

ゲッ歯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

(財) 食品農医薬品安全性評価センター
資料評価情報部

小林克己 大橋信之 大河原一弘

要約：農薬、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（スチューデントt検定、アスピルウェルチt検定等）、決定樹方式多重比較検定（等分散検定、分散検定、パラメトリックおよびノンパラメトリック多重比較検定を含む）、または他の検定法のいずれ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五雑誌について調査した。t-検定は国内1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-検定に代って決定樹方式多重比較検定が近年大半を占めてきている。毒性試験の専門誌であるNTP 報告ではダ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を使用していることに注視したい。t-検定やパラメトリック多重比較検定に比べて、ノンパラメトリックの順位検定は平均値の差の検定ではなく平均順位の差の検定の為、飛び離れた変量等が含まれると有意の判定に難解な場合ある。しかしながら、今後3群以上を設定した毒性試験に対しては、決定樹方式多重比較検定の使用が益々多くなるう。

1. はじめに

農薬、医薬品および新規化学物質等の開発に必須の試験として、実験動物を用いた毒性（安全性）試験が含まれている。これら毒性試験に使用される統計手法は各ガイドライン¹⁻³⁾に明確化されてはならず、試験責任者が何らかの方法を選択している。一般的にゲッ歯類を用いた毒性試験では対照群、低用量群、中用量群および高用量群の四群が設定され、2-4週間の短期間の反復投与試験から2年間の癌原性試験を含めた長期の反復投与の試験までである。これら毒性試験から得られる定量データ（体重、飼料摂取量、飼料効率、血液・生化学検査値、尿検査値および臓器重量検査値）の解析には、大別してt-検定および決定樹方式多重比較検定⁴⁻⁹⁾が使用されている。毒性試験では、薬剤の投薬の影響によって投与群が対照群に対して各データの分散および平均値が著しく変動したり、動物数も大きく変化することがある。従ってこれら変動要因を十分に考察へ反映できる統計手法が望まれる。

本報告では、毒性試験に使用されている統計手法について農薬、医薬品および新規化学物質の申請窓口別に分け、受託機関である(財)食品農医薬品安全性評価センター（以下安評センター）の統計手法の動向を12年間に渡り調査し、さらに、国内で発行されている3種類の雑誌および米国で発行されている2種類の雑誌に記載されている同動向を調査した。しかし、我が国の農薬の毒性試験報告は要約のみの公表であり、従って統計手法については不明であった。

2. 調査材料および方法

1) 安評センターの動向

当センターで試験計画書を締結したものについて農薬、医薬品および新規化学物質に分類し、1980から1992年の12年間の動向をt-検定および決定樹方式多重比較検定に分けて調査した。調査対象試験は、2週間から2年間の対照群を含めて3群以上（ほとんど4群）を用いたゲッ歯類の反復投与試験および生殖・発生毒性試験・催奇形性試験について調査した。

2) 投稿雑誌掲載論文

下記の国内および米国発行誌に掲載された論文の中でt-検定または決定樹方式多重比較検定を年代別に調査した。該当試験は、1)と同様の試験に加え、対照群を含めて3群以上を用いた薬剤の影響試験についても調査した。NTP誌以外の年代別分類は原稿受諾年によった。

a. 国内発行誌

- 日本癌学会誌、Jpn. J. Cancer Res.、1988-1992年
- 医薬品研究、IYAKUHIN KENKYU、1988-1992年
- 応用薬理、PHARMACOMETRICS、1988-1992年

b. 米国発行誌

- Toxicology and Applied Pharmacology, 1987-1992年 (Vol. 17, No. 1 まで)、本誌は農薬・医薬品等の薬剤全般に対する毒性試験、癌原性試験、薬理試験から *in vitro* まで広い分野に渡って掲載されており、供試動物もゲッ歯類、犬、魚類、鳥類と多種に渡っている。
- NATIONAL TOXICOLOGY PROGRAM (NTP), PUBLIC HEALTH SERVICE / National Institutes of Health, 1987-1992 (一部欠損ナンバーあり、1992年はTR329号の一冊のみ)、既存化学物質に対する毒性試験報告の専門誌で、この組織は米国内ラボで実施した急性、4週間、13週間の予備試験および2年間の慢性毒性試験/癌原性試験等に加えて *in vitro* の試験までが一冊となって薬剤毎に報告書化し各国へ配布されている。調査は発行年によった。

