

ISSN 0288-2906

医薬安全性研究会 会報

Bulletin of Japanese Society for Biopharmaceutical Statistics.

July. 1999

〈特集〉 最大対比法

No.44



サイエンティスト社

1999～2000年 医薬安全性研究会
関連スケジュール

- ☆第80回定例会20周年記念大会（総評会館） 1999年10月30日（土）
☆第81回定例会（総評会館） 2000年 1月29日（土）
☆第82回定例会（総評会館） 2000年 4月 1日（土）

〔データ解析講習会〕（数学工房 共催）

- ☆応用数理講習会 -統計モデルへの線型代数の応用-
1999年 8月 6日（金）～ 7日（土）（一泊合宿）

☆線型代数入門

〔第Ⅱ期〕線型写像・線型変換・線型方程式

1999年 9月～12月（全6回+合宿）

〔第Ⅲ期〕計量ベクトル空間、行列式、線型方程式への応用

2000年 1月～ 4月（全6回）

☆確率・統計入門

〔第Ⅱ期〕確率分布と統計量

1999年 9月～12月（全6回）

〔第Ⅲ期〕確率モデル

2000年 1月～ 3月（全6回）

〔新刊予告〕

- 『医学研究における実用統計学』（予約募集中）
『直接閲覧に必要なカルテの見方』（予約募集中）
『システムティック・レビュー』（予約募集中）
『医療経済入門』（予約募集中）
『毒性質問箱No.2』（予約募集中）

「臨床試験のための統計的原則」
をめぐる諸問題
成立の経緯から Q & A まで

1999年4月3日

吉村 功
東京理科大学

はじめのお断り

今日の話は個人的見解

すべては吉村の事実認識と個性の発現
違う見解の統計家も沢山いる

The opinion, findings, and views
expressed in this talk are those of the
speaker and should not be attributed to
@@@.

2

取り上げる話題

1. はじめに
2. 経緯
3. 多施設共同治験
4. 複数の検証的試験
5. 施設と薬剤の交互作用

3

取り上げる話題(2)

6. 非劣性試験
7. 解析対象集団: FA と PP
8. 試験統計家のこと
9. その他の質問
10. おわりに

4

1. はじめに: 全文の入手

◆ ICHの産物

「臨床試験のための統計的原則」が

1998年11月30日

厚生省医薬安全局審査管理課長から
各都道府県衛生主管局長あてに通知

(<http://www.nihs.go.jp/dig/ich/efficacy/e9/>
よりダウンロード可) > 何度も読んでほしい

5

はじめに: 主要な変化

◆ 何が変わる?

試験計画の質の吟味

試験計画の遵守の要求

◆ 過去の試験では? たとえば

選択基準違反も救済?

複数評価基準のどれかで承認?

後付解析が検証?

6

はじめに: 最近の動き

- ◆ 通知に至るまでの議論
厚生省検討班での議論、製薬協の注文
- ☞ 上田班の報告会(1月24日)
- ◆ 吉村のFDA 訪問(3月24日)
- ☞ 医学会総会のパネル討論(4月2日)
- ☞ のマークの抄録は、<http://www.ms.kagu.sut.ac.jp/indexj.html> で入手可

7

2. 成立の経緯

- ◆ ICHトピックの中では最も短期で完成?
 - ◆ 重要性を明示したいという機運
 - ◆ 旧ガイドライン、現臨床試験の質への不満
- ◆ ある程度完成した下敷きの存在
日本、アメリカのガイドライン vs
CPMP のガイドライン
- ◆ ICH 自体の収束傾向

8

成立の経緯: EWGメンバー

- ◆ どちらかという数理工系の素養が大
 - ◆ 議論がかなり論理的
 - ◆ 議論の進め方では日本への配慮があった
 - ◆ でもやっぱりアメリカはアメリカ
- ◆ 内容では押されっぱなし(個人的印象)
 - ◆ 医薬開発、吟味での日本の実力不足
 - ◆ 臨床試験の質は日本の方が劣る
 - ◆ なんといっても言語でのハンディキャップ

9

成立の経緯: 詰めでの時間

- ◆ 山を越えてから1年半
 - ◆ ブリュッセルの会議(4ICH) が1997.6.
 - ◆ step 4 が1998.2.
 - ◆ 通知が1998.11.30.
- ◆ 影響
 - ◆ 経過措置の遅れ
 - ◆ 啓蒙の機会多く、心の準備は十分?

10

3. 多施設治験の被験者数: 記述

意図している患者集団への一般化の根拠が理解でき、説明できることは重要である。
このことによっても、必要な施設(センター)の数と型式(例えば、専門医か一般臨床医か)及び試験の数と型式が影響を受けるであろう。
検証的試験の結果は、条件によらず安定しているべきである。
場合によっては、一つの検証的試験からの証拠だけで十分であることもありうる。

11

この記述の趣旨: 外部妥当性

外部妥当性、一般化可能性、internal validity を確保するという目標を目指して、可能な技術を駆使せよ。
そうすれば自動的に、センターの数と型式及び試験の数と型式の決め方が定まるだろう。
場合によっては、一つの検証的試験からの証拠だけで十分であることもありうる。
たとえば、薬理学的知見、他の臨床試験、規模の大きさが外部妥当性を保証できればよい。

12

多施設治験の被験者数: Q&A

質問: 一般化可能性は、施設あたりの症例数と関連しており、「施設あたりの被験者数が多い数カ所の施設において行う場合があり、」

「一群10例以上が一つの目安である。」の文章及びIRBを設置できている施設を勧奨すると、専門医での多数例と、一般化可能性の、どちらを重要視している?

「施設あたりの被験者数に大きな差が生じることのないよう配慮する必要がある」では、どのくらいの被験者数の違いは許容される?

13

多施設治験の被験者数: Q&A

質問: 「施設あたりの被験者数に大きな差が生じることのないよう配慮する必要がある」では、どのくらいの被験者数の違いは許容可能?

吉村の見解: 計画段階で感度分析を試みる。

重み付き平均の安定性は?

施設間差がどの程度検出できる?

そのために、経験と知識のある試験統計家の参加を義務づけている。

自分でそれに応えられない統計家は失格

14

多施設治験の被験者数: Q&A

1施設当たりの症例数は1群10例以上となっているが、第1相、第2相(初期)も同じと考えて良いか。

A: どう応えればよいか皆で考えよう

第1相試験の目的は? 目的にふさわしい被験者の選び方は?

前期第2相試験の目的は? 目的にふさわしい被験者の選び方は?

15

多施設治験の被験者数: Q&A

「1群10例」は何のため?

10という数は本質的でない。交互作用があったとき、その内容を吟味するには、その程度の反復数が必要だろうという数値

科学的に有効性が検証された後での、

交互作用が出そうもない試験治療なら、少数多施設でも一般化可能性は容認可

16

多施設治験の被験者数: Q&A

Q: 3剤比較の症例設定は、具体的にはどのような計算式を用いればよいか。

参考に1例を示してほしい。

疑問: 3剤比較を何の目的に行いたい?

17

多施設治験の施設の定義: Q

多施設共同治験の枠組みにおける施設という用語の定義は(例えば、治験責任医師か、場所か、地域か)

施設間差の原因には、医師側(治療方針)と患者側(地域など)が考えられる。

問診法などを除けば、主要変数は客観測定、集中測定で偏りを減らせるし、併用薬、投与量などの医師側の要因の影響も減らせる。

地域を主にして施設を定義するのが、実際的と考えられるがどうか?

18

多施設試験での施設の定義:A

吉村の見解:

影響する要因は?

医師側: 治療技術、遵守度、評価基準

患者側: 重症度、未知の予後因子

環境: 食習慣、気候、職業、生活水

全要因において内的均質性が主張できること

それらを確認し、疑問が出されたときデータを持って相手を納得させられること

19

4. 複数試験のこと: Q & A

Q: 3剤比較を行わない場合は、被験薬と実薬対照、及び、被験薬とプラセボとの両方の比較が必要になるのか?

Q: 実薬対照とプラセボはどうなるのか?

疑問: 何のために試験をしようとしている?

20

検定はまず交互作用のないモデル

↑3.2 多施設共同試験:

◆試験治療の主効果は、最初に施設と試験治療の交互作用を含まず施設間差を考慮に入れるモデルを用いて調べられるであろう。(may)

◆試験治療効果の肯定的な効果が判明した場合、結論の一般化可能性に影響する可能性があるため、通常は施設間における試験治療効果の不均一性を探索すべきである。

21

検定はまず交互作用のないモデルで(2)

↑異論・異議あり:

「通常勧められている、先に交互作用の解析を行うアプローチを禁止されるのは不合理である。通常の方法で正しく交互作用解析がなされるなら、それも正しいアプローチであることを認めよ」

◎ガイドラインの記述は非通常?

◎正しいものを正しいとわざわざ認めるの?

22

検定はまず交互作用のないモデルで(3)

↑ガイドラインでの方針の理由:

◆交互作用が大きいときにそれを無視して適当に定めた主効果なるもので承認することは避けたい。

◆交互作用が無いときはこの方針が良い。

◆交互作用があるとこの方式では検出力が落ちる=>それでよい。

◆交互作用についての予備検定方式は有意水準を制御できない。

23

非劣性試験の有意水準は片側 2.5%

↑5.5 推定、信頼区間及び仮説検定:

◆承認申請のための試験では、片側の第1種の過誤を両側検定で慣例的に用いられている値の半分に設定する方法が好ましい。

◆このように設定することにより、試験治療間の差の大きさを推定するために通常用いられている両側信頼区間との整合性を図ることができる。

24

非劣性試験の有意水準は片側 2.5% (2)

◆従来の日本の慣例:

- ◆優越性試験では両側 5%
- ◆非劣性試験(同等性検証と通称)では片側5%

◆これからは両方を片側 2.5%, 信頼水準 95% の両側信頼区間方式

25

非劣性試験の有意水準は片側 2.5% (3)

◆なぜ変更?

- ◆日本が独自の方針をだしてよいが欧米はこの方式で行く、と明言
- ◆日本の非劣性試験だけが別ルール?
☞日本の試験データが外国で使えない!
- ◆外国を日本方式で説得できるか => No!
- ◆日本の甘さが目立つだけ

26

非劣性試験の有意水準は片側 2.5% (4)

◆大きな変更の統計学的な根拠は?

- ◆反問、今までの方式の統計学的な根拠は?
- ◆従来からの批判=有効率が 10% も劣って良いというのは甘過ぎる。
- ◆薬剤の承認を求めるときには、従来より被験者数を約 40% 増しとなる。
- ◆それが欧米と同じコストとなる。

27

7. 解析対象集団は FA? PP?

Full analysis set (FA、最大の解析対象集団):

ランダム化が行われた被験者から偏りを導入することなく除外できる被験者のみを最小限に除外したもの

Per protocol set (PP、治験実施計画書に適合した対象集団):

割付を明らかにする前にすべて明確にし、文書に残しておく。

28

7. 解析対象集団は FA? PP?

FA と PP の特徴の比較:

FAでは被験者数(sample size)が大きくなる

→見かけ上精度が良いという形

→差の信頼区間が狭くなる→同等の見せかけ

PPでは、現実に起こる脱落や違反が除外されることで、薬効の過大評価が起こりやすい。

どちらにしても結果が変わらないことが重要

29

7. 解析対象集団は FA? PP?

「同等性試験又は非劣性試験では、最大の解析対象集団を使用することは一般に保守的ではないため、その役割は十分慎重に考慮すべきである。」

質問: 非劣性の試験においては、主となる解析対象集団は治験実施計画書に適合した対象集団と宣言しても問題ない?

それとも、主要な集団を決めるのに何か考慮すべき点がある?

30

7. 解析対象集団は FA? PP?

吉村の見解: 検証という意味で、主たる解析対象集団をPPと宣言するのは一つの方針

しかし、両方の解析を行って、違いが生じた場合には、それぞれ特徴に基づいて原因を追及することも必要。

それなしに、機械的にどちらかを定めておき、他方の結果を無視するのは適当でない。

外部妥当性の吟味という精神を忘れない!

31

7. 解析対象集団は FA? PP?

「選択基準と除外基準の変更」

質問: 基準を途中で変更した場合、それまで集めた集団と後からの集団とで、何かの違いがあると思われる。

どのような面で整合性があれば、変更した集団での結果を検証的と主張できる?

