

平成 29 年度厚生労働科学研究補助金（化学物質リスク事業）

分担研究報告書

統計学的視点に基づく Multi-ImmunoTox Assay のデータ解析方法の 検討

分担研究者：大森崇

研究要旨

[背景と目的] Multi-ImmunoTox Assay(以下、MITA)は、化学物質がヒトの免疫系に与える影響を *in vitro* で評価することを目標とする試験法である。バリデーションにおける phase1 研究では MITA の判定方法が十分には確立していないため、3 種類の判定方法の候補（方法 1、方法 2、方法 3）が提案され、適用された。しかし、これら 3 種類の統計学的な性能評価は行われていなかったため、本年度の課題として、3 種類の判定方法についてシミュレーションに基づく性能評価を行った。

[方法] 方法 1 は個々の実験で判定を行い、独立した 3 回の実験の判定結果をもって最終判定を行う方法である。方法 2 は、独立した 3 回の実験から得られた測定値を統合したデータセットを作成し、このデータセットに基づきデータ解析を行った結果から最終判定を行う方法である。方法 3 は、方法 1 と同様に 3 回の実験を統合し、最終判定を行う方法であるが、方法 2 のように測定値を使うのではなく、得られる測定値を%換算した値を用いる方法である。いくつかの毒性パターンを想定して、シミュレーションによりこれら 3 種類の判定方法の性能評価を行った。

[結果] 実験間差や各濃度における繰り返しの誤差によらず、毒性がないパターンに対して、方法 1、2 では誤判定は約 5%程度だが、方法 3 では約 18%が誤判定であった。毒性があるパターンに対して、方法 3 は他の 2 方法に比べて謝る確率が高かった。

[結論] 統計学的な観点においては、方法 3 は、第一種の過誤を 5%に抑えることができ、他の 2 つの方法に比べて、様々な毒性パターンに対する高い検出力を有する方法である。

A. 研究目的

Multi-ImmunoTox Assay(以下、MITA)は、化学物質がヒトの免疫系に与える影響を *in vitro* で評価することを目標とする試験法である。MITA では、ある化学物質の IL2 の発現を評価は独立した 3 回の実験から得られる測定値を用いて行われ、immunosuppression、immunoaugmentation、no effect のいずれかの判定がされることが求められている。しかし、その判定方法が十分には確立し

ていなかったため、MITA のバリデーションにおける phase1 研究では、研究に入る前に 3 種類の判定方法により、データ解析が行われた。個々の判定の統計学的視点に基づく性能の良し悪しは、毒性がないものを誤って毒性ありとしてしまう確率がどの程度か、毒性があるものを正しく毒性ありと判定できる確率はどのくらいであるのかを知らずして評価可能である。しかし、これまでこれら 3 種類の判定方法統計学的な性能評価は行われてい

なかった。

本報告では、提案された 3 種類の判定方法について、シミュレーションに基づく性能評価を行った結果をする。

B. 研究方法

MITA のデータの特徴

MITA では、独立に実施された 3 回の実験の結果を用いて判定には濃度 0(DMSO)群と各濃度群との対比較により、最終的な判定を行う。1 回の実験には、96 穴プレートが用いられ、Control 群に加え、濃度 0(DMSO)、濃度 1~濃度 10 までの計 12 段階の濃度群の測定値が得られる。個々の濃度群では 4 回の繰り返しがある。

Phase1 研究で検討した 3 種類の方法について

Phase1 研究において検討をおこなった 3 種類の判定方法（方法、方法、方法）についてまとめる。

方法は個々の実験で判定を行い、3 回の実験の判定結果をもって最終判定を行う方法である。3 回の実験それぞれにおいて、濃度 0 群と各濃度群の間に何らかの統計的に有意な違いがみられるかどうかを判定した後、それら 3 回の判定結果をまとめることで濃度 0 群に対する化学物質の影響を評価する。

それに対して方法は、独立した 3 回の実験で構成されている試験法の特徴を重視し、3 回の実験から得られた測定値を統合したデータセットを作成し、このデータセットに基づきデータ解析を行った結果から最終判定を行う方法である。それぞれの実験の違いは、データセットに含める変数の 1 つとすることで、データ解析では実験間の影響の調整を行ったうえで判定を行う。

