

毒性試験に用いる統計 処理法の模索

財) 食品農医薬品安全性評価センター
小林 克己

科学の進歩に伴い数多くの化学物質が合成され世の中に送り出されている。それらは、医薬、食品添加物、化粧品等直接ヒトに用いられるもの、農薬や動物用薬品のように動植物を介してヒトに摂取される可能性のあるもの等様々な場面でヒトと接触を持つ。

これら化学物質は有効面(薬理薬効、生物活性)の評価と毒性面の評価がなされた後商品となるが、この段階での評価判定には生物統計手法が用いられる。

前者の薬理薬効試験に用いられる統計手法では多くの薬剤濃度が設定できました、繰り返しも多く、再試験が可能で試験期間も比較的短期である。さらに医薬、農薬では薬効の有無は対照群と比較しなくても一目瞭然である場合が多い。しかし動物用薬剤の中の成長促進剤等では対照群と比較して通常でも10%以内¹⁾の増加にすぎなく飼料効率改善効果もほぼ同程度である。このような評価試験では各群の分散もほぼ同程度を示すため、直ちに分散分析法(analysis of variance, ANOVA)やt検定等、いずれの統計手法を採用してもさほど問題はないと思われる^{1), 2), 3)}。しかし後者の毒性試験になると種々の問題が生じる。

一般的に毒性試験では対照群を含め4群を設定し試験を実施する。そして確実中毒量および無作用量を検索するのが常である。

今回受託試験施設がルーチン的に実施でき、毒性試験に適した統計処理法を模索し、いくつかの知見を得たので数種の定量データ(parametric data)を示し報告する。

一般的に統計処理の基礎には Student のt検定^{4), 5), 6), 7), 8)}の様に対照群と被験物質を投与した1群間との比較検定、いわゆる2群間の検定(2つの平均値の差の検定)である。各群の動物数の同一化、等分散を持ったものが好ましいが、これらが異っていても検定可能な Aspin-Welch

検定⁹⁾および Cochran のt検定法^{10), 11)}等の2群間の比較検定が最も多く現在使用されている。この場合、両群の分散の相違によって対照群との差が約2.8%減少していても5%水準で有意差を示し、5.6%減少していても同様の有意差を示す場合が時々見られる(Table 1)。我々は(有意水準値5%、 $p=0.05$)と対照群との差(指数)の

Table 1. Significant Differences by Student's Test and in Percents of Control Group

Sex	Dose level (ppm)	Erythrocyte counts, ($\times 10^6/\text{mm}^3$)
Male	0	8.23 \pm 0.17 (100.0)
	1,000	8.00 \pm 0.28* (97.2)
	10,000	7.91 \pm 0.21** (96.1)
	100,000	7.77 \pm 0.50* (94.4)

*: $p < 0.05$.

** : $p < 0.01$.

(): ln % of control group.

Ten rats in each group at 13 weeks after dosing.

どちらかを重要視すべきかである。これも薬量相関性の有無を含め今後重要な判断材料の規準となろう^{12), 13)}。ともあれ各ガイドライン^{14), 15)}は統計処理法を必要としているが手法については触れていない。そこで試験成績を正確に表現し検出力の高い方法で実施することが必要となるが、この検出力の強さは手法によって異なる為、解釈が大変むずかしい、その理由に少い群数の設定、検出力が高すぎて有意差が出すぎはしないか?、これら検出力の強弱をカバーするため生物学的有意差¹⁶⁾いわゆる背景値(background data)の集積が大切で、これらデータと比較することも大きな手助けとなる。

受託施設では試験依頼者の要望または受託者の推奨する統計手法によって試験結果の検定が実施される。しかし長期試験では、投薬後52~104週に進むにつれて各群間の分散値も極めて異なり動物数も同様に変化してくる。これらの事を考慮し分散値の大きな群が1ないし2群あっても一応に統計処理ができるように Chayne Gad^{9), 10), 11)}らによる決定樹法(decision tree, Fig. 1)やこれらを若干モディファイした方法が使用さ

