のないよう十分高く設定することの保障であるが、これは統計学的裏付けを持たない[7]。

C. 二項法

二項法は、2006 年の FAO/WHO 合同残留農薬専門家会議（JMPR）においてエクセル®表計算ソフトを使用し95 パーセンタイル値を算定する統計的手法として正式に紹介された。二項法に関する JMPR の評価及び表計算ソフト使用に関する勧告は 2006 年の JMPR 報告書に載っている[8]。

その名称が示唆するように、この方法で使われている計算法は、二項分布の性質に基づく特定の結果を観察する確率を有する。この確率は、パーセンタイル値（通常、95 パーセンタイル値より低い値）の「最適推定値」の計算に利用される。より低いパーセンタイル値の算定法は、二項分布の性質を除いて、データの分布という内在的ないかなる仮定にも基づかない。したがって、それは「分布に依存しない」計算法であり、いかなる分布状態（例えば、正規分布、対数正規分布、ガンマ分布など）のデータセットのサンプルサイズであっても、より低いパーセンタイル値の「最適推定値」を提供することを示唆している。より低いパーセンタイル値が算定されると、「外挿係数」が適用され、95 パーセンタイル値及び 95 パーセンタイル値の「最適推定値」が出る。この計算で使用された外挿係数は、著者が収集した（経験的）監督下作物残留試験データから得たものであり、95 パーセンタイル値と 99.5 パーセンタイル値との比率、及び 95 パーセンタイル値よりも低いその他の様々なパーセンタイル値との比率を表している。

1. 専門的詳細

二項法は以下のように提案された。

n 個の監督下作物残留試験データを大きい方から順に{$x_1>x_2>\ldots>x_n$}と並べ替えると、*n* 個のサンプルの少なくとも 1 つの値は 95%の信頼性をもって、選択されたパーセンタイル値 *sp* を上回る。

$$sp = 100 \times (1-0.95)^{\frac{1}{n}} \qquad (4)$$

さらに、*n* 個のサンプルからとった *k* 個のサンプルが *sp* パーセンタイル値を確実に上回る確率は、

$$P_{=k} = (1-\frac{sp}{100}) \times {}^{n}C_k \times (\frac{sp}{100})^{n-k} \qquad (5)$$

ここで、$1 \leq k \leq n$ 。

この確率から、*sp* パーセンタイル値の最適推定値が計算される。

sp パーセンタイル値は次の式で求められる。

$$\frac{\sum_{i=1}^{n} x_i P_{=i}}{\sum_{i=1}^{n} P_{=i}} = \frac{\sum_{i=1}^{n} x_i P_{=i}}{0.95} \qquad (6)$$

---

7 興味深いことに、変動係数が 1 である対数正規分布（圃場残留試験データでは珍しい値でない）において、75 パーセンタイル値の 2 倍の値は 95 パーセンタイル値、93.4 パーセンタイル値に近い値を示す。

8 http://www.fao.org/ag/AGP/AGPP/Pesticid/JMPR/DOWNLOAD/2006_rep/report2006jmpr.pdf

そして、*sp* パーセンタイル値に外挿係数 g_e を乗じ、95 パーセンタイル値が「外挿」される。g_e は既知の経験に基づく分布データより求める。

95 パーセンタイル値は次の式で求められる。

$$\frac{\sum_{i=1}^{n} x_i P_{=i}}{0.95} g_e \qquad (7)$$

2. **考察**

二項法の目的は圃場残留試験データのサンプルに基づく分布の 95 パーセンタイル値での「最適推定値」を出すことである。この計算は不偏点推定法（unbiased point estimate）とも呼ばれ、（正確に計算された場合）50%の例で真の 95 パーセンタイル値を下回ることが期待される。換言すれば、試料のサンプリング（この場合、圃場残留試験より試料を集めること）が 100 回行われ 95 パーセンタイル値での「最適推定値」がそれぞれの試料について計算された場合、そのうち 50 の「最適推定値」は各サンプルを抽出した母集団分布の実際の 95 パーセンタイル値を下回ることが期待されるだろう、といえる。*こうした推定値は、第一の目的を 95 パーセンタイル値での 95% UCL の算定におく EU 方法 I による推定値とは概念上異なるタイプのものである。*こうした場合、95 パーセンタイル値での 95%UCL が、わずかに 5%の例で真の 95 パーセンタイル値よりも低く推定されると期待される。

二項法の目的は、作業グループの目的とは異なる。二項法の目的が 95 パーセンタイル値での UCL よりは、その「最適推定値」を計算するものだとしても、それでもなお二項法は以下に述べる重要な理論上の問題点を含んでいる。

問題点 1

前述の式を使うと、$P_{=k}$ は *n* 個のサンプルからとった *k* 個のサンプルが *sp* パーセンタイル値を確実に上回る確率であり、x_k（k 番目に高い値）が *sp* パーセンタイル値を上回る確率ではない。例えば、$x_1 > x_2$ なので、x_1 が *sp* パーセンタイル値を上回る確率は x_2 が *sp* パーセンタイル値を上回る確率よりも大きい。しかし、$P_{=1} < P_{=2}$ である。$P_{=k}$ と x_k は相関関係に無いので、これら積の和は選択されたパーセンタイル値での最適推定値の期待値の計算とはならない。

問題点 2

$$\frac{\sum_{i=1}^{n} x_i P_{=i}}{\sum_{i=1}^{n} P_{=i}} = \frac{\sum_{i=1}^{n} x_i P_{=i}}{0.95}$$ は *sp* パーセンタイル値の期待値を計算するのに正しい式ではない。なぜなら、

$\sum_{i=1}^{n} P_{=i} = 0.95$ はすべての確率の和でないからである。それは全ての事象発生の確率を包含する $\sum_{i} P_{=i} = 1$ となるに場合のみ、期待値となる。

二項法の理論的基礎に対する懸念から、二項法がバイアスのかかった点推定値を与えるかを判定するためブートストラップ・シュミレーションを実施した。簡潔にいえば、様々な平均値と標準偏差を有する正規分布および対数正規分布の両方からサンプルを抽出して行った複数のシュミレーション（それぞれ 1000 回の反復を含む）に基づいて、二項法はバイアスのかかっていない 95 パーセンタイル推定値を出さないことを作業グループは確認した。これらのシュミレーションに関する更なる詳細および作業グループが明らかにした理論的懸念事項は付属文書Ｂに

紹介する。

Ⅳ. NAFTA 対数正規法

EU 方法Ⅰは、適切な統計的概念を採用し、高めのパーセンタイル値での UCL を計算するものであり、EU 内で国境を越えて採用・許容されており、同じまたは類似のデータが使用される場合に整合性のある推定値を出す。本作業グループは、EU 方法Ⅰの多くの利点を認め、評価する。しかしまた同方法の深刻な欠点も確認した。特に、同方法は残留農薬が正規分布すると仮定している点、同方法はこの仮定を検証または評価する構成要素を含んでいない点と、および，高い値を棄却するための異常値判定法（これもまたデータが正規分布すると仮定する）を採用しているが、残留農薬濃度はしばしば対数正規分布するということを考えれば（または少なくとも右肩上がりとなる）、多くの真の高い値を棄却しかねないという点である。その他に懸念される事項２点が更に明らかにされた。*UCL が、高い方のパーセンタイル値を基にして計算されるならば，小さいサンプルサイズの算定値は非実用的に高くなる* － これは残留物の母集団が正規分布するとの不適切な仮定によってのみ導かれる事実である[9]。また、EU 方法Ⅰでどのくらい小さいサンプルサイズを使うべきか、または EU 方法Ⅱでどの点で点推定すべきかについてのガイダンスがないこと、である。

A. EU 方法Ⅰの修正としての NAFTA 対数正規法の解説

前述のように、EU 方法Ⅰはしっかりとした統計的理論に支えられた手法であるが、圃場残留試験データではあまり見られないデータの正規分布仮説に依拠している。正規性という不正確な仮定のもと EU 方法Ⅰで計算される UCL は低く見積もられる傾向になるだろう。加えてこの過小見積もりを導くバイアスは、その多くが対数正規分布においては適切な観測値であろう「異常値」の誤った棄却により悪化させられることがある。

EU 方法Ⅰの限界を所与のものとして、データが正規分布でなく対数正規分布するとの仮定に基づき、またその仮定の検証がプロトコルの標準的な一部となるならば、この手法は大幅に改善されるだろう。EU 方法Ⅰに採用されているパーセンタイル値の UCL を推定する式は、対数正規分布に改定された仮定に対応するよう容易に修正できる。この場合、MRL は以下のように提案されよう。

$$MRL = \exp(\bar{x} + g_{(1-\alpha;p,n)}s), \qquad (8)$$

ここで、$\bar{x}$ と s は残留物データの自然対数（logs）の平均値と標準偏差。$g_{(1-\alpha;p,n)}$ は Odeh と Owen（1980）の表 1[10]から得られる。

この EU 方法Ⅰの修正版を「NAFTA 対数正規法」[11]と呼ぶ。高い値も含む全ての残留値は 95 パーセンタイル値の 95% UCL を導くために利用される。セクションⅨの B のシミュレーションは、真の 95 パーセンタイル値を下回る推定値が 5%しか生じず対数正規分布を支持しているが、この NAFTA 対数正規法は対数正規分布における 95 パーセンタイル値の 95%UCL のバイアスがかからない推定値を出してくれる。

