

厚生労働科学研究補助金（化学物質リスク事業）

（総合）研究報告書

Multi-ImmunoTox Assay(MITA)における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。バリデーション研究を行うにあたり、マイクロソフト社の Excel で作成するデータシートに平均値の比の 95%信頼区間を実装する必要があった。その際に、Excel で利用できる関数が適用できず近似式が必要となった。また、バリデーション研究を行う時点では、MITA の判定方法が十分には確立しておらず、判定に関する方針には 3 つの候補が考えられた。

[方法] Excel 上のデータシートに導入する平均値の比の信頼区間の検討として、これまでに実施された計 4168 の実データを用い、山内の近似式による 97.5%点と比の 95%信頼区間の下限の値を統計解析ソフト R で算出したこれらの値と比較することを行った。

MITA の 3 つの判定方法に関して、以下の方法 から方法 について検討を行った。方法 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 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 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%換算した値を用いる方法である。MITA のバリデーション研究により得られた実データを用いて、施設内および施設間再現性の検討を行った。また、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] 平均値の比の信頼区間の検討に用いた実データにおける小数自由度の範囲は 3~6 であり、この範囲においてパーセント点も 95%区間の下限も R と山内の近似のどちらもほぼ同様な値を取っていることがわかった。

判定方法に関する実データに基づく検討では、施設間再現性の結果はどの方法も同じであった。施設内再現性は、要約結果だけを比べると方法 が方法 、 に比べて若干優れているように思われるが、濃度反応関係を見る限り、方法 が優れているということではなさそうであった。

判定方法に関するシミュレーションに基づく検討では、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 、 では誤判定は約 5%程度だが、方法 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 は他の 2 方法に比べて謝る確率が高かった。

[結論] 平均値の比の信頼区間を Excel で算出するには、Excel の t 分布のパーセント点を計算する関数である t.inv 関数を用いるのではなく、山内の近似式である(1)式によってパーセント点を計算し、平均値

の比の95%信頼区間を得ればよい。

MITAによる化学物質の評価のために検討した3種類の判定方法については、方法が、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。しかし、実データに基づく検討からは検討した判定方法にはまだ検討の余地がある。

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。MITAのバリデーション研究の開始時点において、方針の異なる3種類の判定方法の候補の可能性が考えられた。

バリデーション研究では、各試験実施からMITAによって被験物質を適用した際のデータを収集するために、マイクロソフト社のExcelでデータシートを作成する必要がある。このデータシートにMITAのデータを入力すると、試験の成立条件が判定され、個々の判定結果を得られるようにすることが望ましい。判定には平均値の比の95%信頼区間を計算する必要があり、多くの統計パッケージがこの計算を行うが、Excelにはこの機能がないため、関数を組み合わせる必要がある。特に比の信頼区間を作成する際にExcelに実装されているt.inv関数がうまく利用できないため、初年度である2016年度は、山内の近似式として知られている近似式に基づく構成方法の実装可能性について検討を行うことを目的に研究を行った。

MITAのバリデーション研究が進み、各施設から得られたデータを用いて、データ解析を行う際には、研究者の間で議論された方針を解析方法として定式化し、実装させる必要がある。2017

年度は、3種類の方針に基づいた判定方法の定式化と実際のデータに適用した結果を報告することを目的に研究を行った。

さらに、それぞれの判定方法について、統計学的視点に基づき、誤って毒性ありとっ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という視点で評価するために、2018年度は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うことを目的に研究を行った。

B. 研究方法

B.1. 平均値の比の信頼区間の構成方法の検討 検討に用いたデータ

検討には、MITAのバリデーション研究のデータが得られていなかったために、IL-8 Luc assayのバリデーション研究を通して得られた379実験分データを用いた。この試験は0濃度と11の濃度の試験からなるため、計4168のパーセントが得られることになる。