れつつある^{17), 18), 19)}。

この決定樹法以前から分散分析法で有意差を示した場合の群間が有意差を持つかを検索するため、Duncan^{10), 20), 21), 22)}およびTurky^{9), 22), 23)}の多重範囲検定法 (Duncan's multiple range test, Turky's multiple range test) や Dunnett の多重比較検定法^{9), 24), 25), 26)} (Dunnett's multiple comparison test) 等で有意差を吟味してきた。これら分散分析法は3群以上の場合に用いられる検定法でANOVA実施前にBartlettの等分散検定^{9), 10)} (Bartlett's test) の必要が重要視されているようである。これはパラメトリックデータとして扱うか否かの第1関門ともなっている。これら多重範囲 (比較) 検定の利点は対照薬剤が入り多群となった場合、有意判定が各群平均値の肩口につけるアルファベットによって敏速に判断ができる特徴を持っている (Table 2)。特に動物用薬剤の薬効および獣医畜産関係の論文に多く用いられている^{27), 28), 29)}。しかし毒性試験では対照群との比較が主で Table 2 に示した有意差マークを用いることがほとんどない。

さてここで最近使用されてきつつある前述の決定樹法を若干モディファイ (Fig. 3) し片側検定によって種々検討した例を示す。

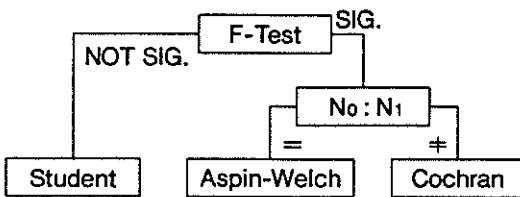


Fig. 2. Statistical Analysis of Among the Two Groups on the An-Pyo Center.

これらはいずれも二年間の長期毒性で、慢性毒性と癌原性試験を組合せた併合試験をラットで、癌原性試験をマウスを用いて実施した。両試験共対照群を含めて4群を設定し、パラメトリックデータの種類と検定項目はTable 3に示した。

調査項目は体重、飼料摂取量、飲水量、血液学検査、生化学検査、尿検査および臓器重量とその体重比等でラットの併合試験では雄雌合計680回、マウスの癌原性試験では雄雌合計434回の検定回数が必要となる (マウスの癌原性試験では上記の調査項目中実施しないものもある)。

この二試験の各測定項目を Fig. 3 の決定樹法を用いて実施した。まずBartlettの等分散検定で吟味 (P=0.05) した結果、ラットの併合試験は Table 4 にマウスの癌原性試験は Table 5 に各々示した。ラットの試験では680項目中186項目の27.3%が有意差を示し等分散でないことを示した。又体重の項目は最も等分散化していた。一方マウスの試験では434項目中114項目の26.2%が有意差を示した。中でも雌体重に高い率を示した。

これらの事からラット・マウス各二年間の試験に於る Bartlett の等分散検定で有意差を示す割合は26~27%前後であった。従ってこの26~27%の統計量がノンパラメトリックデータとしてクラスカル・ワリスの順位検定 (Kruskal-Wallis Nonparametric Analysis of Variance)^{9), 10), 30)} を通らなくてはならない。この検定は周知の如く数の大きい値から並べて行うもので差の大小にかかわらず順位づけを行う検定でパラメトリック手法と比較すると説得性に欠ける様に思える。Bartlett の等分散検定で有意差を認める経時的変化は Table 6 に示した。体重では

Table 2. Expression of the Significant Differences on Various Multiple Comparison or Range Test

Sex	Dose level (ppm)	Erythrocyte counts, ($\times 10^6/\text{mm}^3$)		
		Dunnett's	Duncan's	Turkey's
Male	0	8.23 ^a	8.23 ^a	8.23 ^a
	1,000	8.00 ^{ab}	8.00 ^{ab}	8.00 ^{ab}
	10,000	7.91 ^{ab}	7.91 ^b	7.91 ^{ab}
	100,000	7.77 ^b	7.77 ^b	7.77 ^b

Means with different superscript letters differ significantly ($p < 0.05$).

