

東アジア国際共同試験における各国薬物動態同等性評価基準の研究

A study on equivalence evaluate criteria of Pharmacokinetics

Multi-regional Clinical Trials in East Asia

北里大学大学院薬学研究科 臨床統計部門 修士2年 張 鳴雨

北里大学大学院薬学研究科 臨床統計部門 竹内正弘教授

北里大学大学院薬学研究科 臨床統計部門 高橋史朗講師

Department of Clinical Medicine (Biostatistics & Pharmaceutical Medicine),
School of Pharmacy, Kitasato University)

1、序論

国際統一な新薬承認審査基準の「ICH ガイドライン」によって、国際共同試験で医薬品開発の国際化が加速している。人種間で内因性および外因性民族的要因の影響に大きな違いがあると考えられた。よく影響がある因子は薬物代謝の遺伝多型など遺伝的要因、肝臓腎臓心血管機能など生理的および病理的な要因、医療習慣、臨床試験の実施方法など環境要因である。そのため、国際共同試験の実施するのは有効性、安全性における民族間の差があるかどうかを明瞭に把握できる。また、一斉に様々な人種のデータを集められる。

近年、東アジア諸国国際共同試験の数が増加している。医薬品開発にアジア地域としては必要不可欠の存在になっておる。東アジアよく一つ地域 (yellow people 人種) として国際共同試験に参加するが、民族的要因によって、一部の薬剤において薬物動態が東アジア諸国で異なるといった報告があった。そのため、2007 年から日中韓三国保健大臣会議が開催され始めた。日中韓三国保健大臣会議によって、民族的要因に焦点を当てた有害事象の発見傾向の違い、薬物動態・薬力学に関するデータの収集を日本・中国・韓国で開始した。3 ヶ国臨床試験データの同等性を示すことおよびデータの相互受け入れを目指して、詳しく検討し始めている。2009 年、日中韓大臣声明に基

づいて、厚生労働科学研究一川合班が日中韓米における同一試験計画で、3つの薬物について医薬品の民族差に関する国際共同臨床研究を実施した。

同一試験計画で実施される国際共同試験は審査の迅速化が進み、ドラッグラグの問題が解決される。しかしながら、ここで大きな問題となるのは、3カ国のデータに関する統計的な積極的に薬物動態同等性を評価する判断基準が決めていないことである。

民族的要因と薬物動態の関連を検討した先行研究を調査したところ、2つの同等性を示す方法が見つけた。それが、パラメータの平均の差に対する有意差検定の結果が有意でない場合に同等と結論付ける方法と、積極的に同等性を評価するための後発医薬品開発における生物学的同等性評価法である。本研究では積極的に同等性を判断することができる後者の方法を日中韓の薬物動態試験へ適応することを考えた。

通常の生物学的同等性試験では、クロスオーバー比較試験で試験製剤と標準製剤の血中薬物濃度のAUCの対数値の平均値の差の90%信頼区間が、ガイドラインに定められた $\log(0.80) - \log(1.25)$ の同等領域に含まれるとき、試験製剤と標準製剤は生物学的に同等と判定する。なお、AUC及びC_{max}が正規分布する場合には、試験製剤と標準製剤のパラメータの母平均の差の90%信頼区間を標準製剤の母平均の90%信頼区間に対する比として、 $-0.20 \sim +0.20$ の範囲にあるとき、同等である。この方法の利点は、同一被験者ごとに異なる薬剤の血中薬物濃度AUCの比を算出するため、個体間の違いの影響を受けにくいことが挙げられる。そして、少ない被験者数で試験を実施できるというメリットもある。

本研究で考える状況は多国間の並行群間試験で薬物動態試験同等性を示すことである。そのため、通常の生物学的同等性試験で用いられる少ない被験者数では、以下の2つ異なるポイントがある。

1) 国ごとの異なる被験者の血中薬物濃度を用いるため、個体間の変動を排除できず、従来法の同等領域に90%信頼区間が含まれる可能性が低下である。

2) 複数回の対比較を行うことより偶然に 90%信頼区間が同等領域から外れる可能性が高まるといった問題が考えられた。

以上の理由で、国間薬物動態同等性を検討することを行う際に、後発医薬品の生物学的同等性試験の方法をそのまま用いることはできない。また、多国間での国際共同治験を想定した中では、Cochran の Q 統計量を用いた同等性の検討方法も有力である。しかし、当統計量は参加する国籍数が少ない時や、各国のデータの標準偏差が予測値であるときはその精度が下がるとの報告があり、東アジアでの小規模な国際共同治験には適応しているとは言えない。

もし、精度良い方法を提案され、薬物動態の同等性が示せる判断基準を作ると、今後東アジアで小規模な国際共同治験を実施して、3 カ国の臨床試験データの相互受け入れる事が可能になる。国際共同試験を積極的に推進していくための一助となる。

本研究では、多国に行う並行群間試験により薬物動態同等性の判断基準を提案することである。

2、データの概要

厚生労働科学研究川合班薬物動態試験では、2009年度から2年間で、モキシフロキサシン、シンバスタチン、メロキシカム3つの薬物を試験薬として、日中韓米において同一試験計画で薬物動態試験を実施した。それにより、東アジア民族間の薬物動態特性を厳密に比較すると同時に、差が認められた場合の要因を検討した。

調査した項目は、被験者の民族、例数、年齢、体重、性別、摂食状態、試験薬の剤型、投与量と投与経路などだった。薬物動態学的パラメータとしては最大血中濃度 (C_{max})、血中濃度曲線下面積 (AUC)、半減期 ($t_{1/2}$) などだった。

本研究では、川合班薬物動態試験の薬物動態パラメータの1つである血中濃度曲線下面積 (AUC) の結果に基づき、提案法よりシミュレーションを実施して、閾値の設定および提案法の性質を評価する。

3、同等確率の計算

並行群間試験に対して、薬効同等性の検討を行う際に、後発医薬品の生物学的同等性試験の方法という従来法をそのまま用いることはできない。個体間変動の影響により90%信頼区間が広くなり、その結果として従来法の同等領域に含まれる可能性が低くなる。そこで、私は従来法の問題点を解決するため、信頼区間の代わりに平均値の差が同等領域に含まれる確率を考えた。