B. 対数正規仮定及び関連する統計検定

前掲のように、適切な統計計算には統計的手順の一部であり仮定である分布状態を正式な評価対象とすることが必

---

[9] 前述のように、このことは 75 パーセンタイル値を 2 倍する EU 方法Ⅱが、代替的計算法として導入された理由である可能性がある。

[10] NAFTA 作業グループが作成した MRL 表計算ソフト中の計算式において類似の表が再現、使用されている。表計算ソフトで使われている因子の値は、有意水準 $\alpha = 0.05$ における様々なサンプルサイズの 95 パーセンタイル値及び 99 パーセンタイル値（すなわち $p = 0.95, 0.99$）の信頼限界上限を計算するためのものである。

[11] *圃場残留試験データに基づく農薬トレランス設定のためのガイダンス*という題の SOP 草案では、この手法は「EU 方法Ⅰ対数正規法」と呼ばれている。混乱を避けるため、本文書では「NAFTA 対数正規法」と呼ぶ。SOP 草案と表計算ソフト改定の際は、改定 SOP と MRL 表計算ソフト及び本文書で統一させるため、この新しい名称が使用されるだろう。

要である[12]。対数正規法は評価対象となる圃場残留試験データの分布状態につき、より妥当な仮定を与えるものであるが、データが実際に対数正規分布しない場合は偏った推定値を算出するだろう。したがって、対数正規分布という仮定はMRLを提案する計算を行う前に評価されるべきである。NAFTA作業グループは、対数正規分布の仮定の妥当性を評価する2つの相互補完的な手法を推奨する。確率プロットとシャピロ・フランシア検定統計量である。

1. 確率プロット

確率プロット（Chambers 1983）は、あるデータセットが所与の分布状態に従うか否かを評価する、グラフを使った手法である。特に正規確率プロットは、あるサンプルが凡そにおいて正規分布するかを判定するために使用される。同様に、残留物データの対数の分布状態を評価することは、データが対数正規分布するかを評価することと同義である（対数正規性は数値の対数が正規分布することを意味するため）。

仮にある残留値の順序集合が $X=\{x_1,x_2,\ldots,x_n\}$ と表され、これらの値の対数に対応する集合が $Y=\{y_1,y_2,\ldots,x_n\}$ と表されるとする。対数の値 $Y=\{y_1,y_2,\ldots,x_n\}$を予想される正規順位スコア b_isに対しプロットする。正規分布すると評価されるには、これらの点はおおまかな直線となるべきである。この直線からの逸脱は $Y=\{y_1,y_2,\ldots,x_n\}$が正規性から逸脱していることを示し、これは $X=\{x_1,x_2,\ldots,x_n\}$が対数正規分布に従わないことを意味する。

NAFTA作業グループは、下に示すBlom(1958)により概要が示された正規順位スコア $b_i's$ の計算法を採用した。

$$b_i = \Phi^{-1}\left(\frac{r_i - \frac{3}{8}}{n + \frac{1}{4}}\right) \qquad (9)$$

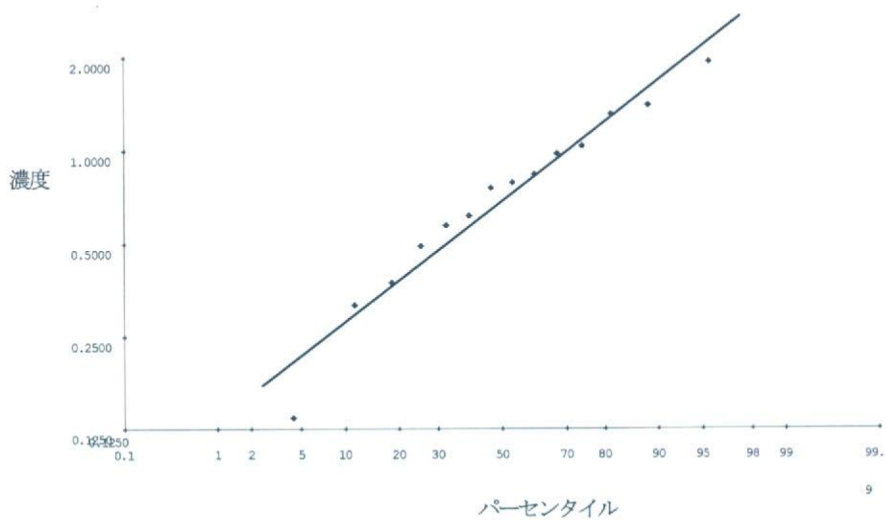
このとき、nはサンプルサイズ、Φ^{-1}は逆正規累積関数、そして r_iが観測値 y_i $1 \le i \le n$ の順位とする。Tukey、Van der Waerden、またはHazenによって提案された他の方法などは、順位から正規スコアを算出するのに使われる。ところが、Blomの方法は真の期待正規スコアにより近い近似値を提示する。

グラフ1は各点が直線に近似する圃場残留試験データの確率プロットの一例であり、対数正規分布が適切なモデルであることを示しており、元のスケールでのデータも対数正規分布することを意味している。グラフを利用した手法は、試験データとそれをあてはめた分布の視覚による比較を可能にする。グラフを利用する手法は定量的ではないが、十分な大きさのサンプルサイズの場合は特定の分布状態の選択を助けること、または分布状態のあてはめの棄却に最も説得力を持つ。この説得力は定量的な適合度検定が持つ欠点から導かれる。全ての定量的評価がそうであるように、主観性の度合いは確率プロットのあてはまりの評価と関連する。適合性の良し悪しに明確な区別がないことは明らかである。データセットが対数正規性を有すると合理的に考えられるためには、線が正確な直線である必要はない。NAFTA作業グループは、目視検査によりデータセットの対数正規性を正確に判定する能力は経験値により向上すると考えるが、その過程に欠点がない訳は無い。したがって次のセクションで述べるような、定量的な適合度検定結果を考慮することも重要である。

[12] 「統計的手順における分布状態の仮定」の量的評価のみでなく、適切な統計計算ではこの仮定がグラフによって評価されることが必要だと主張する統計学者も多くいる。

グラフ１

対数正規確率プロット



2. シャピロ・フランシア検定統計量

前掲のグラフによる方法が対数正規性仮定の妥当性を評価するのに適切な手法だと多くの人が捉えている一方、この仮定をより客観的に評価できる統計計算がある。シャピロ・ウィルクスの検定を修正したシャピロ・フランシア検定統計量はデータセットの正規性を判定する。シャピロ・フランシア検定統計量W'は以下で定義される。

$$W' = \frac{\left(\sum_{i=1}^{n} b_i y_i\right)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \qquad (10)$$

ここで、

$$\boldsymbol{b'} = (b_1, b_2, \ldots, b_n) = \frac{\boldsymbol{m'}}{\boldsymbol{m'm}} \qquad (11)$$

また、$\boldsymbol{m'} = (m_1, m_2, \ldots, m_n)$は標準正規順位統計量の予想される値のベクトルを表す。

W'の計算手順はシャピロ・ウィルクスのW'の計算よりも非常に簡単で簡素化されている。しかし、どちらの検定も敏感度はあまり変わらない。統計量W'はサンプルサイズ n が $n < 50$ の場合のみ使用しうるのに対して、W'は $n > 50$ の場合であっても使用できる。

Devaney(1997)によると、W'と確率プロットのb_i'とy_i'の相関係数である r の間には密接な関係がある。rは、作業グループが作成したエクセル®表計算ソフトで作成される確率プロットから容易に計算できる。データの対数正規性の判定にはW'よりも rが使われる。対数正規仮定を棄却すべきか判定する rの棄却限界は米国標準技術局（the National Institute of Standards and Technology）のウェブサイトで入手できる(http://www.itl.nist.gov/div898/handbook/eda/section3/eda3676.htm)。本ウェブサイト上では rは確率プロット相関係数（the probability plot correlation coefficient; PPCC）と呼ばれる。

こうした統計量は特に初心者にとっては有益であるが、これらの検定には限界があり、対数正規仮定の妥当性を決めるにはグラフを使用した検定（確率プロット）と統計量検定の双方を合わせて検討することが重要であることを認識すべきである。モンテ・カルロ評価法についての OPP ガイダンス文書（米国環境保護庁，1997）には次の様な記述がある。

> 適合度検定結果を過度に解釈、または信頼しないよう注意しなければならない。多数の分布態様候補がある中で適合度検定が持つコンピューターの力と速さを使い、「最適な」あてはまりの良さの分布態様を選択し、その「最適な」あてはめを持つ分布状態が一定の有意水準において棄却されなかったと主張することはあまりにも魅力的である。そうしたやり方は統計学的に誤りであり、回避すべきである〔Bratley et al., 1987, 134 頁〕。適合度[Goodness-of-fit test; GoF〕検定の有益性が低いことは周知の事実であり、一般に当てはまりの良い分布状態を決めるよりは当てはまりの悪い分布を棄却するのに適している。サンプルサイズが小さいものから中くらいのものに関しては、GoF 検定は観察された分布と当てはめられた分布の間の細かい違いに対してあまり敏感でない。一方、大きなデータセットに関しては観察された分布と当てはめられた分布の微小な違いさえも帰無仮説の棄却を導く可能性がある。サンプルサイズが小中サイズのものに関しては、GoF 検定はおおまかな違いを確認する系統だった手法として捉えるのが最も良かろう。GoF 検定はせいぜい分析者が目視検査で気づいた事を確認するのに役立つぐらいであろう。特定の確率モデルが GoF 検定に基づいて棄却されたとしても、確率プロットその他の比較の目視検査により、データに適切にあてはまると判断される場合は当該確率モデルの維持を決定することが適切な場合がある。

