

《会員寄稿》

ビークルを用いた安全性試験から得られる定量データの取扱い
 第三報：安全性試験に用いる nonparametric
 多重比較検定法の有意差検出力について

財団法人 食品農医薬品安全性評価センター 小林克己, 北島省吾

要約：イヌ等の1群の動物数少ない試験に使用される nonparametric 型 Dunnett の多重比較検定は、低用量の値に大きな差が観察されても順位に変換することから有意差が認められない。もし有意差を示すことを期待すれば最低15頭が必要となる。したがって、イヌなどの動物数の少ない試験の統計処理法には、得られた定量値自体を分散分析した後、多重比較（範囲）検定する方法またはt-検定を推奨する。

1. はじめに

我国に於けるゲツ菌類およびイヌを用いた農薬および医薬品の安全性試験に用いられる統計手法は、近年t-検定に代わり図1に示すようなDunnettおよびScheffeの多重比較検定法 [1, 2] を含む決定樹法によって対照群と用量群間差を吟味している。これら安全性試験に供試される動物数は、ゲツ菌類を用いた4~13週間の短期試験が1群10匹以上、また104週間の長期試験が1群50~80匹であ

る。一方イヌを用いた試験では、各ガイドライン [3, 4] が示すように供試動物数が1群3~6頭とゲツ菌類のそれに比べて極めて少ない。特に各群の分散が著しく異なるため Bartlett の等分散検定によって有意差（全群間の分散の一様性がない）が認められた場合、Kruskal - Wallis の nonparametric ANOVA 検定に移行するその結果有意差を示し、標本数が等しい場合は Dunnett 型順位検定、標本数が異なる場合は Scheffe 型順位検定で群間差を吟味している。しかし、外国においては、これら多重比較検定は使用せず、Willcoxon 検定 [5, 6] および Shirley [7] が推奨する順位検定等が使用されている。Dunnett 型順位検定を実施した場合、イヌを用いた試験では対照群に比べて高値を示した群が認められても有意差が認められない場合 [8] が散見される。この現象の理由を解明するために仮想値を設定し nonparametric の順位和検定法がイヌの試験に適しているのか否か検討した。

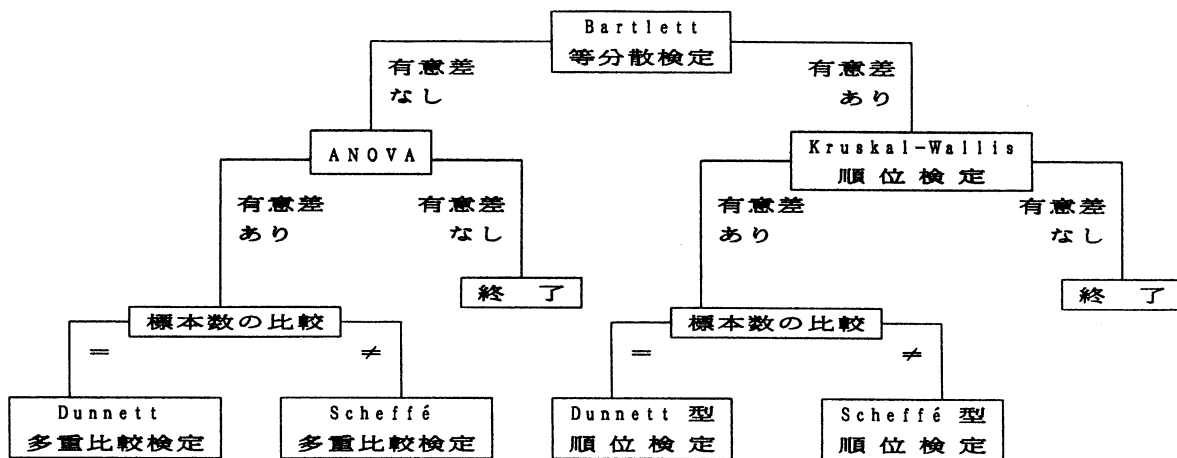


図1 多重比較（範囲）検定の決定樹

2. 方法

使用したデータは1群4頭を用いて4群設定の場合と5群設定の場合から得られた臨床検査値を順位変換した値、つまり対照群に対して用量が増加するにしたがって一定方向に傾く最も高い用量相関性を示すデータを作成し標本数(動物数)の相違によって対照群と各用量群間の順位差の有意差検出パターンをDunnett型順位検定を使用して検討した。次にDunnett型順位検定およびWilliamsのWilcoxon型検定[ユックムス(株)]を使用した場合、低用量群との間に有意差が認められる最低頭数を計算した。