32

7. 解析対象集団は FA? PP?

吉村の見解: 盲検下の独立なモニタリング委員会による検討で、忍容性を考慮して年齢範囲を狭めたり、集積率を考慮して症状を限定することはありそう。

科学的知見が増えある要因を持った患者をじよがいすることもありそう。

一般に、対象集団を狭めるときの方がは問題が少なく、広げるの方が危険?

基準変更の理由が最も重要

33

7. 解析対象集団は FA? PP?

被験薬T、実薬対照A、プラセボP、の3群比較で二つの仮説

H1: T vs A は非劣性

H2: T vs P は優越性

を検証しようとする場合、解析対象としては、H1 に対してはPP、H2 に対してはFAを用いることが推奨されている。総括報告書で結果をまとめるとき不都合が生じそうである。

できればどちらかに揃えたい。どちらを主とすればよいか。

34

7. 解析対象集団は FA? PP?

疑問: 目的が違うのではないかな?

総括報告書で結果をまとめるときどんな不都合が生じる?

できればどちらかに揃えたいというのは何故?

試験の目的は何なのだろう?

実薬の有効性に疑問がある薬剤のFDA への申請なら、プラセボ対照に注目すればよい。事前相談で解決!

35

8. 試験統計家: Q & A

試験統計家は、どのような治験にも加える必要があるか? 例えば、第1相治験は?

吉村の見解: これは自分で考えること。

こういう質問が出る会社にいる試験統計家は、よほど役に立たない人なのでしょう。

強制されたから名前だけ入れるという認識で、まともな試験ができるとは思えない。

36

他の質問: 盲検下レビュー

「本ガイドラインでは、試験完了(最後の被験者の最終観察)から割付を明らかにするまでの間にデータをチェックすることを、データの盲検下レビューと呼ぶ。」

「この形式のモニタリングの期間は、通常試験実施施設が選択されるときに始まり、最後の被験者のデータが収集されクリーニングされたときに終了する。」

試験完了前あるいは途中における「ブラインドのレビュー」は、モニタリング?

37

他の質問: 盲検下レビュー

説明: blind review というのは、このガイドライン議論でも、どちらかという終わり頃に、特別な意味で導入された概念。

key open 前にモニタリングを行って、いろいろな計画変更を行うのはblind review でない。

→「マスク下でのモニタリングでの検討」

(私個人はこういう誤解が生じやすいので blind review という用語をこのように特殊化するのに気が進まなかった。他に適切な用語を英語で対置できず受け入れた。)

38

他の質問: 盲検下レビュー

「盲検下レビューから、治験実施計画書に述べた解析の主要な特徴となる事項の変更を提案しようとする場合、変更の内容は治験実施計画書の改定に記録すべきである。」

説明: この文章で重要なことは、記録するということ。

盲検下レビューというのは、データがこてされた最後の段階での最終吟味

39

他の質問: 盲検下レビュー

盲検下レビューの段階では、更に集積する場合を除いて、主解析を含め解析の項を変更することが多いと考えられるが、解析計画書を変更するだけではないけない?

各施設のプロトコルをその時点で変更するのは、手続き上だけで、本当に意味がある?

それとも治験調整医師などでのプロトコルの変更でかまわない?

疑問: この質問の意味が理解できない。

40

他の質問: 過去の試験条件

「実対照薬は、... 新たに行う試験に関連した、医学又は統計学の進歩を考慮した上で、新たに行う治験計画上の重要な特徴(主要変数、実対照薬の用量、適格基準など)を、実対照薬が臨床的に適切な有効性を明確に示した過去の優越性試験と同じにすべきである。」

質問: 過去の海外臨床試験などで優越性が検討されているとした場合、日本での非劣性試験においても優越性試験と同じような集団を指して適格基準などを同じにすべき?

41

他の質問: 過去の試験条件

説明: 実薬対照での試験で非劣性を検証しただけでは、プラセボに対する有効性を検証したことになる、という主張から来た記述。

実薬対照が確かにプラセボより勝っていることが検証された条件、以外の条件下で実薬対照との非劣性を検証しても、有効性の証明にならないということ。

現実の適用は微妙。どちらかというと事例ごとにFDAに解釈を求めべきこと。

42

他の質問: 誰に対してマスク?

「中央治験事務局の職員に、試験治療コードが知られることがないように、適切なコンピュータログリズムを用いることが望ましい。」

ここで言う「中央治験事務局の職員」には、CROでの電話登録を行うときのCROの担当者が含まれる? =>もちろん

試験依頼会社とは独立していて、電話で受け答えをする人に試験治療コードを知られないようにする必要はある? =>もちろん

43

誰に対してマスク?

★ どんな拷問を受けても知らないことは白状できないというのが理想のブランド

★ 自分を含めたすべての人にマスクするのが当然

(入学試験ではどの採点者も受験者名を知らない! 出題者さえ!)

44

9. その他の質問:

Q:「有害事象のパターンが示されるようなグラフ表示」の具体例がある?

質問に対する疑問:

自然なグラフ表示以外に何か特別なことを求めるような状況が、臨床試験にあるだろうか?

ちよつと思ひ当たらないのだが?

45

10. おわりに

↑ 3月26日の FDA の発表

☞ 問題になっている糖尿病薬の使い方についての制約を強める

↑ それにたいするコメント

☞ FDAは承認を急ぎすぎている?

↑ 日本は?

☞ 遅い理由, 急ぐ理由, 慎重さの維持

46

おわりに(2)

- ★ 目指すところは臨床試験の質の向上
- それを通して, 有効性と安全性の確実化
- 被験者数は最小に=>データの有効利用
- 質の良い試験統計家の関与は有用
- 質の良い統計家の育成が急務

精神をつかんで欲しい

小手先の「受験対策」は本筋にあらず!

47

<特集>最大対比法

用量反応関係の評価するときの 「対比」の役割

医薬安全性研究会 1997.7.5.

吉村 功
(東京理科大学工学部)

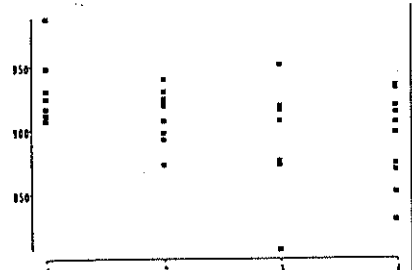
「対比」という用語

- ここで問題にしている「対比」という用語は英語の "contrast" という単語の訳語
- 世の中一般での contrast は？
 - (名詞) the contrast between light and shade
→ 明暗の対照 (研究社)
 - (動詞) to set in order to show difference
 - (名詞) opposition or unlikeness in things compared (Chambers)

「対比」という用語(2)

- 統計学の世界で定義されている用語 technical term としては？
- 例: 4群の実験で各群の測定変数(たとえば赤血球数 RBC を $n = 10$ 匹について測ったもの)の平均を Y_1, Y_2, Y_3, Y_4 とすると、次の三つはいずれも群平均についての対比である。
 - $> Y_1 - Y_2$
 - $> Y_1 + Y_2 - 2Y_3$
 - $> Y_1 - Y_2 - Y_3 + Y_4$

実験データのイメージ



対比の定義

- なぜこれを対比と言うか？
=> 対比の定義に合っているから
- では定義は？
- a 群の実験で各群の群平均 Y_1, Y_2, \dots, Y_a の重み付き和 $c_1 Y_1 + c_2 Y_2 + \dots + c_a Y_a$ において、重みの係数 c_1, c_2, \dots, c_a が $c_1 + c_2 + \dots + c_a = 0$ を満たすとき、上の重み付き和を (a 群の実験における平均についての) 対比という。

対比を考えるときの背景知識

- 各群の平均 Y_1, Y_2, \dots, Y_a が独立に
- $E\{Y_i\} = \mu_i, V\{Y_i\} = \sigma^2$ の分布に従う確率変数であるとする。
- 各群の用量を d_1, d_2, \dots, d_a とする。
- 知りたいのは、用量 d_i と反応の期待値 μ_i の関係である。

対比を考える背景知識(2)

- 期待値 $E(Y_1)$ は, Y_1 の分布の平均(母平均)である.
- 分散 $V(Y_1)$ は, Y_1 の分布の分散(母分散)である.
- 期待値という演算は線形性を持つ.
- $E(a_1 Y_1 + a_2 Y_2 + \dots + a_n Y_n)$
- $a_1 E(Y_1) + a_2 E(Y_2) + \dots + a_n E(Y_n)$
- 観測変数が独立のとき分散という演算は加法性を持つ.
- $V(a_1 Y_1 + a_2 Y_2 + \dots + a_n Y_n)$
- $a_1^2 V(Y_1) + a_2^2 V(Y_2) + \dots + a_n^2 V(Y_n)$

対比を考える背景知識(3)

- a 群の実験での対比の場合
- $E(c_1 Y_1 + c_2 Y_2 + \dots + c_n Y_n)$
- $c_1 E(Y_1) + c_2 E(Y_2) + \dots + c_n E(Y_n)$
- $-c_1 \mu_1 + c_2 \mu_2 + \dots + c_n \mu_n$
- $V(c_1 Y_1 + c_2 Y_2 + \dots + c_n Y_n)$
- $c_1^2 V(Y_1) + c_2^2 V(Y_2) + \dots + c_n^2 V(Y_n)$
- $c_1^2 \sigma^2 + c_2^2 \sigma^2 + \dots + c_n^2 \sigma^2$
- $(c_1^2 + c_2^2 + \dots + c_n^2) \sigma^2$

例にあげた対比では

- $E(Y_1 - Y_2) = E(Y_1) + (-1) E(Y_2) = \mu_1 - \mu_2$
- $V(Y_1 - Y_2) = (1^2 + (-1)^2) \sigma^2 = 2 \sigma^2$
- $E(Y_1 + Y_2 - 2 Y_3) = \mu_1 + \mu_2 - 2 \mu_3$
- $V(Y_1 + Y_2 - 2 Y_3) = (1^2 + 1^2 + (-2)^2) \sigma^2 = 6 \sigma^2$
- $E(Y_1 - Y_2 - Y_3 + Y_4) = \mu_1 - \mu_2 - \mu_3 + \mu_4$
- $V(Y_1 - Y_2 - Y_3 + Y_4)$
- $(1^2 + (-1)^2 + (-1)^2 + 1^2) \sigma^2 = 4 \sigma^2$

対比の役割は?

- もし $\mu_1 - \mu_2 = \mu_3 - \mu_4 = 2\mu$ だったら, 次の三つの対比の中で, 期待値の絶対値が最も大きくなるのはどれ?
 $> 1 Y_1 + 1 Y_2 + (-2) Y_3 \quad \Rightarrow > \mu + \mu - 4\mu = 2\mu$
 $> 1 Y_1 + (-2) Y_2 + 1 Y_3 \quad \Rightarrow > \mu - 2\mu + 2\mu = \mu$
 $> (-2) Y_1 + 1 Y_2 + 1 Y_3 \quad \Rightarrow > -2\mu + \mu + 2\mu = \mu$
- 注意!
- 係数が大きくなると絶対値も大きくなるから比較には標準化が必要.

