非劣性試験デザインに関するドラフトガイダンス [4] を発出している.

通常の試験治療と標準治療の 2 群での非劣性試験において, 試験治療の有効性を証明するためには, 非劣性マージン Δ の設定方法と有効な治療と無効な治療を区別する能力である分析感度の保証が重要な鍵となる. ICH E9 [1] で, 非劣性マージンは, 「臨床的に許容できると判断しうる最大の差であり, 実対照薬の有効性を立証した優越性試験において観測された差よりも小さいものであるべき」と定義され, ICH E10 [2] では, 統計的に否定し

ようとしている「対照治療に対して劣っている程度」とされている。実際には、対照となる標準治療の過去のプラセボとの比較試験成績から設定されることが多い。一方、分析感度の保証とは、非劣性試験で標準治療が従来の治療効果を発揮できているかが重要であり、非劣性試験では直接評価できる内部コントロールがないため、外部情報（標準治療の過去に実施したプラセボ比較試験成績）から判断せざるを得ない (ICH E10 [2])。

そこで、直接的に内部妥当性を評価できるために、試験治療、標準治療に加えてプラセボを含めた 3 群での非劣性試験デザインが提案されている (ICH E9 [1], ICH E10 [2], EMEA ガイドライン [3], FDA ドラフトガイダンス [4])。この 3 群非劣性試験デザインについては、非劣性マージンの設定を 3 群非劣性試験で観測される標準治療とプラセボの治療効果 μ_R と μ_P 、試験の計画時点で事前に決定する割合 f ($0 < f < 1$) を用いた $\Delta_F = f \times (\mu_R - \mu_P)$ として、多くの論文で議論されているが ([5], [6], [7]), 非劣性マージン Δ として事前に設定する方法についてはあまり議論されていない。我々は、実際の非劣性試験で一般的な事前に 1 つの値として非劣性マージン Δ を設定する方法として、評価変数が連続変数, 2 値変数, 生存時間データでの 3 群非劣性試験デザインを提案しており、本稿ではその提案方法を紹介する ([8], [9], [10])。

§2. 3 群非劣性試験の定式化

試験治療群, 標準治療群にプラセボ群を含めた 3 群非劣性試験における主要評価項目の評価変数 X はいずれも互いに独立で同一の分布に従うものとする。

- 試験治療 (E) 群: $X_{E1}, X_{E2}, \dots, X_{En_E} \sim_{i.i.d} \text{パラメータ } \theta_E \text{ をもつ分布 } G$
- 標準治療 (R) 群: $X_{R1}, X_{R2}, \dots, X_{Rn_R} \sim_{i.i.d} \text{パラメータ } \theta_R \text{ をもつ分布 } G$
- プラセボ (P) 群: $X_{P1}, X_{P2}, \dots, X_{Pn_P} \sim_{i.i.d} \text{パラメータ } \theta_P \text{ をもつ分布 } G$

非劣性マージン Δ は通常の 2 群での非劣性試験と同様に、標準治療の過去に実施された (複数の) プラセボ対照試験における同じ主要評価項目の群間差より、臨床的に意味のない最大の値として計画時点で事前に 1 つの値として設定する。

§3 提案する方法

§3.1. 3 群非劣性試験における仮説検定

3 群非劣性試験では、分析感度を保証した下で試験治療の標準治療に対する非劣性を証明することが可能であり、そのため治療効果 (試験治療群: θ_E , 対照治療群: θ_R , プラセボ群: θ_P) には以下の大小関係が成り立つ ([8], [9], [10]).

$$\theta_P < \theta_R - \Delta < \theta_E \quad (1)$$

ただし、ここでは値が大きくなるほど治療効果が大きい場合を示しており、逆に治療効果が小さくなる場合は辺々を -1 倍すればよい。

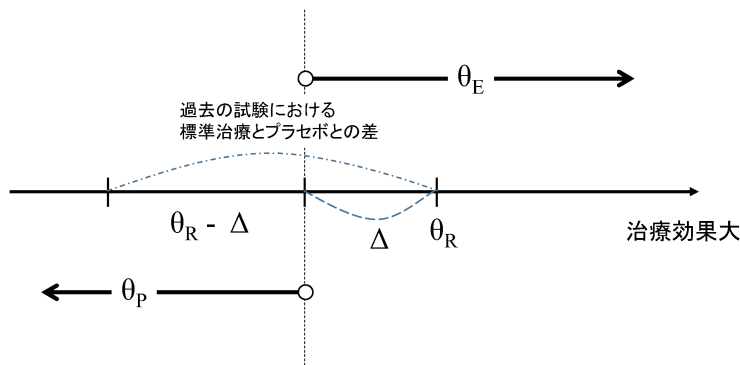


図 1 3 群間の大小関係

この大小関係から、分析感度を保証した下で試験治療の標準治療に対する非劣性を証明する仮説検定問題として

$$H_1 : \theta_E \leq \theta_R - \Delta \text{ v.s. } K_1 : \theta_E > \theta_R - \Delta \quad (2)$$

$$H_2 : \theta_R \leq \theta_P + \Delta \text{ v.s. } K_2 : \theta_R > \theta_P + \Delta \quad (3)$$

