

## 厚生労働科学研究費補助金

### 医薬品・医療機器等レギュラトリーサイエンス政策研究事業

ワクチン等の品質確保を目的とした国家検定の最適化や国際整合化を目指すための研究

#### 分担研究報告書

### 血液製剤の国家検定の見直しに関する研究

-国家検定見直しに用いる定量的試験結果の安定性評価法の開発-

研究分担者 水上 拓郎 国立感染症研究所 次世代生物学的製剤研究センター

研究協力者 谷生 道一 国立感染症研究所 次世代生物学的製剤研究センター

研究要旨：国家検定における試験結果の安定性評価法および自家試験値との同等性評価法の開発を目指し、昨年度までの分担研究において正規分布を仮定した評価法（ $S_1$ ）および矩形分布を仮定した評価法（ $S_2$ ）を独自に考案し、38種の実試験結果について $S_1$ および $S_2$ による評価を行った。その結果、いずれの評価法においても客観的に評価可能であることが分かった。また基準値に近い値を含むデータ群では $S_1$ よりも $S_2$ において厳しい判定になることが分かり、 $S_2$ は異常値検知として有用である可能性が示唆された。一方、検証可能な実試験数には限りがあるため、より一般的かつ多条件におけるデータ群を用いた検証が必要と考えられた。そこで本年度は、一定の規格幅において、サンプルサイズおよび標準偏差値の条件を変えた模擬試験を70グループ（100試験/グループ）作成し、各試験における不合格経験頻度と、算出された $S_1$ と $S_2$ の比較検証を行った。その結果、開発した評価法を用いることで、試験結果の安定性を客観的に評価可能であることが示唆された。

#### A. 研究目的

検定・検査で実施されている試験では数値で結果が得られる定量的試験が多く、得られた試験値が規格を満たすかどうかで合否判定をおこなっている。測定値には、試験方法等に由来するばらつきが含まれ、ばらつき方も試験の性質によって異なる。通常、定量的試験では、試験の目的と測定値のばらつきを考慮した上で、上限値または下限値、あるいはその両方により規格が定められていると考えられる。品質管理の視点では、測定値が規格内に安定して収まっていることを確認することが必要であるが、試験結果の安定性

について、統一的な評価基準はない。これらを踏まえ、令和3年度に正規分布仮定の試験データにおける標準偏差値（SD）の95%信頼区間を用いた評価法（ $S_1$ ）、また令和4年度に測定値の分布範囲を矩形と仮定した標準偏差値を用いた評価法（ $S_2$ ）を考案し、実際の国家検定試験38種について検証を行った。その結果、全ての試験において現場感覚と合致した評価結果が得られ、 $S_1$ および $S_2$ の両方の結果を検証することで、より客観的な安定性評価が可能であることが示唆された。本研究では、シミュレーションにより作成した多様な条件の試験データ群について

S<sub>1</sub>およびS<sub>2</sub>による安定性評価法の有用性について検証を行うことを目的とした。

## B. 研究方法

＜安定性判定基準式の定義＞

評価に用いる S<sub>1</sub> および S<sub>2</sub> は以下で定義する。安定性評価の基準は、それぞれ 1 以下とする。各式の詳細は令和 3 年度および令和 4 年度の報告書で説明している。

$$S_1 = \frac{SD_{95L}}{\frac{|\text{Min}(L_{up} - AV, AV - L_{low})|}{3}}$$

$$S_2 = \frac{\frac{D_{max} - D_{min}}{\sqrt{3}}}{|\text{Min}(L_{up} - D_c, D_c - L_{low})|}$$

ここで、SD<sub>95L</sub> は実データより算出された SD 値の 95%信頼区間の下限値、AV は実データの平均値、L<sub>up</sub> と L<sub>low</sub> はそれぞれ試験規格の上限値と下限値、D<sub>max</sub>、D<sub>min</sub> および D<sub>c</sub> はそれぞれ実データの最大値、最小値および中心値（最大値と最小値の平均値）である。尚、片側規格の場合は上式において L<sub>up</sub> = L<sub>low</sub> と置き換える。