3. 結果および考察

1群4頭で4または5群設定による安全性試験から得られたデータをDunnett型順位検定によって群間差を吟味した結果を表1に示す。4群を設定した場合、低用量群では対照群に比較して大きい数値であったにもかかわらず、順位検定では有意差が検出されない。また、同様に5群を設定した場合には、低および中用量においても有意差は認められない。例数の少ない試験に対してDunnett型順位検定を応用する場合、対照群に比較して低用

量群で有意差が検出できる1群の動物数を検討した結果を表2に示す。Dunnett型順位検定で有意差が示される1群の動物数は4群の場合が15頭、5群の場合が26頭であった。一方、我国では使用例が少ないWilliamsのWilcoxon型検定を使用した場合、低用量で有意差が認められる1群の動物数は4群の場合で8頭、また5群の設定で12頭でありDunnett型順位検定に比べて少数例で有意差を示した。

ゲッ歯類を用いた104週間の慢性毒性および癌原性試験では約30%のパラメータが不等分散を示し順位検定[9]となるが、この場合通常使用頭数が1群50~80匹のため、Dunnett型順位検定が応用できる。しかし、イヌ等の動物数が極めて少ない試験に応用されDunnett型順位検定は、低用量の値に大きな差が認められても順位に変換することにより有意差が認められない。したがって、イヌなどの動物数の少ない試験の統計処理には順位変換せず分散分析したのち多重比較(範囲)検定法やt-検定を推奨する。順位変換を使用する場合には、動物数が少ない場合(3頭/群)でも有意差の検出が可能なMann-WhitneyのU検定を選択したほうが良いように思う。

表1 1群4例の安全性試験から得られた臨床検査値を順位変換しDunnett型順位検定を用いて検討した結果

群:	対照群	低用量群	中用量群	高用量群	最高用量群
	1	5	9	13	17
	2	6	10	14	18
	3	7	11	15	19
	4	8	12	16	20
4群の場合		NS	*	**	
5群の場合		NS	NS	*	**

* p<0.05, **p<0.01 compared with control group.

NS:Not significant difference.

表2 順位検定を使用した場合の群内頭数および群数の差異による有意差検出の変化

	群内の各頭数	対照群	低用量群	中用量群	高用量群	最高用量群
Dunnett 型順位検定	1 4	—	NS	**	**	
	1 5	—	*	**	**	
	2 5	—	NS	**	**	**
	2 6	—	*	**	**	**
Williamsの	7	—	NS	**	**	
	8	—	*	**	**	
Willcoxon 型検定	1 1	—	NS	**	**	**
	1 2	—	*	**	**	**

* $p < 0.05$, ** $p < 0.01$ compared with control group.

NS: Not significant difference.

文 献

- [1] 佐久間明 (1981): 薬効評価 II. pp. 23 - 27, 東京大学出版会, 東京.
- [2] 山崎実, 野口雄治, 丹田勝, 新谷茂 (1981): ラット一般毒性試験における統計的手法の検討, 武田研究所報, 40 (3/4) 163 - 187.
- [3] 農林水産省 (1985): 農薬の登録申請に関わる毒性試験成績の取扱いについて, 59 農蚕第 4200.
- [4] 厚生省 (1989): 医薬品毒性試験法ガイドライン, 薬審 1 第 24 号.
- [5] Dunn, O. J. (1964): Multiple comparisons using rank sums. Technometrics, 6, 106 - 107.
- [6] Gad, C. and Weil S. C. (1986): Statistics for Toxicologists, Principles and Methods of Toxicology. Edition by A. Wallace Hayes, pp. 291 - 292, Raven Press, New York.
- [7] Shirley, E. (1977): A non-parametric equivalent of Williams' test for contrasting increasing dose levels of a treatment. Biometrics, 33, 386 - 389.
- [8] 小李克己, 渡部浩治 (1994): 薬剤 A の統計処理について, 第 58 回定例会, 医薬安全性研究会, 東京.
- [9] 小李克己 (1989): 毒性試験に用いる統計処理法の検索, 医薬安全性研究会会報, No.28, 9 - 19.

