

非劣性, 同等性, 優越性検証のための 多重決定方式による統合的接近法

広津千尋*

A Unifying Approach to Non-inferiority, Equivalence and Superiority Tests via Multiple Decision Processes

Chihiro Hirotsu

新薬と既承認標準薬との比較臨床試験において, 非劣性検証, 同等性検証, 優越性検証を一連の統計的多重決定方式として行う二つの考え方を提案する. その一つは母平均の差に関し, 信頼係数を0.95に保ったまま, 通常の単純な両側信頼区間方式による消費者危険率, および生産者危険率を改良する信頼領域方式である. この方式では, その下限が非劣性マージンを上回ると同時に, 区間内に0を含むことを要請するため, いたずらに例数を増やすことによって非劣性検証を達成しようとする行為は避けられる.

もう一方の閉手順方式では, partitioning principle によって両側検定, 片側検定を自然に統一すると共に, 非劣性検証と優越性検証を統合するいわゆる switching procedure を正当化する. とくに非劣性検証に関して, 旧厚生省ガイドライン (片側5%検定) と ICH E9 ガイドライン (片側2.5%検定) を同時に正当化する考え方を提案する. この方法は臨床試験の結果に応じて弛い非劣性から優越性まで, 被験薬の対照薬に対する有効性に関する相対的エビデンスの強さを変化させる魅力的な方法といえる.

Two approaches of multiple decision processes are proposed for unifying the non-inferiority, equivalence and superiority tests in a comparative clinical trial for a new drug against an active control. One is a method of confidence set with confidence coefficient 0.95 improving the consumer's and producer's risks of the usual approach of the naive two-sided confidence interval. The method requires to include 0 within the region as well as to clear the non-inferiority margin so that a trial with somewhat large number of subjects for proving non-inferiority of a drug which is actually inferior should be unsuccessful.

The other is the closed testing procedure, which combines the one-and two-sided tests by applying the partitioning principle and justifies the switching procedure unifying the non-inferiority, equivalence and superiority tests. In particular regarding the non-inferiority the proposed method justifies simultaneously the old Japanese Statistical Guideline (one-sided 0.05 test, Kouseisho, 1992) and the ICH E9 Guideline (two-sided 0.05 test, Lewis, 1999). It should be very attractive to have such multiple decisions changing the strength of the evidence of the relative efficacy of the test drug against the control according to the achievement of the clinical trial.

* 明星大学理工学部 〒191-8506 東京都日野市程久保2丁目1-1
E-mail: hirotsu@ge.meisei-u.ac.jp

1. はじめに

2 薬剤の比較臨床試験において固有な興味ある話題として、非劣性検証、同等性検証、優越性検証の問題がある。この問題には2通りの設定があり、その一つは例えば錠剤から粉末、顆粒のように剤型変更した際のバイオアベイラビリティ (bio-availability) の同等性を保証する場合である。これは生物学的同等性の検証と呼ばれる。この場合、特性値としては薬剤投与後の最大血中濃度 (Cmax) や濃度推移曲線下面積 (Area Under the Curve) 等が採られる。その特性値に関し2剤の母平均を μ_1, μ_2 として、

$$0.80 \leq \frac{\mu_1}{\mu_2} \leq 1.25 \Leftrightarrow -0.223 \leq \log \mu_1 - \log \mu_2 \leq 0.223 \quad (1.1)$$

が満たされるとき、2剤は同等と見なされる。

これを検証する統計的方法としては、通常、 $\log \mu_1 - \log \mu_2$ に対する信頼係数 0.90 の信頼区間が (1.1) 式の内側に収まることを検証する両側信頼区間方式が採られている。この問題に対し広津 (2004) では μ_1/μ_2 が (1.1) 式の範囲内でないときにこれを同等と見なす消費者危険率を、従来方式同様片側 0.05 に保ちつつ、 $\mu_1/\mu_2=1$ が真のときに、 μ_1/μ_2 を領域内に含む確率を 0.95 に高める多重決定方程式に基づく信頼領域を提唱している。

