

“ABSEPS=”には、求めたい計算精度を、計算値の標準誤差の3倍の値で、“ABSEPS=0.0001;”のような形で入れる。精度を標準誤差で指定するのは、このプログラムが準乱数を用いて関数値の計算点を選び数値積分をしているので、計算値にランダム性が含まれるためである。有意水準を与えて検出力を計算するときは、棄却限界値の計算が必要である。これは逐次近似で求めるので、与えた有意水準と逐次近似で計算した第1種の過誤確率との差として、求めたい精度の値を“EPS1=”に入れる。このプログラムではその精度を0.0001としてある。

“A1=”と“A2=”に入れるのは、その関係が $\alpha_1 = r\alpha_2$ で全体の有意水準を“Alpha”とする2つの最大対比法の有意水準 α_1, α_2 を逐次近似で求めるための α_1 の初期値で、それぞれ十分小さい値と大きい値の二つを指定する。

“Switch=”には、棄却限界値と検出力の両方を出力するとき“Switch=1;”を指定し、棄却限界値のみを出力するときは、1以外の値を入れる。

“ESTPOWER_CMCM”はプログラム内のサブルーチン名である。

このプログラムでは、Bretz and Genzのサブルーチンの制約によって、2つの最大対比法を構成する全ての対比統計量間の相関行列が半正定値行列(positive semi-definite matrix)でなければならない。

上記のプログラムを実行すると出力結果の“N”には、各群のサンプルサイズが、17,17,17,17,17というように出力され、“CRIVAL1”と“CRIVAL2”には棄却限界値2.2011719と1.8945313が出力され、“ALPHA1”と“ALPHA2”には2つの最大対比法の第1種の過誤確率0.0493146と0.0491975が出力される。“POWER”には計算結果の検出力81.74153%が出力される。

3.2 サンプルサイズ計算例

前節と同じ5群のダネット検定とHU/HD法で、有意水準(包括的帰無仮説の下での第1種の過誤確率)2.5%、前節で示した対立仮説に対して検出力が80%になるサンプルサイズ計算プログラムとその出力例を表2に示す。

すでに説明したことを省くと各項目の意味・使い方は次の通りである。

“Beta=”には検出力に対応する第2種の過誤確率すなわち $1 - 0.80 = 0.20$ を入れる。このプログラムでは、“N_ALLOC=”に各群のサンプルサイズの比を入れ、第1群のサンプルサイズを計算するようにしてある。ここでは比を多くの場合と同様に1としたが、比が違っていても計算できる。

“Nmim=”と“Nmax=”に入れるのは、逐次近似における第1群のサンプルサイズの初期値で、それぞれ十分小さい値と大きい値の二つを指定する。“Eps2=”には、逐次近似で求める精度について、指定した検出力と計算した検出力の差の限界値を与える。ここでは0.01を入れている。

“DO UNTIL~END”は、初期値の下限と上限の間で必要なサンプルサイズを求めるためのプログラム、その前の“DIFF=1”、“N1=Nmax”は、そのプログラムのための初期値であり、然るべき条件を満たせば他の値でも差し支えないが、このままにしておく方が無難である。

出力結果では、サンプルサイズが小数で出力されるが、実際はその小数を下回らない整数にしなければならない。この例では、各群のサンプルサイズを17,17,17,17,17にすることになる。このときの検出力が前節で例示した81.7%で、出力として示された小数のサンプルサイズに対する

表 2. 必要サンプルサイズ計算例

```

PROC IML;
  %include '(sub_mvt.sas の保存先パス)sub_mvt.sas';
  Alpha=0.025; Ratio_Alpha=1; Beta=0.2;
  Contrast1={-1 1 0 0 0,
             -1 0 1 0 0,
             -1 0 0 1 0,
             -1 0 0 0 1};
  Contrast2={-26 -1 4 9 14,
             -14 -9 -4 1 26};
  Expect={0 0.25 0.50 0.75 1.00};
  VARIANCE=1; N_ALLOC={1 1 1 1 1};
  ABSEPS=0.0001; Eps1=0.0001; A1=0; A2=0.1;
  Nmin=5; Nmax=50; Eps2=0.01;
  DIFF=1; N1=Nmax;
  Switch=1;
  DO UNTIL(ABS(Power-(1-Beta))<Eps2);
    IF DIFF>0 THEN Nmax=N1; ELSE Nmin=N1;
    N1=(Nmax+Nmin)/2;
    RUN ESTPOWER_CMC(M(Switch,Alpha,Ratio_Alpha,A1,A2,N1,Eps1,ABSEPS,Expect,
                      VARIANCE,CONTRAST1,CONTRAST2,N_ALLOC,N,Power,Crival1,
                      Crival2,Alpha1,Alpha2);
    DIFF=Power-(1-beta);
  END;
  PRINT N,, Crival1 Crival2 Alpha1 Alpha2 Power;
QUIT;

```

出力結果:

出力結果:				
N				
16.25	16.25	16.25	16.25	16.25
CRIVAL1	CRIVAL2	ALPHA1	ALPHA2	POWER
2.1992188	1.8925781	0.0496686	0.0495421	0.0800529

検出力計算値 80.1%より大きくなる。

4. 細胞形質転換試験のデータ解析法への応用

4.1 2段階細胞形質転換試験の概要

がん原性における化学物質のイニシエーション活性及びプロモーション活性を検出するための *in vitro* 試験法に「BALB/c 3T3 細胞による2段階細胞形質転換試験 (cell transformation assay, CTA)」がある。CTA はイニシエータ検出試験とプロモータ検出試験の2種類からなる試験であ

る。試験の詳細は日本規格協会発行のテクニカルレポート (JSA, 2002) に委ねて、データの意味を理解するための概要紹介を行うと以下のようなになる。

イニシエータ検出試験では、BALB/c 3T3 細胞を一定数入れたシャーレにプロトコルに定められた手順で被験物質を添加し、その後プロモータ処理をする。これで反応が陽性であれば、被験物質はイニシエータであることになる。プロモータ検出試験では、当該細胞を一定数入れたシャーレにイニシエータ処理を行った後で、プロトコルに定められた手順で被験物質を添加する。これで反応が陽性であれば、被験物質はプロモータであることになる。反応変数としては、ディッシュごとのフォーカス (単数で focus, 複数で foci) が用いられる。フォーカスとは、「単層培養細胞上に、がん原性物質などの処理によって形態学的形質転換を起こした細胞が高密度に増殖し形成した細胞集団」のことである。

この試験法の多施設バリデーション共同研究 (以下、「共同試験」という) が、19 施設の参加の下に日本で行われ、その結果が Tsuchiya et al. (1999) によって報告されている。この共同研究では、被験物質として、“0-Methylcholanthrene (MCA)” と “12-O-teradecanoylphorbol-13-acetate (TPA)”, “dimethyl sulphoxide (DMSO)” を用いている。MCA はイニシエーション活性を持つ物質、TPA はプロモーション活性を持つ物質、DMSO は溶媒としてよく用いられている物質である。