しかしながら、同等確率を複数回計算すると、少なくとも一つの対で同等ではないと判断する可能性が高くなることをわかった。もし平均値の差の最大値が同等領域に含まれば、その他の対の平均値は必ず同等領域に含まれる。だから、複数の同等確率を避けるため、最大な平均の差の分布を用い、平成22年厚生労働科学研究川合班の薬物動態パラメータの1つであるAUCの結果に基づき、最大な平均値の差が同等領域に含まれる確率がある閾値より大きいときに同等と判断する基準を提案する。

3. 1 最大な平均の差

多国の並行群間試験に対して、パラメータの同等性を示すことでは、多重比較の対比較の問題である。それを応じる Tukey-Kramer 法という解析方法がある。2群ごとの比較を考えると、検定統計量として通常の t 統計量を考えるようにする。今回、全群通してデータの分散 σ^2 が等しいことを仮定しているので、 σ^2 の推定値には4群にデータを全て使ってやるの σ^2 である。つまり、統計量を

$$t_{ij} = |\bar{x}_i - \bar{x}_j| / \sqrt{(1/n_i + 1/n_j) \sigma^2}$$

とした。

多重比較の問題であるから、試験全体としての α エラー (FWE) を制御することの検討が必要である。正しい帰無仮説が多いほうが FWE は大きい。そのため、全ての帰無仮説が正しい状況で FWE を制御しておけば、真の状況が他の場合でも適切と考える。その通り6個の統計量 t_{ij} のうち、最も棄却されやすいもの、つまり最大なものを制

御すればよい、ということになる。

$$T = \max_{i < j} t_{ij} = \max_{i < j} |\bar{x}_i - \bar{x}_j| / \sqrt{(1/n_i + 1/n_j) \sigma^2}$$

$T = \max t_{ij}$ は今回の最大な平均の差の統計量である。

3. 2、同等確率

最大な平均の差と言え、平均値の最大値と最小値間の幅ということである。つまり平均値の範囲である。そのため、Tukey-Kramer のところで出てきた student 化した範囲の分布を考える。

student 化した範囲の分布では棄却限界値より小さい確率を計算できる。そして、提案の最大な平均値の差 $T = \max t_{ij}$ について、分子の部分各群の平均値 $N(0, 1)$ に従うように変形させたら、 $Q_{k,v}^{\alpha}$ と対応つける形 $T = \max_{i < j} T_{ij} = \frac{1}{\sqrt{2}} Q_{k,v}^{\alpha}$ になれる。

それより、 T の下側 95% 点を求める場合、 $Q_{k,v}^{\alpha}$ の下側 95% 点を計算して、それを $\sqrt{2}$ 割ってよい。すなわち、 T 統計量の棄却限界値は student 化した範囲の分布 $Q_{k,v}^{\alpha}$ の棄却限界値 q を $\sqrt{2}$ 割ってよい。

提案した最大な平均値の差 $T = \max t_{ij}$ が同等領域に含まれる確率は、棄却限界値の代わりに同等領域の範囲で確率を検討する。よって、student 化した範囲の分布 $Q_{k,v}^{\alpha}$ で棄却限界値 q より小さい確率の計算と類似している。

提案した最大な平均の差の分布が同等領域 $(-L, L)$ に含まれる同等確率の計算式では、

$$\Pr(\max T_{ij} \leq l) = \int_0^\infty \sum_{i=1}^k \int_{-\infty}^\infty \prod_{\substack{j=i \\ j \neq i}}^k \left\{ \Phi_j(y_i + \sqrt{2}lx) - \Phi_j(y_i - \sqrt{2}lx) \right\} d\Phi_i(y_i) f_v(x) dx \text{ となる。}$$

ここで、 $\Phi_i(\bullet) \Phi_j(\bullet)$ は $\bar{x}_i, \bar{x}_j \sim N(\mu_i, \frac{\sigma^2}{n}), N(\mu_j, \frac{\sigma^2}{n})$ 分布標準化に変形して $N(0, 1)$

に従う累積分布関数、 $f_v(x) = \frac{v^{v/2}}{\Gamma(v/2) 2^{v/2-1}} x^{v-1} \exp(-vx^2/2)$ は $Q_{k,v}^{\alpha}$ 分母の部分ル

ート χ^2 分布割自由度の分布の密度分布関数、 $\Gamma(\bullet)$ はガンマ関数を示す。

3. 3、同等領域

今回、後発医薬品の生物学的同等性試験の同等領域を参考になる。後発医薬品の生物学的同等性試験ガイドラインに定められたのは、クロスオーバー比較試験で試験製剤と標準製剤の血中薬物濃度の AUC の対数値の平均値の差の 90% 信頼区間が、 $\log(0.80) - \log(1.25)$ の同等領域に含まれるとき、試験製剤と標準製剤は生物学的に同等と判定する。なお、AUC 及び $C_{\max}$ が正規分布する場合には、試験製剤と標準製剤のパラメータ母平均の差の 90% 信頼区間を標準製剤母平均の 90% 信頼区間に対する比として、 $-0.20 \sim +0.20$ の範囲になると、同等と判定する。

この群間異なる因子は、試験製剤と標準製剤ではなく、人種である。それでも、変動する因子の数が増減したわけではないため、後発品の生物学的同等性試験と同様の方法を用いることができると考えられる。なお、統計解析を行う上では、今回の血中濃度曲線下面積 (AUC) の結果が対数正規分布に従うことが多いため、対数変換を行って使う。同等領域が後発医薬品の生物学的同等性試験の $\log(0.80) - \log(1.25)$ の範囲に参考される。この範囲から広げて $\pm 20\% (\log(0.8) - \log(1.25))$ から $\pm 50\% (\log(0.5) - \log(2))$ まで 5% ずつ増加させ検討している。

3. 4、同等確率の計算

数字計算の条件：平成 22 年厚生労働科学研究川合班の薬物動態パラメータの 1 つである AUC の結果に基づき、地域数を 4、各地域の AUC に対して等分散の正規分布を仮定し、4 群のうち 1 つの群の平均値が変化する状況を想定した。同等領域に関しては $\pm 20\% (\log(0.8) - \log(1.25))$ から $\pm 50\% (\log(0.5) - \log(2))$ まで 5% ずつ増加させた。(表 1、川合班試験 3 剤の結果)

