

《暫定翻訳》

この翻訳は正式なものではありません。  
取扱はご注意ください。

Pure Appl. Chem., Vol. 74, No. 5, pp. 835-855, 2002.

© 2002 IUPAC

国際純正応用化学連合

分析・応用・臨床・無機及び物理化学部門

分析試験施設のための品質保証計画のハーモナイゼーションに関する部門間作業委員会\*

**HARMONIZED GUIDELINES FOR SINGLE LABORATORY VALIDATION OF  
METHODS OF ANALYSIS  
(IUPAC Technical Report)**

分析法の単一試験施設バリデーションに関する  
ハーモナイズされたガイドライン  
(IUPAC 技術報告書)

IUPAC、ISO、および AOAC インターナショナルの後援のもと、1999年11月4～5日、ハンガリー、ブダペストで開催された、分析試験施設のための品質保証システムのハーモナイゼーションに関するシンポジウムでの結論

作成及び出版

MICHAEL THOMPSON<sup>1</sup>, STEVEN L.R. ELLISON<sup>2</sup>, AND ROGER WOOD<sup>3‡</sup>

<sup>1</sup> Department of Chemistry, Birkbeck College (ロンドン大学), London WC1H 0PP, UK;

<sup>2</sup> Laboratory of the Government Chemist, Queens Road, Teddington, Middlesex TW11 0LY, UK;

<sup>3</sup> Food Standards Agency, c/o Institute of Food Research, Norwich Research Park, Colney, Norwich, NR4 7UA, UK

\* 1997～2000年の調査委員会会員は、以下の通り：

**座長：** A.Fajgelj、1997～、(IAEA、オーストリア)；**会員：** K.Bergknut (ノルウェー)；Carmen Camara (スペイン)；K.Camman (ドイツ)；Jyette Molin Christensen (デンマーク)；S.Coates (AOAC Int., USA)；P. De Bievre (ベルギー)；S.L.R. Ellison (UK)；T.Gills (USA)；J. Hlavay (ハンガリー)；D.G. Holcombe (UK)；P.T. Holland (ニュージーランド)；W. Horwitz (USA)；A. Kallner (スウェーデン)；H. Klich (ドイツ)；E.A. Maier (ベルギー)；C. Nieto De Castro (ポルトガル)；M.Parkany (スイス)；J. Pauwels (ベルギー)；M. Thompson (UK)；M.J. Vernengo (アルゼンチン)；R. Wood (UK)。

‡ 関連著者

---

本報告書の再刊や複製、その保管及び／または電子媒体による宣伝は、すべての参考文献、IUPAC 著作権シンボル©および発行年が記載された謝辞が明確に記されている場合のみ、正式な IUPAC の許可なしに認められる。他言語への翻訳出版では、上記条件に加え、関連 IUPAC National Adhering Organization から、あらかじめ承認を受けることが必要である。

# 分析法の単一試験施設バリデーションに関するハーモナイズされたガイドライン

## (IUPAC 技術報告書)

**要約：** 方法バリデーションは、分析化学における複雑な品質保証システムの必須部分として一般的に認識されている手段の一つである。以前、ISO、IUPAC 及び AOAC インターナショナルは協力して、「方法性能試験の設計、実施及び解釈」[1]、「(化学) 分析試験施設の技能試験」[2]、「分析化学試験施設における内部品質管理」[3]、及び「分析測定における回収率の情報の使用」[4]に関する合意プロトコルまたはガイドラインを作成してきた。これらのプロトコル／ガイドラインを作成した作業委員会は、分析法の単一試験施設バリデーションに関するガイドラインを作成するように IUPAC から依頼を受けた。これらのガイドラインでは、適切な分析法のバリデーションを保証するために採用すべき最低限の推薦事項を提示する。

UK Royal Society (英国学士院) of Chemistry が公表した予稿であるガイドライン草稿は、化学試験施設における品質保証システムのハーモナイゼーションに関する国際シンポジウムで審議されている。

## 目次

1. 緒言
    - 1.1 背景
    - 1.2 既存プロトコル、基準及びガイド
  2. 定義と用語
    - 2.1 一般的用語
    - 2.2 本ガイドのみで使用した定義
  3. 方法バリデーション、不確かさ及び品質保証
  4. 方法バリデーションの基本原理
    - 4.1 規格及びバリデーションの範囲
    - 4.2 仮定の試験
    - 4.3 分析誤差の原因
    - 4.4 方法及び試験施設の影響
  5. バリデーション試験の実施
  6. バリデーション試験の範囲
  7. 勧告
- 添付資料 A：試験の方法性能特性の要求事項に関する注意事項
- A1. 適応範囲
  - A2. 選択性
  - A3. キャリブレーション及び直線性
    - A3.1 直線性と切片
    - A3.2 一般的なマトリックス効果に関する試験
    - A3.3 最終キャリブレーション手順

- A4. 真度
  - A4.1 真度の推定
  - A4.2 真度実験条件
  - A4.3 真度試験に関する基準値
    - A4.3.1 認証標準物質 (CRM)
    - A4.3.2 標準物質
    - A4.3.3 参照方法の使用
    - A4.3.4 添加／回収率の使用
- A5. 精度
- A6. 回収率
- A7. 範囲
- A8. 検出限界
- A9. 測定限界または定量限界
- A10. 感度
- A11. 堅牢性
- A12. 目的適合性
- A13. マトリックス変動
- A14. 測定不確かさ

添付資料 B. 不確かさに関する追加考慮点  
バリデーション試験における推定

- B1. 感度分析
- B2. 判定

参考文献

## 1. 緒言

### 1.1 背景

分析のあらゆる分野において、国や国際的な規則を遵守するためには、信頼できる分析法が必要である。結果的に、試験施設が、求められる品質のデータを提供する能力を持つこと、及び求められる品質のデータを提供することを保証するために適切な手段を講じなければならないことが、国際的に認識された。そのような手段は、以下のものが含まれる。

- ・ 妥当性の確認された分析法を用いること；
- ・ 内部品質管理手順を用いること；
- ・ 技能試験計画に参加すること；及び
- ・ 国際規格、通常 ISO/IEC 17025 に適合するようにすること。

特に ISO/IEC 17025 に準ずるには、一覧に挙げた全てを含む、他の技術及び管理要求事項の様々な範囲が要求されると同時に、測定のトレーサビリティを確立することに焦点が当てられていることに留意しなければならない。

従って、方法バリデーションは、信頼できる分析データを得るために試験施設が実施しなければならない手段の不可欠な要素である。上記の他の局面については、分析試験施設のための品質保証計画のハーモナイゼーションに関するIUPAC部門間作業委員会が、特に、

方法性能（共同）試験[1]、技能試験[2]、及び内部品質管理[3]に関するプロトコル／ガイドラインを作成することで、以前、扱っている。

部門によっては、食品分析に代表されるように、「完全に妥当性の確認された」方法への要求事項が法律に記載される [5, 6]。分析法の「完全な」バリデーションには、通常、試験施設内方法性能試験（共同試験（collaborative study または collaborative trial）として知られている）における方法の性能特性試験も含まれることになっている。国際的に認証されたプロトコルは、国際的にハーモナイズされたプロトコル[1]及びISO手順[7]に代表される、共同試験による分析法の「完全な」バリデーションのために確立された確立されたこれらのプロトコル／基準では、分析法の妥当性を完全に確認する共同試験において、試験施設数及び試験物質の数を最小とするように要求している。しかし、これは分析法の完全なバリデーションをするために必ずしも実用的、あるいは必要であるとは限らない。このような場合、「単一試験施設方法バリデーション」が適切なこともある。

単一試験施設方法バリデーションは、以下の項目を含む一部の状況において適切である：

- ・費用のかかる正式な共同試験を実施する前に、方法の実行可能性を保証する場合。
- ・共同試験でのデータが利用できない場合、または、正式な共同試験が実施できない場合に、分析法が信頼できる証拠を提示する場合。及び
- ・「直ぐに使える」妥当性の確認された方法が、正しく用いられていることを保証する場合。

社内で方法の特徴付けを行う場合は、どの特性を評価すべきかを試験施設が定め、顧客と正確に合意することが重要である。しかし多くの場合、これらの特性は、法律で明言されることもある（例えば、食物中の動物用医薬品の残留物や食品産業における農薬）。試験施設が引き受ける評価の範囲は、法律の要求事項と見合わなければならない。