また、方法は、方法と同様に 3 回の実験を統合し、最終判定を行う方法であるが、方法の

ように測定値を使うのではなく、得られる測定値を%換算した値を用いてデータ解析し、判定を行う点で方法とは異なる。

各実験から得られる測定値について

MITA では、内部標準プロモーターである SLR、INF- プロモーターである SLO、IL-2 プロモーターである SLG が導入された細胞を使い、その発光の度合いで免疫の活性の程度が測定される。1 回の実験において、96 穴プレートの各セルから SLG-LA (SLG ルシフェラーゼ活性)、SLO-LA (SLO ルシフェラーゼ活性)、SLR-LA (SLR ルシフェラーゼ活性) の 3 種類の発光に関する測定値が得られる。化学物質の評価において、第 i 番目の実験 (i=1,2,3) の第 j 番目の濃度 (j=0,1,2,...,10) の第 k 番目の繰り返しの測定値をそれぞれ SLG_{ijk} 、 SLO_{ijk} 、 SLR_{ijk} とする。

これらの測定値を用いて、判定のために以下の 4 種類の値が必要となる。

$$I.I.SLR_{ij} = (\overline{SLR_{ij}}) / (\overline{SLR_{i0}}),$$

$$nSLG_{ijk} = SLG_{ijk} / SLR_{ijk},$$

$$\%supp_{ij} = \left\{ 1 - \frac{(\overline{nSLG_{ij}})}{(\overline{nSLG_{i0}})} \right\} \times 100,$$

$$\text{ただし、} \overline{SLG_{ij}} = \sum_{k=1}^4 SLG_{ijk} / 4,$$

$$\overline{SLR_{ij}} = \sum_{k=1}^4 SLR_{ijk} / 4 \text{ である。}$$

$I.I.SLR_{ij} = (\overline{SLR_{ij}}) / (\overline{SLR_{i0}})$ は細胞の状態が正常かどうかを判断するための指標である。 $nSLG_{ijk}$ は SLG_{ijk} をプレート内の各セルにおいて標準化した値である。また、 $\%supp_{ij}$ は各濃度での 4 回の繰り返しの平均値を使って算出され、各濃度での免疫毒性の程度を%換算した値である。

方法

方法は個々の実験で判定を行い、3 回の実験の判定結果をもって最終判定を行う方法である。

各実験において、 $I.I.SLR_{ij} \geq 0.05$ ($j=0, \dots, J$)をみ
たす濃度の $nSLG_{ijk}$ を以降のデータ解析に採用す
る。 $nSLG_{ijk}$ を用いて、いずれかの濃度群の間に
差があるかどうかを知るために有意水準 5%で
1way ANOVA を行う。この検定の帰無仮説は、「第
i 番目の変数におけるすべての濃度群における
nSLG の母平均は等しい」である。その結果、統
計的に有意な差がみられた場合は、濃度 0 群と各
濃度群の対比較である両側 Dunnett 検定を有意水
準 5%で行う。この検定の帰無仮説は「第 i 番目の
実験における濃度 0 群と第 j 濃度群のそれぞれにお
ける nSLG の母平均は等しい」である。

1way ANOVA と Dunnett 検定の結果に基づい
て、まず個々の実験の判定を suppression(-)、
augmentation(+), no effect(0)のいずれかとする。
3 回の実験の判定結果が 3 回とも suppression(-)
もしくは augmentation(+)で一致した場合は、最
終判定はそれぞれ immunosuppression、
immunoaugmentation となる。

3 回の実験の判定結果が一致しなかった場合は、
それぞれの実験で、絶対値が最大となる
 $\max_{0 \leq j \leq 10} |\%supp_{ij}|$ の値を算出し、それら 3 つの値
を用いて両側 t 検定を有意水準 5%で行う。この検
定の帰無仮説は「 $\max_{0 \leq j \leq 10} |\%supp_{ij}|$ の母平均は 0
である」である。負の方向に統計的に有意差があ
れば immunosuppression、正の方向に統計的に有
意差があれば immunoaugmentation、統計的に有
意差がなければ no effect と最終判定を行う。