対比統計量

- 対比は係数を標準化しないと比較が不公平.
- 標準化には分散の平方根を使うのが自然.
- $t(Y) = \frac{c_1 Y_1 + c_2 Y_2 + \dots + c_n Y_n}{\sqrt{(c_1^2 + c_2^2 + \dots + c_n^2) \sigma^2}}$
- \rightarrow 分散 σ^2 を推定量 V で置き換える.
- \rightarrow 対比統計量という概念

対比の表現法

- 対比は係数だけで決まるので, 係数を並べて示す.
 $> Y_1 + Y_2 - 2 Y_3 \rightarrow (1 \ 1 \ -2)$
 $> Y_1 - Y_3 \rightarrow (1 \ 0 \ -1)$: 真ん中の 0 を忘れるな!
- この表現法では, 変数の並べ方を明らかにしていないと, とんでもない誤解が生じる
- \rightarrow SAS プログラムでは特に注意!
- 対比についての計算は標準化して考えるのが暗黙の了解なので, 係数は整数にするのが好都合
- $(2 \ 2 \ -4) \Leftrightarrow (1 \ 1 \ -2) \Leftrightarrow (0.5 \ 0.5 \ -1.0)$

対比と用量反応関係

- 4群で考えよう。
- 直線的な用量反応関係
 - 一般には $\mu_i = \beta_0 + \beta_1 d_i$
 - $d_1 = 1, d_2 = 2, d_3 = 3, d_4 = 4, \beta_0 = 0$ とすると
 $\mu_1 = \beta_1, \mu_2 = 2\beta_1, \mu_3 = 3\beta_1, \mu_4 = 4\beta_1$
 - 基準化した対比で比べたとき、この用量反応関係に対して期待値の絶対値が最大となる対比は、係数 (c_1, c_2, c_3, c_4) が等間隔の場合:
 - 整数化して $\rightarrow (-3 \ -1 \ 1 \ 3)$

関連する数学的定理

- a 群の実験での対比の場合
 - ◆ $c_1 + c_2 + \dots + c_a = 0$
 - ◆ $c_1^2 + c_2^2 + \dots + c_a^2 = 1$
 - ◆ $\mu_1, \mu_2, \dots, \mu_a$ の値は既知
 - ◆ という条件の下で $k_1\mu_1 + c_2\mu_2 + \dots + c_a\mu_a$ を最大にするには、
 - ◆ $c_1 \propto \mu_1 - \mu$ (係数を平均からの偏差に比例させる。)

浜田論文に登場する対比

| 対照 | 低用量 | 中用量 | 高用量 | 対比の意味 (略号) |
|----|-----|-----|-----|--------------------|
| -3 | -1 | 1 | 3 | 直線的に増加 (linear) |
| -7 | -5 | -1 | 13 | 用量に対して直線的に増加 (reg) |
| -3 | 1 | 1 | 1 | 低用量で飽和 (l-down) |
| -5 | -1 | 3 | 3 | 中用量で飽和 (m-down) |
| -3 | -3 | 1 | 5 | 中用量から立ち上がり (m-up) |
| -1 | -1 | -1 | 3 | 高用量から立ち上がり (h-up) |

```
proc glm data=data outstat=out: class group;
model y=group/est;
contrast 'linear' group -3 -1 1 3;
contrast 'reg' group -7 -5 -1 13;
contrast 'l-down' group -3 1 1 1;
contrast 'm-down' group -5 -1 3 3;
contrast 'm-up' group -3 -3 1 5;
contrast 'h-up' group -1 -1 -1 3;
```

最大対比法

- 推定的発想
 - いくつかの対比統計量を用意しておき、その最大値で把握される用量反応関係を真の用量反応関係と判断するのが合理的であろう。
 - 検定的発想
 - いくつかの対比のどれでもあまり大きな値にはならないなら→目立った用量反応関係はないと判断するのが合理的であろう。
- > 最大対比法というアイデア

最大対比法はどんな場面で 使えるか？

1998. 7. 4.

東京理科大学 吉村 功

はじめに (1)

- 用量反応関係の評価は「医用統計」での最も重要な課題の一つ
- 有害作用の評価ではどの用量からがどの程度有害作用が生じるかが焦点の一つ
- 臨床用量の設定では、どこが最適化を求める基礎資料

はじめに (2)

- 問題は、たかだか4用量程度の実験では、用量反応関係を直接求めることができないこと
- 仮説検定で用量反応関係を推定、評価できるか？ => 無理
- 別の方法が必要 => 対比を利用してある程度の評価ができないものだろうか？

状況設定 (1)

- ★ 群の数が a の一元配置
- ➡ $d_1 = 0$ => 対照群, サンプルサイズ n_1
- ➡ d_2 => 低用量群, サンプルサイズ n_2
- ➡ ...
- ➡ d_a => 最高用量群, サンプルサイズ n_a

想定している状況 (2)

- Y_{ij} = 第 i 群, 第 j 固体の反応
- ★ Y は互いに独立に
- ★ $E\{Y_{ij}\} = \mu_i, V\{Y_{ij}\} = \sigma^2$

群の平均

$$\bar{Y}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} Y_{ij}$$

最大対比法という概念 (1)

- ★ 重み $w_j = (w_{j1}, w_{j2}, \dots, w_{ja}), \sum w_{jk} = 0$ の対比 T_j を次式で定義する。

$$T_j = \frac{w_{j1}\bar{Y}_{j1} + w_{j2}\bar{Y}_{j2} + \dots + w_{ja}\bar{Y}_{ja}}{\text{Estimator of SD}\{\text{numerator}\}}$$

最大対比法という概念 (2)

★ 標準的な分母は次の形

$$\sqrt{\left(\frac{w_{j1}^2}{n_1} + \frac{w_{j2}^2}{n_2} + \dots + \frac{w_{ja}^2}{n_a}\right)\sigma^2}$$

==> いわゆるスチューデント化

★ 重みを一組 (p 個) 決めると

==> 対比統計量が p 個決まる

最大対比法という概念 (3)

★ p 個の対比に対して,

★ $T = \max \{T_1, T_2, \dots, T_p\}$

を判断の基準に用いるのが最大対比法.

★ 仮説検定

★ モデル選択

★ 多重決定

最大対比法の例

★ 4 群の場合

● ダネット (1959)

$w_1 = (-1, 0, 0, 1)$

$w_2 = (-1, 0, 1, 0)$

$w_3 = (-1, 1, 0, 0)$

● ウィリアムズ

(1971)

$w_1 = (-1, 0, 0, 1)$

$w_2 = (-2, 0, 1, 1)$

$w_3 = (-3, 1, 1, 1)$

● Max-t 法 (1979)

$w_1 = (-3, 1, 1, 1)$

$w_2 = (-1, -1, -1, 1)$

$w_3 = (-1, -1, -1, 3)$

歴史的経緯 (1)

有名なのは Scheffe の方法 <== 1953

スチューデント化した対比法

<== (いわゆる max-t 法) 1979

順序制約のもとでの推論 <== 1988

最小有効量を定める複数対比法

<== 1989, Ruberg

歴史的経緯 (2)

第 II 相臨床試験のデータから第 III 臨床試験での臨床用量を求めるための方法

<== 1991?, BIOS での大橋の提案

無毒性量を判定する方法はだいたいこれに属する

<== 1991?, BIOS での吉村の指摘

歴史的経緯 (3)

具体的な仮説検定の手順の提案

= p-値計算法

<== 1992, 1994, 岸本と浜田の論文

拡張したウィリアムズ法としての具体化と拡張

<== 1997, 若菜, 吉村, 浜田

スチューデント化での問題点

- ◆ グループの大きさが異なるとき、標準偏差が大きく変化=>不合理の発生
- 用量反応関係の不平等な比較

$$T = \frac{\text{Max}_j \{ (w_{j1} \bar{Y}_{j1} + w_{j2} \bar{Y}_{j2} + \dots + w_{jk} \bar{Y}_{ja}) / \sqrt{\sum w_{jk}^2} \}}{\text{Estimator of SD(Numerator)}}$$

とする方がよいのではないかと

最大対比法の性能評価を検出力で行うのは不合理

- ◆たとえば、無毒性量を間違えることの損失は異なった用量を与える確率で評価するのではなく、どれくらいずれた値を推定したかで評価すべきではないか？

重みの組を選ぶ適切な規準、やり方があるか？

- ➡ Ruberg 氏の提案
検出力の最小値が最大になるような組を採用したらどうか
=> それにもとづいてデザインを考える

最大対比法は未だ十分には完成していない。

- ◆ いろいろな側面から長所・短所の研究が必要
- ◆ 毒性評価
- ◆ 臨床用量設定
- ◆ 最小有効量
- ◆ 用量反応関係の評価

最大対比法の性能評価

東京大学 薬剤疫学教室

浜田 知久馬

1. はじめに

医薬品を開発する場合、新薬が既存の薬剤と比べて全く同じような薬効しかないなら世に出す意味はない。したがって臨床試験においては、新薬が既存の治療法に比べて何らかの意味で優れていることを証明することが要求される。私は、これに類することが新しい統計手法を開発する際にも必要であると思う。既存の方法と同じような性能しかないのであれば、新たにその方法を提案しても、それは統計家が論文を増やすための手段でしかない。

統計手法の性能を評価する際には、推定においては偏りがない(不偏である)、MSE(mean square error)が小さい等が指標として用いられることが多く、検定においては検出力が代表的な指標である。これらの性能が既存の方法と同等であっても、計算が簡便である、仮定からのズレに頑健(robust)であるなどのメリットがあれば、その統計手法は世の中に送り出す価値がある。あるいは適応拡大のように、2群の場合を多群へ、パラメトリックの問題をノンパラメトリックに拡張して、しばしば新しい統計手法が生まれる。

さて、用量反応関係を評価することを目的とした医薬研究において、用量反応関係を複数の対比(contrast)によってモデル化し、最大の対比に基づいて用量反応パターンを評価する方法として最大対比法がある。最大対比法を医薬研究の中で応用しようとする動きは1990年頃から始まった。Roberg(1989)は minimum effective dose を求めるために最大対比法を適用した。またこれとは独立に、少し遅れて、大橋・浜田等は後期Ⅱ相臨床試験において、用量反応関係を明らかにするため、最大対比法を用い始めた。また吉村・浜田等は、一般毒性試験において、用量反応のパターンを判断する方法として最大対比法を提案した。最大対比法の考え方は、非常に単純ではあるが、p値の計算には高性能のコンピュータが必要となるため、実用化されたのは、このように比較的最近である。このためその性能については十分に検討されていない。そこで本稿では、毒性試験データに最大対比法を適用した場合の性能について評価する。具体的には次の2つの点を明らかにする。

1) プライマリーエンドポイント：検出力

用量相関性の標準的な解析手法である回帰分析と最大対比法の検出力を比較し、同等以上であることを示す。

2) セカンダリーエンドポイント：人の用量反応パターンの判断との一致度

統計解析による評価は客観的なものであるが、そこで得られた結果は、人間の直観とマッチしていることが望まれる。そこで最大対比法による用量反応パターンの判定の人による判断の一致度が高いことを示す。

2. 最大対比法(maximum contrast method)とは

一元配置分散分析型のデータ構造を想定し、 Y_{ij} を第*i*群の*j*番目の観測値を表すものとする。

このとき帰無仮説の下で分散が1になるように基準化した対比統計量(Z)は次のようになる。

$$Z = \sum C_i Y_{i.} / \text{SQRT} \sum \{C_i^2 s^2 / n_i\}$$

ここで C_i : 群*i*の対比の係数 (Z の期待値を0にするため $\sum C_i = 0$ とする)

$Y_{i.}$: 第*i*群の平均値

n_i : 第*i*群のサンプルサイズ

s^2 : 誤差分散 $\sum \sum \{Y_{ij} - Y_{i.}\}^2 / \sum \{n_i - 1\}$

毒性試験では設定する用量範囲がかなり広いため、様々な用量反応関係が生じる可能性があり、次のような対比の係数を設定することが考えられる。

表1 4群の場合の最大対比法の対比の係数(coefficients of Maximum contrasts)

| 対照 | 低用量 | 中用量 | 高用量 | 対比の意味(略号) |
|----|-----|-----|-----|---------------------|
| -3 | -1 | 1 | 3 | 直線的に増加(linear) |
| -7 | -5 | -1 | 13 | 用量に対して直線的に増加(reg) |
| -3 | 1 | 1 | 1 | 低用量で飽和(l-end) |
| -5 | -1 | 3 | 3 | 中用量で飽和(m-end) |
| -3 | -3 | 1 | 5 | 中用量から立ち上がり(m-start) |
| -1 | -1 | -1 | 3 | 高用量から立ち上がり(h-start) |

また対比の群間平方和に対する寄与率 R^2 は次のように計算される。通常 R^2 は全平方和についての寄与を計算するが本稿では、群間平方和について計算する。

$$R^2 = \sum n_i \{b C_i\}^2 / \sum n_i \{Y_{i.} - Y_{..}\}^2$$

ここで $b = \sum n_i Y_{i.} C_i / \sum n_i C_i^2$ である。

このように複数の対比統計量の値を同時に計算し、その最大値 Z_{max} に基づいて、用量反応関係についての情報を得るのが最大対比法である。用量反応関係について評価する際、偶然を越えた意味ある用量反応関係のみを拾い上げる必要がある。このためには検定が有効な手段であるが、複数の対比について検定を行うと、多重性の問題によって、第一種の過誤が増大するため、有意水準を調整した解析が必要になる。任意の複数の対比に対して多重性を調整する古典的な多重比較法として、Bonferroni、Scheffe法が知られているが、どちらも過度に保守的になることが知られている。特にこの例のように対比統計量間の相関が高いときは、非常に保守的になってしまう。これに対し、任意の複数の対比について、等分散性と多変量正規分布を前提として、多重積分によって多重性調整p値を計算するプログラムを岸本・浜田(1994)を作成しているが、指定する対比の種類、サンプルサイズによって棄却限界値は異なるため、汎用的なアプローチとはいえない。

