

# 毒性試験データの解析：無影響量決定問題について

九州大学大学院\* 柳 川 堯

(受付 1997年9月16日；改訂 1997年10月21日)

## 要 旨

慢性毒性をターゲットとしてラットやハムスターを用いて行われる反復投与毒性試験に焦点を当て、しかも特に反応がカテゴリカルデータとして観測される場合の無影響量 (NOAEL = No Observed Adverse Effect Level) 決定問題に話題を制限して、まず毒性データの統計解析になぜ従来の技法が適用できないか、適用してはならないか、をクローズアップし毒性データ解析の本質の一端を明らかにする。特に、反復投与毒性試験では試験に供される動物数は一般にごく限られた数であること、従って無影響量決定に多重性を調整する検定法を適用すると無影響量がゆゆしく過大評価されることになり、毒性や副作用の早期検出を目的とする毒性試験の原理に合致しないことを指摘する。次に、この原理を満たす方法として最近著者等が開発した (Yanagawa et al. (1997)) 赤池情報量基準 (AIC) を用いる新しい技法について述べ、最後に、SAS (SAS Institute Inc.) の PROC MULTTEST で Fisher の直接法 (片側検定) を多重性調整した resampling 技法と新しい技法を実際のデータに適用し、実証的吟味を行う。その結果、新しい技法が毒性試験の原理を満たす方法である事が示唆される。

キーワード：AIC, Dunnett 型検定, 反復投与毒性試験, MULTTEST, 多重比較, 用量・反応関係。

## 1. 序

毒性試験は、環境化学物質、食品添加剤等の化学化合物による曝露、あるいは放射線被爆がヒトの健康にどのようなリスクを与えるかについて評価を行うことを最終の目的としている。何らかの毒性物質に不幸にして被曝してしまったヒトの集団を除けば、直接ヒトを対象とする人体試験は出来ない。そのため毒性試験では、サルやイヌ等の大型の動物、ラットやハムスター等のげっし類、サルモネラ等のバクテリア類、ラット等の培養細胞や小核、あるいはヒトの培養リンパ球等を用いて試験を行い、得られたデータに基づいて毒性の評価を行う。ターゲットとされる毒性は、直ちに反応が出現する急性毒性、長期間微量曝露に関する慢性毒性、催奇性など次世代に影響を与える生殖発生毒性、ガン原性、変異原性などがある。

毒性試験データは上記のような、使用される資料、およびターゲットとされる毒性の組み合わせによって多様多岐にわたって出現し、多くの場合、教科書的な統計技法は適用できない。試験の様相を深く理解し、その特徴をとらえ、かつ個々のデータに即した新しい解析法が必要

\* 数理学研究科：〒812-8581 福岡市東区箱崎6-10-1。

† プロシーディングの残部少しあり、希望者は上記にご連絡下さい。

である。わが国では、医薬安全性研究会がこれらについて研究を蓄積してきた。1991年5月には同研究会が中心となって東大山上会館で国際会議「毒性学における統計的方法」<sup>†</sup>（前頁脚注参照）が開催され、最先端の生物統計学者によって問題点がクローズアップされるとともに新しい技法、考え方が紹介され、また、テキスト「毒性試験データの統計解析」（吉村、大橋（1992））も出版されている。

しかしながら、毒性試験のデータの解析に関してはまだ研究されるべき多くの問題がある。本稿では、慢性毒性をターゲットとしてラットやハムスターを用いて行われる反復投与毒性試験に焦点を当て、しかも特に反応がカテゴリカルデータとして観測される場合の無影響量（NOAEL=No Observed Adverse Effect Level）決定問題に話題を制限する。このことによって、まず毒性データの統計解析になぜ従来の技法が適用できないか、適用してはならないかをクローズアップし、毒性データ解析の本質の一端を明らかにしたい。次に最近著者等が開発した（Yanagawa, et al. (1997)）赤池情報量基準（AIC）を用いる新しい技法について述べ、最後に、従来の技法と新しい技法を実際のデータに適用し適用結果について実証的な考察を行う。