を同時に有意水準 α で帰無仮説 (H_1, H_2) を棄却することに帰着できる。つまり、仮説 (2) については通常の試験治療群と標準治療群の非劣性仮説であり、その群間差が $-\Delta$ より下回らないことを示す一方で、仮説 (3) は標準治療群はプラセボ群より上回るのではなく (単なる優越性の証明ではなく)、プラセボ群 $+\Delta$ より上回ることを示すことになる。ただし、パラメータ θ は、3 群非劣性試験における主要な評価項目の評価変数に応じて、母集団の平均値、比率、ハザードに対応するものとする。

したがって、帰無仮説 (H_1, H_2) の下でそれぞれの評価変数に準じ (漸近的に) 標準正規分布に従う 2 つの検定統計量 (T_1, T_2) を導出し,

$$T_1 > u_{\alpha/2}, \cap T_2 > u_{\alpha/2} \quad (4)$$

が成り立つときのみ、試験治療群の標準治療群に対する分析感度が保証された下での非劣性が証明されたと判断する。ただし、 $u_{\alpha/2}$ は標準正規分布 $N(0, 1)$ の上側 $100 \times \alpha/2\%$ 点とする。

§§3.2. 3 群非劣性試験における被験者数の設定

一方で、3 群非劣性試験における被験者数 (試験治療群 n_E , 対照治療群 n_R , プラセボ群 n_P) について、 $n_E : n_R : n_P = 1 : C_R : C_P$ ($0 < C_R, C_P \leq 1$) の関係が成り立つとすると、総被験者数は $N = n_E + n_R + n_P = (1 + C_R + C_P)n_E$ と表せる。

提案した 3 群非劣性試験における検定方針に対する検出力関数は,

$$\text{Power} = \Pr\{T_1 > u_{\alpha/2} \cap T_2 > u_{\alpha/2} \mid K_1, K_2\}. \quad (5)$$

3 群の被験者数 n_E, n_R, n_P が十分に大きいとき、検定統計量 (T_1, T_2) の同時分布は、漸近的に平均ベクトル μ , 分散共分散ベクトル Σ をもつ 2 変量正規分布 $N_2(\mu, \Sigma)$ に従う。ただし,

$$\mu = \begin{cases} (0, 0)^T, & \text{帰無仮説 } (H_1, H_2) \\ (\mu_1, \mu_2)^T, & \text{対立仮説 } (K_1, K_2) \end{cases} \quad (6)$$

であり、分散共分散ベクトルは $\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$ である。なお、 $\mu_1 = \mu_1(\theta_E, \theta_R, n_E, C_R, \Delta)$, $\mu_2 = \mu_2(\theta_R, \theta_P, n_E, C_R, C_P, \Delta)$ と、各群の治療効果を示すパラメータ θ , 被験者数 n (すなわち試験治療群の被験者数 n_E と割り付け比率 C_R, C_P) および非劣性マージン Δ の関数であり、相関係数 $\rho = \frac{\text{Cov}(T_1, T_2)}{\sqrt{\text{Var}(T_1)}\sqrt{\text{Var}(T_2)}}$ は、 $\rho(\theta_E, \theta_R, \theta_P, C_R, C_P)$ と被験者数 n_E に依存しない関数として表される。

したがって、提案する検定手法の検出関数から漸近的に検出力 $1 - \beta$ を達成するために必要な被験者数 $n_E, n_R = C_R n_E, n_P = C_P n_E$ は、帰無仮説 (K_1, K_2) の下で

$$1 - \beta = \Pr\{T_1 \geq u_{\alpha/2} \cap T_2 \geq u_{\alpha/2} \mid (T_1, T_2) \sim N_2(\mu_1, \mu_2, 1, 1, \rho)\} \quad (7)$$

を, n_E, C_R, C_P について解けばよい. なお, 2 変量正規分布の上側確率については,

$$\begin{aligned} L(h, k; \rho) &:= \Pr\{T_1 \geq h \cap T_2 \geq k\} = \int_k^\infty \int_h^\infty \phi_2(x, y; \rho) dx dy \\ &= \int_{-1}^\rho \phi_2(h, k; t) dt = \int_0^\rho \phi_2(h, k; t) dt + Q(h)Q(k), \end{aligned}$$

とする ([12], [13]). ただし, $\phi_2(x, y; \rho)$ は 2 変量正規分布 $N_2(\mathbf{0}, \Sigma)$ の確率密度関数であり, $\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$, $\phi_1(x)$ は標準正規分布の確率密度関数であり, $Q(h) := \int_h^\infty \phi_1(t) dt$ とする.

ただし, (7) 式から理論的に n_E, C_R, C_P の解を導出することは困難なため, 数値的に C_R と C_P の範囲 ($0 < C_R \leq 1, 0 < C_P \leq 1$) から (C_R, C_P) の組み合わせ (例, C_R, C_P を 0.01 刻みで 100 通り $\times$ 100 通り = 10,000 通り) を与えた下で n_E について解けばよい.