それにもかかわらず、分析領域によっては、同一の分析法を多数の試験施設で用いて、定義されたマトリックス中で安定な化学物質を決定する。これらの試験施設が利用できる適切な共同試験方法が作成できた場合、その方法の妥当性を確認するための共同試験の費用が理にかなっていると考えられるのは当然だということの評価すべきである。共同試験の方法を用いることで、ある方法を定期的にも用いる前に、ある試験施設が投入しなければならない甚大なバリデーション作業をかなり削減できる。本来の目的に適合していることが明らかとなった共同試験の方法を用いる試験施設では、方法に記載された性能特性を達成できることを実証さえすればよい。方法が正しく使用されていることのそのような確認は、完全な単一試験施設バリデーションよりもかなり費用が安い。分析団体にかかる総費用は、特定の方法について共同試験で妥当性を確認し、続いて、その方法を使用したい試験施設がその性能属性の妥当性を確認する方が、多数の試験施設が別々に同一方法の単一試験施設バリデーションを実施するよりも安くなるのがしばしばある。

## 1.2 既存プロトコル、規格及びガイド

方法バリデーション及び不確かさに関する数多くのプロトコル及びガイドライン[8~19]が、特に、AOACインターナショナル、ハーモナイゼーションに関する国際会議（ICH）及び Eurachem 文書において作製されてきた：

- ・共同試験に先がけたの単一試験施設試験に関する指針[13]を含むAOACの統計学マニュアル；

- ・医薬品承認の提出を裏付けるために用いた試験に関する最小限のバリデーション試験要求事項を規定したICHテキスト[15]と方法論[16]；
- ・分析法の目的適合性：方法バリデーション及び関連トピックに関する試験施設ガイド（1998年）[12]；および
- ・分析手段における不確かさの定量化（2000年）[9]。

方法バリデーションは、1997年12月に開催された食糧管理のための分析法のバリデーションに関するFAO/IAEA合同会議委員会でも広く審議され、その報告書は利用可能である[19]。

現行のガイドラインでは、国際的な承認を受けた情報が提供され、また何より、単一試験施設の方法バリデーションにおける最善の実施のための方向性を示すことを目的として、上記文書の重要な科学原則が結集されている。

## 2. 定義と用語

### 2.1 一般的用語

本文書で用いられた用語は、利用できる場合は、ISOおよびIUPACの定義を尊重する。以下の文書には、関連定義が記載されている：

- ・IUPAC：化学用語概論 第二版（「Gold Book」）A. D. McNaught およびA. Wilkinson, 1997年
- ・ISO：度量衡学での国際的な基本語彙及び一般用語、1993年

### 2.2 本ガイドのみで使用した定義

相対不確かさ：相対標準偏差として示された不確かさ。

妥当性の確認された範囲：バリデーションを行った分析法の濃度範囲の部分。

## 3. 方法バリデーション、不確かさ及び品質保証

方法バリデーションは、分析法が基づくあらゆる仮定を試験し、方法の性能特性を確立し、文書化するひとまとまりの試験を利用する。これにより、その方法が特定の分析目的に適合するかが実証される。分析法の典型的な性能特性とは、適応性、選択性、キャリブレーション、真度、精度、回収率、操作範囲、定量限界、検出限界、感度及び堅牢性である。これらに、測定不確かさ及び目的適合性を加えることもある。

厳密に言えば、バリデーションでは「分析法」よりもむしろ「分析システム」を言及しているはずであり、その分析システムには、定義された方法プロトコル、定義された分析物の濃度範囲及び試験物質の特定の種類が含まれる。本文書の目的により「方法バリデーション」を参照することは、概して分析システムを参照することと捉えられるであろう。そういうものとして分析手順を取り扱っているものは、「プロトコル」と呼ばれるであろう。

本文書において、方法バリデーションは、内部品質管理（IQC）または技能試験のような継続活動とは別のもので見なす。方法バリデーションは一回、あるいは方法の作業時

間中、比較的まれな間隔で実施される。それにより、どのような性能を、その方法が将来提供すると期待できるかについて知ることができる。内部品質管理では、その方法が長期に渡ってどのように機能するかを知ることができる。従ってIQCは、IUPACハーモナイゼーションプログラム [3] において別の活動とみなされる。

方法バリデーションにおいて、結果の精度と関連する、興味の対象となる量的特性が得られると考えられる。従って、方法バリデーションは、測定の不確かさを推定する（見積もる）作業に等しいと言ってもそれは、概して真実である。長年、バリデーション目的に関して、上記に一覧で挙げた個々の項目、また、そのパターンを反映するかなりの範囲のこれらのガイドラインを参照することで、方法性能の別の局面を代表することが慣習となっている。しかし、目的への適合性及び結果の信頼性の両者の鍵となる指示薬として、測定不確かさへの依存度が増すと共に、分析化学者が、不確かさの推定を裏付ける測定バリデーションを実施する頻度が増え、実施者によっては、それを直ちに実施したいと考えるかもしれない。それに応じて、測定不確かさについて、添付資料Aでは分析法の性能特性として簡単に取り扱った。一方、添付資料Bでは、他では取り扱われない一部の手順に関する指針を追加的に記載した。

## 4. 方法バリデーションの基本原則

### 4.1 規格及びバリデーションの範囲

バリデーションは、特定の種類の試験物質の濃度範囲を決定するために、特定の分析物及び特定の目的に対して使われた、定義されたプロトコルに適用する。一般に、バリデーションでは、方法が、分析物の濃度範囲に渡り、また、分析物が添加される試験物質に対して、適切に実施されることを確認しなければならない。従って、あらゆる目的への適合性判定基準の記載と共に、どんなバリデーションを実施する前にも、これらの特徴を完全に明記しなければならない、ということになる。

### 4.2 仮定の試験

バリデーション試験は、目的への適合性を示し、バリデーションデータの実用的使用を優先させる結果となる性能係数を提供することに加えて、分析法が基づくあらゆる推定を客観的に試験する役割を担う。例えば、もし結果を単純な直線キャリブレーション関数で算出した場合、自動的に分析は有意な偏りが無く、応答は分析物濃度に比例しており、且つ、偶然誤差の分散が対象範囲に渡って一定であることが仮定される。たいていの場合、そのような仮定は、方法の開発中、またはより長期に渡って累積された経験に基づいてなされ、結果的にかかなり信頼できるものである。それにもかかわらず、的確な測定学は、試験済みの仮説に依存している。これが、数多くのバリデーション試験が統計学的仮説試験に基づく理由である； その目的は、方法の原則についてなされた理にかなった仮説に重大な誤りがないことを基本的に確認することである。

この明らかに難解な注釈には、重要で実質的な示唆がある。特定の仮定が正しいことを「証明」するよりも、信頼性のある仮説からのひどい逸脱がないかを確認する方が容易である。従って、分析物及びマトリックスの範囲に渡る特定の分析技術（例えばガスクロマトグラフィー分析、あるいは酸分解法など）が成功して長期的に使用されている場合、バ

リデーション確認は当然、比較的軽い予備試験の形式を取る。反対に、経験が浅い場合、バリデーション試験では、試験での特定の場合成にされた仮説は適切であるという、確かな証拠を提供する必要がある、一般に十分な範囲で状況の詳細を試験する必要があると考えられる。従って、ある所定の例で要求されるバリデーション試験の範囲は、ある意味、使用した分析技術の累積された経験に依存することになる。

以下の議論において、試験施設は対象の技術を十分に練習すること、また、あらゆる有意検定の目的は、特定のプロトコルが依存する仮説を軽視すべき理由を力説する証拠が全くないことを確認することであるのは当然であると考えられる。読者は、あまり知られていない、あるいはあまり確立されていない測定技術には、より厳格な確認が必要であることを心に留めていなければならない。

### 4.3 分析誤差の原因

分析測定における誤差は、様々な原因\*により、また、組織の様々なレベルで発生する。これらの原因（分析物の特定の濃度について）を示す有益な方法の一つは、以下の通りである[24]\*\*である：

---

\* 大量の目的物から試験施設試料を調製することによる、不確かさの厳密な意味でのサンプリング不確かさは、本文書では考慮しない。試験施設試料からの分析試料の採取と関連する不確かさは、測定不確かさにおいて切っても切れない部分であり、自動的に以下の分析の様々なレベルで含まれる。