山内の近似式

自由度 v の t 分布の97.5%点を $t_{0.975}(v)$ とすると、山内の近似式は $t_{0.975}(v)$ を

$$t_{0.975}(v) \approx u + \frac{y_1(u)}{v} + \frac{y_2(u)}{v^2} + \Lambda + \frac{y_5(u)}{v^5} \quad (1)$$

で近似する。ただし $y_1(u) = (u^3 + u)/4$,
 $y_2(u) = (5u^5 + 16u^3 + 3u)/96$,
 $y_3(u) = (3u^7 + 19u^5 + 17u^3 - 15u)/384$,
 $y_4(u) = (79u^9 + 776u^7 + 1482u^5 - 1920u^3 - 945u)/92160$,

$y_s(u) = (27u^{11} + 339u^9 + 930u^7 - 1782u^5 - 756u^3 + 17955u) / 368640,$
 $u = 1.96$ である。³⁾

パーセント点と95%信頼区間の下限の比較

1468 の比のデータについて、上記の式で得られる山内の近似式により算出される 97.5% 点の値と 95% 信頼区間の下限の値を、統計解析ソフトである R で計算されるそれぞれの値と比較した。

B.2. MITA の判定方法に関する検討

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。

バリデーション研究で検討した 3 種類の方法について

バリデーション研究において検討をおこなった 3 種類の判定方法(方法 1、方法 2、方法 3)について以下にまとめる。

方法 1 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あと、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 2 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 2 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%換算した値を用いてデータ解析し、判定を行う点で方法 1 とは異なる。%換算した値の利用は、実験施設ごとで測定に用いる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。

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。

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

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

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

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

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

$$I.I.SLR_{ij} = (\overline{SLR_{ij}}) / (\overline{SLR_{i0}}) \text{ は細胞の状態が正常か}$$

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_j ≥ 0.05 (j=0,...,J)をみたく濃度の 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_j ≥ 0.05 (j=0,...,J)をみたく濃度を算出し、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,..., 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。

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、Conc.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 (GLM) を用いて計算することができる。

モデルを

$$\%supp_{i,j} = \mu^* + \alpha_i^* + \beta_j^* + \varepsilon_{ij}^*$$

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、 $\hat{\beta}_j^*$ 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 より大きな場合には 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。

バリデーション研究のデザイン

バリデーション研究は、MITA の施設内及び施設間再現性を評価するために行われた。この研究では食品薬品安全性センター (FDSC)、産業技術総合研究所つくば本部 (AIST-Tsukuba)、産業技術総合研究所四国センター (AIST-Takamatsu) の 3 施設が試験実施施設として参加し、5 種類のブラインド化された化学物質分の試験が 3 セット (1st round、2nd round、3rd round) 行われた。

シミュレーション

シミュレーションは、 $nSLGi_{ijk}$ 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 ε_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 $E[nSLGi] = 5 + \varepsilon_i$
B	直線的にゆるやかに毒性が増加 $E[nSLGi] = 5 + \varepsilon_i + 0.3 \text{ 濃度}$
C	曲線的にゆるやかに毒性が増加 $E[nSLGi] = 5 + \varepsilon_i + 0.002 \text{ 濃度}^2$
D	曲線的に強く毒性が増加 $E[nSLGi] = 5 + \varepsilon_i + 0.05 \text{ 濃度}^2$