表 3 及び図 2 に示したのは、この共同研究で得られたイニシエータ検出試験データの一部である。

表 3. 細胞形質転換試験データ

Lab.	イニシエータ MCA ($\mu\text{g/ml}$)	プロモータ TPA ($\mu\text{g/ml}$)	ディッシュ番号											
			1	2	3	4	5	6	7	8	9	10		
A	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0.1	0	0	0	0	0	0	0	0	0	0	0	0	0
	0.2	0	0	0	0	0	1	0	1	0	0	0	0	0
	0.5	0	0	1	0	1	1	0	0	1	0	0	0	0
	1.0	0	0	1	1	1	0	0	0	1	1	0	0	0
D	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0.1	0	0	0	0	0	0	0	0	0	0	0	0	0
	0.2	0	0	0	0	0	0	0	0	0	0	1	0	0
	0.5	0	0	0	0	1	6	0	1	0	0	0	0	0
	1.0	0	1	1	0	1	1	1	0	1	1	0	0	0
F	0	0.1	0	0	0	1	0	0	0	0	0	1	0	0
	0.1	0.1	1	2	2	5	3	2	2	3	2	1	1	0
	0.2	0.1	1	5	0	3	0	2	3	2	0	1	1	0
	0.5	0.1	1	3	0	2	1	1	0	1	3	1	1	0
	1.0	0.1	2	1	0	3	0	6	3	3	1	1	1	0
M	0	0.1	0	0	0	0	0	0	0	0	1	0	0	0
	0.1	0.1	1	1	0	2	1	0	2	1	0	0	0	0
	0.2	0.1	2	2	2	0	1	1	0	1	1	2	0	0
	0.5	0.1	5	3	1	4	1	2	2	3	1	0	0	0
	1.0	0.1	0	2	2	0	0	0	0	0	1	1	1	0

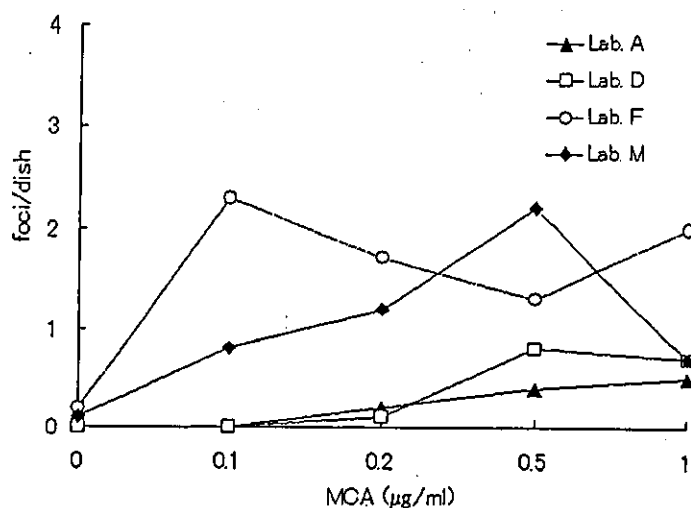


図 2. 細胞形質転換試験データ

る。観測された値は、おおむね1桁の整数でポアソン分布が想定できる。MCAは5用量に設定され、各用量で10ディッシュの反復観測が行われている。プロモータ処理には、プロモータでないDMSOとプロモータであるTPAが用いられている。前者でフォーカス反応が陰性、後者でフォーカス反応が陽性ならMCAがイニシエータ活性を持っていて、プロモータ活性を持たないことになる。

問題はこの種のデータが得られたとき、反応が陰性か陽性かをどのような手法で判定するかである。

4.2 旧西山手順

Tsuchiya et al. (1999) で報告されている共同研究において、本論文の著者らは陽性・陰性の判断に次に示す3ステップの手順を提案し、適用した (Nishiyama et al., 2003)。本論文ではこれを「旧西山手順」と呼ぶことにする。

旧西山手順では、まず、観測値であるフォーカス数 y を $z = \sqrt{y} + \sqrt{y+1}$ に平方根変換する。これを正規分布に従う変数と見なして、以下の手順を適用し、陽性・陰性の判定を下す。(図3の左側の流れ図を参照)

1. (5群の) データにダネット検定を適用し、有意でなければ陰性と判定する。有意であれば次のステップに進む。
2. 頭打ち現象 (down turn) を検出するマーゴリン手順を適用し、高用量で頭打ちが認められた場合には、その用量のデータを除外し、残ったデータに再びマーゴリン手順を適用する。これを繰り返して除外するデータが無くなったとき、残っているデータが3群以下であったら、実験が不十分であったとして、解析を打ち切り再実験を行う。4群以上だったら次のステップに進む。
3. 定義行列が式 (12) で与えられる最大対比法を適用し、有意であれば陽性と判定し、有意でなければ陰性と判定する。

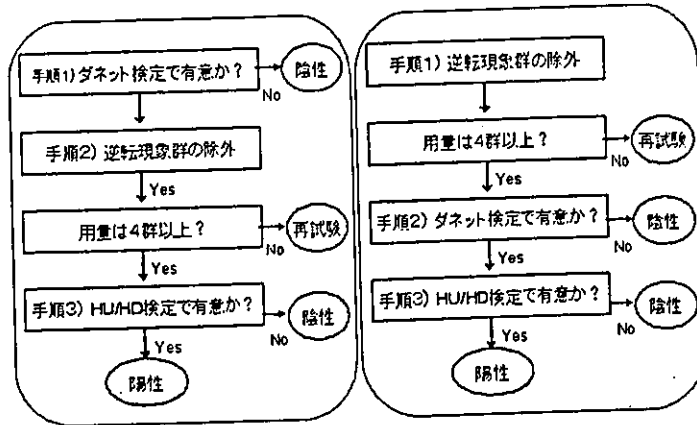


図 3. 新旧西山手順

$$C_{HU/HD} = \left(\begin{array}{l} K(1, K-1, K-1, \dots, K-1) + K(0, 1, 2, \dots, K-1) - 0.5(K-1)(3K-2)(1, 1, 1, \dots, 1) \\ K(0, 0, 0, \dots, K-1) + K(0, 1, 2, \dots, K-1) - 0.5(K-1)(K+2)(1, 1, 1, \dots, 1) \end{array} \right) \quad (12)$$

この手順において、マーゴリン手順とは、Simpson and Margolin (1986) が提案している検定法で、最高用量群の平均をそれより低用量の全群の平均と比較し、有意水準0.50で最高用量群の平均が有意に低ければ、細胞障害という毒性によって試験目的の現象が観測できなかったとするものである。彼らの論文では検定法としてノンパラメトリック検定を採用しているが、上記手順では通常の前平均値のt検定を採用している。引用した論文の筆頭著者がシンプソンであるにもかかわらず、マーゴリン手順と呼ぶのは、この方針がマーゴリンの提案であることが、この分野の研究者の間で周知だからである。定義行列が式(12)で与えられる最大対比法の意義については、Stewart and Ruberg (2000) 及び西山ら (2003) の論文に説明されている。用量反応関係が急上昇型か緩上昇型かが曖昧なときに、比較的頑健な検定法である。

4.3 新西山手順

in vitro 毒性試験での陽性・陰性の判定に、第1章で述べた複合手順を適用したとき、実験を行った毒性家から統計家に出される疑問は、慣例として用いられている0.05あるいは0.01という名義有意水準を採用すると、偽陽性が出過ぎるということである。なぜそうであるかは明確ではないが、かなり良く出される見解は、過分散現象 (overdispersion) があることと、生体 (*in vivo*) の恒常性 (homeostasis) によって *in vitro* で見られる反応が *in vivo* での予測において過大になること、の2つである。実際、Kirkland ed. (1989) では、不均質係数 (heterogeneity factor) H を誤差分散にかけることで、有意になりにくい判定を行うことを勧めている。

しかしながら著者らは、不均質係数の導入に疑問を持っている。実際の実験では、群間に不均質性があるような手順が取られていないし、前述の共同研究のデータではポアソンの分散指数 (Poisson index of dispersion) がほぼ1になっている。Omori et al. (2002) らが解析しているMLAの場合でも、各施設内での実験データでは過分散現象が観察されていない。不均質係数を適切に設定する根拠は一般になく、背景データから定めることも非現実的である。

そこで Hayashi et al. (1989), Omori et al. (2002), Nishiyama et al. (2003) が採用した方針は, 判定手順において総合的な偽陽性確率 (第1種の過誤確率) を慣例的な 0.05 あるいは 0.01 とするのではなく, 毒性家の多くが判定に同意する比較的小さい値に設定することである. 上記の論文では, そのような作業を通して, 判定手順に含まれている個々の検定法の有意水準を調整して, 総合的な偽陽性確率を 0.01 あるいはそれ以下にする手順を採用している.

