

〔一般論文〕

我が国に於いて行われた臨床試験によるジェネリック医薬品と先発医薬品の臨床的同等性評価のレビュー研究Ⅱ：比率を対象にした検定の問題点の検討

Review of Clinical Equivalence Evaluation between Innovator and Generic Products in Japanese Clinical Studies II : Examination of Issues in the Statistical Method Comparing Changes in the Ratio

塩見 真理^{*a}, 伊藤 永久佳^b, 緒方 宏泰^{a,b}MARI SHIOMI^{*a}, TOWAKA ITOH^b, HIROYASU OGATA^{a,b}^a 明治薬科大学薬剤学^b 明治薬科大学大学院臨床薬学専攻〔 Received October 26, 2012
Accepted April 15, 2013 〕

Summary : It has been reported that, of four pravastatin studies that were conducted with similar numbers of subjects and showed similar changes in serum lipid levels, two studies reported significant differences in the comparison of changes in the ratio of laboratory data used as the efficacy criterion, while the other two studies reported no significant differences in the comparison of the differences in laboratory data used as the efficacy criterion. We therefore investigated the causes of the discrepancy in these statistical tests by using a simulation algorithm.

Under the assumption of log-normal distribution for population, two sub-populations were randomly sampled from two hypothetical populations, each of which consisted of 1,000 study data with the same indicators generated by random numbers. To compare these two sub-populations, independent and paired t-tests were conducted on observed values, and a paired t-test was conducted on the ratio of the observed values. Mean value distributions were studied for each indicator of the 1,000 study data generated by random numbers. Distribution was also studied for mean values of the logarithmic transformation of the ratio. Statistical tests were conducted on this data to obtain the power to detect a 20% difference and the probability of detecting a significant difference.

In the analysis of the difference between observed values, the statistical power decreased with an increase in inter-and intra-individual variability, while the probability of detecting a significant difference was maintained at a level of approximately 5% in both independent and paired t-tests. When a test was conducted on the same observed values converted into ratios, although a decrease in the statistical power was again observed with an increase in inter-and intra-individual variability, the probability of detecting a significant difference increased. When a test was conducted on the log-transformed ratios, the power decreased with an increase in inter-and intra-individual variability, while the probability of detecting a significant difference was maintained at a level of approximately 5%. Mean value distributions of differences and log-transformed ratios were symmetric, while that of ratios was skewed and the modes deviated from 1 with an increase in variability.

When a statistical test was conducted using ratio values, despite the statistical power decreasing with an increase in inter-and intra-individual variability, the probability of detecting a significant difference increased. The cause of the seemingly contradictory result was revealed to be the skewed distribution of the ratio values, which was improved by using log-transformation. It was concluded that the significant

* 〒204-8588 東京都清瀬市野塩 2-522-1

TEL & FAX : 0424-95-8869

E-mail : shiomi@my-pharm.ac.jp

differences observed in the statistical test using the ratio values were produced by a type I error whose risks were increased while handling comparative indicators.

Key words : equivalence study, t-test, test on the ratio, statistical power, probability of detecting a significant difference, skewed t-distribution

要旨：プラバスタチンを対象とした4試験では、血清脂質の臨床検査データの変動は同程度であり、被験者数も同程度であったにもかかわらず、効果の指標に比率を用いて比較を行った2試験で有意差が報告され、効果の指標の差を用いて比較を行った2試験では有意差が認められなかったことが報告されている。そこで我々は、この統計検定における評価の食い違いの原因をシミュレーションの手法を用いて検討した。対数正規分布の母集団を仮定し、乱数により発生させたデータ1,000試験分のそれぞれの指標の同一集団からランダムにサンプリングした2群に関し、測定値を用いた対応のない2群比較、対応のある2群比較および測定値の比率を用いた対応のある2群比較を行った。乱数によりランダムに発生させたデータ1,000試験分のそれぞれの指標の平均値の分布を把握した。比率を対数値に変換した平均値の分布も検討した。これらについて統計検定を行い、20%の差を検出する検出力および有意な差が検出される割合を得た。観測値の差で評価した場合、対応のない2群比較、対応のある2群比較いずれにおいても、個体間変動・個体内変動の上昇に伴って検出力は低下したが、有意差検出率は5%程度で維持された。一方、同一の値を比率に変換し検定を行った場合、個体間変動・個体内変動の上昇に伴う検出力の低下が認められたが、有意差検出率は増大した。比率を対数値に変換したところ、個体間変動・個体内変動の上昇に伴い検出力は低下し、有意差検出率は5%程度を維持した。平均値の分布は、差および比率の対数値は左右対称であったが、比率の分布は歪みがある上、最頻値は変動性が大きくなるほど1から外れた値を示した。比率で統計評価した際、個体間変動・個体内変動の上昇に伴って検出力は低下するにも拘わらず、有意差検出率は上昇する一見矛盾する結果の原因は、比率の値の分布に偏りが生じるためであることが明らかになった。この分布の歪みは対数で表現することで改善した。比率で統計検定を行った結果、有意差が認められたのは比較指標の取り扱い上で生み出された第1種の過誤の増大によることが明らかとなった。