## 2. 考察するデータの例と目的

まず、本小論で対象とするデータがどのようなものであるかを例によって示し、この例に即して解析の目的を明らかにする。

表1のデータは、Yellow AB (1-Phenylazo-2-naphthylamine) をそれぞれ0, 500, 1000, 2500 ppm 混ぜた食餌でラットを長期間飼育したあと屠殺剖検し、顕微鏡によって肝臓に出現した反応を観察した記録である。反応はカテゴリ  $C_0$  (None),  $C_1$  (Trace),  $C_2$  (Very Slight),  $C_3$  (Slight),  $C_4$  (Slight to Moderate),  $C_5$  (Moderate) で観測されており、表中の数字は観測された動物数を表す。カテゴリは階層的 (hierarchical) になっており  $C_0, C_1, C_2, C_3, C_4, C_5$  の順で重篤さが増加する。

表1は用量(ppm)が増加するにつれて毒性反応が増加している様子を示唆している。問題は、毒性反応がどの用量 (ppm) から発現しているかを見極めることである。毒性反応発現用量 (ppm) の一つ手前の用量のことを無影響量 (以下 NOAEL=No Observed Adverse Effect Level と略記する) という。解析の目的は、表1のデータから NOAEL を定めることである。以下では、用量を一般的に記号  $d_i, i=0, 1, \dots, k$  ( $d_0(=0) < d_1 < \dots < d_k$ ) で表す。

## 3. 通常解析とその問題点

通常解析では、表1の様なデータはカテゴリ  $C_1, C_2, C_3, C_4, C_5$  を適当に合併し、毒性

表1. Degree of Microscopic Liver Damage. Source of data: Hansen et al. (1963).

| Color fed | Diet (ppm) | No. rats fed | Livers micr. sectioned | Severity |       |             |        |                    |          |
|-----------|------------|--------------|------------------------|----------|-------|-------------|--------|--------------------|----------|
|           |            |              |                        | None     | Trace | Very slight | Slight | Slight to moderate | Moderate |
| None      | 0          | 24           | 19                     | 3        | 13    | 3           | 0      | 0                  | 0        |
| Yellow AB | 500        | 24           | 18                     | 7        | 9     | 2           | 0      | 0                  | 0        |
|           | 1000       | 24           | 16                     | 5        | 6     | 4           | 1      | 0                  | 0        |
|           | 2500       | 48           | 25                     | 0        | 4     | 12          | 5      | 3                  | 1        |
|           |            |              |                        |          |       |             |        |                    |          |

反応を毒性反応「あり」、「なし」の二値反応にまとめ直した上で、Steelの多重比較、Shirley-Williamsの逐次検定、あるいはFisherの直接法をresamplingで多重性を調整する(Westfall and Young (1989))などして、統計的検定の考えでNOAELを決定する。

私どもは、このようなNOAELの決定には問題があると考えている。まず、データを二値反応にまとめ直すことから生じる情報損失がある。表1では、データがまばらで観測値がゼロのセルが多い。このような時、カテゴリーを合併するのが良いと言われている。また、カテゴリーに順序があるときは、カテゴリーを合併してもその影響は少ないという報告さえある。しかしながら、表1をより詳しく良く見ると、観測値ゼロのセルが大きな役割を演じて、用量(ppm)が増加するにつれて毒性反応が増加している様子が示唆されていることに気付く。この情報は有効に利用すべきであろう。さらにこの二値化には、どの反応から先のカテゴリーを毒性反応「あり」とするかについての恣意性もある。

さて、データが二値反応にまとめ直されたとしよう。次の問題点は統計的検定の適用である。これは毒性試験データの本質にかかわる問題点である。目的はAdverse Effect (毒性反応, 副作用)の検出であって、Effect (効果)の検出ではない。Adverse effectの検出は、Effectの検出と本質的に異なることを強調したい。最近、「Adverse Effectsの検出システム」がいくつか提案されているが、これらのシステムは既成の検定を組み合わせただけであって、Adverse effectとEffectの本質的な相違が把握されていない。改めて述べるまでもなく統計的検定は第1種の過誤のみを厳しく押さえる技法である。Effectの検出では、「Effectが無いものを世に出すことを厳しく押さえる」という意味でこの考え方は理にかなっている。しかしながらこの検定をAdverse effectの検出に適用すれば、「Adverse effectがないものを世に出すことを厳しく押さえる」ということになり、理に合わない。毒性試験は「Adverse effectがあるものを世に出すことを厳しくチェックする」という思想の下で行われているからである。つまり、毒性反応を対象にする場合、第二種の過誤を厳しく押さえ、毒性反応出現にいち早く警告の「FLAG」を立てる方法であって欲しい。にもかかわらず統計的検定はこの目的を満たさない。特に、毒性試験の場合、試験に使用される動物数は、極めて限られた少数であって、第二種の過誤が野放しになる。このことはNOAELの値がゆゆしく過大に定められることにほかならない。

