

平成 29 年度厚生労働科学研究費補助金（食品の安全確保推進研究事業）
食品微生物試験法の国際調和に関する研究

分担研究報告書
食品微生物試験法の妥当性評価に関する研究

分担研究者 松岡 英明 東京農工大学 名誉教授
研究代表者 朝倉 宏 国立医薬品食品衛生研究所 食品衛生管理部 部長

我が国が ISO TC34/SC9 の P メンバーになり、今年度、東京で SC9 国際年次総会を実施したことは、食品微生物試験法の妥当性確認に関する諸問題を国際的視点で議論する絶好の機会となった。分担研究者は、既に、先行研究において、ISO 16140:2016 の邦訳、並びにそれに基づく、国内向けバリデーションガイドラインの作成を主導してきたが、特に 2016 改訂版において合理的に理解できない点を多々見出していた。また、同じく先行研究において、バリデーションに密接に関わる「測定不確かさ」に関する諸問題を整理してきた。関連するのは WG3（メソッドバリデーション）および WG2（統計学）であるが、そこから提起された課題は、参照法のバリデーション、ISO 16140 の第 3～6 分冊の編纂、そして測定不確かさの ISO 19036 の改訂が主であった。国内で議論していた内容は、既に過去の議論なのか、あるいは困難なために、取合えずスキップされてきたことなのか、推察するに止まった。本研究では、改めて、我が国で議論してきた問題点の本質を考察した。その結果、現在進行中の SC9 の作業に対する有用な提言になるとの認識が得られた。それによって、WG3 および WG2 のメンバーとの議論も深まるはずである。同時に、我が国としては、WG の作業に積極的に協力することが、P メンバーになったことの権利であり、同時に責務でもある。

A . 研究目的

我が国の微生物試験法の公定法を国際的にハーモナイズさせるためには、試験法ごとに、国際的に認証された試験法を手本（参照法）として、公定法と参照法の同等性を確認しなければならない。その確認スキーム（メソッドバリデーション）の手本とされてきた規格が ISO 16140 である。そこで、先行研究では、ISO 16140:2003 の邦訳、それに基づくガイドラインの策定作業を進めてきた。その過程で、2016 年 6 月に改訂版が出た。変更内容には、それに先立って出された AOAC:2012.2 版ガイドラインの内容と同調した、と推察される部分が少なからずあった。しかし、その部分については、既に、先行研究の中でも情報分析していたので特に大きな問題とはならなかった。

ところが、ISO 16140 の改訂版では、例えば、「ペアード（Paired）試験とアンペアード（Unpaired）試験」、「確定試験（Confirmation）」の要請のように、全く新しい概念の規定が加わったこともあり、その具体的な内容の理解に苦慮していた。

幸い、我が国は平成 27 年（2015）から ISO TC34/SC9 の P メンバーになり、今年度、東京で SC9 国際年次会議を実施した。本年度の第一の目的は、この会議に、日本側委員として出席し、ISO 16140 改訂版策定に直接関わった SC9 のメンバーと情報交換することであった。SC9 ではテーマごとに 20 以上のワーキンググループ（WG）で議論を重ねているが、メソッドバリデーションに関わる WG は、WG3「メソッドバリデーション」と WG2「統計学」である。そこで、本報告書で

は、特にこの2つのWGの報告から洞察された内容を報告する。

ところで、メソッドバリデーションに深く関係しながら、極めて厄介な問題である「測定不確か」は、研究分担者が、かつて研究代表者として調査研究した問題である。それが今回のSC9総会では、WG2(統計学)からISO 19036の改訂案として提示された。そのため、この問題についても、改めて問題点の本質を考察することとした。

なお、ISOの会議内容は非公開となっているので、WGの報告内容に関しては多少曖昧な表現になっていること、ご容赦願いたい。

B．研究方法

(1) ISO16140: 2016 および関連規格における問題点の整理

研究分担者として参画してきた先行研究「食品中の微生物試験法の開発及びその実効性・妥当性評価に関する研究分担研究」では、バリデーションガイドラインの作成に取り組んできた。その過程で抽出し、議論してきた課題を整理した。また、バリデーション最終段階の試験結果の評価において必須となる統計学の問題点を分析した。一方、測定不確かさの問題は、微生物試験に限ったことではないが、理化学試験や生化学試験に比べると、微生物試験の場合は標準物質がないために格段に難しい。周知のごとく、微生物試験では、コロニー計数法に基づく方法を「参照法」として、その結果を「標準」としてきた。そのことに起因する「微生物試験法における測定不確かさ」については、研究代表者として取り組んできた「食品の規格基準に係わる測定値に伴う不確かさに関する研究」で議論を重ねて

