-AOAC-

STAKEHOLDER PANEL ON DIETARY SUPPLEMENTS –
EXPERT REVIEW PANELS

LUTEIN | TURMERIC

The Scientific Association Dedicated to Analytical Excellence®

THURSDAY, DECEMBER 15, 2016

AOAC INTERNATIONAL HQ
2275 Research Boulevard, Suite 300
Rockville, Maryland, 20850, USA

contact: spds@aoac.org

The Scientific Association Dedicated to Analytical Excellence®

AOAC Stakeholder Panel on Dietary Supplements

EXPERT REVIEW PANEL – LUTEIN and TURMERIC METHODS

Thursday, December 15, 2016
12 :30 pm – 4 :00 pm

A G E N D A

EXPERT REVIEW PANEL CHAIR: Darryl Sullivan, Covance

1. Welcome and Introductions (12:30 p.m. – 12:40 p.m.)
   Darryl Sullivan, Covance (ERP Chair)

2. Review
   a. AOAC Volunteer Policies & ERP Proccess Overview and Guidelines (12:40 p.m. – 1:00 p.m.)
      Deborah McKenzie

3. Review of Methods
   For each method the assigned ERP members will present a review of the revised method manuscripts, after which the ERP will discuss the method and render a decision on the status for each method.

   A. Lutein (1 :00 p.m. – 2 :30 p.m.)
      a. LUT-01
      b. LUT-02

   B. Turmeric (2 :30 p.m. – 4 :00 p.m.)
      a. TUR-01
      b. TUR-02

1. Adjourn (4 :00 p.m.)

AOAC Expert Review Panels

2016 ERP Orientation and Logistics

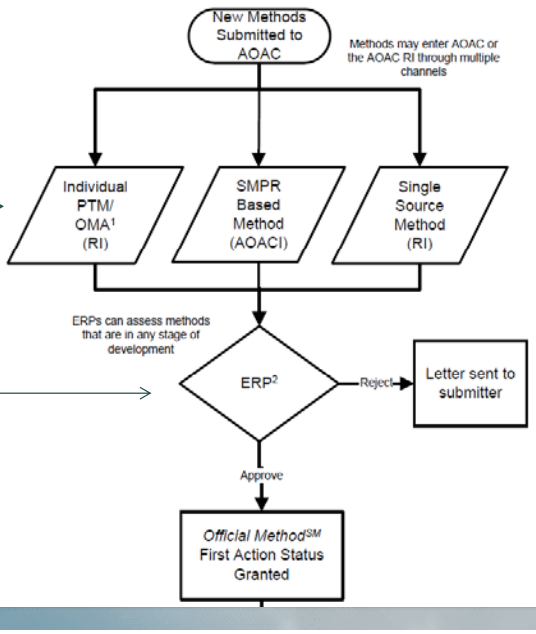
Road to First Action OMA Status

Terms:

- PTM – *Performance Tested Methods℠*
- RI – Research Institute
- ERP – Expert Review Panel
- OMB – Official Methods Board
- SP – Stakeholder Panel
- SMPR – Standard Method Performance Requirement

Three modes of entry and (program administration)

Expert Review Panels will review all methods for all three modes of entry.

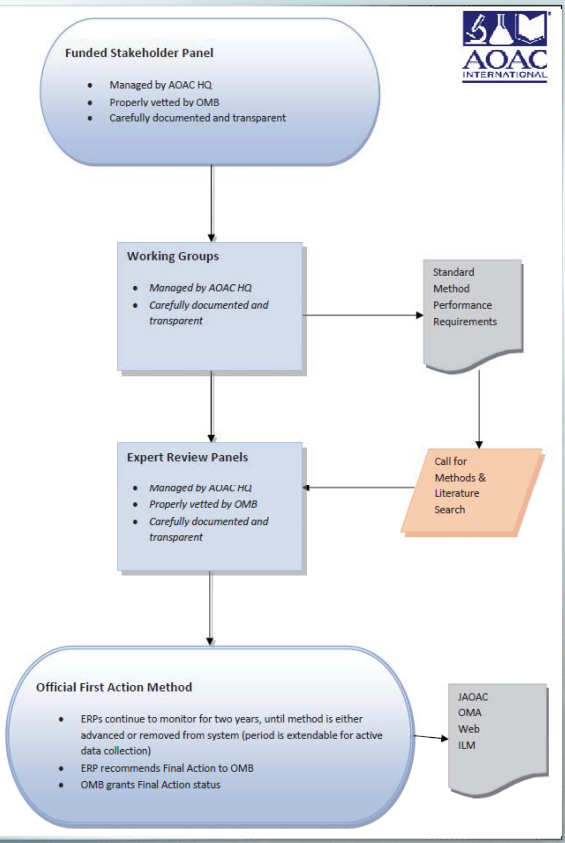


Note: Appeals process always available; see Alternative Pathway Guidelines for appeals process.

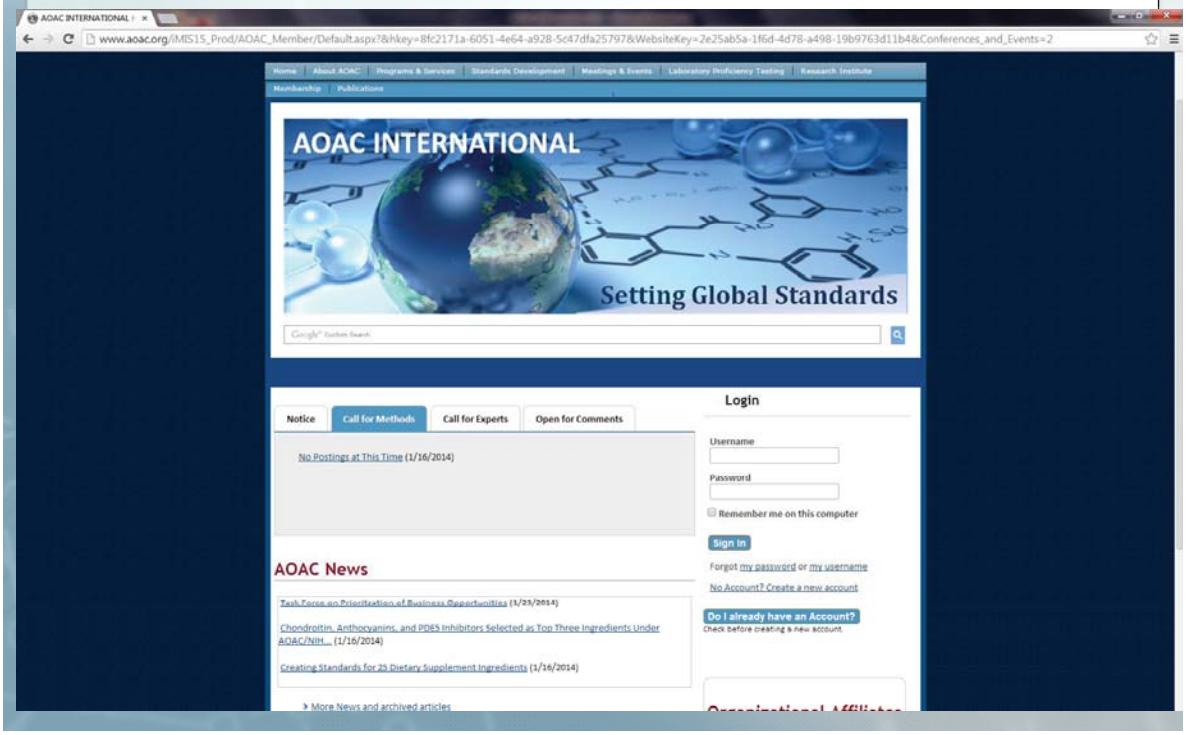
1 PTM certification previously issued, PTM reviewers will be ERP members
2 Unless otherwise provided for under a contractual agreement, AOAC will regularly convene ERPs twice a year: once during the Mid-Year Meeting and again during the Annual Meeting

Recap of the Overall Process for Methods Submitted in response to SMPRs or Call for Methods – aka "alternative pathway"

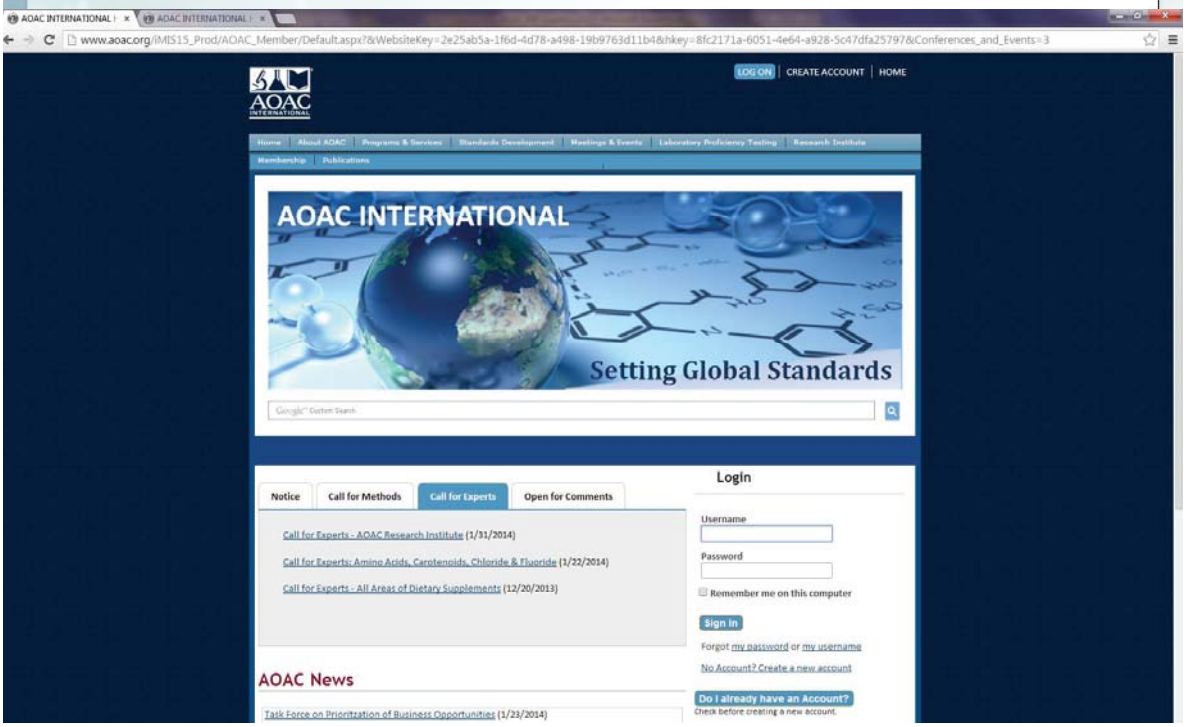
1. Allows AOAC to focus on projects addressing an urgent need of a critical mass of stakeholders.
2. Drives AOAC processes forward faster.
3. Assembles stakeholders (industry, government and academia) to neutral place to articulate and reach consensus on requirements and resolve conflicts.
4. Those requirements are codified and are published as "Standard Method Performance Requirements" (SMPRs).
5. Methods are solicited that purport to meet those requirements.
6. Expert review panels (ERPs) judge the methods against the SMPRs. Method(s) that best meet the SMPRs are adopted and designated "First Action" *Official Method of Analysis*.
7. Process for First Action status to Final Action status follows as the same process for all AOAC First Action *Official Methods*.



Calls for Methods



CALL FOR EXPERTS



Other Forms of Recruitment

- Official Methods Board
- Email Blasts to AOAC network
- Leveraging networks of Advisory Panel members, Working Group Members, AOAC Communities and Sections

REQUIREMENTS FOR ERP SERVICE

- Must have demonstrated expertise in the method, technology, analyte/matrix, etc… Be a subject matter expert.
- Must be able to attend ERP meetings
- Must be able to complete assigned reviews on time
- Must be prepared to speak on the method and share reviews during the meeting
- Must be proactive in tracking assigned First Action *Official Methods*
- Must be able to assist in peer reviewing paper for publication
- Must sign and submit AOAC Volunteer Acceptance Form

AOAC Policies

- AOAC INTERNATIONAL Antitrust Policy
- AOAC INTERNATIONAL Policy On The Use Of The Association Name, Initials, Identifying Insignia, Letterhead, And Business Cards
- AOAC INTERNATIONAL Policy And Procedures On Volunteer Conflict Of Interest
- Volunteer Acceptance Form

Vetting Process

AOAC Chief Science Officer

- Reviews all candidates and supporting documentation for expertise
- Makes a recommendation for an ERP slate

Official Methods Board

- Reviews proposed recommended ERP slate
  - Expertise
  - Balance of panel
  - Conflicts of interest
- Renders decision on proposed ERP members and a Roster is formed.