さらに問題なのは、検定の多重性の調整である。通常、用量 $d_0$ 群 vs. 用量 $d_1$ 群、 $d_0$ 群 vs.  $d_2$ 群、…と検定を繰り返し行い、有意になった用量の一つ手前の用量がNOAELと定められるが、検定を繰り返し行うことから多重性の調整が行われる。端的に言えばSteelの多重比較、あるいはSASのPROC MULTTESTを用いてresamplingによってFisherの直接法の多重性調整を行う技法が適用される。多重性を調整しない検定の棄却限界点とSteel多重比較の棄却限界点を比べれば一目瞭然のように、多重性を調整すると毒性を見落とす危険が増大する。resampling法によって多重性を調整する方法(Westfall and Young (1989))についても同じことである。多重性の調整をすべきでないという指摘は、永田・吉田(1997)にも見られるが、ではどうすればよいか、という具体的な提案が与えられていないのは残念である。

最後に、本稿では発ガン性を除くいわゆる一般毒性反応を対象としていることに注意しておきたい。一般毒性反応に関するNOAEL決定問題ではその背後にthresholdが存在すること、さらに用量がそのthreshold値を越すと毒性反応が発現し、用量を増やすにつれて毒性反応が増加するという前提がおかれているが、これを定式化して導かれるShirley-Williams型の検定を行うと、特にデータ数が小さいときは、離散型検定の特徴が端的に現れ、容易にmisleadingが起り得る。この事に関する詳細はYanagawa et al. (1994)を参照されたい。

#### 4. 「FLAG」法

上のことから示唆されるように、多重性を調整する従来の検定の枠組みで NOAEL を決定するには、第2種の過誤を厳しく押さえればよい。このことを行うことは不可能ではない。例えば、いま3群があるとして、用量を増やすにつれて毒性反応が増加するという前提をおき、NOAELの推定値を $\hat{\delta}$ とするとNOAEL決定問題に関する重要な第2種の過誤は

$$\Pr(\hat{\delta} \geq d_i, \text{ or } d_2 | \text{NOAEL} = d_0)$$

および

$$\Pr(\hat{\delta} \geq d_2 | \text{NOAEL} = d_i)$$

で与えられる。したがってこれらの確率を厳しく押さえるような棄却限界点を求めればよい。Sakata and Yanagawa (1998) はマルコフチェーンモンテカルロ法の適用を提案している。

以下では、最近 Yanagawa et al. (1997) が統計的検定のパラダイムから離れて新しく開発した、AICを用いる「FLAG」法について、詳しい技法やその理論的展開は原著にゆずり、適用例とその解釈、および従来の多重比較の結果との比較に重点をおいて述べる。

##### 4.1 用量・反応関係

一般に、用量・反応関係は反応が二値変数の場合だけにしか定義されていない。表1のようなデータからNOAELを決定するには、多値反応データに対して用量・反応関係をどのように導入するかが第一のポイントとなる。

表1を一般化して、用量のレベルが $k+1$ 、反応カテゴリーの個数が $b+1$ であるとしておき、用量 $d_i$ 、反応カテゴリー $C_j$ のセル確率を $p_{ij}$ 、観測値を $X_{ij}$ で表す。用量・反応関係を次式で一般化する。

$$f_j(d_i) = \frac{p_{ij+1} + \dots + p_{ib}}{p_{ij} + \dots + p_{ib}},$$

$$i=0, 1, \dots, k; j=0, 1, \dots, b-1.$$

$1-f_j(d_i)$ は反応がカテゴリー $C_j$ 又はそれ以上に悪いカテゴリーに属するという条件の下でカテゴリー $C_j$ に属する条件付きリスクを表す。特に $b=1$ のときは、従来の二値反応の場合の用量・反応関係と一致する。また、 $b=2$ のときは $f_0(d_i)$ はRyan (1992) の overall risk と一致する。 $b$ 個の用量・反応関係が導入されていることに注意しよう。この用量・反応関係に対する自然なパラメータ化として

$$f_j(d_i) = F(\theta_{ij}), j=0, 1, \dots, b-1$$

を考える。ただし、 $F$ は所与の分布関数である。以下の応用では $F(x) = 1 - \exp(-x)$ と定めた。この $F$ でモデル化された上の用量・反応関係をYOモデルとよぶ。 $F(x) = e^x / (1 + e^x)$ と定めればContinuation-Ratio Logitモデル(McCullagh (1980))を得る。このパラメータ化から次式を得る。