《会員寄稿》

“話題提供”

毒性試験から得られる定量データに対する
種々の検定法による有意差検出数の差異

財団法人 食品農医薬品安全性評価センター 小林 克己

農薬・医薬品に対する毒性試験に用いられる統計手法は、*t*-検定法、多重比較検定法およびノンパラメトリック検定および多重比較検定を含めた種々の決定樹方式等が使用されているが、これらの検定法が適当かどうかは種々の議論がなされている。これら検定手法によって有意差検出数の差異を調査し、片側検定と両側検定との違いによる有意差検出数の差異も併せて検討した。

財) 食品農医薬品安全性評価センターで実施した農薬に対する慢性毒性/癌原性併合試験(雌雄各80匹の5群, 内サテライト群3群を設定し途中投与26, 52, 78週で各10匹中間剖検, 104週で全生存動物を剖検)の内, 比較的有意差の多く出ている1試験を選び, 3種類の検定手法を使用した即ち,

① *t*-検定 (Student, Aspin - Welch,

Chochran)

② Dunnettの多重比較検定法

③ 自動判別方式(最初にBartlettの等分散検定を実施した。等分散の場合は一元配置の分散分析を行った。分散が有意で各群の標本数が同数の場合はDunnettの多重比較検定, 各群の標本数が異なる場合はDuncanの多重範囲検定で, またBartlettの等分散検定で不等分散の場合はKulskal - Wallisの順位和検定を実施し, 有意の場合はnonparametricのDunnettの多重比較検定)で対照群と各投薬群間の有意差を検定した。検定法①および②は, 各々片側および両側検定で③のDunnettの多重比較検定は片側検定を用い, 有意水準は全て5%で吟味した。

表一 対照群を含む5群設定の慢性毒性/癌原性併合試験から得られたデータに関する検定法の違いによる有意差検出数の差異

項目	総組合せ 検定回数	<i>t</i> -検定 ¹⁾		Dunnettの検定		Bartlett等分散検定		多群間の決定樹による有意差数		
		片側検定	両側検定	片側検定	両側検定	検定回数	有意差数	parametric ²⁾	nonparametric ³⁾	合計
体重	528	246	233	223	212	132	109	54	159	213
飼料摂取量	832	349	279	235	189	208	94	154	248	402
血液学的検査	352	159	126	123	105	88	37	77	42	119
生化学的検査	576	272	235	215	181	144	74	107	104	211
尿検査	64	11	10	7	5	16	3	4	0	4
臓器重量	224	80	61	47	42	56	17	32	13	45
臓器重量体重比	224	104	89	82	67	56	24	48	32	80
合計	2800	1221	1033	932	801	700	358	476	598	1074
(%)		(100)	(85)	(100)	(86)	(100)	(51)	(44)	(56)	(100)

有意水準は5%以下とした。

¹⁾ Student, Aspin-Welch, Chochranの *t*-検定を含む。

²⁾ Dunnettの多重比較検定およびDuncanの多重範囲検定を含む。

³⁾ Nonparametric な多重比較検定。

調査結果は表-1に示した。

1. Dunnettの多重比較検定はt-検定に比較して有意差検出数が少なく、片側および両側検定ともt-検定に比べて各々70-78%を示した。
2. t-検定およびDunnettの多重比較検定法の両側検定は片側検定に比較して有意差検出数が少なく片側検定の約85%を示した。
3. Bartlettの等分散検定の結果、半数の定量データが有意となり不等分散を示した。
4. 最も多く有意差を示した手法はt-検定の

片側検定で、次いでnonparametricのDunnettの多重比較検定を含んだ自動判別方式、t-検定の両側検定、Dunnettの多重比較検定法の片側検定、同じく両側検定の順であった。

従って、毒性試験では、ヒトに対する影響を厳しく評価することと、ほとんどの項目は対照群に対して一定方向に傾くこと。また単に”差がある””差がない”ではなく対照群に対して大きいか小さいかを試験責任者は期待することから、片側検定の使用を推奨する。