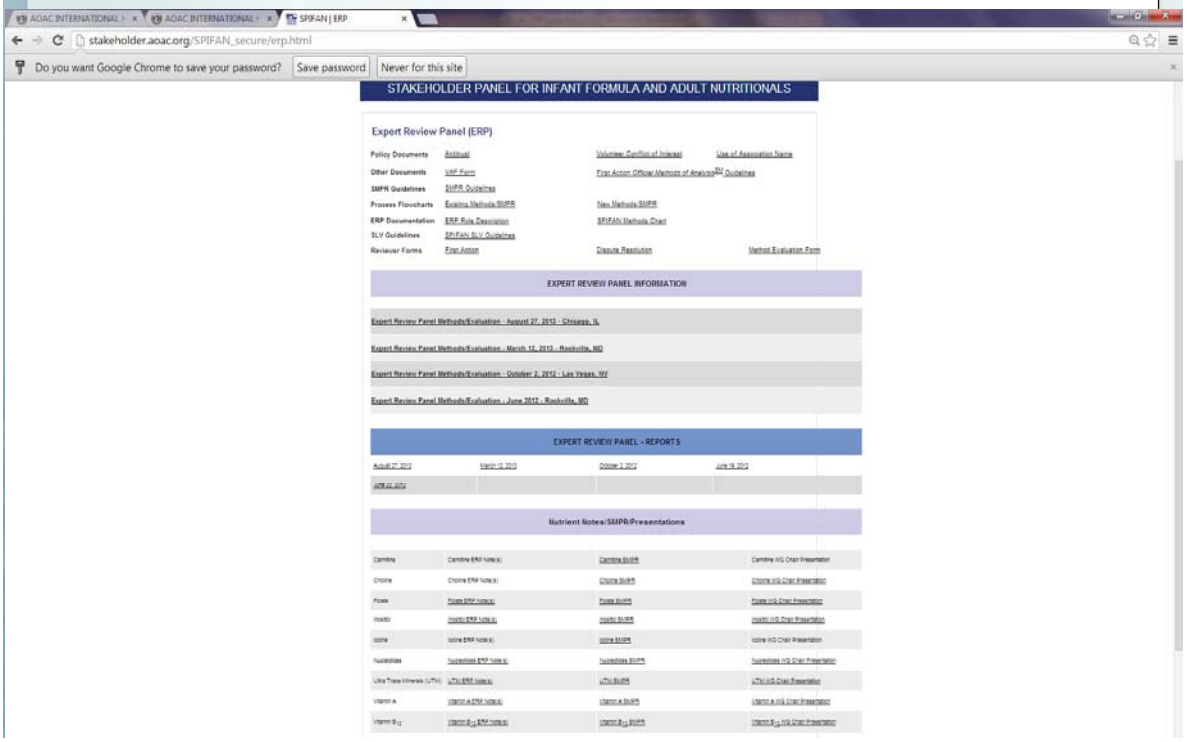
ERP Method Assignments

- A primary and secondary reviewer is assigned to every method.
  - In depth review via review form
  - Prepare to attend and speak on the method and make a recommendation for ERP discussion and consideration.
  - Review forms are completed and returned to AOAC staff in advance of the meeting.
- For Research Institute method submissions:
  - ERP members can participate in the Consulting Service conducting review of protocols – electronically.
- Members of both Committee on Safety and Committee on Statistics serve as advisory resources for all ERPs

ERP REVIEWS

- Primary and Secondary Reviewers or entire ERP (Research Institute ERPs) conduct in-depth review of method and any supporting information.
  - In-depth review is done electronically through password protected website access and is completed prior to the in-person meeting.
  - Deadlines for submission of reviews
  - Depending on the number of methods 15 to 30 days for review
  - Track and present feedback on assigned First Action *Official Methods*.
  - Present on the method during the meeting and can make the motion to adopt the method.
  - Can recommend additional feedback or information for Final Action consideration

ERP REVIEWS



ERP REVIEWS

- **In your judgment, does the method sufficiently meet the Standard Method Performance Requirements (SMPR) or community-based guidance?**

- In your judgment, is the method scientifically sound and can be followed?
- In your judgment, what are the strengths and weaknesses of the method?
- In your judgment, how do the weaknesses weigh in your recommendation for the method?
- In your judgment, will the method serve well the stakeholder community that will use the method?
- In your judgment, what additional information may be needed to further support the method meeting the SMPR or community-based guidance?

ERP Meetings

- ERPs will meet in person at a minimum of twice a year and up to four times per year:
  - AOAC Mid-Year meeting (DC metro area)
  - AOAC Annual Meeting.
  - 2 additional designated times for proprietary method Organziational Affiliates
- At the ERP meeting:
  - Primary and secondary reviewers or entire ERP will present their reviews and makes a motion/recommendation to the ERP whether or not to adopt the method as First Action OMA.
  - ERP discusses the method.
  - ERP renders a decision on First Action status.
  - ERP renders decisions on modifications to First Action methods only.
- If the method is adopted
  - ERP decides on what additional information is needed to recommend the method for Final Action status

ERP MEETINGS

- MEETINGS ARE HELD IN-PERSON, HOSTED BY AOAC
- A QUORUM IS THE PRESENCE OF SEVEN (7) MEMBERS OR 2/3 OF THE TOTAL VETTED ERP, WHICHEVER IS GREATER.

IF NO QUORUM, THEN NO MEETING!

ERP MEETINGS

- REVIEWERS PRESENT THEIR REVIEWS AND MAY INITIATE A MOTION TO ADOPT THE METHOD IF THEY CHOOSE
  - Chair recognizes the reviewers
  - Primary and secondary / ERP reviews are presented.
    - If in favor, they may make and second a motion to adopt or not adopt the method
    - Chair can then entertain discussion on the method
    - Chair can call for a vote once deliberation is complete

ERP MEETING - Discussions

- In your collective judgment, is the method scientifically sound and can be followed as written?
- **In your collective judgment, does the method sufficiently meet the Standard Method Performance Requirements (SMPR)?**
- In your collective judgment, what are the strengths and weaknesses of the method?
- In your collective judgment, do the weaknesses outweigh the strengths in your recommendation for the method?
- In your collective judgment, is the method safe and can it serve well the stakeholder community that will use the it?
- In your collective judgment, is additional information needed to before considering this method for First Action OMA status?

ERP CONSENSUS

- First Action Official Methods status is granted:
- Method must be adopted by unanimous decision of ERP on first ballot, if not unanimous, negative votes must delineate scientific reasons.
- Negative voter(s) can be overridden by 2/3 of voting ERP members after due consideration.
- Method becomes First Action on the date when ERP decision is made.

ERP CONSENSUS

- The ERP may then reach consensus on any additional information that it needs to review to be able to make a recommendation for Final Action *Official Methods* status.
- This is a separate motion.

Post ERP Meeting

- An ERP report with the decisions of the ERP will be drafted
  - Review and approval by ERP chair
  - Posted on website within 15 business days after the ERP meeting
- AOAC staff will send notification to method authors/submitters regarding outcomes on specific methods

ERP Tracking

- Between First Action and Final Action:
  - The primary and secondary reviewers track the methods on behalf of the ERP over this time period.
    - Based on information from method authors, laboratories using the method, general community feedback, additional laboratory work
    - Are ERP recommendations being fulfilled?
    - Is the method meeting the standard criteria more closely?
    - How well is community guidance and OMB guidance being reflected?
  - Updates on the method are given by the primary and secondary reviewers during the ERP meetings.
  - At the end of two years, ERP makes a recommendation to OMB for Final Action status, repeal, or continuance.

Road to Final Action OMA Status

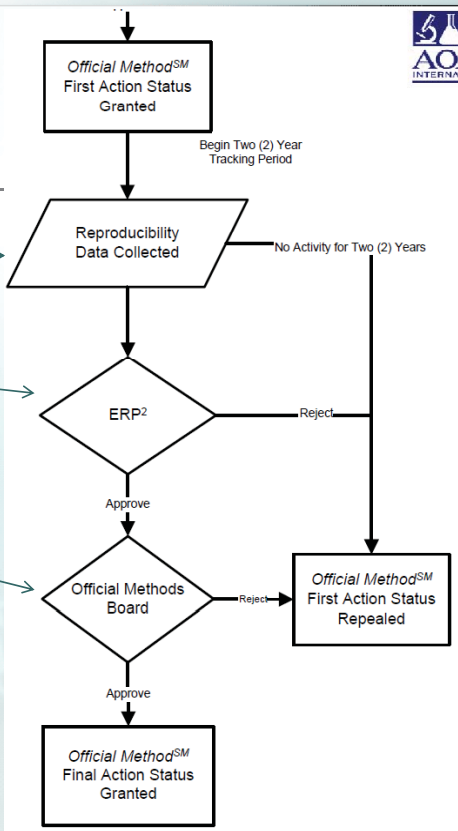
Method reproducibility must be demonstrated before Final Action consideration.

ERP determines if sufficient evidence merits a recommendation for Final Action status or repeal.

•Only the OMB promotes a method to "Final Action" status or repeal the method.

•Methods that did not meet the bar would be repealed.

•Same for all method submissions



Terms:

- PTM – *Performance Tested Methods℠*
- RI – Research Institute
- ERP – Expert Review Panel
- OMB – Official Methods Board
- SP – Stakeholder Panel
- SMPR – Standard Method Performance Requirement

Note: Appeals process always available; see Alternative Pathway Guidelines for appeals process.

1 PTM certification previously issued, PTM reviewers will be ERP members
2 Unless otherwise provided for under a contractual agreement, AOAC will regularly convene ERPs twice a year: once during the Mid-Year Meeting and again during the Annual Meeting

Path to Final Action

Review of ERP Method Recommendations

What to Expect from AOAC Official Method Board (OMB)

Standard Method Performance Pathway

1. Standard Method Performance Requirements authored by Working Groups and established by Stakeholders
2. Expert Review Panel (ERP) vetted by OMB
3. ERP approves methods for First Action
4. Method reproducibility data collected
5. ERP monitors method performance
6. ERP recommendations sent to OMB within 2 years
   - Final Action, First Action continuation, or Repeal

OMB Liaison

- OMB member or designee is assigned to your ERP
- Liaison monitors First Action to Final Action process
- Monitors ERP's documentation of all items in OMB Guidance document (OMA Appendix G)

Method Applicability

- Determine how method meets stakeholder's needs
  - scope, accuracy, precision, etc.
- Are ERP recommendations & improvements implemented?
- Assess method limitations & concerns

Safety Concerns

- Safety review completed for First Action
  - Participation by Safety Committee
- All safety issues identified during 2 year review addressed
  - Participation by Safety Committee

Reference Materials

- Identification of potential reference materials (RM)
  - If none found, define alternative options
- RM performance expectations

Available resource is the AOAC Technical Division on Reference Materials (TDRM)

Single Laboratory Validation

Chemistry

- Linearity
- Accuracy
- Repeatability
- LOD / LOQ
- Matrix scope
- Selectivity

Microbiology

- Inclusivity/Exclusivity
- Robustness
- Repeatability
- POD or equivalent
- Matrix scope

AOAC Committee on Statistics is your resource

Quantitative Reproducibility/Uncertainty

- Experimental designs may vary
  - Collaborative study
  - Proficiency Testing data
  - Multi-lab study variations
- Committee on Statistics
  - is available to discuss new study design protocols
  - Formalized tools were presented at the 2013 Annual Meeting

Qualitative Reproducibility/Uncertainty

- Experimental designs may vary
- Committee on Statistics is available to discuss new study protocols designs

Compare to SMPR

- Method meets Performance Criteria
- Method does not meet Performance Criteria
  - Acceptable or not? List reasoning
- Document acceptability to Stakeholders

Feedback from Users

- Solicit and document user feedback
  - ERP Chair determines mechanism
  - May take form of
    - Proactive calls to users
    - Tally of incoming calls
    - Emails
    - Web surveys

Feedback from Users

- Method performance
- Safety Concerns
  - Warnings
  - Alternatives
- Equipment and supply availability
  - Readily available
  - Practicality
  - Suggested improvements
  - Failures
- Reference material availability

ERP Recommendations

- Supply all documentation to AOAC by established deadline
  - Documentation includes ERP review details
- Representative from ERP present at OMB review meeting
- If method to be repealed, document reasoning

Publication of First Action Methods

- Any approved method(s) along with supporting manuscript(s) and documentation sent to AOAC Publications after the meeting.
  - AOAC Official Methods number assigned.
  - Method and method manuscript prepared for publication in the *Official Methods of Analysis of AOAC INTERNATIONAL* and in Journal of AOAC INTERNATIONAL
  - Updates on methods approved or status changes are published in the Inside Laboratory Management magazine and on the AOAC website

Publication of First Action Methods

NO OMA NUMBER ASSIGNED

UNTIL ALL DOCUMENTATION SUBMITTED

1. Method incorporating ERP revisions (preferably in AOAC Format)
2. Method Manuscript incorporating ERP revisions (in AOAC Format)
3. Signed AOAC Copyright Authorization form

Reports and Documentation

- AOAC staff or designee will capture the decisions and action items into an ERP report.
- The draft report will be sent back to the ERP Chair whose responsibility it is to sign off on the report once approved.
- The report is then distributed to the ERP.
- ERP is responsible for a drafting a written recommendation to the OMB for each method at a maximum of two years following adoption as First Action OMA
- Approved methods from the ERP meetings are published in the OMA and in the *Journal of AOAC INTERNATIONAL*.
- Meeting overviews are published in the AOAC *Inside Laboratory Management* magazine.