キーワード：同等性試験, t-検定, 比率を用いた検定, 検出力, 有意差検出率, t分布の偏り

1. 緒言

我が国において、一部の薬剤で、臨床上的効果の指標マーカーを対象に先発医薬品とジェネリック医薬品との間の比較試験を行い、その結果が報告されている。先に我々は、先発医薬品とそのジェネリック医薬品の一つである医薬品との臨床的な同等性を評価している論文を系統的に評価し、臨床試験デザインや統計の手法について検討した。検討対象とした全ての論文において統計手法は検定法が用いられ、採用されている試験プロトコルの多くに、試験デザイン上の問題点のあることを認めた¹⁾。

特に、プラバスタチンを対象とする4研究^{2,3,4,5)}において、解析に用いられる統計手法においても問題が存在することが認められた。4研究とも観察された臨床検査値の変動の程度は同程度であり、また症例数もほぼ同数で実施されており、被験者数および変動性から検出力は小さいと推定される試験であるものの、4報告のうち、2試験^{2,3)}は効果指標

値に有意差を認めたとする結果を示し、他の2試験^{4,5)}は有意差がないという検定結果を示し、統計評価の結果に相違が認められていた。効果に有意差を認めたとする試験はすべて、ジェネリック医薬品投与後の効果マーカー値をジェネリック医薬品投与直前値に対する比率に変換して検定を行っており、一方、有意差を認めなかった報告は、上記の両測定値の差で検定が実施されていた。本研究では、差あるいは比率という指標によって検定結果が異なる可能性について、本来差がない場合でも有意な差異があるとして統計的に判定される危険率（第一種の過誤； α ）と試験デザイン、統計手法等の関連性をシミュレーションの手法を用いて検討した。

2. 方法

1) 仮定の測定値の発生

正規分布あるいは対数正規分布を仮定し、試験データを発生させた。平均値を150と設定し、対数

正規分布の場合は個体間変動に相当する変動係数を 20, 30, 50, 70 および 100% (該当する標準偏差は, 30, 45, 75, 105 および 150) の条件で, 正規分布の場合は, 変動係数 20, 30% (該当する標準偏差は, 30, 45) の条件で, それぞれ 10,000 点から成る仮想の測定値をモンテカルロシミュレーションにより発生させ, データが全て正の値であることを確認した. 乱数発生は S-PLUS (Version 6.2; TIBCO Software Inc., Palo Alto, CA) を使用した.

2) 対応のない 2 群比較試験を想定したシミュレーション

1) で発生させたデータから 40 点を離散一様分布乱数に基づき, 無作為 (ランダム) 抽出した. 抽出した 40 点のデータの分布は, 仮定したすべての条件で母集団とほぼ同一であり, 偏りなく且つランダムに得られていることを確認した. このデータを用い, 1 群の被験者を 20 名と想定し, 抽出した 40 点のうちランダムに 20 点を対照医薬品群, 残りの 20 点を試験医薬品群の測定値とし, それぞれの群の平均値に対し対応のない t 検定 (第 1 種の過誤の危険率 (α) = 0.05) を行った. 帰無仮説は, 2 群の平均値の差を 0 とした. 検定に使用した t 値 (t_{01}) は, 以下に示した式 (1) - (3) を用いて得た.

$$\sigma_1^2 = \frac{\sum (x_i - \mu_1)^2}{n_1 - 1}, \quad \sigma_2^2 = \frac{\sum (y_i - \mu_2)^2}{n_2 - 1} \quad (1)$$

$$\sigma = \sqrt{\frac{\sigma_1^2(n_1 - 1) + \sigma_2^2(n_2 - 1)}{n_1 + n_2 - 2}} \quad (2)$$

$$t_{01} = \frac{\mu_1 - \mu_2}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad (3)$$

ここで, σ_1^2 , σ_2^2 (式 (1)) は対照医薬品投与群および試験医薬品投与群での分散, x_i , y_i は, 対照, 試験医薬品投与群における i 番目の被験者値, μ_1 , μ_2 は各群の平均値, n_1 , n_2 は各群の例数とした (いずれも 20 例とした).