C. 95/99 ルール

対数正規分布の手法（つまり、95 パーセンタイル値での 95%UCL）の利用は、作業グループが EU 方法 I について持っていた２つの懸念に対応することになる。残留量の分布は一般に正規分布よりも対数正規分布するということと、各検討例で分布仮定が適切であるかを評価することが重要という点についてである。しかし、小さなデータセット（「小さな」がたまたま定義づけられている）または異常に高い最大値を持つデータセットでは、UCL が非実用的に高くなりうる[13]。行政上の規制の観点からも、農薬の誤使用を検出する確率がほとんど無くなるほどに高い MRL を設定してしまわないことが同様に重要である。これは作業グループで最初に議論し、前述した「バランス均衡」である。具体的に言えば、合法的な農薬の散布により生じる作物中残留物が MRL を上回る合理的可能性を排除するのに十分に高い値であり、同時に誤用／非合法の散布を検出する合理的可能性がほとんどなくなるほど高くない値の MRL を設定することが理想的である。

結果として、NAFTA 作業グループは「95/99 ルール」を提案する。「95/99 ルール」は、対数正規分布仮定が適切である場合において、圃場残留試験データセットの 95 パーセンタイル値の中の 95%UCL 値または 99 パーセンタイル値での点推定値のいずれか最小の値に MRL を設定すべきだとする。このルールは以下の内容を含む。

> 母集団分布の真の 95 パーセンタイル値より低くならないことが相当な信頼性を持って（つまり、95%の信頼性で）いえる水準で MRL の設定を行うべきである。しかし、95 パーセンタイル値の上限が 99 パーセンタイル値を上回る場合は、99 パーセンタイル値が MRL を設定するのに使われるであろう。このように、検討対象分布の 99 パーセンタイル期待値を超えるような不適切に高い値に MRL が設定されることはない。

本文書のセクションⅨの B で、95 パーセンタイル値での 95%UCL 値を使用して MRL が算出される場合、MRL の 5%ほどは既知の 95 パーセンタイル値を下回るとシミュレーション結果が示している。また一方、ボックスプロットか

[13] ここで非実用的とは、統計学的観点でなく行政上の規制の観点から使用される。純粋に統計学的観点からいえば、95 パーセンタイル値での UCL は、残留値の 5%以上がそれを上回らないことを 95%の信頼性をもっていえる最適推定値である。MRL の設定が全く統計学的なものと捉えられるならば、95 パーセンタイル値での 95%UCL 推定値は保持されるだろう。

らはこれらのMRLの多くは真の95パーセンタイル値と比較しても相当大きいと知ることができる。95/99ルールを適用する場合、これらの大きな算定値の多くは棄却される。推定値のうち、その一定割合は真の値よりも大きいという信頼性をある程度犠牲にしても、95/99 ルールは異常に高い算定値の影響をある程度緩和してくれる。このことは、小さなデータセットまたは母集団の平均値と標準偏差が適切に算定されていないデータセットにおいては特に有益である。

V. 代替アプローチを要求するデータセット

作業グループは、95/99ルールが全てのデータセットに普遍的に適用可能ではないなど、多くの限界があることを認めた。これまで直接的または間接的に述べたように、対数正規法では次のことが必要である。まず、データセットが対数正規分布に似ていることが必要であり、また、95パーセンタイル値での適切なUCLまたは99パーセンタイル値での点推定が算定可能な十分なサイズのデータセットを必要とする。さらに、利用可能な程度に十分な数の検出例（>LOD）または検出率がある，または定量可能物（>LOQ）が存在することも必要である。これらの点について後述する。

A. 対数正規的でないデータ

95/99 ルールの一部の計算方法もまた、圃場残留試験の残留値が対数正規分布するとの仮定に基づく。したがって前述のように、対数正規仮定が適切であるかを評価することが重要であり、95/99 ルールがこうした仮定の検証をその一部として内包することが重要である。実際には、対数正規仮定を棄却するデータセットもいくぶんある。対数正規分布から抽出したデータであっても、対数正規検定がデータの対数正規性を否定する（特にサンプルサイズが小さいとき）ということも時々現実にある。とりわけ、確率プロットの目視検査及びシャピロ・フランシア検定のどちらによっても対数正規仮定を棄却すべきとされたデータセットにおいては、95/99ルールはMRLの設定に不適切であろう。

B. サンプルサイズの小さなデータセット

小さなサンプルサイズに基づき母集団のパラメーターを推定することは、どんな統計計算にとっても1つの挑戦であり、95/99ルールの手順も例外ではない。小さなサンプルサイズでは、データセットの推定平均値と変数の推定値が真の値に近似しないことがあり、広範囲にわたってデータが分布しうる。その結果、(8) の式で95パーセンタイル値の95%UCLの計算に使われるg係数が小さなサンプルサイズに対しては大きくなりすぎる、つまり、95 パーセンタイル値での算定であるUCLが高くなってしまう。95/99ルールを適用してもMRL算定値が大きくなりすぎてしまうことがある。そのような場合、MRLは非実用的に大きくなり非合法の農薬使用を検出できない結果となってしまうだろう。

どの程度のサンプルサイズまでならサンプルを適切に検討できるのかをより正確に表現するため、作業グループは広範囲に及ぶ一連のブートストラップ・シミュレーションの実施に着手した。このシミュレーション研究では合成データセット（つまり、人為的に作成したデータセット）と実際のデータセットの両方を用いた。期待通り、小さいサンプルサイズではMRLの算定値の分布範囲が相当大きくなることがシミュレーションで明らかにされた。サンプルサイズが15またはそれ以上の場合、MRLの算定値は安定して狭い分布範囲にとどまる。(8) の式において、nが15より大きいときg係数は小さくなるので、これは当然のことである。シミュレーションの実施結果に基づき、15 よりも小さなサンプルサイズは95/99ルールによるMRL値が非実用的となる小さなデータセットと考えられる。

C. 不検出物を含むデータセット

検出下限（limit of detection; LOD）を下回ると報告された残留物データ（「不検出物」とも呼ばれる）含む圃場残留試験のデータセットは多くある。監督下作物残留試験においては作物全体に農薬が使用されるため、LOD を下回る値が真に0であることは期待されない。むしろ<LODとの注記は、実際の残留値が0より大きくLODより小さいことが知られていることを意味すると捉えるべきである。

通常LODを下回ると報告された値は、LODの2分の1の値などのデフォルト値を代入できる。しかし、こうした代入法は対数の値の平均値と標準偏差の偏向した推定値を生じることになる。これはつまり95パーセンタイル値での95%UCLの推定値及び99パーセンタイル値での点推定値に影響を及ぼすということである。Helsel(2005)が以下のような解説を加えている。

> 代入法は、それぞれの不検出物について記録された検出限界の関数とした値をおくことにより、打ち切られたデータの統計量を算出する。次いでこれらの設定した値と、検出限界以上で測定された値とを併せて利用して統計量、仮説検定、回帰モデルを計算により決定する。置換法は広く使用されているが、統計的基礎を持たない。他の2つの手順に比較して、置換法はあまり適切に機能しないということが多数の出版物で明らかにされている・・・・・・数値の代替は容易であるため、最も頻繁に利用される。特別なソフトウェアも必要ではない。しかし、値の打ち立てに正当性を与えることはできない。…代替させる値の決定は恣意的であり、統計結果は通常その代替した値に左右されるからである。
>
> 置換法が最も機能しないのは、複数の検出限界がある場合である。代入する値は、分析精密度やサンプルマトリクス中の妨害物質など、検出限界を決定する条件に左右される。代入する値は必ずしもサンプルの真の値と相関関係を持つものではない。複数の検出限界がある場合、最尤推定法またはノンパラメトリック法のどちらかの方が、置換法よりもはるかに優れた機能を発揮するだろう。

こうした置換法の欠点は、特にLOD未満の値がデータセットのかなりの割合を占める場合に明らかとなる。多様な意見があり、データセットの性質によって意見も異なるが、ひとつの見解としては、このような代替的な「穴埋め的」値がデータセットのおおよそ 10～15%を上回る場合、そうした値は重要性を持ち得、推定平均値と標準偏差推定値を大幅に偏らせる可能性がある（Helsel, 1990; US EPA, 2006）。したがって、圃場残留試験データの相当な割合がLODを下回る場合（10～15%より大きい場合）、データセットのより正確な推定平均値と標準偏差推定値を得るために、より精緻な代替法（imputation）が必要であろう。もしくは、MRLの推定値を算出する他のアプローチや手法があるはずである。

VI. 代替的手法及び補足的手法

NAFTA作業グループは前述の95/99ルールの限界を認識し、MRLの算定のための代替的手法を調査した。データセットが以下の基準のいずれか1つを満たす場合、95/99ルールを補足するために、他の手法が開発された。まず、小さなサンプルサイズのデータセットが対数正規性を有すること、次に対数正規仮定が棄却されること、最後に、データセットがLODより低いサンプルを相当な割合で有していること、である。これら代替アプローチの専門的詳細を以下に議論し、それら計算法の採用の合理的根拠を順に述べる。

A. UCL中央値95法

1. 専門的詳細

NAFTA対数正規法と同様、UCL中央値95法もまた高い信頼性の下で、上限パーセンタイル値の算定を試みるものである。しかしながら、この代替手法は次のことを認識している。サンプルサイズが小さい場合、検討対象母集団の平均値と標準偏差を正確に推定するのがより困難になるため、統計分析の結果の信頼性が低下すること、および母集団の歪度が増加するとき、この推定はより困難となることである。この場合、四分位数は分布の裾に過度に影響されないため、実験で実証された四分位数に基づく推定値がより信頼できるものとなろう。このように、95/99ルールとUCL中央値95法の主要な相違点は、後者が圃場残留試験で観察された残留中央値を使い対数（残留値の対数）の推定平均値を出し、さらに母集団平均値が標準偏差と等しい（すなわち、変動係数が1）と仮定し