方法

方法 は、方法 のように単純に 3 回の実験の
測定値から得られる結果を併合するのではなく、3
回の異なる実験の影響を調整した nSLG の平均の
差を用いて、濃度 0 群と各濃度群の対比較である

両側 Dunnett 検定を有意水準 5%で行うことで最
終判定を行う方法である。この検定の帰無仮説は
「第 i 番目の実験における濃度 0 群と第 j 濃度群の
それぞれにおける nSLG の母平均は等しい」であ
る。

3 回の独立した実験における $I.I.SLR_{ij} \geq 0.05$
($j=0, \dots, J$)をみたす濃度を算出し、3 回の実験のう
ち最も小さい濃度を J_{\min} とする。この方法では j
の範囲を 1 から J_{\min} となる $nSLG_{ijk}$ を用いる。

3 回の異なる実験の影響を調整した Dunnett 検
定は、一般線形モデル (GLM) を用いて計算する
ことができる。モデルを

$$nSLG_{ijk} = \mu + \alpha_i + \beta_j + \varepsilon_{ijk}$$

とすると、独立した 3 回の実験の影響を調整した
 $nSLG_{ijk}$ の平均の差の推定値は $\hat{\beta}_j$ ($j=1, \dots, J_{\min}$)と
して得ることができる。有意水準 5%の両側

Dunnett 検定は、 β_j に対して上記の GLM モデル
の誤差の自由度を伴う両側 Dunnett の棄却限界値
を利用して構成された同時 95%信頼区間用いて判
定を行うことが可能である。この同時 95%信頼区
間の下限が 0 より大きな場合に
immunoaugmentation、上限が 0 より小さい場合
は immunosuppression、同時信頼区間が 0 を含ん
でいる場合は no effect と最終判定を行う。

方法

方法 は、3 回の異なる実験の影響を調整した
 $\%supp_j$ の平均値の差を用いて、濃度 0 群と各濃
度群の対比較である両側 Dunnett 検定を有意水準
5%で行うことで最終判定を行う方法である。この
検定の帰無仮説は「第 i 番目の実験における濃度 0
群と第 j 濃度群のそれぞれにおける $\%supp_j$ の母
平均は等しい」である。方法 と同様に、一般線
形モデルを用いて計算することができる。

モデルを

$$\%supp_{.ij} = \mu^* + \alpha_i^* + \beta_j^* + \varepsilon_{ij}^*$$

とすると、独立した 3 回の実験の影響を調整した

$\%supp_j$ の平均の差の推定値は $\hat{\beta}_j^*$ ($j=1, \dots, J_{\min}$)

として得ることができる。有意水準 5% の両側

Dunnett 検定は、 β_j^* に対して上記のモデルの誤差

の自由度を伴う両側 Dunnett の棄却限界値を利用して構成された同時 95%信頼区間を用いて判定を行うことが可能である。この同時 95%信頼区間の下限が 0 より大きな場合に immunoaugmentation、上限が 0 より小さい場合は immunosuppression、同時信頼区間が 0 を含んでいる場合は no effect と最終判定を行う。

シミュレーション

シミュレーションは、 $nSLG_{ijk}$ に対して、免疫毒性を示さないパターンと、化学物質の濃度が増加するにつれて Immunoaugmentation としての毒性の傾向を示す 3 種類のパターンの計 4 パターンの期待値に、乱数による誤差を加えたデータに基づき行った。検討した 4 パターンをそれぞれパターン A、B、C、D とする。

それぞれの毒性パターンの期待値は、3 回の独立した実験間差である α_i と、各濃度における 4 回の繰り返しの誤差（以下、繰り返し誤差）の標準偏差 σ がそれぞれ小さい場合と大きい場合を想定し、方法 ~ で免疫毒性判定にどのような違いが生じるのかを検討した。シミュレーション回数はそれぞれのパターンに対して 10,000 回行った。誤差には平均 0、分散 σ^2 の正規分布に従う乱数を用いた。検討したパターン A からパターン D の詳細と期待値を表 1 に示す。

