ジェネリック研究　2008;2:141–161

〔総　　説〕

ジェネリック（generic）医薬品――後発医薬品の生物学的同等性試験 guideline を統計学的視点から理解するための要点・急所

Important Points for Understanding the Guideline for Bioequivalence Studies of Generic Products from a Statistical Point of View

足立　堅一　Kenichi Adachi
㈱サンテック

は じ め に

世間では，統計学的な物の見方・考え方や解釈（専門的用語で表現すれば，「検定や推定による統計学的推論」）を中心として，誤解が少なからず見られる．このような誤解は，結果として弊害を世間に撒き散らすことになり，その典型的な例の１つが，「**NS ＝同等**」**論法**であろう．つまり，「**statistically Not Significant（統計学的に有意でない）**」**との検定結果は，「同等」の検証・証左であるとの論法的誤解**である．世間的には統計学的接近法や手法に関するこうした誤用・誤解などは，減少しつつあるものの，遅々としており，極論すれば，根絶には「前途程遠し！」との感を筆者は抱いている．

根絶の一助になればとの思いもあり，最近，S. A. Glantz, “Primer of Biostatistics”, 6th ed, McGraw-Hill,（足立堅一監訳,「基礎から理解できる医学統計学」，篠原出版新社，2008 年 9 月）を出版させて頂いた．原著者は，California 大学医学部教授で，本人自身が医学部出身で，そのことが，この本の特徴として反映されている．つまり，統計学の専門用語をできるだけ使用しないで，日常用語での解説（例えば，有意水準・帰無仮説などの専門用語の使用を可及的に回避）に挑戦したなどの点で特徴のある統計学入門書となっている．そのこともあって，翻訳出版後の読者からの反響によって，本邦でも医師の先生を中心とした愛読者が多いことが判明した．

この Glantz 教授の原著の根底に流れており，かつ，この総説の主題とも関連し，筆者も共感した視点・見解の代表的例を以下に２つだけ紹介する．

① **statistical significant（統計学的に有意）の意味と世間一般的無理解や誤解**
― Glantz は，これを，“highly prized p-value” と呼んでいる．その真意を e-mail での交信で，「ironical な意味を含めた表現か？」と質問したところ，「そうだ！世間では idolized されてしまっている！」との趣旨の回答であった．
―これは，検定で有意となったことに関する誤解であり，前述の「NS ＝同等」論と双璧をなすものである．
―世間的には，未解消であり，根絶には至っていない．

② SD（Standard Deviation，標準偏差）vs. SE（Standard Error，標準誤差）
―後者に対する，世間一般的無理解や誤解
―世間的には，未解消であり，根絶には至っていない．

* 〒 369-0121　埼玉県鴻巣市吹上富士見 2-6-10-2
FAX : 048–548–7027
E-mail: ken11.ada@ezweb.ne.jp

これらの背景には，検定･推定，とりわけ，「**検定＝p 値」が統計解析の最大の「成果物」と言わんばかりの思想**の流布がある．そして，その元凶を発掘すれば，それは極論すれば，「**SE の意味・意義の『誤解・無知』，検定における SE の役割り・意義の『無知』」に由来**するものでもある．

翻って，ジェネリック（generic）医薬品研究分野の現状を観るに，こうした世間一般的な混沌とは無縁であり，少なくとも guideline（指針）に限れば，他分野に比較して，先進的であると断言できる．これは，決してお世辞ではなく，旧 guideline 時代から然りである．先駆的には，江島らの論文[1]や，この専門雑誌の編集委員長でもある緒方宏泰先生などのご尽力も寄与していると思われる．

ジェネリック（generic）医薬品研究分野では，どのような点が先進的かと言えば，ご承知の通り，この分野では切実な統計学的急所・問題と断言できる，「**NS ＝同等」論なる理不尽な論理が「禁じ手」になっていること**である．確かに，初期の論文である江島らの論文などでは，この「禁じ手」に関して，完璧な「封じ手」を設定しているとは言えないものの[2]，既に「禁じ手」の問題点を明確に指摘し，そのための手段は準備されていた．

さて，今回の総説の位置付け・focus について以下に述べる．

①後発医薬品に関連する guidelines を概観して，統計学的視点から要点・急所を中心に解説すること
　― guidelines で推奨されており，極めて妥当かつ実用的と筆者が信じる「線形補間法」などの，妥当性や実用性に関する理由を含めた解説
②この総説の解説は，「後発医薬品の生物学的同等性試験ガイドライン」を中心にし，かつ「生物学的同等性試験」に関する統計学的な事項について解説すること
③「NS ＝同等」論の欠陥（盲点･論理的な誤魔化し）について，その理由も含めて解説すること
　―それを克服した「同等性検証法」について，何故克服されるのかの理由を含めて解説
　―検定法（2 つの片側検定！）と区間推定法との 2 つの接近法とその考え方についての解説
　―他分野での通常の検定法・推定法と比較して，同等性検証での視点的違い・勘違いに陥りそうな点についての解説
④ cross-over 試験について，design 的視点からの特長や，その実体が，しばしば誤解される carry-over effect の素顔などについて解説すること
⑤「対数変換」の理由や処理法，2 薬剤の比較において，「差」だけでなくて「比」が出現する理由，同等性「許容限界」という概念の必要や，その意味や留意点などを解説すること
⑥必要例数（sample size）の設計の急所と指針での扱いや，他分野での算出法と比較した留意点などを解説すること

1．後発医薬品関連 guidelines とその概観

大半の読者には，周知のことではあろうが，最新版の関連 guideline を表にまとめてみた（**Table 1**：以下，Table は 157 ページよりまとめて掲載）．本命は，当然ながら「後発医薬品の生物学的同等性試験ガイドライン（2006 年 11 月 24 日），以下，「後発医薬品 guideline」であり，その他には，それらの正式名称は **Table 1** を参照することとして，「経口固形」製剤の「**処方変更**」，「**含量の異なる**」「経口固形」製剤，それに「**局所皮膚適用**」製剤のためのもの，合計 4 種の guideline である．

製剤の同等性検証用試験方法の視点から眺めると，「後発医薬品 guideline」以外の，「**処方変更**」と「**含量の異なる**」**製剤のもの 2 つは，ヒト試験が免除できる条件を処方の変更程度と「溶出試験」結果から判断する規準を示したもの**であり，「局所皮膚適用」製剤のものは，製剤の多様性から，試験方法も一律ではないことが判明する．なお，「局所皮膚適用」製剤の guideline 以外の，これら 3 種類の guideline については，この専門誌に，当該「**Q&A**」も含め，全文が転載されており，参考になる[3]．

この総説においては，「**溶出試験」については，簡単な解説，「後発医薬品 guideline」での同等性検証用試験方法について，力点を置いた解説**をする．

2．「後発医薬品の生物学的同等性試験ガイドライン」とその概観

これまた，大半の読者には周知のことではあろうが，最新版の「後発医薬品 guideline」の概観を表

にまとめてみた（**Table 2**）.

まず,「バイオアベイラビリティ」を初めとしたkeywordsとしての代表的用語とその定義とを,**Table 2-1**に示す．製剤間の「バイオアベイラビリティ」の比較が困難な場合の，特別な同等性検証試験方法としての「薬力学的試験」や「臨床試験」についての定義もなされている．また,「**溶出試験vs．生物学的同等性試験」の定義と位置付け**も把握しておきたい．つまり,

①**原則，最終的判定手段は健康成人被験者を対象としたヒト試験による「生物学的同等性試験」であること**

②**溶出試験は試験条件の設定（標準製剤ロットの選択，被験者の選択基準）のため，あるいは，一部，評価の補助的手段として用いるのみであること**

である．なお，以後の表中には，筆者の挿入した解説・注であることを他と区別するために，当該箇所に「★」を付与しておく.