もう一つの設定は新薬 (以降被験薬という) の既承認標準薬 (以降対照薬という) に対する非劣性検証である。薬剤の特性は多面的であり、被験薬が対照薬に対し例えば安全性、投与の簡便性等に特長があれば、こと有効性に関しては同等であればよいという考え方は至極合理的とされる。しかしながら、同等性を検証する適切な統計的方法が存在しないため、長らく許認可の現場では同等性の帰無仮説が棄却されないとき同等とみなすいわゆる NS (Non-significance) 同等が採用されていた。しかしそれが不合理なことは、症例数が少なく、精度の低い試験ほど同等性を示すのが容易となることから明らかである。また、一度有効性の不確かな薬剤が許可されると、それを対照として次々と有用でない薬剤が許可される恐れが否定出来ない。実際、1995年にプラセボを対照とした再試験の結果、5種類ものよく知られた脳代謝改善薬が一斉に市場から姿を消したことはまだ記憶に新しい。そこで、より合理的な方法として、被験薬が対照薬に対しある範囲 δ を越えて劣らないことを検証する非劣性検定方式が導入された (例えば広津, 1986 参照)。この方式は正規分布モデルの場合、被験薬の母平均を μ_1 、対照薬のそれを μ_2 として

$$\text{帰無仮説 } H_0: \mu_1 - \mu_2 = -\delta$$

を

$$\text{対立仮説 } H_1: \mu_1 - \mu_2 > -\delta$$

に対し、片側有意水準 α で検定し、 H_0 が棄却された時に非劣性が検証されたとする方式であり、 $\delta (>0)$ は予め定められた非劣性マージンである。ここでは特性値の大きい方が好ましいと設定しているが、 μ_1 に δ を上乗せした後 μ_2 との帰無仮説を検定するために、ハンディキャップ方式とも呼ばれる。この方式では H_1 の棄却を要するため、想定する対立仮説の下での検出力を確保するようしかるべき症例数と精度の高い試験が要求される。この方式は厚生省ガイドライン (1992) で採用され、通常のように $\alpha=0.05$ とされた。一方、1999年に発行された国際統計ガイドライン ICH E9 (Lewis, 1999) でも同様の方式が採用されたが、 α は 0.025 と変更された。この変更は、対照薬と真に同等の被験薬の非劣性を検証するのに、従来に比べ約

1.3 倍の症例数を必要とすることになるので, 慎重に考察されるべき事柄である. 真に同等の場合と同じ症例数だと, 非劣性検証が不成功となる生産者危険率は, 通常の設定で約 0.1 増大する. 一方, この非劣性検証方式に関連して, 通常の

$$\text{帰無仮説 } H_0: \mu_1 - \mu_2 = 0$$

を

$$\text{対立仮説 } H_1: \mu_1 - \mu_2 \neq 0$$

に対して検定する優越性検証方式への switching procedure に興味を持たれる, 例えば Committee for Proprietary Products (2000) 参照.

多重性問題を解消しつつ H_0 および H_0' を同時に検定する方法の妥当性については, Morikawa & Yoshida (1995), Hirotsu & Hothorn (2003) などで考察されている. とくに Morikawa & Yoshida (1995) では, すべての検定を有意水準 α で行う閉手順方式 (Marcus, Peritz & Galeriel, 1976) として,

- (1) 優越性片側検定, 非劣性片側検定,
- (2) 優越性両側検定, 非劣性片側検定,
- (3) 優越性両側検定, 非劣性両側検定,

の組み合わせを正当化している. このうち (2) が旧厚生省ガイドラインの場合に当たる. しかしながら優越性両側検定を有意水準 α , 非劣性片側検定を $\alpha/2$ で行う ICH E9 の正当化は含まれていない. 今のところ ICH E9 は通常の信頼係数 0.95 の両側信頼区間の下限が 0 を越えるとき優越性, $-\delta$ を越えるとき非劣性が言えたとする一貫性で正当化されている. しかしながら, 信頼係数を一定とする時, 単純な両側信頼区間は決して効率がよくなく, その様々な改良が提案されている (例えば, Hayter & Hsu, 1994). そこで本論第 2 節では, 本節の最初に生物学的同等性に関連して述べた多重決定方式に基づく信頼領域をこの switching procedure の問題に適用する. 次に第 3 節では, 旧厚生省ガイドラインと ICH E9 を同時に正当化する閉手順検定方式を提案する. 基本的に正規分布モデルについて述べるが, 第 4 節で 2 項分布モデルの場合を考察し, 5 節にその例を与える.

