

平成10年度 厚生科学研究費補助金
生活安全総合研究事業 実績報告書

「食品の試験・分析の信頼性確保
に関する調査研究」

(H-10生活-001)

財団法人食品薬品安全センター

代表者 柳澤健一郎

食品の試験・分析の信頼性確保に関する調査研究

主任研究者 柳 澤 健 一 郎 （財団法人食品薬品安全センター理事長）

研究要旨 食品衛生検査の重要性に鑑み、検査の信頼性の確保のために食品衛生検査施設の GLP規制が実施され、検査体制の充実と精度管理の徹底が求められることになった。精度管理制度は、平成9年度より実施されたが、その実質的な内容と効率的な運用にはなおいくつかの検討課題を残している。食品衛生検査の精度管理の実態を適切に把握するため、外部精度管理調査が必要である。この研究では、外部精度管理調査のあり方を検討し、適正で実質的な調査を効率的に行うために考慮すべき諸問題を検討した。まず、調査結果の評価のために用いられる各種の統計学的手法の検討、つぎに検査方法と評価指標の再吟味、また検査のための適正な試薬や標準品の確保をいかにすべきかの検討、精度管理調査試料の作製方法の検討を行い、さらに精度管理調査作業の効率化についての検討を行った。

分担研究者

小野 宏 ㈲食品薬品安全センター
秦野研究所・所長
内山貞夫 ㈲食品薬品安全センター
秦野研究所・部長
大隅 昇 文部省統計数理研究所・教授
松田りえ子 国立医薬品食品衛生研究所・
主任研究官

査、精度管理調査のために理化学検査用および細菌学的検査用の試料の作製と使用可能性の評価、それに、調査業務の効率化の諸課題である。

B. 研究方法

(1) 精度管理の統計学的解析方法について、これまでの調査結果の解析では、 $\bar{x}-R$ 管理図を用いる方法と、国際的に利用されている \bar{z} -スコア方式（偏差値方式）とによって評価してきたが、多機関で同時に作られるデータを基にした評価のあり方としては問題があり、こうした方法を検証し新たな統計解析方法を探った。

(2) 検査機関における精度管理の実態を把握するために、全国の食品衛生検査施設についてアンケート調査を行って、分析法バリデーションと内部精度管理の実施状況について検討を行った。

(3) 精度管理において必要となる食品関連化学物質の適正な試薬・標準品を確保し、供給

A. 研究目的

食品検査の精度管理事業を円滑に進める上で、解決すべき問題がまだ残されている。精度管理調査を適切に行うべき検査試料の質を確保し、物品とデータの送受を円滑に行い、結果の解析と評価を適切に行うべきこと、等がそれである。本研究では、それらの課題に応えるために5課題の分担研究を組織した。まず、検査結果の統計処理法の吟味と評価法の検討、参加する検査機関の意識と実態の調

できるような支援体制のためのリファレンスセンターの実現性を調査、検討した。

(4) 外部精度管理調査のためには、適正な調査試料を準備しなければならないが、まず、専門家に聞き取り調査を行って、問題点を数えあげた。理化学的検査のためには、ソルビン酸について各種の食品の調査を行い、細菌学的検査のためには、マッシュポテトを基礎にした試料を試作して、協力検査機関に送付して検査を依頼した。

(5) 調査試料の送付・検査結果の回収・評価する調査機関と参加機関との間のデータ送受の正確性と評価の効率化の検討を行い、電子通信の手段を利用することを試みた。

C. 研究結果

(1) 外部精度管理調査における統計解析・評価方法の再検討を行い、理論的にまた具体的なデータの検証によって、種々の問題点が指摘された。

(2) 検査機関へのアンケート調査により、分析法バリデーション、内部精度管理の実態を調査した結果、それらの分析法管理はすべての機関が完全に対応しているとは言えないことが分かった。また、管理を行うに当たっても管理の目標値の設定や指導が必要と認められた。

(3) 精度管理に使用する試薬や標準品の国内における標準物質の供給の実態を調査した結果、早急に供給体制を整備すべき必要性が認められた。また、提供されている標準品について純度の調査を行ったところ、やや純度の低いものが認められた。

(4) 食品衛生外部精度管理調査のための共通試料の作製に関する検討を行った。これまでの調査についての聞き取り調査を行い、実際の食材に近い試料の開発が必要であり、均質性と安定性の高い調査試料が求められていることが判った。そこで、理化学的検査および微生物検査のための試料の作製を種々試み、

使用可能な食材の見当が得られた。

(5) 外部精度管理調査の実施における調査機関と検査機関の連絡およびデータ転送に電子媒体を用いる方法を試み、モデル連絡調査を行って実用化できる可能性が高いことを認めた。

D. 考察

(1) 現行の外部精度管理調査における統計解析・評価方法の再検討を行い、理論的にまた具体的なデータの検証によって、種々の問題点が指摘された。今後、より適切な評価方法に進めたい。

(2) 検査機関の検査方法の管理について、アンケート調査を行った結果、分析法バリデーション、内部精度管理の現状を調査した。それらの管理の目標値の設定や指導が必要と認められた。

(3) 精度管理業務を円滑に進めるためには、適正な試薬・標準品の確保と安定供給が必須であると考えられるが、国内における標準物質の供給の実態を調査し、早急に供給体制を整備すべき必要性が認められた。また、市場に提供されている標準品について純度の調査を行ったところ、やや純度の低いものが認められ、この点の管理体制の整備も必要と考えられた。

(4) 食品衛生外部精度管理調査において熟練度を調査するための共通試料は聞き取り調査によれば、実際の食材に近い試料の開発が必要であり、均質性と安定性の高い調査試料が求められていることが判った。さらに、理化学的検査および微生物検査の試料について、それぞれ問題解決のための具体的着想も得たが、なお検討が必要である。

(5) 外部精度管理調査の調査機関と検査機関の連絡およびデータ転送に電子媒体を用いる方法は有用である可能性を認めた。これによって、今後より正確性の高い効率的な調査が可能となると見込まれる。

E. 結論

「食品の試験・分析の信頼性確保に関する調査研究」を5つの分担課題に分けて行い、外部精度管理調査のために有益な知見を得た。

(1) 精度管理の統計学的解析方法については、 $\bar{x}-R$ 管理図と、 z -スコア方式とによって評価してきたが、こうした方法を検証し新たな統計解析方法を探った。

(2) 検査機関における精度管理の実態を把握するために、全国の食品衛生検査施設についてアンケート調査を行って、分析法バリデーションと内部精度管理の実施状況について検討を行った。

(3) 適正な試薬・標準品の確保と供給のためのリファレンスセンターの実現性を調査した。

(4) 適正な調査試料の作製方法の開発を試み、とくに実際の食品形態に近い検体の作製方法を検討した。

(5) 調査試料の送付・検査結果の回収・評価の調査機関と参加機関との間の連絡の正確性と効率化を検討した。

F. 研究発表

なし

G. 知的所有権の取得状況

なし

1. 精度管理に必要な統計学的処理の手法に関する検討

大隅 昇 文部省統計数理研究所 調査実験解析研究系 教授

研究要旨

食品の試験・分析における精度管理を行う上で必要とされる各種の統計的手法の適用可能性を探ると共に、測定データ取得環境とデータ処理との関連において生じる諸事象について、統計科学の観点からの総合的な検討を行った。また、精度管理にとって有効な統計手法やデータ解析のあり方についての方法論の基礎的考察を行った。

A 研究目的

精度管理を議論するうえで、測定データの取得環境・実験の計画の検討に始まり、具体的に取得した測定データの統計解析やデータ解析を行う際の方法論を一貫した流れの中で研究することが必要とされ、また要請されてきた。統計科学の見地から、精度管理にとっていかなるデータ測定法やデータ解析法が有効であり、またそれらの適用可能性や限界を知ることがきわめて重要である。本研究は、精度管理にとって適切な測定データの取得環境や統計手法の在り方、精度管理を実施する際の具体的な統計的諸問題の解決策、結果の解釈と吟味方法等についての提案を行うことにその目的がある。