また, 最小検出差をその試験で得られた対照医薬品群平均値 μ_1 を基準としてその 20%, 一群の被験者数を 20, α を 0.05 とした条件で, その試験で得られた対照医薬品群と試験医薬品群の標準偏差を用

いて算出した合成分散の平方根 σ (式 (2)) を用いて検出力 ($1 - \beta$)⁶⁾ を算出した (式 (4)).

$$\Pr\{t_{01} \leq -t(n_1 + n_2 - 2, \alpha)\} + \Pr\{t_{01} \geq t(n_1 + n_2 - 2, \alpha)\} \quad (4)$$

なお対立仮説の条件下, t_{01} が自由度 $\phi = n_1 + n_2 - 2$, 非心パラメータ $\lambda = \sqrt{n_1 \cdot n_2 / (n_1 + n_2)} \cdot ((\mu_1 - \mu_2) / \sigma)$ の非心 t 分布 $t'(\phi, \lambda)$ に従うとした.

以上の操作を 1,000 回繰り返し, それぞれの検定における検出力の平均値 (平均検出力) を求めると同時に, 検定において有意な差が検出された試験数をカウントして有意差検出率を算出した.

3) 対応のある 2 群比較試験を想定したシミュレーション

3-1) 評価項目に測定値を用いた統計検定

1) で発生させた平均値 150, 個体間変動を想定した分散を持つ 10,000 点から成る仮想の測定値から 20 測定点をランダムに抽出し, この値をそれぞれの 20 症例のデータの分布の中央値とした. 中央値は, 仮定したすべての条件で, 母集団から偏りなく抽出されていることを確認した. この中央値に対して, 個体内変動を想定した変動係数をそれぞれ 20, 30, 50, 70 および 100% (該当する標準偏差は, 30, 45, 75, 105 および 150) を有する対数正規分布を仮定し, 1,000 データを発生させ, その同一分布内からランダムに 2 観測値を抽出し対にし, それぞれ, 対照医薬品値, 試験医薬品値とした. 20 症例分の対になったデータを 1 試験とし, 対応のある t 検定 ($\alpha = 0.05$) を行った. 帰無仮説は 2 群の差の平均値が 0 であるとし, t 値 (t_{02}) は以下の式 (5), (6) を用いて算出した.

$$\sigma_d = \sqrt{\frac{\sum (d_i - \bar{d})^2}{n - 1}} \quad (5)$$

$$t_{02} = \frac{\bar{d}}{\frac{\sigma_d}{\sqrt{n}}} \quad (6)$$

ここで, σ_d は 20 名分の被験者データ対における差の標準偏差 (式 (5)) を, d_i は i 番目の被験者におけるデータ対の差, $\bar{d}$ は差の平均値, n は例数 (20 例とした) を示す. 参考にしたプラバスタチンの文献^{2,3)}

では、個々の値を対数変換して比較はされていないため、これを踏襲し対数正規分布を仮定した場合にも対数変換は行わず、値をそのまま比較に用いた。

対照試験値の平均値を基準としてその20%値を最小検出差、対データ総数を20、 α を0.05とした条件で、差の標準偏差を用い、各試験の検出力⁶⁾を算出した(式(7))。

$$\Pr\{t_{02} \leq -t(n-1, \alpha)\} + \Pr\{t_{01} \geq t(n-1, \alpha)\} \quad (7)$$

なお t_{02} が自由度 $\phi = n-1$ 、非心パラメータ $\lambda = \sqrt{n} \cdot \Delta$ の非心 t 分布 $t'(\phi, \lambda)$ に従うとした。ここで $\Delta = d/\partial_d$ とする。

これを1,000回分試験し、それぞれの検定における検出力の平均値(平均検出力)を求めると同時に、検定において有意な差が検出された試験数をカウントして有意差検出率を算出した。

3-2) 評価項目に比率を用いた統計検定

3-1)で得た同一のデータを用い、その比率を算出した。比率(ratio; R)は式(8)に定義した。

$$R = \frac{X_g}{X_i} \quad (8)$$

X_i, X_g は、それぞれ、対照医薬品値および試験医薬品値を示す。検定は、 R または R の対数値を用い、対応のある t 検定、Wilcoxon符号付順位検定(Wilcoxon signed-rank test)を $\alpha = 0.05$ の条件で行った。帰無仮説は、 R の平均値が1、 R の対数値の場合は平均値が0とし、 t 値(t_{03})は以下の式に基づいて計算した。

$$\sigma_R = \sqrt{\frac{\sum (R_i - \bar{R})^2}{n-1}} \quad (9)$$

$$t_{03} = \frac{\bar{R}}{\frac{\sigma_R}{\sqrt{n}}} \quad (10)$$

ここで σ_R は20名分のデータ対における標準偏差を、 R_i は i 番目の被験者の比率、比率の平均値を $\bar{R}$ とした。 R の対数値の場合も同様にした。