Roles and Responsibilities

- Expert Review Panel:
  - Review methods and meet in person to discuss and render decisions on methods for First Action *Official Methods* status.
  - Track First Action *Official Methods*
  - Modify First Action methods if necessary
  - Make recommendations on First Action methods no more than 2 years after adoption to OMB.
- Official Methods Board:
  - Vet and approve ERP membership
  - Assign OMB liaison to be a resource to the ERP
  - Review ERP recommendations and render decisions (*Final Action, Repeal or remain First Action*) on First Action OMAs
- AOAC Staff
  - Coordinate the ERP and meetings, facilitate reviews, document ERP actions/decisions.
  - Issue necessary calls for experts and methods

Expert Review Panels

The ERPs review and approve appropriate methods (as submitted or modified) for adoption as First Action Official Methods or for further validation. ERPs also make recommendations regarding Final Action Official Methods status.

Expert Review Panels

- Must be supported by relevant stakeholders.
- Constituted for the review of methods, not for Standard Method Performance Requirements (SMPR) purposes or as an extension of a Working Group.
- Consist of a minimum of seven (7) members representing a balance of expert stakeholders. **Quorum is a minimum of 7 members present or 2/3 of the total vetted members, whichever is greater.**
- ERP constituency must be approved by the Official Methods Board (OMB).
- Holds transparent public meetings only.
- Remains in force as long as method in First Action Status.

First Action Official Method Status decision

- Must be made by an ERP constituted or reinstated post 2011-03-28 for First Action Official Method Approval (FAOMA).
- Must be made by an ERP vetted for FAOMA purposes by OMB post 2011-03-28.
- Method adopted by ERP must perform adequately against the SMPR set forth by the stakeholders. Or demonstrate performance or characteristics that meet the scope, applicability and/or claims of the method.
- Method must be adopted by unanimous decision of ERP on first ballot, If not unanimous, negative votes must delineate scientific reasons.
- Negative voter(s) can be overridden by 2/3 of non-negative voting ERP members after due consideration
- Method becomes First Action Official Methods on date when ERP decision is made.
- Methods to be drafted into AOAC format by a knowledgeable AOAC staff member or designee in collaboration with the ERP and method author.
- Report of FAOMS decision complete with ERP report regarding decision including scientific background (references etc) to be published concurrently with method in traditional AOAC publication venues.

Method in First Action Status and Transitioning to Final Action Status

- Further data indicative of adequate method reproducibility (between laboratory) performance to be collected. Data may be collected via a collaborative study or by proficiency or other testing data of similar magnitude.
- Two years maximum transition time (additional year(s) if ERP determines a relevant collaborative study or proficiency or other data collection is in progress).
- Method removed from First Action Official Methods and OMA if no evidence of method use available at the end of the transition time.
- Method removed from First Action Official Methods and OMA if no data indicative of adequate method reproducibility is forthcoming as outlined above at the end of the transition time.
- ERP to recommend Method to Official Final Action Status to the OMB.
- OMB decision on First to Final Action Status

Revised October 2013

Online Technical Resources

Method Development, Optimization & Validation

- OMA - Appendix F - Guidelines for Standard Method Performance Requirements
- Homogeneity
- Guide for Writing Methods in AOAC Format
- Statistics Protocol Review Form
- OMA - Appendix D: Guidelines for Collaborative Study Procedures to Validate Characteristics of a Method of Analysis
- OMA - Appendix G: Procedures and Guidelines for the Use of AOAC Voluntary Consensus Standards to Evaluate Characteristics of a Method of Analysis
- OMA - Appendix I: AOAC INTERNATIONAL Methods Committee Guidelines for Validation of Biological Threat Agent Methods and/or Procedures
- OMA - Appendix J: AOAC INTERNATIONAL Methods Committee Guidelines for Validation of Microbiological Methods for Food and Environmental Surfaces
- OMA - Appendix K: Guidelines for Dietary Supplements and Botanicals
- OMA - Appendix L: AOAC Recommended Guidelines for Stakeholder Panel on Infant Formula and Adult Nutritionals (SPIFAN) Single-Laboratory Validation
- OMA - Appendix M - Validation Procedures for Quantitative Food Allergen ELISA Methods: Community Guidance and Best Practices
- Safety Checklist

Method Review

- Examples of Statistical Analysis
- Statistics Manuscript Review Form
- OMA - Appendix A: Standard Solutions and Reference Materials
- OMA - Appendix D: Guidelines for Collaborative Study Procedures to Validate Characteristics of a Method of Analysis
- OMA - Appendix H: Probability of Detection (POD) as a Statistical Model for the Validation of Qualitative Methods

Miscellaneous

- Definition of Terms and Explanatory Notes
- OMA - Appendix B: Laboratory Safety
- OMA - Appendix E: Laboratory Quality Assurance
- OMA - Appendix C: Reference Tables

All resources are accessible at
http://www.aoac.org/vmethguidelines.htm

For questions, please contact:
P 301-924-7077 x157 E dmckenzie@aoac.org

About Expert Review Panels (ERPs)

ERP OVERVIEW:

An Expert Review Panel (ERP) is assembled to review and adopt methods as Official First Action. ERPs will track Official Methods for two years or until such time as reproducibility has been demonstrated and cumulative feedback on method use and performance are obtained. ERPs will make a recommendation regarding Final Action method status for all OMAs to the Official Methods Board (OMB).

All ERP members are expected to serve with the highest integrity and without direct or indirect conflicts of interest. A method assignment can last two years. All members of the ERP are expected to actively participate in ERP meetings and to perform duties and reviews in timely fashion. All members should maintain strict adherence to review timelines and deadlines. AOAC staff documents ERP deliberations.

ESTABLISHING AN EXPERT REVIEW PANEL:

- AOAC staff issues a Call for Experts:
  - Based on voluntary consensus standards and methods submitted to AOAC INTERNATIONAL that may meet the standards.
  - Proprietary and sole source method developers submit individual methods to the AOAC Research Institute.
  - Candidates are asked to submit a CV or information that demonstrates expertise to AOAC staff if not already part of a recognized pool of experts.
- AOAC Chief Scientific Officer (CSO) reviews the documentation for the candidates and make recommends a slate for an expert review panel including the chair to the Official Methods Board.
- The candidate list and supporting documentation are forwarded to the Chair of the OMB who will assign the review to at least two OMB members.
- The OMB reviewers will review the candidates for expertise and perceived conflicts of interest and the OMB may then approve the members of the ERP. A Chair for the ERP is also approved.

EXPERT REVIEW PANEL (ERP):

- Review, discuss and demonstrate consensus on methods for Official First Action method status.
- Participate in the publications process of First Action methods.
- Track and discuss feedback all First Action methods for two years.
- Reach and demonstrate consensus on recommendations for Final Action method status.
- Actively participate in the broader stakeholder effort.

ERP CHAIR:

- Lead ERP discussions in the review and adoption of methods for First Action Official Methods.
- Participate in stakeholder panel activities.
- Review and approve ERP report.
- Work with AOAC staff, working groups and other stakeholder panels to ensure a thorough understanding of the standard method performance requirements and the methods to be assessed.
- Implement the OMB First Action to Final Action Guidelines with the ERP members.
- Advise and review First Action methods and post First Action publications.
- Represent the ERP in presenting the ERPs recommendation to the Official Methods Board regarding Final Action method status.

MECHANICS OF AN AOAC EXPERT REVIEW PANEL

- AOAC CSO assigns methods for review to the expert review panel members.
- For each method, 2 ERP members are assigned as primary and secondary reviewers and present at the ERP meeting.
- All members are expected to actively participate and review methods for First Action Official Method status - conducting thorough and prompt review of methods and being prepared to speak on assigned methods at ERP meetings
- The ERP chair and the 2 reviewers for each method are expected to participate in the publications peer review process for First Action methods.
- ERP reviewers track assigned methods that were adopted as First Action Official Methods and update ERP on method use during two year period between First Action and Final Action
  ERP members are expected to participant in the stakeholder panel activities and/or community at large .

- *ERPs can work with topic advisors (aka, subject matter experts)*
- *OMB can recognize a pool of experts from which ERP members can be selected*

Eligibility Criteria for Expert Reviewers

Be a key expert and/or thought leader of the method or priority under consideration.

- Demonstrated knowledge in the appropriate scientific disciplines.
- Demonstrated knowledge regarding data relevant to adequate method performance.
- Demonstrated knowledge of practical application of analytical methods to bona fide diagnostic requirements.

Be approved by the Official Methods Board

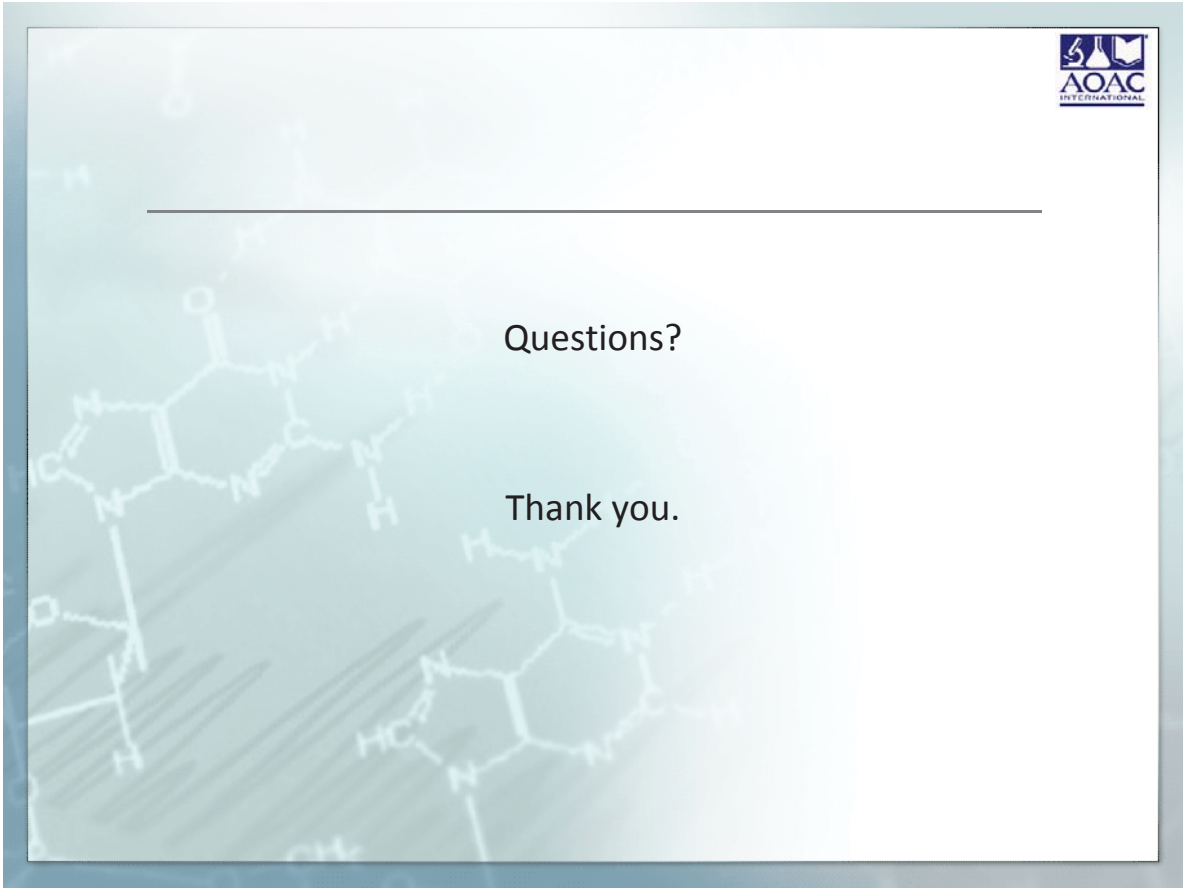
- Qualifications must be clearly described and submitted to AOAC headquarters.

Duties of Expert Reviewers

Members of the Pool of Experts will be called upon to serve on ERPs as needed and to review documents .These documents may include:

- Procedural documents on how methods will be selected and how single laboratory validation studies will be done;
- Methods submitted for consideration as First Action Official Methods;
- Methods submitted for selection for further validation studies;
- Protocols to be used for single laboratory validation studies;
- Selection of methods to be considered for full collaborative studies; and
- Validation study reports

Revised October 2013



LUTEIN EXPERT REVIEW PANEL			
Full Name	**Position**	**Organization**	**Email**
Mr. Darryl M. Sullivan	Chair	Covance Laboratories	darryl.sullivan@covance.com
Dr. Neal E. Craft	Member	Craft Technologies, Inc.	ncraft@crafttechnologies.com
Nour Eddine Es-Safi, Ph.D	Member	Mohammed V University in Rabat	nouressafi@yahoo.fr
AJAI PRAKASH GUPTA	Member	CSIR-IIIM	ajai747@yahoo.co.in
Holly E. Johnson	Member	Alkemist Labs	holly@alkemist.com
Elizabeth Mudge	Member	BCIT	elizabeth_mudge@bcit.ca
Catherine A. Rimmer	Member	NIST	catherine.rimmer@nist.gov
Aniko M. Solyom, Ph.D	Member	GAAS Analytical	asolyom@gaasanalytical.com
Hong You	Member	Eurofins	hongyou@eurofinsus.com
Yang Zhou	Member	Eurofins Scientific Inc.	yangzhou@eurofinsus.com

The Scientific Association Dedicated to Analytical Excellence®

AOAC Stakeholder Panel on Dietary Supplements Expert Review Panel

AOAC Candidate Method #LUT-01

Quantitative measurement of Beta- Cryptoxanthin in Paprika oleoresin, extract and Dietary Supplements (Beadlets).