これに対し、多重性に関する調整p値を resampling (標本再抽出法) によって計算

することを、Westfall and Young(1992)は提案している。この方法は正確な並び替え検定の p 値を近似するものである。また SAS MULTTEST はこの目的のために開発されたプログラムである(浜田・吉田(1992))。

標本再抽出法では、次の手順によって p 値を多重性の問題について調整する。

- 1) 観測データについて、複数の対比のそれぞれについて検定を行い p 値を計算する。
- 2) 観測データから非復元無作為抽出を行って(群番号を並び替えて)、元の観測データと同じサンプル・サイズの擬似的な標本を作成する。
- 3) 擬似的な標本に対し 1) と同じ検定を行い p 値を計算する。このうち最小の p 値を $Min\ p$ 値とする。
- 4) 2), 3) を繰り返して、 $Min\ p$ 値の分布関数 F を十分な精度で推定する。この繰り返し数を標本再抽出回数とよぶ。
- 5) 1) の個々の検定について生の p 値が $Min\ p$ 値より小さくなった回数を再抽出回数で割ったものが調整 p 値となる。

3. 最大対比法の適用例

4 群で行われた毒性試験の赤血球データ(表 2)に基づいて、最大対比法の適用例を示す。この例では薬物投与によって赤血球数の減少傾向が観測されている。

表 2 例題 赤血球データ(単位:mg/dl)

| 群 | 用量 | 生データ | | | | | | | | 平均 | S D | | |
|-----|------|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-------|------|
| 対照 | 0mg | 925 | 917 | 912 | 912 | 949 | 908 | 908 | 989 | 931 | 909 | 926.0 | 25.7 |
| 低用量 | 1mg | 898 | 925 | 908 | 873 | 908 | 941 | 893 | 920 | 922 | 931 | 911.9 | 20.1 |
| 中用量 | 3mg | 874 | 876 | 916 | 908 | 873 | 807 | 874 | 919 | 952 | 916 | 891.5 | 39.8 |
| 高用量 | 10mg | 869 | 919 | 874 | 852 | 830 | 906 | 914 | 898 | 933 | 935 | 893.0 | 35.4 |

表 3 適用結果 (Result of maximum contrast method)

| Contrast | non-adjusted p-value | adjusted p-value | Z | R ² (R-square) |
|----------|-------------------------|---------------------|---------|------------------------------|
| linear | 0.0104* | 0.0258* | -2.7052 | 0.876 |
| m-end | 0.0067* | 0.0177* | -2.8778 | 0.992 |
| l-end | 0.0224* | 0.0532 | -2.3868 | 0.682 |
| reg | 0.0426* | 0.1018 | -2.1017 | 0.529 |
| m-start | 0.0216* | 0.0515 | -2.4013 | 0.695 |
| h-start | 0.1491 | 0.3249 | -1.4742 | 0.260 |

R² :contribution rate for between sum of squares

* :significant at 5% level

6本の対比をあてはめた結果を表3に示した。表中で non-adjusted p-value は多重性について調整せずに、t 分布で近似した両側 p 値、adjusted p-value は SAS の MULTTEST プロシジャによって 10000 回の標本再抽出によって調整した両側 p 値である。また Z と R² は対比統計量と、群間平方和に対する対比の寄与率を示している。調整前の p 値では 5本の対比が、両側 5%の水準で有意であるが、調整した p 値で有意になるのは linear と

m-end の2つのみである。絶対値が最大の対比は m-end で、 R^2 が 0.99 であるので、群間平方和のほぼ全てを m-end という対比によって説明できることがわかる。すなわち中用量までは直線的に赤血球数は直線的に低下し、中用量で飽和に達することがわかる。linear についても有意ではあるが、寄与率は m-end よりかなり劣っている。6本の対比について、調整前と後で p 値の大きさを比較すると調整後は 2-3 倍になっている。Bonferroni では、p 値に用いた対比の数をかけることになるが、この方法は標本再抽出法に比べて、過度に保守的になることが確認できる。

4. 最大対比法の性能評価

以下ではシミュレーション実験により最大対比法の検出力を調べ、回帰分析と比較する。また最大対比法による用量反応パターンの解析結果と人による判断の一致度について評価する。

4. 1 シミュレーション実験による最大対比法の検出力の評価

対照群を含めて4群の場合について表1に示した6種類の対比を設定した場合の最大対比法の検出力を評価した。ここで検出力は、対立仮説の下でいずれかの有意な用量反応関係を検出できる確率として定義した。

以下のシミュレーションでは簡便のため、等分散（分散既知）、正規分布で例数がバランスがとれていることを前提としたが、比較する手法は全て等分散と正規性を仮定したパラメトリック手法であるため、この条件が成り立たない場合でも、性能は相対的には異ならないと考えられる。

標本再抽出法によって個々のシミュレーション実験ごとに、多重性調整 p 値を計算するのは時間がかかるので、まず前提条件に基づき、100万回のシミュレーション実験によって両側 5% 点を求めた。帰無仮説の下で Z 統計量の最大値の棄却限界値は 2.36 となった。

次に示すシミュレーション条件において、最大対比法の検出力と、回帰分析 (-3, -1, 1, 3 の対比を用いた場合)、対立仮説に合せて理論的に最も検出力が高くなるように対比を選んだ場合（以下では最適法と呼ぶ）の検出力を比較した。

・シミュレーション条件

- 1) 群構成：4群（対照群+3用量群）
- 2) サンプルサイズ：1群あたり 10
- 3) 4群の期待値を $\mu_1, \mu_2, \mu_3, \mu_4$ として、SAS の rannor 関数を用いて、 $N(\mu_i, 1)$ 正規乱数を発生させる。期待値の μ_i については表4に示した。
- 4) シミュレーション回数：それぞれの期待値の組み合わせについて 100000 回
- 5) 有意水準：両側 5%（棄却限界値：2.36(最大対比法) 1.96(回帰分析、最適法)）
- 6) 適用手法
最大対比法、回帰分析、最適法（表4に対比の係数を示した）
- 7) 対立仮説：表4に示した 10通りの対立仮説を想定した 1~6 までは最大対比法

で想定した対比に対応する用量反応関係、7は、最大対比法では想定していない中用量で階段状に変化するパターン、8-10は非単調な用量反応関係のパターンである。

表4 シミュレーションで想定した対立仮説 (alternative hypotheses)

| No. | 用量反応関係の内容 | 記号 | $\mu_1, \mu_2, \mu_3, \mu_4$ | 最適な対比の係数 |
|-----|--------------------|-----------|------------------------------|----------------|
| 1 | 直線的 | linear | $\Delta(-1, -1/3, 1/3, 1)$ | -3, -1, 1, 3 |
| 2 | 指数的 | exp | $\Delta(-1, -0.8, -0.4, 1)$ | -7, -5, -1, 13 |
| 3 | 低用量で飽和 | l_end | $\Delta(-1, 1, 1, 1)$ | -3, 1, 1, 1 |
| 4 | 中用量で飽和 | m_end | $\Delta(-1, 0, 1, 1)$ | -5, -1, 3, 3 |
| 5 | 中用量から立ち上がり | m_start | $\Delta(-1, -1, 0, 1)$ | -3, -3, 1, 5 |
| 6 | 高用量から立ち上がり | h_start | $\Delta(-1, -1, -1, 1)$ | -1, -1, -1, 3 |
| 7 | 中用量から立ち上がって、中用量で飽和 | l_m | $\Delta(-1, -1, 1, 1)$ | -1, -1, 1, 1 |
| 8 | 2次曲線型 | quadratic | $\Delta(-1, 1, 1, -1)$ | -1, 1, 1, -1 |
| 9 | 頭打ち1 | downturn1 | $\Delta(-1, 0, 1, 0)$ | -1, 0, 1, 0 |
| 10 | 頭打ち2 | downturn2 | $\Delta(-1, 0, 1, 0.5)$ | -9, -1, 7, 3 |

Δ : 0~1まで0.1刻みで動かした。

・シミュレーションの結果

表5にシミュレーションの結果を示した。

- 1) linear(1)、m_end(4)、m_start(5)の3つの用量反応関係については、回帰分析の方が最大対比法に比べて検出力が勝るが、その違いは大きくても5%程度である。また最適な方法と比べても最大対比法の検出力は、悪くても7%程度下がる程度である。
- 2) exp(2)の用量反応関係については、最大対比法と回帰分析では検出力はほとんど変わらない。最適な方法と比べて、この2つの手法の検出力は5%程度劣る場合がある。
- 3) l_end(3)とh_start(6)の用量反応関係については、回帰分析と比べて最大対比法の方が検出力が高く、大きい場合には10%以上差が生じる。最適な方法と比べて場合でも、最大対比法は10%以上劣ることがない。
- 4) 低用量まで変化がなく、中用量で飽和する用量反応関係 l_m(7)については、最大対比法を構成する6本の対比には、対応するものを含めておらず、この場合には回帰分析の方が検出力が高くなるが、その違いはせいぜい5%程度である。最適な方法と比べても、検出力の違いは大きくても10%強である。
- 5) 2次曲線型の用量反応関係があるときには、回帰分析の検出力は0となるのに対し、最大対比法では、最適な方法と比べて検出力は落ちるものの、ある程度の検出力が保持できる。
- 6) 頭打ちがあるとき(9, 10)では、最大対比法の方が回帰分析より検出力が高く、特に downturn1(9)では、20%以上高くなる場合もある。

表5 シミュレーションの結果(results of simulation)