この方針を採用したとき, 問題になるのは, 手順全体の総合的な偽陽性確率の計算である. 上記の論文では手順を構成する個々の検定法に慣例的な名義有意水準を設定し, 複数の検定のすべてで有意になることを求めることで, 総合的な偽陽性確率を小さくし, モンテカルロ・シミュレーションでその値を評価した. しかしこれでは総合的な偽陽性確率の制御を任意に行うことができない.

本論文で提案している確率計算法は, これに答えようとするものであるが, マーゴリンの手順が入ると, 全体が複合最大対比法にならない.

本論文では, 旧西山法の第1ステップと第2ステップを逆順にすることを考えた. これは, 頭打ち現象が認められた用量群を除外することのある種の予備検定と位置づけ, それ以後を本当の判定手法と考えるものであり, 論理的に妥当である. ただし, 予備検定, ダネット検定, 式(12)の定義行列で与えられる最大対比法の検定統計量の間には相関があるので, 複合最大対比法の枠内には入らない.

以下では, この新しい手順を「新西山手順」と呼ぶことにする. 重複になるが, 念のためその手順を以下に示しておく. (図3の右側の流れ図を参照)

1. (5群の) データにマーゴリン手順を適用し, 高用量で頭打ちが認められた場合には, その用量のデータを除外し, 残ったデータに再びマーゴリン手順を適用する. これを繰り返して除外するデータが無くなったとき, 残っているデータが3群以下であったら, 実験が不十分であったとして, 解析を打ち切り再実験を行う. 4群以上だったら次のステップに進む.
2. データにダネット検定を適用し, 有意でなければ陰性と判定する. 有意であれば次のステップに進む.
3. 定義行列が式(12)で与えられる最大対比法を適用し, 有意であれば陽性と判定し, 有意でなければ陰性と判定する.

4.4 新西山手順の確率計算法

5群の場合について, 個々の最大対比法の名義有意水準を定めて, 新西山手順の第1種の過誤確率, 検出力計算は以下のようなになる.

第1ステップの判定を調べるには, マーゴリン手順のための対比を考える必要がある. 3群になったら再試験であるから, 5群と4群の場合だけを考えればよい. そのための対比ベクトル (定義行列) は次に示す C_{M5}, C_{M4} の2つである.

$$C_{M5} = \begin{pmatrix} -1 & -1 & -1 & -1 & 4 \end{pmatrix}, \quad (13)$$

$$C_{M4} = \begin{pmatrix} -1 & -1 & -1 & 3 \end{pmatrix}, \quad (14)$$

続くステップで用いられる最大対比法の定義行列は, 式(10), (11) (5群のままの場合) 及び次に

示す C_{DT_4}, C_{HUHD_4} (最高用量群が除かれた場合) である.

$$C_{DT_4} = \begin{pmatrix} -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{pmatrix}, \tag{15}$$

$$C_{HUHD_4} = \begin{pmatrix} -15 & 1 & 5 & 9 \\ -9 & -5 & -1 & 15 \end{pmatrix}, \tag{16}$$

前節で説明したように, この手順の確率計算は, 前章で述べた複合最大対比法の枠組で計算できない. この手順の棄却確率は, 第1ステップの条件付きの確率,

$$\sum_{K=4}^5 \{P(t_{max}^{(DTK)} > a_{DTK}, t_{max}^{(HUHDK)} > a_{HUHDK}, t_{max}^{(MK)} > a_{MK}) \prod_{k=K+1}^5 P(t_{max}^{(Mk)} \leq a_{Mk})\} \tag{17}$$

で与えられ, これは包括的帰無仮説の下で第1種の過誤確率, 任意の用量反応関係に対して「その用量反応関係における検出力」である.

ただし, $t_{max}^{(MK)}, t_{max}^{(DTK)}, t_{max}^{(HUHDK)}$ 及び $a_{MK}, a_{DTK}, a_{HUHDK}$ は, それぞれ K 群の場合のマーゴリン手順の対比による検定, ダネット検定, 定義行列が式(12)で与えられる最大対比法の検定統計量及び棄却限界値である. また, $a > b$ のとき $\Pi_a^b(\cdot) = 1$ と定義する.

実際の計算では式(17)を下記のように下側積分で整理して, 各項ごとの計算値の和, 差と積の組み合わせで計算する.

$$\begin{aligned} & \sum_{K=4}^5 [\{ 1 - P(t_{max}^{(DTK)} \leq a_{DTK}) \\ & \quad - P(t_{max}^{(HUHDK)} \leq a_{HUHDK}) \\ & \quad - P(t_{max}^{(MK)} \leq a_{MK}) \\ & \quad + P(t_{max}^{(DTK)} \leq a_{DTK}, t_{max}^{(HUHDK)} \leq a_{HUHDK}) \\ & \quad + P(t_{max}^{(DTK)} \leq a_{DTK}, t_{max}^{(MK)} \leq a_{MK}) \\ & \quad + P(t_{max}^{(HUHDK)} \leq a_{HUHDK}, t_{max}^{(MK)} \leq a_{MK}) \\ & \quad - P(t_{max}^{(DTK)} \leq a_{DTK}, t_{max}^{(HUHDK)} \leq a_{HUHDK}, t_{max}^{(MK)} \leq a_{MK}) \\ & \quad \} \prod_{k=K+1}^5 P(t_{max}^{(Mk)} \leq a_{Mk})]. \tag{18} \end{aligned}$$

式(18)の計算を具体化した新西山手順の棄却限界値, 検出力, サンプルサイズの計算を行うサブルーチンプログラムを作成し, このサブルーチンを「sub_mvt.sas」に追加した. メインプログラム「NewNishiyama.sas」の使い方は, 第3章の説明と同様であるのでここでは省略する. これらも東京理科大学「医薬統計コース」ホームページから取得できる.

4.5 いくつかの手法の性能比較

2段階細胞形質転換試験で観察される用量反応関係として図4の5つの形状を考える. この形状に対して, どのような最大対比法あるいは複合最大対比法を適用するのが良いかは自明ではない. したがって, ここでは以下の4通りの方法を考えて検出力を比較して検討する.