$$p_{i0} = 1 - F(\theta_{i0})$$

$$p_{ij} = \prod_{u=0}^{j-1} F(\theta_{iu})(1 - F(\theta_{ij}))$$

$$j=1, 2, \dots, b-1.$$

##### 4.2 尤度関数の分解

$X_i = (X_{i0}, X_{i1}, \dots, X_{ib})$ が多項分布( $n_i: p_{i0}, p_{i1}, \dots, p_{ib}$ )に従うこと( $i=0, 1, \dots, k$ )、 $X_0, X_1,$

表2. NOAEL (多項反応) 推測のための  $(k+1) \times 2$  表.

| Dose level | Cell Count |                         |                       | Cell Probability     |                  |          |
|------------|------------|-------------------------|-----------------------|----------------------|------------------|----------|
|            | $C_j$      | $C_{j+1} - C_b$         | Total                 | $C_j$                | $C_{j+1} - C_b$  | Total    |
| $d_0$      | $x_{0j}$   | $\sum_{h=j+1}^b x_{0h}$ | $\sum_{h=j}^b x_{0h}$ | $1 - F(\theta_{0j})$ | $F(\theta_{0j})$ | 1        |
| $d_1$      | $x_{1j}$   | $\sum_{h=j+1}^b x_{1h}$ | $\sum_{h=j}^b x_{1h}$ | $1 - F(\theta_{1j})$ | $F(\theta_{1j})$ | 1        |
| $\vdots$   | $\vdots$   | $\vdots$                | $\vdots$              | $\vdots$             | $\vdots$         | $\vdots$ |
| $d_k$      | $x_{kj}$   | $\sum_{h=j+1}^b x_{kh}$ | $\sum_{h=j}^b x_{kh}$ | $1 - F(\theta_{kj})$ | $F(\theta_{kj})$ | 1        |

...,  $X_k$  が互いに独立であることを仮定すると, 観測値  $\{x_{ij}\}$  が得られたときの尤度関数は次式で与えられる.

$$L \propto \prod_{i=0}^k \prod_{j=0}^b d_{ij}^{x_{ij}}$$

ここで,  $\propto$  はパラメータに依存しない定数項を除く項に比例することを表す記号である. これに(1)式を代入して整理すると, 定数項を除く対数尤度関数は次のように表現できることが分かる.

$$l(\theta) = \sum_{j=0}^{b-1} l_j(\theta_j),$$

ここで,  $\theta_j = \{\theta_{0j}, \theta_{1j}, \dots, \theta_{kj}\}'$  は,  $\theta = \{\theta_{ij}\}$  の  $(j+1)$  番目の列を表し, また

$$l_j(\theta_j) = \sum_{i=0}^k x_{ij} \log[1 - F(\theta_{ij})] + \sum_{i=0}^k \log F(\theta_{ij}) \sum_{h=j+1}^b x_{ih}$$

である. この表現より,  $\theta_j$  の統計的推測は, 表2の様な  $(k+1) \times 2$  表に基づいて個別に行えばよいことが分かる. 上で導入した用量・反応関係のうまみはこの1点に集約されている. つまり, 表1からNOAELを定めるには表2のような  $(k+1) \times 2$  表を  $b$  個作成し, 各表からNOAELを決定すればよい.

### 4.3 AICの適用

$(k+1) \times 2$  表からAICを用いてNOAELを決定する方法について述べる. ここでは簡単のため  $j=1$  の場合について考えることとし,  $x_i = x_{i1}$ ,  $m_i = \sum_{h=1}^b x_{ih}$ ,  $p_i = 1 - F(\theta_{i1})$  で表す. すでに注意したようにNOAELを決定する問題ではパラメータ間に順序制約

$$p_0 \leq p_1 \leq \dots \leq p_k$$

が前提とされていることに注意しよう. 説明を簡単にするため, 以下では  $k=2$  とする. 考え方の基本は次の3つのモデル  $H_1, H_2, H_3$  を立てAICを用いてモデル選択を行い,  $H_1$  が選択されれば  $\text{NOAEL} = d_2$ ,  $H_2$  が選択されれば  $\text{NOAEL} = d_1$ ,  $H_3$  が選択されれば  $\text{NOAEL} = d_0$  と推定することである.