次に，製剤の特性に応じた同等性試験方法の選択基準や，試験実施のflowや同等性の判定基準などを,**Table 2-2**に示す．ここでの要点は,1つには,「**経口製剤 vs. 非経口製剤**」の視点であり，非経口製剤では,製剤の特性に応じて試験方法が分かれる．「局所皮膚適用」製剤では,別途,当該guidelineに従い，作用部位が皮膚そのもの，つまり，皮膚角層通過が作用発現に不要な製剤では，例外的に「動物」試験も可となっている．**原則,「静脈注射」製剤だけが,生物学的同等性試験「免除」**であることが分かる.

Table 2-3には,「**生物学的同等性試験**」と「**溶出試験**」に採用されている判定指標を中心に，それらの意味や，合格／不合格の判定基準を，筆者の解説を交えて,「後発医薬品guideline」からpick-upした.

以後,「溶出試験」での基準である「溶出率」や「lag時間」や「f2関数」などについては,「検定」や「推定」などの統計学的な判定手段は採用されていないため，簡単な解説にして，主として生物学的同等性試験について解説する.

合格の判定基準の代表例としては,「**生物学的同等性試験**」では，**AUCやCmaxを指標**として，**両製剤間の「比」**について,その**90%信頼区間が,0.80～1.25**に入ることである．「**溶出試験**」では，多様ではあるが，**代表的には,「lag時間」は,「標準製剤」と「試験製剤」との「差」が10分以内，平均「溶出率」が85%以上**などである.

表中にも「注釈」を挿入したように，lag時間や経時的平均溶出率や，AUCなどの算出方法として,「線形／直線補間法」として「内挿法」が採用されている．筆者は，これらが簡便であり，かつここでの目的に照らして十分に妥当であると同感する者である．**重要なことは，測定pointsを，目的に応じて必要かつ十分な時点数を確保**することが先決である．他の分野でしばしば見られることだが，無配慮にも，極めて不十分な測定時点しか設定･測定せず,その「付け・負債」を，みだりに，いわゆる「**非線形（non-linear）補間法**」とかの高尚な手法での解決に委ねようとする傾向がある．本末転倒である.必要で十分な測定時点を確保・試験すれば,「非線形（non-linear）補間法」も,計算不能などとならず,無理なく機能し，かつ，直線補間法での結果と大差はないものとなる．その逆に，**少なからぬ「付け・負債」の発生する場合には，補間の信憑性・妥当性が懸念されるばかりでなく，しばしば，計算不能が発生**することを警告しておきたい.

3.「生物学的同等性試験」の判定基準などを深く理解するために必要とされる統計解析的概念・視点についての要点・急所

以上の予告編的に既述した話題を，更に具体的に記載･まとめてみた（**Table 3**）．筆者流の「Q & A」方式で，問題の所在を明確化したり，世間にしばしばある偏見・誤解も紹介して，読者がそれらに感染していないかを点検できるように試みている.

以下，各論的に，統計学的な「落とし穴」を指摘することで，読者の「落下」を回避することにしよう．また，参考文献[4]も紹介しておく.

3．1 「NS＝同等」論という「落とし穴」の原因—それは，統計学的検定の仕掛けについての理解不足

「Q & A」方式で，問題の所在にspotを当てながら解説しよう.

Q01：統計学的仮説検定で，有意となるための条件は何か？

A01：代表例として，t 検定について考える．

①２群間に，実際に差 δ が存在すること．有意となり易いためには，δ が大きいこと

②有意となり易いためには，バラツキ，つまり，SD（σ）が小さいこと

③有意となり易いためには，例数が大きいこと

―①を，統計学的用語で解説すれば，どうなるか？と「Q」としたいところであるが，紙面もあるので先を急ぐ．解答は，「**帰無仮説**」：**不成立**，「**対立仮説**」：**成立**である．

上記「A（Answer）」を，**t 検定の式（t 統計量）**から説明しよう．

$$t = \frac{m_x - m_y}{\sqrt{\frac{\sum(x_i - m_x)^2 + \sum(y_i - m_y)^2}{2(n-1)} \times (2/n)}} \qquad \textbf{式（1）}$$

ここで，

x 群：$x_1,\ x_2, \ldots,\ x_i, \ldots,\ x_n$

y 群：$y_1,\ y_2, \ldots,\ x_i, \ldots,\ y_n$

$m_x = \Sigma\, x_i \,/\, n,\ \ m_y = \Sigma\, y_i \,/\, n$

$d_m = m_x - m_y$

さて，「**差 $\delta \neq 0$」つまり，帰無仮説の不成立下でのこの式（1）の挙動**を洞察しよう．

①統計学的に有意となるには，この t が 1.734（絶対値）*とかのように「0」からできるだけ乖離することである．

＊：x 群と y 群の 2 群，例数 n = 10 ／群の時に，両側有意水準 α = 10 % を与える値，percentage point（棄却域）

②上式の分子は，何か？ x 群と y 群との「平均値」の「差 d_m」である．

③問題は，**分母**である．「×」記号の左側は，data のバラツキ，つまり 2 群の SD を併合したものである．これは，**例数 n とは無関係で，母集団のバラツキ σ に依存**する．次に，最大の急所であるが，「×」記号の右側では，**その SD を２／n として，n で除算**していることである．

―詳細は，参考文献[5]を参照してもらい，これの√，つまり**分母全体は，分子 $d_m = m_x - m_y$ の SE（標準誤差）**である．SD と SE とは，ご承知の通り，両者は，以下の関係にある．

$SE = SD \,/\, \sqrt{n}$ **式（2）**

かくして，式（1）において，

④t（絶対値）↑として，「有意」とするためには，

分子：d_m（絶対値）が大

分母：SD を小，n を大

とすれば良いことになる．換言すれば，

⑤統計学的に有意となることの仕掛けと意味とは，これら**３種類の「役者」による「合作」であること，換言すれば，１人の役者の「産物」ではないこと**である．つまり，母集団において，「差 δ」が小であっても，例数↑すれば，有意となり，逆に，「差 δ」が大であっても，SD（σ）が大で，例数小であれば，有意となり難くなることを意味する．

＊：参考までに補足しておくが，通常の統計学的教科書は，「帰無」仮説（つまり，$\delta = 0$）下での検定の仕掛けの解説（**Fig. 1**）で終了していることが多く，その場合には，上記「対立」仮説下での結論とは異なる結論になる．

Q02：統計学的検定で，NS とすることで，「NS = 同等」論に意図的に操作するための方法は何か？

A02：以上の仕掛けの解明を踏まえれば，最早，解答するのは簡単至極である．

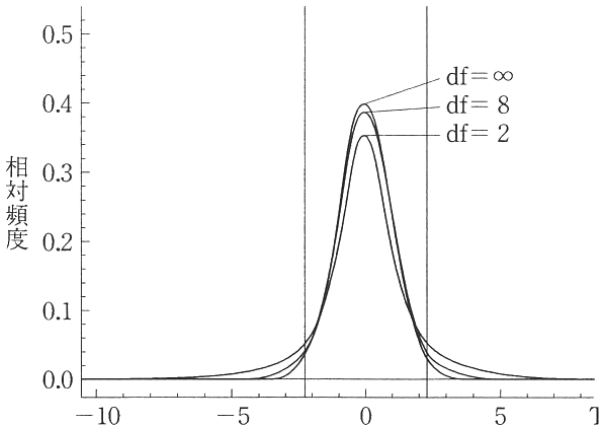
①つまり，例数を極力減らすこと

②SD の制御を極力せずに，大とすること

である．これは，大変に理不尽であり，筆者流の

Fig. 1　帰無仮説下での t 分布の形状と特徴

（図中の df は自由度）



比喩で言えば，「仕事をしなければしない程，給料が上がる」ような矛盾である．

3．2　統計学的検定の仕掛けと「SD vs. SE」

一部復習になるが，data から獲得した SD は，母集団 σ の推定値であり，例数には依存しない．逆に，SE は，**式（2）**から判明するように，例数 n 依存であることに注目したい．こうしたことに関する理解不足に起因して，SE に関する **Table 3** に紹介したような表層的理解や誤解が，世間には少なからず存在する．SE の特性の掌握には，極端な場合を考察した方が理解し易い．$n = \infty$ の場合では，$\sigma \neq 0$ ならなんであれ，$SE \Rightarrow 0$ へと収束する性質であることが判明する．つまり，統計学的有意性と SE，例えば，**式（1）**の分母とは，この観点からも極めて深い関係にあることを銘記したい．2.5 万例という data の解析機会を得たが，大抵の比較対象因子に関して有意となった．論文などの critical review なり critical appraisal には，こうした力量が必須になる．

