

On multiple comparison procedures for mean vectors and partial covariance matrices in clinical trials

(臨床試験における平均ベクトルおよび部分共分散行列の多重比較法について)

平成 27 年 2 月 6 日

理学研究科 数理情報科学専攻
瀬尾研究室 中水流 嘉臣

薬剤の開発において重要な課題の一つは、ある疾患に関して開発している薬剤が同じ疾患に関する他の薬剤に対して長所を持っているとき、データを用いてそれを客観的に示すことである。薬剤を投与したときの効果は、個体間および個体内でバラツキを伴うため、有効性と安全性を確立するための方法論として、我々は統計学を用いている。

臨床試験のための統計的原則 (ICH E9) は、日米欧の 3 極間で、承認申請のための臨床試験に適用する統計的方法論の原則の調和を進めるために作成され、合意したものである [Lewis et al. (1999)]. この ICH E9 ガイドラインは、主として、臨床試験において偏りを最小にし、かつ精度を最大にすることを目的としており、対象集団、評価変数、盲検化などデザインで考慮すべきことから、解析対象集団、欠測値などのデータ解析上で考慮すべきことまで、広範にわたり統計的原則を記している。我々はこの原則に則り、臨床試験から妥当性のある結論を導くために、適切に統計学を用いて、被験薬の評価を実施しなければならない。

ICH E9 ガイドラインの中から特に主要評価変数に関する部分に注目してみると、『主要変数 (「目標」変数、主要評価項目ともいう) は、試験の主要な目的に直結した臨床的に最も適切で説得力のある証拠を与える変数であるべきである。主要変数は通常ただ一つにすべきである。』とある。このように主要変数は一つに定めるのが原則ではあるが、現実には主要な目的に関する複数の変数の中から、主要変数の一つだけ選ぶことができない場合が存在する [Koch and Gansky (1996) など]. そのために、ICH E9 ガイドラインは、主要変数の一つに選ぶことができない場合の方法として、合成変数、総合評価変数、複数の主要変数を用いる際の原則、および長所・短所を明確に記載している。

また、ICH E9 ガイドラインでは副次変数について『主要な目的に関連した補足的な測定値又は副次目的に関連した効果の測定値のどちらかである。治験実施計画書において副次変数を事前に定義し、試験結果の解釈の際に副次変数が果たす相対的な重要性と役割を説明することも重要である。』と記載している。医薬品の開発にあたり、主要な目的である薬効が実対照薬と比較して一定以上劣らないことを示した後、副次的に、安全性がより優れている、効果のバラツキが少なく安定した薬効が期待できるなどの特長を客観的に示すことも医療への貢献という観点から重要である。

本研究では、この状況で二つの課題に焦点を定めた。すなわち、1. 主要評価変数が複数あり、変数の少なくとも一つについては被験薬が対照薬より優れていて、他の変数については一定以上劣っていない（非劣性）ことを確認すること、2. 主要評価変数のいずれかで被験薬のバラツキ（分散）が対照薬のバラツキより小さいことを確認することである。これらの確認には、統計手法として、仮説検定の手法を用いることが考えられるが、どちらの課題でも、第一種の過誤を目標有意水準以下にするという条件下で、できるだけ検出力が大きい仮説検定法を用いることが必要である。そこで本研究では、上記二つの課題について、検出力の大きい仮説検定法を 2 標本問題の枠組みで考案することを試みた。以下にその試みの成果を述べる。

課題 1 は、主要評価変数が複数あり、変数の少なくとも一つについては被験薬が対照薬より優れていて、他の変数については一定以上劣っていない（非劣性である）ことを確認したい場合の検定方法である。仮説を以下に示す。

$$\begin{aligned} \text{帰無仮説 } H_0 &: \left\{ \max_{1 \leq j \leq p} \mu_j \leq 0 \right\} \cup \left\{ \min_{1 \leq j \leq p} (\mu_j + \varepsilon_j) \leq 0 \right\} \\ &\equiv H_0^{(0)} \cup \left\{ H_0^{(1)} \cup \dots \cup H_0^{(p)} \right\} \end{aligned}$$

$$\text{対立仮説 } H_1 : \text{not } H_0 = \text{not } H_0^{(0)} \cap \left\{ \text{not } H_0^{(1)} \cap \dots \cap \text{not } H_0^{(p)} \right\}$$

ただし

$$H_0^{(j)} : \mu_j \leq -\varepsilon_j, j = 1, 2, \dots, p$$

j : j 番目の主要評価変数

μ_j : 主要評価変数 j の 2 群の母平均の差

ε_j : 事前に決めた正の値 (非劣性マージン)