ていることである[14]。サンプルサイズが小さく、歪度が大きくとも、母集団の中央値は相対的に真の値に近く推定できる。平均値が標準偏差と等しいと仮定すると、計算法は単一のパラメーターによる計算で済み、母集団の平均値または標準偏差のどちらの推定値も必要としない。この計算法もまた、データが対数正規分布するとの仮定に基づき公式化されている。

対数正規分布において、中央値（すなわち、50 パーセンタイル値，$p_{0.50}$）は残留農薬濃度の対数の平均の指数関数 $\exp(\mu_{\log})$とほぼ等しい。この指数関数は幾何平均とも呼ばれる。このようにして、以下に示すように中央値を 95 パーセンタイル値の計算に代替的手法として使うことが可能である。

$$\hat{p}_{0.95} = \exp(\mu_{\log} + 1.645\sigma_{\log}) = \exp(\mu_{\log})\exp(1.645\sigma_{\log})$$

$$= p_{0.50}\exp(1.645\sigma_{\log}) \qquad (12)$$

ここで、$\mu_{\log}$と$\sigma_{\log}$はそれぞれ対数の平均値と標準偏差。$p_{0.50}$は 50 パーセンタイル値。

また、対数正規分布では対数の標準偏差は変動係数（coefficient of variation; CV）として表されうる；$\sigma = \sqrt{\ln(1+CV^2)}$。$CV$=1 となる場合、(12) 式は以下のように書き換えることができる。

$$\hat{p}_{0.95} = p_{0.50}\exp[(1.645)\sqrt{\ln 2}] = 3.9\,p_{0.50} \qquad (13)$$

すでに見たように、式（12）では未知のパラメーター $\sigma_{\log}$ があるが、式（13）では $\sigma = \sqrt{\ln(1+CV^2)} = \sqrt{\ln(2)}$ を代入する。この仮定は圃場残留試験データセットで実測された CV により支持されている。

合法的に農薬を処理された作物中の残留物が MRL の値を上回らないよう MRL の値が十分に高いことを保障するため、中央値の 95%UCL 値を計算し、それに 3.9 を乗じる[15]。

$$UCLMedian95th = \exp(\mu_{\log} + 1.645\sigma_{\log} + t_{0.05,n-1}\frac{\sigma_{\log}}{\sqrt{n}})$$

$$= \exp(\mu_{\log})\exp(1.645\sigma_{\log})\exp(t_{0.05,n-1}\frac{\sqrt{\ln(2)}}{\sqrt{n}})$$

$$= 3.9\,p_{0.50}\exp(t_{0.05,n-1}\frac{\sqrt{\ln(2)}}{\sqrt{n}}) \qquad (14)$$

---

14 変動係数は（算術）標準偏差を（算術）平均値で除したものと定義される。変動係数が 1 のとき、標準偏差は平均値と等しい。

15 本文書全体を通じて、この手法は UCL 中央値 95 法と呼称される。ところが、SOP 草案と MRL 表計算ソフトにおいては UPL 中央値 95 法と呼ばれている。使用される計算法との整合性を保つため、本手法は UCL 中央値 95 法と呼ぶ。本手法が開発される過程で、各種の代替的計算法が検討され、そのうちに予測区間上限（upper prediction interval; UPL？）を利用するものがあった。誤称はその時に生じたものである。SOP 草案と表計算ソフトが改訂される際にこの誤称も訂正されるであろう。

ここで、$t_{0.05,n-1}$ はｔ‐統計量の値。n はデータセットのサンプルサイズ。

B.　平均値＋3SD 法

この手法は当初、カリフォルニア農薬規制部門（the California Department of Pesticide Regulation; CDPR）から出席している作業グループのメンバーから提案された。このアプローチでは、サンプルの平均値に、サンプルの標準偏差を３倍した値を加えた値を MRL に採用しようとするものである。

$$MRLs = \bar{x} + 3s \qquad (15)$$

ここで、$\bar{x}$ と s はそれぞれサンプルの平均値と標準偏差。

本手法は当初、正規分布の特性に基づいて開発された。平均値μと標準偏差σを持つ正規分布において、CV が１のときμ＋3σは 99.87 パーセンタイル値を表す。CV が１でない数値をとるときは、μ＋3σは相当高いパーセンタイル値を表す。μとσは未知数なので、推定値 $\bar{x}$ と s で置き換えられる。圃場残留試験データが正規分布することはあまりないので、これらの平均値と標準偏差の推定値が不正確である場合、$\bar{x} + 3s$ はより低い、またはより高いパーセンタイル値を表すであろう。しかしながら、この値は何らかの最高の（high-end）パーセンタイル値を与えると期待される。本手法をチュビシェフの不等式に照らしてみると、その公式は異なる解釈が可能であり、データの分布に関わらず適用できる。

1.　チュビシェフの不等式の解釈

チュビシェフの不等式は、平均値μと標準偏差σを有する任意の X はすべて、いかなる正数 k についても次の式が真であるとする。

$$P(|X-\mu| < k\sigma) \geq \frac{k^2-1}{k^2}$$

$$\Rightarrow \quad P(X < \mu + k\sigma) \geq \frac{k^2-1}{k^2}$$

$$\Rightarrow \quad P(X < \mu + 3\sigma) \geq \frac{3^2-1}{3^2} = 0.89 \qquad (16)$$

チュビシェフの定理ではデータの分布状態に関わらず、平均値と、標準偏差を３倍したものとの和が、母集団の少なくとも 89 パーセンタイル値を表す。繰り返すが、平均値μと標準偏差σは未知であり、サンプルの推定値 $\bar{x}$ と s で置き換えられる。これらの値は真の平均値及び標準偏差を表さないため、その計算は少なくとも 89 パーセンタイル値を得ることを保障しないが、それでも上限パーセンタイル値を与える。

チュビシェフの不等式により、母集団の少なくとも 95 パーセンタイルを捕らえる k の値を計算（そして利用）できる。しかしながらこの経験に基づく推定値は MRL 設定という規制目的にとっては非実用的に高いものである。それより、母集団の少なくとも 89%を捕らえようとする推定値である k=3 の値が選択される。セクションⅨの B のシミュレーションは、μ＋3σの推定値が一定のデータセットでは、他の多くの手法よりも相当程度優れた高いパーセンタイル値を与えることを実証している。本手法の唯一の限界は、95 パーセンタイル値での 95%UCL 推定値でないということだ。しかしながら、母集団の高い割合を捕らえる上限パーセンタイル値ではある。

2.　対数正規仮定の代替法

残留物データは負の値をとらず、正の向きに歪む傾向があるため、一般に正規分布より対数正規分布に接近する。しかし、データの性質の仮定について、その検証をすることなく信頼するのは不適切である。したがって対数正規仮定が妥当であると判断できない場合は、95/99 ルールに従って MRL の計算をするのは不適切であろうし、その他データの対数正規性に依拠するいかなる手法も不適切となろう。こうした状況では、平均値+3SD 法が推奨される。平均値+3SD 法は、チェビシェフの不等式を使用した解釈において、いかなる分布状態にも当てはまるものであり、データの分布に関する仮定が必要ないからである。必要最小サンプルサイズの補足的ガイダンスについてはセクションIXの C を参照。

C. 打ち切られたデータセットの最尤推定法

検出限界（LOD）を下回ると報告された残留物データを含んだ圃場残留試験データ
セットは数多くある。LOD を下回ると報告された値は、通常 LOD の 2 分の 1 の値などのデフォルト値を代入する。ところが、こうした単純な代入法は 95 パーセンタイル値での 95%UCL 推定値及び 99 パーセンタイル値の点推定値を偏らせる結果を導く。特に LOD を下回る値がデータセットの相当な部分を構成する場合にこのことがあてはまる。それゆえこのような場合には、データセットのより正確な推定平均値と標準偏差推定値を得るため、より精緻な代替法（imputation）の採用が必要かもしれない。このような現在利用されている手順のうちのひとつに、LOD を下回ると測定された残留物量への代入のため最尤推定法を利用するものがある。この手法は、LOD を下回る測定への対処法に最も適切であるとして多くの支持者を得ている。

1. 専門的詳細

サンプルサイズが n の残留物データセット X が、LOD を下回る打ち切られたサンプル n_{cen} を有すると仮定し、これらの打ち切られたサンプルに LOD の値が与えられるとする。

$$X = \left\{LOD^{(1)}, LOD^{(2)}, \ldots LOD^{(n_{cen})}, x^{(n_{cen}+1)}, \ldots x^{(n)}\right\} \qquad (17)$$

残留物データが対数正規分布から抽出されたと仮定する。目的はデータセット X の値が抽出された対数正規分布の未知のパラメーター μ と σ の推定である。

データセット X が対数正規分布から抽出されたという仮定は、観測値の自然対数が正規分布から抽出されたという仮定と同義である。仮に Y=lnX がデータセット X からの観測値の自然対数を示すとしよう。

$$Y = \ln X = \left\{\ln(LOD^{(1)}), \ln(LOD^{(2)}), \ldots, \ln(LOD^{(n_{cen})}), \ln(x^{(n_{cen}+1)}), \ldots, \ln(x^{(n)})\right\} \quad (18)$$