2. 信頼領域方式による統合

観測値 y_{ij} が正規分布 $N(\mu_i, \sigma^2)$, $i=1, 2; j=1, \dots, n_i$ に従っているとす. ただし, $i=1$ を被験薬, $i=2$ を対照薬とする. このとき,

$$T_\alpha = \left(\frac{1}{n_1} + \frac{1}{n_2} \right)^{\frac{1}{2}} \hat{\sigma} t_\alpha (n_1 + n_2 - 2)$$

として, 次のような信頼係数 0.95 の信頼領域が得られる (広津, 2004, 図 4.3 参照). ここで, $\hat{\sigma}$ は通常の不偏分散

$$\hat{\sigma}^2 = \sum_i \sum_j (y_{ij} - \bar{y}_i)^2 / (n_1 + n_2 - 2)$$

の平方根, $t_\alpha(\nu)$ は自由度 ν の t 分布の上側 α 点, そして \bar{y}_i , $i=1, 2$ は標本平均である.

信頼係数 0.95 の信頼領域

$$\begin{aligned} \bar{y}_1 - \bar{y}_2 > T_{0.025} \text{ のとき, } 0 < \mu_1 - \mu_2 < \bar{y}_1 - \bar{y}_2 + T_{0.05}, \\ T_{0.05} \leq \bar{y}_1 - \bar{y}_2 \leq T_{0.025} \text{ のとき, } 0 \leq \mu_1 - \mu_2 < \bar{y}_1 - \bar{y}_2 + T_{0.05}, \\ -T_{0.05} < \bar{y}_1 - \bar{y}_2 < T_{0.05} \text{ のとき, } \bar{y}_1 - \bar{y}_2 - T_{0.05} < \mu_1 - \mu_2 < \bar{y}_1 - \bar{y}_2 + T_{0.05}, \\ -T_{0.025} \leq \bar{y}_1 - \bar{y}_2 \leq T_{0.05} \text{ のとき, } \bar{y}_1 - \bar{y}_2 - T_{0.05} < \mu_1 - \mu_2 \leq 0, \\ \bar{y}_1 - \bar{y}_2 < T_{0.025} \text{ のとき, } \bar{y}_1 - \bar{y}_2 - T_{0.05} < \mu_1 - \mu_2 < 0 \end{aligned}$$

この方式による結論は,

$$\left. \begin{aligned} \bar{y}_1 - \bar{y}_2 > T_{0.025} \text{ (両側 5\% 検定で有意) のとき, 被験薬は対照薬に対し優越:} \\ \mu_1 - \mu_2 > 0, \\ T_{0.05} \leq \bar{y}_1 - \bar{y}_2 \leq T_{0.025} \text{ のとき, 同等以上: } \mu_1 - \mu_2 \geq 0, \\ (-T_{0.025} \leq \bar{y}_1 - \bar{y}_2 \leq T_{0.025}) \cap (-\delta \leq \bar{y}_1 - \bar{y}_2 - T_{0.05}) \text{ のとき, 非劣性: } \mu_1 - \mu_2 > -\delta \end{aligned} \right\} \quad (2.1)$$

が言えたとされる。この方式だと、被験薬が対照薬と真に同等の時に非劣性を受け容れる確率(検出力)は、本質的に片側有意水準 0.05 検定と同じで、通常見られる例数では片側有意水準 0.025 検定より 0.1 程度大きくなる。つまり生産者危険率は有意水準 0.05 検定並となる。一方、例えば、 $\bar{y}_1 - \bar{y}_2 - T_{0.05} \geq -\delta$ であっても、 $\bar{y}_1 - \bar{y}_2 < -T_{0.025}$ だと $\mu_1 - \mu_2$ の信頼区間が完全に 0 を下回るので非劣性が検証されたとはされないことは重要である。つまり、非劣性マージン δ を伴う非劣性検証には、 $-\delta < \mu_1 - \mu_2 < 0$ のときに、理論的には例数を大きくしさえすれば非劣性マージンをクリアできるという難点があるが、本方式では