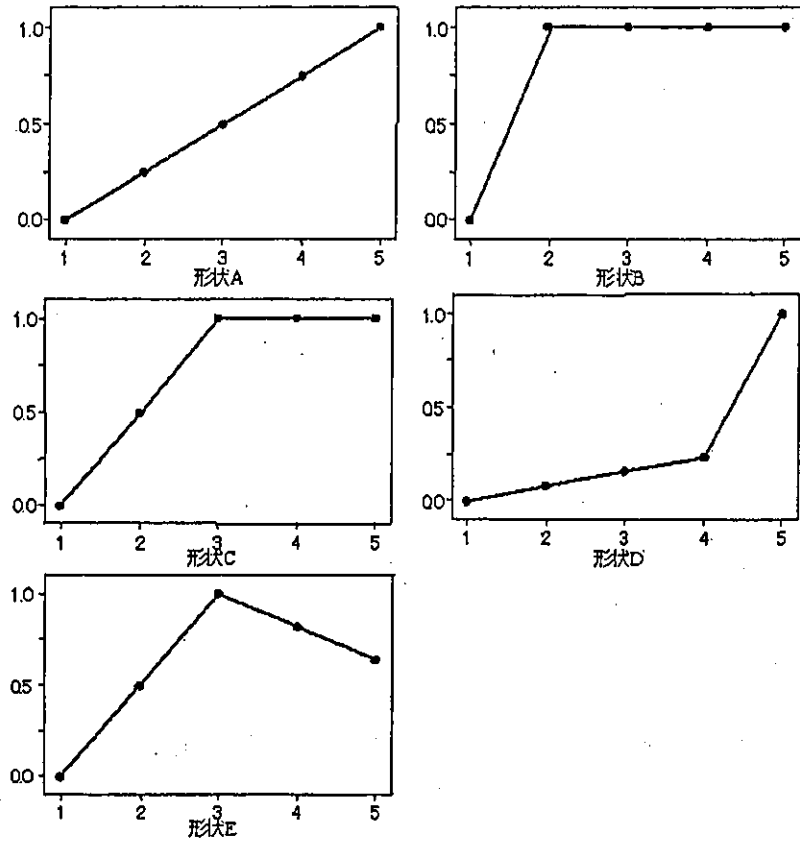


図 4. 例題において検出力を求める用量反応関係

方法 1: 定義行列 $C_{LR} = (-2, -1, 0, 1, 2)'$ の最大対比法 (LR 法), 方法 2: 定義行列がそれぞれ式 (10) と C_{LR} の 2 つの最大対比法による複合最大対比法 (ダネット&LR 法), 方法 3: 定義行列がそれぞれ式 (10) と式 (11) の 2 つの最大対比法による複合最大対比法 (ダネット&HU/HD 法), 方法 4: 第 4.3 節の新西山手順.

図 4 の 5 つの形状に対して, 上記の 4 方法で検出力 80% を保証するサンプルサイズを計算すると, 表 4 が得られる. いずれも, 有意水準は 2.5% で, 用量反応関係の上下方向の最大差は $\Delta = 1$, 誤差分散が $\sigma^2 = 1$ の場合である.

単純な LR 法は, 当然のことながら直線型の形状 A に有利である. また, 直線型に近い形状 C

表 4. 1 群あたり必要サンプルサイズ N (実際の検出力)

	LR		ダネット&LR		ダネット&HU/HD		新西山手順	
	N	検出力	$\alpha_{DT_5} = \alpha_{LR_5}$	N 検出力	$\alpha_{DT_5} = \alpha_{HUHD_5}$	N 検出力	$\alpha_{DT_K} = \alpha_{HUHD_K}$	N 検出力
A	13	0.8007	0.0583	16 0.8069	0.0494	17 0.7996	0.0363	18 0.8030
B	21	0.8186	0.0583	16 0.8047	0.0492	12 0.7983	0.0361	13 0.8117
C	13	0.8007	0.0581	12 0.7903	0.0492	13 0.8005	0.0361	13 0.8008
D	18	0.8172	0.0584	19 0.8045	0.0494	19 0.8063	0.0363	20 0.7970
E	32	0.8117	0.0584	24 0.7977	0.0494	20 0.7976	0.0363	19 0.8019

や形状 D にはロバストであるが、非線形な用量反応関係には不利である。ダネット&LR 法は、ダネット&HU/HD 法や新西山手順と比べると、形状 B の飽和型や形状 E の頭打ち型に対してロバストでない。ダネット&HU/HD 法や新西山手順は、5 つの用量反応関係の全てに対してロバストである。

4.6 適 用 例

(表 3, 図 2) の例に示した試験データに、前節の 4 手法を適用してみよう。すでに述べたように、個々の最大対比法の名義有意水準を慣例的に使われる値 0.05 にすると、毒性家に違和感のある判定結果となる。そこで名義有意水準を 0.01, 0.001, 0.0001 と変えて、4 手法を適用すると表 5 の結果が得られる。“*” は、陽性と判定されたことを示している。

表 5. 実データへの適用

	LR			ダネット&LR			ダネット&HU/HD			新西山手順		
	0.01	0.001	0.0001	0.01	0.001	0.0001	0.01	0.001	0.0001	0.01	0.001	0.0001
A	*	*		*			*			*		
D	*	*		*			*			*		
F				*			*	*		*	*	
M	*			*	*		*	*		*	*	*

有意水準 0.001 の場合のダネット&HU/HD 法と新西山手順が、Lab.A, D は陰性で、Lab.F, M は陽性という毒性家の経験的判断と一致する。2 段階形質転換試験では名義有意水準を慣例的に使われる値 0.05 や 0.01 などになると、毒性学的には偽陽性が出過ぎることが分かる。

新西山手順でマーゴリンの手順を含むのは、図 2 の形状 E のような高用量で頭打ち現象のある用量反応関係でも検出力を高く保つことを意識しているが、ダネット&HU/HD 法は HU/HD 法がある程度の頭打ち現象に対してロバストな検定法であるので Lab.M を陽性と判定したと考えられる。ただし、有意水準を 0.0001 とすると、ダネット&HU/HD 法は Lab.M を陰性と判定し、新西山手順は陽性と判定しているのので、表 4 の検出力による検討で 2 方法の性能が似通っていることから、マーゴリンの手順を含む新西山手順のほうが、頭打ち現象が多く観察されるような試験では、よりロバストな判定法となることが期待できる。

有意水準 0.001 が十分適切な値であるか、新西山手順が最も適切な手法であるかは、より多くの実験データに適用してみても経験を積み重ねる必要があるが、そのためにも本論文での確率計算法は重要な道具になるものと考えられる。

5. 考察とまとめ

5.1 複合最大対比法と最大対比法

本研究の動機は、Nishiyama et al. (2003) が提案したデータ解析法の性質を検討することであった。その過程で、解析法の性質は、本論文で複合最大対比法と名付けたものに一般化して考えると分かりやすい、という判断に至った。この概念は本論文で初めて提案したものであり、複数の最大対比法のすべてで有意差が認められたときに、包括的帰無仮説を棄却するという手法である。

複合最大対比法が最大対比法の枠組みに入らないのには, 2つの側面がある. 1つは, 最大対比法が複数の対比統計量の1つでも大きな値を取るものがあると, 包括的帰無仮説を棄却するのに対し, 複合最大対比法は個々の最大対比法のすべてで有意差が認められたときのみ, 全体として有意差があると判定する. 前者は「または」であるのに対し, 後者は「かつ」である. このため, 帰無仮説の受容域 (棄却域の補集合) が多角形にはなるが凸集合にならない. これは積分領域の設定を複雑にしている.

もう1つは, 複合最大対比法を構成する最大対比法に別々の名義有意水準・棄却限界値を設け得ることである. これによって手法の幅が広がり, 試験法ごとにかなり違った手法を構成できることになる. ただし, それを全く自由にすると手法の選択に自由度が生じすぎるので, 本論文では, 複数の最大対比法の有意水準に一定の関係を与えることでその自由度を1つにする, という方針を採用した. もちろんその関係をどう定めるかにもある種の自由度があるが, これは統計学の問題ではなく, 手法を適用する臨床試験・毒性試験でどのようなものをシャープに検出したいかという問題である.