R の最小検出差は $\bar{R}$ 値の0.2とした。 R の対数値においても同様に、20%の相違を検出する最小検出差として、 $\ln(1.2/1) = \ln(1.2) = 0.182$ および $\ln(1/0.8) = \ln(1.25) = 0.233$ から、より厳しい基準で

ある0.182を採用した。対データ総数は20、 α を0.05とした条件で、 R あるいは R の対数値の標準偏差を用い、各試験の検出力⁶⁾を算出した(式(7))。なお式(7)の条件で、比率は $\Delta = R/\partial_R$ 、比率の対数は $\Delta = \ln(R)/\partial_{R, \log}$ とした。1,000回分試験し、それぞれの検定における検出力の平均値(平均検出力)を求めると同時に、検定において有意な差が検出された試験数をカウントして有意差検出率を算出した。

3-3) 対応のある2群比較における個体内相関の寄与

3-1)で発生させたデータに個体内相関を仮定し、相関を有する条件下で、評価項目の違いが有意差検出率へどのように影響を与えるかを検討した。

二つの変数 x と e がそれぞれ分散の等しい母集団から独立に抽出された無作為標本であるとき、相関 ρ をもった標準正規乱数を x, y とすると、 y は x と e の線形結合 $Y = \rho X + ae$ で表現される。 a は任意の実数、 X および e は標準正規確率変数である。確率変数 Y は $N(0, (a^2 + \rho^2)^{1/2})$ 、 $(a^2 + \rho^2)^{1/2} = 1$ 、 $a = (1 - \rho^2)^{1/2}$ より、 $y = \rho x + (1 - \rho^2)^{0.5}e$ と表現できる⁷⁾。この関係を利用して、データ x に対して、相関をもつ新規データ y を発生させた。元になるデータ x は3-1)の検討で使用した対照医薬品のデータ X_i をそのまま使用し、 e は X_i の同一母集団から無作為抽出して得た。上記の関係をを用いて相関係数を0.5から0.8まで変化させ、データ x に対応する0.5から0.8の相関を有するデータ y を発生させ、同一被験者での対照医薬品値(元データ X_i)と、試験医薬品値(x_i に対して相関を仮定して発生させた X_g)として対にし、被験者1の試験結果とした。これを20症例分実施し、1試験とした。

評価項目に測定値の差を用いた統計検定では、3-1)と同様に、式(5)、(6)を用いて t 値(t_{02})を算出し、対応のある t 検定($\alpha = 0.05$)を行った。評価項目に比率を用いた検定は、3-2)と同様に比率は式(8)のように定義した。検定は、 R または R の対数値を用い、対応のある t 検定を $\alpha = 0.05$ の条件で行った。なお t 値の算出には式(9)、(10)を用いた。

3-4) 観測値の差、比率および比率の対数値の分布の確認

シミュレーション研究で得られた、対応のある2群比較における対照医薬品値と試験医薬品値の差、比率および比率の対数値の平均値の分布を確認する

ために、被験者 20 名分の平均値を試験毎に算出した。これらの平均値を 1,000 試験分集計して、度数をプロットし、その分布の形状を把握した。度数分布の階級区切り（度数の区切り、横軸の刻み）は、カーネル密度分布を用いた確率密度分布⁸⁾を参考にしながら、目視により区切りを決定した。また各試験で得られた t 値についても同様に、確率密度分布の区切りを参考にして目視で度数を決定し、t 値の分布をプロットした。 α が 0.05 以下の確率となる t 値の割合を分布から確認した。

3. 結果

1) 対応のない 2 群比較試験

正規分布または対数正規分布を仮定し、対応のない 2 群比較において、1 群の被験者数を固定した条件で、t 検定による有意差検出率および対照医薬品平均値の 20% の値を有意な差として検出する検出量の平均値と個体内変動との関係を検討し、その結果を Fig. 1 に示した。

正規分布あるいは対数正規分布のいずれにおいても、1 群の被験者数を固定した条件で、個体内変動の大きさに関係なく、有意な差ありと判定される比率はおおよそ 5% と一定の値を示し、一方、検出力は個体内変動の上昇に伴い低下した。これらの関係は、想定通りの妥当な結果である。