Submission Information:

- Author(s): Jyotish Srivastava, Jayanthi Chavan, Sachin Salunkhe, Ratna Upadhyay
- Submitted by: Jyotish Srivastava, OmniActive Health Technologies
- Attachments: 1 (Three tabs of an Excel workbook consolidated to PDF for ERP)
- Submitter notes: The methodology is for quantification of Beta cryptoxanthin in Paprika Oleoresin, Extract & dietary supplements (Beadlets)

ERP Reviewers

- Primary Reviewer: Yang Zhou
- Secondary Reviewer: Aniko Solyom

Candidate Method Location

- https://goo.gl/r1mieQ (Must be an ERP member signed into AOAC website to access)

General Information	
Reviewer Name:	Yang Zhou
Email:	yangzhou@eurofinsus.com
Organization:	Eurofins
Method Reviewed:	LUT-01
Method Title:	QUANTITATIVE MEASURMENT OF BETA-CRYPTOXANTHIN IN PAPRIKA OLEORESIN, EXTRACT AND DIETARY SUPPLEMENTS (BEADLETS)
Applicable SMPR	2016.004
I. Summary	
Summary of Method:	Based on the sample matrix (paprika oleoresin, extract, or dietary supplement) and detection instrumentation (HPLC or UV), samples are extracted with corresponding extraction solvent. Saponification is performed for paprika oleoresin. The extraction solvent is then quantified by HPLC or UV reading.
II. Review of the Method Only	
1. Does the Applicability of the Method Support the Applicability of the SMPR? If not, please explain what is missing.	No. This method didn't demonstrate its capability to separate and quantify beta-cryptoxanthin isomers, lutein and zeaxanthin.
2. Does the analytical technique(s) used in the method meet the SMPR? If not, please specify how it differs from what is stated in the SMPR.	Yes
3. Are the definitions specified in the SMPR used and applied appropriatly in the method? If no, please indicate how the terms are used.	No. In this method, it specified dietary supplements in beadlets form.
4. Does the method, as written, contain all appropriate precautions and warnings related to the method's regaents, components, instrumentation, or method steps that may be hazardous? If no, please suggest wording or option(s).	No. Suggested wording: In this method, various organic solvents (hexanes, chloroform, etc.,) along with several other hazardous chemicals are used and special care should be taken when handling or disposing of these chemicals. Wear gloves, safety glasses and use in a fume hood. Follow all laboratory safety precautions. Refer to Safety Data Sheets (SDS) for any questions relating to any chemicals in use.
III. Review of Information in Support of the Method	
1. Are the definitions specified in the SMPR used and applied appropriately in the supporting documentation (manuscripts, method studies, etc…)? If not, please explain the differences and if the method is impacted by the difference.	No. Reproducibility was replaced by word "ruggedness", which is a completely different property of a method.

2. Is there information demonstrating that the method meets the SMPR Method Performance Requirements using the Reference Materials stated in the SMPR? If not, then specify what is missing and how this impacts demonstration of performance of the method.	No. No reference material was used in the validation.
3. Is there information demonstrating that the method performs within the SMPR Method Preformance Requiements table specifications for all analytes in the SMPR applicability statement? If not, please specify what is missing and whether or not the method's applicaiblity should be modified.	No. Only total beta-cryptoxanthin was reported.
IV. General Submission Package	
1. Based on the supporting information, were there any additional steps in the evaluation of the method that indicated the need for any addional precautionary statements in the method?	No.
2. Does the method contain system suitability tests or controls as specified by the SMPR? If not, please indicate if there is a need for such tests or controls, and which ones.	No. System suitability is needed for beta-cryptoxanthin in all matrices. It can be performed with beta-cryptoxanthin standard.
3. Is there information demonstrating that the method system suitability tests and controls as specified in the SMPR worked appropriately and as expected? If no, please specify.	No. No system suitability was performed in validation.
4. Based on the supporting information, is the method written clearly and concisely? If no, please specify the needed revisions.	Yes
5. Based on the supporting information, what are the pros/strenghts of the method?	Procedure is well described for each matrix. Extraction and quantification of beta-cryptoxanthin in dietary supplement is straight and simple. Peaks are well separated in chromatogram.
6. Based on the supporting information, what are the cons /weaknesses of the method?	Not sure if this method is capable to quantify other analytes in the SMPR such as lutein and zeaxanthin. Not sure if this method is capable to separate the cis and trans isomers of beta-cryptoxanthin. No criteria are defined for linearity requirement and system suitability. LOQ is missing.

7 . Any general comments about the method?	The method is straightforward and clear for all three matrices. For dietary supplement, the procedure is simple. Peaks are well separated in chromatogram. This method is lack of separation of beta-cryptoxanthin isomers and identification of lutein and zeaxanthin. Beta-cryptoxanthin standard is included in the method but the purpose isn't clear.
V. Recommendation for the Method	
Do you recommend this method be adopted as a First Action and published in the Official Methods of Analysis of AOAC INTERNATIONAL? Please specify rationale.	No. This method doesn't demonstrate the capability to separate beta-cryptoxanthin isomers, as well as other carotnoids such as lutein and zeaxznthin. Thus it doesn't meet the requirement of SMPR. In additional, the single laboratory validation result was lack of certain data such as reference material, system suitability, LOQ etc.

General Information	
Reviewer Name:	Aniko Solyom
Email:	asolyom@gaasanalytical.com
Organization:	GAAS Analytical
Method Reviewed:	LUT-01
Method Title:	Quantitative measurement of Beta- Cryptoxanthin in Paprika oleoresin, extract and dietary supplements
Applicable SMPR	2016.004
I. Summary	
Summary of Method:	The method quantifies the amount of beta-cryptoxanthin and total xantophyllis in paprika oleoresin, extracts and dietary supplement beadlets.
II. Review of the Method Only	
1. Does the Applicability of the Method Support the Applicability of the SMPR? If not, please explain what is missing.	Partly. (i) According to the SMPR the method should be able to separately quantify cis and trans isomers of each analytes - the method quantifies beta-cryptoxanthin, without cis/trans distinction. (I believe the SMPR is a bit too ambitious with this requirement.) (ii) The method quantifies only beta-cryptoxanthin but not the rest of the analytes (lutein and zeaxanthin), but it can be still acceptable.
2. Does the analytical technique(s) used in the method meet the SMPR? If not, please specify how it differs from what is stated in the SMPR.	Yes
3. Are the definitions specified in the SMPR used and applied appropriatly in the method? If no, please indicate how the terms are used.	Yes
4. Does the method, as written, contain all appropriate precautions and warnings related to the method's regaents, components, instrumentation, or method steps that may be hazardous? If no, please suggest wording or option(s).	There is no separate safety section in the description, but the chemicals and instruments are commonly used in a routine analytical laboratory.
III. Review of Information in Support of the Method	
1. Are the definitions specified in the SMPR used and applied appropriately in the supporting documentation (manuscripts, method studies, etc…)? If not, please explain the differences and if the method is impacted by the difference.	Yes, except the analytical range.

2. Is there information demonstrating that the method meets the SMPR Method Performance Requirements using the Reference Materials stated in the SMPR? If not, then specify what is missing and how this impacts demonstration of performance of the method.	None of the reference materials stated in the SMPR were used. Beta-cryptoxanthin from Sigma was used to make the standard solution(s), but there is no indication of the purity of the standard. On page 4 there is a chromatogram that supposedly the chromatogram of the beta-cryptoxanthin standard solution that shows a very inferior standard purity.
3. Is there information demonstrating that the method performs within the SMPR Method Preformance Requiements table specifications for all analytes in the SMPR applicability statement? If not, please specify what is missing and whether or not the method's applicaiblity should be modified.	Some, but not enough. (i) Analytical range: Does not specify exactly: "on the basis of linearity", but the linear range is 4-20 ppm, much narrower than the required 0.0005-100% (ii) Limit of quantitation: Meets the requirement, but not quite clear how it was calculated. (Six injections of 1 ppm solution, when the lowest concentration of the linear range is 4 ppm...) (iii) Recovery and Repeatibility: meets requirement for the two higher ranges. No data was provided for the two lower ranges.
IV. General Submission Package	
1. Based on the supporting information, were there any additional steps in the evaluation of the method that indicated the need for any addional precautionary statements in the method?	No
2. Does the method contain system suitability tests or controls as specified by the SMPR? If not, please indicate if there is a need for such tests or controls, and which ones.	No system suitability test, although there is a description in the method for the preparation of blanks.
3. Is there information demonstrating that the method system suitability tests and controls as specified in the SMPR worked appropriately and as expected? If no, please specify.	NA
4. Based on the supporting information, is the method written clearly and concisely? If no, please specify the needed revisions.	More details and clarifications are needed. Is the chromatogram on page is the chromatogram of the standard? What is the purity of the standard? How the purity was determined? (They are usually not stable) What were the storage conditions? What kind of filters were used? Would be helpful to see the chromatograms of the samples.
5. Based on the supporting information, what are the pros/strenghts of the method?	Relatively simple method.
6. Based on the supporting information, what are the cons /weaknesses of the method?	(i) Not quite clear what is the analytical range (ii) There are data for only two matrices, extracts and beadlets (iii) No data for dietary ingredients (iv) Without seeing actual chromatograms, it is hard to judge how good is the separation. (v) Only one analyte, no cis/trans separation (minor issue)

7 . Any general comments about the method?	It has promise, but more data needed
V. Recommendation for the Method	
Do you recommend this method be adopted as a First Action and published in the Official Methods of Analysis of AOAC INTERNATIONAL? Please specify rationale.	It is promising, but I would like to see more supporting data and chromatograms.

The Scientific Association Dedicated to Analytical Excellence®

AOAC Stakeholder Panel on Dietary Supplements Expert Review Panel

AOAC Candidate Method #LUT-02

Determination of Lutein and Zeaxanthin Esters and Their Geometric Isomers in Carotenoid Ester Concentrates Used as Ingredients in Nutritional Supplements: Validation of a Combined Spectrophotometric-HPLC Method

- Author(s): Wagner O. Lombeida, Fernando Rubio, and Luis W. Levy
- Submitted by: Luis W. Levy, Inexa CA
- Attachments: 1 – Applications in various matrices

ERP Reviewers

- Primary Reviewer: Neal Craft
- Secondary Reviewer: Hong You

Candidate Method Location

- https://goo.gl/r1mieQ (Must be an ERP member signed into AOAC website to access)

General Information	
Reviewer Name:	Neal Craft
Email:	ncraft@crafttechnologies.com
Organization:	Craft Technologies
Method Reviewed:	LUT-02
Method Title:	Determination of Lutein and Zeaxanthin Esters and Their Geometric Isomers in Carotenoid Ester Concentrates Used as Ingredients in Nutritional Supplements: Validation of a Combined Spectrophotometric-HPLC Method
Applicable SMPR	2016.004
Information about the Method Only	
Summary of Method:	The method uses a combination of spectrophotomety to measure total carotenoids and normal-phase HPLC on a saponified sample to estimate the percentage of free lutein and zeaxanthin content in carotenoid ester concentrates, including their main geometrical isomers. The unique part of this method is the estimate of a composite-specific absorbance due to the differing absorbances of the geometric isomer distribution. The method is reportedly applicable to carotenoid ester concentrates in oil suspensions and dosage forms. The sample is first dissolved in hexane–2-propanol (95+5, v/v) for spectrophotometric measurement at the maximum absorption wavelength of ~445 nm. Subsequently, a sample is saponified then separated using normal-phase HPLC to determine the relative percentage of the main geometric isomers of lutein and zeaxanthin.
Method Scope / Applicability	The method is reportedly applicable to carotenoid ester concentrates in oil suspensions and dosage forms.
General Comments About the Method	The method principles are sound. Normal-phase chromatography is an excellent method to separate xanthophylls and geometric isomers. Spectrophotometry has been used for decades to estimate carotenoid content. The specific addition in this method is to account for geometric isomer distribution of the samples and adjust the calculation of content for the specific isomer content. This should result in a more accurate estimate of xanthophyll esters. It makes some the assumption that saponification does not alter the original isomer distribution of the esters.
Method Clarity	The execution of the method is adequately clear. The calculation is more complicated and makes some assumptions that may not be accurate or substantiated.
Pros / Strengths	Normal-phase HPLC is a good method to separate the xanthophylls. Spectrophotometry is a simple straight-forward measurement of total carotenoids. There is an effort to more accurately assess the xanthophyll content.
Cons / Weaknesses	There are assumptions made that could bias the results. Total carotenoids includes everything that absorbs at 445nm. This could include hydrocarbon carotenes and more polar xanthophylls which may not be accounted for by the HPLC. It assumes that the saponification does not alter the isomer distribution which is incorrect. The calculation of composite-specific absorbance is theoretically sound but may not be fully substantiated with foundational data. It assumes a single fatty acid ester and uses E1% values that are not well documented or generally accepted. The HPLC conditions are not current technology. Use of neat lutein and zeaxanthin standards would be beneficial. Method is limited to lutein and zeaxanthin ester products. It does not include beta-cryptoxanthin or beadlet products.
Review of Information in Support of the Method	
General Comments about Supporting Data	There is a substantial amount of precision , linearity and selectivity data from a single lab.
Method Optimization	There is no discussion of HPLC or saponification optimization.