解析には、統計解析ソフト SAS ver. 9.4 (SAS Institute Inc., NC) を用いた。

表 1. 検討した毒性パターンとその期待値

パターン	期待値のモデル
A	毒性を示さない $E[nSLGi] = 5 + \alpha_i$
B	直線的にゆるやかに毒性が増加 $E[nSLGi] = 5 + \alpha_i + 0.3 \text{濃度}$
C	曲線的にゆるやかに毒性が増加 $E[nSLGi] = 5 + \alpha_i + 0.002 \text{濃度}^2$
D	曲線的に強く毒性が増加 $E[nSLGi] = 5 + \alpha_i + 0.05 \text{濃度}^2$

$E[nSLGi]$ は期待値を示す。各パターンそれぞれについて、実験間差が小さい場合と大きい場合をそれぞれ

$$\alpha_i = 0, \quad \sigma = 1, \quad \sigma = 2$$

$$\alpha_i = 0, \quad \sigma = 50, \quad \sigma = 100$$

と設定し、繰り返し誤差の標準偏差 σ が小さい場合と大きい場合をそれぞれ「 $\sigma = 0.08$ 」「 $\sigma = 0.8$ 」として検討した。繰り返し誤差の標準偏差の値は、実際の MITA のデータを参考に設定した。

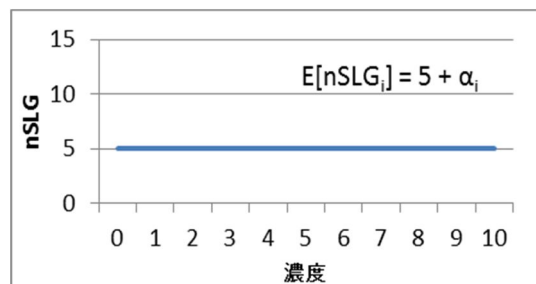


図 1. パターン A

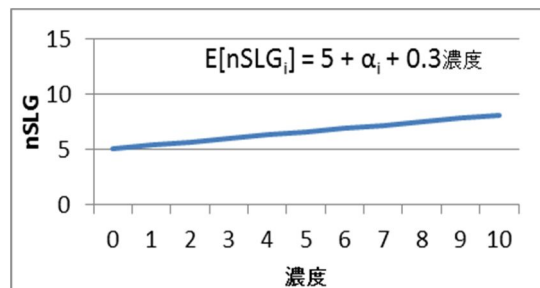


図 2. パターン B

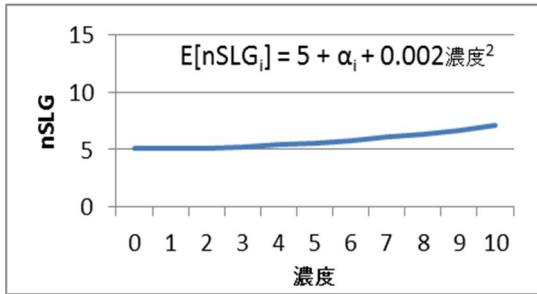


図 3 . パターン C

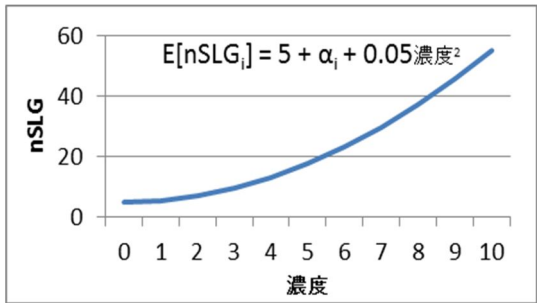


図 4 . パターン D

C . 研究結果

表 1 で示した各毒性パターンにおいて α_1 と α_2 の大きさの設定を変え、10000 回ずつ実施したシミュレーションによる各方法での判定結果を、毒性パターンごとに割合でまとめた結果を表 2~5 に示す。