\*\* 多くの代替グループ分け、または「誤差の分配」は、より詳細に、または状況の異なる範囲に渡って誤差の特定原因を研究する際に可能であり、有益なこともある。例えば、ISO5725の統計モデルでは、一般に試験施設の影響とラン（操作）の影響とを組み合わせる。一方、ISO GUMにおける不確かさの推定手順では、一つ一つの影響、及び測定可能な結果への影響を評価するのにうまく適している。

- 
- ・ 測定の偶然誤差（併行精度）
  - ・ ランの影響（一つのランに関する偏り、及び数回を超えるランの偶然差異と見なされる）
  - ・ 試験施設の影響（単一試験施設による偏りとして見なされる）
  - ・ 方法の偏り
  - ・ マトリックス変動の影響

これらの異なる原因が、必ずしも独立していないかも知れないが、この一覧により、所定のバリデーション試験が取り扱う誤差の原因の程度を確認する有益な方法が提示される。

併行精度（ラン間）の用語には、良く知られた重量及び容量誤差、試験物質の不均一性、及び分析の化学的処理段階での変動を含む、ラン内で変化する手順のあらゆる部分からの寄与が含まれ、繰り返し分析の分散において容易に認められる。ランの影響は、分析者、試薬のバッチ、機器の再キャリブレーション及び試験施設環境（例えば気温の変化）の変更といった、分析システムにおける追加的な日差変動の理由を説明付ける。単一試験施設バリデーションでは、典型的に、多くの個々のランで適切な物質を繰り返し分析するように設計された実験を実施することによりランの影響を推定する。試験施設間変動は、キャリブレーション標準における変動、プロトコルの現地毎での解釈の違い、機器または試薬入手元の変更、または平均気象条件の違いといった環境的要因から発生する。明らかに試

験施設間変動は、共同試験（方法性能試験）及び技能試験の結果において現実のものであると見なされ、方法間変動は、後者の結果において識別できることも度々ある。

一般に、併行精度、ランの影響、及び試験施設の影響は比較可能な大きさであるため、バリデーションにおいて無視できない。過去、特に不確かさの情報を推定して報告する際に、それらの局面が無視される傾向にあった。このため、不確かさの間隔が小さすぎる結果となった。例えば、通常通りに実施される共同試験では、方法の偏り及びマトリックス変動が起因する不確かさの寄与が推定されず、共同試験では別々に扱われなければならない（通常は単一試験施設試験が優先する）ため、完全な事実が得られない。単一試験施設バリデーションにおいて、試験施設の偏りも見過ごされるという特定の危険性があり、そして通常、この試験施設の偏りが、上記一覧の中で、不確かさへ最も大きく寄与する要因である。従って、単一試験施設バリデーションでは、試験施設の偏りに特に注意を払わなければならない。

上述の問題に加えて、方法のバリデーションはその適用範囲に制限される。つまり、方法は試験物質の特有の階級へ適用される。もし定義された階級内でマトリックスの種類がかなり多くある場合は、階級内のマトリックス効果に起因する変動が、追加的な原因になると考えられる。もちろん、その物質に関して、定義された階級外（すなわちバリデーションの範囲外）でも引き続きその方法を使用した場合は、分析システムが妥当であると見なすことはできない。計り知れない大きさの余分な誤差が、測定過程に導入される。

また、分析者は、方法性能が分析物の濃度の関数として変化する方法を考慮することも重要である。ほとんどの例において、結果の分散は濃度と共に確実に上昇し、回収率は高濃度及び低濃度で実質上異なることもある。従って、結果と関連する測定不確かさが、これらの影響と他の濃度依存性要因との両者に依存することも度々ある。幸運にも、性能と分析物濃度との間で単純な相関を推定することが合理的であることもよくある。たいてい誤差は、分析物濃度に比例する\*。しかし、大幅に異なる濃度における方法の性能が興味の対象である場合、性能と分析物濃度との間で推定される関係を確認することが重要である。これは典型的に、予想される範囲の両極端、あるいは2～3点の選択した濃度で性能を確認することにより実施される。直線性の確認でも、同様の情報を得ることができる。

---

\* 検出限界の10倍未満での濃度では適用できないこともある。

#### 4.4 方法及び試験施設の影響

単一試験施設方法バリデーションにおいて、方法の偏りと試験施設の影響を考慮することが非常に重要である。これらの偏りがほとんど無いと見なすことのできる特別な施設を持つ一部の試験施設があるが、この状況は完全に例外である。（しかし、一つの試験施設が特有の分析のみを実施している場合は、方法の偏り及び試験施設の偏りは異なる観点を呈する。）通常、方法及び試験施設の影響は不確かさの量に含まれなければならないが、場合によっては併行精度誤差及びランの影響よりも取り扱いが難しいこともある。一般に、個々の不確かさを評価するために、試験施設が独立して集めた情報を用いることが

必要である。そのような情報の最も一般で有用な情報源は、(i) 共同試験から得た統計値（単一試験施設方法バリデーションの多くの状況には利用できない）、(ii) 技能試験から得た統計値、及び (iii) 認証標準物質の分析結果である。

共同試験では直接、試験施設間の偏りの変動を推定する。この様な試験設計には理論的な欠点があるかもしれないが、これらの変動の推定は、実質的な多くの目的に対しては適切である。結果的に、不確かさの推定値と共同試験で得られた室間再現精度の推定値を比較することにより単一試験施設バリデーションを試験することが、いつも有益である。単一試験施設の結果がかなり小さい場合は、不確かさの重要な原因が無視された可能性がある。（そうではなく、実際には特有の試験施設が共同試験で認められたものよりも小さい不確かさで作業した可能性がある；そのような試験施設は、そのような要求を正当化させる特別な手段を取らなければならないと考えられる。）特定の方法／試験物質の組み合わせで共同試験が実施されていない場合、約120 ppbを超える分析物濃度  $c$  での室間再現標準偏差  $\sigma_H$  の推定値は、通常、両者の変数を質量分率として、Horwitz function； $\sigma_H = 0.02c^{0.8495}$  を用いて得ることができる。（Horwitz の推定値は、通常、共同試験結果で観察された約2のファクター内である。）Horwitz function は、約120 ppb 未満の分率では不正確で、修正関数がより適切であることが認められている。[21と25]。この情報のすべてが、単一試験施設区域に伝達され、最小限の変更がなされることもある。

技能試験から得た統計データから、一般に試験施設の大きさ、及び組み合わせた方法の偏りに関する情報が得られ、参加者は特定の場合におけるより全体的な誤差に関する情報が得られるため、この技術試験で得られた統計データは特に興味深い。原則的に、共同試験、言い換えれば、単一試験施設バリデーションにより得られた個々の推定値と比較するために、全体的な不確かさに関するベンチマーク（基準点）を得ることから得た室間再現精度と同様の方法で、参加者は、1回の試験で得た分析物に関する結果の堅牢性標準偏差といった統計データを使用できる。実際には、技能試験で得られた統計データを共同試験のように体系的に一覧表にしたり、公表したりしないため、これらのデータにはよりアクセスしづらいかもしいが、参加者のみが利用できるようになっている。もちろん、そのような統計データを使用する場合には、適切なマトリックス及び分析物の濃度について言及しなければならない。技能試験への個々の参加者は、参加者が報告した結果と成功した回に割り当てられた値とを比較することより、推定された自分たちの不確かさの妥当性を評価することもできる[26]。しかし、これは継続する活動であり、従って、厳密には（一回限りの出来事である）単一試験施設バリデーションの範囲内ではない。

適切な認証標準物質が利用できる場合、単一試験施設試験により、試験施設が認証標準物質（CRM）を何回も分析して試験施設の偏り及び方法の偏りを合わせて評価することが可能となる。組み合わせた偏りの推定値は、平均の結果と認証値との間の相違である。

適切なCRMがいつも利用できるわけではないため、他の物質も必然的に使用しなければならない。技能試験で残った物質は、時折、この目的に役立ち、物質に割り当てられた値は、不確かさに関しては疑問があるかもしれないが、それらを使用することにより、全体的な偏りに関する確認ができる。具体的に言えば、技能試験で割り当てられた値は、一般に最小限の偏りで推定できるように選択されるため、そのような物質に対する有意な偏りについて試験することは賢明な措置である。更に別の代替方法は、これらの技術と関連する不確かさの測定できない原因が存在する可能性があるが、添加及び回収率の情報[4]を用