Performance Characteristics	Analytical Range: 0.2 to 4.2 mg/L in measured solutions but could not be determined per AOAC. LOQ: 0.2 mg/L Accuracy / Recovery: Recovery was 97%. Could not be ascertained in the absence of cis-isomer standards. Precision: RDS was 0.23%. Reproducibility: RSD <0.92%
System Suitability	Conditions not provided to allow suitability to be replicated. —All instrument configuration; instrument details, including serial numbers and acquisition information; and column details, including identification, were properly recorded during laboratory operations. Equipment calibration, including signal description and baseline noise determination, was performed before tests were executed.
Recommendation for the Method	
Do you recommend this method be adopted as a First Action and published in the Official Methods of Analysis of AOAC INTERNATIONAL? Please specify rationale.	This is the basis of a functional method. Several things could improve and expand the method. It may be more applicable to measuring free lutein and zeaxanthin than the esters. I would recommend that some things be adjusted before recommending first action status. Sample prep to address additional products. Assessment of the composite-specific absorbance. Inclusion of standards and beta-cryptoxanthin.

General Information	
Reviewer Name:	Hong You
Email:	hongyou@eurofinsus.com
Organization:	
Method Reviewed:	LUT-02
Method Title:	Determination of Lutein and Zeaxanthin Esters and Their Geometric Isomers in Carotenoid Ester Concentrates Used as Ingredients in Nutritional Supplements: Validation of a Combined Spectrophotometric-HPLC Method
Applicable SMPR	2016.004
I. Summary	
Summary of Method:	The combined spectrophotometric-LC method was published in August 2016 at J AOAC and submitted to the ERP for SPDS Set 4 Ingredients in response to the call for methods. A three-step procedure is used, involving (1) an HPLC chromatogram of the saponified sample to calculate the composite-specific absorbance value based on its carotenoid profile and documented or estimated extinction coefficients (1%) of predominant geometrical isomers of lutein and zeaxanthin esters, (2) measurement of the optical absorbance of the unsaponified sample and calculation of its specific absorbance at 445 nm, and (3) calculation of the total carotenoid ester content of the sample. No calibration standards are needed for this protocol.
II. Review of the Method Only	
1. Does the Applicability of the Method Support the Applicability of the SMPR? If not, please explain what is missing.	This method is able to separately determine the fatty acid esters of the cis and trans isomers of lutein and zeaxanthin in ingredients and dietary supplements. However, it cannot distinguish the esterified and nonesterified forms of these xanthophylls. In addition, this method did not document any information about beta-cryptoxanthin.
2. Does the analytical technique(s) used in the method meet the SMPR? If not, please specify how it differs from what is stated in the SMPR.	Yes, the analytical techniques used in the method meet the SMPR requirements.
3. Are the definitions specified in the SMPR used and applied appropriatly in the method? If no, please indicate how the terms are used.	Authors used “Precision” as the term to document both “Repeatability” and “Reproducibility” procedures. Authors established “Recovery” by comparing the analytical results of commercial samples with values from suppliers’ Certificate of Analysis, while the SMPR requires comparing the analytical results from spiked sample with their theoretical value.
4. Does the method, as written, contain all appropriate precautions and warnings related to the method's regaents, components, instrumentation, or method steps that may be hazardous? If no, please suggest wording or option(s).	Method's reagents, components, instrumentation, or method steps are generally safe.
III. Review of Information in Support of the Method	
1. Are the definitions specified in the SMPR used and applied appropriately in the supporting documentation (manuscripts, method studies, etc…)? If not, please explain the differences and if the method is impacted by the difference.	Yes.

2. Is there information demonstrating that the method meets the SMPR Method Performance Requirements using the Reference Materials stated in the SMPR? If not, then specify what is missing and how this impacts demonstration of performance of the method.	Authors did not use the reference materials recommended by the SMPR.
3. Is there information demonstrating that the method performs within the SMPR Method Preformance Requiements table specifications for all analytes in the SMPR applicability statement? If not, please specify what is missing and whether or not the method's applicaiblity should be modified.	Analytical range: Submitted method has its analytical range between approximately 0.1 to 100%. The SMPR requires 0.0005 to 100%. LOQ: Method did not document LOQ. Recovery: For samples that have their range >1%, the submitted method has 97% as its recovery, while the SMPR requires recovery as 98-102%. Note: the definition of recovery is different between the submitted method and SMPR. Authors established "Recovery" by comparing the analytical results of commercial samples with values from suppliers' Certificate of Analysis, while the SMPR requires comparing the analytical results from spiked sample with their theoretical value. Repeatbility: Submitted method has its repeatability RSDr
IV. General Submission Package	
1. Based on the supporting information, were there any additional steps in the evaluation of the method that indicated the need for any addional precautionary statements in the method?	Yes. Authors may need to admit that the extinction coefficients in table 1 are not precisely applicable to this method because they were established at their λmax not 445 nm.
2. Does the method contain system suitability tests or controls as specified by the SMPR? If not, please indicate if there is a need for such tests or controls, and which ones.	Yes. The method contains system suitability tests.
3. Is there information demonstrating that the method system suitability tests and controls as specified in the SMPR worked appropriately and as expected? If no, please specify.	Yes.
4. Based on the supporting information, is the method written clearly and concisely? If no, please specify the needed revisions.	Yes.
5. Based on the supporting information, what are the pros/strenghts of the method?	The method is easy to conduct and does not require calibration standards for the routine quantification. The method has high accuracy for the determination of total xanthophyll ester contents. Also, the method is robust when cis-isomers of lutein and zeaxanthin are present in a significant amount.
6. Based on the supporting information, what are the cons /weaknesses of the method?	The method cannot distinguish the esterified and nonesterified forms of xanthophylls. In addition, this method did not document any information about beta-cryptoxanthin.

7 . Any general comments about the method?	Based on the above comments, I do not recommend this method to be considered for further steps.
V. Recommendation for the Method	
Do you recommend this method be adopted as a First Action and published in the Official Methods of Analysis of AOAC INTERNATIONAL? Please specify rationale.	The method cannot distinguish the esterified and nonesterified forms of xanthophylls. In addition, this method did not document any information about beta-cryptoxanthin. Therefore, I do not recommend this method to be considered for a First Action.

TURMERIC EXPERT REVIEW PANEL			
Full Name	**Position**	**Organization**	**Email**
Mr. Darryl M. Sullivan	Chair	Covance Laboratories	darryl.sullivan@covance.com
Gerardo Albornoz	Member	McCormick & Company, Inc.	Gerardo_Albornoz@mccormick.com
Dr. Neal E. Craft	Member	Craft Technologies, Inc.	ncraft@crafttechnologies.com
Nour Eddine Es-Safi, Ph.D	Member	Mohammed V University in Rabat	nouressafi@yahoo.fr
AJAI PRAKASH GUPTA	Member	CSIR-IIIM	ajai747@yahoo.co.in
Holly E. Johnson	Member	Alkemist Labs	holly@alkemist.com
Elizabeth Mudge	Member	BCIT	elizabeth_mudge@bcit.ca
Melissa Meaney Phillips	Member	NIST	melissa.phillips@nist.gov
Catherine A. Rimmer	Member	NIST	catherine.rimmer@nist.gov
Aniko M. Solyom, Ph.D	Member	GAAS Analytical	asolyom@gaasanalytical.com

The Scientific Association Dedicated to Analytical Excellence®

AOAC Stakeholder Panel on Dietary Supplements Expert Review Panel

AOAC Candidate Method #TUR-01

Curcuminoids in Turmeric Roots and Supplements: Method Optimization and Validation

- Author(s): Elizabeth Mudge, Michael Chan, Sylesh Venkataraman, Paula Brown
- Submitted by: Elizabeth Mudge
- Attachments: 1
- Submitter notes: None

ERP Reviewers

- Primary Reviewer: Melissa Phillips
- Secondary Reviewer: Holly Johnson

Candidate Method Location:

- https://goo.gl/r1mieQ (Must be an ERP member signed into AOAC website to access)

General Information	
Reviewer Name:	Melissa Phillips
Email:	melissa.phillips@nist.gov
Organization:	NIST
Method Reviewed:	TUR-01
Method Title:	Curcuminoids in Turmeric Roots and Supplements: Method Optimization and Validation
Applicable SMPR	2016.003
I. Summary	
Summary of Method:	A methanol extraction followed by LC-absorbance for determination of CUR, BDMC, and DMC in turmeric materials. A full method validation is described including optimization of multiple experimental conditions.
II. Review of the Method Only	
1. Does the Applicability of the Method Support the Applicability of the SMPR? If not, please explain what is missing.	Yes
2. Does the analytical technique(s) used in the method meet the SMPR? If not, please specify how it differs from what is stated in the SMPR.	Yes
3. Are the definitions specified in the SMPR used and applied appropriatly in the method? If no, please indicate how the terms are used.	Yes
4. Does the method, as written, contain all appropriate precautions and warnings related to the method's regaents, components, instrumentation, or method steps that may be hazardous? If no, please suggest wording or option(s).	No safety information is included. Could potentially include a statement regarding flammability of solvents, acidity of mobile phase, as well as use of elevated column temperature, but this is not critical.
III. Review of Information in Support of the Method	
1. Are the definitions specified in the SMPR used and applied appropriately in the supporting documentation (manuscripts, method studies, etc…)? If not, please explain the differences and if the method is impacted by the difference.	Yes

2. Is there information demonstrating that the method meets the SMPR Method Performance Requirements using the Reference Materials stated in the SMPR? If not, then specify what is missing and how this impacts demonstration of performance of the method.	Reference standards were acquired from a different vendor (Chromadex) than those listed in the SMPR; however, purity of standards was evaluated by qNMR and is therefore sufficient. Matrix reference materials listed in SMPR were not used in this study (although they are not yet available from NIST).
3. Is there information demonstrating that the method performs within the SMPR Method Preformance Requiements table specifications for all analytes in the SMPR applicability statement? If not, please specify what is missing and whether or not the method's applicaiblity should be modified.	LOQs reported: 0.3 mg/g (0.03%) for BDMC and DMC 1.6 mg/g (0.16%) for CUR SMPR states 0.1%; met for BDMC and DMC, slightly high for CUR Recovery reported: 96.6-103.3% SMPR states 95-110%; met Analytical ranges reported: 0.097-0.943% for BDMC (1-120 ug/mL cal range) 0.056-8.096% for DMC (1-100 ug/mL cal range) 0.377-88.21% for CUR (5-300 ug/mL cal range) SMPR requests 0.1% to >50%; based on levels in products and the calibration ranges demonstrated, I have no concern about this method working in this range. RSDrs reported in supporting information are within range in SMPR (LT 5% at 0.1-50%; LT 3% at >50%) with one exception. BDMC at 0.373% in a tablet has RSDr of 5.5%.
IV. General Submission Package	
1. Based on the supporting information, were there any additional steps in the evaluation of the method that indicated the need for any addional precautionary statements in the method?	No.
2. Does the method contain system suitability tests or controls as specified by the SMPR? If not, please indicate if there is a need for such tests or controls, and which ones.	Yes (precision <5% for replicate samples throughout the run)
3. Is there information demonstrating that the method system suitability tests and controls as specified in the SMPR worked appropriately and as expected? If no, please specify.	Yes (table 2 of supplemental information)
4. Based on the supporting information, is the method written clearly and concisely? If no, please specify the needed revisions.	Yes; will likely need to be reformatted from existing publication and supplemental information into a single document for ease of reader

5. Based on the supporting information, what are the pros/strenghts of the method?	The method is straightforward and relatively simple. It has been tested on a number of matrices (as suggested in Table 3 of SMPR) and in the presence of numerous other botanicals (including 2 from Table 2 of SMPR). The validation is solid and the method performance meets nearly every aspect of the SMPR.
6. Based on the supporting information, what are the cons /weaknesses of the method?	Would prefer an internal standard approach to the external standard approach presented here. LC method has a very minor gradient step (28-30%); is this necessary? Seems that an isocratic method with a wash step might be more robust. Not all dietary ingredients from Table 2 in SMPR were included in testing (e.g., cayenne, carotenoids, Ca2+). Would like to see demonstration of method on products including these to meet SMPR.
7 . Any general comments about the method?	Very strong method, well validated!
V. Recommendation for the Method	
Do you recommend this method be adopted as a First Action and published in the Official Methods of Analysis of AOAC INTERNATIONAL? Please specify rationale.	Yes, I would support this method for First Action status, but would recommend additional work on investigation of remaining dietary ingredients listed in SMPR before MLT (and also include a mixed product containing these in the MLT).