B 研究方法

本年度は、平成10年度に実施された「平成10年度食品衛生外部精度管理調査報告書」（参考文献〔1〕、以下で「外部精度管理調査」と言う）で報告された調査結果のデータセットのうち、残留動物用医薬品調査で用いられたスルファジミジン測定データを利用して、具体的に統計的データ処理を進める際の諸問題について考察した。

なお、本研究の目的は、外部精度管理調査で取得されるような、一般的なデータ測定環境における取得データを前提としたときに見られる統計的データ解析の対応法あるいは考慮すべき諸事項を、数値事例を通して検証することにある。

従って、取得データの存在を前提とした議論となるが、断るまでもなく、統計的デ

ータ解析を行う上での基本原則は、その測定データがいかなる実験計画の下でいかにして測定されたかというデータ取得環境や取得履歴が明示的に見えること必要とされる。

また、実際の取得データと、それに統計的データ解析の方法論を適用するうえで生じる矛盾や乖離をいかに解消しあるいは回避しながらデータ解析を進めるかの指針を検証することにある。

さらに、ここでは前述のように、「食品衛生外部精度管理調査」に関連して取得された複数の特性データのうち、残留動物用医薬品調査で用いられたスルファジミジン測定データを利用した（後述）。

C 研究結果

ここで、統計的データ解析の方法論や詳細な解析結果を詳述する余白はないので、用いた事例データを引用しながら、精度管理における統計的データ解析手法を適用するうえでの問題点や留意事項はいかなるものであるかの要点を述べる。とくに「外部精度管理調査」で用いられている統計的データ解析の諸手法の特徴（長所、短所）を、適用上の留意事項としてデータの分析事例をもとに述べる。

(1) 測定値の取得環境の留意事項

精度管理に限らず、各種の測定データの分析に統計的手法を適用する際に留意すべき事項として、その測定データの集まりがいかなる測定環境で取得されたかを明示的に示すことが重要である。しかし、現実に

は統計理論が要求するようなデータ取得環境を実現することには無理があって、結果として統計解析手法の適用に際しての拘束や制限が生じることとなる。しかも、統計ソフトウェア等の普及により、測定値としての数値の並びさえあれば、あたかも統計解析が正しく行われたかのごとく計算結果だけが示される（一人歩きする）ことに問題があると言わねばならない。

統計的手法適用上のもっとも厄介な問題として観測値（測定値）の取得方法がある。理論上は、ある「母集団」を仮定してそこから無作為に抽出された標本が測定対象となることが求められる（いわゆる無作為抽出による「無作為標本」）。しかし、現実にはこの操作はほとんど不可能に近いと言わねばならない。仮にある検査施設・機関内における内部精度管理においてさえ、これを精密に遵守したデータ測定はかなり困難と考えねばならない。

次の問題として、取得した測定値から統計的推論を行う上での仮定を設けるか否かがある。とくに、母集団や標本について統計的な意味での分布を仮定するか否かの事がある。これは、統計的有意性検定や推定を行う場合に必要とされる操作である。一般には、正規分布等を前提とするが、その主たる理由は数理的操作を容易にするためのものであって、必ずしも測定値から客観的な結論や解釈を得るための手段とはならないことである。換言すれば、現実の測定値に正規分布や適切な母集団構造を仮定すること自体にかなりの無理があると言わねばならない。このような統計的推論を進める上での多くの制約の中で、現実の測定値の示す特徴なり傾向を知るために統計的データ解析をいかに適用するか、あるいは適用可能であるかの議論がまずなされるべきである。

ではここで、今回の分析対象としたスルファジミジン測定データについて、具体的にこの状況を考えよう。このデータセットの場合、各サンプルすなわち各参加機関（検査施設）を統計的にどう考えるかがある。仮にいま形式的に国内の全検査機関の一覧リストが取得できたものとして、ここから

無作為に抽出するという方法が想定されるが、実際には調査への参加の意思決定は各機関の判断で行われるものであるから、既にこの操作が成り立たない。また別の事として、内部精度管理の各機関の変動の差異が予想されることがある。仮に個々の機関内での内部精度管理が整っていたとしても、その機関で用いる検査機器類の差異（とくに、観測値の精度）、検査担当者の技能の差等の影響は免れないし、それらの影響が混入した結果が取得データであると考えべきである。つまり、個々の機関の測定値には様々の要因の影響が異なる変動を伴って潜在的に累積・混入した結果であると考えねばならない。しかし、これらの影響を客観的に評価するためには、考えられるあらゆる要因を列記したうえで、精密な実験計画のもとにデータを測定することが必要となるが、現実的にはきわめて難しいことと思われる（きわめて大規模かつ緻密なコラボレーションを計画できたとしても非常に難しいことと考えられる）。この内部精度管理に、各機関を比較する外部精度管理の諸事項への配慮が加わることになる。統計的な言い方をすれば、各機関内の変動（級内変動）と各機関間の変動（級間変動）をいかに捕捉するかが課題となるのであるが、実際の取得データに対してこの考え方を適用することは難しく、したがってこれが統計的データ解析における制約ともなる。

(2) 観測値の特徴

次に、観測値（測定値）の測定反復数と数値の測定精度も無視できない。スルファジミジン・データについては、反復数が5回と指定されているが、提供試料から必ずしもこの反復回数を守ることが可能とは限らない。また逆に、必要回数以上を測定して、測定値の記録時に必要に応じてある種の操作（間引き、はずれ値除去など）を行うこともあり得るであろう（ここらを正確に掌握するためには、調査時における測定環境の履歴を詳細記録することが必要とされる）。記録数値の精度（測定可能な数値の桁数）も考慮すべき点である。記録すべき有効桁数の指示があったとしても、その

機関で利用している測定器の精度、S/N 比などに応じて、また場合によっては検査担当者の技能水準の差異を含めて、記録された測定値の有効桁数に違いがあることを前提としたデータ解析となることを容認せねばならず、これが平易に統計ソフトウェアや応用ソフトウェア（エクセル等）による処理では済まないことを示している。ちなみに、スルファジミジンの測定データ（5回の反復測定）のごく一部を無作為に抜き出してみると以下のようにになった。

測定回数				
1回	2回	3回	4回	5回
0.0665	0.0702	0.087	0.103	0.0905
0.1016	0.084	0.0837	0.0687	0.0798
0.0737	0.0893	0.0737	0.0779	0.122
0.14	0.14	0.13	0.12	0.22
0.114	0.144	0.0954	0.114	0.150
0.136	0.148	0.0824	0.0891	0.165
0.123	0.157	0.13	0.165	0.154
0.14	0.13	0.12	0.12	0.22
0.166	0.158	0.158	0.149	0.153
0.144	0.159	0.198	0.139	0.145
0.15	0.16	0.16	0.148	0.168
...