いてこれらの偏りの推定値を得ることである。

現在、バリデーションにおいて最も小さいと認識されている影響は、試験物質の定義された階級内でのマトリックス変動に起因する影響である。この不確かさの要因の推定に関する理論的要求事項は、1回のランで分析することになっている試験物質の代表的な採取に対するものであり、それらの個々の偏りを推定し、これらの偏りの変動を算出する。(1回のランでの分析とは、より高い濃度での偏りが変動に影響しないことを意味する。包括される範囲がかなり広い場合は、濃度による偏りの変化に対する処置をしなければならない。) 代表的な物質がCRMである場合、偏りは、結果と基準値との相違として直接、推定することができ、全体の手順は簡単である。利用できるCRMの数が不十分であるという、よくありそうな場合においては、十分に注意して、典型的な試験物質のある範囲で実施する回収試験が頼られることもある。現在、いくつかの例において、この原因から生じた不確かさの大きさに関する定量的な情報は、大きいと疑われているものの、ほとんど無い。

## 5. バリデーション試験の実施

バリデーション試験の詳細設計及び計画の実行は、他で大部分を述べたので、ここでは繰り返さない。しかし、主要原理は該当し、以下で考慮する。

バリデーション試験が代表的であることは重要である。すなわち、可能な限り試験は、方法の範囲内での濃度範囲と試料の種類を網羅することと同様、通常の方法の使用中に作用している影響の数と範囲を現実的に概観できるように実施しなければならない。例えば、ある要因(室温など)が代表として、精度試験中に不規則に変化する場合、その要因の影響は、観察された変動に直接現れるため、更に他の方法の最適化が望ましい場合以外は、追加試験は必要でない。

方法バリデーションとの関係で「代表的変動」とは、要因が、議論されているパラメータの予想範囲に適切な値の分布を取らなければならないことを意味する。継続的な測定可能パラメータに関して、これは認められた範囲、あるいは予想された範囲で、不確かさと言われる。断続的な要因、または試料マトリックスのような予測できない影響を持つ要因に関しては、代表的な範囲は、方法の通常の使用において認められたまたは直面した種類の多様性、あるいは「要因レベル」と対応する。理想的には、代表性は、値の範囲だけでなく、それらの分布にまで拡大する。残念なことに、様々なレベルにおける多数の要因のすべての変化に対して計画を立てることは、経済的でないことがしばしばある。しかし、最も実用的な目的のためには、予想される範囲の両極端に、あるいは予想よりも大きな変化に基づく試験が、認められる最低限度となる。

変動に関する選択要因において、できる限り大きな影響が「発揮される」ことを保証することが重要である。例えば、日間の変動(恐らく、再キャリブレーションの影響が原因と考えられる)が、併行精度と比較してかなり大きい場合、5日間のうちそれぞれ2回測定するほうが、2日間の各々で5回測定するよりも室内再現精度をよりよく推定できると考えられる。十分なコントロールを必要とする、別々の日に単一の測定を10回する方がなお良いが、これでは、日内併行精度に関する追加情報が得られないと思われる。

明らかに、有意性確認の計画において、いかなる試験も、それらが事実上、重要となる前に、そのような影響を検出するための十分な能力を持たなければならない(すなわち、不確かさの最も大きな要因と比較できる)。さらに、以下の考慮点が重要であると考えら

れる：

- ・ 要因は、相互作用することが知られている、または疑いがある場合、相互作用の影響を説明できることを保証することが重要である。これは、相互作用するパラメーターを異なるレベルから無作為に選択して保証することにより、あるいは「相互作用」の影響または共分散を得るための慎重に系統づけられた設計により達成できると考えられる。
- ・ 全体的な偏りの試験を実施する際に、標準物質及び基準値が、定期試験における物質に適切であることが重要である。

## 6. バリデーション試験の範囲

試験施設が、新規、修正またはあまり知られていない方法のバリデーションを行わなければならない範囲は、ある程度、方法の既存状態、及び試験施設の能力に依存する。異なる条件に対するバリデーションの範囲、及び検証手段に関する提案を以下に示す。ここで述べたこと以外は、定期的な使用が方法の目的であると推定される。

- ・ 試験施設は、「完全に」妥当性の確認された方法を使用することになっている。方法は共同試験で試験されるため、試験施設は公表された方法の性能特性を達成できる能力があること（またはそうでなければ分析業務の要求事項を遂行できること）を証明しなければならない。試験施設は、堅牢性試験のような一部の試験を省略できるかもしれないが、精度試験、偏り試験（マトリックス変動試験を含む）及び恐らく直線性試験を実施しなければならない。
- ・ 試験施設は、完全に妥当性の確認された方法を使用することになっているが、新規マトリックスを使用することになっている。方法は、共同試験で試験されるため、試験施設は、新規マトリックスがその系に誤差を引き起こす新たな原因とならないことを証明しなければならない。以前と同様のバリデーションの範囲が要求される。
- ・ 試験施設はうまく確立された方法を使用することになっているが、共同的に試験された方法を使用しないことになっている。以前と同様のバリデーションの範囲が要求される。
- ・ 方法は、一部の分析特性と共に科学論文で発表されている。試験施設は、精度試験、偏り試験（マトリックス変動試験を含む）、堅牢性及び直線性試験を引き受けなければならない。
- ・ 方法は、経験に基づく。経験に基づく方法は、予想される量が、単に以下に述べる手順で認められた結果となる方法である。これは、試料中の特定の分析物の濃度のような、方法から独立した量を評価することを目的とした測定とは異なる。ここでは方法の偏りは慣例的に0であり、マトリックス変動（すなわち定義された階級内）は無関係である。試験施設の偏りは無視できないが、単一試験施設実験で推定することは難しいようである。更に、標準物質が利用できそうにないと考えられる。共同試験データの無い状態では、試験施設間精度は、特別に設計された堅牢性試験から一部推定でき、あるいは、Horwitz functionを用いることで推定できる。
- ・ 分析は「ad hoc（その場限り）」である。「ad hoc」分析は、莫大な費用を費やすことなく、且つ重要性が低い値の一般的な範囲を確立するのに時々必要である。バリデーションを行う労力は、結果的には厳格に制限される。偏りは、回収率の推定や分析物の添加、及び繰り返しによる精度といった方法で試験しなければならない。
- ・ 試験従事者及び機器の変更。重要な例には、主要機器の変更、非常に変わりやすい試薬

(例えば、ポリクローナル抗体) の新規バッチ、試験施設内でなされた変更、新規従事者が初めて使用した方法、または、廃止時期後に採用した、妥当性の確認された方法などが含まれる。ここで、変更により悪影響を与えることがなかったことを実証することが必須の作業となる。最低限の確認は、単一偏り試験、すなわち、典型的な試験物質またはコントロール物質に関する「前後」実験である。一般的に、実施された試験は、変更が分析手順へ及ぼす予想される衝撃を反映していなければならない。

## 7. 勧告

単一試験施設方法バリデーションの使用に関して、以下の勧告が出される：

- ・可能性のある、及び実施できる時はいつでも、試験施設は、国際プロトコルに準拠する共同試験を通して評価された性能特性を持つ分析法を使用しなければならない。
- ・そのような方法が利用できない場合、その方法を使用して顧客用の分析データを得る前に、社内でその方法の妥当性を確認しなければならない。
- ・単一試験施設バリデーションでは、試験施設が、評価のために以下の項目から適切な特性を選択することが要求される：適応性、選択性、キャリブレーション、精確さ、精度、範囲、定量限界、検出限界、感度及び堅牢性。試験施設は、どの特性を測定するかを選択する際に、顧客の要求を考慮しなければならない。
- ・試験施設の顧客が要求した場合に、これらの特性が評価されたとの証拠を顧客が利用できるようにしなければならない。

### 添付資料A：試験の方法性能特性の要求事項に関する注意事項

方法の個々の性能特性に関する一般的な要求事項は、次の通りである。

#### A1. 適応範囲

バリデーション後に、あらゆる性能仕様に加えて、以下の情報を記載した文書を作成しなければならない。

- ・必要に応じて、種分化を含む（例えば「全ヒ素」）、分析物の同一性；
- ・バリデーションで網羅される濃度範囲（例えば「0～50 ppm」）；
- ・バリデーションにより網羅された試験物質（例えば「海産食品」）のマトリックス範囲の仕様書；
- ・機器、試薬、手順（特定の指示において認められる変動を含む、例えば「100±5℃で30±5分間」加熱）、キャリブレーション、品質手順、及び要求されたあらゆる特別な安全措施を記載したプロトコル；
- ・意図されている用途及びその重大な不確かさの要求事項（例えば「スクリーニング目的の食品分析」。結果  $c$  の標準不確かさ  $u(c)$  は、 $0.1 \times c$  未満でなければならない）。