$$H_1: p_0 = p_1 = p_2$$

$$H_2: p_0 = p_1 < p_2$$

$$H_3: p_0 < p_1 < p_2$$

しかしながら, パラメータ間に順序制約が設けられており, 最尤推定値がこの順序制約を満たせば問題がないが, そうでない場合はAICは妥当性を失い適用できない. そこでまず始めに最尤推定値を吟味し, 順序制約を満たさなければあらかじめこの時点で順序制約の破れに応じてNOAELを決定するというバイパスを設ける. 例えば  $\hat{p}_0 > \hat{p}_1 > \hat{p}_2$  なら  $\text{NOAEL} = d_2$  と推定する手順を入れる. 図1に, この考えに基づくNOAEL決定法を流れ図で与えた. この決定法を

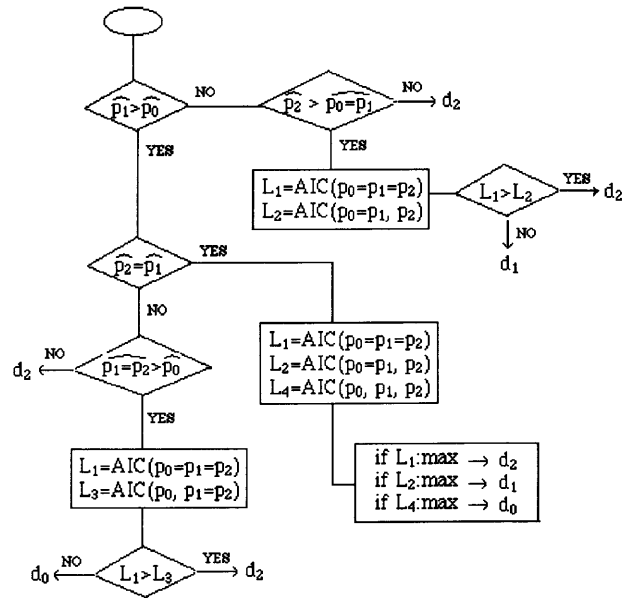


図1. 「FLAG」法の概要.

「FLAG」法とよぶ。

第2の難点は、これも上述したように毒性試験データ、特に表1のような多値反応データでは観測値ゼロのセルが多出することである。観測値ゼロのセルがあれば尤度の計算に支障が生じる。これをバイパスする方法については改めて次節で述べる。

#### 4.4 ゼロセルの対処

隣接した用量のそれぞれにゼロセルがある場合、以下のようにゼロセルの出現パターンに応じて合併、あるいは微量量での置きかえを行う。まず、前節で導入した記号で表すと  $m_a - x_a = m_{a+1} - x_{a+1} = \dots = m_b - x_b = 0$ 、または  $x_c = x_{c+1} = \dots = x_d = 0$  となる場合はゼロセルをもつ用量を合併する。つぎに、そうして得られた新しい表はもはや表2とは同一ではないが、説明の都合上、表2に対して導入した前節の簡略化記号を用いてその他のパターンの調整法について述べる。

$$I = \{i : x_i = m_i, 0 \leq i \leq k\}$$

$$\varepsilon_i = 1 / \left( m_i \sum_{u=0}^k m_u \right), \quad \delta_j = m_j \sum_{u \in I} \varepsilon_u / \sum_{v \notin I} m_v,$$

とおく。Iが空でなければ、Iに属するすべての*i*に対して  $m_i - x_i (=0)$  を  $\varepsilon_i$ 、 $x_i (=m_i)$  を  $x_i - \varepsilon_i$  で置きかえる。この調整によって行和は不変である。次に列和が不変になるようにIに属さないすべての*j*について  $x_j$  を  $x_j + \delta_j$ 、かつ  $m_j - x_j$  を  $m_j - x_j - \delta_j$  で置き換える。この調整法の背後にある原理と妥当性については Yanagawa et al. (1997) を参照されたい。

#### 5. 計算プログラムと適用例

「FLAG」法によるNOAEL決定のためのWindow 95上で作動するソフトを菊池泰樹氏(佐世保高専)が開発しインターネット上のSTATLIBに登録手続き中である。このソフトは、実