3．3　統計学的検定の仕掛けと「統計学的有意性 vs. 専門学的有意性」

ここまで理解が至ると，「t 値大 = p 値小」が，比較群間の「差」の大きさを一意的に示す指標にはならないとの洞察に至ることができる．たとえ，p 値 = 0.00001 であっても，推定された「差 d」がどの程度であるのかが問題なのである．分別を必要とする case には，例数の多くない治験や臨床試験では，殆ど遭遇しないであろうが，data mining の世界では日常茶飯事であるということが，今や容易に推測できよう．理由は，**膨大・大量の「例数（情報）」のために，微小な「差」であっても全て有意**となるのである．その逆も然りで，観測された「差 d」がかなりの大きさなのに，例数小の場合には，p = 0.15 もある．

こうした悟りの境地に至れば，然るべき「差 δ」とは，どの程度かとか，その差を検出するに，必要かつ十分な例数はどの程度かとの発想・智慧の重要性を悟る道が間近に見えて来る．

再度，明確に，かつ力説したい点は，**専門学的有意性という概念導入の推奨**であり，それは，当該専門分野毎に，「専門学的有意性」として，例えば，「意味のある差は如何ほどか」との感覚を磨くことであり，できるだけ，それを公的基準化することであろう．そして，もう 1 つ重要な点は，これは，**統計学者が決定するものでなくて，各分野の専門家が決定すべきもの**なのである（しかし，残念ながら，概して，彼ら専門家にそうした意識はないのが普通である）．

Q03：同等性試験での基準として，このような「専門学的有意性」の基準は何であろうか？　考察すること

A03：まず，以上の仕掛けの解明を踏まえて，読者にこうした意識を持ってもらうことにする．後ほどの解答としよう．

3．4　検定 vs. 推定

―「推定」が「検定」よりも，情報量が多いという意味とは？

―「推定」vs.「検定」の「等価的」側面とは？

検定において，Glantz 流表現では，“highly prized” p 値，筆者流表現では，「吹けば飛ぶような p 値の数に！」懸命になる危険性は理解頂けたであろう．彼は，原著において，「検定よりも推定が more informative」と記述している．筆者も参考文献 [5] において，それを知らずに独立に，偶然に，同一用語を使用している．**検定の有するこうした限界を，完全とは言えないまでも，それなりに克服するのが，実は「推定」**なのである [6]．

「**推定**」のための道具としては，「**点推定値**」と「**信頼区間（Confidence Interval，CI）**」がある．後者は，検定統計量も勘案して算出・作成される．それ以外に必要なものは，t 検定での例では，既述した，当該自由度における **percentage point（棄却域）**である．難解なことではなくて，t 統計量の例では，2 群，例数 n = 10 例 / 群であれば，自由度 df = 2（n − 1）= 18 の時には，両側 90% では，$t_{df=18} = 1.734$ となる．帰無仮説下では，この大きさの値を t 値の絶対値が超えることは，10% 以下となる値である（帰無仮説下なので，**式（1）**の分子，従って，t 値の期待値は，「0」になることに注意すること）．

検定で有意となることと，この t_df との間には，検定をした者は，容易に理解できるというか，当然

であるが，**式（1）**の data から算出した t 値を，t_cal とすると，

t_cal ＞ 0 のとき，t_cal ＞　t_df ⇒ 有意　　**式（3）**
t_cal ＜ 0 のとき，t_cal ＜ − t_df ⇒ 有意　　**式（4）**
とする．

ここで，更に，**式（1）**において，

分子 = $m_x - m_y = d_m$

分母 = SE_d_m（d_m の SE との意味）　　**式（5）**

と表記すると，**式（3）**は，

t_cal ＞　t_df
⇒　$t\_cal = d_m / SE\_d_m$ ＞　t_df
⇒　d_m / SE_d_m ＞　t_df
⇒　d_m ＞　$t\_df \times SE\_d_m$
⇒　$d_m - t\_df \times SE\_d_m > 0$　　**式（6）**

ここでの要点は，

① $t\_df \times SE\_d_m$ を「**信頼限界（Confidence Limit, CL）**」と呼ぶ．そして，

$CI = d_m - CL \sim d_m + CL$　　**式（7）**
CI 下限　　**CI 上限**

このとき，○%信頼区間は，有意水準 α と連動しており，100% − α % =「○%信頼係数」の関係がある．例えば，有意水準 10%の場合には，100% − 10% = 90%信頼区間となる．

②**検定：有意となること = 差の信頼区間：「0」を含まないこと**

―**式（6）**の例では，信頼区間下限が「0」よりも上にあり，含まない．

―これを，筆者は「**等価的**」**側面**と呼んでいる．

Q04：**式（4）の場合について，同様に考察すること**
A04：結果は省略

さて，いよいよ，**推定が more informative との意味**を理解しよう．**Fig. 2** を用いて説明する．図では，言うまでもないが，d_m ⇔「●」，その下に伸びた「ヒゲ」が，**式（6）**の「$- t\_df \times SE\_d_m$」，つまり，**式（7）**の「− CL」に該当する．上に伸びた「ヒゲ」は，「$+ t\_df \times SE\_d_m$」である．

Q05：**Fig. 2 の 4 つの場合について，どんな情報が読み取れるか，考察すること**

A05：解釈は，一意的ではない．

①上記の判明事項から，「0」を跨がないこと⇔有意であるので，図左から，第 1 番目，第 2 番目，第 4 番目は「有意」，第 3 番目は「NS」であることが，p 値なしで判明する．

②**信頼区間（限界）「$t\_df \times SE\_d_m$」の SE_d_m** に注目すること．例数↑すれば，これが↓して，**例数 n = ∞では，棒は消滅**する．

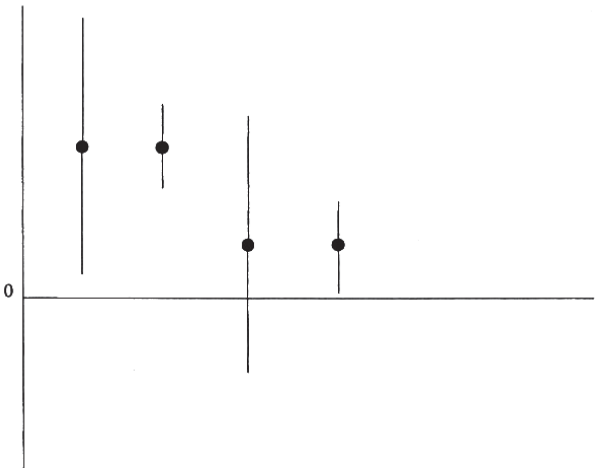
―これから推察すれば，第 1 番目と第 4 番目とは，95%信頼区間であれば，計算される p 値は，ほぼ 5%，例えば，4.35%というようなものであろう．これに対して，第 2 番目は，p = 0.0001 などであろう．

③もう 1 つの注視すべき点は，前者 2 つ（第 1 番目と第 2 番目）は，差は同程度で，後者 2 つ（第 3 番目と第 4 番目）も同程度であるが，前者＞後者でありそうなことも読み取れる．差 δ が観測されたものと同様として，バラツキ σ も 4 者で同様とすると，第 4 番目は，かなり無理して，小さい差を例数↑して有意としたであろうと考察できる．

④別の視点，つまり，「**NS = 同等**」**論の意図的操作をするとしたら，第 1 番目の例において，例数をもっと↓するという手抜き試験をすれば，不正に（？）「同等」が検証できる**ことになってしまうが，**信頼区間での監視をすれば，そうした不正は察知可能**なことも理解できる．

―ここでの急所は，**こうした情報は，p 値だけに依存する限り，獲得できない**のである．

Fig. 2　差の区間推定および検定での S・NS との関係
d_m：観測された差



4.「生物学的同等性試験」の判定基準などを深く理解するために必要とされる統計解析的概念・視点についての要点・急所　―　各論・具体論

同等性試験と直接的に関連してくる問題についてまとめてみた（**Table 4**）.

