

《 会 員 寄 稿 》

安全性試験に用いられる多群間検定の特徴と問題点

バートレットの等分散検定、分散分析、ダンネットの多重範囲検定および
クラスカル・ワリスの順位検定等を応用した場合の各事例

財団法人 食品農医薬品安全性評価センター 小林克己・大橋信之・大川原一弘

1. はじめに

医薬品・農薬の登録申請に必須化されている試験には実験動物を用いる非臨床試験、所謂安全性(毒性)試験が各ガイドライン [1, 2] に明確化されている。これら安全性試験の中には、薬量設定のための4週間程度の短期連続投与から慢性・癌原性試験等の78~104週間の長期連続投与の試験まで用途・期間によって試験の規模が決められ、供試動物としてげっ歯類およびイヌ等が指定されている。これら安全性試験から得られる定量データは体重・飼料摂取量等の飼育管理、血液・生化学・尿等の臨床検査および臓器重量と極めて多く、その検定項目は800から900で、検定回数は検定法によってその3~6倍となる。これら膨大な定量データは、一般的に二群間の検定、所謂t検定系またはノンパラメトリック検定と分散分析(ANOVA)を含めた多重比較・範囲検定の決定樹によって吟味され群間の差が有意かどうか5%水準以下で判定している。しかし、近年の傾向はt検定系の使用が減少し逆に分散分析を応用した多重比較・範囲検定法が増加してきた [3]。

小林 [4] によると2年間の長期毒性試験では30%の検定項目がバートレットの等分散によって不等分散を示しノンパラメトリック検定の順位検定となる。このように、2年間と長期の試験期間を経過した各パラメータは対照群を含め極めて予想もつかない数値が散見される。これらいかなる場面でも応用できるようにガッド[5]らによって多群間用の決

定樹が発表され、各試験機関・試験責任者はこれを一部応用したりモディファイして使用している。今回この決定樹を使用し安全性試験から得られたデータまたは極めて近い値を仮定し有意差の検出パターンに興味ある知見を得たので報告する。

2. 検定法、事例および解説

1) 検定法

図-1に示した多群間用決定樹を用いた。また比較検討用にスチューデントのt検定(該当事例は等分散および同一標本数を示したため)を併せて一部使用した。

2) 事例および解説

a. 事例1. ある1群の変動が大きい小さいと有意差が出にくい

F344ラット、投薬後52週の血漿グルコース値(mg/dl)

	対 照 群		低 用 量 群		中 用 量 群		高 用 量 群	
	152	145	147	143	146	147	132	135
	160	170	149	145	149	148	128	138
	155	163	148	160	170	146	139	143
	153	173	166	155	160	140	137	142
	149	157	135	140	145	135	131	142
標 本 数	10		10		10		10	
不 偏 分 散	80		87		97		26	
平均値±標準偏差	158 ± 9		149 ± 9		149 ± 10		137 ± 5	
有意差(ワリスの順位検定)	-		-		-		**	
t-検定による有意差	-		*		*		***	

- : 有意差なし。
対照群に対して*p<0.05, **p<0.01, ***p<0.001の危険率で有意差を示す。

バートレットの等分散を通過しダンネットの多重比較検定法の結果、高用量群のみ有意差を示した。またこの群は他の群に比較して分散が小さいことから全平方和を利用する分散分析では対照群に比較して6%減少してい