| No. | D-R shape | Method | Δ | | | | | | | | | | |
|-----|-----------|--------|----------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| | | | 0 | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 | 1 |
| 1 | linear | MCM | 0.049 | 0.072 | 0.138 | 0.261 | 0.420 | 0.599 | 0.759 | 0.877 | 0.947 | 0.982 | 0.994 |
| 1 | linear | REG | 0.050 | 0.077 | 0.155 | 0.293 | 0.469 | 0.655 | 0.806 | 0.910 | 0.964 | 0.989 | 0.997 |
| 1 | linear * | OPT | 0.050 | 0.077 | 0.155 | 0.293 | 0.469 | 0.655 | 0.806 | 0.910 | 0.964 | 0.989 | 0.997 |
| 2 | exp | MCM | 0.049 | 0.072 | 0.142 | 0.269 | 0.440 | 0.628 | 0.788 | 0.901 | 0.961 | 0.988 | 0.997 |
| 2 | exp | REG | 0.050 | 0.075 | 0.147 | 0.274 | 0.440 | 0.619 | 0.773 | 0.886 | 0.950 | 0.983 | 0.995 |
| 2 | exp | OPT | 0.051 | 0.079 | 0.165 | 0.317 | 0.505 | 0.696 | 0.840 | 0.933 | 0.977 | 0.994 | 0.999 |
| 3 | l_end | MCM | 0.049 | 0.073 | 0.149 | 0.293 | 0.484 | 0.689 | 0.846 | 0.941 | 0.982 | 0.996 | 0.999 |
| 3 | l_end | REG | 0.050 | 0.072 | 0.135 | 0.247 | 0.395 | 0.565 | 0.719 | 0.844 | 0.924 | 0.970 | 0.988 |
| 3 | l_end | OPT | 0.049 | 0.084 | 0.195 | 0.376 | 0.591 | 0.780 | 0.907 | 0.971 | 0.992 | 0.999 | 1.000 |
| 4 | m_end | MCM | 0.049 | 0.075 | 0.152 | 0.295 | 0.478 | 0.674 | 0.830 | 0.929 | 0.976 | 0.994 | 0.999 |
| 4 | m_end | REG | 0.050 | 0.080 | 0.166 | 0.318 | 0.506 | 0.697 | 0.842 | 0.933 | 0.977 | 0.994 | 0.999 |
| 4 | m_end | OPT | 0.050 | 0.083 | 0.181 | 0.349 | 0.551 | 0.746 | 0.880 | 0.957 | 0.987 | 0.997 | 1.000 |
| 5 | m_start | MCM | 0.049 | 0.075 | 0.153 | 0.295 | 0.480 | 0.676 | 0.831 | 0.929 | 0.976 | 0.994 | 0.999 |
| 5 | m_start | REG | 0.050 | 0.080 | 0.166 | 0.318 | 0.506 | 0.697 | 0.842 | 0.933 | 0.977 | 0.994 | 0.999 |
| 5 | m_start | OPT | 0.051 | 0.083 | 0.181 | 0.350 | 0.553 | 0.747 | 0.881 | 0.955 | 0.987 | 0.997 | 0.999 |
| 6 | h_start | MCM | 0.049 | 0.074 | 0.149 | 0.294 | 0.489 | 0.691 | 0.846 | 0.941 | 0.982 | 0.996 | 0.999 |
| 6 | h_start | REG | 0.050 | 0.072 | 0.135 | 0.247 | 0.395 | 0.565 | 0.719 | 0.844 | 0.924 | 0.970 | 0.988 |
| 6 | h_start | OPT | 0.050 | 0.086 | 0.195 | 0.375 | 0.591 | 0.783 | 0.906 | 0.969 | 0.992 | 0.998 | 1.000 |
| 7 | l_m | MCM | 0.049 | 0.081 | 0.178 | 0.358 | 0.579 | 0.785 | 0.913 | 0.974 | 0.994 | 0.999 | 1.000 |
| 7 | l_m | REG | 0.050 | 0.089 | 0.203 | 0.396 | 0.617 | 0.808 | 0.923 | 0.977 | 0.995 | 0.999 | 1.000 |
| 7 | l_m | OPT | 0.050 | 0.098 | 0.241 | 0.475 | 0.714 | 0.885 | 0.966 | 0.993 | 0.999 | 1.000 | 1.000 |
| 8 | quadratic | MCM | 0.049 | 0.065 | 0.117 | 0.214 | 0.361 | 0.548 | 0.732 | 0.877 | 0.955 | 0.988 | 0.998 |
| 8 | quadratic | REG | 0.050 | 0.050 | 0.050 | 0.050 | 0.050 | 0.050 | 0.050 | 0.050 | 0.051 | 0.049 | 0.050 |
| 8 | quadratic | OPT | 0.050 | 0.096 | 0.245 | 0.474 | 0.714 | 0.887 | 0.966 | 0.994 | 0.999 | 1.000 | 1.000 |
| 9 | downturn1 | MCM | 0.049 | 0.061 | 0.096 | 0.165 | 0.266 | 0.397 | 0.545 | 0.690 | 0.811 | 0.898 | 0.949 |
| 9 | downturn1 | REG | 0.050 | 0.059 | 0.087 | 0.135 | 0.204 | 0.292 | 0.395 | 0.507 | 0.620 | 0.722 | 0.806 |
| 9 | downturn1 | OPT | 0.049 | 0.073 | 0.145 | 0.269 | 0.428 | 0.609 | 0.765 | 0.880 | 0.948 | 0.980 | 0.994 |
| 10 | downturn2 | MCM | 0.049 | 0.067 | 0.119 | 0.217 | 0.357 | 0.524 | 0.691 | 0.826 | 0.917 | 0.967 | 0.989 |
| 10 | downturn2 | REG | 0.050 | 0.068 | 0.121 | 0.215 | 0.343 | 0.493 | 0.644 | 0.775 | 0.874 | 0.940 | 0.973 |
| 10 | downturn2 | OPT | 0.049 | 0.076 | 0.153 | 0.288 | 0.460 | 0.647 | 0.799 | 0.906 | 0.963 | 0.988 | 0.997 |

MCM:最大対比法 REG:回帰分析 OPT:最適法

*:linear な用量反応関係があるとき、最適な対比の係数は-3, -1, 1, 3 となり、回帰分析の結果と同じになる。

以上の結果をまとめると、単調性のある様々な用量反応関係について、最大対比法と回帰分析で検出しやすいパターンに違いはあるものの、全体的にはその差は小さく、最大対比法の検出力は、回帰分析と比べて、同等であり、また、最適な対比を1本のみ用いる場合と比べてもそれほど性能が落ちるものではない。2次曲線的な変化や、頭打ち現象などが起こり、単調性が成り立たない場合については、最大対比法の方が回帰分析より検出力がかなり高くなる。

4. 2 最大対比法による用量反応パターンの解析と人の判断の一致度の評価

最大対比法による用量反応パターンの判定と毒性家の判断の一致度を評価した。

・方法

- 1) 毒性学の専門家7人と、統計解析の業務従事者5人をノミネートした。
- 2) 背景データベース(Hamada et al. (1997))から、4群で行われた試験のデータを選択した。4)で示すa)～f)の用量反応関係を各2通り(寄与率が95%以上と90%前後なるようなものを選択した)、計12通り、g)のタイプとして、中用量から立ち上がって、中用量で飽和するタイプを2通り、非単調な場合を2通り、h)として有意な用量反応関係がない場合を2通り、計18項目を選択した。
- 3) 18項目について表1に示した6本の対比統計量を計算し、各対比の調整しないp値、多重性について調整したp値、群間平方和に対する対比の寄与率を計算した。
- 4) 1)の12人が独立に、無作為に順序を並べ替えた2)の18項目の、生データ、要約統計量(平均値、SD)、散布図に基づき、a)～h)の8種類の用量反応関係に分類した。
 - a) 直線的に増加(等差的に増加)
 - b) 指数的に増加(等比的に増加)
 - c) 低用量で飽和
 - d) 中用量で飽和
 - e) 中用量から立ち上がり
 - f) 高用量から立ち上がり
 - g) a)～f)以外の用量反応関係がある
 - h) 用量反応関係は認められない

・結果

表6に人による判断と最大対比法による判定が比較できる形で結果をまとめてみた。12人の評価者は7人の毒性家と5人の統計の業務従事者から構成されているが、用量反応パターンの判定について専門分野間で顕著な差がみられなかったため、12人を求めた結果のみを以下では示す。最大対比法によってa)～f)と判定される用量反応関係が1～12まで示されており、13, 14は最大対比法では想定してない、対比の係数を-1, -1, 1, 1とったとき高い値をとる用量反応関係である。13と14については参考として、下段に1, -1, 1, 1の対比の寄与率が示されている。15, 16は低用量で上がって、中用量で下がって、高用量で再び上がる非単調な用量反応関係がある場合である。このようなケースでは最大対比は有意となるものの、寄与率は50%前後と低く、最大対比法で想定した用量反応関係とは異なることがわかる。17, 18は用量反応関係が認められないケースで、最大対比の調整p値が0.05以上になる。

1) 用量反応関係がa)～f)のケース

評価者間で判定にばらつきがあることがわかる。12人の評価者間で結果が完全に一致するのは12項目中2項目(ROVARYとTP)のみである。残りの10項目の中で、6番のA_Gについては12人中9人までは、低用量で飽和(c)、残りの3人は用量反応関係なし(h)と判断しており、用量反応パターン自体は1通りにしか判断されていない。11番のRKIDNEYと12番のBW2についても12人中11人までは、高用量から立ち上がる(f)と判

断しており、異なる判断をしたのは1人のみであり、ほぼ評価者間で一致した結果が得られている。残りの項目については、例えばASTのように調整p値(0.0000)が高度に有意でかつ、対比の寄与率(0.9997)が90%を超える場合であっても、評価者間で判定が異なる場合がある。この原因については個別に説明するが、評価者間の判定のパラツキはあるものの、7番のSPLEENを除いて、最大対比法の判定は、評価者間の最多数の判定に一致している。SPLEENについても、最大対比d(m-end)に対し、a(linear)の判定が6人と最も多くなっているが、5人はd(m_end)と判定している。1~12の12項目については、それぞれの用量反応関係に対応した最大対比の調整p値は5%水準で有意であり、寄与率も90%を越えているので、この2つの条件が満たされれば、最大対比法による判定は、ほぼ評価者の間の最多数の判定に一致すると考えてよいことがわかる。

1番のASTについては、a(linear)6人と、f(h_start)3人に判定が分れている。平均値はきれいに直線的に上がっているためlinearの対比の寄与率は0.9997と高いが、散布図をみると、中用量までは大きな変化はなく、高用量において多少外れ値気味に2個体の値が上昇している。この個体のみ変化していると考えた3人は高用量のみ変化があるf(h_start)と判定したと考えられる。

3番のRADR2と9番のBWについてはb(exp)とe(m-start)に判定が分かれているが、この2つは対比の係数が-7, -5, -1, 13, -3, -3, 1, 5と似ているため、区別しにくかったのが原因と考えられる。RADR2ではexpの寄与率が0.9998に対し、m-startのそれは0.9305、BWではexpの寄与率が0.9407に対し、m-startのそれは0.9978と、両方の対比の寄与率とも90%を越え、大きな値をとっている。

4番のALPについては、b(exp)5人、f(h_start)3人、g(その他)3人、a(linear)1人の4通りに判定が分れている。4群でそれぞれの平均値は80, 96, 94, 134となり、低用量と中用量とで平均値が逆転し、平均値は用量に対して単調には変化してない。この原因として、低用量群に高い方で外れた値が存在し平均値を引っ張っていることが上げられる。この外れた値以外では、用量の増加に伴い、指数的に反応が上昇している。外れた値の影響を割り引いて評価した人は、b(exp)と判定し、平均値の値を重視した人は、高用量のみ変化するf(h_start)と判定したと考えられる。

7番のSPLEENについては、a(linear)6人とd(m-end)5人に判定がほぼ2つに割れているが、これは、この2つの対比の寄与率が高く似通った値を取っているためである。寄与率はd(m-end)が0.9978、a(linear)が0.9407となっている。

8番のRLIVERについてはd(m-end)7人、a(linear)2人、c(l-end)1人、f(h_start)1人、g(その他)1人の5通り、10番のTHYMUSについてはe(m-start)7人、g(その他)4人、f(h_start)1人の3通りに判定が割れている。各群の平均値はRLIVERでは2.69, 2.72, 2.81, 2.80、THYMUSについては211, 224, 199, 162となり、どちらも単調性が成り立っていない。RLIVERについてはほぼ単調性が成り立っているような印象も受

けるが、高用量群に値の高い外れ値が存在し、平均値を引っ張っており、この値を除くと頭打ちのような現象が観測される。THYMUS については低用量で一旦値が上昇してから下降する傾向が、グラフから読み取れ、これが4人もg(その他)と判断した原因であると考えられる。

以上の結果から、調整p値が有意であり、寄与率が90%を超えるときは最大対比法による判定は、評価者間の最多数の判定に一致するといえるが、外れ値が存在する、平均値が用量に対して単調に変化してない、似たような寄与率の対比が複数存在するケースでは、評価者間で判定が分れる場合がある。

2) 用量反応関係がgのケース

13番のRHEARTと14番のRADRは、最大対比法を構成する6本の対比では含まれていない-1,-1,1,1の対比が最も大きくなる場合である。このようなケースでも、最大対比法で最も寄与率が高くなった対比に対応する用量反応関係が多く選択されている(RHEART:d(m-end)6人、RADR:e(m-start)8人)。しかしながらg(その他)の判定も、それぞれ5人と3人と多く、a)~f)以外の用量反応関係があると判断した人も多かったことが判る。中用量で階段状に変化する用量反応関係が多く存在するのであれば、-1,-1,1,1の対比も最大対比法の構成対比に加える必要があるが、通常用量反応関係は連続的であり、このような用量反応関係はかなり用量範囲を広くとらない限りはあまり現実的とはいえない。

非単調な用量反応関係がある15番のTCと16番のRPITの場合では、最大対比は5%水準で有意になるものの寄与率は低い、このようなケースで最大対比に対応する用量反応関係を選択する評価者は12人中1人のみである、残りの評価者は、h(用量反応関係なし)か、g(その他)と判定している。

3) 用量反応関係がhのケース

有意な用量反応関係がない17番のTGと18番のPLTは場合では、最大対比に対応する用量反応関係c(1-end)を選択する評価者も、TG,PLTでそれぞれ、1人、2人存在するものの残りはh(用量反応関係なし)を選択しており、最大対比法で有意とならない結果を反映したものと見える。