これを観察すると、明らかに記録方法に差異があり、これの統計処理を行う際には十分配慮が必要であることを示唆している（たとえば、平均値や標準偏差の算出桁数など）。

(3) z スコアの特徴と適用上の制約

「外部精度管理調査」（参考文献 [1]）にあるように各機関での測定データに基づき統計的に比較・評価が行われる。このときに利用される判定指標の一つに z スコアがある。なお同報告書では、この他に統計的データ解析に先だって行うべき事前処理（データ・クリーニング、ランドリ）、管理図の代用方法、用いる初等統計量（統計値）等について述べているが、ここでは触れない。

z スコアとは、測定値に単に標準化処理（測定値を平均が 0、分散が 1 と変換する

操作）を行って得られる数値のことであり、それ以上のものではない。つまりこれを求めることがそのまま正規性の判定やはずれ値の厳密な判定規準を与えるものではない（あくまでも測定値が示す分布の特徴を記述する一つ指標にすぎない）。

つぎに、この z スコアをどう利用するかということがある。「外部精度管理調査」では、各機関における 5 回の反復測定値から求めた「平均値の分布」にこの z スコアを適用している。ここでは、スルファジミジン・データを例としての観察・探査から得られた、統計的解析の観点から留意すべき事柄のいくつかについて述べる。

a) 測定値の代表性と等質性

まず重要な事として、各機関からの測定値に基づいて得られた平均値の集まりを一つの（統計的な意味での）分布とみなして評価することを前提としているが、ここに既にある約束（前提、仮定）がある。それは「各機関は原則として統計的に（無作為に）抽出され、また各機関内での測定環境が同等ないしはそう見なせる状況にある（内部精度管理が同等の管理状態にあると見なせる）」ことを前提として集められたデータであることを暗黙のうちに容認した上での z スコア他の適用となることがある。つまり、取得データの代表性と等質性の問題がある。しかし現実にはこの仮定に対する強固な裏付けが必ずしも得られがたい状況下での分析となることを知っておくことが肝要である（つまり、分析結果の解釈時にこれらの前提から出発したことを意識することが必要となる）。

b) z スコアの限界（適用の範囲）

z スコアを用いる際に考慮すべき事として、対象データセットの分布（ヒストグラムを想定すればよい）としての特徴を事前に知る必要がある。とくに、分布について、

- ・単峰性とならない場合
- ・歪みが大きい場合
- ・同じく、測定値が特定の数値の範囲に集中するような場合

・はずれ値が多数見られる場合

などが観察される場合には、 z スコアの適用には十分な注意を要すると考えられる。スルファジミジンの場合、この中の幾つかの要件を満たすので、そのまま z スコアを適用することには若干の無理がある（後述の分析例を参照）。このことは、測定値の初期スクリーニングに加えて、解析対象データセットについても、各種統計値を算出したりその他の統計手法を用いた探索的アプローチが必要であることを示唆している（初動探査の重要性）。少なくとも基本統計量（平均値、分散、歪度、尖度、分位数など）に加えてグラフィカル手法（箱髭図、幹葉図、形式的な管理図法など）を併用したデータ探査が必要となる。

c) 正規性との関連

分布が比較的単峰性を満たし、しかも歪みや尖りが少なく対称性も満たされるような場合には、 z スコアは有効に機能する。こうした状況が保証されるのであれば、形式的には正規性の検定を行うことも考えられるが、前述のことから、測定値の集団を統計的な意味での（標本）分布と見なせるかという条件をどの程度クリアできるかという問題がある。つまり、原則としては、これらの要件がそれぞれ満たされて初めて、正規性の検定などの操作が有効に機能する。しかしながら、スルファジミジンに限らず「外部精度管理調査」の対象となるような各種の検査項目（安息香酸、ソルビン酸、カドミウム、鉛など）は、経験的に見て、総じてこの条件を満たすとは限らず、したがって z スコアを適用した評価方式のみを適用することには慎重な対応が求められる。

(4) はずれ値の問題への対処

はずれ値や異常値への手当をどのように行うかはきわめて重要な課題である。しかし、これを行うための適切な統計的手法があるとは言えない。元来これは、所与のデータセットの観測状況、測定状況に応じて対応が異なるものであり、その結果として汎用的な方法論が期待できることではない。

無数の方法論や技法が提案されていることから、これがいかに難しいか推測される（参考文献 [[4] , [5] ）。

さらに、このはずれ値の問題を統計的方法論で処理するとき、今までの議論と同様に、ある種の前提・仮定を想定した中での対処方法に限られる。たとえば、無作為標本であること、正規性が満たされること、等々である。こうした制約を多々設けることで、数理的な論理展開は容易となるが、実用手法としての適用可能性や有用性からは益々遊離するというジレンマが生じる。この両者の兼ね合い（均衡）をいかに図って解析を進めるかが（現時点における）精度管理のための統計手法のあり方と思われる。少なくともスルファジミジンのデータに関しては、このはずれ値の操作を行わない原始データのみによる評価はほとんど意味を持たないことが観察される（後述）。

(5) 初動探査の有効性とその検証

このように、精度管理においては、単に測定値が取得できたからといって、即応的、短絡的に統計手法が適用できる状況にはないことが類推される。本研究の提案の一つの方向として、以上に述べたことを「データ取得環境の透明性」と「初動探査の重要性」という表現で代表することがここでの主張の要点であり、これは精度管理における具体的なデータ探査の操作として、以下のようなプロセスに要約される。

i) データの取得・測定環境の透明化

取得データの背景情報の確保と評価、測定環境の記録収集、検査担当者の技能評価の問題等

ii) 回収データのスクリーニング方式の検討 フィルタリングの閾値の検討、測定項目特性に合わせた閾値の設定等

iii) 初動探査によるデータの視察

各種統計値（少なくとも、平均値、分散・標準偏差、歪度、尖度；四分位数、メジアン等）およびグラフィカル表現法（箱髭図、ドットプロット図、ヒストグラム、管理図、幹葉図等）の併用による探査

iv) とくに、はずれ値の評価

箱髷図，刈り込み平均，ノンパラメトリックな方法等の検討

v) 以上（とくに，iii），iv）の検証を終えて得られた解析対象データによる統計解析

vi) 必要に応じた上記の過程の反復検証

本年度の研究として，これらのすべてを検証できたわけではない。しかし，統計的手法を適用する際の考え方については参考文献の [2] ~ [4] に，その一部を述べた。

D 事例解析とそれに基づく考察

ここで以上の主な過程を検証するために，スルファジミジン・データを例として，簡単な統計解析を試みる。

(1) 研究対象としたデータセットの構成

前述のように，「外部精度管理調査」に関連して取得された複数の検査項目データのうち，残留動物用医薬品検査調査で用いられたスルファジミジン測定データを利用した。これの概要は以下の通りである（詳細は参考文献 [1]）。

回収データ数（施設機関数）：199

有効サンプル数：197

ここで，除外の 2 サンプルについては，調査報告書にもあるように，第 1 次スクリーニングにおいて測定に誤りがあったと判断されるものであったので，これを除外した 197 サンプルを対象とした。なお，測定データの数値についてはここでは記載を省略する。

(2) 試料の特徴

まず，試料の具体的な調整法であるが，これは調査報告書にあるように，作成濃度を豚肉中に 0.28 ($\mu\text{g}/\text{g}$) として混入させて調整された。これは，統計的な意味での「真値」に相当すると思われる（真値そのものではない）。次に，この調査試料中のスルファジミジン濃度の確認として，20 試料の濃度測定を行いこれの平均値が 0.196 ($\mu\text{g}/\text{g}$)，標準偏差が 0.004 ($\mu\text{g}/\text{g}$) とな

ったと報告されているが，これが一つの「指定値・目標値」と考えられる。したがって，回収率は 70% ($0.196 \div 0.28 = 0.7$) となる。

しかしここで，統計的見地から，いくつかの事項に留意せねばならない。まず，いずれの値を統計的な意味での「母数」に相当する（見なせる）値とできるかがある（母平均値，母分散を含めて）。上では仮に，真値と書いたが，試料作成時に既に多くの攪乱要因が混入することは明らかである（汚れ等）。つぎに，実際に検査機関に供与・配布される試料の確認時の観測値の示す値（上で指定値とした数値）の測定時にも当然のごとくに多くの要因が混入した結果であると考えねばならない（検査担当者の技能，測定器の精度等）。このように考えると，統計的な意味での概念としての「母数」（真値）を仮想的には設定はできても現実の数値としてのそれは掌握できないことを前提として，統計解析手法を議論するというきわめて深刻な状況から出発せねばならない。