	各群症例 数 (例)	Geometric mean (h・nM)	Sd of Geometric mean (h・nM)	Mean (h・nM)	Sd (h・nM)
モキシフロ キサシン	40	40	9	3.6888	2.1972
シンバスタ チン	20	30	18	3.4012	2.8904
メロキシカ ム	30	40	13	3.6888	2.5649

同等確率が計算された結果は次の図 1-3 になります：

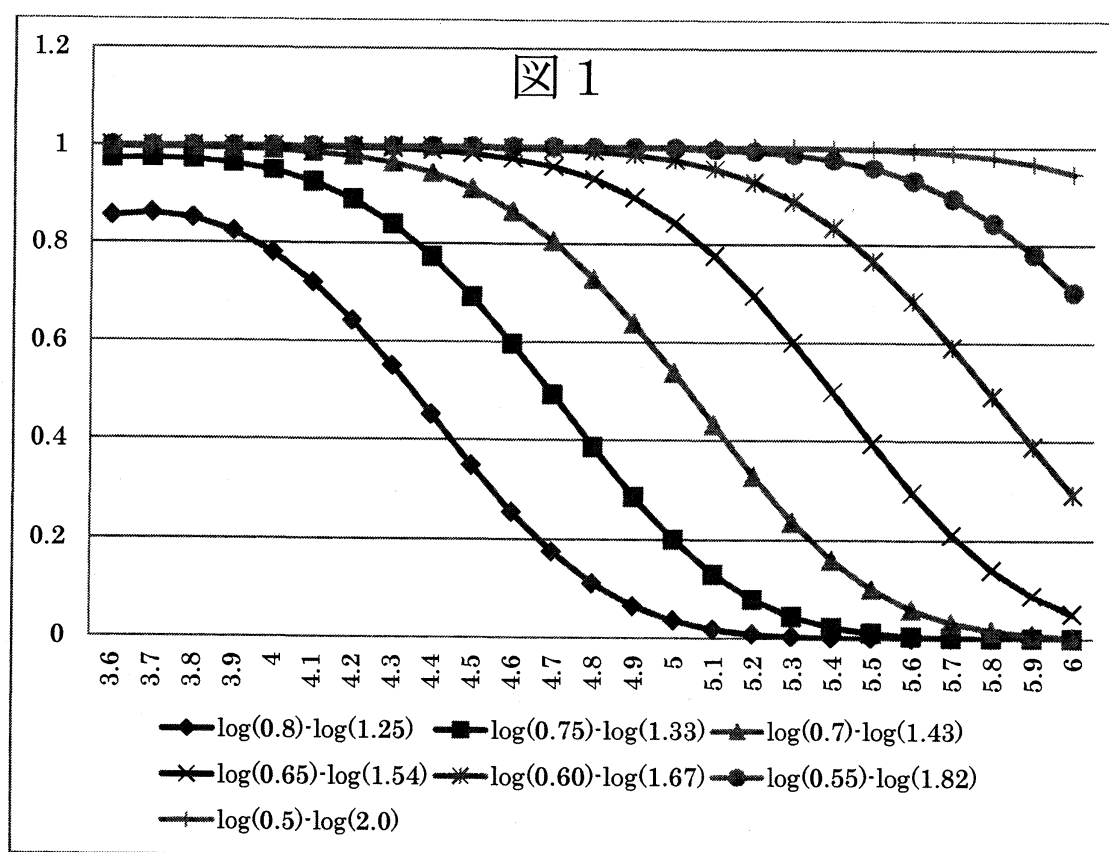


图 2

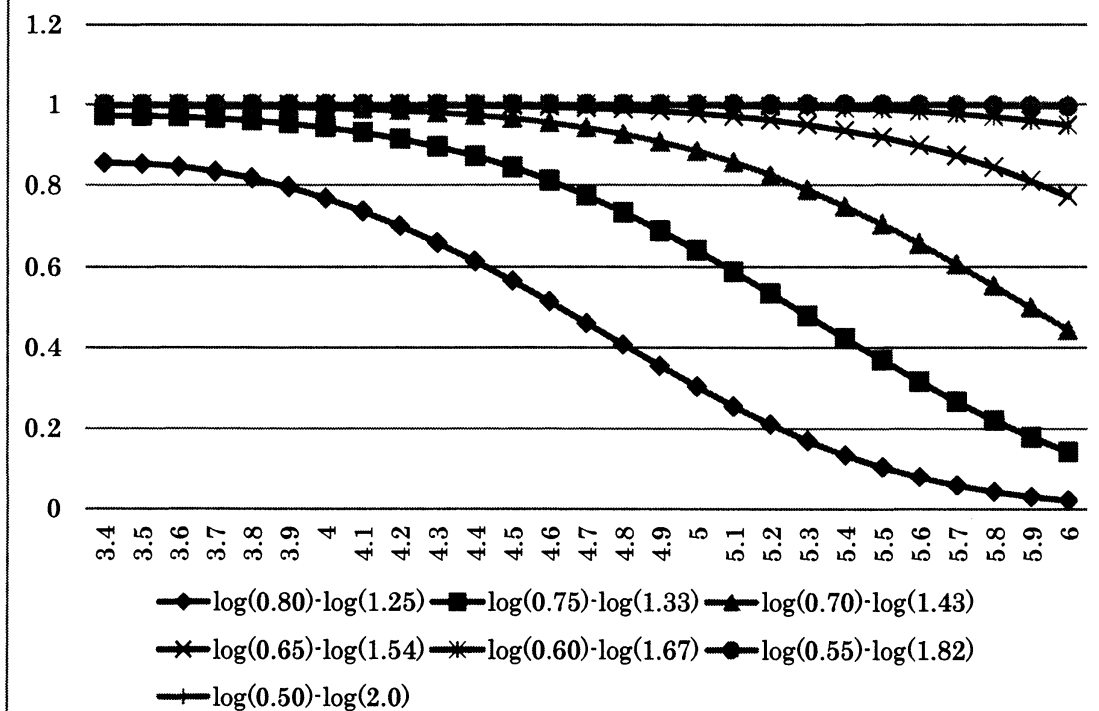


图 3

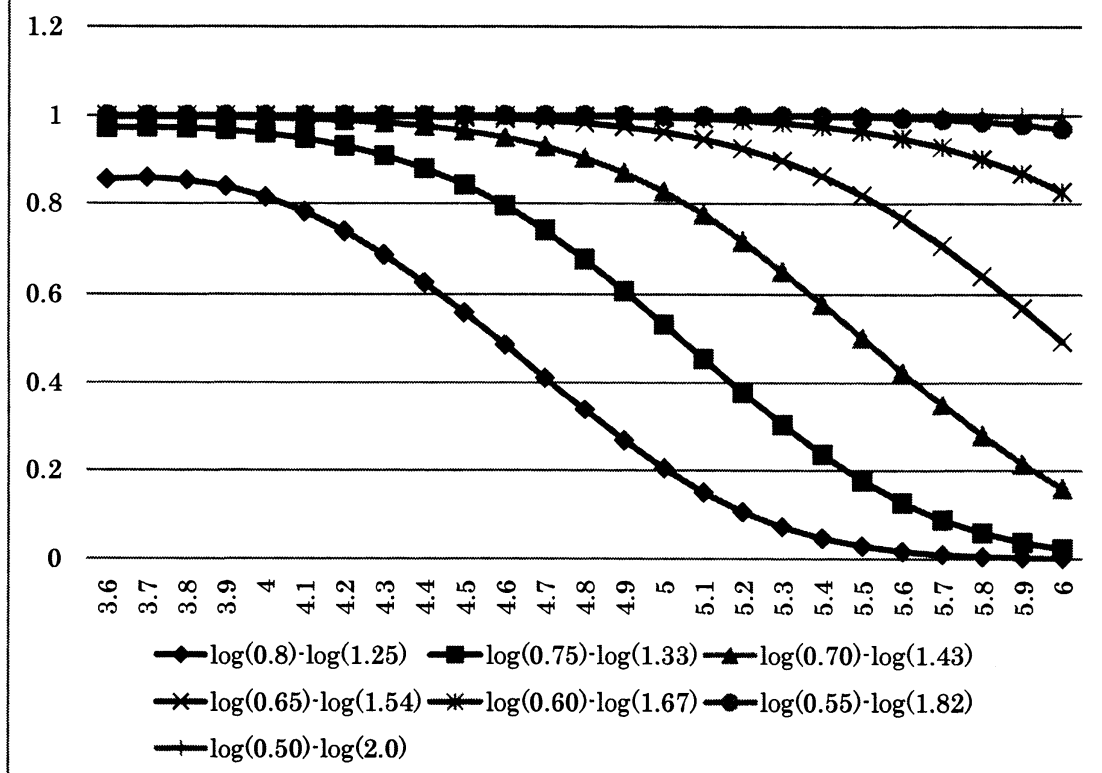


図1はモキシフロキサシン、図2はシンバスタチン、図3はメロキシカムを表している。

いずれの図では、横軸は1つ群の平均値、縦軸は同等確率を示している。いずれの同等領域の場合はそれぞれ色の線を示している。平均値は3.4から6まで0.1ずつ増加し、いずれの領域に含まれる同等確率が計算された。

いずれの薬剤でも、平均変化しながら、同等確率小さくなる。同等領域は広くなるに連れて、計算された同等確率も大きくなることをしめしている。

4、閾値の予測、性質の評価

ここまで実際データに基づいて、同等確率 (P) を計算した。最大な平均値の差が同等領域に含まれる確率がある値より大きいときに同等と判断する基準を提案するため、特定のカットオフ値 (閾値) の予測が必要である。

そして、同等領域に応じた閾値はどのように予測すべきか、一つの考えは Meta 解析において異質性を評価する指標である Cochran Q 統計量が異質と判断したときに、提案法も異質と判断する可能性が高くなるように設定することである。具体的に4.2方法の部分で2つ方法を比較し、同等確率の閾値を予測することを記述している。

さらに、提案法と Cochran Q 統計量の性質について、同等領域を変化させ評価することは4.3提案法の性質の部分で記述している。

4.1、比較対象

ここで、比較対象である Cochran Q 統計量と Cochran Q 統計量から派生した偽陽性率を用いた方法 Higgins I^2 統計量を紹介する。