ラットで79~104週に渡り若干増加したがマウスで0~26週に最も多く、その後減少した。飼料摂取量ではラットおよびマウスで週が進むにつれて増加が認められた。血液学検査はラットおよびマウス共大きな変化が認められなかった。生化学検査はラットのみのデータから週が進むに従い増加の傾向を示した。臓器重量ではラットのみのデータではあるが試験期間を通して大きな変化を示さなかった。しかし臓器重量体重比は52週以降若干増加を示した。

従って52週以降の78および104週時になれば必ずいくつかの群で分散が極めて不均等となり Bartlett の等分散検定で有意差を示してしまう。これらの原因は高用量群の分散が大きい場合、また小さい場合、反面对照群の分散が極めて小さい場合等さまざまであった。さて前述のように26~27%のデータをノンパラメトリック扱いとして進めても良いのだろうか、Bartlett の等分散検定の重要性は毒性試験のように対照群の平均値に対して始めから数10%も減少する事が分かっておりまた分散値も投薬群で著しく変化することが当然である場合、この検定法が有用なのか大変疑問を持たざるを得ない。

ここでこの二つの試験の有意差検定の解析を Fig. 2 に示した t-検定(片側検定)、Bartlett の等

分散検定で有意差を示さなかった項目について Dunnett の多重比較検定、Bartlett の等分散検定をすることなく Dunnett の多重比較検定、最後に Fig. 3 に示した決定樹法の4手法によって試みた。有意差はいずれも5%水準値で吟味し、その結果をラットについて Table 7、マウスについて Table 8 に各々示した。

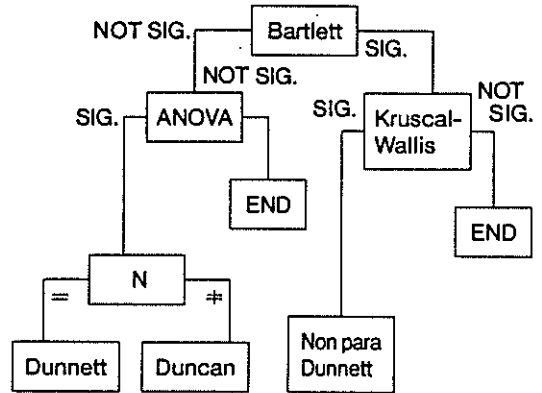


Fig. 3. Decision Tree of the An-pyo Center for Toxicological Study.

ラットの慢性および癌原性併合試験では、最高に有意差を示す可能性 2,040 個中、t-検定が 513 個 (25%)、Bartlett の等分散検定を通過後 Dunnett の多重比較検定法が 268 個 (13%)、Dunnnett の多重比較検定法のみが 320 個 (16%)、決定樹法が 308 個 (15%) であった。

Table 3. Number of Items on Statistical Analysis for 2 Years Chronic Feeding and Oncogenicity Studies in Rats and Carcinogenicity Studies in Mice

Items	80 rats/group × 4 groups			70 mice/group × 4 groups		
	Male	Female	Total	Male	Female	Total
Body weight	58	58	116	58	58	116
Food consumption	120	120	240	120	120	240
Water consumption	6	6	12	Not analysis		
Hematology	11 parameters × 4 times		88	10 parameters × 3 times		
	44	44		30	30	60
Blood chemistry	17 parameters × 4 times		136	Not analysis		
	68	68				
Urinalysis	2 parameters × 4 times		16	Not analysis		
	8	8				
Absolute organ wt.	20 (5×4times)	16 (4×4times)	36	5 (5×1time)	4 (4×1time)	9
Relative organ wt.	20 (5×4times)	16 (4×4times)	36	5 (5×1time)	4 (4×1time)	9
Total	344	336	680	218	216	434

Body weight was measured every 4 weeks from 26 weeks after dosing.