とくに，上に挙げた各数値（指定値，回収率等）のいずれもが一つの値として確定されるものではなく，統計的に考えるとある「分布」に従う変動を伴う量であるという点にあるが，現実にはこれを精密に測る妥当な方法が見あたらないことに問題の本質がある（形式的に統計手法を適用して済むということではないという困難性）。ここに，統計的手法を適用する際の大きな限界がある。

一つの解決策は，ある種の要因実験を設計し，これに複数の検査機関が参加したコラボレーションを計画し，測定環境で生じるであろう各種要因の影響度を評価することであろう。しかし，その結果がそのまま一般的性をもって通常の現場の検査環境に適用できるとは限らない（別の影響要因の混入があり得るから）。ここでは，ここの議論はさらに慎重に行わねばならないことであると指摘するにとどめる。

(3) z スコアと諸統計値によるデータの探査

以上に見るように数値として集積された

測定値（観測値）は、様々の要因の影響を潜在的に含んだ変動を伴う数値であると考えねばならない。また、取得データのサンプリング操作に関して、統計的手法を適用する際に考慮すべき母集団や無作為標本の想定（仮定）が十分に成り立たない環境下で得られたデータセットであることも、分析に際して意識せねばならない。これらの制約を念頭に以下の分析を進めた。

a) z スコアによる検討

もっとも簡単な z スコアと各種の統計値を求めることで、スルファジミジンの測定データの特徴を観察する。とくにここでは、各検査機関から回収された「大きさ 5」の測定値の機関別平均値を求め、これの集まり（大きさ 197 の観測値）についての機関別 z スコアをスクリーニングのフィルタとして利用して、以下のような操作を試みた。なお、ここで言う z スコアとは、

$$(z \text{ スコア}) = [(\text{各機関の 5 個の測定値の平均値}) - (\text{各機関の 5 個の測定値の平均値の平均値})] \div (\text{各機関の 5 個の測定値の平均値の標準偏差})$$

を言う。

- ①全サンプル（197 機関）についての平均値から諸統計値を求める。また、各機関の z スコアによりその絶対値が「2 以上」（2 シグマ限界）のサンプルを除外（フィルタリング）する。
- ②この操作で 191 サンプルが残るので、これについて①と同様の操作を行う。
- ③以下、同じ操作を数回反復する。
- ④ここでは以上の操作により、元の 197 個のデータセットの他に 3 組のデータセット（個数が 191, 181, 168 となった）をそれぞれ生成した。

しかし統計学的な見地からは、この操作には問題がなくはない（かなり大胆な操作である）。反復的にサンプルを除外することで、標本抽出を経て得られたと仮定した操作に作為が入るからである。しかしここ

では、元々のサンプル取得が統計的な意味で無作為化がなされていないという状況を考えて（つまりある種の作為標本であるから）、少なくともはずれ値として混在する「大方の測定値が示す（であろう）データの特徴」から、外れた値を（機械的に）フィルタリングすることの必要性もなくはない（かなり苦しい解釈ではある）。こうして得られた結果を表 1, 2 に示した。

表 1 は、k シグマ限界（k=1, 1.5, 2, 3）を設定したときにその枠（閾値）を外れた測定値の数である。これを見ると、あたかもサンプル数が大きいほど z スコアのはずれ値が少ないように見える。しかし、これは、

- ・ごく“少数の”異常に大きい測定値があること
- ・z スコアの算出式から明らかなように、こうした（平均値から遠い位置にある）少数の異常値の影響を受けやすいこと
- ・平均値の分布が大きく歪んでいることから起こる z スコアの判定機能の限界

等から生じる（当然の）現象であり、ここに z スコアの判定機能の限界が見える。これをより明らかに知るために、各データセットの統計値を求めて表 2 とした。同時に、サンプル数が 191 機関と 181 機関についてはグラフィカル表現法を用いてヒストグラム、箱髭図、正規確率プロット図等を併用した（図 1, 2）。これらを観察することで、上記の指摘の特徴が検証される。たとえば、以下のような傾向が読みとれる。

- ・平均値、標準偏差、歪度、尖度のよう、その算出式の構造から、（平均値から遠く離れた少数個の）はずれ値・異常値の影響を受けやすい統計値の変化が大きい。つまり、シグマ限界方式（2 シグマとして）で除外したデータの影響をこれら統計値が受ける。

- ・メジアン、四分位数を見ると、その変化はさほど大きくはない。これはデータの数値の大きさ（順位）に注目した統計値であることから、直接的にははずれ値の影響を受

けにくいことによる。

・ごく少数のはずれ値の影響から、分布の形に（見かけ上の）大きな歪み・尖りが生じる。当然のことであるが、除外データが増えるにしたがってこれが緩和される。

・同時に、分布の両裾が切り捨てられるので、結果として数値が平均の周りに集まり分布形が対称に近づきシグマ限界方式ではずれ値が分布の両側に現れるようになる。

・分布の形が「単峰性でない場合」には、 z スコアを用いることには無理が生じる。この数値例ではこの傾向はさほど顕著ではないが、サンプル数を181としたときに僅かにこの傾向が見られる（正規確率プロット図の観察から、図2を参照）。

以上のことから、 z スコアを用いた判定方式においては「分布の非対称性や際だって外れている数値の影響」を何らかの方法でチェックし、かつ場合に応じてこれを除外する方法を適用するといった段階的操作が必要となることが示唆される。これには、一次スクリーニングの方式の見直しも含まれる。しかし、本質的にはどの程度、良く管理された内部精度管理の整った測定環境でデータが取得されたか（あるいはされているか）を知る別途の情報取得がなければ、ここでの議論はすべて仮定的なこととなる。

b) はずれ値の検討

通常、統計的手法で考える「はずれ値」とは（ある母集団分布から）抽出した（無作為）標本から得られた観測値についての分布から、大きくはずれた値を取る数値の存在を言う（様々な概念・表現があろうが詳細は類書に譲る；文献[5]）。しかし、「外部精度管理調査」の場合にはこうした状況を期待することははかかなり難しいことは既に指摘した。

測定時に、各機関の5個の測定値のはずれ値チェックを行ったうえで残った測定値についての平均値を分析対象とすることも考えられるが、この場合には「はずれ値として除外対象となった測定値」の理由付け・意味付けを検討せねばならないので、機械的な操作でデータをチェックすること

にも無理がある。

「外部精度管理調査」では、測定値のスクリーニング処理（一次処理）として、範囲チェックのフィルタ「（目標とした）添加量の1/10あるいは10倍」を；それぞれ下限閾値および上限閾値として設定している。スルファジミジン・データの場合では、添加量（ここでは、0.28 ($\mu\text{g/g}$)) を目安としたわけだが、この閾値設定の根拠は、測定値の記録単位が（上下に）一桁以上ずれることがあればやはり異常値と考えるべきであろうといった単純な理由である。しかし後の分析で明らかとなるように、この設定だけでは必ずしも適切とは言えない症状が実際の観測値には現れる。

(5) 機関内変動と機関間変動の評価

「外部精度管理調査」では、各機関から5個の測定値の提出を求めている。したがって、形式的には機関内変動と機関間変動の両者を調べられるはずである。これについていくつかの考察を行った。

a) 平均値と標準偏差の関係

まず一つの簡便法は「平均値と標準偏差の関係」に注目することである。（標本）標準偏差の算出において平均値を用いることから、平均値と標準偏差の2つの統計値の間に何らかの関係（ある種の相関）が予想されることである。つまり平均値が大きく動いた場合、それに連動して標準偏差の変動も変化するかということである。 z スコアや管理図法を適用する際は、その統計的な仕組みから、両者の間に関連があることは好ましくない性質である。これの観察をスルファジミジン・データについて試みた。これの例が図3および図4である。図3はサンプル数が191、図4は168の場合である。なお図中の楕円は（2変量正規分布と見なした場合の）90%信頼限界を表す単なる目安に過ぎない（相関の傾向を見るためであって正規分布が関連するものではない）。これらの図の示す特徴から、平均値と標準偏差の関係は総じて関連性が無いように見えるが、同時に以下のようないくつかのグループの混在が予想される。つまり