Cochran Q 統計量は、複数の国の治験データの平均値が全て同じであることを帰無仮説とする分散分析の考え法を用いた指標である。同統計量は、参加する国の数が少ないときや、各国のデータの標準偏差が予測値になる場合などでは、その精度が落ちるとの報告がある。しかしながら、現在の Meta 解析や国際共同治験で用いられる統

計量として最も有力であることも事実であり、今回は Cochran Q 統計量と比較することとした。

具体的な方法は以下である。国籍 i (今回 $i=1, 2, 3, 4$) の j 番目の患者における AUC を Y_{ij} で表した時、 $Y_{ij} \sim N(u_i, \sigma_i^2/N_i)$ であると仮定する。但し、 N_i 、 u_i 、 σ_i^2 はそれぞれ国籍 i におけるサンプルサイズと平均値と分散とする。 $N = \sum_{i=1}^s N_i$ は全ての国のサンプルサイズである。全ての国のサンプルサイズにおける国籍 i のサンプルサイズの割合 f_i は $f_i = N_i/N$ となる。各国のサンプルサイズで重み付けをした AUC の重み付き母平均値を $\hat{Y}$ とすると、 $\hat{Y} = \sum_{i=1}^s f_i u_i$ と表れる。ここで、帰無仮説 $H_0: u_1 = u_2 = u_3 = u_4$ としたとき、Cochran Q 統計量は以下のように表れる。

$$Q = \sum_{i=1}^s \frac{(u_i - \hat{Y})^2 N_i}{\sigma_i^2} \sim \chi_{s-1, 1-\alpha}^2$$

但し、 α は有意水準である。この Cochran Q 統計量の $\chi_{s-1, 1-\alpha}^2$ 分布における p 値を $p\text{-value-Q}$ とする。

また、Cochran Q 統計量に基づいて、異質程度を評価するため、Higgins I^2 という偽陽性率を用いた方法がある。この Higgins I^2 統計量は $I^2 = \max\{0, 100\% \times [Q - (k-1)]/Q\}$ と表す、 k が比較する国籍の数、 Q が Cochran Q 統計量である。Higgins I^2 統計量が 0% となったという最も高い数値である場合、異質ではないと想定される。25%、50%、75% になると、軽度異質、中度異質、重度異質と想定される。