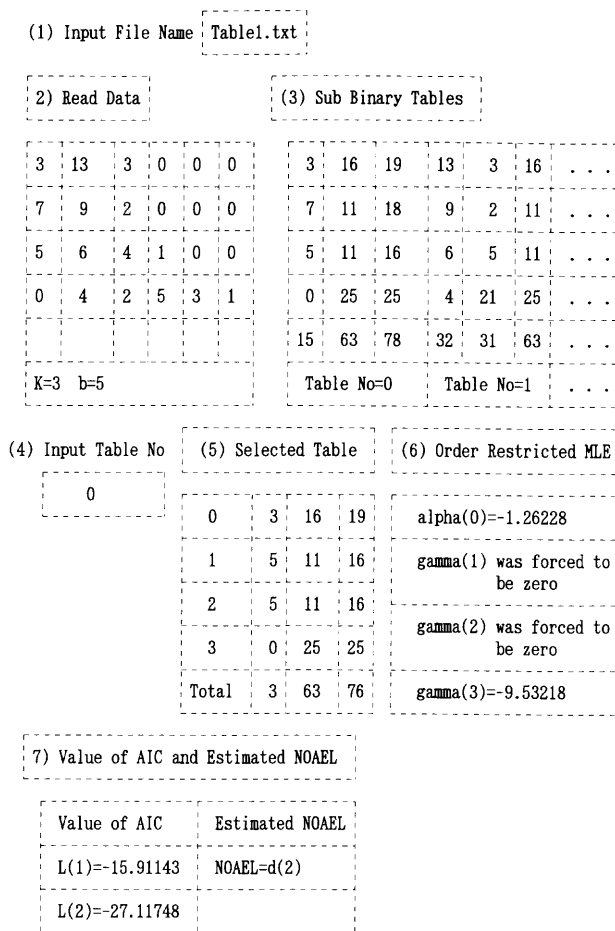


図2. プログラム NOAELAIC.EXE の実行結果.

行用プログラム NOAELAIC.EXE と検証用のデータ test.dat を納めている。

表1で与えられた Yellow AB の毒性試験データにこのソフトを適用して計算した結果を以下に示す。なお、このソフトでは、YO モデルによる解析結果だけが与えられる。

実行プログラムは NOAELAIC.EXE である。まず、test.dat を開きテストデータを確認するとともに、同一のディスク上に Table 1.txt という名のファイル準備し、test.dat と同じ形式でデータをインプットしておく。次に NOAELAIC.EXE をクリックすると図2のような画面が開く。そこで、(1) Input File Name の欄に Table 1.txt とインプットし、(2) Read Data をクリックするとデータが読み込まれる。次に(3) Sub Binary Table をクリックすると図2に見られるような周辺和つきの5枚の4×2表が作成される。さらに(4) Input Table No に、どの Sub Binary Table から NOAEL を求めるかを指定するため、(3)の各表の下に記された表の番号(例えば Table 1 なら1)をインプットし、(6) Order Restricted MLE をクリックし、続いて(7) Value of AIC and Estimated NOAEL をクリックすると、最尤推定量 (MLE) に関する数値結果と AIC の値、及び NOAEL の推定値が提示される。

表3. 表1から作成される4枚の4×2表.

| 用量    | Sub Table 0 |         |    | Sub Table 1 |         |    |
|-------|-------------|---------|----|-------------|---------|----|
|       | None        | T以上     | 計  | T           | VS以上    | 計  |
| $d_0$ | 3           | 16(.84) | 19 | 13          | 3(.19)  | 16 |
| $d_1$ | 7           | 11(.61) | 18 | 9           | 2(.18)  | 11 |
| $d_2$ | 5           | 11(.69) | 16 | 6           | 5(.45)  | 11 |
| $d_3$ | 0           | 25(1)   | 25 | 4           | 21(.84) | 25 |

| 用量    | Sub Table 2 |        |    | Sub Table 3 |        |   |
|-------|-------------|--------|----|-------------|--------|---|
|       | VS          | S以上    | 計  | S           | SM以上   | 計 |
| $d_0$ | 3           | 0(0)   | 3  | 0           | 0(-)   | 0 |
| $d_1$ | 2           | 0(0)   | 2  | 0           | 0(-)   | 0 |
| $d_2$ | 4           | 1(.20) | 5  | 1           | 0(0)   | 1 |
| $d_3$ | 12          | 9(.43) | 21 | 5           | 4(.44) | 9 |

( ): response rate

表4. Yellow AB データ(表1)から推定された NOAEL.

|                   | AIC   | MULT-F (Test Size) |        |        |
|-------------------|-------|--------------------|--------|--------|
|                   |       | (0.01)             | (0.05) | (0.10) |
| None vs. beyond T | $d_2$ | $d_3$              | $d_3$  | $d_2$  |
| T vs. beyond VS   | $d_1$ | $d_2$              | $d_2$  | $d_2$  |
| VS vs. beyond S   | $d_2$ | $d_3$              | $d_3$  | $d_3$  |
| S vs. beyond SM   | $d_3$ | $d_3$              | $d_3$  | $d_3$  |
| SM vs. M          | —     | —                  | —      | —      |

Dose level:  $d_0=0$ ,  $d_1=500$  ppm,  $d_2=1000$  ppm,  $d_3=2500$  ppm.