仮に $F[y, \mu, \sigma]$ と $f[y, \mu, \sigma]$ がそれぞれ累積分布関数(cumulative distribution function; CDF)と確率密度関数(probability density function, PDF)を表すとし、平均値 μ と標準偏差 σ を持つ正規分布の点 y において残留物の対数が抽出されるとしよう。ここでの目的は、これら μ と σ のパラメーターの値を求めることである。*LOD* 未満と測定されたサンプルについて、$y=\ln(LOD)$ の点における CDF、すなわち $F[\ln(LOD), \mu^*, \sigma^*]$ が、パラメーターの値が $\mu=\mu^*$, $\sigma=\sigma^*$ であるとしたときに $\ln(LOD)$ を下回る値が観測される尤度を決定するのに使われる。*LOD* を上回るサンプルについても同様に、$y=\ln(x)$ の点における確率密度関数つまり $f[\ln(x), \mu^*, \sigma^*]$ は、パラメーターの値が $\mu=\mu^*$, $\sigma=\sigma^*$ としたときに $\ln(x)$ と等しい値が観測される尤度を決定するのに使われる。

尤度関数は、サンプルサイズ n の観測から、CDF と PDF の積として表される。

$$L[\ln X,\mu,\sigma]=\prod_{i=1}^{n_{cen}}F\left[\ln(LOD^{(i)}),\mu,\sigma\right]\times\prod_{i=n_{cen}+1}^{n}f\left[\ln(X^{(i)}),\mu,\sigma\right]\qquad(19)$$

尤度関数 L を最大化する 2 つのパラメーターの値、$\mu=\mu_{MLE}$ と $\sigma=\sigma_{MLE}$ を最尤推定値と呼ぶ。尤度関数の対数をとることによって上記の式の結果を和で表すことができる。この和を対数尤度関数 ℓ と呼び、以下のように表す。

$$\ell[\ln X,\mu,\sigma]=\ln L=\sum_{i=1}^{n_{cen}}F\left[\ln(LOD^{(i)}),\mu,\sigma\right]+\sum_{n_{cen}+1}^{n}f\left[\ln(X^{(i)}),\mu,\sigma\right]\qquad(20)$$

尤度関数を最大化する MLE（最尤推定法 the Maximum Likelihood Estimation）パラメーターは対数尤度関数をも最大化するので、対数尤度関数の計算効率を評価する。対数尤度関数を最大化する値を見つけるため、μ と σ に対する ℓ の導関数を以下のように 0 にする。

$$\begin{cases}\dfrac{\partial\ell}{\partial\mu}=0\\[2mm]\dfrac{\partial\ell}{\partial\sigma}=0,\end{cases}\qquad(21)$$

式（21）にニュートン-ラフソン法（Newton-Raphson technique）のような数値計算法を適用すると、対数尤度関数を最大化するパラメーター μ と σ の値を推定できる。関数 l を最大化する MLE パラメーターの推定値 μ_{MLE} と σ_{MLE} は、不検出値（すなわち、LOD を下回る測定値）の代わりに代入する値を計算するのに使われる。こうした代入する値は MLE のパラメーター μ_{MLE} と σ_{MLE} で特定される対数正規分布に常にあてはまる。n_{cen} の代入する値は以下のように決定される。

$$F(i^{\text{th}}\,\text{fill-in value},\mu_{\text{MLE}},\sigma_{MLE})=\frac{i}{n_{cen}+1}\times F[\ln(LOD),\mu_{\text{MLE}},\sigma_{MLE}]\qquad(22)$$

$$i=1,2,\ldots,n_{cen}$$

これらの代入する値は対数変換されたデータに使うものなので、元の尺度での代入する値を得るには、まず累乗（e の累乗）しなければならない。実測データと MLE パラメーター推測値に基づき求められた代入する値を組み合わせる方法は「ロバスト法（Robust Method）」と呼ばれる。国際生命科学機関（International Life Science Institute）の文書の*総暴露量評価*（*Aggregate Exposure Assessment*（ILSI, 1998））と題する付属文書に、こうした代入する値を計算するロバスト法と称される手順が解説されている。

2. 考察

MLE に基づくロバスト法は、打ち切られたデータに対処する高度な代入法である。MLE の利点は、分布パラメーターを推定する際に打ち切られた値の、データセットに対する割合に関する情報の取り込みに、累積密度関数を使用できる点である。打ち切られたデータの割合に応じ、データセットにおける検出下限、分布仮定、代入する値がそれぞれのデータセットで異なる。MLE に基づく代入する値で補完されたデータセットは、LOD の 2 分の 1 の値を不検出値に代入したものよりも元の分布をより正確に表している。このようにして、対数正規仮定に基づく MRL の算定（及びその他の統計計算）はより正確な推定値を与える。

MLE が打ち切られたデータの分析に最適ではあるが、データセット中の打ち切られたデータの割合が増加するにつれ MLE 推定値の偏りが増すことがシミュレーションにより実証された。データセット中の打ち切られたデー

タの割合が大きければ大きいほど、推定値は安定性を欠く。したがって、打ち切りの割合が高い場合は MLE に基づき算出された代入する値には注意すべきである。MLE のもうひとつの欠点は分布仮定を必要とすることである。NAFTA 表計算ソフトにある MLE は、データセットが対数正規分布仮定に基づいている。残留物データが対数正規分布に従わない場合は、対数正規性を仮定する MRL は偏った推定値を生む。

Ⅶ. NAFTA の MRL 方法論

NAFTA 作業グループは、広範囲なブートストラップ･シミュレーション及び圃場残留試験データセットを利用した実践を活用し、多数の異なる MRL 設定の手法を調査した。予期したように、異なる状況下では異なる手法がより優れている。合法的な農薬処理をした作物の押収を避けるのに十分に高い値であり、しかし同時に非合法の農薬使用の検出の見込みをなくすほど非実用的に高すぎない値の MRL の設定という作業グループの目的を、完全かつ一貫して満たす手法はひとつとしてなかった。これを受け、作業グループは一定の統計的及びその他の基準を満たす場合に、各種の手法やアプローチの仕方を組み合わせて利用する一連の手順を作成した。この方法論は、それぞれの手法の利点と限界を考慮するもので、ある検討対象データセットの特性が与えられた場合に最適な手法を選択するよう構築されている。次のセクションで、各手法を組み合わせることについての概括的な論理的根拠を検討し、各種の手法及びアプローチがどのような場合に使われるべきかを示したフローチャートを紹介する。

A. 各種の手法を組み合わせる論理的根拠

MRL の算定には様々な統計的アプローチの仕方があり、それぞれが異なる状況において一定の利点を有する。特定のデータセットに対して適切な手法を選択するには、データの分布特性、データセットの大きさ、及び打ち切られた値がデータセットに占める割合の検討を要する。すべてのデータセットに適している普遍的な手法はひとつもない。データセットが異なれば、データセットが満たすアプローチの仮定も異なるだろう。したがって、手法の性能を最も左右するところのデータセットの性質を確認した後に、使用する手法を選ぶのが最も妥当であろうと NAFTA 作業グループは結論付けた。これらのことに基づき、作業グループは MRL の設定に以下の４つのアプローチを検討すべきと考える。95/99 ルール、UCL 中央値 95 法、平均値＋３SD 法、そして MLE に基づく代替法である。

理想的な状況は、すべての圃場残留試験データセットが相当数のサンプルから成り、その全部が検討対象農薬の検出下限を満たし、対数正規分布に近似するときである。そのような状況では、95 パーセンタイル値での 95%UCL 値を提供する NAFTA 対数正規法以外の手法を検討する必要性は殆どない。すでに述べたが、95 パーセンタイル値が過小評価されないことを保障するために設定する高い信頼度（95%）の故に、非実用的に高い MRL を設定してしまうのを避けるために 95/99 ルールは考案された。サンプルサイズが小さくなれば、95 パーセンタイル値での 95%UCL の計算に使われる係数（Hahn と Meeker が *g* とする係数）は必然的に大きくなる。これはもっぱらサンプル数に起因する UCL の相対的増加という結果を招く。小さいサンプルサイズ（すなわち 23 を下回るサンプル数[16]）では、95 パーセンタイル値での 95% UCL よりも 99 パーセンタイル値が好んで選択される。99 パーセンタイル値での最適推定値が選択される場合、95/99 ルールから考えて、より高いパーセンタイル値を得る代わりに、高い信頼性（95 パーセンタイルを捕らえていることに対する信頼性）を犠牲にすることも許容される。

95/99 ルールは、UCL の計算に使われる *g* 係数に対する小さなサンプルサイズの影響を軽減する。しかし、推定平均値及び標準偏差推定値に小さなサンプルサイズが与える影響を軽減することにも本作業グループは興味が引かれた。ひとつのサンプルから計算される平均値と標準偏差は、そのデータが抽出された元である母集団の真の平均値と標準偏差の推定値にすぎない。サンプルが小さければ、それだけ推定値も不正確となりやすい（これは g 係数が大きくなる理由を説明

---

16 99 パーセンタイル値の推定値と 95 パーセンタイル値での 95%UCL の推定値を計算する公式は極めて似ている。概して、これらの計算は標準偏差と平均値の和の倍数である。99 パーセンタイル値での点推定計算に使用される値は、サンプルサイズが 23 を下回る場合に 95 パーセンタイル値での 95%UCL を計算するのに使う g 係数より小さい。

している)。これは特に、対数正規分布などのひずみのある分布状態において問題となる。この影響を最小化するため、作業グループはUCL中央値95法を作成した。CVが1であると仮定することにより、特定のパーセンタイル値におけるUCLを計算するのに1つのパラメーターの推定のみを必要とする。CVが固定される場合、対数正規分布の性質に基づき95パーセンタイル値とその他のパーセンタイル値と間の様々な相関関係を定式化しうる。中央値はサンプルサイズから受ける影響が最も少ないため、この計算法で利用している。このようにしてサンプル数が15を下回るデータセットに関しては、標準偏差推定値が非実用的に高いMRLを与えないよう保障するため、代替法が使われる。UCL中央値95法の結果と95/99ルールの結果を比較し、どちらか値の小さいほうをMRLとして選択する。