今回の提案法と Cochran Q 統計量を比較した上で、Higgins I^2 統計量の異質程度より、閾値を調べる。

4. 2、閾値の予測

4. 2. 1、方法

Meta 解析で異質性を評価する方法である Cochran Q 統計量を用いて、新たな方法の閾値を設定する。Cochran Q 統計量は異質性の検出力が高いと示す時点で、提案法同等性を示しにくいと考えられた。よく Cochran Q 統計量の p 値 (p-value-Q) が 0.10、0.25、0.50 となる場合、高度異質、中度異質、低度異質、同等のカットオフ値として使われている。いずれの p 値が 80% 検出力で応じている平均値が計算される。こういう平均値は時点とし、提案法より計算された同等確率が提案法の閾値になる。(表 2、Cochran Q 統計量異質程度)

p-value-Q	0.10	0.25	0.50
Cochran Q 統計量	6.251	4.108	2.359
異質程度	高度と中度閾値	中度と低度閾値	低度と同等閾値

4. 2. 2、結果

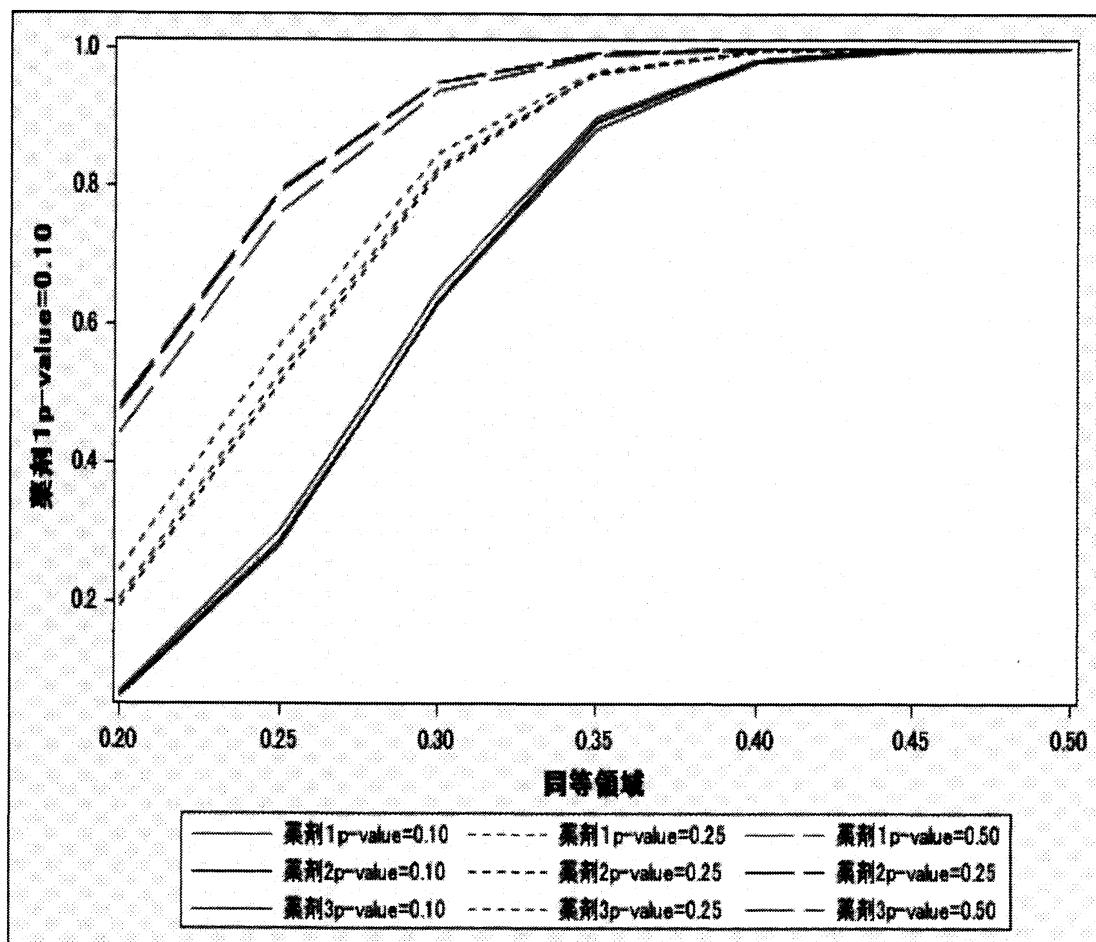


図4で、横軸は同等領域、縦軸は同等確率を示している。赤い線はモキシフロキサシン、青い線はシンバスタチン、緑の線はメロキシカムを表している。いずれの同等領域に含まれる同等確率はこの図を示している。実線の場合は高度異質と中度異質の閾値で、短点線の場合は中度と低度異質の閾値で、長点線の場合は低度異質と同等の閾値になる。いずれの領域に応じている閾値は以下の表を示している：

(表3、いずれの領域応じた閾値の結果)

p-value-Q	0.10	0.25	0.50
Cochran Q 統計量	6.251	4.108	2.359
異質程度	高度と中度閾値	中度と低度閾値	低度と同等閾値
$\log(0.80)-\log(1.25)$	0.05	0.20	0.45
$\log(0.75)-\log(1.33)$	0.25	0.60	0.75
$\log(0.70)-\log(1.43)$	0.65	0.85	0.95
$\log(0.65)-\log(1.54)$	0.88	0.97	0.99
$\log(0.60)-\log(1.67)$	0.98	0.995	0.999
$\log(0.55)-\log(1.82)$	0.998	0.999	0.999
$\log(0.50)-\log(2.00)$	0.9998	0.99998	0.99999

4. 2. 3、考察

図4によって、3種類の薬剤の結果を用いて計算したところ、薬剤間に大きな違いは認められません。同等領域は広くなるに連れて、計算された同等確率も大きくなり、閾値によって異質程度を区別することが難しいである。そのため、35% ($\log(0.65)-\log(1.54)$) 以上広い同等領域は適切ではないと考えている。また、計算された確率が大きいときに同等、小さいときには異質といった判断基準を作るとすると、20%、25%、30%の中で、25% ($\log(0.80)-\log(1.25)$) が最適な同等領域と考えた。