$E[nSLGi]$ 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

「 $\varepsilon_1 = 0, \varepsilon_2 = 1, \varepsilon_3 = 2$ 」

「 $\varepsilon_1 = 0, \varepsilon_2 = 50, \varepsilon_3 = 100$ 」と設定し、

繰り返し誤差の標準偏差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。

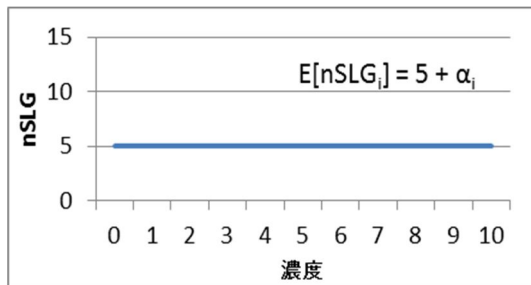


図 1 . パターン A

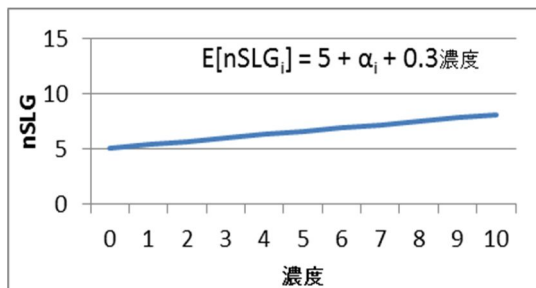


図 2 . パターン B

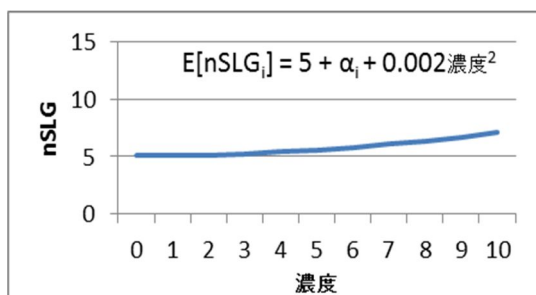


図 3 . パターン C

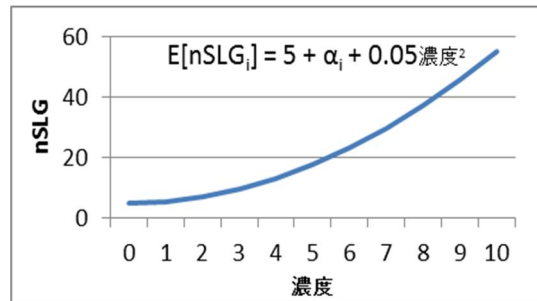


図 4 . パターン D

C . 研究結果

C.1. 平均値の比の信頼区間の構成方法の検討

小数自由度の要約

検討に用いた 1468 の比のデータについての要約統計量を表 2 に示す。

表 2 小数自由度の要約統計量

サイズ(n)	平均値	標準偏差	最小値	中央値	最大値
4169	4.49	1.12	3.00	4.58	6.00

自由度の範囲は 3 ~ 6 の範囲にあることがわかる。

97.5%点の比較結果

検討に用いた 1468 の比のデータについて R により得られる 97.5% を横軸に、山内の近似式により得られる 97.5% を縦軸にとった散布図を図 5 に示す。

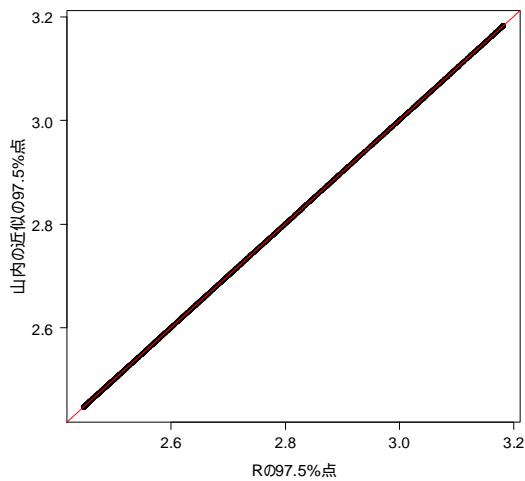


図5 Rと山内の近似の97.5%点

95%信頼区間の下限の比較結果

検討に用いた1468の比のデータについてRにより得られる95%信頼区間の下限を横軸に、山内の近似式により得られる95%信頼区間の下限を縦軸にとった散布図を図6に示す。図において点線で示した参照線は、それぞれ1の値のところであり、これは95%信頼区間の下限を用いたときの統計的有意差の基準となる値である。