#### A2. 選択性

選択性は、ある方法が夾雑物の存在下で分析物を正確に定量できる程度である。理想的に選択性は、存在すると考えられるあらゆる重要な夾雑物に関して評価しなければならない。化学原則において、試験に応答する可能性のある夾雑物を確認することが、特に重要であ

る。例えば、アンモニアの比色試験では、主に脂肪族アミンが反応するとの予想が、最も理にかなっている。潜在的な夾雑物を全て考慮、あるいは試験することは実行不可能であることもある；これが、最悪と考えられるケースを確認することが推奨される状況である。一般原則として、選択性は、いかなる夾雑物も無視できるだけ十分に良好なものでなければならない。

分析の多様な種類において、選択性は、本質的に有意性に基づく、またはそうでないにしても、夾雑物に対して適切な試験に基づく定性的評価である。しかし、有益な定量的手段がある。特に、定量測定の一つは、選択性インデックス  $b_{an}/b_{int}$  であり、夾雑物を定量的に測定できる。ここで  $b_{an}$  は、方法の感度（キャリブレーション関数の傾き）、 $b_{int}$  は潜在的な夾雑物により発生する独立した応答の傾きである。 $b_{int}$  は、マトリックスブランク及び適切な一濃度で潜在的な夾雑物を添加した同様のブランクについて操作を行うことで大まかに決定できる。マトリックスブランクが利用できず、典型的な物質を代わりに使用する場合、 $b_{int}$  は、相互マトリックス効果が無いという仮定においてのみ、そのような単純な実験により推定することができる。 $b_{int}$  は、分析物の感度自体が夾雑物（マトリックス効果）により影響を受ける場合、夾雑物の別の種類と影響が混乱する可能性があるため、分析物の存在しない状態の方がより容易に決定できることに留意が必要である。

### A3. キャリブレーション及び直線性

キャリブレーション物質の調製の際の大きな誤差を除いて、キャリブレーション誤差は、通常（しかし、いつもとは限らない）、不確かさの総量に対し小さい要素であり、「トップダウン」方法により推定される様々なカテゴリーに包括させても安心である。例えば、キャリブレーションで生じた偶然誤差は、全体的に評価されるランの偏りの一部となる。その一方、原因が試験施設の偏りとして現れることがある系統的誤差は、総括的に評価されると考えられる。それにもかかわらず、手順の最適な開発方針に影響を及ぼすことから、方法バリデーションの着手時に知っている有効な一部のキャリブレーション特性がある。この様な階級においては、キャリブレーション関数が、(a)直線、(b)原点を通る、及び(c)試験物質のマトリックスには影響を受けない、のいずれに最も近いかが問題である。ここに記載した手順は、必然的に、定期分析中に引き受けるキャリブレーションよりも厳格となるバリデーションにおけるキャリブレーション試験と関連する。例えば、一度バリデーション試験で、キャリブレーション関数が直線で原点を通ると確立されれば、定期的な使用では、更に簡単なキャリブレーション方針を用いることができる（例えば、2点繰り返し設計）。このより簡単なキャリブレーション方針から生じる誤差は、通常のバリデーション目的での高いレベルの誤差に包括されると考えられる。

#### A3.1 直線性と切片

直線性は、適切なキャリブレーションセットにおいて、応答の直線回帰により得る残差を濃度に対してプロットして調べることで、非公式に試験できる。いずれの曲線パターンも、非線形のキャリブレーション関数のため適合しないことを示す。有意検定は、適合度変動を、純粋な誤差の変動と比較することで実施できる。しかし、適合しない原因は、分析キャリブレーションのある種類において起こる可能性のある非曲線であること以外にも存在するため、有意検定は、残差プロットと組み合わせて使用しなければならない。これは現在、適合する品質の指標として広く使用されているが、相関係数は、直線性に関する試験

としては誤りが起きやすく適切でないため、使用すべきではない。

非線形性とドリフトとを混同し易いことから、設計は、適合度に関する試験では極めて重要である。繰り返し測定は、独立した推定がなされていない場合、純粋な誤差を推定するのに必要である。特定の指針がない状態では、以下の項目を（単変量直線キャリブレーションに）適用しなければならない：

- ・6点以上のキャリブレーション標準がなければならない；
- ・キャリブレーション標準は、対象範囲の濃度に渡って等しい間隔をあけなければならない；
- ・濃度範囲は、遭遇すると予想される濃度の0～150%または50～150%で、より適切と思われる範囲を網羅しなければならない。
- ・キャリブレーション標準は、少なくとも2連で、好ましくは3連以上で、無作為の順番でランしなければならない。

単純な直線回帰へ試験的に適合させた後、明確なパターンについて残差を調べなければならない。分析キャリブレーションにおいて専ら異分散が一般的であり、異分散の示すパターンは、キャリブレーションデータを加重回帰で最もうまく取り扱えることを意味する。この様な状況で加重回帰を用いない場合、キャリブレーション関数の低濃度末端で、誤差が誇張して生じる可能性がある。

適合度に関する試験は、単純なまたは加重回帰により実施できる。適合しないことが有意でない場合、0とは有意に異なる切片に関する試験をこのデータに関しても実施できる。

### **A3.2 一般的なマトリックス効果に関する試験**

分析物の単純溶液としてキャリブレーション標準を調製できる場合、キャリブレーションは非常に簡略化する。この方針が採用された場合、予想される一般的なマトリックスが不釣り合いであることの影響をバリデーションで評価しなければならない。一般的なマトリックス効果に関する試験は、分析物を典型的な試験物質から作成した試験液へ添加する方法を適用して（「標品添加」とも言われる）実施できる。試験は、通常の手順で作成するものと同じ最終希釈となるような方法で実施しなければならない。また、添加範囲は、手順で定義されたキャリブレーションのバリデーションと同じ範囲を網羅しなければならない。キャリブレーションが直線の場合、通常のカリブレーション関数の傾きと分析物の添加プロットとで、有意差について比較できる。有意差が無いことは、検出可能な一般的なマトリックス効果が無いことを意味する。キャリブレーションが直線でない場合、有意検定には、より複雑な方法が必要であるが、同じ濃度で視覚的に比較することで、通常は十分だと考えられる。この試験で有意性が無いことは、マトリックス変動効果（A13項）も無いと考えられることを意味することもある。

### **A3.3 最終キャリブレーション手順**

手順で規定したようなキャリブレーション方針は、含まれる誤差が組み合わさって、推定された不確かさに寄与する可能性もあるが、別々に妥当性を確認することも必要かもしれない。ここで重要なことは、直線性等に関する特定の設計から推定された不確かさは、手順プロトコルで定義されたより簡単なキャリブレーションから導かれるものより小さい可能性があることである。

## **A4. 真度**

#### A4.1 真度の推定

真度は、試験結果と適切に測定された認証基準値との一致の近さである。真度は「偏り」の観点から定量的に記載され、偏りが小さいほど、真度が大きいことを示す。偏りは、典型的に、標準物質に対する方法の応答と、その物質に割り当てられた既知の値との比較により決定される。有意検定が推奨される。基準値の不確かさが無視できない場合、統計的変動性と同様、標準物質の不確かさを考慮して、結果を評価しなければならない。

#### A4.2 真度試験に関する条件

偏りは、例えば、ランの偏り、試験施設の偏り及び方法の偏りといった、分析システムにおける組織の様々なレベルで発生する可能性がある。偏りを取り扱う様々な方法によって、これらのうちのどれを対処するかを覚えていることが重要である。特に：

- ・1回のラン内ですべて実施された標準物質の一連の分析の平均から、特定のランに対する方法、試験施設及びランの全てを合わせた影響についての情報が得られる。ランによってランの影響は無作為であると予想されるため、結果の観察されうる分散から予想されるよりも、結果はランによってより変化に富むと考えられ、これを結果の評価（例えば、別に調査したラン間の標準偏差に対する測定された偏りを試験することにより）において考慮する必要がある。
- ・数回のランにおいて標準物質を繰り返し分析した平均から、特定の試験施設における方法及び試験施設の偏りを組み合わせた影響が推定される（特定の方法を用いて値を割り当てた場合を除く）。