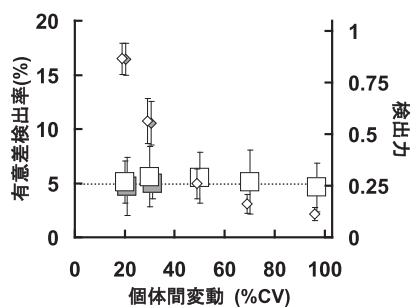


Fig. 1 対応のない 2 群比較における有意差検出率、平均検出力と個体内変動の関係

母集団：正規分布および対数正規分布を仮定。

■；正規分布データによる平均有意差検出率，□；対数正規分布データによる平均有意差検出率
◆；正規分布データによる平均検出力，◇；対数正規分布データによる平均検出力

破線は有意差検出率 5% を示す

2) 対応のある 2 群比較試験

同一被験者に対し対照医薬品および試験医薬品を投与し得られた測定値の差あるいは測定値の比率を指標として t 検定を行い、有意差検出率を求めると共に、20% の差を有意な差として検出する検出力の平均値を求めた。Fig. 2 に、母集団に対数正規分布を仮定した条件で、個体内変動を 100% に固定し個体内変動を 20 から 100% に変化させた場合の、有意差検出率および平均検出力の変化を示した。

測定値の差を指標とした場合 (Fig. 2- (a)) では、対応のない 2 群間比較 (Fig. 1) と同様に、有意な差ありと判定された比率はおおよそ 5% と一定の値を示し、一方、検出力は個体内変動の上昇に伴い低下し、妥当な結果が示された。一方、比率を指標とした比較 (Fig. 2- (b)) では、個体内変動が大きくなるに従い、検出力は低下する傾向が認められているにも拘わらず、有意差検出率は 5% より、更に大きくなる傾向にあることが認められた。さらに統計手法を母集団の分布に依存しないノンパラメトリック法である Wilcoxon 符号付順位検定で検定した場

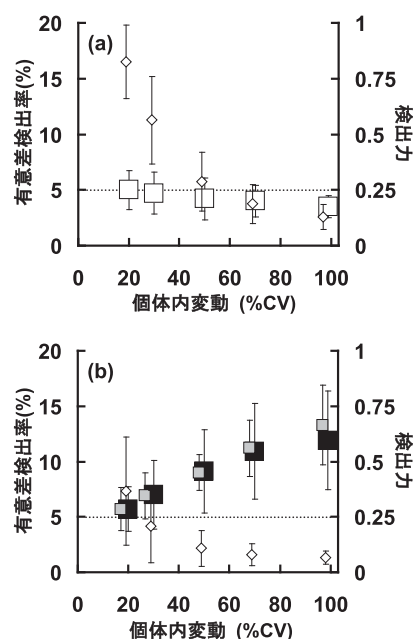


Fig. 2 同一被験者内比較における測定値の差 (a) または比率 (b) の有意差検出率、平均検出力と個体内変動の関係

母集団：対数正規分布を仮定

□；指標に差を用いた平均有意差検出率，■；指標に比率を用いた t 検定にて算出した平均有意差検出率，■；比率を Wilcoxon 検定にて算出した平均有意差検出率
◇；平均検出力，破線は有意差検出率 5% を示す

合でも、t 検定の結果と同様の傾向を示した (Fig. 2-(b)).

個体内変動を固定した条件における、個体間変動と有意差検出率との関係は、先の個体内変動と有意差検出率の関係と同様の傾向を示した (図は省略). ただし、個体内変動がより大きく有意差検出率に影響を与えていた. また、母集団を正規分布に仮定したシミュレーション結果も対数正規分布と同様の傾向を示した (図は省略).

以上の検討は、同一分布内からランダムにサンプリングして得たデータを検定対象にしたが、実際の臨床データにおいては、個体内において相関性を有する可能性がある. そこで、同一被験者内で相関を持つデータを発生させ、個体内相関の程度を 0.8 から 0.5 と変化させて、測定値の差あるいは測定値の比率で比較し、有意差検出率への個体内相関の寄与を検討した. 結果を Fig. 3 に示した.

個体内相関性が高いほど比較に用いる指標が測定値の差か測定値の比率かに関係なく、有意差が検出される比率が大きくなった. しかし、差と比率の比較では、結果に乖離が認められ、同一被験者内に相関を設定しても、比率で比較すると差を指標とした場合と比べて有意差検出率が上昇する傾向は変わらなかった.