$$\bar{y}_1 - \bar{y}_2 \geq -T_{0.025} \quad (2.2)$$

を要求することにより、その難点を克服している。この場合、例えば $n_1 = n_2 = 300$ もあると、よく見られる非劣性マージン $\delta/\sigma = 1/3$ の場合に、実際に $\mu_1 - \mu_2 = -\delta$ である試験で非劣性を受け容れる危険率は 0.017 となり、単純な片側危険率 0.05 方式はもとより、同 0.025 方式をも改良する。なお、 δ/σ を大きく 1/2 に取ると $n = 124$ で危険率が 0.025 以下になる。すなわち非劣性マージンを大きく取っても、劣った薬の非劣性が言えてしまう危険率は低く押さえられる方式となっている。通常、非劣性マージンをクリアできても、ハンディキャップ無し検定で有意に負けては意味がないと定性的に言われているが、本方式はそれを統計的方法で定式化したものと言える。なお、この意味で、 δ を臨床的に許容出来る最大の差と定義することには疑問があるが、本論で詳細に論じることはしない。

3. 閉手順検定方式による統合

第 2 節と同様に正規分布モデルを仮定し、同じ記号を用いる。まず母数空間を

$$H_1: \mu_1 - \mu_2 < 0,$$

$$H_2: \mu_1 - \mu_2 = 0,$$

$$H_3: \mu_1 - \mu_2 > 0,$$

と 3 分割し、さらに

$$H_{\delta 1}: \mu_1 - \mu_2 < -\delta,$$

$$H_{\delta 2}: \mu_1 - \mu_2 = -\delta,$$

という 2 領域を考える。ここで、 $H_{\delta 1}$ と $H_{\delta 2}$ が排反的で、かつともに H_1 に含まれる、すなわち、

$$H_{\delta 1} \cap H_{\delta 2} = \phi \text{ (空集合)}, H_1 \supset H_{\delta 1}, H_1 \supset H_{\delta 2}$$

に注意する. さらに,

$$H_1 \cap H_2 = H_{\delta 1} \cap H_2 = H_{\delta 2} \cap H_2 = \phi$$

である. そこで, Marcus, Peritz & Gabriel (1976) の閉手順方式に従い, H_1 の前にまず $H_{\delta 1}$, $H_{\delta 2}$ の検定を行うが, これらは排反的なので有意水準の調整は要しない. すなわち, まず $H_{\delta 1}$ に右片側 α 検定を適用し, それが棄却されないなら検定はここで終了し, 非劣性は検証されない. $H_{\delta 1}$ が棄却されたなら, 次に $H_{\delta 2}$ に両側 α 検定を適用する. もしそれが棄却されないなら, 検定をここで終了し, 弱い非劣性 ($\mu_1 - \mu_2 \geq -\delta$) のみ言えたとする. $H_{\delta 2}$ も棄却されたなら, H_1 の右片側 α 検定に進む. もしそれが棄却されないなら, 強い非劣性 ($\mu_1 - \mu_2 > -\delta$) が言えたとする. H_1 が棄却されたなら H_2 の両側 α 検定に進み, それが棄却されないとき同等以上 ($\mu_1 - \mu_2 \geq 0$), 棄却されたとき優越性 ($\mu_1 - \mu_2 > 0$) が言えたとする. この検定方式の危険率は α である. 通常のように $\alpha = 0.05$ として以上の手順をまとめると, 次のようになる.

- ステップ 1. $\bar{y}_1 - \bar{y}_2 - T_{0.05} < -\delta$ のとき, $\mu_1 - \mu_2 < \delta$ (被験薬の対照薬に対する非劣性は検証されない).
- ステップ 2. $\bar{y}_1 - \bar{y}_2 - T_{0.025} \leq -\delta \leq \bar{y}_1 - \bar{y}_2 - T_{0.05}$ のとき, $\mu_1 - \mu_2 \geq -\delta$ (被験薬は対照薬に対し弱い非劣性).
- ステップ 3. $(\bar{y}_1 - \bar{y}_2 < T_{0.05}) \cap (\bar{y}_1 - \bar{y}_2 - T_{0.025} > -\delta)$ のとき, $\mu_1 - \mu_2 > -\delta$ (強い非劣性).
- ステップ 4. $T_{0.05} \leq \bar{y}_1 - \bar{y}_2 \leq T_{0.025}$ のとき, $\mu_1 - \mu_2 \geq 0$ (同等以上).
- ステップ 5. $\bar{y}_1 - \bar{y}_2 > T_{0.025}$ のとき, $\mu_1 - \mu_2 > 0$ (優越性).