よって、4群試験で同等領域 25%に応じた閾値 0.75, 0.50, 0.25 とする判断基準を提案する。

4. 3、提案法の性質

4. 3. 1、方法

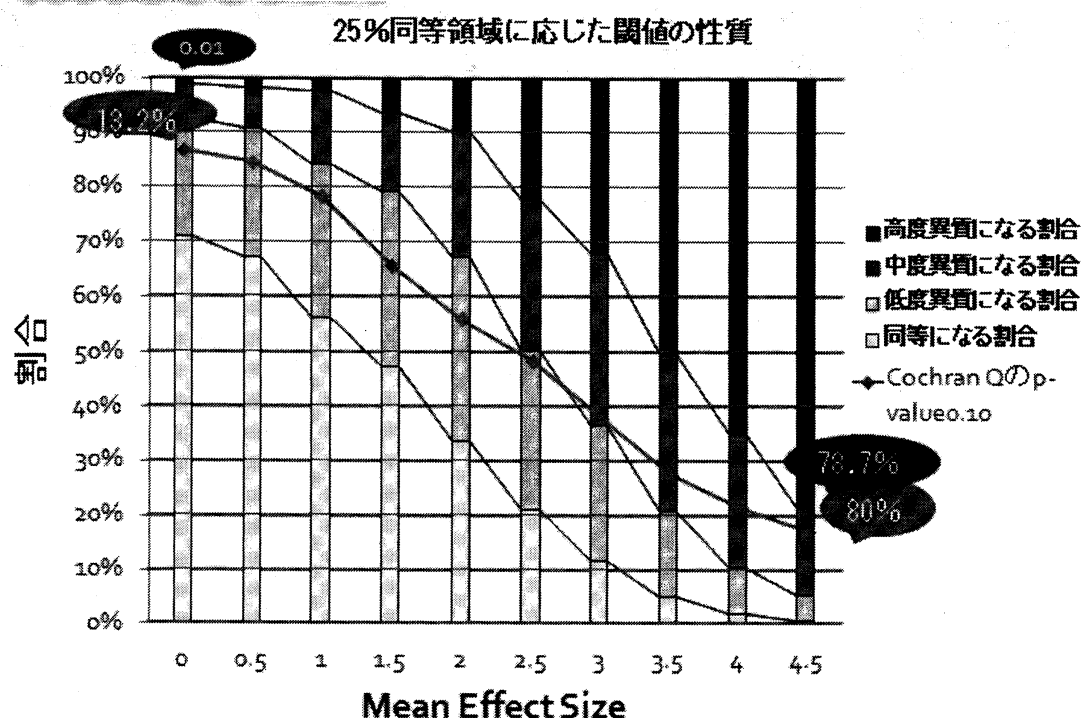
提案法の性質と基準の Cochran Q 統計量 2 つ方法の性質を比較したいである。

4 群の中 1 つ群の平均値が増加しながら、5000 回シミュレーションを実施する。そのうち、Cochran Q 統計量の p-value-Q がカットオフ値以下になる割合が変化している。同じ状況で応じている提案法各異質程度に含まれる割合も変化している。

平均値を変化させ、提案法各異質程度に含まれる割合と基準法 Cochran Q 統計量の p-value-Q がカットオフ値以下になる割合はどのように変化すべきであるか、性質な指標として比べることを検討する。

4. 3. 2、結果

ここで、提案した 25%同等領域に応じた閾値を用いて、性質を比較する。結果は図 5 を示している。



横軸は平均値の Effect Size、縦軸は各異質程度に含まれる割合である。色が深いほど異質程度が高くなる。Cochran Q 統計量の p-value-Q0.10、各時点の検出力がオレンジ色の線を示している。

4. 3. 3、考察

まず、図 5 によって、帰無仮説の状況から見ると、提案法の高度異質になる割合が約 1% である。提案法高度異質に対応した状況は Cochran Q 統計量の p-value-Q0.10 になる状況である。すなわち、帰無仮説になっても、Cochran Q 統計量より 10% 誤って異質と判断になっている。

更に、対立仮説の部分では、Mean Effect Size 4.5 ぐらいになるとき、Cochran Q 統計量の p-value-Q0.10 が 80% 検出力の性能が持っている。提案法高度異質になる割合が 78.7% である。負けてない性能が持っている。

また、提案法は簡単な二値判断ではなくて、対立仮説の部分 Cochran Q 統計量が同等と判断したが、提案法はグレイゾンにちゃんと低度異質と中度異質を切り離れた。