5.2 複合最大対比法と多重対比法

本論文が取り上げた手法に対して, Genz and Bretz (1999) や Stewart and Ruberg (2000) は多重対比法 (multiple contrast method) という名前を用いて各種の議論を行っている. これに対して著者らは, あえて複合最大対比法という名前を用いることにした. これは同じ手法に別の名前を付けたのではなく, 多重対比法のうちの特別な形式のものを対象にしているためである.

その違いは, 複数の対比統計量を利用するとき, 対比統計量の最大値を検定統計量にするかどうかである. たとえば第4.3節で取り上げた新西山法の最初のマーゴリン手順は, 多重対比法には含まれるが最大対比法ではなく, したがって複合最大対比法の枠内に収まらない. 対比をある順番で, 閉手順的・逐次的に利用しているのであって, 最大値を検定統計量に使っていないからである.

実際, 式(13)の定義行列に対応する対比統計量が有意であれば, 頭下がり現象は無かったものとして次の段階に進むので, 式(14)の定義行列に対応する検定は実施しない. したがってマーゴリン手順の部分は複合最大対比法の枠の外になる. 新西山手順の全体は, マーゴリン手順法の結果に従属させて最大対比法を変えるから, 互いに独立な最大対比法を複数組み合わせるという複合対比法の枠に収まらない.

5.3 複合最大対比法と p 値の関係

複合最大対比法では, 検定統計量が複数あるので帰無仮説の分布での検定統計量の実現値より極端な領域の確率という p 値の考え方が当てはまらない. 複合最大対比法は含まれる複数の最大対比法が有意であるかどうかによって判定する意思決定法である.

通常の p 値のような意味は持たないが, 仮に全体の有意水準に対して有意であるかを判定する指標を作るならば, 与えた有意水準の関係 $\alpha_1 = r\alpha_2$ に矛盾しない以下の確率 p が考えられる. ただし, 以下の記述法では $r \geq 1$ とする.

観測データに基づく2つの最大対比法の p 値, p_1, p_2 に対して, $\Pr(t_{\max}^{(1)} > t_{\max}^{(1)'}) = \max(p_1, rp_2)$, $\Pr(t_{\max}^{(2)} > t_{\max}^{(2)'}) = \max(p_1, rp_2)/r$ となる $t_{\max}^{(2)'}, t_{\max}^{(2)'}$ を新たに定めると, 包括的帰無仮説の下で

の確率

$$p = P(t_{max}^{(1)} > t_{max}^{(1)'}, t_{max}^{(2)} > t_{max}^{(2)'} | H_0) \quad (19)$$

は矛盾無く ($p_1 < \alpha_1, p_2 < \alpha_2$) $\iff p < \alpha$ となる.

5.4 プログラム

ホームページ“www.rs.kagu.tus.ac.jp/yoshilab/iyaku/top.html”に展示したプログラムは, Kirkland ed., (1989) が提案している各種の手法あるいは, Hayashi et al., (1989), Kim et al., (2000a; 2000b), Omori et al., (2002) などが提案している他の複合最大対比法にも修正して利用することが可能である.

謝 辞

本論文の完成には査読者及び編集理事からの示唆・助言によるところが多い。ここに記して心から感謝の意を表する。

参考文献

- Dunnett, C. W. (1955). A multiple comparison procedure for comparing several treatments with a control. *Journal of the American Statistical Association*, 50, 1096-1121.
- Genz, A. and Bretz, F. (1999). Numerical computation of multivariate t-probabilities with application to power calculation of multiple contrasts. *Journal of Statistical Computation and Simulation*, 63, 361-378.
- Hayashi, M., Yoshimura, I., Sofuni, T. and Ishidate Jr., M. (1989). In vivo rodent erythrocyte micronucleus assay. *Environmental and Molecular Mutagenesis*, 13, 347-356.
- Kim, B. S., Cho, M. and Kim, H. J. (2000a). Statistical analysis of in vivo rodent micronucleus assay. *Mutation Research*, 469, 233-241.
- Kim, B. S., Zhao B., Kim, H. J. and Cho, M. (2000b). The statistical analysis of the in vitro chromosome aberration assay using Chinese hamster ovary cells. *Mutation Research*, 469, 243-252.
- Kirkland, D. J. ed. (1989). *Statistical Evaluation of Mutagenicity Test Data*. Cambridge: Cambridge University Press, pp. 102-140.
- 永田靖, 吉田道弘 (1997). 統計的多重比較法の基礎. サイエンティスト社.
- 日本工業標準調査会標準部会環境・試験循環専門委員会 (2002). TR Z 0023 化学物質のがん原性を予測するための BALB/c 3T3 細胞を用いる短期 2 段階形質転換試験. 日本規格協会.
- Nishiyama, H., Omori, T. and Yoshimura, I. (2003). A composite statistical procedure for evaluating genotoxicity using cell transformation assay data. *Environmetrics*, 14, 183-192.
- 西山智, 柳原宏和, 吉村功 (2003). 最大対比法を活用するための SAS/IML プログラム. 計量生物学, 24, 57-70.
- Jpn J Biomet Vol. 25, No. 1, 2004

- Omori, T., Honma, M., Hayashi, M., Honda, Y. and Yoshimura, I. (2002). A new statistical method for evaluation of L5178Ytk+/- mammalian cell mutation data using microwell method. *Mutation Research*, **517**, 199-208.
- Simpson, D. G. and Margolin, B. H. (1986). Recursive nonparametric testing for dose-response relationship subject to downturns at high dose. *Biometrika*, **73**, 589-596.
- Stewart, W. H. and Ruberg, S. J. (2000). Detecting dose response with contrasts. *Statistics in Medicine*, **19**, 913-921.
- Tsuchiya, T., Umeda, M., Nishiyama, H., Yoshimura, I., Ajimi, S., Asakura, M., Baba, H., Dewa, Y., Ebe, Y., Fushiwaki, Y., Hamada, S., Hamamura, T., Hayashi, M., Iwase, Y., Kajiwara, Y., Kasahara, Y., Kawabata, M., Kitada, E., Kubo, K., Mashiko, K., Miura, D., Mizuhashi, F., Mizuno, F., Nakajima, M., Nakamura, Y., Nobe, N., Oishi, H., Ota, E., Sakai, A., Sato, M., Shimada, S., Sugiyama, T., Takahashi, C., Takeda, Y., Tanaka, N., Toyozumi, C., Tsutsui, T., Wakuri, S., Yajima, S., Yajima, N. (1999). An interlaboratory validation study of the improved transformation assay employing Balb/c 3T3 cells: Results of a collaborative study on the two-stage cell transformation assay by the non-genotoxic carcinogen study group. *ATLA*, **27**, 685-702.
- Williams, D. A. (1971). A test for differences between treatment means when several dose levels are compared with a zero dose control. *Biometrics*, **27**, 103-117.
- Yoshimura, I., Wakana, A. and Hamada, C. (1997). A performance comparison of maximum contrast methods to detect dose dependency. *Drug Information Journal*, **31**, 423-432.