きたが、その結論を改めて見直し検討課題として整理した。

(2) 国際動向調査

上記により整理された課題を念頭に、文献情報の他、今年度はISO TC34/SC9の年次総会に委員として出席し、公式、あるいは非公式に国際動向の情報取得に努めた。関連するWGの委員と対面での意見交換を行った。

C．研究結果

(1) バリデーション実施システム

バリデーションガイドラインは最終的には「案」とするに留めた。公開が難しかった最大の理由は、我が国には未だ公的なバリデーション実施システムが無かったからである。ガイドラインの中には「専門家の判断を要する事項」が少なからずあるが、この場合の「専門家」とは、単に専門的な経験や知識をもった者というだけでなく、その判断をする権限をも与えられた者、という意味である。具体的には、国際的バリデーション実施機関であるAOAC、AFNOR、NordValなどが組織した専門委員会等である。我が国には、それに相当する組織が無い。そのため、この組織の構築を目指す議論や活動も行ったが、難しい状況は変わらなかった。そのため当面は、当該の厚労科研グループのなかで、アドホック組織を作り対応していくことが現実的な解ではないか、との意見などが出されていたが結論には至らなかった。

本年度に至り、HACCPに対応するために衛生指標菌に対する標準試験法の確立が喫緊の課題となった。しかし、衛生指標菌についての国際的な定義はなく、したがって国際的に認証された試験法もな

い。そこで、我が国としては国際的に通用する衛生指標菌試験法の開発において、是非、先導的役割を果たすべきであると考えられる。このような状況を背景に、公的なバリデーション実施システムを構築すべきとの声が高まることが期待される。

(2) 同等性の判定基準

微生物試験法では標準物質がない。参照法として認証されている試験法の結果が標準物質の代わりとなる。前提として、参照法の結果は常に正しい、とされる。しかし、参照法の多くは特定の固相培地で形成されるコロニーの数を数える方法であって、数%程度のバラツキは避けられない。したがって、参照法の結果と比較する試験法(代替法や迅速法)が同等であると断定することは容易ではない。それでも、他に方法がないので、可能な限りプロトコルをきめ細かく規定してバラツキの少ない、再現性のある結果が得られるようにしているのが実情である。

問題は、そのようにして得られた結果から、同等性を判定するための基準の意味についてである。例えば、最終的には統計学的判断をするので、測定データの数が問題となる。そのため、プロトコルでは同一条件での繰り返し数を規定している。また、バリデーションの最終段階で実施される共同試験(コラボスタディ)では、何ヶ所の試験所で実施すべきかが規定されている。例えば、12ヶ所で5回繰り返しのデータを得た場合と、15ヶ所で4回繰り返しのデータを得た場合を考える。単位データの数だけでいうなら、どちらも60データで同じ条件になる。しかし、バラツキの原因を考えると、試験所間のバラツキと、試験所内でのバラツキは区別して考える。

煩雑ではあるが、このような数は実施段階で考えられることなので容易に揃えることはできる。実際は、どこかの試験所でのデータのバラツキが極端に大きい場合は「おかしい」と判断し、外れ値と称して除外する。その場合の判断基準もやはり統計学的に決められる。その結果、折角、当初は数を揃えておいても、最終的に比較する段階では数が揃わなくなる場合もある。t-検定では、両者の数がそろっていなくても問題はない。むしろ重要なのは有意な差と判断する確率である。多くの場合、95%が採用される。すなわち、両者の結果が95%以上の確率で同等(有意差がある確率は5%以下)と判断されるとき、両者は同等であると結論する。問題はその先である、この95%に何の根拠があるのか、何も説明がないのである。精度の高い化学分析でならともかく、参照法自体のバラツキ(この問題は「測定不確かさ」として別途論じる)が数%もあることを考えれば、95%が果たして妥当な判定基準となりうるか、との疑問が生じる。