この方式は ICH E9 の非劣性検証をステップ 3, 優越性検証をステップ 5 として含んでいる. また, 非劣性検証を精密化し, 弱い非劣性 (ステップ 2) として旧厚生省ガイドライン方式を含んでいる. さらにハンディキャップ無しの片側検定に応じた同等以上 (ステップ 4) を含んでいる. すなわちここで提案した方法は, 竹内 (1973) による片側検定, 両側検定の統合方式を活用して, 強弱の非劣性, 同等性, 優越性を危険率 0.05 の一つの多重決定方式にまとめ, 旧厚生省ガイドラインと ICH E9 ガイドラインを同時に正当化している. なお, 例数や δ の取り方によっては $T_{\alpha/2} - \delta > T_\alpha$ ということがあるかも知れない. しかし, その場合は単にステップ 3 の事象が起り得ないだけで, この手順に本質的な影響はない.

4. 閉手順方式の 2 項分布モデルへの適用

非劣性検証は有効率比較を目的として 2 項分布モデルで用いられることも多い. いま, 被験薬および対照薬に対する観測値 y_i , $i=1, 2$, が 2 項分布 $B(n_i, p_i)$ に従っているとす. 前節同様に, まず母数空間を

$$H_1: p_1 - p_2 < 0,$$

$$H_2: p_1 - p_2 = 0,$$

$$H_3: p_1 - p_2 > 0,$$

と 3 分割し, さらに

$$H_{\delta 1}: p_1 - p_2 < -\delta,$$

$$H_{\delta 2}: p_1 - p_2 = -\delta,$$

を定義する. ここで症例数 n_1, n_2 は十分大きく, 正規近似が使えるものとする. このとき, 前節に無かった難点は, 通常の場合 H_i の検定とハンディキャップ仮説 $H_{\delta i}$ の検定で, 統計量の規準化に用いる分散が異なることである. すなわち, H_i の検定では規準化統計量

$$\frac{y_1/n_1 - y_2/n_2}{\sqrt{\hat{V}_0}},$$

$$\hat{V}_0 = \left(\frac{1}{n_1} + \frac{1}{n_2} \right) \frac{y_1 + y_2}{n_1 + n_2} \left(1 - \frac{y_1 + y_2}{n_1 + n_2} \right)$$

を用いるのに対し, $H_{\delta i}$ の検定では規準化統計量

$$\frac{y_1/n_1 + \delta - y_2/n_2}{\sqrt{\hat{V}_1}},$$

$$\hat{V}_1 = \frac{1}{n_1} (\hat{p}_2 - \delta)(1 - \hat{p}_2 + \delta) + \frac{1}{n_2} \hat{p}_2(1 - \hat{p}_2)$$

が用いられる. ここで \hat{p}_2 は $H_{\delta 2}$ の下での確率分布

$$\binom{n_1}{y_1} ((p_2 - \delta)^{y_1} (1 - p_2 + \delta)^{n_1 - y_1}) \binom{n_2}{y_2} p_2^{y_2} (1 - p_2)^{n_2 - y_2}$$

に対する最尤推定量であるが, n_1 と n_2 がそれ程違わないときは

$$\hat{p}_2 = \frac{y_1 + y_2 + n_1 \delta}{n_1 + n_2}$$

を用いることができる (Dunnett & Gent, 1977).