4. 1「許容限界」という概念とその導入の必要性

「許容限界」という概念は，通常の検証，つまり，「勝った」というようなことを主張する「**優越性（superiority）」検証**では，さほど表面化しない．しかし，優越性検証においてもこの概念は，既述したように，「どの程度の差」であれば「専門学的有意性」があるのかという概念と関連している．そして，その場合の適切な用語としては，「最小」有意差（これより「大」であれば，これ以上「小」でなければ，専門的観点からも「有意」）なる用語であろう．ただ，**優越性の検証での差が大きい程，好ましいとの視点と異なり，同等性の検証の場合では，差がないのが理想的であり，「許容」限界**との用語になる．筆者流に，両者での概念を表現すると，前者が「最小」「許容限界」（当該限界値より『大』であるべし！）に対して，後者および次に紹介する非劣性検証では，「最大」「許容限界」（当該限界値より『小』であるべし！）となる．

他分野であるが，**非劣性の検証**でも，「負けても，この程度以下」との視点から「**許容限界**」との命名が相応しい[7]．非劣性（non-inferiority）なる概念は，最近誕生したものであり，10 年程度が経過しているに過ぎないことに注意したい．また，許容限界なる概念と推定，つまり信頼区間法との相性も良い．**Fig. 3** を用いてそれを解説しよう．

①「非劣性」の場合には，「勝つことへの歯止め」は不要である．負け方の度合いが主題とされ，これが，「非劣性許容限界」となる．図では，縦軸「0」のもっと下側の－Δ位置の横線「―」が，それである．第 3 番目，第 5 番目は，NS であるものの，－Δより上に，その信頼区間（CI）下限があるので，「非劣性」が検証されたことになる．第 7 番目はできなかったことになる．

②最後の第 8 番目を注視したい．これは，「負けた」にもかかわらず，「非劣性」検証ができたことになる．このとき重要なのは，やはり，「**許容限界」の意味を再度考察することである．負けても，－Δ以内の程度であれば，許容できたとの意味では不自然とは言えない．**

③また，②を信頼区間の視点から考察すると，こうした data について，**非劣性を検証するには，例数↑が要求されるために，現実的には，実施が困難になることでの歯止めもかかる**であろう．つまり，「**NS ＝同等」論での例数↓する程，同等が主張できる状況とは，事情が正反対であることにも注目**したい．これは，同等性検証のときも，同様に考えることができる．

4. 2 「同等性」検証での「許容限界」という概念とその導入の必要性

ここまで理解できれば，「同等性」「許容限界」は，簡単である．上記の－Δに対応して，上下±Δが必要になるだけである（**Fig. 4**）．前節の場合と同様に，**「許容限界」に収まっていれば，たとえ，「0」を跨がないでも，「同等」と判定されることを納得**されたい．例えば，**Fig. 4** での第 8 番目は，当該許容範囲に収まるので，「0」を跨がず，有意に負けているものの，同等性が検証されたことになる．ここでも，銘記すべきは，例数が不足すると，ヒゲとしての CI が伸びてしまい，±Δ内に収まらなくなることである．

Fig. 3　差の区間推定およびΔと非劣性の検証

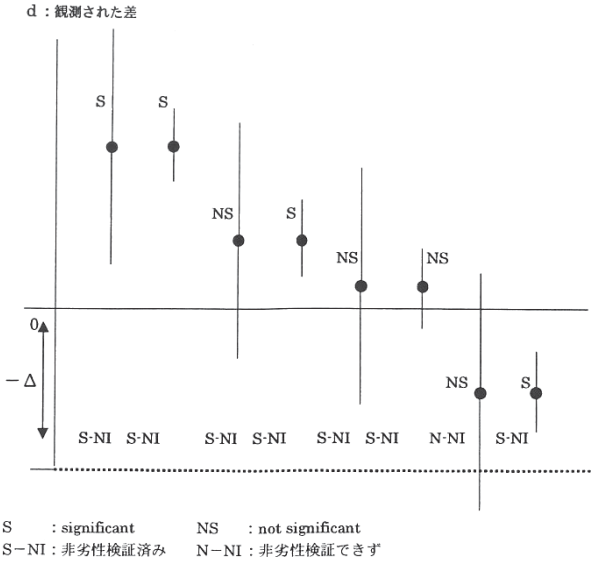
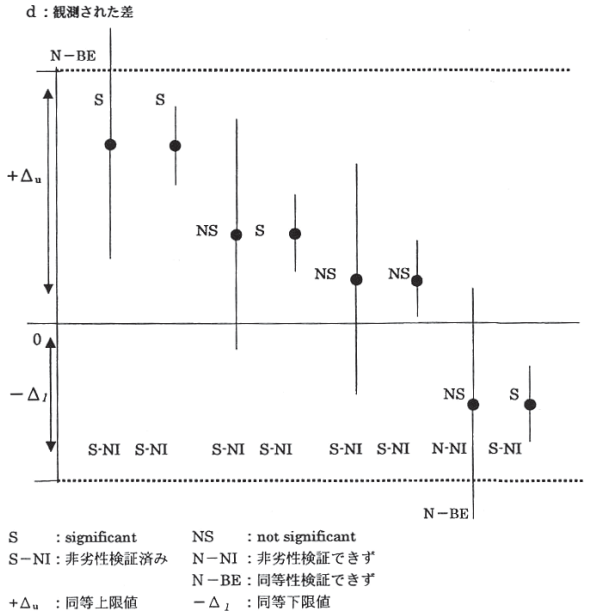


Fig. 4　差の区間推定およびΔ_u・Δ_lと非劣性・同等性の検証



「許容限界」なる概念を修得できたので，以下では，途中ではそれと関係のない，統計学に関連する概念・手法に関する解説も加えながら，最後にこの分野で特有な「2つの片側検定」の意味や意義などの統計学的検証法を解説する．

4．3 「cross-over design」の妥当性 vs. carry-over effect

これについては，**Table 4** の第1項目で解説を含めて記載した．注視すべきは，

① **carry-over effect（持ち越し効果）**なる（薬理学的）概念と**交互作用（interaction）**という実験計画法上の概念との関連性などで，世の教科書にも，用語的・概念的にも少なからぬ混乱があると感じる．

②更に，始末の悪いことに，交互作用とは，類似点はあるものの，**別の概念である薬理学での「interaction」は，一般には「相互作用」**と訳されており，両者を混同して交互作用を相互作用としている訳本もある．

③更には，「**時期効果**」と「**順序効果**」との混同も見られる．

以上，用語の定義の問題ではあるものの，やはり，弁別・識別の困難な類似概念であることが，こうした混乱・混同の大きな原因であろう．これらについての解説は，ここでは割愛する．関心のある読者は多かろうと思うので，参考文献[8]を参照されたい．

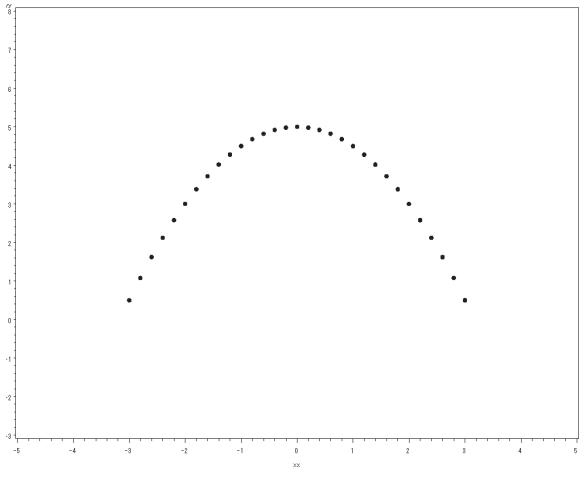
4．4 「線形補間法」としての「内挿法」の妥当性

validation なる概念は，当然ながら，valid や validity の同類語である．validation は，「検証」などと訳されることが多いが，「目的」に対応して，そのために必要とされる性能を確保することである．性能の具体的内容としては，目的に見合う「精度」や「正確度」であり，それを確保することである．その意味からの訳語としては，「妥当性」の保障・確保が validation との解釈が，それこそ，「妥当」であろう．validation の考え方に関しては，分析法などで，別件ながら，この guideline でも記載されている．