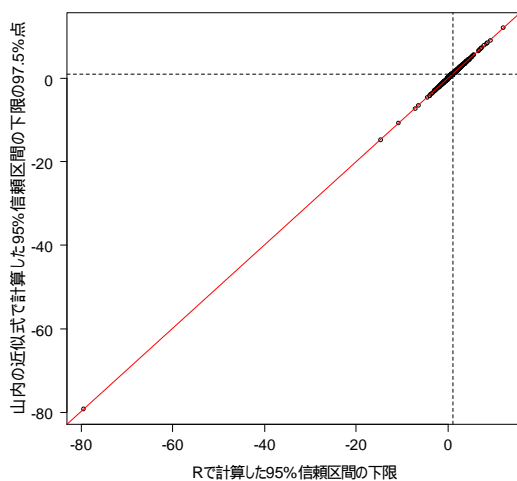


図6 Rと山内の近似の95%信頼区間の下限

C.2. MITAの判定方法に関する検討

バリデーション研究により得られた実データに基づく検討結果

バリデーション研究により得られた実データを用いて、方法 から方法 のそれぞれで検討した結果を以下の別紙付録図A1、別紙付録表A1からA3に示す。

(図A1は別紙に示す)

図A1は、各施設の各物質ごとに判定で用いている濃度反応関係を描いている。Criteria1、Criteria2、Criteria3はそれぞれ方法、方法、方法に対応している。方法は3回の実験の一つ一つについて判定がなされるために、3つの濃度反応曲線を描いた。一方、方法、方法は3回の実験を統合したデータセットを作成し、解析を行うために濃度反応曲線は1つとなる。

(表A1は別紙に示す)

表A1のS、A、Nは、3回の独立した実験結果をもってそれぞれ「Immunosuppression」「Immunoaugmentation」「No effect」の最終判定結果を示す。また3回の独立した実験結果は括弧の中に左から順に示す。

また、表A1の赤色と青色で色をつけた部分は3回の独立した実験結果をもってそれぞれ「Immunosuppression」、「No effect」と施設内で判定が一致したことを示す。

(表A2は別紙に示す)

表A2中の赤色と緑色で色をつけた部分は3セットの実験結果をもってそれぞれ

「Immunosuppression」、「Immunoaugmentation」

と施設内で判定が一致したことを示す。

(表 A3 は別紙に示す)

表 A3 中の赤色で色をつけた部分は 3 セットの試験結果をもって「Immunosuppression」と施設内で判定が一致したことを示す。

実験間差	誤差	判定	方法 (%)	方法 (%)	方法 (%)
$i_1 = 0$	= 0.08	Immunosuppression	33.26	0	0
		Immunoaugmentation	49.34	99.57	99.35
		No effect	17.40	0.43	0.65
$i_2 = 1$	= 0.8	Immunosuppression	12.68	1.43	1.24
		Immunoaugmentation	6.12	5.46	5.43
		No effect	81.20	93.11	93.33
$i_3 = 2$	= 0.08	Immunosuppression	0	0	0
		Immunoaugmentation	49.12	99.57	1.90
		No effect	50.88	0.43	98.10
$i_2 = 50$	= 0.8	Immunosuppression	0	1.58	0.03
		Immunoaugmentation	0	4.98	0.10
		No effect	100	93.44	99.87

シミュレーションによる検討結果

表 1 で示した各毒性パターンにおいて i_1 と i_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 3~6 に示す。

表 3 . パターン A の判定結果

実験間差	誤差	判定	方法 (%)	方法 (%)	方法 (%)
$i_1 = 0$	= 0.08	Immunosuppression	8.90	2.55	2.22
		Immunoaugmentation	9.26	2.76	2.53
		No effect	81.84	94.69	95.25
$i_2 = 1$	= 0.8	Immunosuppression	7.88	2.39	2.11
		Immunoaugmentation	9.69	2.39	2.29
		No effect	82.43	95.22	95.6
$i_3 = 2$	= 0.08	Immunosuppression	0	2.41	0.04
		Immunoaugmentation	0	2.65	0.06
		No effect	100	94.94	99.9
$i_2 = 50$	= 0.8	Immunosuppression	0	2.46	0.08
		Immunoaugmentation	0	2.62	0.04
		No effect	100	94.92	99.88