MRLの計算に影響を与えるもうひとつの要因は、検出下限(LOD)[17]を下回ると報告されたサンプル数である。こうしたデータはより一般的には打ち切られたデータと呼ばれる。打ち切りの度合い、及びこれら打ち切られた値に代入する手法次第で、打ち切られたデータに基づくパラメトリック計算は偏りをもつ。打ち切られたデータには、通常LODの2分の1の値を代入する。しかしながら作業グループは、前のセクションで論じた最尤推定法(MLE)に基づく、より複雑な代入法の使用を推奨する。MLEでは、検討対象データの分布仮定の特定が必要であり、この場合は対数正規分布仮定である。MLEでは、対数正規パラメーターの推定に検出可能な残留物濃度の値とLODを下回るデータの割合を使う。これら対数正規パラメーターの推定値を計算して、対数正規分布にあてはまる代入値をだす。打ち切りデータの割合が10%を超える場合は、0でも、LODの2分の1の値でも、LOD自体の値でもないMLEに基づく代入値[18]をLODを下回るデータに代入する。次いで、これら代入値と検出可能な濃度の値を使い、通常通りMRLの値を計算する。MLE法は打ち切られたデータセットのパラメーターの推定に有効であるが、打ち切りの度合いが上がるにつれこの手法の信頼性は低下する。データセットにおける打ち切られたデータの割合に関する情報は有益であり推定の過程でも活用されるが、打ち切りそのものはやはり情報の欠如を意味する。情報の欠如が大きすぎると、MLEパラメーターの推定値の信頼性は下がる。60%を上回るデータが打ち切られている場合、この手法の利用には注意が必要である。

ここまでに検討したMRLの計算法－－95/99ルール、UCL中央値95法、MLEに基づく代入法－－は、すべてデータセットの対数正規分布仮定に基づくものである。この仮定は残留物濃度データにとっては妥当なデホルト仮定であるが、我々作業グループはこの基準を満たさないデータセットがあることに気が付いた。そうした場合、対数正規仮定に基づく計算法の使用は不適切となろう。対数正規仮定の妥当性が確認されない場合、MRLの設定に使う上限パーセンタイル値の計算にはその他の手法を使うことが必要となろう。これを受け、作業グループは対数正規仮定が適切か否かを確認する2つの相互補完的な手法の利用を提案した。この仮定の適切性を視覚的に評価する定性的手法として、確率プロットを選んだ。確率プロットから得た相関係数(r)に基づく正規性検定(対数変換したデータで行う)を、確率プロットを補完する定量的手法として選択した。これら2つの手法の結果に基づき、最終的にレビューアーがデータの対数正規性を決定する。

対数正規仮定の適切性が確認されない場合、作業グループは代替的手法として平均値+3SD法の利用を提案した。同手法はもともとCDPRの作業グループのメンバーから提案されたものである。前のセクションでより詳しく検討したが、この手法では算術平均値に標準偏差の3倍の値を加えるもので、経験的に有効性が確認されている。作業グループのブートストラップ・シミュレーションの実施から、平均値+3SD法が上限パーセンタイル値を与え、そのパーセンタイル値はデータを抽出した母集団の分布の性質の影響を受けることが明らかであると分かった。チュビシェフの不等式は、いかなる分布状態においても真であるところの、平均値、標準偏差、及び分布におけるパーセンタイル値の相関関係を明らかにする。チュビシェフの不等式を使って平均値+3SD法による結果を解釈すると、少なくとも89%の分布が平均値の標準偏差を3倍したものにあてはまる、つまり、結果は分布の*少なくとも*89パーセンタイル値を表していることが分かる。仮に計算に使われた値が*真の*平均値と標準偏差である場合、値が不正確である確率が0なので、上述の内容に従うと、89パーセンタイル値での100%UCLとなろう。けれども、このようなデータの真の平均値及び標準偏差が得られることは

---

17 LODが現在検討中の打ち切りの閾値であるが、定量限界(limit of quantitation; LOQ)を閾値とすることもまた可能であり、LODと同様の議論があてはまる。

18 作業グループは、こうしたMLEに基づく代入値を計算するエクセル®表計算ソフト(MRL表計算ソフトとは別)を作成した。

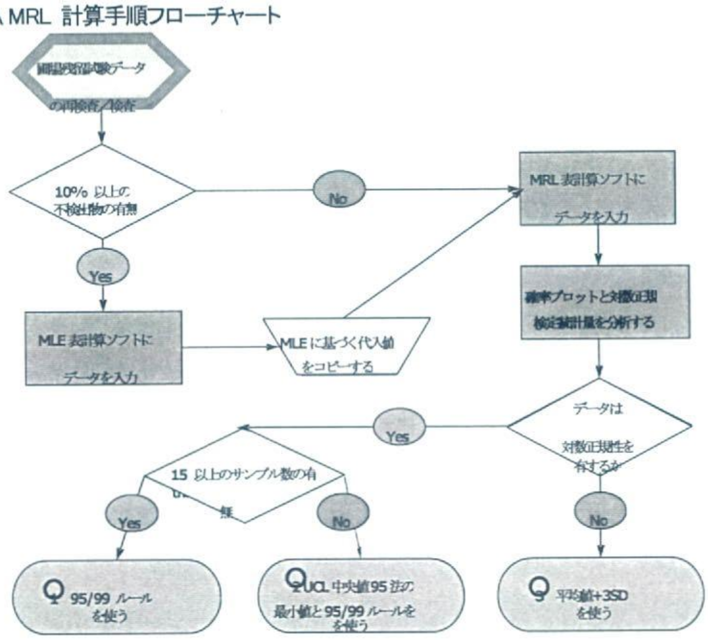
ないため、平均値及び標準偏差の推定値が使われる。このように、対数正規仮定の適切性が確認されない場合、平均値＋3SD 法を MRL の計算に使う。データセットの平均値と標準偏差の推定値が、それらの真の値に的確に近似しているとの仮定に基づく解釈であるとの認識に立てば、この MRL の値は、データが抽出された分布の少なくとも 89 パーセンタイル値であると解釈される。

要約すれば、農薬の MRL の設定のために圃場残留試験データを評価する際生じうるデータに固有の懸案事項や課題に対処するため、作業グループは本質的に異なる手法を理論的に分かりやすく組み合わせた決定アルゴリズムを考案した。この方法論により、MRL の計算に最適な手法の選択を左右するデータセットの性質を評価する体系的なアプローチが可能となる。作業グループは、この方法論にしたがって得られた MRL の計算法の適切性を評価するために相当な努力を払った。こうした努力の結果から、従来行われていたよりも一層調整され調和の取れた方式で、統計的基礎を有し、科学的正当性を主張できる MRL の設定に使用しうる方法論の開発に成功したものと作業グループは自負する。

B. フローチャート

圃場残留試験データに基づく農薬トレランス設定のためのガイダンス（NAFTA, 2005）と題した SOP の草案で、作業グループはデータセットの特質に基づき MRL の計算に最適な手法を選択する決定アルゴリズムを図示するフローチャートを紹介した。
図は、若干修正を加えた、前のセクションにおける検討に基づき適切な計算法を決定する過程を示したフローチャートである[19]。

図2. NAFTA MRL 計算手順フローチャート



C. NAFTA の MRL 方法論における「カットオフ（最小）」サンプルサイズ

統計計算においては、適切なサンプルサイズというものが常に課題であり関心の対象である。サンプルサイズが異なれば、適切に機能する手法も異なりうる。NAFTA　MRL 表計算ソフトは多数の手法を組み合わせているので、「カットオフ（最小）」サンプルサイズは取り組むべき重要な問題である。適正な「カットオフ（最小）」サンプルサイズの決定は主

[19] SOP 草案に載っている元のフローチャートにフォーマットの修正を加えただけである。内容の（したがって MRL 表計算ソフトも）変更はない。

観の枠を出ない。一般に、最小のサンプルサイズからも、作業グループの方法論に従い信頼性のあるしっかりとしたMRL推定値を与える必要があった。作業グループは適切な最小サンプルサイズを決定するのを助けるシミュレーションを行った。セクションIXのBのシミュレーションは、NAFTA　MRL表計算ソフトが10という小さなサンプルサイズから計算したMRLが相対的に狭い範囲にとどまり、妥当な推定値を提供したのを実証している。データセットが10を下回るデータの点しか有しないとき、NAFTA　MRL表計算ソフトによるMRLの計算は、真の95パーセンタイル値を低く見積もる確率が高く、また、その計算はあまり正確でない。*それでもなお、NAFTAの手法は合成及び現実の状況下のデータセットにおいて、その他の手法（UE方法、最大値を「丸める」方法、二項法を含む）に比較しはるかに性能が高い。さらに作業グループはサインプルサイズ6まで適切性を保持できると考える。*これらのシミュレーション結果は本文書のセクションIX（シミュレーションとその結果）で紹介する。

Ⅷ. その他の統計的問題

NAFTA　MRL表計算ソフトが取り組んでいない、または正確に検討していない統計的問題が数多くある。このセクションでは、より重要な概念のいくつかを簡単に紹介し、こうした問題点が統計結果またはMRL表計算ソフトを使用してMRLを設定する手法の厳密さに対し、実際はわずかな影響しか与えないと作業グループが考える理由を説明する。これらの問題点が将来的に再検討され、より全体的な検討が加えられる可能性はあるが、表計算ソフトでの経験は、圃場残留試験データの大部分に適切なMRLの提案をするという事実を証言している。