内挿法について，guideline では，測定時点の数・時期についても適切な指示がなされており，それを遵守する限り，妥当性は確保できると思われる．

当然ながら，**直線補間法は，本来的には，線形でない，つまり非線形な現象に関して，線形・直線近似をすること**である（別段，線形的現象の補間に対しても OK）．ここでは，**lag 時間，経時的平均溶出曲線，AUC という非線形な現象についての補間**ということになる．変曲点前後でない個所においては，十分な測定点があれば，補間精度は妥当なものとなることが期待できる（**Fig. 5**）．図では，非線形な

Fig. 5 「線形補間法」としての「内挿法」の妥当性



2次曲線の例．横軸 x，縦軸 y

理論曲線として，単純な 2 次曲線について，線形補間を考えてみる．近接する 3 点の中央の点が，欠測などの場合でのその「補間」を想定すると，

前の点の座標（x_b，y_b），

後の点の座標（x_a，y_a），

その間に位置する任意の欠測値存在点（x_r，y_r）として，y_r を欠測値とすると，補間は，

$\Delta_x = x_a - x_b$，$\Delta_y = y_a - y_b$

$\Delta_{r\_x} = x_r - x_b$，$\Delta_{r\_y} = y_r - y_b$

$\therefore \quad \Delta_x : \Delta_y = \Delta_{r\_x} : \Delta_{r\_y}$

$\Delta_y \times \Delta_{r\_x} = \Delta_x \times \Delta_{r\_y}$

$\Delta_{r\_y} = \Delta_{r\_x} \times (\Delta_y / \Delta_x)$

$y_r - y_b = \Delta_{r\_x} \times (\Delta_y / \Delta_x)$

$y_r = \Delta_{r\_x} \times (\Delta_y / \Delta_x) + y_b$　　**式（8）**

となる．x_r が欠測値でも，当然，同様に計算される．図の例では，近接する 3 点で，中点が欠けても十分な補間が可能であり，もう 1 つ間隔を空けた欠測値でも，精度の↓は，微々としたものであることが分かる．

なお，ここで「欠測値」としたものは，測定すべき時点であるにもかかわらず，測定しなかったという本来的定義よりは，広義の意味で用いた．一方で，同等性検証試験とは離れた他の分野の試験，とりわけ，臨床研究では，本来的欠測値は，必発するとして過言ではない．その典型例で，しかも最も悩ましいのが，「経時的測定 data」である．これに対しては，分散分析法を発展させた解析方法が世間一般的には採用されている．この手法は，数学的には高尚で精緻な手法かも知れないものの，混合効果模型を採用しない従来の解析法では，欠測値を有する個体は，当該個体の残り data 全てが解析に使用されないことになる．これらの手法の何れにせよ，最大の難点は，分散分析手法が有する，帰無仮説としての「一様性」の検定から得られる結論，つまり「一様でない＝どこかが違う・pattern が違う」というものと，研究の現場が要求する結論とに乖離があることであろう．こうしたことを考察すれば，AUC，Cmax などのような着眼点に立脚した指標は，多少の欠測値に対しては robust であり，しかも，このような乖離も軽減する方向であるとの点から，筆者は，他の分野でも，積極的に活用すべきだと考えている．

4．5 「対数変換」の必要性とその是非の基準

その必要性とその是非の基準は，極めて単純明快である．**対数変換することで，原 data の分布が正規分布に接近するかどうかが基準**である．統計学的理論は，t 検定に代表される如く，正規分布することを前提にして構築されているので，標的現象がそうでない場合には，理論的不整合が発生する．このことが理解できれば，常用対数変換すべきか，自然対数変換すべきかとの類の疑問にも，即答可能となる．逆に，本来正規分布する data・現象を対数変換してはならない．

生物現象は，対数正規分布をするものが意外と多い[9]のであるが，そのことが，特に他の生命科学分野の研究者には，これまた意外と認識されていない．

4．6 「0」や「定量限界値」の「対数変換」

これについてだけは，細かい点ではあるが，guideline での指針とは異なる方針を提案する．**Table 4** の第 4 項に提示した通りである．つまり，対数変換後の値をも「0」のままにするか，または，**定量（測定）限界値／2 とすることが better** である．

Q06：**定量（測定）限界値／2 とすることの論理的根拠を考えたことがあるか？　考察すること**

A06：実は，これまた，**1 種の「線形補間」法と筆者は考える．**定量限界値の処理法については，参考文献[10]を参照（しかし，そこでは 1／2 は紹介していない）

4．7 「推定」と「検定」との等価性の simulation による検証

筆者個人としては，推定の検定との等価的側面を活用して，「許容限界」などが，明快に解説できると確信しているし，ここでもそうして来た．しかし，信頼区間法に加えて，この分野で特有に出現する，「2 つの片側検定[11]」について，その素顔を解明しておくことは必要であろう．まず，**推定と検定との等価性を，simulation によって検証**しよう．既に，「3. 4　検定 vs. 推定」において，式を用いた解説をした．適宜復習されたい．

この simulation では，同一の正規分布に従う母集団（平均 $\mu = 0$，バラツキ，つまり標準偏差 $\sigma = 1$. 但し，cross-over design ではなくて，通常の design を想定している）から，10 例／群を無作為抽出して，**式（1）**などを利用して，t 値や信頼区間を算出･作成したものである．このそれぞれの抽出を「試行」とする．○：mean，上側●：両側 90% CI 上限，下側●：両側 90% CI 下限，☆：t_cal 計算値をそれぞれ示す（**Fig. 6-1**）.

次に，考察に直接関係のない，○：mean を除去した **Fig. 6-2** で，以下を考える.

①t 値が上下の縦「—」を逸脱するとき，「有意」となること

②信頼区間が「0」を跨がないことが「有意」となること

に注目すると，

—信頼区間が「0」を含まない⇔t 検定で有意との等価性の注目！

具体的には，

下限＞0⇔t＜両側 90% CI 下限．図では，左端の 4 つの case

上限＜0⇔t＞両側 90% CI 上限．図では，右端の 4 つの case

という「等価性」が確認できる.

4．8 「同等性検証」vs.「信頼区間」法

「優越性」の検証においての，上記の「検定」との等価性の概念を活用した「信頼区間」を指標とする方法は，「同等性」の検証では，どのようになるのであろうか？ それは，極めて理解が容易であろう．つまり，**±許容限界範囲に，その信頼区間が収まっていることを同等性が検証されたとする**というものである．この基準に従うならば，**Fig. 6-2 において，例えば，「± t_df = ± 1.734」を「許容限界」とした場合**には，**左端の 1 個と右端の 2 個を除けば，同等性が検証**されたことになる．これに対して**「0」を含まないのは，左右合計 8 個であり，「±許容限界範囲に，その信頼区間が収まっている」ものと「『0』を含まない」ものとの両者の関係は，実は，Fig. 6-2 のように設定した例では，概略，倍増・半減**の関係にある．理由は，上限が 0 ＜となったものであっても，それら値に対応する下限値達が下限棄却値外に突出することは，確率的に低下し，その逆の場合も，下限が＞0 となったものについても，同様に，上限が上限棄却値外に突出することは，確率的に低下する.