3. 調査結果

1) 安評センターの動向

当センターで1980-1992年間に試験計画書の締結した内、t-検定および決定樹方式多重比較検定に分けて調査した結果を表-1に示した。

農薬に対する毒性試験では、1980-1986年の間ほとんどt-検定が使用されてきたが、1987年から決定樹方式多重比較検定が使用され始め、その後少しずつ増加し1991および1992年の使用率は100および67%となった。

医薬品では1982年を除き1981-1985年の間t-検定の使用率が100%で、その後t-検定が暫時減少し、その反面1986-1992年の間決定樹方式多重比較検定の使用率が増加し始め、1992年の使用率は57%となった。

新規化学物質について通産・厚生両者によって作成された毒性試験のガイドラインは1986年に公表され、当センターでは1986年からそれに沿った試験が行われた。表中ではその他で記載されているが、その中には食品添加物等の試験も若干含まれる。1986および19887年の間のt-検定の使用率は100%、1989年のt-検定の使用率は57%で、その後暫時減少し1992年の使用率は8%となった。

これら農薬、医薬品および新規化学物質等全体の動向は、1980-1983年の間t-検定の使用率が80-100%とほとんどを占めていたが、その後減少し始め1992年はその使用率が31%にまで減少し、代って決定樹方式多重比較検定が7割となった。

2) 国内発行誌の動向

国内発行誌の動向は、t-検定および決定樹方式多重比較検定に分けて調査した結果を表-2に示した。

日本癌学会誌は1988-1992年の間ほとんどt-検定が使用されている。

医薬品研究誌は1988-1990年の間t-検定の使用率が33-78%であったが、その後の2年間は決定樹方式多重比較検定の使用率が100%となった。

応用薬理誌は1988-1990年の間t-検定の使用率が72-80%であったが、その後減少して1991および1992年はt-検定の使用率が48および50%となった。

3) 米国発行誌の動向

米国発行誌の動向は表-3に示した。

Toxicology and Applied Pharmacology誌はt-検定の使用率が1987年に42%で、その後減少し続け1981年の使用率が17%、1992年では使用されなかった。その半面決定樹方式多重比較検定の使用率は1987年の58%から年々増加し、1991年の使用率が83%、1992年が100%となった。NATIONAL TOXICOLOGY PROGRAM (NTP) の報告では1987-1989年に渡り、主に体重と一部の臓器重量の値に対してほとんどダンネットの多重比較検定¹⁰⁾が用いられており、その後減少し1990年の使用率は約60%で1991および1992年では使用されなくなった。1989年から1992年に渡りDunn¹¹⁾およびShirley¹²⁾のノンパラメトリック多重比較検定法で対照群と各投薬群間の差を吟味している。これら検定法の使用率は1989年が12%、1990年が43%で1991年以降は100%となった。NTPの報告は、他の雑誌に見られるt-検定を一切使

用しないで分散分析を含めたダンネットの多重比較検定を使用してきたが、近年DunnおよびShirleyのノンパラメトリック多重比較検定法で対照群と各投薬群間の差を吟味していることに注視したい。

3. 考察

長期毒性試験から得られる定量データの種類と検定対象を表-4に示した。これら多項目で大量の検定回数を速やかに処理する為に、Shayne and Carrol(1986)によって決定樹方式による自動判別法⁴⁾が提唱された。すなわち2群の設定時は、分散比の検定後、分布の状況と標本数の差異によって3法のt-検定を用いる。3群以上の多群を設定した場合は、バートレットの等分散検定による分散の一様性の検定により、分散が一様な場合は分散分析を実施する。群間に有意差が認められた場合、各群の標本数が同一であるならばダンネットの多重比較検定法により、同一でなければダンカンの多重範囲検定法によって群間の平均値の差を比較する。分散が一様でない場合はクラスカル・ワリスの順位検定で群間に有意差を認めた場合、ノンパラメトリックの多重比較検定を実施する(図-1)。

一般的に薬物の各種影響試験および特に毒性試験は通常3群以上を設定する。しかし、これら試験報告によってはt-検定の使用も多く見られることから多重比較検定を含め、どのような方法でこれらに対処しているのか調査した結果、農薬、医薬品および新規化学物質等の毒性試験を受託している安評センターの12年間の動向としては、1980-1988年においてはt-検定を多く使用し、その後1991-1992年の2年間ではt-検定が減少して44%に対して決定樹方式多重比較検定が66%と増加した。使用被験物質の分類別に見ると、1992年は決定樹方式多重比較検定の使用率が化審法の反復投与試験で90%以上で、次ぎに農薬の67%、医薬の57%と半数以上を占めている。このことから、今後は決定樹方式多重比較検定の使用の増加が示唆される。