このような統計学的判断は、察するに、多くの人が問題であると考えていたようだ。ISO 16140: 2016の中には、そのような判定基準の決め方に、今なお逡巡している様子が見てとれる。

具体的には、定量試験結果の判定基準を規定した箇所である。同等性の判定基準として、今までには無かった β 期待許容区間(β expectation tolerance interval; β -ETI)を採用している点である。参照法の平均値(複数試料について測定した結果の平均値)と代替法の平均値の差がバイアスである。バイアスがゼロであれば、参照法と代替法(は同一の結果を示したことになる。しかし、平均値が同一であっても、統計的なバラツキはあるはずで、そのバラツキを考慮した、統計学的な同

等性を示す必要がある。通常、t-検定で95%以上の確率で有意差がある場合を検定することが多いが、今の有意差がないこと、すなわち同等であることを80%の確率でいえることを判定する。何故、この場合は、95%ではないのか？科学的な理由はない。さらに、80%の確率ではあるが、バイアス値 $\pm\beta$ -ETI(80%)が、 ± 0.5 Log units 以内なら同等であると結論される。さらにその先があって、もし、部分的にでも ± 0.5 Log units を超える場合があったとしても、参照法の標準偏差の4倍以内であれば同等でるとしてよい、となっている。

80%という数値を決めるに当たっては、「同等であるのに、同等ではないと判断してしまう間違い」をできるだけ少なくしたい、という趣旨と理解されるが、そのために益々複雑な統計学的解析をすることに、果たして合理性があるか非常に疑わしい。肝心の「コロニー計数法」の精度を上げるための技術革新の法が春賀に科学的で合理的であると考えている。

(3) 確定試験

定性試験において、結果次第でさらに確認のための念押し試験が要請されるようになった。率直に言って「なぜ？」である。参照法ではネガで代替法ではポジの場合は、擬陽性である。その場合は、さらに別の試験法で確認しろ、という。その逆に、擬陰性の場合は、確認試験は不要という。擬陽性と擬陰性に重み付けをするような発想は全く理解できない。検出することが、試験法の本来の要請であるから、検出できなかった、ということはリスク管理上、問題である。それに比べて、擬陽性はフェイルセーフの観点から問題にはならない、ということか。しかしバリデーションは本来、同等性のみを評価するもので、どちらかが検出感度

が高いとか、検出スペクトルが広いとかを判断するものではない。さらに、この確定試験の必要性の有無が、ペアード試験かアンペアード試験かによって左右されるという。

念を入れるという気持ちは理解できるが、その確定試験法がどのようなものかについては、参照法の確認手順、代替法の確認手順、あるいは両者を混合した手順などの何れかの手順に従って、と記載されている。バリデーション実施者は、ここで確実に戸惑うはずである。参照法と代替法が同一結果になるか否かを調べる規格であるから、異なった値が出ることは当然想定されているはずである。それにも拘らず、敢えて、第3の試験手順、しかもそのプロトコルが任意性のあるもので再確認せよ、としているのは、自己撞着としか言いようがない。しかも、「代替法の擬陰性による陰性偏差、陽性偏差、または陽性一致」、「代替法の擬陽性による陽性偏差、陰性偏差、または陰性一致」の6つのケースが定義されているが、その区別が分かりにくい。実際、発行されたISO 16140: 2016版中でも明らかに誤記と思われるままになっている。

このように複雑で難解な手順が組み込まれるようになった背景を推察するに、同等性を調べた際に、できるだけ合格(同等であるという結論)させたいという思いがあったのではないか。同等といっても、そもそも、参照法自体に相当の不確かさがあるのだから、その不確かさの故に、代替法が、本来、正確な結果を出していたはずなのに、参照法の結果と異なる結果であった、という理由で、不合格になってしまうことが、不合理であると考える者が、少なからずいたということではないか。もし、それが事実であればそのことを明確に書くべきである。規格を読んで、それに従ってバリデーション