S<sub>1</sub> ≤ 1 の場合、実データは規格範囲内に約 99.7%以上含まれる事が期待される。規格外（不合格）値を含む場合は、D<sub>max</sub> ≥ L<sub>up</sub> または D<sub>min</sub> ≤ L<sub>low</sub> であるため、定義式より必ず S<sub>2</sub> > 1 となる。また S<sub>2</sub> ≤ 1 の場合、

$$\frac{\text{規格幅}}{\text{実データ分布幅}} = \frac{L_{up} - L_{low}}{D_{max} - D_{min}} \geq \frac{2}{\sqrt{3}} = 1.15 \dots$$

であるため、実データ分布が 15%以上余裕を持って規格幅内に収まっていることが保証される。

＜模擬試験データ作成＞

定量的試験では目標値があり、測定値は目標値を中心としてばらつく状況が一般的であるため、本シミュレーションで作成する試験値は、正規分布を仮定した。

両側規格試験モデルとして、目標平均値 = 100、規格 = 100 ± 20 (L<sub>up</sub> = 120、L<sub>low</sub> = 80) を設定した。SD 値は、規格幅 = 目標平均値 ± 2 SD、± 2.5 SD、± 3 SD、± 3.5 SD、± 4 SD、± 4.5 SD および ± 5 SD となる 7 条件を設定した (± k SD の場合 SD 値 = 20/k)。各試験のロット数 (サンプルサイズ) は、5、10、15、20、25、30、50、100、250 および 500 の 10 条件を設定した。シミュレーションデータの作成および解析は、Microsoft Excel for Mac ver. 16 で行った。模擬測定値は乱数関数 RAND() を確率変数として、NORM.INV(RAND(), 100, SD) により生成した。同一の SD 値およびロット数設定において 100 試験作成し、これを 1 グループとした (合計 70 グループ作成)。各試験における規格外 (OOS) 数、S<sub>1</sub> および S<sub>2</sub> をそれぞれ算出し、各グループで OOS ≥ 1、S<sub>1</sub> > 1 および S<sub>2</sub> > 1 となる試験数をそれぞれ集計した。本研究で用いた両側規格試験のシミュレーションデータ作成概要を図 1 に示す。

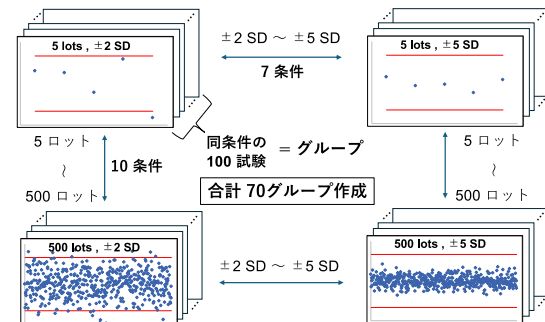


図 1. 両側規格模擬試験データ作成概要。偶発性を考慮し、同じ SD 値およびロット数の設定条件で 100 試験を作成し、これを 1 グループとした。

片側規格試験の場合でも、データが正規分布を仮定出来る場合は、本質的に両側規格試験と同様の結果が得られることが期待される。一方、含量の有無を確認する試験など、検出限界付近の値を含む場合は、試験値の正規分布仮定は不適切と考えられる。このような、検出限界付近の値を含む片側規格試験を想定し、片側上限規格値=100、目標平均値=0、検出限界値  $\leq 0$  の条件によるシミュレーションを行った。検出限界値以下の測定値を0とするため、 $NORM.INV(RAND(), 0, SD)$ で模擬値を生成後、IFS(模擬値>0, 模擬値, 模擬値 $\leq 0, 0$ )により変換した値を模擬測定値とした。SD値およびロット数の設定は、両側規格と同様にそれぞれ7条件および10条件とした。尚、SD値は、規格=0 $\pm$ 100と置き換えて算出した値を用いた( $\pm k$ SDの場合SD値=100/k)。

(倫理面への配慮) 該当しない

### C. 研究結果

#### <両側規格試験>

シミュレーションで作成した両側規格試験の各グループ内において、 $OOS \geq 1$ となった試験数を、ロット数に対して設定SD別にプロットしたグラフを図2に示す。