そして、提案法は Cochran Q 統計量より精度良い方法と考えている。

5、3 群試験および 5 群試験へ拡張

群の数を変わったら、対比較の回数が変わるため、閾値が変わると考えられた。更に、3 群試験および 5 群試験の状況を検討する。

5. 1、3 群試験の設定

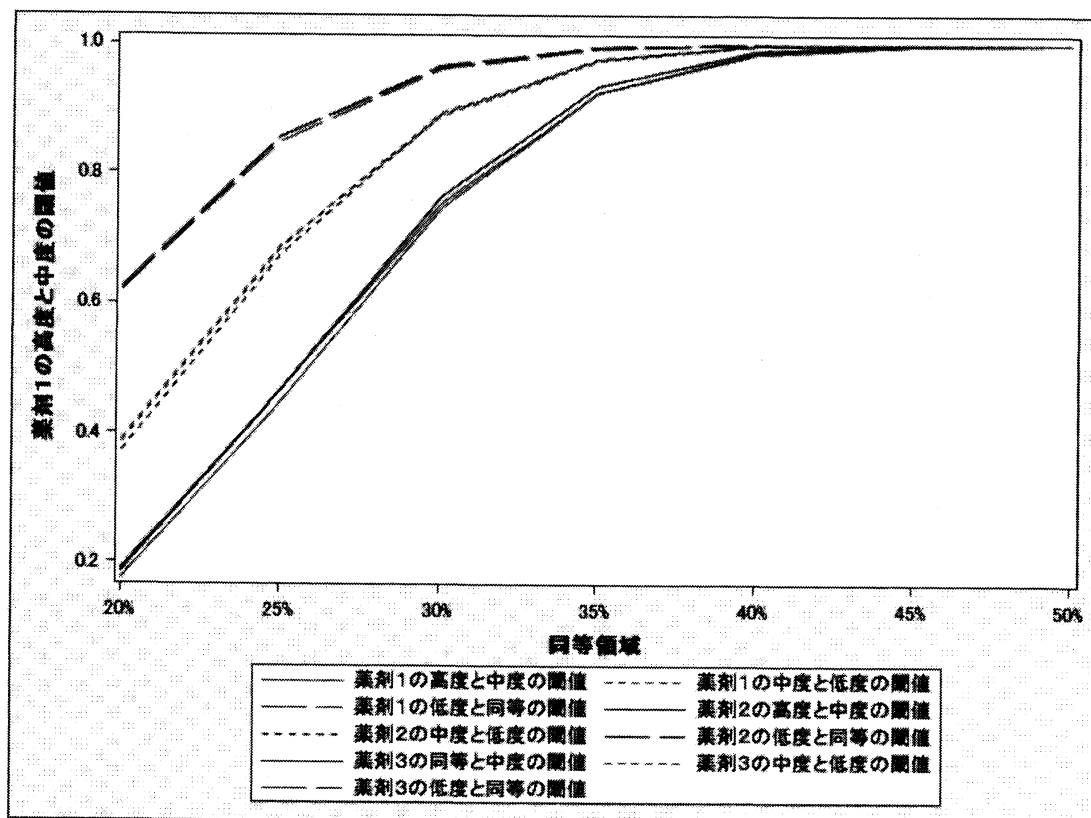
3 群のうち 1 つ群の平均値が変化することを想定して提案法の閾値を設定する。同様に Cochran Q 統計量は異質性の検出力が高いと示す時点で、提案法同等性を示しにくいと考えられた。Cochran Q 統計量の p 値 (p-value-Q) が 0.10、0.25、0.50 となる場合、高度異質、中度異質、低度異質、同等のカットオフ値として使われている。いずれの p 値が 80% 検出力で応じている平均値が計算される。こういう平均値は時点とし、提案法より計算された同等確率が提案法の閾値になる。(表 4 では 3 群の場合、

Cochran Q 統計量異質程度)

p-value-Q	0.10	0.25	0.50
Cochran Q 統計量	4.605	2.773	1.386
異質程度	高度と中度閾値	中度と低度閾値	低度と同等閾値

5. 2、3群試験の結果

次のページの図6を示している。同様に、横軸は同等領域、縦軸は同等確率を示している。赤い線はモキシフロキサシン、青い線はシンバスタチン、緑の線はメロキシカムを表す。いずれの同等領域に含まれる同等確率はこの図を示している。実線の場合は高度異質と中度異質の閾値で、短点線の場合は中度と低度異質の閾値で、長点線の場合は低度異質と同等の閾値になる。



いずれの領域に応じている閾値は以下の表を示している：

(表5、いずれの領域応じた閾値の結果)

p-value-Q	0.10	0.25	0.50
Cochran Q 統計量	4.605	2.773	1.386
異質程度	高度と中度閾値	中度と低度閾値	低度と同等閾値
$\log(0.80)-\log(1.25)$	0.20	0.40	0.60
$\log(0.75)-\log(1.33)$	0.45	0.70	0.85
$\log(0.70)-\log(1.43)$	0.75	0.90	0.95
$\log(0.65)-\log(1.54)$	0.93	0.975	0.993
$\log(0.60)-\log(1.67)$	0.985	0.997	0.999
$\log(0.55)-\log(1.82)$	0.998	0.9995	0.9999
$\log(0.50)-\log(2.00)$	0.9999	0.9999	1