を実際に行おうとする人は、ほとんど、そのような背景を理解しているわけではない。その結果、実施できずに当惑するか、逆に適当に解釈して本筋とは異なる手順で行ってしまうか、であろう。我が国で、当に指導的専門家集団である標準法検討委員会で検討しても理解できない、ということが、その証左である。

(4) 定性試験での判定基準：感度

定性試験では、次の3項目を調べる。感度 (Sensitivity)、 相対検出レベル (Relative level of detection; RLOD)、 代替法の包含性 (Inclusivity) と排他性 (Exclusivity)。この中で、感度は、参照法と代替法の結果を比較して、各々が陽性か陰性かを調べることある。この時、両方とも陽性になるか、陰性になれば良いが、異なったときにどうするか、というときに、上記の(3)確定試験で述べたような判断が入る。最終的には、4つの場合、すなわち陽性一致、陰性一致、陽性偏差 (参照法が陰性で、代替法が陽性の場合)、陰性偏差 (参照法が陽性で、代替法が陰性) に分かれ、各々の場合の数から、感度が計算されるが、この計算結果が同じになるのは、陽性偏差と陰性偏差の数が等しいときである (なお、この場合の「数」には、各々、擬陽性による陰性偏差や擬陰性による陽性偏差も含める数であるので、混同しないように「総数」と表記する)。また、いずれの偏差にしても、その数が多ければ、両方法が同等とは言えない。したがって、両方法が同等であると判断できる基準として、陽性偏差と陰性偏差の各々の総数の差が規定数以下、であって、かつ 陽性偏差と陰性偏差の各々の総数の和が規定数以下の場合、となっている。理屈は理解できる。問題は、この規定数の根拠がわからないことである。わからなくても具体的数値がきめられているので、バリデー

ションの実施は機械的にできるようにはなっている。しかし、厄介なことに、陽性偏差と陰性偏差の数自体を決める際に、上記の確認試験の結果が必要になる。例えば、陽性偏差の総数は、「陽性偏差 + 擬陰性による陽性偏差」、陰性偏差の総数は、「陰性偏差 + 擬陽性による陰性偏差」、である。この判断が煩雑で、合理的とも思えず、規格本文で誤記されていたりして、混乱していることは上記の通りである。

(5) 定性試験での判定基準：RLOD

定性試験の第2の調査項目はRLODである。3段階の菌濃度、すなわち無菌、低濃度、高濃度で、参照法と代替法の陽性率を比較する。この場合の低濃度とは、参照法での陽性率が25~75%になるような濃度、というように規定されているが、そのような試料を調製することは意外に難しい。予め、予備実験で検討しておくことは必須である。書いてある手順通りにバリデーションしようとしても、このような条件設定があると、そこで止まってしまうことにもなりかねない。さらに、こうした条件を設定できて結果が得られたとしても、その先の解析法は、規格ではブラックボックスになっている。すなわち EXCEL 上で自動計算するプログラムが ISO から無償でウェブサイト (<http://standards.iso.org/iso/16140>) に公開されている。参照法検体数、代替法検体数、参照法陽性数、代替法陽性数、を入力すると、RLOD、その信頼区間、対数変換後の標準偏差、そして参照法と代替法の検定結果 (p 値) などの統計学的数値が直ちに得られる。一見便利ではあるが、ブラックボックスであることに変わりはない。

これまでの記述からも容易に考えられることであるが、同等性の判断、すなわち合否判定基準の根拠は、必ずしも科学的

合理的に決められたものとは限らない。肝心の統計学的計算の過程を理解することが、専門外の人には難しい、ということがしばしば言われるが、それは正しくない。わかるような表現になっていない、というべきであろう。難解なこととして、ブラックボックス化することは極力避けなければならない。ISO としてはサービスのつもりでも、ユーザーとしては理解の妨げになっていると言わざるを得ない。

(6) 東京会議での WG3 (メソッドバリデーション) の報告・結論

(イ) ISO 16140 では手本とすべき参照法の存在を前提としているが、それが無い場合は、そもそも、そのような参照法自体を開発しなければならない。また既に参照法がある場合でも、その内容を改変したりする場合、どのようなバリデーションをすれば良いのか、という問題提起は以前からなされていた。それに関わる規格が ISO 17468: 2016: Microbiology of the food chain Technical requirements and guidance on establishment or revision of a standardized reference method. である。この規格に対する改訂案がだされた。元々、本文 4 ページに Annex A として図が 1 個 (フローチャート) 付いているだけの簡単な内容であったから、随時、改訂が重ねられていくことは予想された。新しい参照法を開発する場合に対しては、次の 5 段階から成るスキームを提示している。