ここでも、収集データの等質性あるいは代表性についての保証が求められることになる（データをいかに集めたかの実験計画、検査環境のあり方の不透明性に起因する）。

- ・ 平均値も偏り、かつ標準偏差（バラツキ）も大きい場合
（正確でなくかつ精度も悪い）
- ・ 平均値のみが偏る場合（正確でない）
- ・ 標準偏差（バラツキ）だけが大きい場合
（精度が悪い）

ただし、これらを検証するための情報取得を実際にどう行うか、あるいは論拠をどこに求めるかの検討は、今回の手元のデータだけからは難しく、上記の特徴を断片的に指摘するにとどめるしかない。

b) 変動係数による観察

もう一つの指標として変動係数（C.V. = [標準偏差 ÷ 平均値] × 100）がある。これは測定における相対的精度を知る指標としても重要である（異なる測定環境や測定項目間の相互の変動の比較評価を行うとき等）。表 2 の統計値の一覧に示すように、4 組のデータセットの変動係数にはかなりの差異がある。つまりそれだけではずれ値除去の影響があったということである。精度の相対的な大きさの目安としては、数%から 10 数%程度であろうが、「外部精度管理調査」に示された他の測定項目（とくに重金属検査項目、保存料検査項目）と比べるとスルファジミンの変動係数には大きな差異がある（鉛が 14.2%、安息香酸が 10.9%、ソルビン酸が 8.2%程度であった）。表 2 に見るように 181 サンプルで 16.7%弱、168 サンプルの場合でも 13.4%弱となっている。

c) 管理図法の代用の考え方

標本（検査機関）のサンプリングのあり方から、統計理論をそのまま適用することには壁がある。「外部精度管理調査」では、各機関（各群）の 5 個の測定値に対して、（統計的に独立に観測されたとの仮定の下に）管理図法の考え方を代用しているが、

これはむしろ（添加量に対して）「各機関内（郡内）変動」と「各機関間（群間）変動」をモニタリングするチャートと考えるべきである（各機関の平均値としての差異を平均値管理図で、また個々の機関内の変動の推移をレンジ（範囲）あるいは標準偏差の管理図で代用して、それぞれモニタリングする）。本来は、CODEX 等のレポートにあるように、ある特定機関内の特定な検査環境で時間経過に伴って、しかも時系列的に測定されるようなデータの評価、あるいはコラボレーションを行っている良く管理された機関で取得したデータの検証を行うような場合に適用されるべき方式と考えられる。

ここでは参考として、図 5 にサンプル数が 181 の場合を例として、いわゆるシュハート流の 3 シグマ法に基づく「 \bar{x} -R 管理図」を代用したチャートを示した。なお、横軸の並び（機関）の順は、管理図本来の特性とは異なり時系列的な意味はない。単にデータの並び順に従って並べただけである。従って管理図法で利用される点の並びの傾向（トレンド、連の検定等）を評価する方法は適用できない。前述のように平均値と標準偏差の各機関別の比較検証の道具（むしろ平均値と標準偏差との連続的な検定を行っている）と考える方が自然である。なお、この図は管理図そのものではないので、各点の結線は行わずに縦棒で表した。この図から、以下のような特徴が読みとれる。

まず、各機関間の差異を観察する \bar{x} 管理図では多くの機関が管理線外にある。元来、管理図の仕組みは同じロットあるいは母集団から逐次的にまた統計的に独立にサンプリングされたランダム標本を前提としているのであり、こうした状況とは異なる回収データにとっては、個々の機関の平均値が変動することはむしろ当然といえる。一方、各機関内の変動を示す R 管理図では、幾つかの機関が上部管理線を越えるものの大方は管理線内に入っている。これを総合すると、スクリーニングした結果の 181 機関のデータセットについては、大方の機関の精度（バラツキ）は整ってはいるものの、目

標値を当てるといふ「正確さ」の点でかなり変動があることが観察される。これを、回収率が各機関でかなり異なるようだとはい換えることができるかもしれない。また仮に、配布試料についての（予定した）添加濃度の0.28 ($\mu\text{g/g}$) から見ると、総じて各機関の平均値はこれを下回る（回収率が下回る）。しかし作成試料の濃度確認として得られた、20 試料の平均値0.196 ($\mu\text{g/g}$) と比較した場合には、平均値の分布がこれの周りに集まっておりそれほどおかしな結果ではない。ここで、回収率と目標値の関係をどう考えるかの問題が生じるのであるが、これについてはここで取り上げるだけの情報がないので、指摘にとどめ言及は避ける。

E 結論

本年度は、一組のデータセット（合成抗菌剤：スルファジミジン）を例として、これの具体的な観察において見られた諸事象から、そこに内在する問題点を指摘するにとどまった。次年度に向けては、さらに別の検査項目についての測定値の統計的評価や、さらには測定値の取得環境のあり方、検査方法と測定値との関連等について言及したい。なお、精度管理における統計的データ解析のあり方については、ここで述べた主張も含めて参考文献に挙げた記事も参考とされたい。

重要と思われることは、限られた情報の中での検討とならざるを得ないこと、つまり高度な統計解析手法を適用するだけの情報量がない、あるいは測定値を評価する際に考慮すべき要因があまりにも多すぎて、結果としてノイズ要因があまりにも多すぎると思われることがある。

また、ここではスルファジミジン・データを例としたが、ここでの指摘は「外部精度管理調査」で取得したすべてのデータに共通した事項と考えられる。とくに、測定項目によって、そこで得られる測定値の等質性、はずれ値の扱い方等については、各項目の性格に応じた独自の有効な分析方法があり、しかもそれぞれが異なるのではな

いかと思われる。例えば、今回の「外部精度管理調査」における理化学検査項目だけを見ても、重金属、保存料、残留農薬、残留動物用医薬品とあるが、それぞれが特有の特徴を示す事が見えるわけで、これを一括した同一の方式で評価することにやや無理があるものと思われる。

加えて、参加施設のタイプ（保健所、指定検査機関、地方衛生研究所、検疫所、食肉衛生検査所、市場衛生検査所）、各機関における検査施設機器類の性能、検査担当者の技能等々、考慮すべき要因は多々ある。これらの影響を網羅的にチェックするだけの情報量が取得データの中にあるかという根本的な問題への鍵を模索し、簡単なデータ解析を通じて問題を提起したことが本報告の結論である。

【参考文献および発表】

[1] (財)食品薬品安全センター秦野研究所(1999), 「平成10年度食品衛生外部精度管理調査結果報告書」.

[2] 大隅昇(1998), 精度管理における統計的データ解析(1) - 測定値の変数変換とzスコア -, 食品衛生学雑誌, 39巻, 4号, J-352-J332.

[3] 大隅昇(1998), 精度管理における統計的データ解析(2) - zスコアの代替指標, 変動の観察 -, 食品衛生学雑誌, 39巻, 5号, J-384-J389.

[4] 大隅昇(1998), 精度管理における統計的データ解析(3) - グラフィカル表現によるデータの初動探査 -, 食品衛生学雑誌, 39巻, 6号, J-451-J456.

[5] 大隅昇(1999), 精度管理における統計的データ解析(4) - グラフィカル表現によるはずれ値の探査 -, 食品衛生学雑誌, 40巻, 2号(印刷中).