General Information	
Reviewer Name:	Holly Johnson
Email:	holly@alkemist.com
Organization:	Alkemist Labs
Method Reviewed:	TUR-01
Method Title:	Curcuminoids in Turmeric Roots and Supplements: Method Optimization and Validation
Applicable SMPR	2016.003
I. Summary	
Summary of Method:	Presented is a validated HPLC-UV method for quantitation of individual curcuminoids in turmeric raw materials and finished products. Method parameters were modified in response to a previous ERP; extraction and separation have been optimized, with extraction solvent and column temp being significant factors. Factorial studies were used to guide optimization and a Single Lab Validation was performed on 12 materials with the modified method.
II. Review of the Method Only	
1. Does the Applicability of the Method Support the Applicability of the SMPR? If not, please explain what is missing.	Yes
2. Does the analytical technique(s) used in the method meet the SMPR? If not, please specify how it differs from what is stated in the SMPR.	Yes
3. Are the definitions specified in the SMPR used and applied appropriatly in the method? If no, please indicate how the terms are used.	Yes
4. Does the method, as written, contain all appropriate precautions and warnings related to the method's regaents, components, instrumentation, or method steps that may be hazardous? If no, please suggest wording or option(s).	N/A
III. Review of Information in Support of the Method	
1. Are the definitions specified in the SMPR used and applied appropriately in the supporting documentation (manuscripts, method studies, etc…)? If not, please explain the differences and if the method is impacted by the difference.	Yes

2. Is there information demonstrating that the method meets the SMPR Method Performance Requirements using the Reference Materials stated in the SMPR? If not, then specify what is missing and how this impacts demonstration of performance of the method.	The SMPR states 4 USP chemical references and 2 NIST reference materials. The method uses 3 Chromadex chemical ref standards and turmeric study materials from AHP and other sources.
3. Is there information demonstrating that the method performs within the SMPR Method Preformance Requiements table specifications for all analytes in the SMPR applicability statement? If not, please specify what is missing and whether or not the method's applicaiblity should be modified.	Yes for LOQ, accuracy/recovery, repeatability, and intermediate precision.
IV. General Submission Package	
1. Based on the supporting information, were there any additional steps in the evaluation of the method that indicated the need for any addional precautionary statements in the method?	No
2. Does the method contain system suitability tests or controls as specified by the SMPR? If not, please indicate if there is a need for such tests or controls, and which ones.	Yes
3. Is there information demonstrating that the method system suitability tests and controls as specified in the SMPR worked appropriately and as expected? If no, please specify.	Yes
4. Based on the supporting information, is the method written clearly and concisely? If no, please specify the needed revisions.	Yes
5. Based on the supporting information, what are the pros/strenghts of the method?	Validation data indicate good specificity, precision, and accuracy in a variety of materials.
6. Based on the supporting information, what are the cons /weaknesses of the method?	High concentration materials that are common in commerce were not evaluated, >90% curcuminoids.
7 . Any general comments about the method?	A simple rapid method with excellent performance in the matrices evaluated.
V. Recommendation for the Method	

Do you recommend this method be adopted as a First Action and published in the Official Methods of Analysis of AOAC INTERNATIONAL? Please specify rationale.	Yes. This method demonstrates excellent performance with a variety of common materials in commerce, has been optimized for separation and extraction efficiency, meets all SMPR requirements, and can be used as a rapid method for routine cGMP compliance

The Scientific Association Dedicated to Analytical Excellence®

AOAC Stakeholder Panel on Dietary Supplements Expert Review Panel

AOAC Candidate Method #TUR-02

Quantitation of Curcuminoids

- Author(s): Aniko Solyom, Phil Koerner, Scott Krepich, Zeshan Aqeel, Peter Angeli
- Submitted by: Aniko Solyom
- Attachments: 0
- Submitter notes: None

ERP Reviewers

- Primary Reviewer: Nour Eddine Es-Safi
- Secondary Reviewer: Elizabeth Mudge

Candidate Method Location:

- https://goo.gl/r1mieQ (Must be an ERP member signed into AOAC website to access)

General Information	
Reviewer Name:	Nour Eddine ES-SAFI
Email:	nouressafi@yahoo.fr
Organization:	Mohammed V University in Rabat
Method Reviewed:	TUR-02
Method Title:	QUANTATION OF CURCUMINOIDS
Applicable SMPR	AOAC SMPR 2016.003
I. Summary	
Summary of Method:	The TUR-02 method presents results dealing the extraction of curcuminoids in dried plant material, extracts, tablets and capsules. The further quantification of curcumin, demethoxycurcumin and bisdemethoxycurcumin were investigated through HPLC-UV analysis
II. Review of the Method Only	
1. Does the Applicability of the Method Support the Applicability of the SMPR? If not, please explain what is missing.	The applicability of the method support the applicability of the AOAC SMPR 2016-003.
2. Does the analytical technique(s) used in the method meet the SMPR? If not, please specify how it differs from what is stated in the SMPR.	Yes the analytical technique used in the method meet the SMPR
3. Are the definitions specified in the SMPR used and applied appropriatly in the method? If no, please indicate how the terms are used.	Yes, the definitions specified in the SMPR were used and applied appropriately in the proposed method
4. Does the method, as written, contain all appropriate precautions and warnings related to the method's regaents, components, instrumentation, or method steps that may be hazardous? If no, please suggest wording or option(s).	Yes
III. Review of Information in Support of the Method	
1. Are the definitions specified in the SMPR used and applied appropriately in the supporting documentation (manuscripts, method studies, etc…)? If not, please explain the differences and if the method is impacted by the difference.	The definitions specified in the SMPR were used and applied appropriately in the supporting documentation

2. Is there information demonstrating that the method meets the SMPR Method Performance Requirements using the Reference Materials stated in the SMPR? If not, then specify what is missing and how this impacts demonstration of performance of the method.	Yes standard curcuminoids (curcumin, demethoxycurcumin and bisdemethoxycurcumin) were used.
3. Is there information demonstrating that the method performs within the SMPR Method Preformance Requiements table specifications for all analytes in the SMPR applicability statement? If not, please specify what is missing and whether or not the method's applicaiblity should be modified.	The LOD, LOQ and repeatability requirements were investigated and their corresponding parameters are reported in the proposed methods. In the repeatability given results, some points are out of ranges specified in the SMPR. Parameters concerning recovery and reproducibility are missing.
IV. General Submission Package	
1. Based on the supporting information, were there any additional steps in the evaluation of the method that indicated the need for any addional precautionary statements in the method?	Yes, recovery and reproducibility performance of the method are to be done.
2. Does the method contain system suitability tests or controls as specified by the SMPR? If not, please indicate if there is a need for such tests or controls, and which ones.	Yes the method include injections of blank and standard samples.
3. Is there information demonstrating that the method system suitability tests and controls as specified in the SMPR worked appropriately and as expected? If no, please specify.	Yes
4. Based on the supporting information, is the method written clearly and concisely? If no, please specify the needed revisions.	Yes the proposed method is clearly and concisely written.
5. Based on the supporting information, what are the pros/strenghts of the method?	The method seems to be simple giving a good separation of the three investigated curcuminoid compounds.
6. Based on the supporting information, what are the cons /weaknesses of the method?	Repeatability must be improved
7 . Any general comments about the method?	No
V. Recommendation for the Method	

Do you recommend this method be adopted as a First Action and published in the Official Methods of Analysis of AOAC INTERNATIONAL? Please specify rationale.	I think that the proposed method could be a candidate for the quantification of curcuminoids giving good separation of the three explored compounds (curcumin, demethoxycurcumin and bisdemethoxycurcumin). However it suffers from some limitations, such as the lack of supporting data on recovery in addition to the improvement of repeatability results. Therefore and in the prsented form, I don't think that the prtoposed method could be adopted as a First Action

General Information	
Reviewer Name:	Elizabeth Mudge
Email:	emudge@bcit.ca
Organization:	BCIT
Method Reviewed:	TUR-02
Method Title:	Quantitation of Curcuminoids
Applicable SMPR	2016.003
I. Summary	
Summary of Method:	This method is to quantify curcuminoids (curcumin, demethoxycurcumin and bis-demethoxycurcumin) in turmeric roots and finished products using HPLC with detection at 425 nm.
II. Review of the Method Only	
1. Does the Applicability of the Method Support the Applicability of the SMPR? If not, please explain what is missing.	Yes - this method is applicable to quantitation of the three major curcuminoids in turmeric roots and finished products
2. Does the analytical technique(s) used in the method meet the SMPR? If not, please specify how it differs from what is stated in the SMPR.	Yes.
3. Are the definitions specified in the SMPR used and applied appropriatly in the method? If no, please indicate how the terms are used.	Yes.
4. Does the method, as written, contain all appropriate precautions and warnings related to the method's regaents, components, instrumentation, or method steps that may be hazardous? If no, please suggest wording or option(s).	Yes - most of this method does not require significant safety warnings
III. Review of Information in Support of the Method	
1. Are the definitions specified in the SMPR used and applied appropriately in the supporting documentation (manuscripts, method studies, etc…)? If not, please explain the differences and if the method is impacted by the difference.	For the most part, yes - with the exception of repeatability. The studies performed here specify 6 replicates, but in this case these are actually triplicate samples with duplicate injections, treated as six replicates. This data is therefore somewhat biased to improve the precision of the method because replicate injections of the same sample will improve repeatability compared with six actual replicate samples which take into account the entire sample preparation.
2. Is there information demonstrating that the method meets the SMPR Method Performance Requirements using the Reference Materials stated in the SMPR? If not, then specify what is missing and how this impacts demonstration of performance of the method.	No reference materials were used in this study report.

3. Is there information demonstrating that the method performs within the SMPR Method Preformance Requiements table specifications for all analytes in the SMPR applicability statement? If not, please specify what is missing and whether or not the method's applicaiblity should be modified.	For linearity - yes. For analytical range - mostly, but limited to 25%, but does not meet the range requirements of greater than 50% (there are products out there that high). For LOQ - the SMPR specifies that the LOQ should be less than 0.1% in the sample materials. The LOQ specified in this method is ug/mL. using the lowest dilution for this method the curcumin (which is the highest LOQ - 2.86 ug/mL) would back calculate to 0.0143 %, assuming that the sample mass is 500 mg, which is not actually specified in the method (taken from optimization information). So Yes, LOQ is sufficient. Accuracy/Recovery - Not specified in the document. Repeatability - As specified above the methodology used biases the repeatability data due to between injection and between sample being treated as replicates. Additionally, the CurcuViva product is consistently not within the <5% RSDr as specified in the SMPR. All other products appear to meet SMPR repeatability requirements, but I would recommend re-treating the data and only taking the triplicate measurements to ensure true repeatability is obtained. Some of the other dietary ingredients listed in Table 2 have also not been used in any of the multi-component products.
IV. General Submission Package	
1. Based on the supporting information, were there any additional steps in the evaluation of the method that indicated the need for any addional precautionary statements in the method?	No
2. Does the method contain system suitability tests or controls as specified by the SMPR? If not, please indicate if there is a need for such tests or controls, and which ones.	The system suitability requirements of the SMPR specifies that check standards should be run throughout the validation - in the case of this run, there were only blank samples, but no QC/check standards run. System suitability was not performed at the beginning of the validation either to ensure that the system was working properly.
3. Is there information demonstrating that the method system suitability tests and controls as specified in the SMPR worked appropriately and as expected? If no, please specify.	Linearity was sufficient, which would ensure that the system was suitable, but there was no actual system suitability performed.
4. Based on the supporting information, is the method written clearly and concisely? If no, please specify the needed revisions.	There is some information that is missing in the method preparation. The actual sample masses of the samples (or atleast recommended ranges of masses) are missing. This would cause a lot of confusion adopting the method. Additionally, several different extraction volumes were used. It would make sense to specify which volumes were suitable for what types of samples (or what expected ranges or ingredients). As for softgels, the results obtained from this would likely be mg/capsule, rather than mg/g, as you should not typically include the mass of the softgel in the analysis.
5. Based on the supporting information, what are the pros/strenghts of the method?	The strengths: simple sample preparation fast chromatographic separation baseline separation of the curcuminoids suitable for many sample types, ingredients, etc.
6. Based on the supporting information, what are the cons /weaknesses of the method?	The weaknesses: lack of sufficient information for sample preparation requires a significant amount of glassware extraction solvent selection (see below)

7 . Any general comments about the method?	The selection of extraction solvent seemed to be based more on the precision of the replicates rather than on analyte recovery. When looking at the error, the error of the methanol is slightly larger, but there is a significant difference between the quantity of curcumin and DMC extracted with methanol compared with the 80% methanol. Without any recovery data in this study, it appears that the accuracy of this method may be insufficient. There were also re-extraction repetitive dilution study, but without descriptions of the actual methodologies, it is difficult to understand how this graph explains that only 1 extraction is required.
V. Recommendation for the Method	
Do you recommend this method be adopted as a First Action and published in the Official Methods of Analysis of AOAC INTERNATIONAL? Please specify rationale.	No. The method appears to be missing key information including method recovery/accuracy. The repeatability data was calculated using duplicate injections as replicates and the selection of extraction solvent appears to impact the recovery of the curcuminoids.