次に 3-2) で発生させた比率を対数に変換し、t 検定を行い、1,000 試験分の検出力の平均値と有意差検出率を検討した. 比率の対数値における有意差検出率と平均検出力の関係を、個体間変動を 100% に固定し個体内変動を 20 から 100% に変化させた場合を例として Fig. 4 に示した.

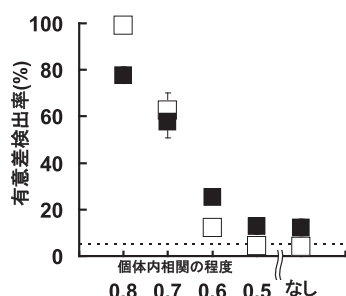


Fig. 3 個体内相関を考慮した同一被験者内比較における有意差検出率

母集団：対数正規分布を仮定

□；指標に差を用いた平均有意差検出率，■；指標に比率を用いた平均有意差検出率

比率を対数値に変換し、個体内変動を上昇させた場合、個体内変動が大きくなるに従い、検出力は低下した. また有意差検出率は、測定値での比較と同じ約 5% 程度で、一定値を示し、比率での比較で認められた個体内変動の上昇に伴う有意差検出率の上昇は認められなかった. 個体間変動を変化させた場合も同様の結果であった (図は省略).

以上の結果から、測定値の比率を用いた場合にのみ、有意差検出率と個体内変動との関係が同一方向に変化するという異常な結果を示すことが明らかとなった. その原因は比率の平均値の分布による可能性を考え、これを検証するために、シミュレーションにより 1,000 試験分発生させた測定値の差、比率および比率の対数について、平均値の度数をグラフ化することにより、それぞれ分布の形状を確認した. 各指標における平均値の 1,000 試験分度数分布を個体間変動は 100% とし、個体内変動は 20% から 100% に変化させてプロットした (Fig. 5).

観測値の差の平均値 (Fig. 5-(a)) は、個体内変動の大きさに関係なく、最頻値はほぼ 0 であり、かつ 0 を中心にほぼ左右対称の分布を示した. しかし比率の分布 (Fig. 5-(b)) は歪みがある上、変動性が大きくなるほど最頻値が 1 から正の方向に外れた. この比率を対数で表現した場合 (Fig. 5-(c)) は、観測値を差で比較した場合と同様に 0 を中心とした左右対称の分布を示し、分布の歪みは改善した. これらの指標における平均値の分布の形状が、t 値の分布にどの程度影響するか確認するために、平均値を標準誤差で標準化し、t 値の度数をプロットした (Fig. 6).

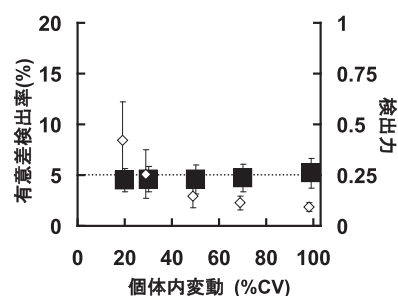


Fig. 4 同一被験者内比較における比率の対数値の有意差検出率、平均検出力と個体内変動の関係

母集団：対数正規分布を仮定

■；平均有意差検出率，◇；平均検出力
破線は有意差検出率 5% を示す

観測値の差の t 値 (Fig. 6- (a)) および比率の対数値の t 値 (図は省略) は、理論どおり、0 を中心とした正規分布に近い分布を示した。ところが比率の t 値 (Fig. 6- (b)) は、個体内変動に伴って最頻値が 0 から外れ、平均値の分布の歪みの影響が見られた。この偏りのため、比率の t 値は個体内変動が大きいほど、 $p < 0.05$ を示す t 値から外れる確率が大きくなることが分布から確認できた。

4. 考察

プラバスタチンの臨床効果を血漿脂質濃度の変化によって比較検討した 4 報告^{2, 3, 4, 5)} は、変化の程度と変動性がほぼ同程度であったにも拘わらず、観測値の比較を t 検定で行っている 2 報では、被験者