一方マウスの癌原性試験では最高に有意差を示す可能性1,302個中、t-検定が412個(32%)、Bartlettの等分散を通過後Dunnettの多重比較検定法が201個(15%)、Dunnettの多重比較検定法のみが282個(22%)そして決定樹法が213個(16%)であった。いずれの試験もt-検定が最も有意差を多く検出し、次いでDunnettの多重比較検定法、決定樹法そしてBartlettの等分散検定後Dunnettの多重比較検定法の順であった。しかしt-検定を除く3検定法間では大きな差が見られず、これら3検定法に比べてt-検定法は約1.5倍の有意差を示したことになり今回検討した4法中t-検定が最も高い検出力を示した。

毒性試験では二年間の内、後半以降にいずれかの群で分散値が著しくなる。動物数が各群いずれも10匹以上の場合、パラメトリックデータはパラメトリックデータとして扱いたいのが人情である。またBartlettの等分散検定で有意差が出る場合の多くは、ある1群に原因のある場合が多い。従ってBartlettの等分散検定を除外し分散分析法を実施しても、分散の大きい群は全平方和から全体の誤差項に吸収されてしまう。時にはこの方法でも著しく大きい分散を持つ群が混入することにより、低薬量群においても有意差水準がt-検定に比べて甘くなる。

分散分析法では全数値を用いて全平方和を作成し、各自由度で割った分散の誤差項を分母にし、各調査項目間の分散を分子にして算出し、その値(分散比)をF表を用いて判定する。この方法では数10年前のように分析機器の未発達時代の用手法時代や、測定に長時間を用いた時代等に於ける人為的測定誤差が誤差項に十分考慮されている。最近では血液分析法も多項目を短時間で処理し、体重、飼料等は640頭を各々1時間以内で測定可能で、もはや人為的疲労による誤差は極めて少い。また分散分析法は二元配置および多元配置法によって多項目を一度に吟味できるのも特徴であるが、毒性試験では多元配置法はなく、二元配置法が時おり見られる³¹⁾。

従って対照群、低薬量群、中薬量群および高薬量群は各々独立して存在していると考えられる。これらの事からt-検定系が被験物質の影響

を適確に検出する能力を持っていると推測できる。もし1群のみ分散値が極めて大きい、または小さい場合、また $n_1 \neq n_2$ の場合でも他への群への影響は全く無いのが特徴である。

前述の慢性毒性試験では定量データの内の約30%弱がBartlettの等分散検定で有意差を示し、ノンパラメトリック検定に進むが同データをFig.2に示したF-検定で有意差($P < 0.05$)を示し、Aspin-WelchおよびCochranのt-検定を実施した率をTable 9に示した。Bartlettの等分散検定に比べて、体重はラットで7%、マウスで21%、飼料摂取量はラットで13%、マウスで17%と、Bartlettの等分散検定に比較して極めて少い頻度で有意差を示した。他の血液学検査および臓器重量などについても、Bartlettの等分散検定に比べ同等またはそれ以下で有意差を示すことから、t-検定系での検索の方が妥当のように思われる。

以上の事から毒性試験に用いられる公平且つ検出力の高い定量データの統計処理法はFig.2に示したt-検定系と考える。この検定法で有意差の認められてしまう軽微な変化は背景データや対照群間との差で判断することが最適である。

本報告は安評センター所報、第2巻、1988に掲載したものをそのまま引用した。

財団法人

食品農医薬品安全性評価センター

(略称 安評センター)

〒437-12 静岡県磐田郡福田町塩新田字荒浜 582-2

TEL (0538)58-1266

FAX (0538)58-1393