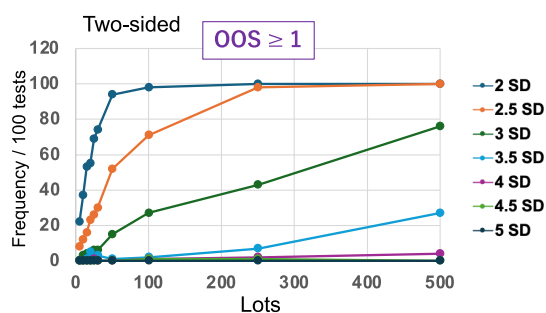


図2. 同一SD条件グループ内において不合格を経験( $OOS \geq 1$ )した試験数のロット数に対する変化(両側規格)

図2より、規格内(100 $\pm$ 20)に $\pm 2$ SD $\sim$  $\pm 3.5$ SDと設定した場合、ロット数増加に伴い、 $OOS \geq 1$ 、即ち不合格値を1回以上出した試験数は増加することが分かった。一方、 $\pm 4$ SD以上を設定した場合、ロット数が増えても不合格を経験する試験はほとんど無かった。実際の試験では、OOSが出た場合、再試制度があると考えられるため、 $\pm 4$ SD以上の試験では不合格はほとんど発生しないと期待される。尚、 $\pm 5$ SD設定において $OOS \geq 1$ となる試験数は0であった。

図3に、各グループにおいて $S_1 > 1$ または $S_2 > 1$ となった試験数を、ロット数に対して設定SD別にプロットしたグラフを示す。

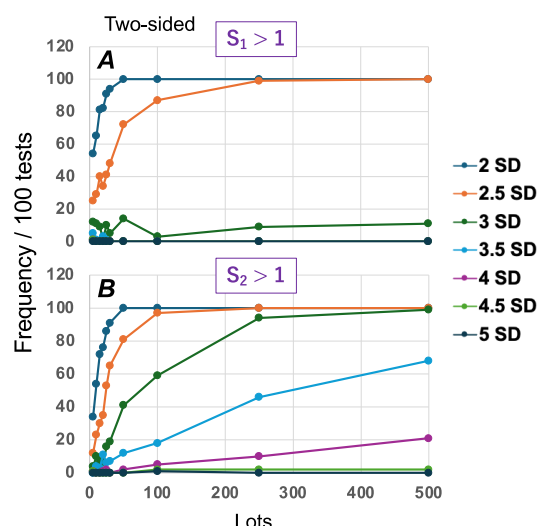


図3. 同一SD条件グループの $S_1 > 1$ (A)および $S_2 > 1$ (B)となった試験数のロット数に対する変化(両側規格)

図2と図3Aの比較から、 $\pm 2$ SD $\sim$  $\pm 2.5$ SDの場合、ロット数増加に伴い $S_1 > 1$ となる試験数が増加し、同じSD設定条件において $[OOS \geq 1 \text{の試験数}] < [S_1 > 1 \text{の試験数}]$ となることが分かった。一方、 $\pm 3$ SD $\sim$  $\pm 5$ SDでは、 $S_1 > 1$ となる試験数が低下し、 $OOS \geq 1$ でも $S_1 < 1$ となる試験があった。 $S_2$ に

については、同じ SD 条件下において、常に  $[OOS \geq 1 \text{ の試験数}] < [S_2 > 1 \text{ の試験数}]$  であった (図 2、3B)。設定 SD 別の各判定結果を表 1 に示す。 $S_2$  の定義上、 $OOS \geq 1$  の条件では、必ず  $S_2 > 1$  であるため、 $OOS \geq 1$  かつ  $S_2 \leq 1$  となる試験は存在しない (表 1; 検出失敗および  $S_2$  未検出)。

#### < 検出限界値を含む片側規格試験 >

シミュレーションで作成した、検出限界値を含む片側規格試験の合計 70 グループについて、各グループ内において  $OOS \geq 1$ 、 $S_1 > 1$  および  $S_2 > 1$  となった試験数を、ロット数に対して設定 SD 別にプロットした各グラフを図 4 に示す。