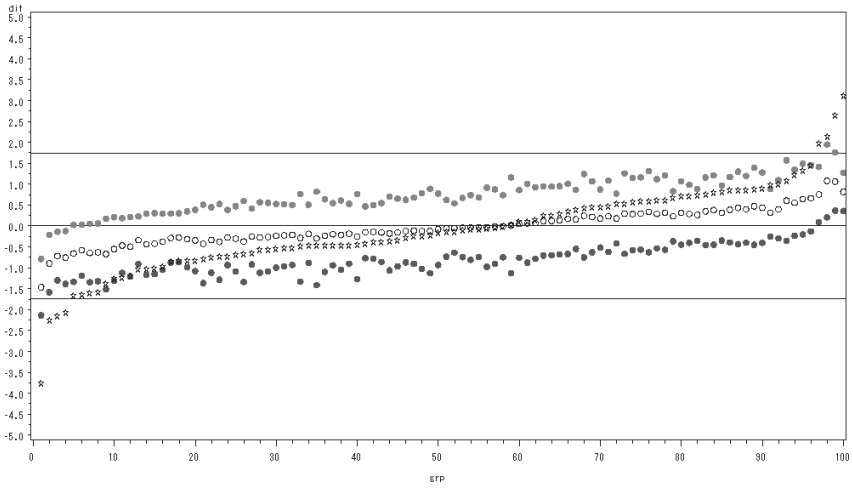
ここで**注意・弁別すべきこと**がある．それは，**上記の「± t_df = ± 1.734」を「許容限界」とした例に限定して言えば，同等性検証では，通常の検証，つまり「優越性」検証の時と比較して，検証される個数が変化し，増加する**ことである．この変化は，検証基準を，

①信頼区間が「0」を含まない（検定では，棄却範囲内である）こと

から

②±許容限界（上記の例では「± t_df = ± 1.734」）

Fig. 6-1　○：mean，上側●：両側 90% CI 上限，下側●：両側 90% CI 下限，☆：t_cal 計算値との 3 者の関係

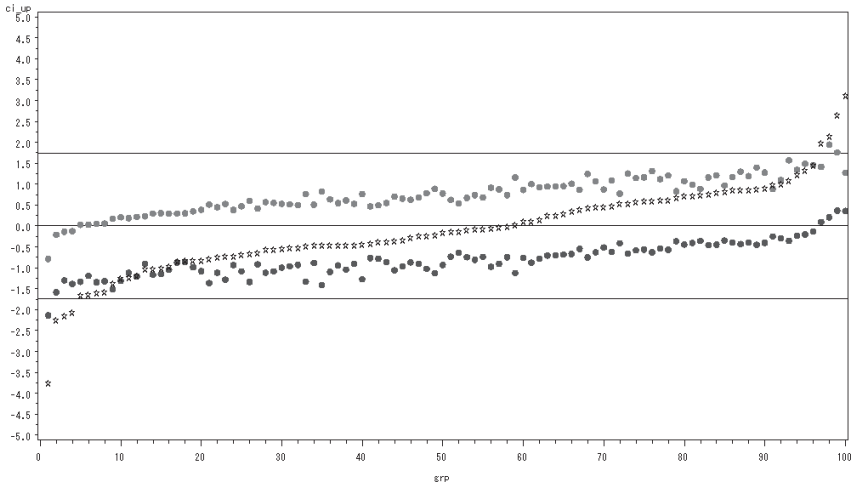


CI 両側 90%（片側 95%）：信頼限界値 CL ＝ t_df × SE_ δ_μ ＝ 0.776（df ＝ 18⇒t_df ＝ 1.734）．縦軸の上下の「—」で示す範囲の値は，t_df ＝± 1.734

—横線：「0」，± t_df ＝± 1.734 の 3 本．横軸は，第 1 試行～第 100 試行

—図は，見易いように，t_cal 値で sort

Fig. 6-2　上側●：両側 90% CI 上限，下側●：両側 90% CI 下限，☆：t_cal 計算値との 3 者の関係



CI 両側 90%（片側 95%）：信頼限界値 CL ＝ t_df × SE_d$_m$ ＝ 0.776（df ＝ 18 ⇒ t_df ＝ 1.734）．縦軸の上下の「—」で示す範囲の値は，t_df ＝± 1.734．第 4．8 節では，これらの± t_df ＝± 1.734 を「許容限界」の一例として解説した．
—○：mean を除去

範囲に，その信頼区間が収まっている
へと変更したことに起因する．

既に，意識されたであろうが，前述の優越性検定基準は，**同等性の検証のためには，変更する必要がある．つまり，上記②に対応すべく変更が必要**なのである．換言すれば，**Fig. 6-2** の t_cal 値☆は，帰無仮説：$\delta = 0$ のためのものであり，そのままでは流用不可なのである．

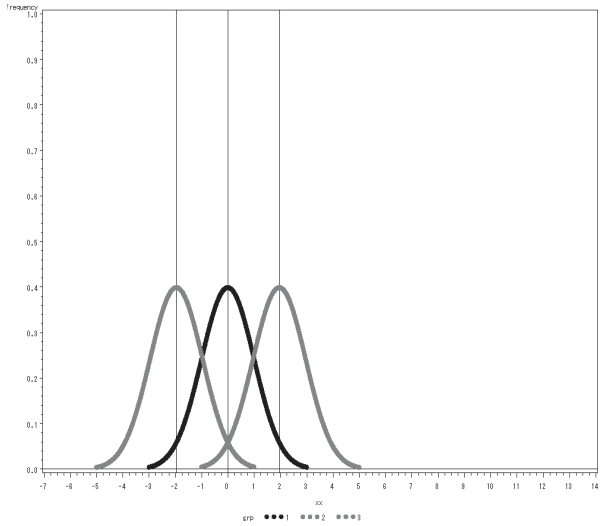
4．9　「信頼区間」vs.「2 つ」の「片側検定」との関係とは？

2 つの片側検定については，式とその操作は，**Table 4** の第 6 項に提示したが，考え方（ここでの解説は考え方についてであり，本物については，次の節で解説する）は，この **Fig. 6-2** と **Fig. 7** との 2 つを照合しながら解説する．**Fig. 7** において，**右端の分布の peak を上側限界 $\theta > 0$，左端の分布の peak を上側限界 $-\theta < 0$** とする．θ は，換言すれば「**許容限界**」であることに注意したい．また，上側と下側で，θ が異なっていても一向に構わない．この模式図では，$\theta = 1.96$ としたが，これまた何でも構わない．また，まず，別の視点からの信頼区間の意味と対応付けた独自の解説を試みる．

右端分布では，それから左側に離れる程，右端分布に属する確率は↓する．例えば，横軸「0」以下の値の出現は，片側確率 5% などとなる．当該値，つまり，横軸「0」以下の値が出現した時，「上側限界 $\theta >= 0$ の分布由来との仮説」を断念して，差 δ（原著では，（$\mu_T - \mu_R$）$< \theta$ の分布由来とする．信頼区間に関しての解釈として一般的な解釈*（次の Q07 で解説）とは別のもう 1 つの解釈は，このような視点からのものである．つまり，「0」から CL ＞ 0 離れた位置にある分布に属する集団の実現値が，0 ＞＝となる確率は，5%（片側）である．換言すれば，それが片側 95% CL の意味であるとの解釈である．

この「0」の点を，**Fig. 6-2** と対応付けることにするが，まず，検定と等価である信頼区間を用いて説明する．その図の縦軸で「0」の位置の横棒が対

Fig. 7　2 つの片側検定の考え方



— Fig. 6-2 と照合すること
右端の分布の peak を上側限界 $\theta > 0$，左端の分布の peak を上側限界 $-\theta < 0$　とする．

応しており，これより下の値の信頼区間**上限**（実現値）が発生した場合には，当該帰無仮説（**通常の帰無仮説とは，異質であるので注意！**）が棄却される．図では，左端の5つ（図が明快でないが，0線上の1個も入れて5つ！）の試行が，0＞＝であり，棄却される．この時点で，それ以外は棄却されず，「不合格＝非同等」となる．

同様に，**Fig. 7**で，今度は，逆端である**左端分布**で考察する．**Fig. 6-2**で，信頼区間**下限**0＝＜を眺めると，既述した通り，右端の4個は満足しているので，「下側限界－θ＝＜0の分布由来との仮説」が棄却される．それ以外は棄却されず，「不合格＝非同等」となる．

