臨床研究セミナー記録

日本臨床外科学会・日本外科学会共催（第79回日本臨床外科学会総会開催時）

第19回臨床研究セミナー

第1部 臨床研究の基礎講座

1. 臨床研究のための研究デザイン—非劣性試験：考え方と実臨床への導入—

横浜市立大学大学院医学研究科臨床統計学

山中 竹春

（2017年11月25日受付）

キーワード 非劣性試験, 優越性試験, 非劣性マージン

I. はじめに

本稿では外科手技や薬物療法の試験において汎用される非劣性試験の考え方と実臨床への導入について解説する。

II. 優越性試験と非劣性試験の統計学的な違い

優越性試験と非劣性試験の二つの違いを検定の面から説明する。検定の背景には帰無仮説と対立仮説が存在する。優越性試験（片側検定とする）では、

帰無仮説：新治療の効果 $\leq$ 対照治療の効果
（新治療は対照治療と同等または劣る）

対立仮説：新治療の効果 $>$ 対照治療の効果
（新治療は対照治療より優れる）

である。帰無仮説は棄却され、対立仮説が採用されることが期待される。一方、非劣性試験では、

帰無仮説：新治療の効果 $\leq$ 対照治療の効果 $- \Delta$
（新治療は対照治療より Δ を超えて劣る）

対立仮説：新治療の効果 $>$ 対照治療の効果 $- \Delta$
（新治療は対照治療より Δ を超えて劣らない）となる。この Δ を非劣性マージン（non-inferiority margin）と呼び、根拠をもって研究者が決定する。統計学的にはこの仮説の右辺の違いが、優越性試験と非劣性試験の違いである。信頼区間を用いて同じことを説明する。一般に、 $(100 - \square\square\%)$ 信頼区間とは「この区間が真値を含んでいる確率が $(100 - \square\square\%)$ である」という区間である。「有意水準 $\square\%$ の検定で有意」とは、 $(100 - \square\square\%)$ 信頼区間が帰無仮説で示される値を含まない、ということに同値である。例として、図1に点推定値と信頼区間を4通り示した。横軸がハザード比、三角が点推定値、実線が95%信頼区間である。優越性試験の場合、結果（a）は95%信頼区間が帰無仮説で示される値（HR=1）を含まないので、統計的に優れていると結論する。一方、結果（b）は、点推定値は1より小さいものの95%信頼区間が1を含んでいるため、新治療が統計的に優れているとは結論されない。非劣性試験は、この帰無仮説の値に関して非劣性マージン Δ 分だ

DESIGN OF CLINICAL TRIALS: NON - INFERIORITY DESIGN AND ITS APPLICATIONS -

Takeharu Yamanaka

Department of Biostatistics, Yokohama City School of Medicine, Yokohama, Japan

図1 試験の結果と検定のマージン

横軸がハザード比，三角が点推定値，実線が95% 信頼区間．結果が左に位置するほど（1を下回るほど）新治療が優れていることを表す．縦の破線は，左が優越性試験の帰無仮説，右が非劣性試験の帰無仮説を表す．

け緩めた検定を行う．95% 信頼区間全体がマージン Δ を下回れば帰無仮説を棄却し、「新治療の対照治療に対する非劣性が示された」と結論する．図1でいうと，(a)と(b)は非劣性が示された状況である．(c)は信頼区間がマージンを含んでいるので，非劣性は示されていない．(d)は非劣性が示された結果であるが，同時に有意に劣っている結果であり，臨床的な解釈は議論のあるケースとなる．

III. 非劣性試験におけるマージンの設定と妥当性

非劣性試験では非劣性マージンが重要なパラメータである．非劣性マージンの決定には，1)新治療の clinical importance に基づき専門家集団，あるいは規制当局を交えて協議する，2)対照治療 vs プラセボ（または無治療）の既存の試験で観測された群間差をもとに，当該試験において，プラセボ以上の効果が間接的に示されるような値に設定することが考えられる．米国規制当局のガイダンスには後者の方法の活用が色濃くでている．例えば，対照治療 vs プラセボ（または無治療）の既存の試験で観測された対照治療のプラセボに対する効果が $HR=0.8$ であった場合，対照治療からみたプラセボの効果は $M1=1.25 (=1 \div 0.8)$ となる．非劣性マージンを $M1$ とするとき，対照治療からみて $M1$ を超えて劣らないということは，プラセ

ボを超えて劣らないことを意味する．さらに対照治療の効果の何%が維持されるか？という観点から， $1.0 < M2 < M1$ となる値 $M2$ をマージンとし， $M1$ のうち残せた分を維持率 (retention rate) とする考え方がある． $M2$ を超えて劣らないということは，対照治療のプラセボに対する効果の半分以上を維持できることを意味する．マージンを厳しくしすぎると症例数が非現実的な数になるため実施可能性とのバランスを考える必要があるが，非劣性試験では「仮想プラセボに対して優越している」ことを保証するデザインとすることも重要である．

IV. 実施・解釈における注意点

割付と異なる治療を受けた，あるいはプロトコルに規定された治療計画を遵守しなかったなどの症例が多くいる場合に ITT 集団で解析を行うと，治療効果の差が本当にあっても観測される群間差は薄まり，群間差を見出す力（検出感度 assay sensitivity）は低下する．優越性試験では，検出感度が低下すると結果が保守的（有意差を言いにくい）になるため，ITT 集団での解析が主たる解析とされる．一方非劣性試験の場合は，検出感度が低下するとかえって非劣性を示しやすく（有意差を言いやすい）になってしまう．

非劣性試験では，対照治療群がプラセボ（または無治療）に比べてどれだけよいかを当該試験内

で検討することができない。マージン設定に参照した研究ほど対照治療の効果が無い、あるいは昔の研究を参照したが、その時点に比べて医療環境の改善等によりプラセボ（または無治療）の予後が良くなっていることがありうる。マージンの設定には過去の試験を質的に評価して、可能な限り類似した試験を複数用いることが望ましい。

優越性試験では、結果が有意ではなかったとき、非劣性を主張することはできない。帰無仮説を後付けで変更することになり、第Ⅰ種の過誤（エラー）が増大するためである。非劣性を結論するためには科学的に決定したマージンをもつ非劣性試験の

実施が必要で「優越性試験で有意差がなかった」と「非劣性が示された」ことは根本的に異なることに留意しなければならない。

V. おわりに

非劣性試験には優越性試験と比べて統計的な構造に違いはないが、試験を実施する上での留意点がいくつかある。正しい理解による実施と解釈が求められる。

利益相反：なし