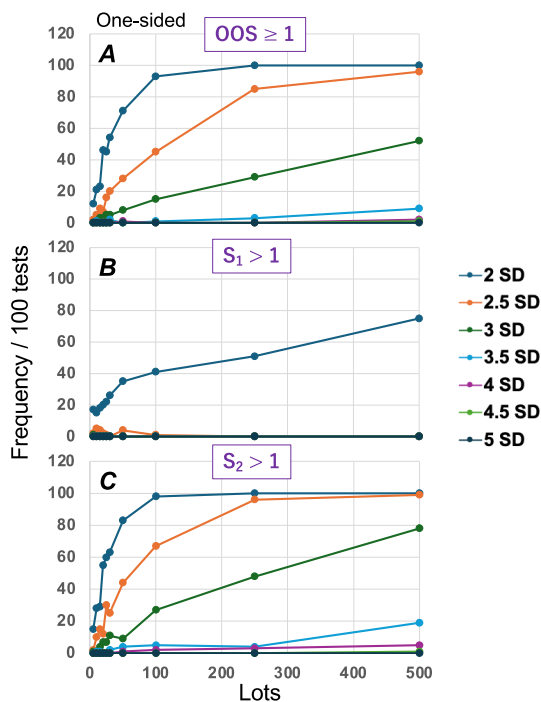


図 4. 同一 SD 条件グループの  $OOS \geq 1$  (A)、 $S_1 > 1$  (B) および  $S_2 > 1$  (C) となった試験数のロット数に対する変化 (片側規格)

図 4A より、片側規格においても規格内 ( $0 \pm 100$ ) に  $\pm 2$  SD  $\sim \pm 3.5$  SD となるように SD 値を設定した場合、ロット数増加に伴っ

て不合格を経験した試験数が増加した。但し、両側規格 (図 2) に比べ、片側規格では、1 試験あたりの OOS 数は約半分になると期待されるため、 $OOS \geq 1$  となる試験数の増加勾配は緩やかになったと考えられる。同様に、両側規格 (図 3B) に比べ、片側規格の  $S_2$  においても  $S_2 \geq 1$  となる試験数の増加勾配は緩やかになったが (図 4C)、両側規格試験の場合と同様に、同じ SD 条件下において、常に  $[OOS \geq 1 \text{ の試験数}] < [S_2 > 1 \text{ の試験数}]$  であった (図 4A、4C)。

一方、 $\pm 2$  SD の場合を除いて、 $S_1 > 1$  となる試験はほとんど無く (図 4B)、 $\pm 2$  SD の場合でも、 $[OOS \geq 1 \text{ の試験数}] < [S_1 > 1 \text{ の試験数}]$  となった (図 4A、4B)。

#### D. 考察

##### < 両側規格試験 >

今回検証した両側規格試験では、全ての条件において  $[OOS \geq 1 \text{ の試験数}] < [S_2 > 1 \text{ の試験数}]$  となっており、 $S_2$  は不合格 ( $OOS \geq 1$ ) のリスク検出に有用であることが分かった。これは定義上、 $S_2 \leq 1$  が、 $[\text{規格幅}]/[\text{実データ分布幅}] \geq \sim 1.15$  と等価である、即ち規格幅の方が実データ分布幅よりも常に 15% 以上高いことを評価基準としているため妥当と考えられた。これらの結果は、 $S_2 \leq 1$  であることをもって、試験成績は余裕をもって規格に適合していると結論可能であることを示している。一方、 $\pm 4$  SD、 $\pm 4.5$  SD および  $\pm 5$  SD の設定で、 $OOS = 0$  かつ  $S_2 > 1$  となった試験が、それぞれ 33、4 および 1 件検出された (表 1; 注意  $S_2$ )。一般的に規格幅内に  $\pm 4$  SD 以上の精度で実施されている試験は、安定した試験であると判断されるため、 $\pm 4$  SD 以上の試験で  $S_2 > 1$  となる結

果は、偶発的に発生した外れ値に起因したものと予想される。これは、 $S_2 > 1$  であることのみで試験結果が不安定と評価するのは不適切であることを示している。