[6] Vic Barnett and Toby Lewis, *Outlier in Statistical Data - Third Edition -* (1994), John-Wiley & Sons.

[7] Stephen B. Vardeman and J. Marcus Jobe (1999), *Statistical Quality Assurance Methods for Engineerings*, John-Wiley & Sons.

表1 シグマ限界による判定

k シグマ限界		197 サンプル	191 サンプル	181 サンプル	168 サンプル
1 シグマ限界	上に	2	10	15	14
	下に	0	12	13	17
1.5 シグマ限界	上に	0	1	8	10
	下に	0	7	5	7
2 シグマ限界	上に	0	3	3	3
	下に	0	4	8	5
3 シグマ限界	上に	6	3	0	0
	下に	0	0	0	0

表2 サンプル数を変えたときの各種統計値

統計値	サンプル数を変えた場合			
	197 サンプル	191 サンプル	181 サンプル	168 サンプル
平均値	0.2268	0.1945	0.1913	0.1944
分散	0.0405	0.0023	0.0010	0.0007
標準偏差	0.2012	0.0475	0.0319	0.0260
歪度	6.2350	1.7153	-0.3779	-0.1067
尖度	13.7959	9.2355	0.2305	-0.1389
メジアン	0.1956	0.1942	0.1942	0.1954
第1四分位数	0.1746	0.1722	0.1746	0.18025
第2四分位数	0.2160	0.2126	0.2101	0.21015
変動係数 (%)	88.7099	24.4007	16.6918	13.3612