分担研究報告所

「バリデーションデータの統計解析」（平成 16 年度研究）報告

分担研究者：大森崇

研究要旨

- ・ **【背景と目的】**バリデーション研究では、複数の施設が同じ手順に従い実験を行うので、結果を報告する際には、データの記載方法を統一する必要がある。しかしながら、この認識は十分に理解されていない。日本で過去に実施された眼刺激性代替法バリデーション研究では、データクリーニング作業に 2 年の歳月を費やす結果となった。近年行われた光毒性試験代替法バリデーション研究ではこの経験を反映させて、研究前からデータクリーニングの準備が行われた。この研究では、光毒性試験代替法バリデーション研究のデータクリーニングで、どのような準備を行ったかを示すこと、眼刺激性試験代替法のバリデーション研究との結果を比べること、結果を踏まえて今後の代替法バリデーション研究のデータクリーニング作業への提言を行うことを目的とする。
- ・ **【方法】**試験前にデータ記入のためのデータシートファイルを作成することにした。これらのデータシートファイルをチェックし、誤りや不明な点がある場合は問い合わせを行いこととした。問い合わせの項目を眼刺激性試験代替法バリデーション研究のデータクリーニングと同じ項目を用いることで、効果を計ることとした。
- ・ **【結果】**データシートファイルは眼刺激性代替法バリデーション研究のデータクリーニング結果の考察を生かし作成された。結果は、眼刺激性試験代替法バリデーション研究での提出ファイル数に対する問い合わせのベファイル数が 1.13 倍（1535/1742）であったのに比べ、光毒性試験代替法バリデーション研究では 0.53（222/420）であった。また、問い合わせ項目の割合は 2 つの研究で大きく異なっていた。
- ・ **【結論】**上記の結果は、2 つの研究のさまざまな違いを考慮しても、バリデーション研究を開始する前から研究の計画段階から準備を行うことが有効であることを示している。バリデーション研究を実施する際には、データファイルシートを工夫する、技術移転でデータの入力についての説明する、誤解のないようにプロトコル・実施手順書・データファイルシートを吟味する、データクリーニングの手順を確立する、などのことが必要であることが参加者全員に広く認識されるべきである。

A. 研究目的

本研究は、動物実験代替法開発におけるバリデーション研究（以下、バリデーション研究）にお

いて、現時点において具体的に必要とされている統計解析における諸問題について検討を行う。この検討は、3 年間でデータマネジメント、データ解

析方法、研究の計画について検討を行う予定であり、本年度はデータマネジメントについての検討を行ったのでその結果を報告する。

通常、バリデーション研究の研究手順は、新たに開発された試験方法について、試験方法の手順書が作成され、その手順書に基づき *in vivo* 試験の成績が知られている複数の物質を用いて、複数の施設で実施した試験結果の再現性や *in vivo* 試験の予測能力などが検討される。これらの検討は、複数の施設で行われた試験結果を一箇所の施設に集めて評価されることになる。もしも複数の施設で実施する研究ではなく、単に個々の施設のみでの試験結果を報告する研究であるならば、試験結果として得られるデータの扱い、つまり、測定値の記入方法やまとめ方などは、施設独自のものであっても特に大きな問題は生じないであろう。しかしながら、バリデーション研究を行う場合には、複数の施設が共通の試験を実施するため、試験手順書のみならず、得られるデータの記録方法なども統一する必要がある。

残念なことにデータの記録方法の統一が重要であることは十分には認識されていない。このことを示す一つの例は、過去に日本で実施された47の試験実施施設が参加した大規模な眼刺激性試験代替法のバリデーション研究であろう。この研究では、事前におおまかな記録方法が設定されていたにもかかわらず、個々の試験実施施設から提出されたデータは、記載方法が統一されていない、必要なデータが記載されていないといったものであつた。これらのデータを、最終的な結果を導くためのデータベースを作成するために、この研究の研究者は2年以上のデータクリーニング作業を費やすことになった。この経験は、施設から集められるデータに誤りがなく、集められたデータからデータベースを作成するまでに実施するデータクリーニングをいかに効率よく実施するかが、バリ

デーション研究を実施する際の一つの課題であることを示しているといえる。

近年多くの代替法のバリデーション研究の結果が報告されているが、報告の中にデータクリーニングやデータマネジメントについて記載されているものは少なく、記述してある場合でも詳細は記載されていることはない。また、わが国で実施された眼刺激性試験代替法のバリデーション研究以外でデータクリーニングやデータマネジメントについて記載している文献はほとんどない。

現在、わが国では光毒性試験代替法のバリデーション研究が行われ、その研究成果は厚生労働科学研究報告書(大野(2004))で報告されているが、このバリデーション研究では、過去に実施された眼刺激性試験代替法のバリデーション研究の経験を踏まえ、施設から集められるデータに誤りを少なくし、集められたデータからデータベースを作成するまでに実施するデータクリーニングの効率を高めることが課題の一つとされていた。

本研究では、

- ・ 光毒性試験代替法で、眼刺激性試験代替法のデータクリーニングの経験を反映させた点を明示すること、
- ・ このバリデーション研究におけるデータクリーニングの結果を眼刺激性試験代替法バリデーション研究のデータクリーニングの結果と比較した結果を示すこと、
- ・ さらに、今後のバリデーション研究でのデータクリーニングのあり方について提言を行うこと

を目的とする。

B. 研究方法

B.1. 眼刺激性試験代替法バリデーション研究のデータクリーニングの要約

試験の概要やデータクリーニングの詳細はすで

に報告されている(Ohno ら (1995)、Omori ら (1995))。ここではこれらの文献の要約を示す。

先に示した眼刺激性試験代替法のバリデーション研究では、7 種類の被験物質を用いて、16 種類の細胞毒性試験が評価された。試験実施施設数は 47 施設であった。ただし、すべての施設がすべての代替法を評価したわけではない。物質ごとに用量と吸光度を含む試験結果が記録された電子媒体のデータファイルが各施設から提出され、全部で 1,535 のデータが集められた。データクリーニングにより、何らかの問合せや訂正を行うこととなったファイル数は、のべ 1,742 データファイルとなった。

多くのデータファイルが問合せや訂正となった理由を探るために、問合せや訂正の内容は以下に示す項目に分類された。

- (1) 細胞毒性の指標を計算するために必要なデータが入力されていない (44%)
- (2) 規定したフォーマット形式にデータが入力されていない (19%)
- (3) プロトコルや事前に決められたルールに適合していない (17%)
- (4) 単純な入力ミス (7%)
- (5) 試験実施施設からの誤りであったとの報告 (6%)
- (6) 紙媒体で提出されたデータと入力された結果が一致しない (5%)
- (7) その他 (2%)

() 内のパーセントは、のべ問合せ数についての割合を示しており、このうち(1)から(3)までの内容が、全体の約 80%を占めていた。さらに、それらの項目が多くなっている原因として、

- ・ 個々の試験法に適したデータファイルのフォーマットがきちんと定められていない、
- ・ データシートの入力方法に多くの誤解があった、

- ・ プロトコルで用いられた用語が誤解されやすく、試験によりその定義が異なっていた、
- ことなどが考察されている。

B.2. 光毒性試験代替法のバリデーション研究の各試験方法

新たに実施された光毒性試験代替法のバリデーション研究の対象となっている試験法は、酵母光生育阻害試験と赤血球光溶血試験の 2 つの試験法を実施して、両方の結果に基づき光毒性を判定するという方法である。したがって、2 種類の試験法がすべての試験実施施設数で実施される。試験実施施設は 6 施設で、9 種類の被験物質が評価された。この研究では研究計画書が作成されており、ここでは、各施設が 1 回の試験結果を 2 回以上実施して提出することとなっている。個々の試験法については、Sugiyama ら (1994a)、Sugiyama ら (1994b)、Sugiyama ら (2002)、大野 (2004) に詳細が記載されているが、データクリーニングの説明と、バリデーション試験の研究手順書での決められていたこととして必要となる最小限の記載を以下に示す。