A.　圃場残留試験サンプルの独立性

米国とカナダでは、圃場残留試験は複数の場所で行われている。通常、各試験地から2つの分析試料が採取される。天候やその他の条件から、同試験地で採取された試料は、異なる試験地で採取されたものよりも類似する傾向がある。換言すれば、同試験地で採取した試料には相関性があり、互いに独立でない。MRL表計算ソフトで行われる統計計算はすべて、分析対象サンプルは独立であると仮定している。圃場残留試験データに観察される非独立性は、より明確には階級内の相関関係と表現できよう。もしくはこの場合より具体的には、圃場残留試験地内でみられる相関関係といえよう。同じ試験地から得られた値の相関関係が強ければ（すなわち試験地内の相関関係が強ければ）、それだけパラメーター推定値の偏りも大きくなるだろう。しかしながら、作業グループはそうした2つの試料を平均化することは情報を失う結果となりうるため、不適切であろうと結論した。

各圃場残留試験地から作物試料が2つしか採取されないため、作業グループは特に95/99ルールにおいて、統計計算に相関関係が及ぼす影響は無視できるほど小さい可能性が高いと判断した。前のセクションで述べたように、95/99ルールは2つの（競合する）推定値の小さい方を選択することから成る。95パーセンタイル値での95%UCLと99パーセンタイル値での点推定である。大きなサンプルでは、99パーセンタイル値での点推定よりも95パーセンタイル値での95%UCL[20]の方が小さいため、通常MRLは95パーセンタイル値での95%UCLで設定される。これは、サンプルサイズが大きくなるにつれ信頼限界上限を計算するg係数が小さく、また一定の割合で減少するという事実による。たとえば、サンプルサイズが16から32に増加するとg係数は概ね2.5から2.2へと減少する。一方小さなサンプルでは、95%UCLよりも99パーセンタイル値での点推定の方が小さいため、99パーセンタイル値での点推定でMRLの設定が行われることが多い。これは、サンプルサイズが小さくなるにつれg係数がより大きく、また一定の割合で増加することによる。たとえば、観測値の数が8から4へ減少すると、g係数は3.2から5.1へと増加する。点推定の計算は、信頼限界の計算に比べてサンプルサイズから受ける影響が少ない。繰り返すが、このように各圃場残留試験地から採取された2つのサンプルがMRL推定値に及ぼす影響は取るに足らないものと作業グループは判断した。こうした当初の見解を検証するため、階級内の相関関係がMRL表計算ソフトの性能に与える影響の度合いを究明する、合成データを使用したシミュレー

20　「通常」という語を使用したのは、対数正規仮定が棄却され（対数正規性を有するデータにおいてさえ）、平均値+3SD法が選択される、またはサンプルの平均値に比較し相対的に標準偏差が大きい場合はUCL中央値95法が使われる場合があるためである。

ションを行った。セクションIXのCにこのシミュレーションを載せた。

B. 作物群間での圃場残留試験の一体化

同じ作物群における食品は、農薬処理が同様に行われれば類似の残留物を有するものと期待される。複数の代表的作物の圃場残留試験データの評価の際、残留農薬の結果が同程度である場合、作物群全体に対する単一のMRLを設定することもある。作物群にMRLを設定するというコンセプト及びその利点が、Food and Feed Crops of the United States (Markle, Baron, Schneider, 1998)) に簡潔に述べられている。

> 各種の農薬利用に関し、1つ1つの食品及び飼料作物の残留物データを集積するよりも、作物群を構成している小数の代表的作物または指標作物のデータを集積する方が、特に規制目的においては時間的、経済的、規制の質の保証の面においてより効率的である。作物群の構成には植物学的関連性が必ずしも有益ではなく、規制目的においては植物学的関連性を持たない作物を同じ作物群に含めることが有益な場合がある。このように、作物群を形成する枠組みの作成には、植物学的関連性、可食部の比較及びその利用法、文化慣習、地理的分布態様、生産方法、飼料、及び加工品が検討要素となる。

作物群MRLの設定をするか否かを決定する際、米国とカナダはそれぞれの作物の適切なMRLの値を決定するため各代表的作物の圃場残留試験データを独立に評価する。異なる作物間において最大残留値がある程度近似している場合（通常5倍を超えない範囲）、作物群のMRLが設定されよう。MRLの設定に利用する観測値の例数を増やすため、作物群内における圃場残留試験データを統合できることが理想的である。サンプルサイズの増加は、母集団の分布のパラメーターの推定の質を向上させることが多いため、MRLの設定に統計手法を利用する場合は特に有利となる。信頼限界に基づきMRLの設定をする場合は（NAFTA対数正規法のように）、サンプルサイズが大きくなれば一般にMRLは小さな値をとる。このことは、母集団パラメーター（例えば、95パーセンタイル値）推定値が真の値により近づくことの信頼性が高いことを反映している。しかしながら、同じ作物群に属する作物の圃場残留試験データに基づく統計計算を行う場合（すなわち、異なる作物の圃場残留試験サンプルを同じ作物からとれたように扱う場合）、残留物は単一の共通の分布状態から導かれたものだと暗黙に仮定する。これが真でないという程度まで、得られたパラメーター推定値は偏りを持つだろう。したがって、異なる作物の圃場残留試験データを一体化する前に、共通の分布仮定の妥当性を検証することが重要である。データセットが同じ母集団分布から抽出されたものである可能性が高いか否かを検証する手法及び／または様々なパラメーター推定値（例えば、中央値または平均値）が相当程度異なるか否かを検証する様々な（パラメトリック及びノンパラメトリックな）統計手法が存在する。

<u>NAFTA　MRL表計算ソフトは、同じ作物群における各種の代表的作物の圃場残留試験データを統合化することの妥当性を決定する統計手順を提供するものではない[21]。</u>将来、作業グループがこの点について取り組むこともあろうが、その頃にはガイダンスも作成されるだろう。それまでは、MRL表計算ソフトは各代表的作物について個々にMRLの値を計算するのに使用されるべきである。米国環境保護庁とPMRAのガイダンスに従い、これら個々のMRLの相互類似性が十分に判定される場合は、作物群全体について単一のMRLが一般に使用される標準として設定されるだろう。

C. 圃場残留試験における地理的／気候的差異

米国（及びカナダ）は特定の作物に基づくMRLを設定するのに必要とするサンプルの数及び地理的位置を決定するガイドラインを持っている（US EPA, 1996）。各地の生産地域間で想定される地理的及び気候的相違に起因する残留物の差異を反映させるため、圃場残留試験は各地で行うことが必要である。また、各地理的領域で必要とされるサンプルの数は、その地域における生産状況を相対的に反映させるよう考えられている。NAFTA　MRL表計算ソフトに

---

21 NAFTA　MRL表計算ソフトに不可欠の部分を成す、対数正規仮定に基づく確率プロットは、2つ以上の分布態様がその形、位置、及び広がりにおいて類似性を有するかを検討するのに理想的であり、その判定を下す統計的検定は容易に利用できるものと考える。

取り入れられている統計手順は、圃場残留試験データが共通の分布状態から抽出されていると仮定する。したがって、地理的／気候的差異による残留値の差異は、母集団内の変動に包含される。地理的範囲が異なれば残留物分布も異なる結果となる可能性もあるが、GAP に従って農薬処理がなされた場合の様々な北アメリカの地理及び気候から集められた圃場残留試験残留物に対数正規分布がよいモデルを与えることを作業グループは確認した。圃場残留試験で残留物が異なる地域／気候帯（ゾーン）から集められた場合に、重要な体系的相違がみられるかを統計分析する計画を作業グループは現在進めている。地域／気候帯内におけるある作物に特定の残留物の差異が、地域／気候帯間における差異とその大きさにおいて類似している場合、地域／気候帯そのものが人工的なものであって、現在の慣行と同様に地域／気候帯ごとの残留物データは統合が可能であることを示唆していよう。しばらくは、米国とカナダは NAFTA の MRL の設定のため北アメリカにおける様々な地理的位置での圃場残留試験データの一体化を継続して行う予定だ。

IX. シミュレーションと結果

作業グループは、どのような状況で各種手法の性能を最も発揮できるかを判定するため、合成データ及び実際の圃場残留試験データ（ブートストラップ・サンプリングを通じて）の両方を利用した多数のシミュレーションを行った。このセクションでは様々なシミュレーションの結果を紹介するが、実際に行ったシミュレーションの結果のすべてを含むものではない。しかし、シミュレーションで使用された SAS コード®は注記されており、本文書の付属書類で参照できる。このように作業グループの行った結果の再現が可能であり、また、作業グループが実施したが本セクションに紹介されていない、変動係数やサンプルサイズの値を変えたものや実際の圃場残留試験データにブートストラップ法で計算したデータを付加的したデータセットなどを使って、他の多様なシミュレーションを行うことも可能である[22]。

このセクションではシミュレーションの結果をまとめるのにフェイリア率、平均平方誤差、及びボックスプロットを大いに活用する。ここでの概括は、調査対象の各種手法の性能評価において主要な手段としての役目を果たす。本文書中で「フェイリア率（failure rate）」とは、ある手法が計算する MRL が、真の 95 パーセンタイル値（合成データつまり、「作られた」データの場合）または経験的 95 パーセンタイル値（実際の圃場残留試験データの場合）を下回る頻度を指す。フェイリア率は割合で報告される。95 パーセンタイル値での 95%UCL が計算された場合、フェイリア率は概算で 0.05、すなわち 5%となる。フェイリア率は可能な限り低い方が望ましい。ところが、フェイリア率が低下すると、それに伴い 95 パーセンタイル値を大幅に高く見積もる計算結果が増えてしまう[23]。そこで、手法の性能の判定に平均平方誤差（mean square error; MSE）をもうひとつの判定基準に使用する。これらシミュレーションにおける MSE は、MRL 算定値と、真の 95 パーセンタイル値（合成データの場合）または経験的 95 パーセンタイル値（実際の圃場残留試験データの場合）との差を二乗したものの平均値として計算される。