## 6. 計算結果の吟味

表3は前節SUB TABLES 0, 1, 2, 3を再現したものである。( )内の数字は比率である。表4にこれらの表から推定したNOAELを記載した。また、比較のためFisherの直接法(片側検定)を用いてDunnettタイプの群間比較を行いresampling法によって多重性調整して得られたNOAELの推定値も併せて記載した。この計算には、SAS(SAS Inst Inc.)のPROC MULTTESTを用いた。後者を以下ではMULT-F検定とよぶことにする。SUB TABLE 0について、検定の水準を10%に定めると「FLAG」法とMULT-F法はともに69%から100%への比率のジャンプを摘発しNOAEL= $d_2$ と決定している。しかし、検定の水準を1%や5%に定めるとMULT-F検定はこのジャンプを検出できない。比率が31%も増加するのであるから、ここではNOAEL= $d_2$ とするのが妥当ではないだろうか。次にSUB TABLE 1について、 $d_0$ と $d_1$ の反応率が逆転しているので「FLAG」法はこの調整を行うが、 $d_1$ から $d_2$ への約27%の比率の増加を摘発しNOAEL= $d_1$ と推定している。これに対してMULT-F法は有意水準10%でもこの増加を見逃しNOAEL= $d_2$ と推定している。SUB TABLE 2についても、「FLAG」法は



$d_2$  から  $d_3$  への 23% の比率のジャンプを摘発し  $\text{NOAEL} = d_2$  としているのに対して MULT-F 検定ではこのジャンプを見逃している。標本数が少ないことから MULT-F 検定では検出力が低いことがその理由であろう。最後に SUB TABLE 3 についてであるが、これくらいの比率のジャンプがあれば、「FLAG」法であれば、MULT-F 検定であれ同じ結果を導くのは当然であろう。

## 7. 考 察

一般に、データの大きさが表 1 の程度の場合、有意水準を 10% の MULT-F 検定から定まる NOAEL の値は「FLAG」法から推定される NOAEL 値より大きいか等しい。有意水準をもっと大きく取る、あるいは試験に用いられる動物数を増やすことによって検出力を上げれば MULT-F 検定も十分役に立つ。しかしながら、実際には動物数に制限があるし、適正な有意水準をいかに定めればよいかという問題点がある。このような現実の制約の中でいち早く警告の「FLAG」を立てるという意味で「FLAG」法は有意義であると思われる。また、上では比較のために各 SUB TABLE に MULT-F 検定を適用したが、表 1 のような多重反応データから NOAEL を推定するという方法そのものがこれまで存在しない。

さて、「FLAG」法は、 $b$  個の NOAEL を決定する。そのどれを最終的に NOAEL として決定するかについては、どのカテゴリーから先の反応を Adverse effect とみるか毒性学者の見解にまかせたい。例えば表 4 では最小の NOAEL 値  $= d_1$  で T (Trace) と VS (very slight) 以上のカテゴリーの比較の場合に与えられている。よって、もし、T が無視できる毒性反応ならば  $\text{NOAEL} = d_1$  とすればよい。また、もし T と VS が無視できる反応なら  $\text{NOAEL} = d_2$  とすればよい。さらに T, VS, 及び S (Slight) が無視できる反応なら  $\text{NOAEL} = d_3$  とすればよい。ここで反応カテゴリーの重篤度が増加するに連れて定められる NOAEL 値が増加するという期待にそった結果が得られていることに注意しておきたい。

発ガン性の動物実験データでは、ふつう反応曲線が想定され、与えられた一定の危険値に対して信頼区間を構成しその下限をもって許容量とする。本論文で対象とした一般毒性の場合にも類似した方法が可能であろう。そもそも、与えられた用量の中から NOAEL を選ぶというのは、毒性学での常套的な考えのようであるが、違和感がある。しかしながら、反応曲線を少ないデータから推定するには重大な問題点があり、通常の毒性試験のデータのサイズからは私どもが開発した「FLAG」法に甘んじるほかなさそうに思う。