General Information	
Reviewer Name:	Melissa Phillips
Email:	melissa.phillips@nist.gov
Organization:	NIST
Method Reviewed:	TUR-02
Method Title:	Quantitation of Curcuminoids
Applicable SMPR	2016.003
I. Summary	
Summary of Method:	An 80:20 methanol:water extraction followed by LC-absorbance for determination of CUR, BDMC, and DMC in turmeric materials. Optimization of some extraction parameters was discussed.
II. Review of the Method Only	
1. Does the Applicability of the Method Support the Applicability of the SMPR? If not, please explain what is missing.	Yes
2. Does the analytical technique(s) used in the method meet the SMPR? If not, please specify how it differs from what is stated in the SMPR.	Yes
3. Are the definitions specified in the SMPR used and applied appropriatly in the method? If no, please indicate how the terms are used.	Yes
4. Does the method, as written, contain all appropriate precautions and warnings related to the method's regaents, components, instrumentation, or method steps that may be hazardous? If no, please suggest wording or option(s).	No safety information is included. Could potentially include a statement regarding flammability of solvents, acidity of mobile phase, as well as use of elevated column temperature, but this is not critical.
III. Review of Information in Support of the Method	
1. Are the definitions specified in the SMPR used and applied appropriately in the supporting documentation (manuscripts, method studies, etc…)? If not, please explain the differences and if the method is impacted by the difference.	Yes
2. Is there information demonstrating that the method meets the SMPR Method Performance Requirements using the Reference Materials stated in the SMPR? If not, then specify what is missing and how this impacts demonstration of performance of the method.	Reference standards were acquired from a different vendor (Chromadex) than those listed in the SMPR; purity of standards should be evaluated and confirmed or corrected in-house. Matrix reference materials listed in SMPR were not used in this study (although they are not yet available from NIST).

3. Is there information demonstrating that the method performs within the SMPR Method Preformance Requiements table specifications for all analytes in the SMPR applicability statement? If not, please specify what is missing and whether or not the method's applicaiblity should be modified.	LOQs reported: 0.92 ug/mL (0.09%) for BDMC 1.33 ug/mL (0.13%) for DMC 2.86 ug/mL (0.29%) for CUR SMPR states 0.1%; met for BDMC, slightly high for DMC, high for CUR. *Calculations to % done without inclusion of methanol density, which would actually make them higher and further from requirement No recovery information provided. Analytical ranges reported: 0.09-1.66% for BDMC (0.6-15 ug/mL cal range) 0.39-10.3% for DMC (1.7-42 ug/mL cal range) 0.00016-25.2% for CUR (4-100 ug/mL cal range) SMPR requests 0.1% to >50%; based on levels in products and the calibration ranges demonstrated, I have no concern about this method working in this range. RSDrs for one BDMC test, one CUR test, and 3 DMC tests are outside of the range in SMPR (LT 5% at 0.1-50%; LT 3% at >50%). Values range from 6-7%. 6 samples were tested.
IV. General Submission Package	
1. Based on the supporting information, were there any additional steps in the evaluation of the method that indicated the need for any addional precautionary statements in the method?	No
2. Does the method contain system suitability tests or controls as specified by the SMPR? If not, please indicate if there is a need for such tests or controls, and which ones.	No
3. Is there information demonstrating that the method system suitability tests and controls as specified in the SMPR worked appropriately and as expected? If no, please specify.	N/A
4. Based on the supporting information, is the method written clearly and concisely? If no, please specify the needed revisions.	Yes
5. Based on the supporting information, what are the pros/strenghts of the method?	The method is straightforward and relatively simple. It has been tested on a number of matrices (as suggested in Table 3 of SMPR) and in the presence of numerous other botanicals (including 2 from Table 2 of SMPR).

6. Based on the supporting information, what are the cons /weaknesses of the method?	No tinctures or liquids were tested. Would like to see demonstration of method on these types of products to meet SMPR. Method was not tested in the presence of carotenoids, piper nigrum, or Ca. Would like to see demonstration of method on products including these to meet SMPR. Would prefer an internal standard approach to the external standard approach presented here. No recovery data - would need recovery data to meet SMPR. Several concerns about extraction optimization. 100% methanol gives 10%+ better extraction efficiency, yet 80:20 methanol:water was selected. Also, additional curcuminoids are being extracted in 2nd-4th extractions. This could be because 80:20 methanol:water is not as effective as 100% methanol, but would like to see the same study with 100% methanol.
7 . Any general comments about the method?	No
V. Recommendation for the Method	
Do you recommend this method be adopted as a First Action and published in the Official Methods of Analysis of AOAC INTERNATIONAL? Please specify rationale.	Not at this time. I think significant work is needed on the extraction protocol and more products need to be explored before this method could be First Action.

First Action Method Updates

Expert Review Panel Tracking and Recommendations of First Action Methods

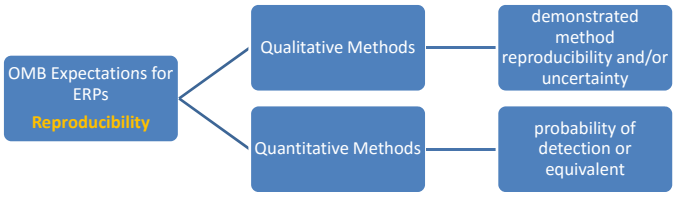
AOAC Policies & Procedures

- Policy on Antitrust
- Policy on Use of Association Name, Identifying Insignia, Letterhead, Business Cards
- Policy on Volunteer Conflict of Interest
- Expert Review Panel Policies and Procedures
- OMA Appendix G

OMA, Appendix G

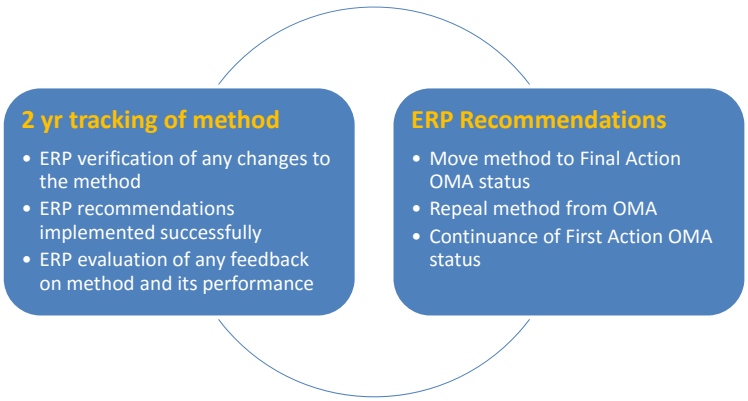
Further data indicative of adequate method reproducibility (between laboratory) performance to be collected. Data may be collected via a collaborative study or by proficiency or other testing data of similar magnitude.

- ERP is looking to verify if method reproducibility has been appropriately assessed and satisfactorily demonstrated



OMA, Appendix G

Two years maximum transition time (additional year(s) if ERP determines a relevant collaborative study or proficiency or other data collection is in progress).



OMA, Appendix G

Method removed from Official First Action and OMA if no evidence of method use available at the end of the transition time.

First Action OMA Tracking

- Tracking period is ≤ 2 years and begins on the date of the ERP's decision to adopt a method for OMA First Action status.

No Use in 2 Years

- Repeal from OMA

OMA, Appendix G

Method removed from Official First Action and OMA if no data indicative of adequate method reproducibility is forthcoming as outlined above at the end of the transition time.

First Action OMA Tracking

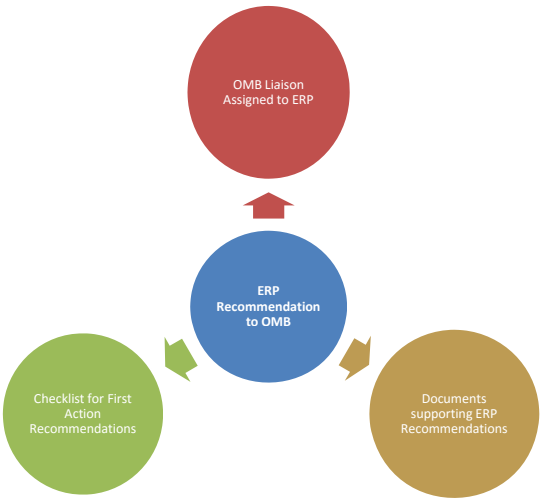
- Tracking period is ≤ 2 years and begins on the date of the ERP's decision to adopt a method for OMA First Action status.

No Demonstration of Method Reproducibility in ≤ 2 Years

- Repeal from OMA

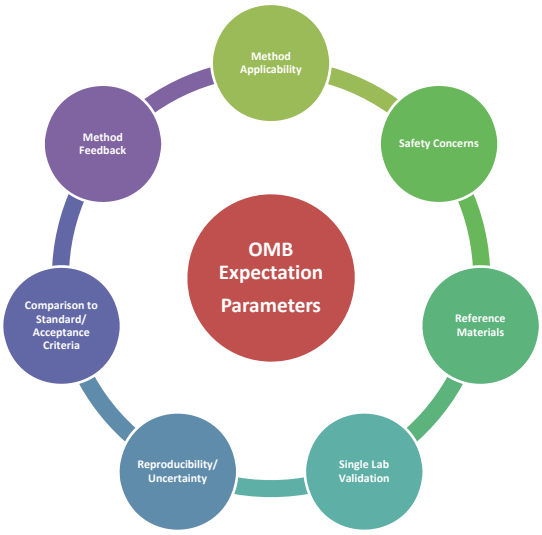
OMA, Appendix G

ERP to recommend Method to Official Final Action Status to the OMB.

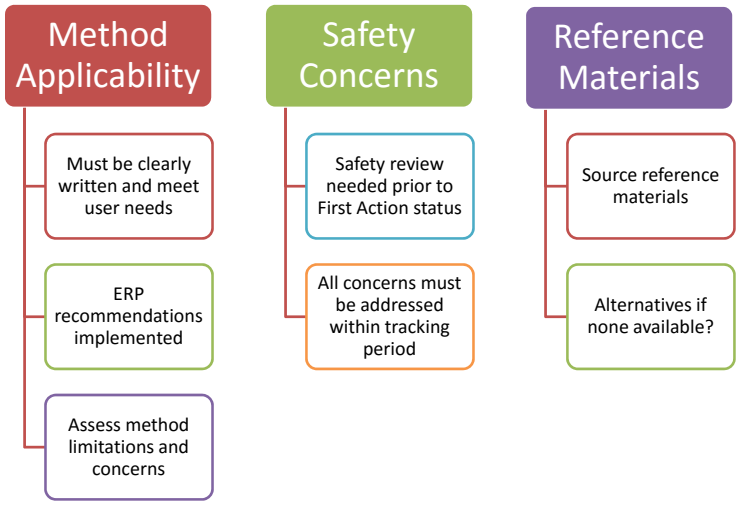


OMA, Appendix G

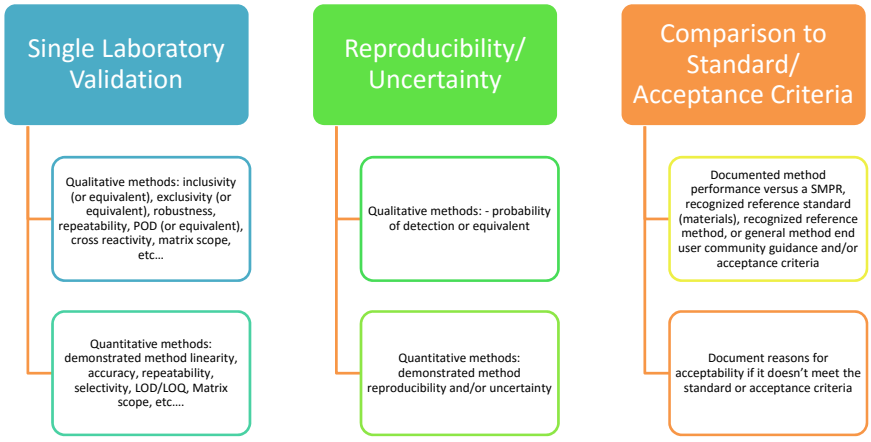
First Action to Final Action Methods: Guidance for AOAC Expert Review Panels



OMB Expectation Parameters



OMB Expectation Parameters



OMB Expectation Parameters

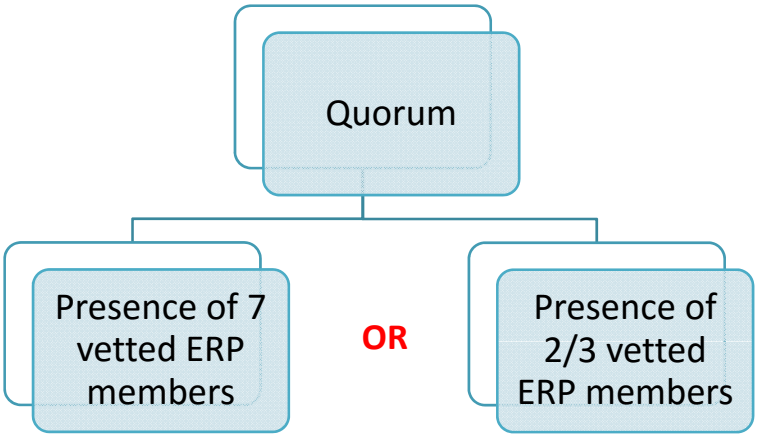
Method Feedback from End Users

Consider any positive or negative feedback on overall method performance, applicability, availability of reference materials, matrix scope, method component sourcing, robustness or ruggedness parameters.