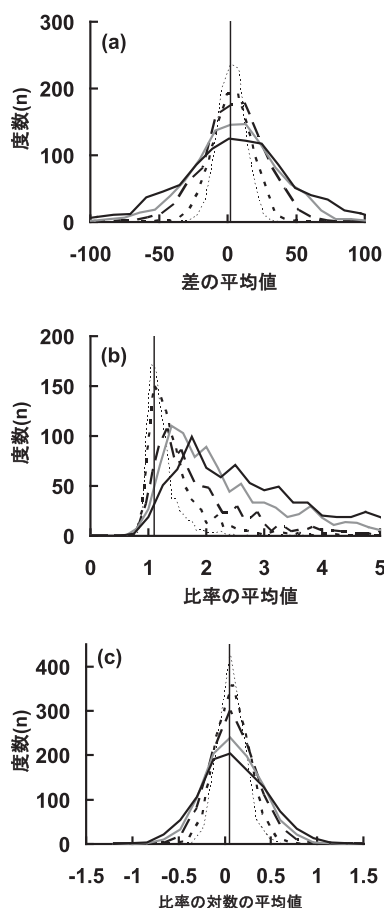


Fig. 5 観測値の差 (a), 比率 (b) および比率の対数値 (c) における平均値の度数分布
母集団：対数正規分布
個体内変動は ; 20%, ; 30%, --- ; 50%, — ; 70%, — ; 100% で示した。黒の実線は (a), (c) では 0, (b) では 1 を含む度数を示す

数が少ないこともあって有意差が認められないという結果になっていたが、比率の比較を行った場合、有意な差が検出されていた。その評価結果の食い違いを生じさせた可能性は、2 つ考えられる。

一つは、測定された平均値の差は 20% 以内であり、しかも、臨床検査値の正常範囲内の推移となっていることから、比較項目を比率に取ることで検出力が過大となった可能性があると考えられる。二つ目は、測定値および比率の分布の問題が挙げられる。正規分布を仮定した同一母集団から得た観測値の差は正規分布を示す。一方比率で表現すると、試験医薬品の効果が対照医薬品より小さい場合は、0 から 1 の範囲の値しかとらないのに対し、試験医薬品の効果が対照医薬品より大きい場合は、1 以上のあらゆる値を取り、分布は左右対称にならないと推定される。更に変動性の大きな集団では、分布の歪みの影響を大きく受け、最頻値が 1 からずれていく可能性が考えられる。 t 検定による比較検定は、正規分布を仮定した検定法であるため、この偏りが t 値の分布の歪みを生み、指標に差を用いたときより、比率で比較した方が有意と判定される確率が上昇している可能性がある。本研究は、これらの可能

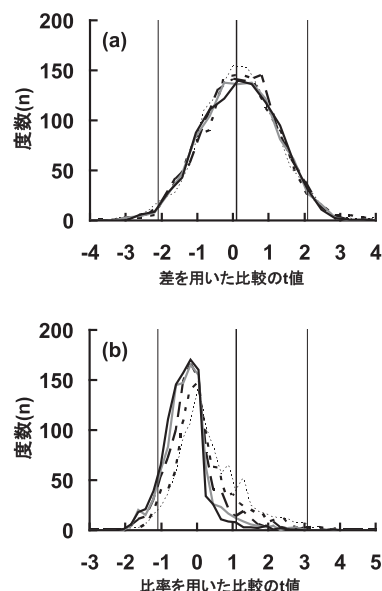


Fig. 6 観測値の差 (a), 比率 (b) における t 値の度数分布
母集団：対数正規分布
個体内変動は ; 20%, ; 30%, --- ; 50%, — ; 70%, — ; 100% で示した。黒の実線は、—— ; (a) は $t = 0$, (b) は $t = 1$, — ; $p = 0.05$ の t 値, (a) は $t = \pm 2.093$, (b) は $t = -1.093, 3.093$ を示す。

性をシミュレーションの手法により検討した。

シミュレーション研究により、同一分布内からデータを20例ずつサンプリングし、それらの差の平均値の比較を行うと、対応のない2群比較 (Fig. 1), 対応のある2群比較 (Fig. 2-(a)) の方法に関係なく、個体内変動および個体間変動の増大に伴って、20%の相違を検出する検出力は低下し、また条件設定通り5%前後の有意差検出率を示した。一方、同一データを比率で表現し検定を行った場合、検出力は個体内変動および個体間変動の増大に伴って低下した。しかし同一被験者内比較では5%より高い比率で有意な差が検出され、個体内変動および個体間変動が大きくなるほど、有意差検出率はむしろ上昇する傾向が認められた (Fig. 2-(b))。この傾向は、個体間変動より個体内変動の程度に大きく影響を受ける傾向にあった。比率の比較で認められた有意差検出率の上昇は検出力の上昇とは関係がなく、観測値を比率に変換したことにより検出力が過剰に増大した可能性は否定された。