表 2 . パターン A の判定結果

実験間差	誤差	判定	方法 (%)	方法 (%)	方法 (%)
$\alpha_1 = 0$	= 0.08	Immunosuppression	8.90	2.55	2.22
		Immunoaugmentation	9.26	2.76	2.53
		No effect	81.84	94.69	95.25
$\alpha_2 = 1$	= 0.8	Immunosuppression	7.88	2.39	2.11
		Immunoaugmentation	9.69	2.39	2.29
		No effect	82.43	95.22	95.6
$\alpha_3 = 100$	= 0.08	Immunosuppression	0	2.41	0.04
		Immunoaugmentation	0	2.65	0.06
		No effect	100	94.94	99.9
$\alpha_3 = 100$	= 0.8	Immunosuppression	0	2.46	0.08
		Immunoaugmentation	0	2.62	0.04
		No effect	100	94.92	99.88

表 3 . パターン B の判定結果

実験間差	誤差	判定	方法 (%)	方法 (%)	方法 (%)
$\alpha_1 = 0$	= 0.08	Immunosuppression	0	0	0
		Immunoaugmentation	100	100	100
		No effect	0	0	0
$\alpha_2 = 1$	= 0.8	Immunosuppression	0.47	0	0
		Immunoaugmentation	99.06	99.97	99.99
		No effect	0.47	0.03	0.01
$\alpha_3 = 100$	= 0.08	Immunosuppression	0	0	0
		Immunoaugmentation	100	100	0
		No effect	0	0	100
$\alpha_3 = 100$	= 0.8	Immunosuppression	0	0	0
		Immunoaugmentation	98.81	99.99	2.42
		No effect	1.19	0.01	97.58

表 4 . パターン C の判定結果

実験間差	誤差	判定	方法 (%)	方法 (%)	方法 (%)
$\alpha_1 = 0$	= 0.08	Immunosuppression	33.26	0	0
		Immunoaugmentation	49.34	99.57	99.35
		No effect	17.40	0.43	0.65
$\alpha_2 = 1$	= 0.8	Immunosuppression	12.68	1.43	1.24
		Immunoaugmentation	6.12	5.46	5.43
		No effect	81.20	93.11	93.33
$\alpha_3 = 100$	= 0.08	Immunosuppression	0	0	0
		Immunoaugmentation	49.12	99.57	1.90
		No effect	50.88	0.43	98.10
$\alpha_3 = 100$	= 0.8	Immunosuppression	0	1.58	0.03
		Immunoaugmentation	0	4.98	0.10
		No effect	100	93.44	99.87

表 5 . パターン D の判定結果

実験間差	誤差	判定	方法 (%)	方法 (%)	方法 (%)
$\alpha_1 = 0$	= 0.08	Immunosuppression	0.05	0	0
		Immunoaugmentation	99.95	100	100
		No effect	0	0	0
$\alpha_2 = 1$	= 0.8	Immunosuppression	1.13	0	0
		Immunoaugmentation	98.68	99.81	99.95
		No effect	0.19	0.19	0.05
$\alpha_3 = 100$	= 0.08	Immunosuppression	0	0	0
		Immunoaugmentation	99.96	100	0
		No effect	0.04	0	100
$\alpha_3 = 100$	= 0.8	Immunosuppression	0	0	0
		Immunoaugmentation	98.83	99.65	0.07
		No effect	1.17	0.35	99.93

D . 考察

毒性がない物質を誤って毒性ありと判定してしまう確率（第一種の過誤）

3 回の実験間差や各濃度における繰り返しの誤差によらず、毒性がないパターン A の場合は No effect と判定されることが望ましい。方法 1、2 では誤判定は約 5%程度だが、方法 3 では約 18%が誤判定であり、第一種の過誤確率が高いことがわかる（表 2）。方法 3 は 2 段階の判定としているため