#### A4.3.1 真度試験に関する基準値

##### A4.3.1 認証標準物質（CRM）

CRMは、既知の不確かさを用いて国際基準で追跡可能であり、従って、マトリックスの不一致が無いと仮定すれば、偏りの全ての局面（方法、試験施設及び試験施設間）を同時に扱うために使用することができる。結果的に CRMは、そうすることが実行可能な場合、真度のバリデーションにおいて使用すべきである。認証値の不確かさが、重要な大きさの偏りを検出できるだけ十分に小さいことを保証することが重要である。そうできない場合でもCRMの使用が推奨されるが、追加的な確認が必要である。

典型的な真度検定では、標準物質に関する平均応答を発生させる。結果を解釈する際には、試験施設における統計学的変動から生じる不確かさと共に認証値と関連する不確かさを考慮しなければならない。後者は、ラン内、ラン間または実験の意図に依存する試験施設間標準偏差の推定に基づくこともある。認証不確かさの値が小さい場合、通常、適切な精度条件を用いて、student'sのt検定を実施する。

実施可能な場合、必要に応じて、適切なマトリックス及び分析物濃度を用いて多数の適切なCRMを試験しなければならない。これが実施され、分析結果に関する不確かさよりも認証値に関する不確かさの方が小さい場合は、単純回帰を使用してその結果を評価しても、ほぼ安全であるといえる。この様に、偏りは、濃度の関数として表すことができ、0 以外の切片（「暫定の」または定数の偏り）、あるいは単一性でない傾き（「回転の」または比例の偏り）として現れる可能性がある。マトリックス範囲が大きい場合、結果の解釈時に注意を払うべきである。

#### A4.3.2 標準物質

CRMが利用できない場合、またはCRMに追加するときは、偏りが有意でないことが必ずしも0の偏りとなる証拠とはならない可能性があること、また、いずれの物質に関しても有意な偏りがある場合はいつも調査の対象となることを念頭におきながら、目的(標準物質[10])に対して十分に特徴づけられたあらゆる物質を使用すると考えられる。標準物質には以下の例が含まれる：標準物質の作製者が特徴付けを行うが、それらの値は、不確かさが記載されないか、定性されていない物質；物質の製造メーカーが特徴付けを行った物質；標準物質として使用する試験施設において特徴付けを行った物質；制限された総当たり実験、または技能試験に割り当てられた物質。これらの物質のトレーサビリティは疑わしいかもしれないが、偏りに関する評価を全く行わずに使用するよりもずっと良いと考えられる。明確な不確かさがなければ、いずれの有意検定も完全に結果の観察可能な精度に依存するが、物質は、CRMとほぼ同様に使用されると考えられる。

#### A4.3.3 参照方法の使用

参照方法は、原則的にバリデーションにおける別の方法で偏りに関して試験するために使用できる。これは、既に妥当性が確認され、試験施設で使用されている、確立された標準法の代替または修正を確認する際に有効な選択である。両方法は、多数の典型的な試験物質を、好ましくはかなり均等に有効な濃度範囲を網羅しながら分析するために使用される。適当な統計的手法(例えば、変動と正規性の均一性に関する正常な確認のための対応 t-検定)を用いた、その範囲に渡る結果の比較により、方法間のあらゆる偏りが示されると考えられる。

#### A4.3.4 添加／回収率の使用

標準物質がない場合、あるいは標準物質試験を裏付けるため、偏りは、添加及び回収率により調べることができる。典型的な試験物質は、その元の状態及び試料へ既知量の分析物を加えた(添加)後の試料の両者について、バリデーションを行っている方法を用いて分析する。添加した物質量の割合としての二つの結果の相違は、代替回収率、または、場合により限界回収率と呼ばれる。回収率は、偏りが方法に影響していることを示す単位とはかなり異なる。厳密に、ここに記載した回収試験では、添加した分析物について作用する影響による偏りのみを評価する。同様の影響が、必ずしも天然分析物に対して同程度に当てはまるわけではなく、追加的な影響が天然分析物に適用されることもある。添加／回収試験では、結果的に、良好な回収率だからと言って真度は保証されないが、低い回収率では、明らかに真度がないとの所見が非常に強く持たれやすい。添加／回収データを取り扱う方法は、別の箇所[4]で詳細を述べている。

### A5. 精度

精度は、規定された条件下で得られた独立した試験結果どうしの間の一貫性の近さである。通常、標準偏差または相対標準偏差という言葉で定義される。精度と偏りの区別は基本的であるが、分析システムの検討されるレベルに依存する。従って、単回測定の見点から、そのランについてキャリブレーションに影響するあらゆる偏差は、偏りと見なされるであろう。長年の作業を検討している分析者の見点から、ランの偏りは毎日異なり、関連精度

と共に確率変数の様に作用すると考えられる。精度を推定するための規定条件では、この変化を考慮する。

単一試験施設バリデーションに関して、二組の条件が関連する：(a) 期待値0、標準偏差 $\sigma_r$ として単回ラン中に観察された変動を示している、併行精度条件下の精度、及び(b) 期待値0、標準偏差 $\sigma_{run}$ として、ランの偏りにおける変動 $\delta_{run}$ を示しているラン毎の条件下の精度。通常、これらの誤差の原因はいずれも、個々の分析結果において操作され、従って、合成精度 $\sigma_{tot} = (\sigma_r^2/n + \sigma_{run}^2)^{1/2}$ を持つ。ここで、 $n$ は、1回のラン内で繰り返した結果を、報告値とするために平均した数値である。2つの精度推定値は、多数の連続ランにおいて、選択した試験物質を二連で分析することで最も簡単に得ることができる。別の変動要素は、その後、変動に一元配置分散分析を適用することで求められる。二連分析では、個々の試料に適用する手順を独立して実施しなければならない。別の方法として、合成精度 $\sigma_{tot}$ は、連続ランにおいて試験物質を一度分析し、通常式から標準偏差を推定することで直接推定できる。(観察された標準偏差は、一般的に $s$ の記号で表し、母集団偏差 $\sigma$ とは区別することに注意すること。)

精度の値は、予想される試験条件の代表であることが重要である。まず、ラン間の条件における変動は、その方法を定期的使用する条件下の試験施設において、通常起こると考えられることを代表していなければならない。例えば、試薬バッチ、分析者、および器具における変動は、代表的なものでなければならない。次に、使用する試験物質は、マトリックス及び(理想的には)粉砕状態の観点から、また、定期的な適用で遭遇すると考えられる物質の観点から典型的でなければならない。従って、実際の試験物質、またはそれほど重要性はないがマトリックスが適合した標準品は適切であり、分析物の標準溶液は適切でないこともある。CRM及び調製された標準物質をしばしば典型的な試験物質よりかなりしっかりと均一にするため、それらの分析で得られた精度は、それに応じて、その試験物質に関して観察されると予想される変動を過小評価する可能性があることにも留意しなければならない。

精度は、分析物濃度により非常に変動しやすい。典型的な推定は、(i) 分析物レベルで精度に変化がない、または(ii) 標準偏差は、分析物レベルに比例、または直線的に依存する、である。いずれの場合も推定値に関して分析物レベルが実質的に変動する可能性があるか(すなわち、中央値から約30%以上)について確認が必要である。最も経済的な実験は、分散の違いに関する適切な統計学的試験と共に、操作範囲の両極値で、あるいはその付近で、単純に精度を評価することと考えられる。F-検定は、正規分布の誤差に適切である。

精度データは、最低限の併行精度、及びここで示したラン間の条件に加えて、異なる一連の条件の幅広い変動に関して得られる可能性もあり、追加情報を得るのに適切なこともある。例えば、別々の操作者及びランの影響、日間または日内の影響または一つあるいは数種の機器を用いて達成できる精度の指標を持つことは、結果を評価するのに、または、測定を改良するのに有効であると考えられる。異なる設計範囲及び統計学的分析技法が利用可能であり、慎重な実験計画がこの様な試験全てにおいて強く推奨される。