以上、最大対比法で考慮した用量反応関係があるケース、想定しない用量反応関係があるケース、用量反応関係なしのいずれの場合においても、評価者間で多少の判定のバラツキはあるものの、最大対比法による用量パターンの解析結果は、評価者間で最も多い判定に一致するといえる。

最大対比法では、回帰分析と比べて同等以上の検出力を持っており、また調整p値が有意かつ寄与率が90%以上であるときは、最も多くの人が判断する用量パターンに一致することがわかった。以上の結果は用量反応パターンを統計学的に客観的に判断する手段としての最大対比法の妥当性を示すものといえる。

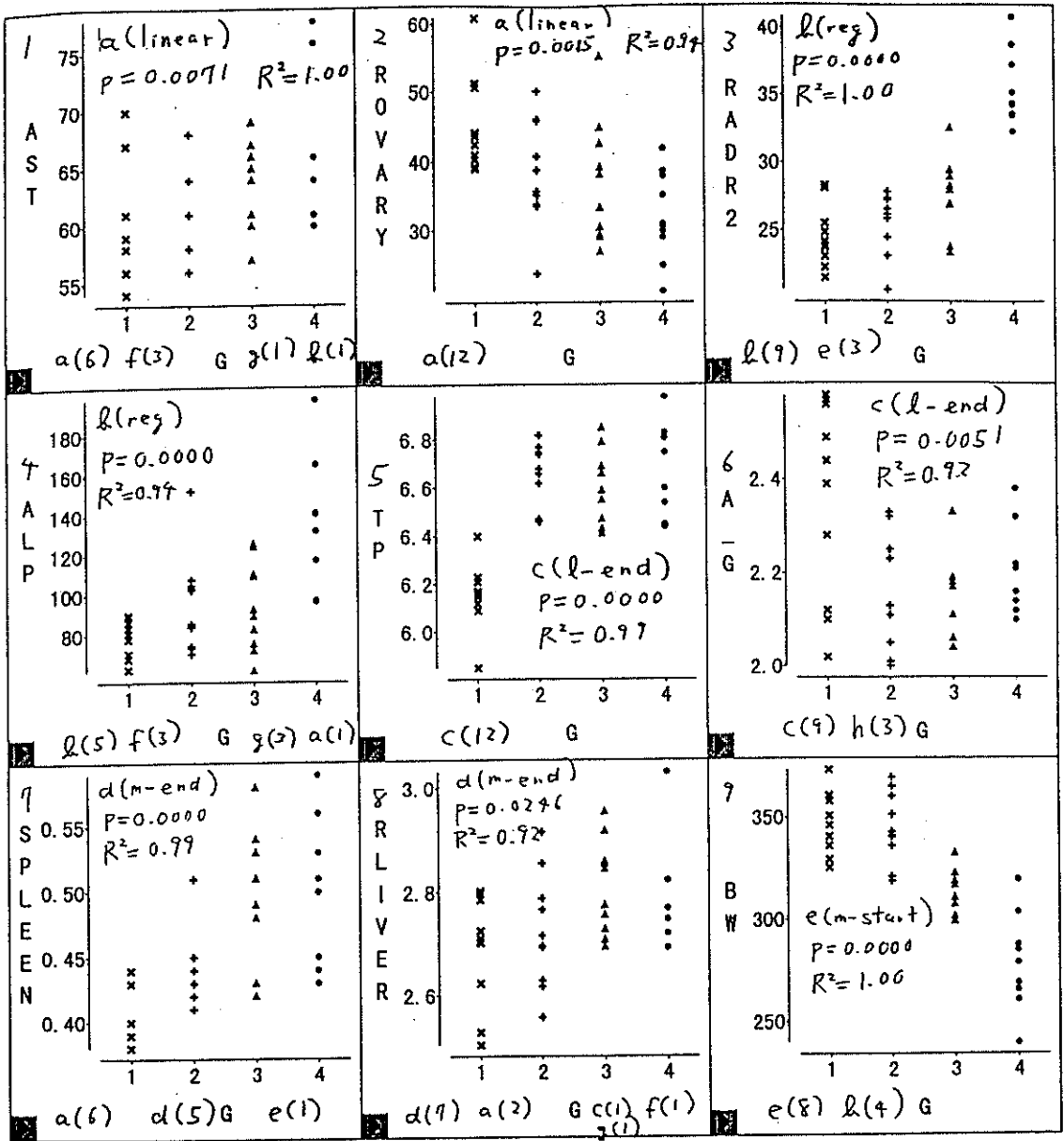
表6 用量反応パターンの判定(Judgment for dose-response patterns)

| No | Item | Content of item | Judgment by researchers | Maximum contrast | Raw-p value | Adjusted p value | R ² |
|----|------------|-----------------------------|-----------------------------|-------------------------|------------------|------------------|------------------|
| 1 | AST a) | aspartate amino-transferase | a(6) f(3) g(1) b(1) h(1) | a(linear) | 0.0024 | 0.0071** | 0.9997 |
| 2 | ROVARY a) | relative ovary w. | a(12) | a(linear) | 0.0004 | 0.0015** | 0.9443 |
| 3 | RADR2 b) | r. adrenal w. | b(9) e(3) | b(reg) | 0.0000 | 0.0000** | 0.9998 |
| 4 | ALP b) | alkaline phosphatase | b(5) f(3) g(3) a(1) | b(reg) | 0.0000 | 0.0000** | 0.9445 |
| 5 | TP c) | total protein | c(12) | c(l-end) | 0.0000 | 0.0000** | 0.9923 |
| 6 | A_G c) | albumin globuline ratio | c(9) h(3) | c(l-end) | 0.0015 | 0.0051** | 0.9188 |
| 7 | SPLEEN d) | spleen w. | a(6) d(5) e(1) | d(m-end) | 0.0000 | 0.0000** | 0.9893 |
| 8 | RLIVER d) | r. liver w. | d(7) a(2) c(1) f(1) g(1) | d(m-end) | 0.0092 | 0.0246* | 0.9181 |
| 9 | BW e) | body weight | e(8) b(4) | e(m-start) | 0.0000 | 0.0000** | 0.9978 |
| 10 | THYMUS e) | thymus w. | e(7) g(4) f(1) | e(m-start) | 0.0000 | 0.0000** | 0.9305 |
| 11 | RKIDNEY f) | r. kidney w. | f(11) e(1) | f(h-start) | 0.0000 | 0.0000** | 0.9429 |
| 12 | BW2 f) | body weight | f(11) h(1) | f(h-start) | 0.0034 | 0.0114* | 0.9020 |
| 13 | RHEART g) | relative heart w. | d(6) g(5) h(1) | d(m-end) -1 -1 1 1 | 0.0006 0.0005 | 0.0013** | 0.9040 0.9178 |
| 14 | RADR g) | r. adrenal w. | e(8) g(3) a(1) | e(m-start) -1 -1 1 1 | 0.0000 0.0000 | 0.0001** | 0.8572 0.9881 |
| 15 | TC g) | total cholesterol | h(8) g(3) c(1) | c(l-end) | 0.0050 | 0.0143* | 0.4536 |
| 16 | RPIT g) | r. pituitary w. | h(6) g(5) f(1) | f(h-start) | 0.0030 | 0.0094** | 0.5136 |
| 17 | TG h) | triglyceride | h(11) c(1) | c(l-end) | 0.3718 | 0.6784 | 0.8772 |
| 18 | PLT h) | platelet count | h(10) c(2) | c(l-end) | 0.0389 | 0.0673 | 0.9719 |

() : the number of person

参考文献

- Ruberg, S. J. (1989) Contrasts for identifying the minimum effective dose, JASA, 84, 816-822
- 浜田知久馬・岸本淳司・大塚芳正(1993). 臨床試験における至適用量設定のための統計学的アプローチ 応用統計学会予稿集, 40-44
- 吉村功、浜田知久馬(1996) 医薬データの統計解析における最大対比法の活用、統計学会予稿集 42-43
- 岸本淳司、浜田知久馬(1994) 任意の対比群について多重比較を行う数値積分プログラム 計算機統計学 7, 147-154
- 浜田知久馬・吉田道弘(1992). MULTTEST プロシジャの紹介. SUGI-J'92 論文集, 357-370
- Hamada C et al. (1997) A study on the consistency between statistical evaluation and toxicological judgment. Drug Information Journal, 31-2



[質問]

尿中蛋白濃度の測定結果が 2種の尿検査紙で違ったら？

東北労災病院 健康診断センター
石原 信夫

尿中蛋白の有無の検査を実施しなければならない場合

1: 仕事の内容に係わらず実施しなければならないもの
(=一般健康診断)

- 1) 雇入時の健康診断
- 2) 定期健康診断

2: 仕事の内容に応じて実施しなければならないもの
(=特殊健康診断)

- 1) 特定化学物質(44種)の取り扱い
- 2) 有機溶剤
- 3) 鉛及びその無機化合物

→いわゆる健康診断においては殆どの場合、尿中蛋白の有無の検査が行われる。また、病院・診療所においては初診時検査の一つとして広く行われている。

→検査手法は殆ど全ての場合「試験紙法」

「東北労災病院健康診断センター」

尿中蛋白の有無の検査→陰性(-): 腎障害は無いと判断

↓
陽性(+)

- 1) 数日後に再検査 → 陰性(-): 腎障害は無いだろう
→ 陽性(+以上)

- 2) 病院等に紹介、専門医による診察

↓
疑陽性(+/-): 通常陰性(-)と同じに扱う。

但し、他の所見があれば再検査。

尿蛋白陽性(+以上)の人がよその医療機関を受診した場合には、再度、尿蛋白の有無の検査を受けるが、当センターの結果と一致しなくても、問題は起きない。

→医療機関が異なれば、検査結果が一致しなくとも、必ずしも不思議ではない(分析方法、検査担当者の習熟度などが異なるから)。また、「あそこの検査はあまりあてになりませんな」との陰口は少ない。

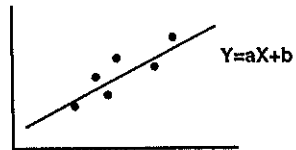
東北労災病院を受診した場合

- 1) 外来での尿検査は病院検査科で行う。
(センターとは試験紙が異なる)
結果が食い違くとセンターの悪口がでる。
- 2) 健康診断センターで健康診断を受けた人が、病院で行っている一泊二日の人間ドックに入る場合が多い。
人間ドックでは、外来よりも、センターにおける検査結果が重視される。

→検査結果が不一致の場合の雑音が一層激しい

健康診断センターの臨床検査の責任者と病院検査科の責任者は同じ

↓
病院検査科での
検査結果



センターでの検査結果

のような相関図めいたものを作り、直線回帰式を求め、センターでの測定結果を病院での測定結果に換算(?)し、食い違いに対処したいとの提案があった。

私の意見：「試験紙が異なれば結果が異なるのは当たり前であり、同じ責任者の下で、異なる二つの方法が使用されていることの方が問題である。むやみやたらと回帰式による換算をすればよいという性質の問題ではない、やめてくれ」

「相関図で二つの方法による結果が良く一致しているというが、そんなことは当たり前。もし試験紙の種類が異なれば、測定結果も大きく異なるというのであれば、試験紙による測定自体が無意味になる」

会議での大勢：もっともらしい数式の力は偉大（遠大！）で、私の意見は少数派。

「二種類の尿検査紙で尿中蛋白濃度の測定結果が違った場合、どう対処すればよいか」

尿中蛋白濃度の測定結果が 2種の尿検査紙で違ったら？ (回答)

大鵬薬品・開発情報センター
DM部 佐野 正樹

A社、B社の尿蛋白データ

| | | A社の尿蛋白値 | | | | |
|---------|-------|---------|-------|-----|------|-------|
| | | (-) | (-/+) | (+) | (++) | (+++) |
| B社の尿蛋白値 | (-) | 47 | 2 | 2 | 1 | 2 |
| | (-/+) | | 7 | 4 | | 1 |
| | (+) | | | 9 | | 1 |
| | (++) | | | 2 | 4 | 1 |
| | (+++) | | | 1 | | 1 |