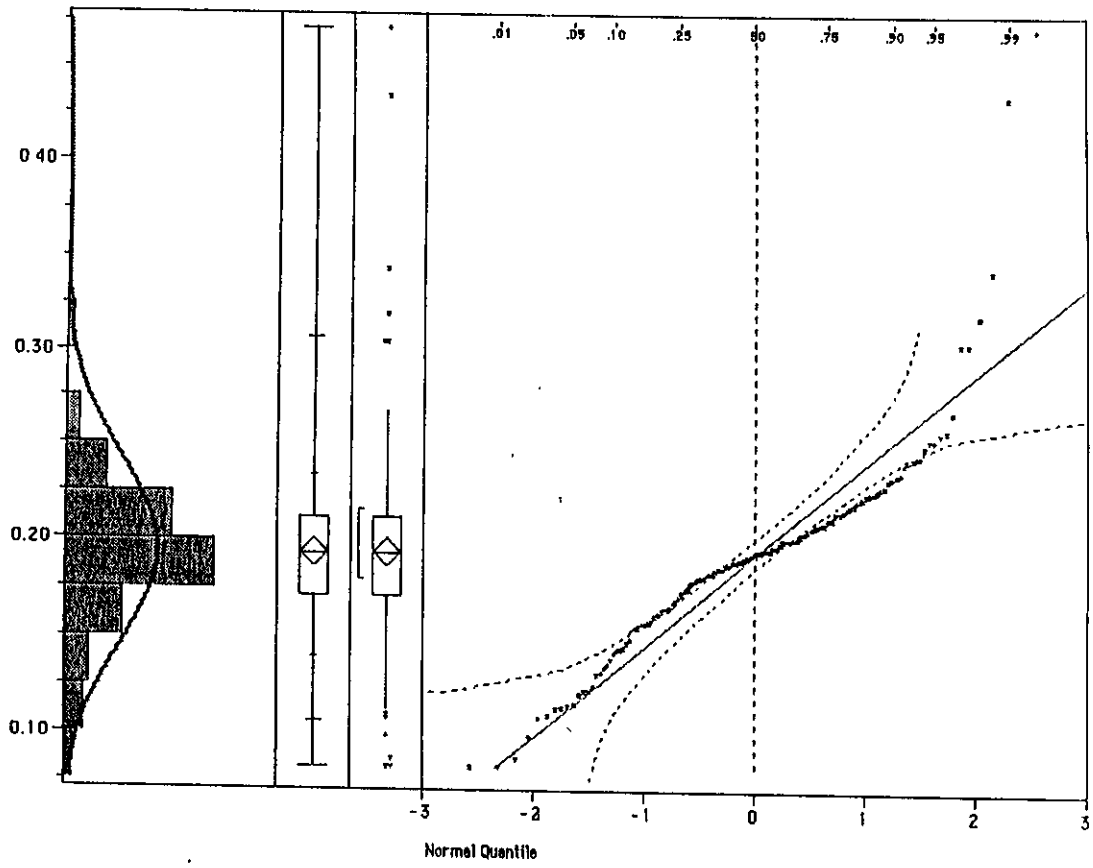


図1 191 機関の平均値の分布

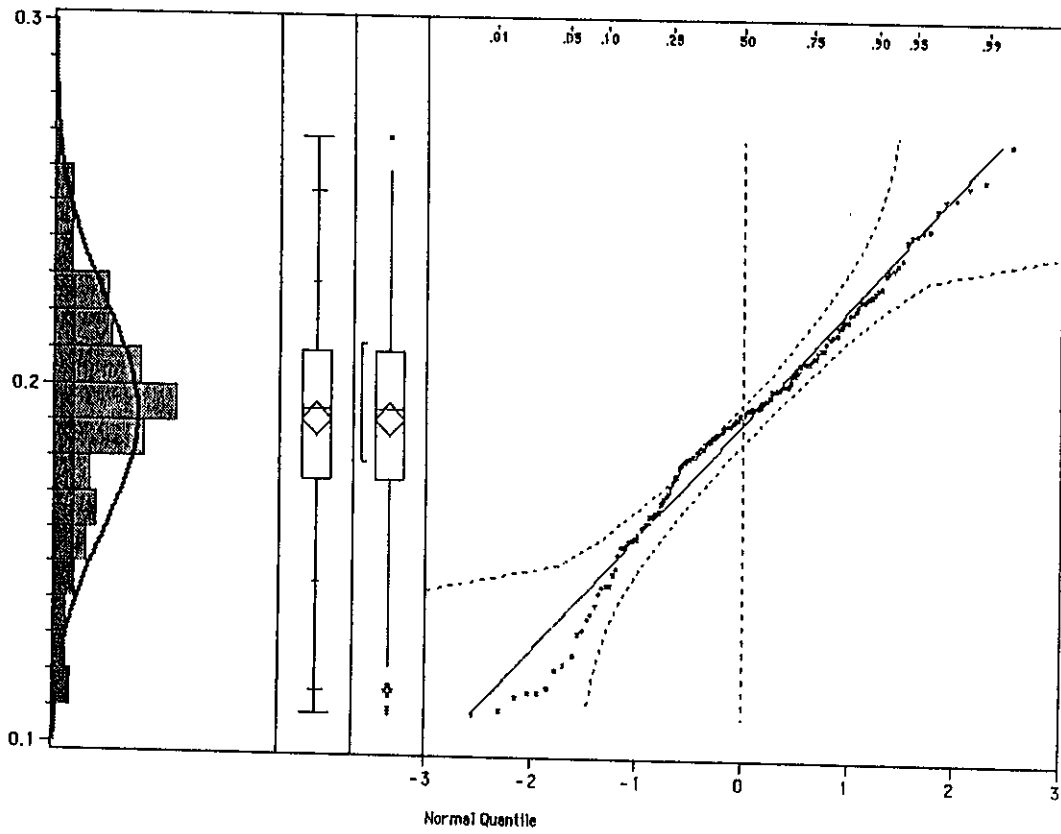


図2 181 機関の平均値の分布

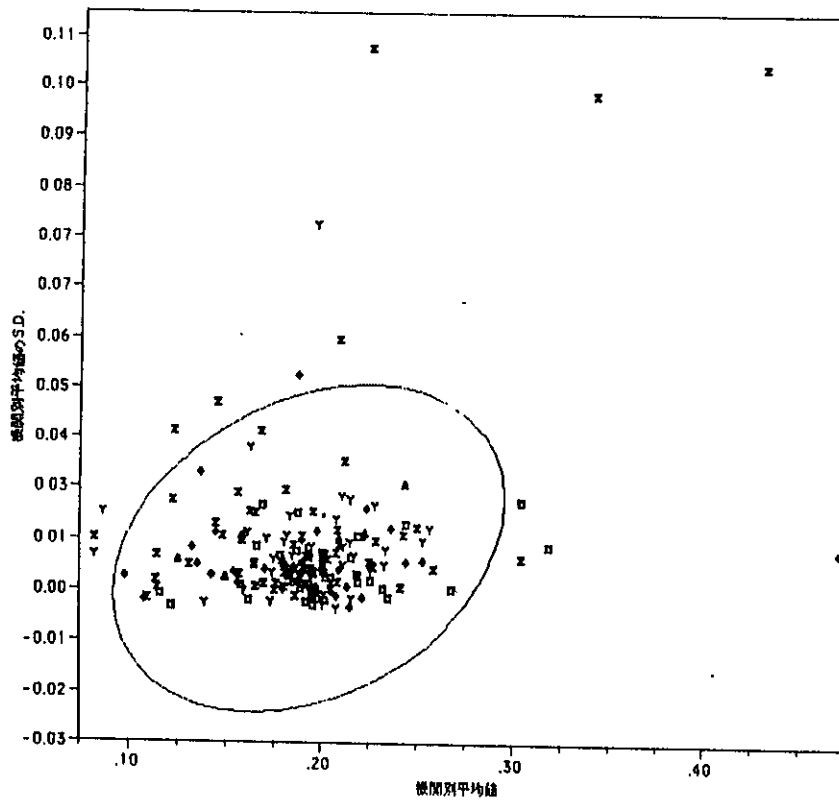


図3 機関別の平均値と標準偏差との散布図 (191 機関のとき)

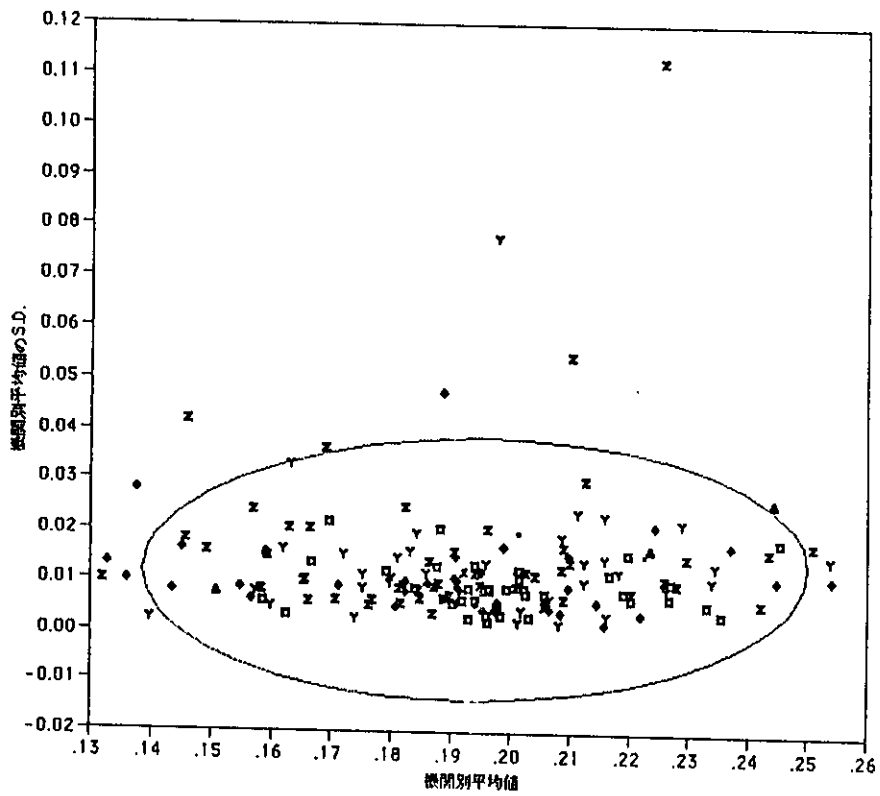
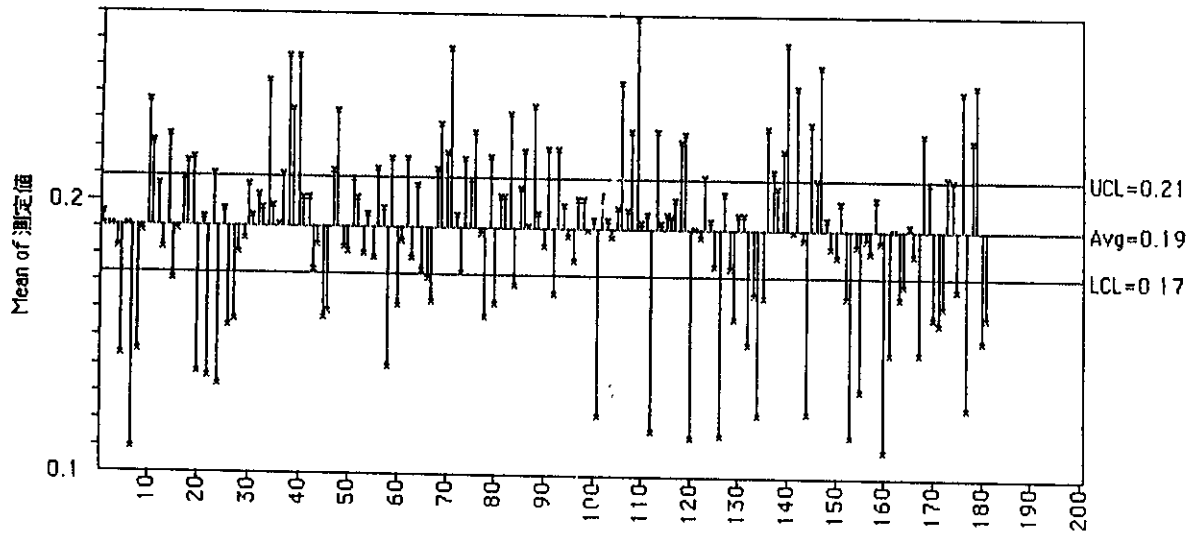
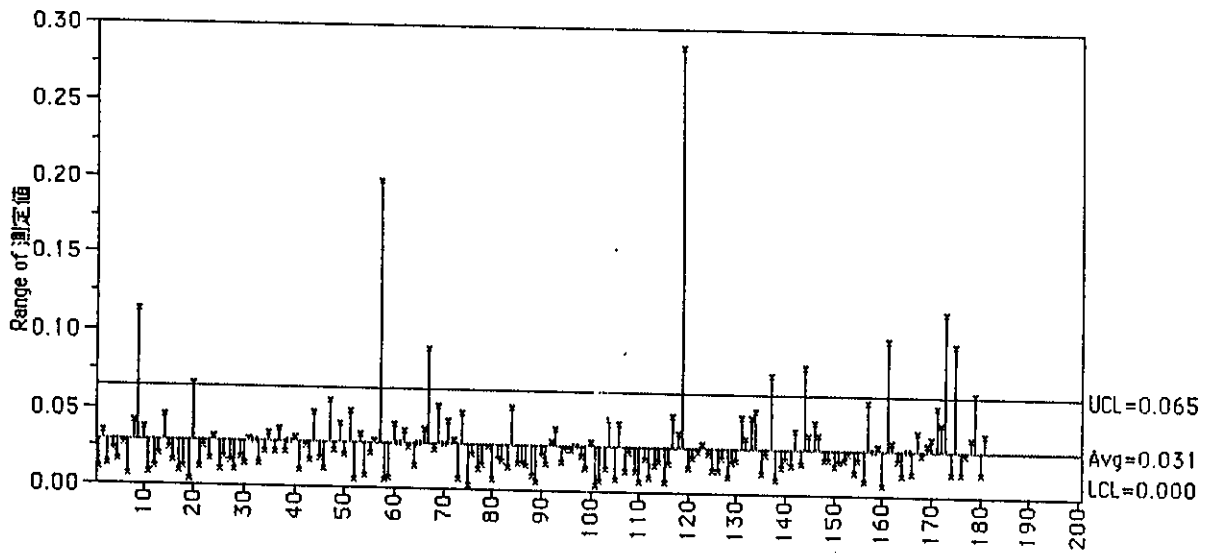