## A6. 回収率

回収率を推定する方法は、真度を推定する方法（上記）と合わせて議論する。

## A7. 範囲

妥当性の確認された範囲とは、方法が妥当と見なすことのできる範囲内での分析物濃度の間隔である。この範囲が、必ずしもキャリブレーションの有効範囲と一致するわけではないことを認識することが重要である。キャリブレーションは、広い濃度範囲を網羅することがある一方、バリデーションの残差（そして通常、不確かさの観点から非常に重要である）が網羅する範囲は、更に限定される。実質的には、ほとんどの方法は、濃度の1、2点のレベルでしか妥当性が確認されないであろう。妥当性の確認された範囲は、濃度スケールに関するこれらの点から外挿することが合理的であると見なされることもある。

検出限界より十分上の対象濃度に焦点を絞って方法を使用する場合は、一点の重要なレベル付近のバリデーションが適切であると考えられる。ほとんどが個々の分析システムに依存するため、この結果を分析物の他の濃度へ外挿する一般的に安全な方法で定義することは不可能である。従って、バリデーション試験の報告書には、バリデーションの実施者が熟練した判断をもって、推定された不確かさが真実であると見なす、バリデーションを実施する危険値付近の範囲を記載しなければならない。

対象の濃度範囲が0または検出限界に近づく場合、一定の絶対不確かさまたは一定の相対不確かさのいずれかを推定することは間違いである。この一般的な状況では、正の切片をもつ、不確かさ  $u$  及び濃度  $c$  との間の直線関数相関を推定する概算方法が有効である。この式は、

$$u(c) = u_0 + \theta c \quad \text{である。}$$

ここで、 $\theta$  は、検出限界よりも十分上のいくつかの濃度で推定された相対不確かさである。 $u_0$  は、濃度0で推定された標準不確かさであり、場合によっては、 $c_L/3$ として見なすことができる。これらの状況において、有効性の確認された範囲を、0から、最も高いバリデーションポイントの小さい整数倍数まで拡張したと見なすのが適切であると考えられる。

## A8. 検出限界

大まかに、検出限界は、0とは確実に区別できる試料中の分析物の最少量、または最低濃度である[22、23]。バリデーション範囲が含まれない、またはその付近でない分析システムでは、検出限界はバリデーションの一環とする必要はない。

考えが明らかに単純であるが、検出限界の全体的な課題を取り巻く問題の概略を以下に示す：

- ・その問題に対して、いくつかの可能な概念アプローチが存在し、それぞれが検出限界について幾分異なる定義を示している。問題を明確にしようと試みることは、更に混乱を招くように思われる。
- ・これらの各アプローチは、濃度 0 付近での精度の推定値に依存するが、これが併行精度の条件を示すと見るか、推定のためのいくつかの別の条件を示すと見るかが明確でない。
- ・莫大な量のデータが集められない限り、検出限界の推定では、かなり大きな偶然誤差が

生じやすい。

- 操作的な要因のため、検出限界の推定値は、低濃度側に偏りやすい。
- 検出限界と関連する統計的推測は、少なくとも低濃度では疑問視される、正規性を推定することに基づく。

方法バリデーションにおける最も実用的な目的に関して、方法が有効かの大まかな指針としてのみ使用する、迅速に推定ができるようにする簡単な定義を選択する方が良いように思われる。しかし、方法開発において推定される検出限界は、完全な分析法を特徴づけるために使用されるものとは、概念または数値において一致しない可能性があることを認識しなければならない。例えば、文献または機器カタログで引用され、希釈で調整される「機器の検出限界」は、「実用的な」検出限界よりもかなり小さいことがよくあり、方法バリデーションには適さない。

結果的に方法バリデーションに関して、用いた精度推定値 ( $S_0$ ) では、典型的なマトリックスブランク、または低濃度物質において、0 または負の結果を検閲せずに、少なくとも異なる 6 濃度で完全に分析物濃度を測定した結果に基づくことが推奨され、検出限界の近似値は  $3S_0$  として算出される。推奨された自由度の最小値と共に、この値は、かなり不確実であり、容易にファクター2の誤差となりうることに注意を払う必要がある。より厳格な推定が必要な場合 (例えば、検出に基づく、または物質の決定を裏付ける目的)、適切な指針を参照すべきである (例えば、参考文献22、23)。

#### **A9. 測定限界または定量限界**

ある濃度以下では、その分析法では許容可能な精度を持って操作できないその濃度について述べるのが有益なことが度々ある。その精度は、任意に10%RSDと定義されることがよくあり、場合によってその限界は、同様に任意に、検出限界の一定の倍数 (典型的には2倍) とすることもある。それは、そのような限界以上で安心して操作する程度ということになるが、我々は、それは濃度スケールの人工的な両断理論であることを認識しなければならない；そのような限界以下での測定が情報量に欠けるわけではなく、目的に適合させるのはもっともと考えられる。従って、バリデーションにおけるこのような種類の限界の使用は、ここでは推奨しない。測定の不確かさを、濃度の関数として表し、試験施設及び顧客またはデータの最終的な使用者との間で承認された、目的に適合した基準を持つ関数と比較することを試みる方が好ましい。

#### **A10. 感度**

方法の感度は、キャリブレーション関数の勾配である。通常これは、機器設定に依存して任意であるため、バリデーションでは有効でない (しかし、品質保証手順では、機器が基準に一致して、十分に機能しているかどうかを試験することが有効なこともある。)

#### **A11. 堅牢性**

分析法の堅牢性は、手順に書かれた実験条件からわずかに逸脱した際に、その分析法を用いて得られた結果の変化に対する耐性である。実験パラメーターの限界は、方法プロトコルに記載しておかなければならない (以前は、必ずしもこのことは実施されていなかった)。また、このような認められる逸脱は、別に、またはいくつかを組み合わせ、得られた結

果に重大な変化をもたらしてはならない。（「重大な変化」とは、ここでは、その方法が、目的適合性を定義している不確かさの承認された限界内で操作できないことを意味する。）結果に影響を及ぼす可能性のある方法の局面を識別し、堅牢性試験を用いて、それらの方法性能への影響を評価しなければならない。

方法の堅牢性は、手順への小さな変更を慎重に行い、結果への影響を調べることにより試験される。方法の数多くの局面を考慮する必要があるが、これらの大部分は、影響がほんのわずかであるため、通常いくつかを一度に変更することが可能である。Youdenは、分割要因計画に基づく経済的な実験について報告している[13]。例えば、7つの変動要因の8つの組み合わせを用いる手法を公式化すること、すなわち、ちょうど8個の分析結果を用いて7つのパラメーターの影響を見ることが可能である。単変量手法も適しており、ここでは、一度にひとつの変量のみを変化させる。

堅牢性試験を取り扱える要因の例は、以下の通りである：機器、操作者あるいは試薬銘柄の変更；試薬の濃度；溶液のpH；反応温度；工程の完了までに要した時間等である。

#### A12. 目的適合性

目的適合性は、分析者とデータの最終使用者との間で承認された、判定基準と適合する方法性能の程度であり、最終使用者の要求を述べている。例えば、データにおける誤差は、定義された低い可能性よりも高い頻度で不正確な決定をもたらす可能性のある程の大きさであってはならないが、小さすぎて最終使用者に不必要な出費をさせるものであってはならない。目的適合性の判定基準は、本添付資料に記載した一部の特性に基づくことができるが、結局のところ、許容可能な合成不確かさの観点から表現されると考えられる。

#### A13. マトリックス変化

多くの部門でマトリックス変化は、分析測定において最も重要なものの一つであるが、ほとんど知られていない誤差の原因の一つである。例えば、試料のマトリックスを特定することで妥当性を確認する分析システムを定義する場合、定義された階級内でかなりの変動範囲があるかもしれない。極端な例を引用すると、「土壌」試料の階級は、粘土、砂、チャーク、紅土（主に $\text{Fe}_2\text{O}_3$ と $\text{Al}_2\text{O}_3$ ）、泥炭等から、あるいはこれらの混合物から構成される。これらの種類の各々が、原子吸光分析法のような分析法に対して独自のマトリックス効果を寄与させる可能性があると考えられる。自分たちが分析している土壌の種類に関する情報が無い場合、この変動しやすいマトリックス効果により、それ相応の余分な不確かさが生じると考えられる。