る低・中用量群に有意差が認められない。しかし、*t*-検定は2群の数値のみ計算式に取りこまれることから対照群に対して低・中・高用量群間に有意差が認められた。

b. 事例2. 群数が大きくなるに従って有意差が出にくい

投薬量 (ppm)	対照群				2.0 ppm				20.0 ppm				1,000 ppm				1.0 ppm	
	13.9 14.3	14.0 13.3	14.0 13.8	14.1 13.9	14.0 14.0	14.0	14.0	13.7 13.8	14.5 14.1	13.5 14.1	14.2 14.1	14.1 13.5	13.9 13.7	14.1 13.9	14.2 13.8	14.3 14.4	14.1 14.2	13.8 14.1
13.9 13.5	13.8 13.4	14.1 14.0	14.4 14.4	10	10	10	10	10	10	10	10	0.06	0.16	0.05	0.03	0.08	0.08	
平均値±標準偏差	13.9±0.3	13.9±0.4	13.9±0.2	14.2±0.2	14.0±0.3	14.0±0.3	13.9±0.2	14.2±0.2	14.0±0.3	14.0±0.3	14.0±0.3	13.9±0.3	13.9±0.4	13.9±0.2	14.2±0.2	14.0±0.3	14.0±0.3	
計算値		0.41	0.00	2.52														
<i>F</i> 値の分布表の値 (<i>K</i> =3)		2.45	2.45	2.45														
有意差		-	-	-	*	*	*	*	*	*	*	*	*	*	*	*	*	
算値		0.40	0.00	2.51													0.81	
<i>F</i> 値の分布表の値 (<i>K</i> =4)		2.53	2.53	2.53													2.53	
有意差		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	

-:有意差なし。
対照群に対して **p*<0.05 の危険率で有意差を示す。

4群間(対照群, 20, 200および1,000ppm群)の検定結果、バートレットの等分散検定を通過しダンネットの多重比較検定法の結果、高用量群のみ有意差を示した。この4群に加えてほぼ同様な平均値および分散を持った10ppm群を仮に設定していたとして、この群を加え計5群で同様の検定を実施したところ、対照群と各用量群との間に有意差は認められなかった。この理由は分布表にある、即ち*t*-検定は動物数が増加すれば分布表の値は小さくなるため有意差が出やすいが、ダンネット、ダンカン、ターキーおよびシェフェなどの分布表は群数が増加すると分布表の値が大きくなるため有意差が出にくい [5]。

c. 事例3. 高用量群の動物数が他群に比べて極めて少ない場合、ノンパラメトリック検定を選択せざるをえない。

	対照群		低用量群		中用量群		高用量群															
	4.2 3.3	1.6 2.6	2.7 1.4	0.7 0.6	4.3 1.8	2.6 1.5	3.7 1.2		1.3 2.5	2.5 1.2	4.7 0.5		2.7 3.9	1.6 1.8	1.7 0.4		3.3 2.0	1.9 0.6	2.0 2.0			
標本数	10	10	10	2	10	10	2	10	10	10	2	10	10	10	2	10	10	10	2	10	2	
不偏分散	1.1	0.4	1.8	0.0	1.1	0.4	1.8	0.0	1.1	0.4	1.8	0.0	1.1	0.4	1.8	0.0	1.1	0.4	1.8	0.0	1.1	0.0
平均値±標準偏差	2.9±1.0	1.8±0.6	2.0±1.4	0.7±0.1	2.9±1.0	1.8±0.6	2.0±1.4	0.7±0.1	2.9±1.0	1.8±0.6	2.0±1.4	0.7±0.1	2.9±1.0	1.8±0.6	2.0±1.4	0.7±0.1	2.9±1.0	1.8±0.6	2.0±1.4	0.7±0.1	2.9±1.0	0.7±0.1
バートレットの等分散検定の結果、不等分散を示した。																						
順位平均	22.9	13.9	15.2	4.3	22.9	13.9	15.2	4.3	22.9	13.9	15.2	4.3	22.9	13.9	15.2	4.3	22.9	13.9	15.2	4.3	22.9	4.3
順位総和	229	139	152	9	229	139	152	9	229	139	152	9	229	139	152	9	229	139	152	9	229	9
有意差 (ノンパラメトリック検定)		-	*	-		-	*	-		-	*	-		-	*	-		-	*	-		-
<i>t</i> -検定による有意差		**	-	*		**	-	*		**	-	*		**	-	*		**	-	*		**

-:有意差なし。
対照群に対して **p*<0.05, ***p*<0.01の危険率で有意差を示す。

高用量群の標本数および不偏分散が極めて小さいため群間に不等分散を示したことから順位検定へ移行せざるを得ない、その結果、中用量群のみ対照群と有意差を示した。しかし、この高用量群の影響がはいるこまない*t*-検定法で吟味した場合、対照群に対して低・高用量群に有意差を示した。この様に高用量群の1群のみに死亡率が高く、分散が極めて大小に分布することによって順位検定に移行せざるを得ない。