データが少ないとき AIC の信頼度が問題になるという指摘もある。確かにそのとおりであるが、このとき統計的検定がもつ問題点との比較で考えれば、上の例に見られるように、あるいは私どもが行った多くの数値的検証の結果でもそうであるが、「FLAG」法の方がはるかに良い挙動を示す。なお、「FLAG」法による第 1 種、第 2 種の過誤については Yanagawa et al. (1994) で正当率を評価し、統計的検定のそれと比較しているので参照されたい。

本稿では、慢性毒性をターゲットとする反復投与毒性試験に焦点を当て問題点をクローズアップした。表 1 と同様な表にまとめられる毒性試験データであっても、中には用量がある一定量をこえると反応が減少傾向をしめす、いわゆる down turn data とよばれるものがある。また、越智 (1998) に見られるように、生殖発生毒性試験では、共通の親ラットから生まれる胎児が対象とされることから、個体反応の独立性が成り立たない。これらのデータに対しては本稿の方法は適用できない。新しい方法の開発が望まれる。最後に、毒性試験データの統計解析は、NOAEL 決定問題以外にもいくつもの大きな統計的研究課題があることを指摘しておきたい。

## 参 考 文 献

- Hansen, W. H., Nelson, A. A. and Fitzhugh, O. G. (1963). Chronic toxicity of yellow AB (1-Phenylazo-2-naphthylamine) and yellow OB (1-o-Tolylazo-2-naphthylamine), *Toxicology and Applied Pharmacology*, **5**, 16-35.
- McCullagh, P. (1980). Regression models for ordinal data (with discussion), *J. Roy. Statist. Soc. Ser. B*, **42**, 109-142.
- 永田 靖, 吉田道弘 (1997). 統計的多重比較法の基礎, 統計学会チュートリアルセミナー, 1-28.
- 越智義道 (1998). 超多項変動を持つデータの解析, 統計数理, **46**, 201-221.
- Ryan, L. (1992). Quantitative risk assessment for developmental toxicity, *Biometrics*, **48**, 163-174.
- Sakata, T. and Yanagawa, T. (1998). Determination of the no-observed-adverse-effect-level by controlling over-estimation probability, Statistics Research Report, No. SRR 007-98, The Australian National University, Canberra.
- SAS (1994). *SAS User's Guide: Statistics*, Ver. 6.07, SAS Institute Cary, North Carolina.
- Westfall, P. H. and Young, S. S. (1989). P-value adjustments for multiple tests in multivariate binomial models, *J. Amer. Statist. Assoc.*, **84**, 780-786.
- Yanagawa, T., Kikuchi, Y. and Brown, K. G. (1994). Statistical issues of the no-observed-adverse-effect level in categorical response, *Environmental Health Perspectives*, Supplements, **102**, 95-101.
- Yanagawa, T., Kikuchi, Y. and Brown, K. G. (1997). No-observed-adverse-effect levels in severity data, *J. Amer. Statist. Assoc.*, **92**, 446-454.
- 吉村 功, 大橋靖雄 (1992). 『毒性試験データの統計解析』, 地人書館, 東京.

## Statistical Analysis in Toxicology : Issues and a New Method on the Determination of the No-observed-adverse-effect Levels

Takashi Yanagawa

(Graduate School of Mathematics, Kyushu University)

Important statistical issues in the analysis of toxicological data are discussed by focusing on the determination of the no-observed-adverse-effect level (NOAEL) for severity data. Conventional statistical methods employ multiple comparison techniques in particular and statistical tests in general for the determination of the NOAEL. However, the use of statistical tests have shortcomings, particularly when the sample size is small. It is stressed from the principle of detecting the adverse effect as early as possible that the II<sup>nd</sup> kind of errors must be controlled rather than the I<sup>st</sup> Kind of errors. The [FLAG] method that is developed recently by using Akaike Information Criterion (AIC) (Yanagawa et al. (1997) is illustrated as a method to satisfy this principle. The method is compared with the Dunnett-type multiple comparison with one-sided Fisher exact test using the MULTTEST procedure (SAS (1994)). It is indicated that the [FLAG] method behaves well for the determination of the NOAEL.

---

Key words : AIC, dose-resonse relationship, Dunnett-type test, long-term animal bioassay, multiple comparison, MULTTEST.