結局，**2つの片側検定的発想においては，これら2つの仮説とも棄却された，つまり，ANDとして棄却されたものだけを，「合格＝同等」**とする．これが，**「同等性検証」用，2つの片側検定の仕掛け**ということになる．

この2つの検定結果から，両者を満足する，つまり，棄却するcaseとしては，**AND論理**であるので，0例となる．**注意すべき点は，この発想法と通常の片側検定のそれとの異同**である．通常の検定であれば，上側が左の5個で「有意」（負け），下側が右の4個で「有意」（勝ち）となり，これは，それぞれ，信頼区間法での上限「0」以下，下限「0」以上と等価になる．つまり，「有意（優・劣）」に関して，**OR論理**でOKとするならば，上下，計9個が有意となる．

以上の解説では，許容限界値の例として1.734を示したものの，simulationとしては，帰無仮説下の母集団からの抽出をした標本であり，1.734に連結するdataとはなっていない．

結局，以上では，「2つの片側検定」を信頼区間に関する「別」の解釈から，その等価性を誘導したことになる．つまり，

とりあえず，許容限界値Δを無視して，あるいは，許容限界値をΔ＝±0として，考察すると，

①右端分布由来を棄却することは，2つの片側検定の内の1つである，上側の片側検定を棄却することであり，これは，信頼区間下限が＜0となることと等価であること

②同様に，左端分布由来を棄却することは，2つの片側検定の内の1つである，下側の片側検定を棄却することであり，これは，信頼区間上限が＞0となることと等価であること

③上記において，双方とも棄却されることは，信頼区間で考察すると，信頼区間が許容限界値±Δ内に含まれていることと等価となる．

この2つの片側検定的方式では，**同等許容範囲域（規格）外特性製剤を「同等」としてしまう過誤（消費者risk）を，片側α＝5%の時には，上下各5%以下に制御**している．（次の節で解説する本物では，θと表記される許容限界値Δを－したり＋したりして検定するために，Δ＝0とするここでの例とここでの結論と異なることに注意！）

ここまで説明して，読者は，先ほどの**第4．8節**で信頼区間法で同等とされた個数とここでの個数が違うのをどうしてくれるのか!? と言われそうである．ここでは，全て，**通常の**帰無仮説下での信頼区間や検定なのである．この信頼区間に対応するような仮説検定での「仮説」は，既述した如く，変更する必要がある．

Q07：**上記「＊」の一般的な解釈とは，どんなものか？**

A07：**当該1－α信頼区間を繰返し作成するとき，真値をそれら中に含む頻度は，1－αである．例えば，100－α（片側）%信頼区間であれば，α＝5%とするならば，作成した100個の信頼区間の内，平均的に90個（片側の場合！　両側では95個）で真値を含む．**

Q08：**Fig. 6-2では，2つの片側検定的発想において，2つとも棄却された，つまり，ANDとして棄却されたものとはどれか？　考察すること**

A08：省略．各自考察すること

4．10　「信頼区間」vs.「2つ」の「片側検定」との関係とは？　―結論

Fig. 6-2において，t_cal値（☆）での説明をしなかった理由を以下に述べる（以下では，**Fig. 6-2**ではなくて，**Fig. 6-3**を参照）．

Fig. 7の右端分布と検定とを関連付けるには，その時の帰無仮説，「上側許容限界θ＞＝0の分布由来との仮説」は，

H_0（**通常の帰無仮説とは異なる！**）：

$\theta = < \mu_T - \mu_R$　　**式（9）**

となる．この仮説に対応するt検定は，**式（1）**のt統計量に変えて，その分子を，

$$t = \frac{\theta - (m_T - m_R)}{\sqrt{\frac{\sum(x_i - m_T)^2 + \sum(y_i - m_R)^2}{2(n-1)} \times (2/n)}}$$　　**式（10）**

とする．これを例えば，10例／群，df = 18 ⇒棄却域1.734で検定すると，$(m_T - m_R) = 0$ の時でも，$\theta \neq 0$ であればpeakが0でないt分布となる．そのために，例えば，$\theta = 1.734$ の場合には，θ ／SE_ δ_μ のところにpeak（期待値）を有するt分布を形成する．ここで，SE_ δ_μ（δ_μ つまり，d_m の期待値のSE，との意味），図の例では，SE_ $\delta_\mu = \sigma\sqrt{2}/\sqrt{n} \Rightarrow \sqrt{2}/\sqrt{10} = 1/\sqrt{5}$，従って，**式（10）**の期待値は，$t = \theta/(1/\sqrt{5}) = \theta \times \sqrt{5}$，$\theta = 1.734 \times \sqrt{5} = 3.88$ である．**Fig. 6-3** では，縦軸3.88に位置する最上端の横線がこれである．

ここで，注目すべき点は，この場合の上側棄却限界の線（1.734）とこれらt_cal値達，そして先ほどの信頼区間との関係である．それは，以下の関係になる．

①**t_cal値に関して，上側棄却限界の線（1.734）「以上」が，2つの片側検定の内の上側検定では，仮説 H_0: $\theta = < \mu_T - \mu_R$ が棄却され，「有意」＝「同等」となる．図の例では，左端の2個以外は，「同等」となる．**

②このt検定結果と，信頼区間**下限**との関係である．それらは一致している．つまり，検定：有意＝信頼区間**下限**が，**下側棄却限界の線（－1.734）の上に収まるとの関係が成立**していること

③**片側5%であるので，通常の検定では，「0」の上下に逸脱する例についての確率は，10%となるが，（ややこしいが，この「0」であるとの帰無仮説が棄却された場合でも），2つの片側検定では，「± t_df = ± 1.734」を「許容限界」とした例に限定して言えば，その半数で同等性検証が保持されるため，それを逸脱する確率は，概略，半減して5%**となること

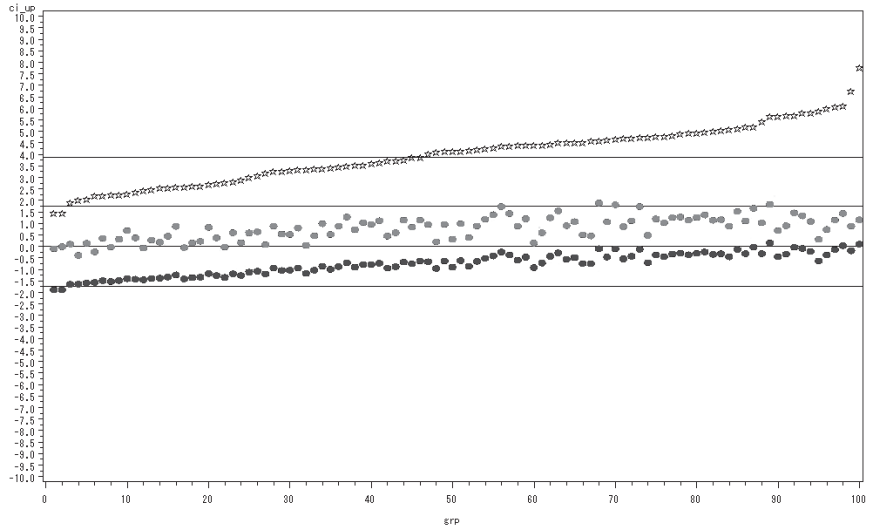
である．

もう1つの，2つの片側検定，つまり，**式（10）**の分子 $\theta - (m_T - m_R)$ を，今度は，$-(\theta - (m_T - m_R)) = (m_T - m_R) - \theta$ としたものでも，同様なことになる．

Fig. 6-2 では，既述した通り，片側95%（両側90%）CIとしている．t検定は，普通の**式（1）**でのものを，帰無仮説下で片側 $\alpha = 5\%$ で実施している．これら両者の結果は等価となる．事実，**文献[11]**によると，2つの片側 α での検定は，**信頼区間では，$1 - \alpha$ ではなくて $1 - 2\alpha$** と等価であるとの記載がある．以下に結論を整理しておこう．