表 4 . パターン B の判定結果

実験間差	誤差	判定	方法 (%)	方法 (%)	方法 (%)
$i_1 = 0$	= 0.08	Immunosuppression	0	0	0
		Immunoaugmentation	100	100	100
		No effect	0	0	0
$i_2 = 1$	= 0.8	Immunosuppression	0.47	0	0
		Immunoaugmentation	99.06	99.97	99.99
		No effect	0.47	0.03	0.01
$i_3 = 2$	= 0.08	Immunosuppression	0	0	0
		Immunoaugmentation	100	100	0
		No effect	0	0	100
$i_2 = 50$	= 0.8	Immunosuppression	0	0	0
		Immunoaugmentation	98.81	99.99	2.42
		No effect	1.19	0.01	97.58

表 5 . パターン C の判定結果

表 6 . パターン D の判定結果

実験間差	誤差	判定	方法 (%)	方法 (%)	方法 (%)
$i_1 = 0$	= 0.08	Immunosuppression	0.05	0	0
		Immunoaugmentation	99.95	100	100
		No effect	0	0	0
$i_2 = 1$	= 0.8	Immunosuppression	1.13	0	0
		Immunoaugmentation	98.68	99.81	99.95
		No effect	0.19	0.19	0.05
$i_3 = 2$	= 0.08	Immunosuppression	0	0	0
		Immunoaugmentation	99.96	100	0
		No effect	0.04	0	100
$i_2 = 50$	= 0.8	Immunosuppression	0	0	0
		Immunoaugmentation	98.83	99.65	0.07
		No effect	1.17	0.35	99.93

D . 考察

D.1. 平均値の比の信頼区間の構成方法の検討

表 2 によると、比に関する 1468 のデータの自由度の中央値は 4.58 である。この値を用いて Excel2010 の t.inv 関数で、97.5%点を計算するとエラーもなく 2.776445 という値が返される。しかし、この値は自由度を 4 として t.inv 関数で計算しても同じ値である 2.776445 が返されてしまう。一方、R のパーセント点を計算する関数 qt を用いた場合、自由度が 4.58 の場合には 2.643129 が、自由度が 4 の場合には 2.776445 が返され、小数自由度に対して適切な値を返していることがわかる。

図 5 よりパーセント点は R と山内の近似のどちらもほぼ同様な値を取っていることがわかる。また、図 6 より 95%区間の下限も R と山内の近似のどちらもほぼ同様な値を取っていることがわかる。よって、山内の近似式は統計解析ソフト R を用いて解析すると同様の結果を得ることができるとがわかった。

よって、Excel でデータシートを構築する場合には、Excel の t 分布のパーセント点を計算関数である t.inv 関数を用いるのではなく、山内の近似式である(1)式によって計算し、比の 95%信頼区間を得ればよい。この結果を今後 modified MITA のバリデーション研究で行う際に作成するデータシートを反映させることにする。

Excel は広く普及しているソフトウェアである。R はフリーのソフトウェアであるという利点があるものの、実験実施者にとって統計ソフトは馴染みがあるソフトウェアではないため、R での解析を実験実施者に求めるべきではないであろう。また、Excel の t.inv 関数の問題は Excel のバージョンが更新されることで問題は解消されるかもしれない。しかしながら、どのような実験施設も常に最新の Excel を有し、利用しているとは考えにくい。

D.2. MITA の判定方法に関する検討

バリデーション研究により得られた実データに基づく検討結果

図 A1 からは 3 回の実験で濃度反応関係が異なっても、併合すると濃度反応関係は施設内、施設間ともに似た曲線になっていることがわかる。このことより、MITA は 3 回の実験によって判定するという開発の方針は支持できると言えるであろう。