d. 事例4. 外れ値が含まれると Non-parametric ANOVA を選択せざるを得ず時には、同一平均値で有意差が出るため結果の解釈が困難となる。

	対照群		低用量群		中用量群		高用量群															
	0.70 0.68	0.72 0.64	0.56 0.59	0.51 0.59	0.70 0.74	0.66 0.66	0.66 0.68	0.49 0.60	0.60 0.65	0.88 0.68	0.57 0.67	0.58 0.62	0.65 0.72	0.70 0.51	0.70 0.83	0.51 0.57	0.63 0.78	0.65 0.63	0.86 0.68	0.60 1.78		
標本数	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	
不偏分散	0.00	0.01	0.01	0.15	0.00	0.01	0.01	0.15	0.00	0.01	0.01	0.15	0.00	0.01	0.01	0.15	0.00	0.01	0.01	0.15	0.15	
平均値±標準偏差	0.69±0.05	0.67±0.09	0.68±0.10	0.69±0.39	0.69±0.05	0.67±0.09	0.68±0.10	0.69±0.39	0.69±0.05	0.67±0.09	0.68±0.10	0.69±0.39	0.69±0.05	0.67±0.09	0.68±0.10	0.69±0.39	0.69±0.05	0.67±0.09	0.68±0.10	0.69±0.39	0.69±0.39	
バートレットの等分散検定の結果、不等分散を示した。																						
順位平均	25.8	22.8	22.6	10.9	25.8	22.8	22.6	10.9	25.8	22.8	22.6	10.9	25.8	22.8	22.6	10.9	25.8	22.8	22.6	10.9	25.8	10.9
順位総和	258	228	226	109	258	228	226	109	258	228	226	109	258	228	226	109	258	228	226	109	258	109
有意差		-	-	*		-	-	*		-	-	*		-	-	*		-	-	*		*

-:有意差なし。
対照群に対して **p*<0.05 の危険率で有意差を示す。

対照群および高用量群の平均値は共に0.69を示したが両群間の差は5%以下の危険率で有意差を示した。この理由については、(1)クラスカル・ワリスの順位検定で実施したことによる。この検定は、パラメトリック検定のように実測値を検定するのではなく、4群すべてのデータを小さい値から順位を付け、この順位を指標に吟味する。(2)対照群と高用量群の平均値が同一なのは、高用量群の動物番号10の1.78が極めて大きい値を示している為で順位が小さくとも平均値は対照群と同一となる。即ち、この1.78はたとえ0.89となっても順位は変化しないが、平均値は小さくなり、有意差が見られても納得することができる。

改善策としては、4群の1.78を何らかの統計手法で外れ値(異常値)として除外し、ダンネットの多重比較検定法で実施すれば両群の差と有意水準は、パラレルに吟味できるであろう。

この様な例は、バートレットの等分散検定で有意差を示したことによって外れ値の存在が確認される。これをt検定やダンネットの多重比較検定法で実施した場合はこの1.78の存在に気がつかず有意差がないと判断される。またノンパラメトリックの順位検定は、対照群間と単に順位に対して有意差があるとしか読み取ることができない。

e. 標本数が1の場合バートレットの等分散検定は計算不能

毒性試験では時折被験物質の影響で数週間後に高用量群でほとんど死亡し残り1頭になることが極めて稀にある。この場合この検定法は計算不能となる。この様な場合に対応できることを確認することが重要である。この理由は下記に示した計算式 [6] 中の分母の動物数の項にn-1がありこれによって分母がゼロとなるためである。

Bartlett's is based on the calculation of the corrected X² (chi square) value by the formula:

$$X_{corr}^2 = 2.3026 \frac{\sum df \left(\log_{10} \left[\frac{\sum [df(S^2)]}{\sum df} \right] \right) - \sum [df(\log_{10} S^2)]}{1 + \frac{1}{3(K-1)} \left[\sum \frac{1}{df} - \frac{1}{\sum df} \right]}$$

$$\text{where } S^2 = \text{variance} = \frac{N \sum X^2 - (\sum X)^2}{N - 1}$$

- X = individual datum within each group.
- N = number of data within each group.
- K = number of groups being compared.
- df = degrees of freedom for each group = (N - 1).