マトリックス変動の不確かさは、バリデーション過程において他では考慮していないため、別に定量することが必要である。適切な範囲での分析物濃度を用いて、定義された階級内で遭遇する可能性のある一連の代表的なマトリックスを集めることでその情報が得られる。プロトコルに従って物質を分析し、結果における偏りを推定する。試料がCRMでない限り、偏りの推定は通常、添加・回収率推定の手法により実施しなければならない。不確かさは、偏りの標準偏差により推定される（注意：この推定には、繰り返し分析からの変動の寄与も含まれる。これは、添加が用いられる場合、大きさ  $2\sigma_r$  を持つと考えられる。

厳格な不確かさの量が必要な場合は、これを、マトリックス変動による変動からは差し引き、二重に計測することを避けなければならない。)

#### A14. 測定不確かさ

測定不確かさ推定の正式な手法では、式または数式モデルから測定不確かさの推定値を算出する。方法バリデーションとして記載された手順は、全ての種類の偶然誤差が十分に与えられて、結果を推定するために使用された式が全ての認識された、また全ての結果への有意な影響を具体化する妥当な表現であることを保証するように設計される。要するに、以下で更に詳細を述べた一つの警告と共に、バリデーションを受けた式または「モデル」は、測定不確かさを推定するために、直接使用されることもある。これは、「不確かさの伝播法」に基づく、確立された以下の原則によりなされる。ここでは、独立したインプット効果は、

$$u[y(x_1, x_2, \dots)] = \sqrt{\sum_{i=1, n} c_i^2 u(x_i)^2} \quad \text{である。}$$

ここで、 $y(x_1, x_2, \dots, x_n)$ は、いくつかの独立変数 $x_1, x_2, \dots$ の関数であり、 $c_i$ は、 $c_i = \partial y / \partial x_i$ として評価された感度係数、 $x_i$ 、 $u(x_i)$ 及び $u(y)$ に関する偏微分 $y$ は、標準不確かさである。つまり、測定不確かさは、標準偏差の形式で表現される。 $u[y(x_1, x_2, \dots)]$ は、いくつかの別々の不確かさの推定値の関数であるため、合成標準不確かさと言われる。

従って、結果を求めるために用いた式  $y = f(x_1, x_2, \dots)$  から測定不確かさを推定するために、まず、各条件  $x_1, x_2$ , 等で、不確かさ  $u(x_i)$ を確立することが必要であり、次に、これらを、バリデーションで認められた通りの変量効果を表現するのに必要な追加条件と組み合わせることが、最後に、更にあらゆる追加的な影響を考慮に入れることが必要である。先に述べた精度の議論において、示された統計学的モデルは、

$$y = f(x_1, x_2, \dots) + \delta_{\text{run}} + e \quad \text{である。}$$

ここで、 $e$  は、特定結果に対する偶然誤差である。精度実験から、 $\delta_{\text{run}}$  と  $e$  はそれぞれ、標準偏差  $\sigma_{\text{run}}$  と  $\sigma_e$ を持つことが知られているため、これらの後者の条件(または厳格には、それらの推定値 $s_{\text{run}}$  及び  $s_e$ ) は、これらの追加的な条件と関連する不確かさである。個々のラン内の結果を平均すると、これらの二つの条件(以前与えられたもの)と関連する合成不確かさは、 $s_{\text{tot}} = (s_e^2/n + s_{\text{run}}^2)^{1/2}$ となる。精度の条件が分析レベルにより変化することが示された場合は、所定の結果に対する不確かさの推定値では、そのレベルに適切な精度の条件を採用しなければならない。不確かさの推定値に関する偏りは、結果として、バリデーションにおいて推定され、試験された統計学的モデルから直接得られる。この推定値に、不均一性及びマトリックス効果(セクションA13参照)を(特に)説明するのに必要な条件を更に付け加えなければならない。最終的に、求めた標準不確かさに「包括係数」 $k$ を乗じて、拡張不確かさ、すなわち「測定量に起因する可能性のある値の分布の大きな端数を網羅するために予想された間隔」を得る[8]。統計学的モデルがうまく確立されている場合、分布は正規になることが知られており、推定値と関連する自由度の数は高くなり、一般に $k$ は2とされる。従って、拡張不確かさは95%信頼区間に相当する。

ここで、一つ付け加えるべき重要な警告がある。推定統計学的モデルを試験する際、不

完全な試験が必然的に使われる。これらの試験では、いずれの影響も等しくゼロであることを証明できないことを指摘してきた；それらでは、影響が小さすぎて、有意性に関する特定の試験と関連する不確かさの範囲内で検出できないことを示せるだけである。特に重要な例は、試験施設の偏りに関する有意検定である。明らかに、真度のみを確認するために実施した試験であれば、その方法は実際に偏りがないか否かに関する一部の残差不確かさを持つに違いない。そのような不確かさがそれまで算出された不確かさの観点から有意である場合、追加的な処置を行わなければならない。

不確実な基準値の場合、最も簡単な許容度は、適用された試験での統計学的不確かさを組み合わせた、その物質に関する規定の不確かさである。これを完全に議論することは、本文の範囲外である；詳細を知りたい場合は、参考文献9を参照すること。しかし、推定された統計学的モデルから直接推定した不確かさは、分析結果と関連できる最小の不確かさであるが、ほぼ確実に、過小評価されると考えられること；拡張不確かさは、同様の考えに基づき、 $k=2$ を用いることでは、十分な信頼性は得られないことに注意を払うことが重要である。

ISOガイド[8]は、信頼性を向上させるため、条件を任意に追加するよりも、要求に応じて $k$ の値を増加させることを推奨する。妥当性の確認された統計学的モデルに基づくが、モデルにおいて更に信頼性を与えるためのバリデーション試験を超える証拠がない不確かさの推定値に関して、 $k$ は3未満であるはずがないことが、実際の経験により示された。バリデーション試験が、複雑であると疑う強い理由がある場合、 $k$ は、要求された値よりも増加するはずである。

## 付録B. 不確かさに関する追加考慮点

### バリデーション試験における推定

#### B1. 感度分析

不確かさの推定において用いられる具体的な表記、

$$u[y(x_1, x_2, \dots)] = \sqrt{\sum_{i=1, n} c_i^2 u(x_i)^2}$$

では、「感度係数」 $c_i$ が必要である。不確かさの推定において、所定の影響因子 $x_i$ は、既知の不確かさ $u(x_i)$ を持っているが、係数 $c_i$ は十分には特徴づけられない、または結果の式からは簡単には得られないと分かるのが一般的である。特に、影響が測定式に含められない場合が一般的である。何故ならば、それが通常は有意でない、あるいは正しさを判定するのに十分に関係が理解されていないからである。例えば、抽出操作での室温への溶液温度の影響 $T_{sol}$ について、詳細に調べることは滅多にない。

そのような影響に関連する結果において、不確かさを評価することが望ましい場合、実験的に係数を決定することが可能である。これは最も単純に、 $x_i$ を変化させて、基本的な堅牢性試験とほぼ同様の方法で、結果への影響を観察することにより実施する。たいいていの場合、最初の実験で基準値以外のせいぜい2つの $x_i$ の値を選び、観察された結果からおおよその勾配を算出することで十分である。続いて勾配からおおよその $c_i$ の値を求める。従って、条件  $c_i u(x_i)$ は、決定できる（これは有意性、または結果への予想される影響を実証する実用的な方法の一つであることに注意すること）。

そのような実験において、観察された結果における変化は、信頼できる $c_i$ の算出には十分であることが重要である。これは予め予想することが難しい。しかし、影響量 $x_i$ に関する許容範囲、またはその量に関して拡張された不確かさ、すなわち有意性のない変化という結果の予想を仮定した場合、より広い範囲で $c_i$ を評価することが明らかに重要である。結果的には予想された範囲  $\pm a$  の影響量に関して（ここで  $\pm a$  は、例えば、認められた範囲、拡張された不確かさの間隔、または95%信頼区間である可能性がある）、可能ならば、信頼できる結果を保証するために、少なくとも $4a$ の変化を感度実験で採用することが推奨される。

## B2. 判定

影響が認識され、有意性がある可能性がある場合、不確かさの信頼できる推定値を得ることがいつも可能であるとは限らない。そのような状況において、ISO ガイドでは、専門的に考慮した不確かさの推定値の方が、不確かさを無視するより好ましいことを明確にしている。従って、不確かさの推定値が、潜在的に起こる可能性のある重要な影響に利用できない場合、分析者は予想される不確かさの判定に最善をつくし、合成不確かさの推定の際にそれを適用しなければならない。参考文献 8 では、不確かさの推定における判定基準の使用に関する指針を更に述べている。