- ・ステップ 1 : 試験法の選択、
 - ・ステップ 2 : 試験法の評価、
 - ・ステップ 3 : 実際のマトリクスを用いた複数試験所で試験、
 - ・ステップ 4 : ステップ 1 ~ 3 までのデータに基づき、その先の評価を進めるべき試験法と実施する者の選択、
 - ・ステップ 5 : コラボスタディ。
- これに引き続いて、既存の参照法を改定

する場合の留意事項が述べられている。そして、今回、提案された改訂は、この後半の部分に関してであり、具体的な事例を加えるなど、出来るだけ分かりやすくしようとの意図が感じられた。このような議論は、具体的な事例が出されると、議論が発散する恐れがあるとはいえ、我が国としても最も重要な内容であるので、注視していくことが重要である。

(ロ) ISO 16140: 2016 は、今後出版される予定の分冊も含め、6 分冊構成である。第 1 分冊「用語」、第 2 分冊「参照法に対する代替法 (市販キット等) の妥当性確認の手順」が出版済みであり、第 3 分冊「単一試験所で実施される参照法および妥当性確認済代替法の検証手順」、ベリフィケーションの仕方、各試験所で導入する場合に必須、第 4 分冊「単一試験所試験法の妥当性確認手順」、第 5 分冊「市販キット等になっていない試験法のための部分的試験所間妥当性確認手順」、第 6 分冊「微生物確定試験および識別のための代替法の妥当性確認手順」が計画中である。こうした背景により、各規格の作成、あるいは改訂作業、特に、実際に実験によって確認する作業を行ってくれるメンバーを募る、などの提案がなされた。ボランティアとしての協力をどのように行なってゆくべきか、我が国としての課題である。

(7) 東京会議での WG2 (統計学) の報告・結論

測定不確かさの問題が議論された。かつて、この問題について長時間議論した経験とその結果を反芻しながら、改めて、この難問についての議論内容を考察した。測定不確かさの基本概念は、恐らく 1993 年に発行された文書 GUM によって提示されたのが最初であろう。その後、議論、改訂を経て、2008 年 ISO/IEC Guide 98-3

としてまとめられた内容が、当時としての最も明確な表記であると考えられる。その骨子の中の2項目、測定・分析における「真値」の概念を排除し、そこから派生する「かたより(バイアス)」という概念を全て「測定不確かさ」という分散成分に置き換える、その測定不確かさを評価するために、測定・分析のプロセスを明示することを要求する、は基本概念としては理解できるものであった。しかし、例えば、にしたがって測定・分析のプロセスを個々の段階に分解し、その各段階の測定不確かさを推定して積み上げようとしても、見積もりが難しい段階があれば、結局は、全体の積み上げができないことになる。そこで、そのような段階は除外して考えよう、ということが当時のISOの方針であった。それでは厳密な意味で全体の測定不確かさが決められないことになってしまう。しかし第3項、統計的実験に基づかない、「エキスパート・オピニオン」による主観的評価を導入することによって「適当に」判断する、とした。一見、科学的判断を半ば諦めたような内容ともとれるが、元来、複雑な測定・分析のプロセスの全てが科学的に扱うことができる、と考える方が非現実的であり、そのような時こそ十分な経験と知識を持った「エキスパート」の判断が頼りである、という常識的な結論を示しているとも考えられるのである。

平成20～22年度(2008～2010)に実施された厚労科研：食品の安心・安全確保推進研究事業「食品の規格基準に係わる測定値に伴う不確かさに関する研究」(研究代表者：松岡英明)では、理化学試験、生化学試験、微生物試験の全てが対象となっていたが、特に微生物試験に対する議論が集中的に行われた。上記の国際動向調査に関連して、ISO TC69(統計学的方法)/SC6(測定方法と測定値)の国際