近年の国内発行誌は、日本癌学会誌がt-検定を多用している。逆に医薬品研究はここ2年間で全て決定樹方式多重比較検定を使用している。応用薬理誌ではt-検定と決定樹方式多重比較検定の使用率が同率で、今後日本癌学会誌を除き決定樹方式多重比較検定の使用が大半を占めると考えられる。

Toxicology and Applied Pharmacology誌では、近年ほとんど決定樹方式多重比較検定の使用が主流となっている。しかし、本誌の中には下記の検定法の使用が散見される。

①Newman-Keulsの多重範囲検定法。

②分散分析法で有意差を示した場合は、t-検定で2群間の比較を行なう。

③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群間の比較を行なう。

また、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。

NTPの報告では、1990年以前が全てダンネットの多重比較検定法を使用してきたが1991年からノンパラメトリックの多重検定が使用されている。これら定量データを始めからノン

パラメトリックの順位検定で吟味することの理由には、長期毒性試験から得られる定量データの一部は標本数、分散および平均値等が群間で著しく変化する為、一般的に決定樹方式で吟味しているがこれらの統計手法流れの繁雑性を避けること又、群間の一様性がある場合でもパラメトリック検定とノンパラメトリック検定を比べた場合その検出力に極めて大きい差がない為と思われる。さらに癌原性試験では最終的に病理組織学所見が重要視されることも理由となろう。小林¹³⁾によると2年間の長期毒性試験では、定量データの約30%がパートレットの等分散検定で有意差を示しノンパラメトリックの順位検定で吟味しなくてはならないと報告している。

長期の毒性試験から得られる定量データの内、特に52週齢以降のデータには加齢による個体差の増強が加わり、また薬剤の影響に群間個体数の差異が生じ(表-5)時には著しく外れた測定値も見られることから、各群間の分散が著しく異なる。

まれに決定樹方式多重比較検定で実施した為に2群間の平均値が同一であるが有意差を示す場合がある(表-6)。この理由は投薬群に飛び離れた変量があることからパートレットの等分散検定では一様性を示めさず順位検定に進んだことが一因となった為である。

製薬協のアンケート調査¹⁴⁾による試験施設間の調査結果によると近年t-検定が22-35施設から3-5施設に激減し、逆に決定樹方式多重比較検定が7-9施設から23-26施設に増加したこと、またt-検定は生物学的に問題とすべきでないような差をも有意と判定することがあり成績の評価を混乱させることもあると報告している。

今後、これら毒性試験の統計処理法は決定樹方式多重比較検定の使用が多くなる傾向にある。しかしながら、t-検定およびノンパラメトリックの順位検定の応用の利点も考えたい。

4. 参考文献

- 1) 厚生省薬務局審査課：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）、1989。
- 2) 農林水産省農蚕園芸局物防疫課：農薬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、1985。
- 3) 通商産業省：新規化学物質に係わる試験の方法について、環保業第700号、薬発第1039号、61基局第1014号、1986年。
- 4) Shayne Gad and Carrol S. Weil: Statistics and Experimental Design for Toxicologists. New Jersey: The Telford Press, INC., 1986.
- 5) 吉村 功（編）：毒性・薬効データの統計解析。サイエンティスト社、1987。
- 6) Colquhoun, D.: Lectures on Biostatistics. Oxford University Press, 1971: Iyaku journal Co. Ltd., Tokyo.
- 7) 佐久間 昭：薬効評価—計画と評価—IおよびII、東京大学出版会、1981。
- 8) 吉田 実：畜産を中心とする実験計画法、養賢堂、東京、1980。
- 9) 医薬品評価委員会・基礎研究部会：実験動物技術者のための教育資料、日本製薬工業協会、1984。
- 10) Dunnett, C. W.: A multiple comparison procedure for comparing several treatments with a control, J. Am. Stat. Assoc. 50, 1096-1122, 1955.
- 11) Dunn, O. J.: Multiple comparisons using rank sums, Technometrics 6, 241-252, 1964.
- 12) Shirley, E.: A non-parametric equivalent of Williams' test for contrasting increasing dose levels of a treatment, Biometrics 33, 386-389, 1977.
- 13) 小林 克己：毒性試験に用いる統計処理法の模索、医薬安全性研究会、No. 28、9-12。1989。
- 14) 日本製薬工業会医薬品評価委員会、基礎部会第5分科会：生殖・発生毒性試験における統計処理及びデータの取扱いについて、医薬品研究、23 (5) 656-663、1992。