$S_1$  については、 $\pm 2 SD$  および  $\pm 2.5 SD$  設定で、 $[OOS \geq 1 \text{ の試験数}] < [S_1 > 1 \text{ の試験数}]$  となったが、 $\pm 3 SD$  以上の設定では、ほとんどの試験で  $S_1 \leq 1$  となった。 $S_1$  の定義では、規格範囲内に平均値  $\pm 3 SD$  が収まることを評価基準にしていることから、 $\pm 3 SD$  以上の設定で  $S_1 > 1$  となる試験数が低下することは妥当と考えられた。これらの結果は、 $S_1 > 1$  であることをもって、試験成績は不安定であると結論可能であることを示している。一方、 $\pm 3.5 SD$ 、 $\pm 4 SD$  および  $\pm 4.5 SD$  の設定で、 $OOS \geq 1$  かつ  $S_1 \leq 1$  となった試験が、それぞれ 45、8 および 2 件検出された (表 1 ;  $S_1$  未検出)。 $S_1 \leq 1$  の場合、実データの 99.7% 程度以上が規格範囲内に含まれることが期待されるため、 $S_1 \leq 1$  かつ  $OOS \geq 1$  の試験では、偶発的に発生した外れ値が  $OOS$  となったと考えられる。同様に、 $S_1 \leq 1$  かつ  $S_2 > 1$  の場合も、偶発的に発生した外れ値が含まれていると考えられるため、「 $S_1 \leq 1$  かつ  $S_2 > 1$ 」(表 1 ; 注意  $S_2$ ) の条件が、外れ値検出に有用であると考えられる。

#### <検出限界値を含む片側規格試験>

検出限界値を含む片側試験のシミュレーション結果においても、常に  $[OOS \geq 1 \text{ の試験数}] < [S_2 > 1 \text{ の試験数}]$  となったことから、両側試験の場合と同様、 $S_2$  は不合格のリスク検出に有用であることが分かった。一方、常に  $[OOS \geq 1 \text{ の試験数}] > [S_1 > 1 \text{ の試験数}]$  であったことから、 $S_1$  は検出限界値を含む試験の評価には不適であることが分かった。

この結果は、 $S_1$  が正規分布仮定の基で定義されていることから、妥当であると考えられる。一方、 $S_2$  は、データ分布の最小値と最大値を使って定義されているため、ばらつきの少ない試験では適切な評価を与えると期待されるが、1 ロットでも外れ値が含まれると  $S_2$  値は大きくなるため、検出限界値を含む試験では、外れ値の評価法について別途検討が必要である。

## E. 結論

本研究結果により、両側規格試験において、 $S_1 > 1$  の場合、試験成績は不安定であり、 $S_2 \leq 1$  の場合、試験成績は安定して規格に適合していると結論可能であると考えられた。また「 $S_1 \leq 1$  かつ  $S_2 > 1$ 」の場合は外れ値を含む可能性が示唆された。一方、検出限界値を含む試験では正規分布仮定が成立しないため、 $S_1$  による評価は不适当であるが、 $S_2$  による評価は可能であることが分かった。但し、 $S_2$  は外れ値の影響を受けるため、検出限界値を含むと予想されるデータでは外れ値を別途評価する必要があると考えられた。

## F. 研究発表

### 1. 論文発表

1) Choi CW, Choi Y, Maryuningsih YS, Wibisono B, Kim JW, Ramondrana D, Mizukami T, Ochiai M, Samat AA, Mangorangca C, Thi DL, Van HP, Shim SB, Seong SK, Shin IS. Report for the Eighth Asian National Control Laboratory Network meeting in 2023: Self-sufficiency strategy of plasma-derived medicinal products and regulatory harmonisation. *Biologicals*.

2024; 85; 101754.

G. 知的財産権の出願・登録状況 なし

2. 学会発表 なし

表 1. 両側規格 7000 試験 (70 グループ) における判定結果まとめ

判定	正常	検出失敗	S <sub>1</sub> 未検出	S <sub>2</sub> 未検出	注意 S <sub>1</sub>	注意 S <sub>2</sub>	警告	異常
S <sub>1</sub>	≤ 1	≤ 1	≤ 1	> 1	> 1	≤ 1	> 1	> 1
S <sub>2</sub>	≤ 1	≤ 1	> 1	≤ 1	≤ 1	> 1	> 1	> 1
OOS	0	≥ 1	≥ 1	≥ 1	0	0	0	≥ 1
± 2 SD	111	0	6	0	76	16	95	696
± 2.5 SD	325	0	38	0	79	62	98	398
± 3 SD	613	0	143	0	32	158	15	39
± 3.5 SD	818	0	45	0	7	127	1	2
± 4 SD	958	0	8	0	0	33	1	0
± 4.5 SD	993	0	2	0	0	4	1	0
± 5 SD	999	0	0	0	0	1	0	0
総試験数	4817	0	242	0	194	401	211	1135