ここで幸いなことに, n_1, n_2 が 50 以上でほぼ等しく, かつ $\delta = 0.1$ 程度のときに, 2 通りの標準偏差 $\sqrt{\hat{V}_0}, \sqrt{\hat{V}_1}$ の相対誤差は近似的に

$$\frac{|\sqrt{\hat{V}_0} - \sqrt{\hat{V}_1}|}{\sqrt{\hat{V}_1}} = \frac{\delta}{8\hat{p}(1-\hat{p})} \left\{ 2(2\hat{p}-1) \frac{|n_1 - n_2|}{n} + \delta \right\} \quad (4.1)$$

で与えられ (広津 2004), 通常の臨床試験に当てはめると 0.02 程度である. ただし,

$$n = \frac{n_1 + n_2}{2}, \quad \hat{p} = \frac{y_1 + y_2}{2n}$$

である. このため近似的には H_i および $H_{\delta i}$ の検定に共通に \hat{V}_0 を用いることができ, 次のような多重決定方式が得られる. ここで, 簡単のために

$$T_\alpha = \sqrt{\hat{V}_0} K_\alpha, \quad K_\alpha \text{ は標準正規分布の上側 } \alpha \text{ 点}$$

と置いた.

ステップ 1. $\frac{y_1}{n_1} - \frac{y_2}{n_2} - T_{0.05} < -\delta$ のとき, $p_1 - p_2 < -\delta$ (被験薬の対照薬に対する非劣性は検証されない).

ステップ 2. $\frac{y_1}{n_1} - \frac{y_2}{n_2} - T_{0.025} \leq -\delta \leq \frac{y_1}{n_1} - \frac{y_2}{n_2} - T_{0.05}$ のとき, $p_1 - p_2 \geq -\delta$ (被験薬は対照薬に対し弱い非劣性).

ステップ 3. $\left(\frac{y_1}{n_1} - \frac{y_2}{n_2} < T_{0.05} \right) \cap \left(\frac{y_1}{n_1} - \frac{y_2}{n_2} - T_{0.025} > -\delta \right)$ のとき, $p_1 - p_2 > -\delta$ (強い非劣性).

ステップ 4. $T_{0.05} \leq \frac{y_1}{n_1} - \frac{y_2}{n_2} \leq T_{0.025}$ のとき, $p_1 - p_2 \geq 0$ (同等以上).

ステップ 5. $\frac{y_1}{n_1} - \frac{y_2}{n_2} > T_{0.025}$ のとき, $p_1 - p_2 > 0$ (優越性).

5. 例 題

表 1 の有効性データは慢性蕁麻疹に対する第 III 層 2 重盲検無作為化試験の結果 (Kawashima, Kukita, Harada et al., 2002) である. この例では

$$\hat{V}_0 = \left(\frac{1}{128} + \frac{1}{124} \right) \times \frac{180}{252} \times \frac{72}{252} = 0.00324$$

であり,

$$T_{0.025} = \sqrt{\hat{V}_0} \times K_{0.025} = 0.111$$

となる. 標本有効率差は

$$\frac{y_1}{n_1} - \frac{y_2}{n_2} = \frac{103}{128} - \frac{77}{124} = 0.184$$

と得られ, これが $T_{0.025}$ より大きいので被験薬の対照薬に対する優越性が言える.

表 1 慢性蕁麻疹に対する第 III 相試験の結果

	改良	不安・悪化	計
被験薬	103	25	128
対 照	77	47	124

なお, この例では標本有効率に大きな差 0.184 があるが, それでも $\hat{V}_1 = 0.00308$ であり, 相対誤差 (4.1) は 0.026 程度である.

6. 考 察

かつて厚生省新薬許認可の現場でまかり通っていた, 同等性の帰無仮説が棄却されないとき同等とみなす, いわゆる NS 同等は, 厚生省ガイドライン (1992) において非劣性検証 (ハンディキャップ方式) に置き替えられた. ここでは, 優越性検証両側有意水準 0.05 検定に対し, 非劣性検証は片側有意水準 0.05 検定として導入されている. この方式は, Morikawa and Yoshida (1995) で議論されているように, 非劣性が検証されたときのみ優越性検証を行うという閉手順検定方式として正当化される. 厚生省ガイドライン (1992) では, 有効率比較の場合に, 併せて, 非劣性マージンの一つの目安として $\delta = 0.1$ が示されている. これは有効率が極端でないとき (例えば 0.6~0.8 のとき), 1 群 100 例程度の試験で標本有効率差がほぼ 0 のときに非劣性が言えるという意味で合理的と考えられていた (広津, 2004).