酵母光生育阻害試験

酵母を用いて試験法であり、照射、非照射それぞれの条件下で、濾紙に含ませた被験物質の曝露により酵母の寒天培地上で得られる阻止帯の長さが試験の評価項目となる。この試験では、1 物質あたり 4 濃度が、溶媒対照、陽性対照とともに 6 穴プレートが用いられる。これらが 2 系列で実施され、それらについてそれぞれ照射した場合と非照射の場合とで行われるため、1 回の試験で 6 穴プレートは 4 枚分の結果が得られることになる。

赤血球光溶血試験

赤血球を用いた試験法であり、照射、非照射の

それぞれの条件下で、完全溶血に対する被験物質投与により溶血した割合が評価項目となる。この試験では1物質あたり4濃度が、溶媒対照、完全溶血、PBS（-）添加対照とともに実施される。24穴プレートが使用されるため、最大で3物質を同時に実施することができる。これらが照射、非照射についてそれぞれ2系列で実施される。溶血度の測定は、上清を96穴プレート2枚に複製して移し、吸光度計により行われる。24穴プレート1枚分の照射、非照射の結果を96穴プレートで測定するため、1回の試験で96穴プレートは4枚分の結果が得られることになる。このバリデーション研究では、540nmの波長で測定された吸光度が主解析として採用されることになっているが、525nmかもしくはそれに近い波長も同時に測定され報告することとされていた。

B.3. データファイルシートの作成と説明

先に記したように、光毒性代替法バリデーション研究では、短期間で研究結果を示すようにするために、眼刺激性試験代替法バリデーション研究の経験を反映させるようなデータ収集方法が必要であった。このため、実験結果を記入する各施設の実験実施者が、誤りや誤解することなくデータを記入するための電子媒体のデータシートファイルを作成した。

データシートファイルについて実験実施者に理解が得られるように作成したデータシートファイルとその入力方法をバリデーション研究の技術移転の際に時間を設けて各試験実施者に説明した。また、各施設に入力用のデータシートファイルを配布する際に、入力例を示したデータシートファイルを配布した。このバリデーション研究では、技術移転に参加しないものは試験を実施する資格を与えられなかった。

B.4. データクリーニングの範囲と手順

光毒性試験バリデーション研究では、データが正しく入力されているかどうかを確認するためにデータシートファイルとは別にデータシートファイルに入力する基となったデータの紙媒体での提出を求めた。これら紙媒体のデータは試験実施中に試験実施者が測定した値の記入や、吸光度計から得られるプリントアウトである。酵母光生育阻害試験で阻止帯の長さを測定するノギスや、赤血球光溶血試験で用いられる吸光度計の種類は施設により異なるため、紙媒体の資料を統一することはしなかった。データクリーニングは、紙媒体に基づき、データシートファイルの入力項目の中の被験物質名、本試験・予備試験の有無、試験回数、系列、測定値、吸光度の波長（赤血球光溶血試験の場合のみ）がチェックの対象とされた。データシートファイルで入力が必要な施設名、実施者、溶媒名、用量の値などはデータファイルに入力されているかはチェックの時点で確認するものの、紙媒体との一致を求めるまでのチェックは行っていない。

眼刺激性試験代替法のバリデーション研究では、データクリーニングの手順について特に事前に定められた規定はなかった。この研究ではデータクリーニングの作業を効率よく行うために、施設からのデータが提出される以前にデータクリーニングの手順を確立した。この手順では、原則として施設から届いたデータシートファイルのクリーニングを1週間以内に実施すること、疑問が生じたデータは電子メールで確認を行うことなどが決められた。また、データクリーニングのための各データシートファイルシートのチェックリストを作成しチェックを行った。

B.5 2つのバリデーション研究のデータクリーニングの比較検討法

上記のように準備をしたデータクリーニングの
手順が有効か否かを確認するために、この研究で
は、限刺激性試験代替法バリデーション研究のデ
ータクリーニング作業で用いられた分類に沿った
以下の項目に該当したデータファイルのベファ
イル数を調べることにした。これにより、限刺激
性試験代替法バリデーション研究の時のデータク
リーニング作業で生じた問題の内容の質がこのバ
リデーション研究でどのように異なっているかを
おおまかに比較することができる。そこで、こ
の研究では各問い合わせの以下の項目に分類した。
すなわち

- (1) 何らかの必要なデータが入力されていない
 - (2) 規定したフォーマット形式にデータが入力されていない
 - (3) プロトコルや事前に決められたルールに適合していない
 - (4) 単純な入力ミス
 - (5) 試験実施施設からの誤りであったとの報告
 - (6) 紙媒体で提出されたデータと入力された結果が一致しない
 - (7) その他
- である。

度数は、問い合わせ回数とその問い合わせに該当するファイル数とする。後者は限刺激性試験代替法バリデーション研究の場合の数え方に対応している。

C. 研究結果

C.1. 作成したデータファイル

光毒性試験バリデーション研究で作成したデータファイルについて記載する。

データシートファイルは Microsoft 社の Excel2002 を用いた。Excel を用いた利点は、このソフトウェアの(1)汎用性一どの試験実施施設でも利用することができ操作に慣れていること、(2)その機能一シートの保護、入力規制、表計算機能、

によっている。

限刺激性試験代替法バリデーション研究の経験を具体化するために、以下の点を考慮してデータシートファイルを作成した。

- (1) 酵母光生育阻害試験と赤血球光溶血試験のそれぞれ異なるデータシートファイルを作成した。
 - ・どちらの試験のデータシートファイルも、1回の試験の1系列分の照射と非照射について1データシートファイルに入力されるようにした。
 - ・1つのデータシートファイルには、施設名、被験物質、溶媒対照、濃度、測定値などの情報を入力するようにした。
- (2) 試験実施者がデータシートファイルのどの位置に何を入力するのかということを理解しやすいようにした。
 - ・試験をしている状態に合うように、プレートの形に沿った入力方法とした。
 - ・入力に必要な箇所は色をつけた。
 - ・マウスを近づけると何を入力すべきかのコメントが表示されるようにした。
- (3) 誤った入力や予期しない入力をなくすようにした。
 - ・施設名や溶媒など事前に規定されているものは、選択するようにして、入力できないようにした。
 - ・数値の入力のところには数値以外は入力できないようにした。
 - ・入力する必要のない場所には入力できないようにした。
 - ・シートが改変されないように注記した。
- (4) 各試験実施者による試験の結果の確認ができるようにした。
 - ・同じファイルの別のワークシートに照射と非照射の結果から得られる光毒性の判定が示されるようにした。

データシートは技術移転の際に説明された。この際、赤血球光溶血試験のデータシートファイルは、各施設の吸光度計の出力をカットアンドペーストできるように改変することが提案され、改変がされた。

最終的に使用されたデータファイルシートを図1、図2に示す。

<図1 酵母光生育阻害試験のデータシートファイル>

<図2 赤血球光溶血試験のデータシートファイル>

C.2. 提出されたデータファイル数

提出されたデータシートファイル数を表1に示す。

<表1 提出されたデータシートファイル数>

酵母光生育阻害試験と赤血球光溶血試験で、データファイル数が異なる理由は、試験回数はそれぞれの試験法で統一するという研究計画書になっていないこと、赤血球光溶血試験では、1枚のデータシートファイルに最高3物質分のデータを入力することが可能なこと、赤血球光溶血試験では540nmの波長で測定された吸光度のデータと、525nmもしくはそれに近い波長で測定された吸光度のデータは別に入力されるためである。

C.3. データクリーニング作業の結果

眼刺激性試験代替法バリデーション研究のデータクリーニング項目について、問い合わせ回数と該当データファイル数を表2に示す。

<表2 各データクリーニング項目における問合

せ回数と該当データファイル数>

表1と表2からわかるように、光毒性代替バリデーション研究で集められたデータファイル数は合計420ファイルであり、問合せをしたのべデータファイル数(表中II)が222ファイルであった。