①2つの片側検定での判定の有意水準と信頼区間法

Fig. 6-3　両側90% CI vs. 2つの片側t_cal計算値との3者の関係



CI：両側90%（片側95%）：信頼限界値 CL = t_df × SE_d_m = 0.776（df = 18 ⇒ t_df = 1.734）．縦軸での上下の「—」で示す範囲は，± 1.734，3.88（意味は本文参照）も付与．但し，t_cal値は，2つの片側検定の内の片方しか提示していないことに注意！
具体的には，
—実現値の「差」の両側90%信頼区間下限が－1.734を下回らない
＝t_cal計算値が1.734を上回る
—○：meanを除去

での「区間内」判定法での信頼区間（信頼係数）との関係は，

有意水準をαとすると，対応する信頼係数は，$1-2\alpha$である．

②両側90％信頼区間（CI）であったとして，**信頼区間が「0」を含まないとの基準，つまり，通常の検定と等価である基準を**，同等性の検証用としての，**±許容限界範囲に，その信頼区間が収まっているとの基準に変換することで，AND論理での2つの片側検定による検証法と等価になる．**

③±θを付加した状態で検定するという，特有な，2つの片側検定での判定についても，5％片側は，最終的には当該両方の検定でそれから逸脱する，つまり，有意とならない確率は，両者で，10％でなくて，**もっと低下**する．上下双方での逸脱は，**概略，5％に制御される．ただし，これは，あくまでも，許容限界±1.734とした場合に限定した特殊例での結論であることも強調しておこう．一般論としては，設定する許容限界値に依存することになる．**

4．11　両薬剤の「差」vs.「比」，双方存在の理由とは，その関係とは？

この分野では，対象とする現象が対数正規分布することまでも察知されていることが，「差」だけでなくて，「比」が登場する理由だと考えて良かろう．つまり，対数正規分布を対数変換した後の世界で「差」を取れば，それは，「比」を対数化したものとなる（**Table 4：第9，10項**）．

対数化した世界では，結果を読むのに，直感的な理解に苦労するために，再度，対数化前の尺度に戻すことになる．これに付随して発生するのが，「差」での「対称信頼区間」vs.「比」での「非対称信頼区間」と理解すれば良かろう（**Table 4：第9，10項**）．

4．12　同等性の検証 vs. 必要例数

世間でよく見かける誤解など（**Table 3, Table 4, 第11, 12項**），guidelineでの指針（**Table 4, 第13項**）については，それらを，また，例数算出・設計式については，別途，参考文献[12]などを参照されたい．ここでは，**同等性検証での例数決定因子達とそれらの挙動が，優越性検証・非劣性検証とはどこが異なるか**を注視しておきたい．

①両製剤の「差δ」

―通常の検証と異なり，「小」程，必要例数↓

―参考文献[12]などでは，「比」となっていることにも注意！

つまり，差＝0⇔比＝1.0

②同等性許容限界

―「大」程，必要例数↓

―①と②との弁別できる感覚育成を！

③バラツキσ

―「小」程，必要例数↓

―参考文献[12]などでは，「CV」となっていることにも注意！

である．

Q09：参考文献[12]などで，例数決定因子とその挙動と必要例数との関係を考察すること

A09：原論文のTable 1では，差がない，比＝1.0で一番例数「小」となっていること，差があれば，比が0.8や1.20となると例数が↑すること，CVも，「大」程，例数↑，原論文の式（1）～式（3）でも同様

4．13　有意水準と許容限界との識別・理解の重要性とその考察

既述した通り，同等性の検証法としての，2つの片側α検定では，下側許容限界値を下回る（逸脱する）という帰無仮説を棄却して，下回らない（逸脱しない）との対立仮説を採択し，それと同時に，上側許容限界値を上回る（逸脱する）という帰無仮説を棄却して，上回らない（逸脱しない）との対立仮説を採択し，この2つが同時に成立した場合にのみ，同等性が検証されたとの手順を採用している．つまり，ANDやORの論理で言えば，ANDの論理ということになる．信頼区間法でも，当然ながら，これと等価な検証がなされる．

4．13．1　有意水準α vs. 優越性の検証・同等性の検証

まず，議論すべきは，一般論としての有意水準α（信頼区間法では，信頼係数$1-\alpha$）の意味につ

いてである．通常の検定，つまり，優越性の検証において，周知の通り，αは，通常は両側5%（0.05）と設定されている．その意味を復習すると，「差無し」のものを，誤って「差有り」とすることが，検定を仮想的に反復するとき，100回中5回の割り合いで発生することは許容しようというのがα＝5%（0.05）の意味である．

αを，同等性の検証で考察すると，2つの片側α検定というANDの論理により，下側許容限界値と上側許容限界値とが形成する範囲を想定した時，その範囲外のものを，誤って範囲内としてしまうことが，100回中5回の割り合いで発生することは許容しようということである．***注意すべきは，ANDの論理の採用により，「片側」から事実上，一般のものと同様に，「両側」検定となっている点***である．

αは，通常の検定では，5%で認知されているが，同等性の検定では，5%ではなくて，増減すべきとの異論を持つ者がいるであろうか？ 例えば，5%でなくて，厳しくして1%とすべきだという類の異論である．これについては，筆者が推測するに，95%（？）の者が異論はないであろう．通常の検定では，α過誤は，「無効」な（場合によっては，毒にはなるが，薬にはならない）物が，「有効」として認知（認可）され，世間に出回るという種類のものである．同様に，同等性の検定では，α過誤は，「規格外の物が，「規格内」として認知（認可）され，世間に出回るという種類のものである．***αの意味は，双方とも無用な物を世に出してしまう／認知してしまうという方向では，同じ***である．異論を持つ者は，両者において一律5%ではなくて，前者が5%で良いこと，後者で変更すべきこととの両方の根拠を示すべきであろう．「世間では，前者は5%でやられているので理由はない」などというのは，根拠にはならない．

4．13．2　NS同等論 vs. 指針などでの同等性の検証

今回の指針に象徴される，第2世代（？）の同等性の検証は，従来のNS同等論からすれば，隔世の感がある程に工夫をされた，優れたものであることも既に解説した通りである．比喩的に言えば，NS同等論はザル法であり，極論すれば悪法である．努力しない程，報われるという側面・危険性を有するからである．第2世代の同等性の検証法では，これに対しても規制ができることも既述した．両者を同一次元で議論すべきではない．

4．13．3　同等性の検証での下側／上側許容限界値についての考察

前述の有意水準αに関する議論とこれら許容限界値との議論とが，現場で混同されていることを筆者は危惧する．極めて混同し易いからである．現象（data）が正規分布をする場合，下側許容限界は，100%－20%，上側許容限界は，100%＋20%とされ，対数正規分布するときには，これと同様な趣旨で，下側許容限界は，100%－20%，上側許容限界は，100%＋25%とされ，これは世界標準とされている．前者で議論すれば，この±20%という値に異論があるという議論であれば，それは一理あり，議論すべきである．

その時に注意すべきは，***この±20%の是非・妥当性は，統計学的観点から誘導されるものだと世間的にしばしば誤解されている***ことである．そうではなくて，現場が，筆者流に言わせて頂くなら，感性と直観・経験を踏まえて，妥当な根拠を考慮すべきものである．統計学的推論，とりわけ，「検定」の盲点が看破できず，それに盲従（？）する習慣が身についてしまった者には，そこからの脱却は容易ではないことは理解できる．しかし，感性と直観・経験を踏まえて，妥当な根拠を現場で体得している貴重な例も皆無ではない．周知の例としては，臨床検査値が「範囲外」にあっても，医師は，「この程度の外れなら問題ない！ 様子を見ましょう！」というような直観・洞察や，「肺動脈せつ入圧が計測できない場合には，右房圧で代用できる」といった直観である．後者に関しては，筆者はかつて，その相関係数rを計算したことがある．rが0.8程度であったと記憶している．***これは，生命科学・生命現象での，「意味のある」相関度***ということになる．ここでも注意したいのは，***「統計学的な有意」と混同しないこと***である．例数を増やせば，r＝0.01であっても，「統計学的」には「有意」となるのである．また，cyto-toxicな制癌剤の早期第Ⅱ相試験において，hurdleが有効率20%というのも，こうした例