§§3.3. Optimal Allocation

実際の臨床試験における被験者数については, 倫理的に必要な最小限となる被験者を対象とすること, またプラセボ群に割り当てる被験者数も可能な限り最小限にしたいため,

1. 総被験者数 $N = n_E + n_R + n_P = n_E(1 + C_R + C_P)$ を最小にする
2. プラセボ群の被験者数 $n_P = C_P n_E$ を最小にする (C_P を最小にする)

を満たす, (C_R, C_P) の組み合わせを Optimal Allocation と定義する.

これまで, さまざまな評価変数およびパラメータの設定の下で数値的な検討を実行した結果, $C_R = 1$ の場合にほぼ上記の条件を満たす Optimal Allocation が得られた ([8], [9], [10]). つまり, 試験治療群と標準治療群の被験者数が同数 ($n_E = n_R$) であり, 分析感度を保証した下での非劣性を証明する目的からしても理に適っている.

§4 結論と考察

試験治療群, 標準治療群にプラセボ群を加えた 3 群非劣性試験デザインについては, 内部妥当性, 分析感度の評価が可能な非劣性デザインであることから各種のガイドラインで強く推奨されている. しかし, 通常の 2 群の比較にプラセボ群を含めることにより, 3 通りの対比較が可能となるため, 3 群非劣性試験を実施する目的や立場により, 対比較の優先順位や解釈が異なる. また, 3 群非劣性試験デザインに関する既存研究の多くが, 非劣性マージンを 3 群非劣性試験で観測される標準治療とプラセボの治療効果の差のある割合 (試験

の計画時点で事前に決定する割合)として設定しており、観測される群間差に応じて、 Δ としては変動してしまうことから、非劣性マージン Δ の定義からすると十分とは言えない。

そこで、本稿では特に主要評価項目の評価変数を一般化した記載にしたが、医学研究で一般的な連続型変数、2 値変数、生存時間データの場合について、事前にある一つの値として選択した非劣性マージン Δ を用いて、分析感度を保証した下での試験治療群の標準治療群に対する非劣性を証明するための解析手順と必要な被験者数の設定および各群への Optimal Allocation を提案し、評価変数に依らず、ほぼ試験治療群と標準治療群の必要被験者数が同じ場合に総被験者数とプラセボ群の被験者数が最小になることを示した。なお、各評価変数に対する 2 つの検定統計量 (T_1, T_2) 、2 つの対立仮説の下での漸近 2 変量正規分布の平均ベクトル (μ_1, μ_2) 、相関係数 ρ の導出の詳細、実際の数値的な検討結果と提案手法の妥当性については、Hida & Tango ([8], [9], [10]) を参考にされたい。

参考文献

- [1] ICH Harmonised Tripartite Guideline. Statistical Principles for Clinical Trials. 1998.
- [2] ICH Harmonised Tripartite Guideline. Choice of Control Group and Related Issues in Clinical Trials. 2000.
- [3] Guideline on the choice of the non-inferiority margin (Doc. Ref. EMEA/CPMP/EWP/2158/99). 2005.
- [4] FDA Guidance for Industry Non-Inferiority Clinical Trials. 2010.
- [5] Pigeot I, Schäfer J, Röhmel J, Hauschke D. Assessing non-inferiority of a new treatment in a three-arm clinical trial including a placebo. *Statistics in Medicine* 2003, **22**: 883–899.
- [6] Mielke M, Munk A and Schacht A. The assessment of non-inferiority in a gold standard design with censored, exponentially distributed endpoints. *Statistics in Medicine* 2008, **27**: 5093–5110.
- [7] Kombrink, K., Munk, A., Friede, T. Design and semiparametric analysis of non-inferiority trials with active and placebo control for censored time-to-event data. *Statistics in Medicine* 2013, **32**: 3055–3066.
- [8] Hida E. and Tango T. On the three-arm non-inferiority trial including a placebo with a prespecified margin. *Statistics in Medicine* 2011, **30**: 224–231.
- [9] Hida E. and Tango T. Three-arm noninferiority trials with a prespecified margin

for inference of the difference in the proportions of binary endpoints. *Journal of Biopharmaceutical Statistics* 2013, **23**: 774–789.

- [10] Hida E. and Tango T. Design and analysis of a three-arm non-inferiority trial with a prespecified margin for the hazard ratio. 2015, submitting.
- [11] Jung, S.H. *et al.* Sample size computation for two-sample noninferiority log-rank test. *Journal of Biopharmaceutical Statistics* 2005, **15**: 969–979.
- [12] Kotz S, Balakrishnan N, Johnson NL. *Continuous Multivariate Distributions Volume 1 : Models and Applications* (2nd edn). Wiley: New York, 2000.
- [13] 柴田義貞. 正規分布-特性と応用. 東京大学出版会. 1981.