3. 検定法の動向

農業、医薬品および新規化学物質等に対する毒性試験から得られる定量データに対する統計処理をt検定系、分散分析系または他の検定法のいずれかを使用しているのか受託機関の一つである(財)食品農医薬品安全性評価センターおよび内外の5雑誌について調査した結果、t検定系は国内一雑誌のみで使用され、他の3雑誌および安評センターでは、t検定系に代って分散分析系が近年大半を占めてきている(図-2)。毒性試験の専門誌であるNTP報告ではダンネットの多重比較検定に代ってノンパラの多重比較検定法 [6, 7] を応用していることに注視したい。

Toxicology and Applied Pharmacology誌では、近年ほとんど分散分析系の使用が主

流となっている。しかし、本誌の中にはNewman-Keulsの多重範囲検定法、分散分析法で有意差を示した場合はt検定で2群間の比較およびバートレットの等分散検定後、一様性を示した場合はt検定で2群間の比較を行ない、一様性を示さない場合は順位検定あるいはSnedecor and CochranのQ検定等が紹介されている(表-1)。

4. 結 論

実験動物を用いた安全性試験から得られる定量データの中には時折種々の予想のつかないデータつまり、ある群で大きい小さい分散を示す、動物数が極めて少ない、極めて大きい外れ値が含まれる、動物数が1頭等が見られる。しかしながら、これらのデータに対して多群間検定の決定樹を用いることによって被験物質の影響を甘く評価したり見落す場合が示唆される。これらの問題点について表-2に集約した。

これらの問題の対応法として試験計画書に想定される事例を全て書き入れるか、その他の適切な統計処理を使用する場合があること等を明記することが大切である。近年t検定に対して多群間検定の決定樹を用いる方法が主流を占めてきた。上記に掲げた問題点はt検定の使用によって解決できるが、第一種の過誤が大きいことが欠点とされている。しかし、これらの問題に対して多群間検定の決定樹による有意差検出の甘さの方が大きいように思われる。

参考文献

- [1] 厚生省薬務局審査課(1989). 医薬品の製造(輸入)承認に必要な毒性試験のガイドラインについて, 厚生省(薬審第24号).
- [2] 農林水産省農蚕園芸局物防疫課(1985). 農業の登録申請に係る毒性試験成績取り扱いについて, 59農蚕第4200号.
- [3] 小林克己(1993). ゲツ菌類を用いた毒性試験から得られる定量データに使用される統計処理法の動向, 医薬安全性研究会, No. 37, 32-44.
- [4] 小林克己(1988). 毒性試験に用いられ

- る統計処理法の模索, 医薬安全性研究会, No. 28, 9-19.
- [5] Hothorn L (1991). 毒性試験における多重比較検定法, 毒性試験に関する国際計量生物会議, 東京.
- [6] Gad S. and Weil C. S. (1986). Statistics and Experimental Design for Toxicologists. New Jersey : The Telford Press, INC.
- [7] Dunn, O. J. (1964). Multiple comparisons using rank sums. Technometrics 6, 241-252.
- [8] Shirley, E. (1977). A non-parametric equivalent of Williams' test for contrasting increasing dose levels of a treatment. Biometrics 33, 386-389.

表—1. 米国誌に掲載されているゲッ歯類を用い三群以上設定した動物試験(安全性試験)の統計手法

- 1) ANOVA (分散分析法) で有意差 → t-検定
 - 2) ANOVAで有意差 → Snedecor and Cochran Q-test
 - 3) バートレットの等分散検定で等分散 → t-検定
バートレットの等分散検定で不等分散 → 順位検定
 - 4) ANOVAとt-検定の混合
 - 5) Newman-Keuls multiple-range test
 - 6) Nonparametric multiple comparison procedures of Dunn and Shirley
- 1-5) Toxicol. Appl. Pharmacol.
- 6) National Toxicology Program (NTP) Technical Report/NIH, U. S. A.

表—2. ゲッ歯類を用いた毒性試験に常用されるANOVA系(Non-parametric Anova含めて)の特徴と問題点

1. ある一群の変動が大きいと有意差が出にくい。
 2. 多重範囲・比較検定法は、群数が大きくなるに従って有意差が出にくい。
 3. 高薬量群の動物数が1~2頭と極めて少ない場合、Non-parametric Anovaを選択せざるを得ない。
 4. 外れ値が含まれると、Non-parametric Anovaを選択せざるを得なく時には、結果の考察が困難(同一平均値でも有意差が出る)。
- 二年間の長期毒性試験では30%の検定項目がバートレットの等分散検定の結果、不等分散を示す。