に、検定を複数回行っていることにより、検定の多重性が生じた結果といえる。

濃度の増加に伴い毒性が明確になる物質を正しく検出できる確率（検出力）

実験間差が小さい場合、パターン B とパターン D の結果は、方法 1、2、3 のすべての方法において、毒性ありである Immunoaugmentation と判定されており、いずれの方法でも高い検出力を保持して判定できていることが分かった。パターン C においては、検出力の高さの視点からは方法 1 と方法 2 に違いはなかったが、方法 3 においては正しく Immunoaugmentation と判定できる割合は他の 2 方法に比べて小さかった（表 3、4、5 それぞれの上 2 段）。実験間差や誤差が小さくても、方法 3 は方法 1、2 に比べて検出力が低く、適切に判定できないことを示している。このことは、方法 3 において判定を行う際、3 回の各実験における個々の判定結果が一致せず suppression 側に偏った傾向にあれば、最終判定を行う際に Immunosuppression と判定されやすくなってしまふことが要因ではないかと考えられる。

次に実験間のばらつきが大きい場合に注目すると、パターン B、C、D のいずれの場合も、方法 3 を用いた時の判定結果がほぼ No effect となってしまうことがわかる（表 3、4、5 それぞれの下 2 段）。これは、方法 3 では応答が 0 から 1 の範囲に収まる $\%supp_j$ に基づき計算していることによる。 $\%supp_j$ の換算した値の利用は、実験施設ごとに測定機械が異なることによって生じる測定値のスケールの大きさの違いの影響を排除することができると思われる。しかし、この換算の定義より、常に各実験の濃度 0 群の値が常に 0% となる一方で、何らかの原因で実験間差が生じた場合に化学物質の濃度が増加するに従い実験間のばらつきは大きくなってしまふ。方法 3 は $\%supp_j$ を用いて判定を

行う解析に、線形モデルを想定しているため、この状況でモデルが適切とはいえない。この現象は実験間差が大きい場合に特に顕著になるであろう。シミュレーションで設定した実験間差の大きな状況は誤差の標準偏差と比べてもはるかに大きく、現実の実験では生じることがないような大きさであるものの、 $nSLG_{ijk}$ の実験間差が濃度によらず一定の大きさで生じるような場合には、方法 1 よりも方法 2 の方がより精度よく判定を行うことができるといえる。

本研究の限界

本研究の弱点は、シミュレーションで用いたデータは、方法 3 で使用したモデルと同じモデルを用いて発生させており、このことによって方法 3 の判定結果が他の方法よりも有利であることである。しかし、上記に示したように方法 3 は応答が 0 から 1 の範囲に収まる $\%supp_j$ に対して線形モデルを想定しており、想定するモデルの仮定は適切ではないであろう。

検討したシミュレーションで発生させたデータは、実際に実験により得られる SLG_{ijk} 、 SLO_{ijk} 、 SLR_{ijk} では無く、これから計算される $nSLG_{ijk}$ に基づいている。このため、本検討のシミュレーションには、I.I.SLR_{ij} に基づく判定が行っているわけではない。しかしながら、 $nSLG_{ijk}$ に基づくデータで検討を行うことにより、判定方法間の本質的な特徴を知ることができているといえる。

検討した判定方法が実際に毒性の判定に用いることができるかどうかは別の問題である。方法 3 は検出力が高い方法であることが本検討のシミュレーションにより明らかになったが、統計的検定のみに基づく判定のみで判定することが、毒性の判定として適切であるかどうかは別の問題である。毒性判定として実用化するためには、この分野の

毒性家との議論を必要とする。

なし。

E . 結論

MITA による化学物質の評価のために判定方法の候補となる 3 種類の方法の統計的性能をシミュレーションにより評価した。統計学的な観点においては、方法 1 は、第一種の過誤を 5%に抑えることができ、他の 2 つの方法に比べて、様々な毒性パターンに対する高い検出力を有する方法である。

謝辞：本検討を進めるにあたり解析の多くの部分に協力してくれた神戸大学医学研究科の小林真弓氏に感謝いたします。

F . 健康危険情報

G . 研究発表

なし。

H . 知的財産権の出願・登録状況

なし。

参考文献

- 1) Littell, R. C., Stroup, W. W., Freund, R. J. SAS for linear models. SAS, 2002.
- 2) 永田靖、吉田道弘. 統計的多重比較の基礎. サイエンティスト社, 1997.