一方, 1998 年に発行された ICH E9 ガイドライン (Lewis, 1999) では優越性検証両側有意水準 0.05, 非劣性検証片側有意水準 0.025 と改められ, 序論で述べたように, 対照薬と真に同等の被験薬の非劣性を検証するのに従来より約 1.3 倍の症例数を要することとなった. 本来, 非劣性検証の有意水準 α と非劣性マージン δ は関連しており, その方式を長く続けた時, 平均的にどういう有効性の新薬を世に送り出すことになるかという視点から決められるべきもの

である。しかしながら、これを正式に論ずるのは難しく、とりあえず ICH E9 の妥当性は、非劣性検証と優越性検証共通に単純な信頼係数 0.95 の両側信頼区間を用い、その下限が $-\delta$ または 0 を越えることで、それぞれ検証されることとされている。しかしながら、ある母数に関する信頼係数 0.95 の信頼領域は目的に応じていろいろに構成され、単純な両側信頼区間に限る必然性はない。

本論の第 2 説で述べた方法は、生産者危険率（あるいは検出力）を片側有意水準 0.05 検定並に保ったまま、ある程度大きな試験で消費者危険率を片側有意水準 0.025 検定より改良する。すなわち、検出力は旧厚生省ガイドライン並、危険率は設定によって ICH E9 並という、両者の利点を併せ持つ興味ある方式となっている。一方、第 3 節、第 4 節で述べた多重決定方式は危険率 0.05 を保ったまま、旧厚生省ガイドラインと ICH E9 ガイドラインを自然に統合・包括し、臨床試験で得られた被験薬の対照薬に対する相対的有効性のエビデンスに応じて、弱い非劣性、強い非劣性、同等以上、優越性と主張の強さを変更する魅力的な方法と言える。いわゆる 0, 1 型の検定ではなく、データの解釈を与えている。この 3 節の方法による非劣性検証において、2 節と同様に (2.2) 式を課すことは保守的な変更である。そうすることによって強弱非劣性検証の検出力は実質的に変わらず、消費者危険率が改善される。

なお、非劣性検証は通常、真に同等以上であるという見積りの下に計画される。すなわち、実際に δ 近く劣っているものをパスさせようという方式ではなく、やはり真に同等以上のものをパスさせようという方式であると考えられる。安全性にとくに優れる薬剤については、別の考慮が必要であろう。

参 考 文 献

- Committee on Proprietary Medical Products (2000). Points to consider on switching and non-inferiority. Committee on Proprietary Medical Products, London.
- Dunnett, C. W. and Gent, M. (1977). Significance testing to establish equivalence between treatments with special reference to data in the form of 2×2 tables. *Biometrics* 33, 593-602.
- Hayter, A. J. and Hsu, J. C. (1994). On the relationship between stepwise decision procedures and confidence sets. *J. Amer. Statist. Assoc.* 85, 778-785.
- 広津千尋 (1986). 臨床試験における統計的諸問題(1)—同等性検定を中心して—. *臨床評価* 14, 467-475.
- 広津千尋 (2004). 医学・薬学データの統計解析—データの整理から交互作用多重比較まで—. 東京大学出版会, 東京.
- Hirotsu, C. and Hothorn, L. (2003). Impact of the ICH E9 Guideline : Statistical principles for clinical trials on the conduct of clinical trials in Japan. *Drug Information J.* 37, 381-395.
- Kawashima, M., Kukita, A., Harada, S. et al. (2002). The efficacy and safety of SCH 9851 (Loratadine), a long-acting and selective histamine H_1 receptor antagonist, for chronic urticaria in a phase III, randomized, double-blind, multi-center comparative study. —Investigation of superiority of SCH 29851 10 mg once daily to SCH 29851 1 mg once daily and of non-inferiority of SCH 29851 10 mg once daily to ketotifen fumarate 1mg twice daily—. *Clinical Medicine* 18, 881-905.
- 厚生省 (1992). 臨床試験のための統計ガイドライン. 厚生省, 東京.
- Lewis, J. A. (1999). Statistical principles for clinical trials (ICH E9): an introductory note on an international guideline. *Statistics in Medicine* 18, 1903-1942.
- Marcus, R., Peritz, E. and Gabriel, K. R. (1976). On closed testing procedures with special reference to ordered analysis of variance. *Biometrika* 63, 655-660.
- 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.
- 竹内 啓 (1973). 数理統計学的方法的基礎. 東洋経済新報社, 東京.