である。

この仮説に対する検定方法については、Bloch et al. (2001)らの検定を改良した Perlman and Wu (2004)によるインターセクション・ユニオン検定 (IUT) が提案されている。しかしながら、Perlman and Wu (2004)の方法は、IUT に用いる片側尤度比検定において、未知の共分散行列が局外母数であるため検定統計量や帰無仮説の下での理論分布の計算が複雑で難しいことが知られており、実際的ではない。IUT に用いる片側尤度比検定の代替としては、Glimm et al. (2002)の近似片側尤度比検定などが考えられるが、Glimm et al. (2002)の近似片側尤度比検定は、帰無仮説を原点、対立仮説をすべて正の空間としていることから、主要評価変数間に正の相関があり、主要評価変数のいずれかが負であるような場合に、第一種の過誤確率を目標有意水準以下に制御できないという問題がある。そこで、本研究では、Glimm et al. (2002)の近似片側尤度比検定に対し、主要評価変数間に正の相関があり、主要評価変数のいずれかが負であるような場合の好ましくない性質を改善できるような修正をおこなった検定を開発し、それを IUT に用いることにより、簡便でより保守的な検定を与えた。モンテカルロシミュレーションにより、提案する検定は、主要評価変数が2つである場合は正の相関が0.8未滿, 3つである場合は正の相関が0.6未滿であれば、棄却域が一樣でないにも関わらず、第一種の過誤確率を目標有意水準内に制御したうえで、Perlman and Wu (2004)と比較して、検出力を向上できることを確認した。Offen et al. (2007), Sankoh et al. (1999)によると、複数の主要評価変数を持つ臨床試験において、過去の経験より、変数間の相関は0.4の近傍が多く、相関係数は0.2~0.8の範囲であるとしている。これらを鑑みると、提案した方法は実際的であると考えられる。

課題 2 は、主要評価変数のいずれかで被験薬のバラツキ (分散) が対照薬のバラツキより小さいことを確認したい場合の検定方法および区間推定である。分散、共分散行列に関する均一性の検定や多重比較、同時信頼区間について、1変量における k 標本問題については、任意の2標本の分散の対比較、および対照比較に対する同時信頼区間が、Hartley (1950)により、ユニオン・インターセクションの原理に基づいて議論されている。また、O'Brien 検定、Brown-Forsythe 検定、Levene 検定、Bartlett 検定も等分散性の検定として利用可能である [Proust (2012)]。多変量の場合には、2標本問題を含む k 標本問題における分散共分散行列の均一性を評価するための尤度比検定さらには修正尤度比検定などが与えられている [Nagao (1973), Muirhead (1982) など]。このように分散、共分散行列に関する均一性の検定や多重比較、同時信頼区間はさまざまな方法が提案されているが、本研究では、Siotani, Hayakawa and Fujikoshi (1985)で提案されている以下の仮説に対する検定に注

目する. なお, ここでも課題 1 と同様に 2 標本問題の枠組みで議論する.

$$\text{帰無仮説 } H_0 : \frac{\mathbf{a}'\Sigma_1\mathbf{a}}{\mathbf{a}'\Sigma_2\mathbf{a}} = 1, \text{ for all non - zero vector } \mathbf{a}$$

$$\text{対立仮説 } H_1 : \frac{\mathbf{a}'\Sigma_1\mathbf{a}}{\mathbf{a}'\Sigma_2\mathbf{a}} \neq 1, \text{ for at least one non - zero vector } \mathbf{a}$$

ただし

Σ_1, Σ_2 : 第1群, 第2群の母共分散行列 ($p \times p$)

\mathbf{a} : 任意のベクトル ($p \times 1$)

である.

この仮説検定は, \mathbf{a} の成分に適当な数値を用いることにより, 重み付き検定とすることができ. 一方で, \mathbf{a} を任意としていることにより検出力が著しく低い検定となっている. 実際には比較の数は限定されていることが多いことから, 本研究では, \mathbf{a} の数を制限することで検出力の向上を目指すこととした. \mathbf{a} の数が 1 つである場合, 帰無仮説の下での検定統計量は正確に F 分布にしたがうことを示し, \mathbf{a} の数が複数である場合は, F 分布から求める棄却限界値を Bonferroni の方法, Sidak の方法で多重性の調整をおこない, モンテカルロシミュレーションにより, Siotani らの方法と比較して, 第一種の過誤確率を目標有意水準内に制御したうえで検出力を向上させることを確認した.