二つの測定値が独立に正規分布あるいは対数正規分布に従う場合でも、その測定値の比率に変換したデータは、もとの測定値データの分布と同一であるとはみなせない可能性があるため、ノンパラメトリックな統計手法でもあわせて検討した。しかしWilcoxon符号付順位検定にて検定した結果でも、t検定で検定した結果と同様に、比率で比較すると個体内変動が大きいほど有意な差として検出される頻度は上昇する傾向が得られた (Fig. 2-(b))。有意差を検出する確率にパラメトリックかノンパラメトリックかによる検定手法の相違が与える寄与は、小さいものと考えられた。

同一被験者内の比較では、同一被験者の前後の値に相関性が存在する可能性がある。この寄与の程度を確認するために、個体内相関を組み込んだシミュレーションもあわせて実施した。相関を有するデータを対にし、対応のあるt検定にて検定を行ったところ、個体間・個体内変動を大きく(70%)設定した場合、個体内相関の程度の低下に伴い、測定値を用いた場合と比率を用いた場合での有意差を検出する比率の乖離が認められた。従ってデータの個体内に相関性を規定しても、比率を指標とする比較で有意差検出率が上昇する傾向は変わらなかった

(Fig. 3)。

比率での比較において、個体内変動の増大に伴い20%の差を検出する検出力は低下しているが、有意差を検出する確率は上昇するという一見矛盾した結果を、各指標の平均値およびt値の分布を確認することで説明を試みた。有意差検出率の検討で得られたデータの各指標の平均値の度数およびt値の確率密度をプロットしたところ、比率の平均値およびt値の分布は、観測値の差の分布と異なり、左右対称でなく、最頻値が1からずれる偏りが見られた (Fig. 5-(b), Fig. 6-(b))。この偏りは個体内変動の増大に伴って大きくなった。比率で比較を行うと、t値の分布が、有意差を示す $p<0.05$ の確率にシフトしてしまい、差の場合と比べて有意差検出率が増大することが原因であることが明らかとなった (Fig. 6)。またこの分布の偏りは、比率を対数値に変換することにより改善することが明らかになった。そこで比率を対数に変換して同様に検出力、有意差検出率を評価したところ、個体内変動、個体間変動がそれぞれ大きくなると検出力は低下し、有意差検出率はほぼ5%となった。対数に変換することで、本来差がないデータの場合と同様の結果が得られることが検証できた (Fig. 4)。ノンパラメトリックな検定も同様に行ったところ、比率を対数変換した指標は、変動性の増大に関係なく有意差検出率は5%前後を示し、これらの結果は統計手法の影響を受けないことがわかった。

以上より、プラバスタチンの効果を対象にした比較試験において、比率で統計検定を行った結果認められた検定結果は、比較指標の取り扱いが統計解析上適切でないことにより、第1種の過誤の増大を引き起こしていた可能性が確認できた。比率の値を用いて検定手法により評価を行う場合には、対数に変換した後に検定を行うべきであることが明らかになった。この結果は、生物学的同等性に限らず、普遍的に複数群間の比較を検定で行う場合に適用できると考える。

引用文献

- 1) 塩見真理, 伊藤永久佳, 緒方宏泰. 我が国に於いて行われた臨床試験によるジェネリック医薬品と先発医薬品の臨床的同等性評価のレビュー研究 I:

- 試験デザインの評価. ジェネリック研究. 2009;3:18-26
- 2) 平野勉. プラバスタチンナトリウム先発医薬品から後発医薬品への切替えによる血清脂質の変動. *Prog Med.* 2005;25:2415-7
 - 3) 一森伸二, 三森史朗, 三森佳子ほか. 糖尿病合併高脂血症患者におけるプラバスタチンナトリウム先発医薬品から後発医薬品への変更に伴う血清脂質の変動. *Ther Res.* 2006;27:2271-4
 - 4) 國領俊之, 松本名美, 菅原義生ほか. プラバスタチン製剤先発医薬品と後発医薬品における臨床効果の比較. 医療薬学. 2006;32:912-6
 - 5) 松本名美, 國領俊之, 菅原義生ほか. プラバスタチン Na 製剤の先発品と後発品の臨床的比較. 公立甲賀病院紀要. 2005;8:27-32
 - 6) 永田靖. サンプルサイズの決め方. 東京:朝倉書店; 2003. p.105-28
 - 7) 山田剛史, 杉澤武俊, 村井潤一郎. Rによるやさしい統計学. 東京:オーム社; 2008. p.326-9
 - 8) Venables WN and Ripley BD. *Modern Applied Statistics with S-PLUS 4th ed.* New York: Springer-Verlag; 2002. p.126-32