および国内委員とも連携して議論が進められたが、最終的には、次のような結論を出した。すなわち、「測定不確かさ」と表記されても、それが意味する内容は同じとは限らない、「測定不確かさ」はその推定の目的によって、具体的方法が異なる、「測定不確かさ」は、その具体的な値を必要とする状況の緊急度に応じて、その趣旨に合った推定法で実施することが必要、「測定不確かさ」を議論する対象となる分析法や試験法は、バリデーションされていること、あるいはバリデーションするに足る方法であることを前提としている、多くの場合、トップダウン法に頼らざるを得ないので、そのために必要となる標準物質、あるいは標準法は不可欠、であるとの結論に至った。そして、自ら、不十分ながら、その生菌の標準物質の開発に取り組み、その原理とプロトタイプの開発に成功して今日に至っている。

以上の結論の中で、にいう「トップダウン法」は各段階の不確かさを積み上げていく「ボトムアップ法」に対比される考え方である。言うまでもなく、当面見積もりが困難な段階については考慮の対象外にしておくとしても、最終的に、科学的な議論のためにはボトムアップ法でなければ成らないであろう。その考えに基づく規格が、ISO/TS 19036:2006 “Microbiology of the food chain—Guidelines for the estimation of measurement uncertainty for quantitative determination” 「フードシステム(生産・加工・流通・保管・販売)の微生物学定量分析における測定不確かさ(Measurement uncertainty; MU)を見積もるためのガイドライン」であり、2006.2.1に第1版が出され、2009.2.1に補正1が出されている。

東京会議でのWG3からの提言は53項目

に及んでいるが、多くは、不確かさの要因をできるだけ詳細に、かつ合理的に分類すること、その上で、個々に不確かさをできるだけ、小さくするための技術的改善を図ること、に関係している。しかし、例えばコロニー計数をイメージング装置で高精度に行うという、不確かさを減じるためには最も有効と思われる方法の導入が全く議論されていない（ように思えるのだが）ことは、2010年当時と全く変わらず、不可解としか言いようがない。イメージング技術は熟練者の何倍ものスピードで、各コロニーの時空間的情報を精確に捉えることができるのだから。

D . 結論

ISO TC34/SC9 の P メンバーとして、規格策定に直接関わっている専門家と議論できるようになったことは、エポックメイキングである。そこで、これまで長年、ISO 16140 を筆頭に既存の規格の邦訳、それに基づく国内向けガイドラインの作成に取り組んできた過程で直面した疑問点を改めて整理した。それを元に、年次総会中に公式、あるいは非公式に議論を試みた。ところが、こちらが疑問に思う論点が、相手にとっては、既に過去のことなのか、議論が噛み合わない。その状況は、我が国にとっての疑問点と年次総会での提案事項の比較表（下記）からも明らかである。しかし、この認識の違いをより明確にして、発信し続けることが、次の段階で我が国からの提言を速やかに賛同してもらうために極めて重要であると思われる。また、提言だけでなく、例えば ISO 16140 の第 3～6 分冊の編纂作業への実質的な貢献についても、P メンバーになったことの責任の一部と考えるべきであろう。こうした認識が得られ、国際的動向の中で我が国が果たすべき責務を明確にすることができたことは、大きな成果である。

なお、生菌標準物質に関する研究に関しては、今年度の研究計画に含まれてはいなかったが、本研究に密接に関わる内容であるので、自主的に実施し、その成果を学会等で発表したことを付記する。

「メソッドバリデーション」及び「統計学」に関して、我が国で議論されてきた課題	ISO TC34/SC9 東京年次総会で、WG3: メソッドバリデーション、及び WG2: 統計学、から提案された事項
測定不確かさ	測定不確かさ
バリデーション実施システム	-----
同等性の判定基準	-----
確定試験	-----
定性試験での判定基準: 感度	-----
定性試験での判定基準: RLOD	-----
参照法のバリデーション	参照法のバリデーション
-----	ISO 16140 の第 3～6 分冊の編纂

E . 健康危害情報

該当なし。

F . 研究発表

（国内学術集会）

齊藤美佳子、松岡英明：大腸菌の死菌が生菌の増殖に及ぼす影響．第 44 回日本防菌防黴学会年次大会、大阪（2017.9.26）．

G . 知的所有権の取得状況

該当なし。