財団法人 食品農医薬品安全性評価センター

〒437-12 静岡県磐田郡福田町塩新田字荒浜582-2、

電話 0538-58-1266、ファックス 0538-58-1393。

Table 1. Changes in Rodents Toxicological Study Analyzing t-tests or Multiple Comparison/
Range Test Designed Three Groups or More in the An-Pyo Center

Year*	Percentages (Numbers of study)													
	Pharmaceutical				Pesticides				Others				Total	
	t-tests	MC/RT	t-tests	MC/RT	t-tests	MC/RT	t-tests	MC/RT	t-tests	MC/RT	t-tests	MC/RT	t-tests	MC/RT
1980	-Z(0)	-Z(0)	100Z(15)	0Z(0)	-Z(0)	-Z(0)	-Z(0)	-Z(0)	-Z(0)	100Z(15)	0Z(0)	100Z(15)	0Z(0)	
1981	100Z(15)	0Z(0)	100Z(18)	0Z(0)	-Z(0)	-Z(0)	-Z(0)	-Z(0)	-Z(0)	100Z(33)	0Z(0)	100Z(33)	0Z(0)	
1982	50Z(4)	50Z(4)	100Z(12)	0Z(0)	-Z(0)	-Z(0)	-Z(0)	-Z(0)	-Z(0)	80Z(16)	20Z(4)	20Z(4)	20Z(4)	
1983	100Z(12)	0Z(0)	100Z(9)	0Z(0)	-Z(0)	-Z(0)	-Z(0)	-Z(0)	-Z(0)	100Z(21)	0Z(0)	100Z(21)	0Z(0)	
1984	100Z(7)	0Z(0)	86Z(12)	14Z(2)	-Z(0)	-Z(0)	-Z(0)	-Z(0)	-Z(0)	90Z(19)	10Z(2)	10Z(2)	10Z(2)	
1985	100Z(24)	0Z(0)	71Z(5)	29Z(2)	-Z(0)	-Z(0)	-Z(0)	-Z(0)	-Z(0)	94Z(29)	6Z(2)	6Z(2)	6Z(2)	
1986	72Z(13)	28Z(5)	100Z(7)	0Z(0)	100Z(1)	100Z(1)	100Z(1)	100Z(1)	100Z(1)	77Z(20)	23Z(6)	23Z(6)	23Z(6)	
1987	79Z(11)	21Z(3)	50Z(1)	50Z(1)	100Z(7)	0Z(0)	0Z(0)	0Z(0)	0Z(0)	83Z(19)	17Z(4)	17Z(4)	17Z(4)	
1988	95Z(19)	5Z(1)	60Z(9)	40Z(6)	-Z(0)	-Z(0)	-Z(0)	-Z(0)	-Z(0)	80Z(28)	20Z(7)	20Z(7)	20Z(7)	
1989	93Z(13)	7Z(1)	64Z(9)	36Z(5)	57Z(4)	43Z(3)	43Z(3)	43Z(3)	43Z(3)	74Z(26)	26Z(9)	26Z(9)	26Z(9)	
1990	85Z(11)	15Z(2)	29Z(2)	71Z(5)	50Z(1)	50Z(1)	50Z(1)	50Z(1)	50Z(1)	64Z(14)	36Z(8)	36Z(8)	36Z(8)	
1991	85Z(22)	15Z(4)	0Z(0)	100Z(9)	22Z(2)	78Z(7)	78Z(7)	78Z(7)	78Z(7)	55Z(24)	45Z(20)	45Z(20)	45Z(20)	
1992	43Z(9)	57Z(12)	33Z(1)	67Z(2)	8Z(1)	92Z(11)	92Z(11)	92Z(11)	92Z(11)	31Z(11)	69Z(25)	69Z(25)	69Z(25)	

* Year the study protocol concluded.
MC/RT: Multiple comparison/range test.