同様に、3群試験で同等領域 25%に応じた閾値 0.85, 0.70, 0.45 とする判断基準を提案する。

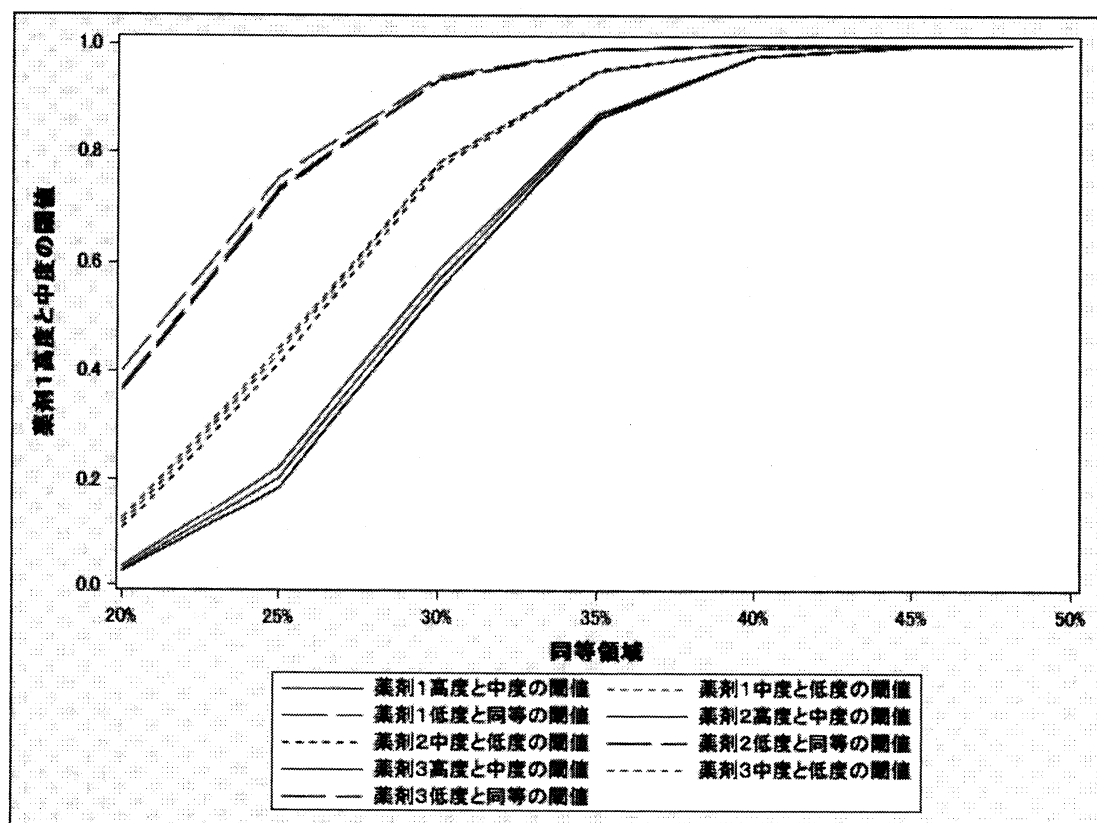
5. 3、5群試験の設定

5群のうち1つ群の平均値が変化することを想定して提案法の閾値を設定する。同様に Cochran Q 統計量は異質性の検出力が高いと示す時点で、提案法同等性を示しにくいと考えられた。Cochran Q 統計量の p 値 (p-value-Q) が 0.10、0.25、0.50 となる場合、高度異質、中度異質、低度異質、同等のカットオフ値として使われている。いずれの p 値が 80%検出力で応じている平均値が計算される。こういう平均値は時点とし、提案法より計算された同等確率が提案法の閾値になる。(表 6 では 5 群の場合、Cochran Q 統計量異質程度)

p-value-Q	0.10	0.25	0.50
Cochran Q 統計量	7.779	5.385	3.357
異質程度	高度と中度閾値	中度と低度閾値	低度と同等閾値

5. 4、5群試験の結果

次のページの図7を示している。同様に、横軸は同等領域、縦軸は同等確率を示している。赤い線はモキシフロキサシン、青い線はシンバスタチン、緑の線はメロキシカムを表す。いずれの同等領域に含まれる同等確率はこの図を示している。実線の場合は高度異質と中度異質の閾値で、短点線の場合は中度と低度異質の閾値で、長点線の場合は低度異質と同等の閾値になる。



いずれの領域に応じている閾値は以下の表を示している：

(表7、いずれの領域応じた閾値の結果)

p-value-Q	0.10	0.25	0.50
Cochran Q 統計量	7.779	5.385	3.357
異質程度	高度と中度閾値	中度と低度閾値	低度と同等閾値
$\log(0.80)-\log(1.25)$	0.02	0.10	0.40
$\log(0.75)-\log(1.33)$	0.20	0.40	0.70

$\log(0.70) - \log(1.43)$	0.55	0.75	0.95
$\log(0.65) - \log(1.54)$	0.85	0.95	0.99
$\log(0.60) - \log(1.67)$	0.978	0.99	0.999
$\log(0.55) - \log(1.82)$	0.998	0.9995	0.9999
$\log(0.50) - \log(2.00)$	0.9999	0.99999	1

同様に、5群試験で同等領域 25%に応じた閾値 0.70, 0.40, 0.20 とする判断基準を提案する。

5. 5、考察

いずれの同等領域に応じた閾値の判別能力が異ならなかったことが考察された。群の数を変わったら、対比較組み合わせの数が変わるため、閾値が変わった。群数多くなると、同等になる閾値が下がるため、その時、同等領域広い範囲を選択して、性質も同じである。

6、総括

今回提案した、最大な平均値の差が同等領域に含まれる確率の閾値より大きい場合、同等と判断する基準は、川合班の薬物動態パラメータの1つである AUC の結果をもちいて、異質程度および同等領域により、同等確率計算を通じて閾値を変更することで、Cochran Q 統計量に近い性質を持つことを分かった。そして、データのバラツキや、各国症例数の条件があるものの、4群試験で同等領域 25%に応じた閾値 0.75, 0.50, 0.25 とする判断基準は実際の現場でも使用可能な指標であることを示した。

今後の課題として大きく3つを挙げられる。

第一に、川合班の実データに基づき、生物学的同等性評価法の判断基準と本研究提案した判断基準に当てはめて、こういった結果になっているかを比較することを考えておる。

第二に、様々な状況を想定してシミュレーションの必要がある。今回のシミュレーションは人種ごとに代謝酵素や生活習慣といった諸因子を出来る限り母集団と類似させることで、人種間変動の影響を最小限に抑えると考えて、同分散の正規分布という条件から実施した。しかし、実の国際共同試験で、国ごとの分散は多少違いがある。人種間のバラツキにより、各国同分散の状態と想定されたことが十分ではないと考えた。今後、国ごとのバラツキも入れて分散を変動しながら、シミュレーションを実施し、同等確率の変化を見るべきである。そして、各国でその症例数が異なることによる同等確率の変化を確認するため、そうした場合についても検討する必要があると考えた。

第三に、本研究の閾値設定では Cochran Q 統計量を用いて設定したが、今後、同等性を示す方法から閾値がどれくらい変わるかを検討したいである。

7、参考文献

海外臨床データを受け入れ際に考慮すべき民族的要因について
厚生省医薬安全局審査管理課長（1998）

外国で実施された医薬品の臨床試験データの取り扱いについて
厚生省医薬安全局長（1998）

臨床 PK/PD と統計解析—ブリッジング試験と NONMEN—朝野芳郎 薬物動態, Xenobio, Metabol. and Dispos. 15(4):363-370(2000)

後発医薬品の生物学的同等性試験ガイドライン（2006）

Chen et al., (2010) Assessing consistent treatment effect in a multi-regional clinical trial. PHARMACEUTICAL STATISTICS Pharmaceut. Statist.

Y. Hochberg et al., (1987) Some probability Distributions and Tables Useful in Multiple Comparisons.

E.O.Bayman et al., Detecting qualitative interaction: A Bayesian approach.

Statistics in Medicine (2009MAR)

統計的多重比較法の基礎 永田靖 第 9 章 多重比較法で用いる確率分布—各種確率分布と正確な棄却限界値の求め方 (p145-147)

多重比較について (その 2-1 : 具体的手法) 土居正明

Testing for Homogeneity in Meta-Analysis1. The One-Parameter Case: Standardized Mean Difference BIOMETRICS67, 203-212 March 2011

H. M. James Hung et al. Consideration of regional difference in design and analysis of multi-regional trials.

John Cook et al., Assessing the Generalizability of results from multinational and multicenter clinical trials.