Total No = 85

東北労災病院の多数意見

1. A法とB法の順位相関は大きく、かつ有意に高い
 $r=0.8064$, $P<0.001$
2. 両法の傾きも切片も推定できる
3. すなわち、両法の測定結果は相互に変換できるので、まったく問題はない

石原先生の質問

1. A社製試験紙の方が高く出るようだ。会議で説得できる証明法があるか？
2. 相関係数や回帰直線を求めることに、意味があるだろうか？
3. どういう行動をとるべきか？

はじめに

この回答は、第73回医薬安全性研究会定例会(1998年1月31日)で発表した内容に、少しだけ修正と追加をしたものです。

石原先生の質問書には、左に示すような尿蛋白測定データが添付されていました。

健康診断に来られた方々の尿が、A社の尿試験紙とB社の尿試験紙でそれぞれ測定されました。各セル内の数字は人数(尿サンプル数)を示し、対角方向の数字は測定値が一致した尿サンプルということになります。

石原先生が勤務されている病院では、大多数の方々がこの結果に問題はないと考えられたようです。

2つの方法で測定した結果は高度に有意な相関を示していますので、一方の測定値を他方に容易に変換できそうです。

何か議論の方向が違うのじゃないかと石原先生は感じました。しかし、統計学的解析という壁に阻まれ、職員の皆さんを説得することは出来ませんでした。

石原先生のご質問を簡単にまとめると左のようになりますので、回答に移りたいと思います。

両社で、仮に差がなかったら

| | | A社の尿蛋白量 | | | | |
|---------|-------|---------|-------|-----|------|-------|
| | | (-) | (-/+) | (+) | (++) | (+++) |
| B社の尿蛋白量 | (-) | 47 | 1 | | | |
| | (-/+) | 1 | 7 | 2 | | |
| | (+) | | 2 | 9 | 2 | 2 |
| | (++) | | | 2 | 4 | 2 |
| | (+++) | | | 1 | 2 | 1 |

Total No = 85

もし、A社とB社の測定に違いがなければ、一致しなかった尿試料は対角セルの両側にほぼ均等に現れます。たとえば、左のような感じの表になるはずです。

つぎに、この表と石原先生が示された下表を見比べてください。

A社では高くでているようだ

| | | A社の尿蛋白量 | | | | |
|---------|-------|---------|-------|-----|------|-------|
| | | (-) | (-/+) | (+) | (++) | (+++) |
| B社の尿蛋白量 | (-) | 47 | 2 | 2 | 1 | 2 |
| | (-/+) | | 7 | 4 | | 1 |
| | (+) | | | 9 | | 1 |
| | (++) | | | 2 | 4 | 1 |
| | (+++) | | | 1 | | 1 |

Total No = 85
Upper = 14
Lower = 3

A社の測定値が高めに偏っているのが見て取れます。

一方、両社ともに(-)となった試料が55% (47人分)も占めています。どんなに偏った検査法でも0は0になりますから、ここで相関が高いといっても当てにはなりません。試験紙法での偏りは重大な問題となるのです。

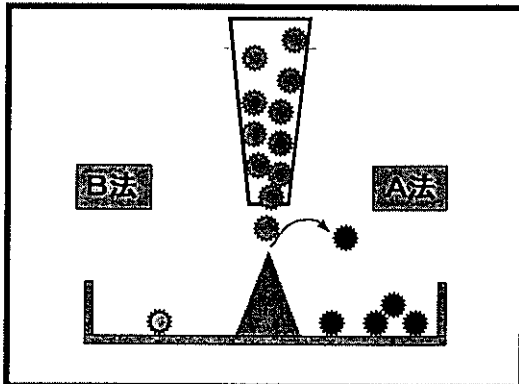
統計的アプローチ

- 1) 対角線に対する対称性の検定
 - (1) 二項分布を前提とした検定
Upper = 14, Lower = 3
 - (2) スコアを使った検定
- 2) 対応あるデータの検定

これらの測定値に偏りがあることを左に示す方法で検討します。なお、それぞれの方法の特徴については、最後で簡単に議論します

まず、A法に偏った検査値が14人分、B法に偏った検査値が3人分あることに注目し、二項分布を前提とする検定を行います。

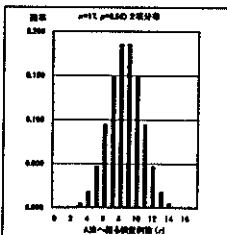
二項分布による検定をイメージで説明します。ここに17個の金平糖があり、姉弟でランダムに分け合うことになりました。姉の提案で、金平糖を筒に入れて三角錐の積み木の手前に落ちたら姉の取り分とすることにしました。結果は、姉には14個、弟には3個だけだったので、弟は「筒の置き方がおかしいよ」と喧嘩になってしまいました。



対称性の検定(二項検定)

$n=17, p=0.5$ の理論二項分布

| A法への 偏りの数(r) | 確率 | 累積確率 | 上側確率 |
|-----------------|----------|----------|----------|
| 0 | 0.000006 | 0.000006 | 1.000000 |
| 1 | 0.000130 | 0.000137 | 0.999863 |
| 2 | 0.001038 | 0.001175 | 0.998963 |
| 3 | 0.005188 | 0.006363 | 0.993637 |
| 4 | 0.018158 | 0.024521 | 0.975479 |
| 5 | 0.047211 | 0.071732 | 0.928298 |
| 6 | 0.094421 | 0.166153 | 0.833847 |
| 7 | 0.148378 | 0.314529 | 0.685471 |
| 8 | 0.185471 | 0.500000 | 0.500000 |
| 9 | 0.185471 | 0.685471 | 0.314529 |
| 10 | 0.148378 | 0.833847 | 0.166153 |
| 11 | 0.094421 | 0.928298 | 0.071732 |
| 12 | 0.047211 | 0.975479 | 0.024521 |
| 13 | 0.018158 | 0.993637 | 0.006363 |
| 14 | 0.005188 | 0.998963 | 0.001175 |
| 15 | 0.001038 | 0.999863 | 0.000137 |
| 16 | 0.000130 | 0.999994 | 0.000006 |
| 17 | 0.000006 | 1.000000 | 0.000000 |



Excelの関数 BINOMDIST BINOMINVT
(r,17,0.5,false) (r,17,0.5,true)

正副標準 $Z = \frac{(r - np) - 1/2}{\sqrt{npq}} = 2.425356$
上側標準 = 0.007647

筒の位置が真ん中だったら、姉側に金平糖が落ちる確率はいつも0.5のはずです。n個のうちr個が姉側(A法)に偏る確率は、(n=17,p=0.5)の二項分布としてExcelを使って簡単に求められます。A法へ14個以上偏る確率は0.0064(正規近似で0.0076)であり、偏りは高度に有意と判定されます。

統計的アプローチ

- 1) 対角線に対する対称性の検定
 - (1) 二項分布を前提とした検定
 - (2) スコアを使った検定
対角からの外れをスコア化
-4, -3, -2, -1, 0, 1, 2, 3, 4
- 2) 対応あるデータの検定

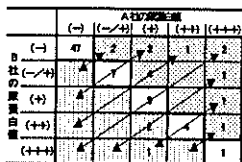
対角線から離れる程度をスコアで重みづけた検定も可能です(対角線から、A法側にある試料をプラス、B法側にある試料をマイナスとします)。

A法とB法とが同等であるという帰無仮説のもとでは、両法の平均の差は0となるはずですが、これ以降は、すべて両側検定で行うことにします。

対称性の検定(スコア検定)

対角からの外れをスコア化

-4, -3, -2, -1, 0, 1, 2, 3, 4



| | |
|----------|--------|
| N | 85 |
| Mean | 0.2 |
| S. D. | 0.6 |
| t | 2.9047 |
| P | 0.0047 |
| Test | ** |
| Wilcoxon | 132.0 |
| Z | 2.7767 |
| 修正 Z | 2.7517 |
| P | 0.0055 |
| Test | ** |

85人分の尿測定値を使って、それぞれt-検定と Wilcoxon 検定をしました。

P値をみると、t-検定; 0.0047

Wilcoxon 検定; 0.0055

ですから、両検定ともにA法への偏りは高度に有意であることを示します。

統計的アプローチ

- 1) 対角線に対する対称性の検定
 - (1) 二項分布を前提とした検定
 - (2) スコアを使った検定
- 2) 対応あるデータの検定

会場の皆さんの中には、「対応のある検定」をまず試みられる方も多いのではないのでしょうか。

対応あるデータの検定

(-), (-/), (+), (++) , (+++)

↓ ↓ ↓ ↓ ↓
1, 2, 3, 4, 5

のランク付けを行って、
A法とB法の間で対応ある
両側検定をする。

| Pair | |
|----------|---------|
| N | 85 |
| Mean | -0.3 |
| S. D. | 0.9 |
| t | -2.7570 |
| P | 0.0072 |
| Test | ** |
| Wilcoxon | 21.5 |
| Z | -2.6531 |
| 修正 Z | -2.6290 |
| P | 0.0080 |
| Test | ** |

対応のあるデータに対する(ピンク本の1群の検定を参照), t-検定とWilcoxon 検定を行いました。

P値をみると, t-検定; 0.0072

Wilcoxon 検定; 0.0080

ですから, 両検定ともにA法への偏りは高度に有意であることを示します。

統計的結論

- A社の尿蛋白濃度が, 相対的に高く出ていることは, 間違いないといえる。
- どちらがおかしいかは, 定量測定値と比べなければ, 判定できない。

つまり, A社の尿測定値が高く出ていることは統計的に間違いないと言っているでしょう。

しかし, これだけのデータで一方がおかしいと判定することはできません。決着をつけるには, 十分な精度をもつ定量測定値をスタンダードとして比較する必要があります。

尿蛋白値の変動要因

1. 試料と経過時間
尿pHの自然上昇
施設間のsample移動順序
2. 定量再現性
3. 測定方法による違い
TBPB量(A:0.24mg, B:0.37mg)
測定装置
妨害物の混入(無機塩, 洗剤, 殺菌防腐剤)

A社とB社の尿蛋白測定値が違ってくる要因を思いつくまま列挙しました。この回答は, 原因を探ることが目的ではないので, これ以上の検討はしませんでした。

相関係数や回帰直線を求めることに意味があるだろうか?

- 半定量法では, 偏りが大きな問題
- 健康診断の性格上, マイナスが殆ど
- マイナスは, 偏った方法でもマイナス
- 相関や回帰はこれらに強く影響される
- つまり, 相関や回帰係数を調べることは, 方法論として誤りである

ここで, 病院の多数の方が賛同された相関係数や回帰直線による統計解析の妥当性に話を戻します。

これらの方法は定量測定値には一般に適切に適用できます。しかし, 今回のような状況のデータに対しては, 手法としての妥当性を失うために, 適用してはならないのです。

どういふ行動をとるべきか

- 1) より良い検査方法を選択する
- 2) この方法に統一する

石原先生の最後の質問に対する回答です。そもそも一つの病院の関連部門で二つの異なる検査紙を使っていることに問題があります。互いに協力し合って、より良いと判断される検査紙に統一するのが最も適切ではないでしょうか。従来の意見の違いを越えて、最良の検査紙を選択されることを期待します。

おわりに

1. 石原さんの質問にみる統計解析の光と影

石原先生の質問を初めて世話人会で伺ったとき、統計解析の持つ影の部分がまだまだ生き続けているのを感じました。統計手法はデータ解析の強力な武器ですが、それが形式的に適用されるときは害毒もまたいっそう強力です。わたくしも大学を出て小さな毒性試験の受託会社に勤めはじめた頃、大手製薬企業の方からの「測定データの処理はまず棄却検定をして不自然なデータを取り除いてからにするのが常識です」という意義明快な意見に翻弄させられた経験があります。大昔の製薬企業には若者と弱者いじめの上手な統計ツワモノが徘徊していたものです。そんな環境のなかで、わたくしが今まで大きな間違いをせずに済んだとすると、正しいことは正しいと主張することを指導してくれた上司(奈良間功・摂南大学教授)、吉村先生、そして研究会の仲間達のお陰ではないかと思えます。その当時から比べると随分とすっきりとした良い時代になったという印象をもっておりましたところ、今回の質問を受けてちょっと違った社会変化を感じました。それは、研究指導部長という肩書きを持っておられる石原先生が統計解析を突きつけられて慥然としている構図から受けるものです。もしかすると最近、年輩者と強者いじめの上手な統計ワカモノが徘徊するようになってきたのでしょうか？ そうであれば非常に良いことであり、是非、この研究会に来て持論をぶつけていただきたいと心から願うのですが。