表 1~3 から施設内再現性については、方法 1 は 3/5 となる物質はなかったが、施設間再現性はいずれの方法も 3/5 であったことがわかる。被験物質 1、3、4 はいずれの方法でも判定結果が同じであり、被験物質 2 と 5 の判定が異なっており、これら 2 物質の結果の詳細が必要である。

これらの 2 つの物質について図 A1 の特に方法 1、2 に着目すると濃度反応関係は試験間によって大きく異なっているとは言えず、判定結果の違いは、判定方法に起因する微細な部分であると考えた方

がよいと考えられる。

方法 1 は、独立した 3 回の実験の判定結果が異なった場合に行う t 検定において、異なる濃度の値を用いてしまう可能性に対する懸念など、最終判定に至るまでの複雑さある。また、方法 1、2 では Dunnett 検定において濃度 0 群に対してある 1 点の濃度で有意な差があれば immunosuppression もしくは immunoaugmentation と判定されるため、用量依存性を考慮すべきかもしれない。

シミュレーションによる検討結果

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 3）。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。

シミュレーションによる検討結果

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 4、5、6 それぞれの上 2 段）。実験間差や誤差が小さくても、方

法 は方法 、 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 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 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 では応答が 0 から 1 の範囲に収まる %supp_{ij} に基づき計算していることによる。%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が期待されている。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 は %supp_{ij} を用いて判定を行う解析に、線形モデルを想定しているので、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 よりも方法 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

シミュレーションによる検討において、作成したデータは、方法 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 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 は応答が 0 から 1 の範囲に収まる %supp_{ij} 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。

また、シミュレーションによる検討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} に基づく判定が行え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。実データに基づく検討の考察として記載したように、どの判定方法においても、判定結果の再現性を保つための改良が必要である。特に濃度反応関係の組み入れた判定方法が必要であろう。MITA に詳しい毒性家とのさらなる意見交換が必要である。

MITA は 3 回の独立した実験結果に基づき判定を行う試験法として提案された。本検討では検討した判定方法は 3 回の異なる実験結果が得られたことを前提として検討を行ってきた。しかし、他の *in vitro* 毒性試験では、2 回の判定結果が同じであった場合、3 回目の実験を行わずに 2 回の判定結果に従うことが多い。このような判定についての検討は行えなかった。

E. 結論

MITAのように、化学物質の評価に必要な平均値の比の95%信頼区間をExcelで作成するデータシート上で行うには、Excelのt分布のパーセント点を計算する関数であるt.inv関数を用いるのではなく、山内の近似式である(1)式によってパーセント点を計算し、平均値の比の95%信頼区間を得ればよい。

MITAの判定方法の検討として検討した3種類の方法の間では、統計学的な観点からは、方法が第一種の過誤を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。しかし、実データの検討からはいずれの方法も判定結果を再現することが難しいため、濃度反応関係を考慮した判定が必要となる。

謝辞：本検討を進めるにあたり解析のためのデータシートの作成とデータの整理を協力してくれた同志社大学の丸谷あおいさん、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

なし。

G. 研究発表

Kimura Y., Yasuno R., Watanabe M., Kobayashi M., Iwaki T., Fujimura C., Ohmiya Y., Yamakage K., Nakajima Y., Kobayashi M., Omori T.,

Kojima H. and Aiba S. (2017). Multi-Immuno Tox Assa (MITA): the creation of its data set and the results of validation studies. *Alternatives to Animal Testing and Experimentation* 22 (Supplement), 60.

H. 知的財産権の出願・登録状況

なし。

参考文献

- 1) Fieller, E. C. (1944). A fundamental formula in the statistics of biological assays and some applications. *Quarterly Journal of Pharmacy and Pharmacology*, 17, 117–123.
- 2) 永田靖. 統計的方法のしくみ-正しく理解するための30の急所-. 日科技連出版社, 1996.
- 3) 芳賀敏郎、橋本茂司. 統計解析プログラムの基礎. 日科技連出版社, 1980.
- 4) Littell, R. C., Stroup, W. W., Freund, R. J. SAS for linear models. SAS, 2002.
- 5) 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.