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(P_{95} - m_i\right)^2$$

ここで、n はシミュレーションが反復して行われた回数であり（通常 n=1000）、P_{95} は 95 パーセンタイル値、m_i は i 番目のシミュレーションにおける MRL 算定値。

グラフ表示された様々なシミュレーション結果を、ボックスプロット（箱ヒゲ図とも呼ぶ）の形式でも表した。各ボックスプロットは特定の手法による MRL 推定値の分布を視覚的に表現したものである。箱の上辺と下辺がそれぞれ上限四分位数と下限四分位数（すなわち、25 パーセンタイル値と 75 パーセンタイル値）で、箱の内部の線が中央値（すなわち、50

[22] NAFTA　MRL 表計算ソフト（http://www.pmra-arla.gc.ca/english/pdf/mrl/method_calc.xls）の「データ・ライブラリ」ワークシートに実際の圃場残留試験のデータセットがいくつか入っている。ソフトに入っている大きなデータセット（n>25）を使った付加的なブートストラップ・シミュレーションを行うこともできる。

[23] 許容可能なフェイリア率は任意の低さに設定しうる。しかしながら、望ましいフェイリア率を引き下げると（例えば 5%から 1%へ）、当該手法において生じる高い値の数が付随して不可避的に増加してしまう。ゆえに、許容可能なフェイリア率を 5%に設定することは、MRL の設定に内在する「バランス均衡」の別の形での表れである。

パーセンタイル値）である。上限四分位数と下限四分位数の差を 4分位範囲（interquartile range; IQR）と呼ぶ。上部の囲いは、IQR の 1.5 倍分を上限四分位数より上にとったものであり、下部の囲いは IQR の 1.5 倍分を下限四分位数より下にとったものである。「ひげ」の両端は囲いの上部／下部にある最大／最小観測値を示す。縦軸は対数目盛表示された MRL 算定値を表し、横軸は手法の名称またはサンプルサイズのどちらか一方を表している。横軸に水平の直線は真の 95 パーセンタイル値（合成データの場合）または経験的 95 パーセンタイル値（実際の圃場残留試験データの場合）を表す。ボックスプロットはシミュレーション結果を定性的にまとめるものであり、フェイリア率や MSE といった他の定量的な概括法に、図式による概括という補完材料を与える。

作業グループの目的は、95 パーセンタイル値の低い見積もりが頻繁に起こらず、かつ、大幅に高く見積もることもない方法論の構築である。フェイリア率は MRL 算定値が 95 パーセンタイル値を低く見積もる頻度を計るひとつの方法であり、MSE は MRL 算定値の 95 パーセンタイル値との近接の度合いを定量的に測るひとつの方法である。このことから、MRL の計算に上手く機能する手法というのは、低いフェイリア率と比較的低い MSE の*双方を有する*ものだということが分かる。しかし、フェイリア率と MSE は反比例の関係にある。つまり片方の最小化は他方の増加を招くため[24]、両者は競合する基準として使われる。理想的なのは、ひとつの手法が広範囲なサンプルサイズと変動係数のもとでも上手く機能することである。もちろん他の手法より優れている手法というものはある。それゆえ、作業グループはデータセットの多様な性質（サンプルサイズ及び変動係数を含む）に対しても、全体として上手く機能するよう、複数の手法を組み合わせることを試みたのである。

A. 95/99 ルールにおける「カットオフ（最小）」サンプルサイズ

前のセクションで検討したように、95/99 ルールは 95 パーセンタイル値での 95%UCL と 99 パーセンタイル値での点推定のどちらか小さい方を選択するものである。95 パーセンタイル値での 95%UCL のフェイリア率は概算で 0.05 となり、95 パーセンタイル値が過小評価されないことを保障することになる。しかし、サンプルサイズが小さくなれば、95% UCL はそれに伴って大きくなり、95 パーセンタイル値が過小評価されないのを保障することになる。これは逆に MSE の増加を導くことになるが、95/99 ルールを使うと MRL の過大評価を防ぐことになる。95%パーセンタイルでの UCL が 99 パーセンタイル値を超える場合、95/99 ルールではデフォルトの 99 パーセンタイル値をとるが、そのことによって 95 パーセンタイル値より大幅に高く MRL を見積もらないことを保障しており、また、フェイリア率が 0.05 よりも大きくなることを意味している。シミュレーションの目的は、フェイリア率または MSE が許容できないほどに大きくなるサンプルサイズを判定することである。

シミュレーションには、対数正規分布から様々なサイズのサンプルを抜き出すことが含まれる。それぞれのサンプルから 95/99 ルールを使い MRL が計算される。この過程が各サンプルサイズについて 1000 回繰り返し行われる。最後に、これらシミュレーションの結果が、母集団の対数正規分布の真の 95 パーセンタイル値に基づいてフェイリア率及び MSE としてまとめられる。作業グループは様々な CV（CV＝0.75, 1.0, 1.25, 及び 1.5）を持つ対数正規分布のシミュレーションを行った。様々なサンプルサイズについての多様な CV を持つ対数正規分布のフェイリア率及び MSE を表 1 に載せた。様々なサンプルサイズ及び CV におけるシミュレーション結果のボックスプロットは図 1 にある。

表 1 と図 1 の結果から、サンプルサイズが小さくなればフェイリア率と 95/99 ルールの MSE がそれに伴い増加することをシミュレーションにより実証されたことが分かる。しかしながら、サンプルサイズが 11 よりも小さい場合（すなわち、n<11）、フェイリア率はさらに高い増加率で増加する。CV が 0.75 のとき、n<11 では MSE の変化率は顕著に増加す

24 例えば、同じ相対範囲にわたる推定値（例えば、最大推定値が最小推定値の 2 倍の大きさの場合）を生じる 2 つの手法を考えてみる。第 1 の手法では推定値が大体 95 パーセンタイル値付近に集中しており、第 2 の手法では推定値が中央を外れてその 90%が 95 パーセンタイル値を上回っている。明らかに、第 1 の手法のフェイリア率（約 0.50）は第 2 の手法のフェイリヤー（約 0.10）よりも高くなろう。しかしながら、第 1 の手法においては 95 パーセンタイル値と最大及び最小推定値の間の距離がほぼ等しくなるため、MSE は低くなろう。一方、第 2 の手法においては 95 パーセンタイル値と最大推定値の間の距離がかなり大きくなるため（実際の MSE の算出には距離を二乗する）、MSE は比較的高くなるだろう。

る。しかし、CV が 1.0 のとき、n が 15 から 14 に減少すると MSE は大幅に変化する（30 ポイントまで）。

1000 回の計算を行い、そのうち既知の 95 パーセンタイル値を下回る MRL 算定値の割合であるフェイリア率と、既知の 95 パーセンタイル値との誤差を二乗したものの平均である MSE は、95/99 ルールがどのサンプルサイズまで適切な MRL を推定できるかを判定する 2 つの判断基準である。特定のサンプルサイズにおける高いフェイリア率は MRL を低く見積もる確率が大きいことを意味し、また逆に低いフェイリア率はその確率が小さいことを意味する。高い MSE は推定値が真の値から大きく逸脱していることを意味する。これら 2 つの判断基準のうち、フェイリア率を「カットオフ（最小）」サンプルサイズを決定する第一の手段とすべきである。

*シミュレーションの結果に基づき、95/99 ルールにおける「カットオフ（最小）」サンプルサイズは 15 に選択された（NAFTA の方法論全体におけるものではない）。*データセットの対数正規性が妥当だと認められる場合、n≥15 のときは NAFTA　MRL 表計算ソフトで MRL の算出に 95/99 ルールを使うことを作業グループは決定した。15 以上のサンプルのデータセットについては、95/99 ルールは相当程度まで過大見積りすることなく、許容できるフェイリア率で 95 パーセンタイル値を安定して与える。

表1.

対数正規分布　μ ＝1				
サンプルサイズ	フェイリア率 (MSE)			
	CV＝0.75	CV＝1.00	CV＝1.25	CV＝1.50
n＝4	0.34 (333)	0.34 (936)	0.34 (3543)	0.34 (11030)
n＝6	0.25 (134)	0.26 (444)	0.26 (1405)	0.26 (3674)
n＝8	0.21 (87)	0.21 (330)	0.21 (1004)	0.21 (2544)
n＝10	0.17 (70)	0.17 (255)	0.17 (751)	0.17 (1848)
n＝11	0.14 (61)	0.14 (218)	0.14 (614)	0.14 (1436)
n＝12	0.13 (58)	0.14 (206)	0.14 (581)	0.14 (1363)
n＝13	0.12 (58)	0.12 (204)	0.12 (568)	0.12 (1319)
n＝14	0.11 (58)	0.10 (192)	0.10 (528)	0.10 (1205)
n＝15	0.10 (54)	0.08 (169)	0.08 (453)	0.08 (1021)
n＝16	0.10 (53)	0.09 (168)	0.09 (453)	0.09 (1021)
n＝17	0.10 (49)	0.08 (170)	0.08 (458)	0.08 (1031)
n＝18	0.07 (47)	0.07 (149)	0.07 (392)	0.07 (867)
n＝19	0.08 (47)	0.07 (152)	0.07 (403)	0.07 (893)
n＝20	0.06 (45)	0.06 (139)	0.06 (363)	0.06 (795)
n＝25	0.06 (35)	0.05 (111)	0.05 (285)	0.05 (612)