眼刺激性試験代替法試験では、集められたファイル数に比べ問い合わせのべファイル数が上回っていたことと比べると問い合わせの数は減らされていることがわかる。また、特に眼刺激性試験代替法のバリデーション研究で問題となった項目は少なくなっており質的に問い合わせの内容が異なった結果となっていたと考えられる。

特に、「(7) その他」とした項目の問い合わせファイル数が多くなっている。この項目の具体的な内容を表3に示す。

<表3 データクリーニングで問い合わせた内容>

表3の内容から、データシートファイルの設計段階では予想しきれなかったことが生じていることがわかる。これらは今後のバリデーション研究に生かされるべきであろう。

D. 議論

D.1. 眼刺激性試験代替法バリデーション研究におけるデータクリーニングの内容との相違

本研究の結果は、眼刺激性試験代替法バリデーション研究でのデータクリーニングで作成された分類に基づき、光毒性試験代替法バリデーション研究のデータクリーニングの結果をまとめることで、データシートを作成した効果を評価している。結果は、事前にデータシートファイルを作成したことで、問い合わせの割合は減少し、眼刺激性試験代替法で問題となった問合せ項目を減少させてい

ると考えられる。ただし、2つの研究を比較する場合にさまざまな状況がかなり異なっていることに注意する必要があるであろう。

- ・ 時間の違いを含めた環境要因（例えば、眼刺激性試験代替法バリデーション研究を行った時には Excel の普及ほとんどの施設で用いていなかった）
- ・ 試験実施施設の数
- ・ 試験方法の違い
- ・ データクリーニングの範囲

などである。したがって、本稿では、データクリーニングの効率は何%改善したという定量的な議論を行うことはしていない。上記の要因をそろえた公正な比較を行うには、実験的な研究を行うことであろうが、そのような研究を実施することは、現実には困難であるだけでなく、実際の研究の場で生じる諸問題を反映できるとも思えない。したがって、この研究では、様々な要因が異なるということも考慮した上で、眼刺激性試験代替法バリデーション研究の経験を生かしたデータシートファイルの設計とその効果を評価せざるをえない。上記のような限界を考慮したとしても、「(1) 必要なデータがない」ということが一切なく、「(2) 規定したフォーマットで送付されていない」というものがわずか1ファイルのみ（誤って送信された）であったことは、試験開始前からデータクリーニングを意識し検討を行い、その準備をすることがバリデーション研究を実施する際に重要であることを示していると考えられる。

D.2. データクリーニング作業の効率

眼刺激性試験代替法バリデーション研究で2年以上の時間を費やしたデータクリーニング作業をいかに効率よく短時間で行うかは、光毒性試験代替法バリデーション研究を始める段階で大きな課題であった。この点を改善するために、我々は事

前にデータクリーニングの手順を確立した。施設から提出されるデータにどのような疑問が生じるかは不明であったが、眼刺激性試験代替法のバリデーション研究で報告された問合せの項目から想定される疑問や問合せ事項を予測したものから作成した、1データファイルに1枚の簡単なチェックシートを用いてデータクリーニングの作業を進めた。各施設から送られてきた試験結果の時間差があったこともあり、送付されてから1週間以内に送付されたデータシートファイルのデータクリーニングに着手ができた。また、問合せの返答の2日以内に確認や訂正作業を実施することができた。手順の確立とチェックリストの作成の効果を比較することは不可能であるが、データクリーニング作業の効率という点についてかなり貢献したと考えている。このことは、第三者が研究の評価を行う際に実際に行ったデータクリーニング作業の範囲を示すこともできる利点もある。一方で、チェックリストの作成は、チェックリストにない項目の誤りを見落とす可能性がある。しかし、データクリーニング作業が長くなれば、担当者が担当を変わってしまい、連絡がつかなくなるなどということも起こりうる。また、いかなる場合においても完全なチェックなどというのは不可能であるのだから、長い時間をかければよいということは決してないであろう。

作業の確立やチェックリストを作成するためには、どのような誤りが生じうるかということ予測する必要がある。そのためには、過去の研究の経験やその研究の試験実施計画書の把握は重要である。研究の参加者全員のデータクリーニングへの認識が必要である。

D.3. 試験の複雑さの影響

16試験を評価した眼刺激性試験代替法バリデーション研究では、複雑な試験法ほどデータクリー

ニングで問合せや訂正を行うデータシートファイル数が多い傾向があった。このバリデーション研究でも、酵母光生育阻害試験に比べて赤血球光溶血試験は操作やデータの入力方法が複雑である。これは表 2 の問合せの回数やデータシートファイル数でもその傾向がうかがえる。特に、赤血球光溶血試験のプロトコル逸脱であったデータシートファイル数は 48 ファイルあったが、これは 1 施設がプロトコルを誤解したために生じたものであった。広く普及する代替法を考える場合には試験法の簡易さというは重要な要因となるであろうが、これはデータの質の確保という点からも認識されるべきである。

研究計画書や試験実施手順書などは、誤解がなく、理解しやすくなるように再考を重ねるべきである。試験計画時に、そのための時間を確保するように心がけるべきである。

D.4. 改善すべき課題

データクリーニング作業における我々の最終的な目標は、眼刺激性試験代替法バリデーション研究に比べて何%問い合わせ数が改善したかではなく、問い合わせ数がなくなるようにするような方法論を構築していくことであろう。本稿で記したような計画段階からデータクリーニングを考慮した光毒性代替法バリデーション研究でも、データクリーニング作業で完全には問合せや訂正のデータファイルをなくすことができていない。この研究のデータクリーニング作業を通して、今後の研究で参考になるであろう、改善できる点を以下に考察する。

データシートファイルへの対応

赤血球光溶血試験では、3 つの被験物質を同時に試験することが可能である。この研究では、試験をするプレートを意識してデータシートファイル

を作成したため、1 枚のシートで最高 3 物質分の結果を入力することを可能とした。しかし、1 物質のみしか入力しない場合の空いた欄についての処理をどのような入力とするかは規定していなかった。このため施設により、その施設には配布されない被験物質のコードを選択したり、ブランクの値を入力したりするという異なる対応が取られ、問合せの対象となっている。データシートファイルの設計時に被験物質が 1、2 物質の場合の入力方法を考えておくべきであった。

試験実施手順書の用語

データクリーニング作業を通じて、プロトコルからの逸脱が 1 つみつかった。赤血球光溶血試験で 24 穴プレートから 96 穴プレートに移す際に 2 系列に移すことが実施されていなかったのである。赤血球光溶血試験の実実施手順書には、24 穴プレートで試験を行う際に duplicate で実施する旨が記載されており、さらに 24 穴プレートから 96 穴プレートに移す際も duplicate という用語が用いられていた。同じ用語を使用した施設の試験実施者にプロトコルからの逸脱を起こさせる要因となった可能性は大きい。また、ある施設は系列という用語の誤解をしていることがわかったため、データクリーニングの段階で入力しなおす作業が生じた。これは、上記の他に実施手順書で「duplicate」という記載がなされ、データシートファイルでは同じことを「系列」と記載していたことによるかもしれない。プロトコル等で用いる用語に注意すべきであることは眼刺激性試験代替法のバリデーション研究の反省としても報告されていることである。研究が始まる前の試験プロトコル、試験手順書の確認の際には、用語の定義や書類によって異なる用い方がされていないかも含めて検討されるべきであろう。