図4 機関別の平均値と標準偏差との散布図 (168 機関のとき)



5-1 \bar{X} 管理図方式による機関別平均値の比較



5-2 R管理図方式による機関別範囲の比較

図5 $\bar{X}-R$ 管理図の代用をによるチェック (181 機関について)

2. 試験法の評価方法に関する研究

分担研究者 松田りえ子 国立医薬品食品衛生研究所
食品部主任研究官

研究要旨

全国の食品衛生検査施設 938 カ所に、調査票を送付し、分析法バリデーション及び内部精度管理の実施状況について質問した。回答数は 587 であり、そのうち実際に検査を実施している施設は 379 カ所であった。これらの施設からの回答を解析した結果、ソルビン酸等の検査頻度の高い検査項目については 70 % 近くの施設において、バリデーションあるいは内部精度管理が行われていた。

A 研究目的

平成 7 年に食品衛生法が改正され、食品衛生に関する各種試験検査機関への業務管理基準（GLP）の導入が義務づけられた。大部分の機関において、GLP の実施に必要な信頼性確保部門の設置、及び各種標準作業手順書（SOP）の作成は完了しているが、分析値の信頼性に大きく関与する、分析法のバリデーション、内部精度管理については未だ確立しているとは言い難い状況である。

そこで、本研究では検査頻度の高い項目について、分析法バリデーション及び内部精度管理の実施状況及び、実際に行われている検査の回収率・再現性を把握することを目的とした調査を行った。

B 研究方法

全国の食品衛生検査施設に、調査票を送付し回答を依頼した。回答の期限は 1 月とした。

調査を実施した検査項目は以下の通りである。

検査実施頻度の高いもの

安息香酸

ソルビン酸

パラオキシ安息香酸エステル
類

BHA

鉛

カドミウム

二酸化イオウ

プロピオン酸

プロピレングリコール

チアベンダゾール

ポリソルベート

マラチオン

ジクロロボス

ディルドリン

オキシリン酸

スルファジミジン

違反頻度の高いもの

アフラトキシン

T B H Q
ビテルタノール
クロルピリホス
ピリミホスメチル
シペルメトリン
ダミノジット
パラチオンメチル
E P N
臭素
シアン化合物

各検査項目ごとの質問事項は以下の通りである

- バリデーションを実施しているか
- 実施している場合
 - 回収率について
 - 試料作物、実施濃度、回収率 (%)
 - 再現性について
 - 試料作物、実施濃度、RSD (%)
 - 定量下限濃度
 - 検出下限濃度
- 内部精度管理を実施しているか
- 実施している場合
 - 回収率について
 - 試料作物、実施濃度
 - 回収率 (%), 標準偏差
 - くり返し再現性について
 - 試料作物、実施濃度、RSD (%)

C 研究結果

1) 回収状況

全国 938 カ所の施設に調査用紙を送付し、587 カ所から結果が返送された。回収率は地方衛生研究所は 80 %、保健所は 59 %、食肉検査所は 67 %、指定検査機関は 70 %、その他（検疫所、食

品検査所等）は 59 % であり、全体では 63 % であった。

また、回答を送付した保健所のうち、207 カ所は、検査業務を行っておらず今回の調査対象とはならない機関であったため、これを除いた計 379 機関について、以下の解析を行った。対象機関の内訳は、地方衛生研究所 64 カ所、保健所 180 カ所、食肉検査所 75 カ所、指定検査機関 44 カ所、その他 16 カ所であった。

2) 検査の実施頻度

各検査項目の検査実施頻度を Table 1 に示した。また、対象機関分類中の実施割合も Table 1 中に示した。

食品添加物の検査中では、ソルビン酸、安息香酸、二酸化イオウ、パラオキシ安息香酸エステル類の実施割合が高く、食肉検査所を除く衛生研究所・保健所・指定検査機関・その他の施設では 50 % 以上の実施率であった。

農薬の検査実施率は全体としては 40 % 以下であるが、衛生研究所では 50 % 以上の実施率であった。ダミノジッドの実施率は衛生研究所でも 13 % と低かった。これは、他の農薬が同時に分析可能であるが、ダミノジッドは個別に分析が行われるためと考えられる。

金属の検査は、衛生研究所・指定検査機関では 50 % 以上が行っているが、この検査を行っている保健所は 30 % 以下であった。

残留動物薬のオキシリン酸・スルファジミジンの検査は衛生研究所・食肉検査所での実施率が高く、保健所での実施は非常に少なかった。

3) バリデーション・内部精度管理の実施状況

Table 2 に各検査項目における、分析法バリデーション及び内部精度管理の実施状況を示す。表に示した数字は、その検査を実際に行っている施設における、バリデーションと内部精度管理の実施率である。

全体を通して最も実施率の高かった検査項目はソルビン酸で、全体として70%の機関がバリデーションあるいは内部精度管理を行っていた。次いで、安息香酸、ディルドリン、クロルピリホス、スルファジミジンの実施率が高く、60%以上であった。ソルビン酸・安息香酸は検査を実施している施設数も多く、検査頻度が高いことため、GLP導入の最初、にバリデーションが行われたと考えられる。ディルドリン・クロルピリホスは保健所ではほとんど検査されていないため、全体としての検査率は低いが、衛生研究所の80%が検査を実施している。60%の衛生研究所でバリデーション・内部精度管理が行われている。スルファジミジンの検査は、衛生研究所、食肉検査所で多く行われており、半数以上がバリデーション・内部精度管理を実施している。

逆に、バリデーション・内部精度管理が進んでいない項目は、臭素(26%)・シアン化物(20%)であった。これらは検査件数が少ないために、まだバリデーションが行われていないと考えられる。

機関別では、バリデーションの実施率に大きな差はなく、半数以上の機関でバ

リデーション・内部精度管理の実施が開始されていた。

バリデーションと内部精度管理の実施状況を比較すると、内部精度管理を行っている機関の方が多い傾向が見られた。特に保健所では、この傾向が顕著で、ソルビン酸では、バリデーションを行っている保健所が29%であるのに対し、内部精度管理を行っているのは51%に達している。厚生省から精度管理のガイドラインが出され、実施方法が明らかであるので、内部精度管理の導入が先行したと考えられる。

調査対象外の検査項目についてバリデーション又は精度管理を実施していた機関は95カ所であった。検査項目数は171と多岐にわたっていた。亜硝酸根、サッカリンナトリウム、オキシテトラサイクリンの検査については20カ所以上の機関で、バリデーション・内部精度管理が行われていた。

D 考 察

検査頻度の高い検査項目においては、60%以上の機関でバリデーション・内部精度管理が実施されていた。特に、バリデーションよりも内部精度管理が先行していることが特徴的である。平成7年に行った指定検査機関を対象とした調査では、内部精度管理を行っている機関はバリデーションを行っているところよりも少なかった。これは、精度管理ガイドラインで実施方法を具体的に示したため、導入が促進されたと考えられるので、バリデーションについてもガイドライン等で実施方法を明らかにすれば、さ