2. 森野さんへの回答

定例会後、石原産業中央研究所の森野さんから質問が事務局に届きました。その内容を要約すると次の通りです。

1. 対応のある Wilcoxon 検定(符号付き順位和検定)のP値が違う
2. 対角からのスコアを1標本t検定すると、 $t=2.757$, $P=0.0072$ であった
3. 対応のある Wilcoxon 検定では、差が0となる対が多いときに妥当でなくなるとい

われている。これに対して、1標本t-検定は0となる対も含めて評価するので結論の信頼性が高いと考えてよいか。

まず、第1番目の指摘に関しては、わたしがソフトの使い方を誤ったため、今回の回答では修正してあります。

第2番目に関しては、対角セルに対して対称となる距離でスコア一化し、両法に差がないとする帰無仮説(=全試料のスコア平均が0となる帰無仮説)の下で、t-検定を行ったものです。今回の回答では Wilcoxon 検定の結果も比較のため付けておきました。もちろん、この場合にはデータが対になっているという情報は使っていませんので、森野さんのいうP値とはすこし違ってきます。

最後の質問については、対応のある Wilcoxon 検定は、符号付き順位和検定と呼ばれるように、2項検定(符号検定)の拡張であり、0を含まない正負のデータが0を挟んで対称となることを前提にしています。森野さんは、差が0となるデータを使っていないので、その結論は信頼できないとの疑いをお持ちのようです。その一つの理由として、かりに差がない対のデータが200人分加わったとしても、2項検定や Wilcoxon 検定はP値に変化がないが、t-検定ではこの影響を評価(P値が増加する)できることを指摘されておられます。この例え話は、対応のある2群の検定といわれる t-検定と Wilcoxon 検定が全く同じものを検定しているわけではないことを注意させるために、しばしば使われています。しかし、現実の状況では、測定試料数を増加させれば、対角セル内の試料数が増加すると同時に、それ以外のセルでも試料数が増加しますので、2項検定や Wilcoxon 検定は妥当性を失いません。

なお、石原先生に追加調査していただいたのですが、A法とB法との違い原因を突き止めることができませんでした。両法の測定結果にどのような関係があるかがはっきりしない現時点では、強い統計的前提を解析に持ち込むべきではないというのが私の結論です。この観点からして、もっとも確実な証拠を与える(保守的な)解析は、対角セルに対する対称性のみを検討することですので、データの2項分布による検定か、Wilcoxon 検定(符号付き順位和検定)がより適切であったと考えています。この意味では、2項分布による片側検定P値である 0.0064 だけで検定の説明を打ち切り、質問の趣旨に集中すべきだったかもしれません。しかし、もし両側検定で考えるべき内容であったならP値は 0.0127となり、かりに13人分の試料しかなかったのであれば両側P値は 0.0494となってようやく5%有意に過ぎないところです。そこで、対称性の仮定がより強いスコア化による両側 t-検定($P=0.0047$)と、2項分布の拡張である符号付き順位和検定(Wilcoxon 検定, $P=0.0080$)を補足することにしました。いずれにしても測定の偏りは明らかなのですが、わずかな情報を付け加えるだけで検定効率を上げることができることを示したかったわけです。

以上

Medical Writing

株式会社Biostatistical Research
古川敏仁

1

きっかけ

- 1996年2月「CDM」の翻訳時
- Medical Writer ?
- 7人中6人は知らず
- 1人のみ、「そうゆう職業があるそうだ」

- 「クリニカル・データ・マネジメント」 発売中

2

きっかけ2

- 1998年6月 ASCO(ロス)下記の書籍発見
「Medical Writing」

a Prescription for Clarity

A self-help guide to clearer medical English
Neville W. Goodman and Martin B. Edwards

Cambridge University Press,
1991 First published, 1997 Second Edition

3

[Medical Writing]を読むと

- 英語力が付きます 正しい単語がわかる
- 合理的(当たり前)のことが裏付けされます

- 時間に余裕があり、英語の論文を書く(読む)人には是非勧めます。
- 製薬企業の方 手っ取り早い本ではないので、内容を楽しむ人以外には.. でも是非読むべきです。

4

さらに

- 1998年7月4日 SASユーザー会
「統括報告書」

- 大橋先生らのコメント

5

引用1

良いメディカルライティングのための
単純な格言があるとすれば、それは、

「ほとんどどのような場合でも、
短い単語ほど、短い文章ほど良い」

6

引例 1

- 下地となる知識を同程度持つ者が、文章から正確な意味をくみ取れないとすれば、それは文章作成上に問題がある。
- MWは同程度の知識を持つ者に、正確に情報を伝える技術

7

引例 2

- 明瞭で正確な考え方と、明瞭で正確な文章とは同じことである
- おまけ
明瞭で正確な考え方と、明瞭で正確な発表とは同じことである

8

引例 3

だめなMedical Documentが氾濫しているからといって、一概に書き手(経験の浅い)が悪いとはいえない...

読み手が良い経験を重ねないと、悪い書き方は伝染する。

その分野固有の書き方が、効率的な場合もある。

9

Medical Writing とは

- 医師、医学分野の研究者、その他の者が
- 医学関連事項を文書化する技術
- 専門家 Medical Writer

10

Medical Writer

- 医学論文の作成
- 臨床試験統括報告書、Expert Reportなどの作成
- 申請書類、その他の文書作成

11

調査1: 発表者の最初の出会い

- 1986年(12年前)
- 「Medical Writing: Writing for Biomedical journals」 Dr. Robert E. Bjork著
「実例による医学英語論文の書き方」
引地岳雄訳 メジカルビュー社

12

Dr. Robert E. Bjork

- 1971年 Pomona College卒業
- 1979年 University of California at L.A
Adjunct Lecturer(歯学・医学・公衆衛生学
のWriting担当)

欧米では遅くとも1960年初頭にはWritingが
専門分野として確立していたらしい。

13

ICHでは

- E3: STRUCTURE AND CONTENT OF
CLINICAL STUDY REPORTS
- E6:GUIDELINE FOR GOOD CLINICAL
PRACTICE E9,E8 etc
- Medical Writer,Writerの記載なし
VS
- Statistician,Investigator 多数の記載

14



September 24-25, 1998
Hotel Sofitel Paris Forum Rive Gauche
PARIS, France

Medical/Technical Writing & Associated Technologies

PROGRAMME
CODE:49735255

Betty R. Kahnerl, Ph.D.
Web-Apex Research, USA

Suzanne Wodderkopp, MSc (Pharm)
Novo Nordisk AS, Denmark

Workshop Objectives

Every pharmaceutical company has to produce scientific reports and summary documents for regulatory agencies. Some companies have writers who perform this task, other have published or clinical scientists, statisticians, regulatory affairs and or medical research groups who prepare documents. Some companies have very structured and efficient review processes, whereas in others, it may take longer to prepare a document than it did to perform the original study. In the changing world of the pharmaceutical industry, each company needs to develop the most efficient processes and processes for preparation and review of high quality documents.

In this Workshop, speakers and agency reviewers will address issues relating to preparing documents in today's environment of global drug development and technical submission. Sessions will cover management of the writing task and technical perspectives on document preparation. Individual participants will benefit of major aspects of document preparation and review.

Ample opportunity will be provided for attendees to interact intensely with speakers and other attendees with similar interests.

Exhibition

The DIA will provide the opportunity for pharmaceutical industry support organizations to exhibit their products and services at this Workshop. Companies interested in exhibiting should contact:

Dr. Ellen ROTH
DRUG INFORMATION ASSOCIATION
201 Office Center Drive, Suite 130
FISHT WASHINGTON, PA 19024-2211
USA
Tel: +1 215 424 2266
Fax: +1 215 641 1229
E-Mail: dia@diaonline.org

15

Drug Information Journal

- 29巻1号(1995)~32巻1号(1998)
約5000項 約500論文中
- Medical Writing がテーマ1
- 意外に少ない

16

A Team-oriented Approach To The Analysis and Reporting Process

Beth B. Haury, MA Manager, Medical Writing
and Regulatory Documents

Vol.31 pp.665-670 Drug Information Journal

Presented at DIA Workshop "Medical Writing
,October 30-31 1995,Philadelphia

17

Final,Integrated Report作成の 5段階

- Analysis Plan
- Preresults report
- Bullet Outline of results
- Draft Report
- Final Report

Medical Writing を請け負うCROのやり方

18

Analysis Plan(解析計画書)

- 作成: 生物統計家 協力: 医師
- 編集・編纂 Medical Writer ,
Programmer

19

Preresults Report

- 報告書に盛る内容を定義
- 案作成 Medical Writer
- 協力・訂正 Monitor
- 社内規約に基づいて別のWriterがチェック

20

Bullet Outline of Results

- 結果のあらまし、問題点、全ての図表
- 作成: Medical Writer
- 生物統計家: 有意差の意味などの説明
- Programmer: 図表作成の手伝い
- 医師: データの医学的解釈

21

Draft Version of the Report

- Bullet OutlineをもとにDraft報告書を作成
- 生物統計家と医師が下文章を作成
- Medical Writerが編集
- 関係者全員で査読

22

Final Version of the Report

- Writerが会社の意見を取り入れて修正
ただし、最小ではあるが、必ず関連部門の
意見や協力を仰ぐ

23

日本の現状

- 製薬企業
外資系の一部の企業でMedical Writerの
専門職を設けるうごき
- CRO 有りと言えども...・まだ、ない

24

質問

- Pfizerでの現状
欧米では
日本では
- 担当業務は 守備範囲は

25

質問、質問

- Medical Writerを専門化することで
業務効率は上がりますか
- Medical Writerは専門化すべきか

26

質問、質問、質問

- 社内での組織上の位置づけ
地位
給与

27

質問、質問、質問、質問

- 今後、日本でのMedical Writerに関
してあなたはごどう思いますか。
発展、定着するのか
技術的なレベルの向上は必要か
社会的なコンセンサスは
その他

28

引用4

- 良い文章というものを正確に教示すること
は難しいが、
- 悪い文章の例を指摘することは簡単である。
- 悪い点を改善することで、良い文章――

29

日本の実例の問題点

序文 ・・能測定の臨床的意義は不明であり、真の値も不明である

結論 本方法は従来法と比較して測定精度を大幅に向上させ、統計的にも有意であった、ゆえに、臨床的意義が認められた

30

こう直したが、むなし

校正前 臨床的意義が認められた

改訂 本法は従来法と比較して、

- 平均値(正確度)同じ=臨床的意義は不明
- 精度は大幅に向上
- よって従来法よりもbetter
- 従来法認可 ならば 本法は当然認可
- 精度向上により、臨床的意義の解明期待

31

結論

- Medical Writerの専門化は急速に進む
- 対象は全ての申請資料+学術論文
- 専門技術の習得
- 質的な向上

32

[事務局だより]

会報 No. 44がまとまりましたのでお送りします。

78回定例会（99年 4月 3日）での吉村先生の刺激に充ちた講演「『臨床試験のための統計的原則』をめぐる諸問題」の詳細な講演原稿を初め、定例会で何回も取り上げられた最大対比法は特集にいたしました。その他話題のQ & Aや、新しいテーマであるメディカル・ライティングなど、充実した編集ができたのではないかと思います。

定例会は力のこもった発表が続いております。会報も、1号1号出していきたいようにしたいと思います。

（大野）

医薬安全性研究会

会報 No. 44

Bulletin of Japanese Society for Biopharmaceutical Statistics.

1999年 7月10日 発行

定価 1,050円（本体 1,000円）

編集・発行 （株）サイエンティスト社

〒101-0062 東京都千代田区神田駿河台3-2 山崎ビル

TEL. 03(3253)8992

FAX. 03(3255)6847

振替 00180-1-71335

印刷・製本 （株）エポ

医薬安全性研究会 会報

第44号 1999年7月

■ 目 次 ■

* 医薬安全性研究会と関連スケジュール

- 「臨床試験のための統計的原則」をめぐる諸問題
- 成立の経緯からQ & Aまで -
吉村 功 1

- <特集> 最大対比法
- 用量反応関係を評価するときの「対比」の役割
吉村 功 9
- 最大対比法はどんな場面で使えるか？
吉村 功 12
- 最大対比法の性能評価
浜田知久馬 15

- [Q & A] 尿中蛋白濃度の測定結果が2種の尿検査紙で違ったら？
Q : 石原信夫 / A : 佐野正樹 27
- Medical Writing
古川敏仁 35

- 事務局だより 41

編 集

医薬安全性研究会

定価 1,050円