Documentation Needed

- Method Safety Evaluation
- Reference Materials
- Evidence of Single Laboratory Validation or equivalent
- Evidence of Reproducibility Assessment
- Published First Action OMA
- Method Performance versus SMPR *or* acceptance criteria
- Final draft of First Action OMA to be considered for status update
- Rationale or Justification for Repeal or Continuance of First Action OMA

ERP Meetings



WHICHEVER IS GREATER

ERP Meetings

METHOD AUTHOR: present any method feedback obtained and any resulting changes to the method, any reproducibility information, any implemented ERP recommendations, final draft of method proposed for decision

ERP MEMBERS: present any method feedback obtained and discuss any resulting changes to the method, any reproducibility information, any implemented ERP recommendations, review and agree upon final draft of method proposed for decision, and make a recommendation to OMB.

CONSENSUS: 2/3 vote in favor of a motion. Abstentions do not count towards vote; in case of multiple abstentions. Staff will monitor and record consensus voting.

STAFF: Will organize and coordinate meeting, record ERP actions and decisions, draft ERP report and distribute after chair approval, work with chair and OMB liaison to complete checklist and assemble recommendation package for OMB.

Questions?

Thank you.

Appendix D: Guidelines for Collaborative Study Procedures To Validate Characteristics of a Method of Analysis

{*Note*: These guidelines incorporate symbols, terminology, and recommendations accepted by consensus by the participants at the IUPAC Workshop on Harmonization of Collaborative Analytical Studies, Geneva, Switzerland, May 4–5, 1987 [*Pure Appl. Chem.* **60**, 855–864(1988); published as "Guidelines for Collaborative Study of Procedure to Validate Characteristics of a Method of Analysis," *J. Assoc. Off. Anal. Chem.* **72**, 694–704(1989)]. The original guidelines were revised at Lisbon, Portugal, August 4, 1993, and at Delft, The Netherlands, May 9, 1994, *Pure Appl. Chem.* **67**, 331–343(1995). These revised, harmonized guidelines have been adopted by AOAC INTERNATIONAL as the guidelines for the AOAC Official Methods Program, *J. AOAC Int.* **78**(5), 143A–160A(1995). Although the directions were developed for chemical studies, some parts may be applicable to all types of collaborative studies.}

Summary Statement of AOAC Recommendation for Design of a Collaborative Study

Minimum Criteria for Quantitative Study

Minimum number of materials (see Note 1 on p. 4).—Five (only when a single level specification is involved for a single matrix may this minimum be reduced to 3).

Minimum number of laboratories.—Eight reporting valid data for each material (only in special cases involving very expensive equipment or specialized laboratories may the study be conducted with a minimum of 5 laboratories, with the resulting expansion in the confidence interval for the statistical estimates of the method characteristics).

Minimum number of replicates.—One, if within-laboratory repeatability parameters are not desired; 2, if these parameters are required. Replication should ordinarily be attained by blind replicates or split levels (Youden pairs).

Minimum Criteria for Qualitative Analyses

Ten laboratories reporting on 2 analyte levels per matrix, 6 test samples per level, and 6 negative controls per matrix. (*Note*: AOAC criteria for qualitative analyses are not part of the harmonized guidelines.)

1. Preliminary Work (Within One Laboratory)

1.1 Determine Purpose and Scope of the Study and Method

Determine purpose of the study (e.g., to determine attributes of a method, proficiency of analysts, reference values of a material, or to compare methods), the type of method (empirical, screening, practical, reference, definitive), and the probable use of the method (enforcement, surveillance, monitoring, acceptance testing, quality control, research). Also, on the basis of the relative importance of the various method attributes (bias, precision, specificity, limit of determination), select the design of the collaborative study. The directions in this document pertain primarily to determining the precision characteristics of a method, although many sections are also appropriate for other types of studies.

Alternatives for Method Selection

(*1*) Sometimes obvious (only method available).

(*2*) Critical literature review (reported within-laboratory attributes are often optimistic).

(*3*) Survey of laboratories to obtain candidate methods; comparison of within-laboratory attributes of candidate methods (sometimes choice may still not be objective).

(*4*) Selection by expert [AOAC-preferred procedure (selection by Study Director with concurrence of General Referee)].

(*5*) Selection by Committee (ISO-preferred procedure; often time-consuming).

(*6*) Development of new method or modification of existing method when an appropriate method is not available. (Proceed as a research project.) (This alternative is time-consuming and resource-intensive; use only as a last resort.)

1.2 Optimize Either New or Available Method

Practical Principles

(*1*) Do not conduct collaborative study with an unoptimized method. An unsuccessful study wastes a tremendous amount of collaborators' time and creates ill will. This applies especially to methods that are formulated by committees and have not been tried in practice.

(*2*) Conduct as much experimentation within a single laboratory as possible with respect to optimization, ruggedness, and interferences. Analysis of the same material on different days provides considerable information on variability that may be expected in practice.

Alternative Approaches to Optimization

(*1*) Conduct trials by changing one variable at a time.

(*2*) Conduct formal ruggedness testing for identification and control of critical variables. *See* Youden and Steiner (pp 33–36, 50–55). The actual procedure is even simpler than it appears. (This is an extremely efficient way for optimizing a method.)

(*3*) Use Deming simplex optimization to identify critical steps. *See* Dols and Armbrecht. The simplex concept can be used in the optimization of instrument performance and in application to analytical chemical method development.

1.3 Develop Within-Laboratory Attributes of Optimized Method

(Some items can be omitted; others can be combined depending on whether study is qualitative or quantitative.)

Determine calibration function (response vs concentration in pure or defined solvent) to determine useful measurement range of method. For some techniques, e.g., immunoassay, linearity is not a prerequisite. Indicate any mathematical transformations needed.

Determine analytical function (response vs concentration in matrix, including blank) to determine applicability to commodity(ies) of interest.

Test for interferences (specificity): (*1*) Test effects of impurities, ubiquitous contaminants, flavors, additives, and other components expected to be present and at usual concentrations. (*2*) Test nonspecific effects of matrices. (*3*) Test effects of transformation products, if method is to indicate stability, and metabolic products, if tissue residues are involved.

Conduct bias (systematic error) testing by measuring recoveries of analyte added to matrices of interest and to extracts, digests, or other treated solutions thereof. (Not necessary when method defines property or component.)

Develop performance specifications for instruments and suitability tests for systems (which utilize columns or adsorbents) to ensure satisfactory performance of critical steps (columns, instruments, etc.) in method.

Conduct precision testing at the concentration levels of interest, including variation in experimental conditions expected in routine analysis (ruggedness). In addition to estimating the "classical" repeatability standard deviation, s_r, the initiating laboratory may estimate the total within-laboratory standard deviation (s_e) whereby s_e is the variability at different days and with different calibration curves, by the same or different analysts within a single laboratory. This total within-laboratory estimate reflects both between-run (between-batch) and within-run (within-batch) variability.

Delineate the range of applicability to the matrices or commodities of interest.

Compare the results of the application of the method with existing, studied methods intended for the same purposes, if other methods are available.

If any of the preliminary estimates of the relevant performance of these characteristics are unacceptable, revise the method to improve them, and re-study as necessary.

Have method tried by analysts not involved in its development.

Revise method to handle questions raised and problems encountered.

1.4 Prepare Description of Method

Note: A collaborative study of a method involves practical testing of the written version of the method, in its specific style and format, by a number of laboratories on identical materials.

Prepare method description as closely as possible to format and style that will be used for eventual publication.

Always express reagent concentrations in terms of mass (or volume) per volume (or mass); never in terms requiring the analyst to recalculate or look up formula weights, e.g., moles. Moles may be used, particularly with volumetric standards, but only in addition to mass and volume. Many errors are caused by incorrect recalculation of formula weights.

Clearly specify requirements for chromatographic materials, enzymes, antibodies, and other performance-related reagents.

Clearly describe and explain every step in the analytical method so as to discourage deviations. Use imperative directions; avoid subjunctive and conditional expressions as options as far as possible.

Clearly describe any safety precautions needed.

Edit method for completeness, credibility (e.g., buffer pH consistent with specified chemicals, volumes not greater than capacity of container), continuity, and clarity.

Check for inclusion of performance specifications and system suitability tests, defined critical points, and convenient stopping points. Incorporate physical or chemical constants of working standards solutions, e.g., absorptivities, half-scale deflections, recoveries, etc., or properties of operating solutions and chromatographic materials, e.g., pH, volumes, resolution, etc., and any other indicators (e.g., sum equals 100%) that suggest analysis is proceeding properly.

If time and resources are available, conduct pilot study involving 2–3 laboratories.

1.5 Invite Participation

Selection of Collaborators/Candidate Laboratories

Laboratories invited to participate should have personnel experienced in the basic techniques employed; experience with the method itself is not a prerequisite for selection. Lists of possible participants can be developed through personal contacts, technical societies, trade associations, or literature search, and advertisements in the Referee section of AOAC's magazine, *Inside Laboratory Management*. Collaborators are chosen by the organizers of the collaborative study from a diversity of laboratories with interest in the method, including regulatory agencies, industry, and universities.

Letter of Invitation

Address a formal letter to the individual responsible for assignment of laboratory effort. State reason for selecting that laboratory (e.g., as a volunteer or has responsibility or familiarity with the problem or method), estimated number of person-hours required for performance, number of test samples to be sent, number of analyses to be required, expected date for test sample distribution, and target date for completion of the study. *Emphasize the importance of management support in assigning the necessary time for the project*. Enclose a copy of the method and a return form or card (with postage affixed, if appropriate), requiring only a check mark for acceptance or refusal of the invitation, a signature, space for address corrections, telephone and fax numbers, e-mail, and date.

Laboratory Coordinator

With large studies, involving several analysts per laboratory, several familiarization samples, receipt of items at different times, or similar recurrent situations, acceptance of the invitation should be followed by a letter suggesting that a Laboratory Coordinator be appointed. The Laboratory Coordinator should be responsible for receiving and storing the study materials, assigning the work, dispensing study materials and information related to the study, seeing that the method is followed as written, accumulating the data, assuring that the data are correctly reported, and submitting the collaborative study manuscript within the deadline.

1.6 Instructions and Report Forms

Carefully design and prepare instructions and forms, and scrutinize them before distribution. A pilot study is also useful for uncovering problems in these documents.

Send instructions and report forms immediately on receipt of acceptance, independent of study materials, if selection of

laboratories is not to be based on performance in pilot or training studies. The instructions should include in bold face or capital letters a statement:

THIS IS A STUDY OF THE METHOD, NOT OF THE LABORATORY. THE METHOD MUST BE FOLLOWED AS CLOSELY AS PRACTICABLE, AND ANY DEVIATIONS FROM THE METHOD AS DESCRIBED, NO MATTER HOW TRIVIAL THEY MAY SEEM, MUST BE NOTED ON THE REPORT FORM.

Include instructions on storage and handling, markings, and identifications to be noted, any special preparation for analysis, and criteria for use of practice or familiarization samples, if included. Pre-code the form for each laboratory and provide sufficient space for as much sequential data as may be required for proper evaluation of the results, including a check of the calculations.

The initiating laboratory should indicate the number of significant figures to be reported, usually based on the output of the measuring instrument.

Note: In making statistical calculations from the reported data, the full power of the calculator or computer is to be used with no rounding or truncating until the final reported mean and standard deviations are achieved. At this point the standard deviations are rounded to 2 significant figures and the means and relative standard deviations are rounded to accommodate the significant figures of the standard deviation. For example, if the reproducibility standard deviation $s_R = 0.012$, the mean is reported as 0.147, not as 0.1473 or 0.15, and RSD_R, relative reproducibility standard deviation, is reported as 8.2%. If standard deviation calculations must be conducted manually in steps, with the transfer of intermediate results, the number of significant figures to be retained for squared numbers should be at least 2 times the number of figures in the data plus 1.

When recorder tracing reproductions are required to evaluate method performance, request their submission both in the instructions and as a check item on the form. Provide instructions with regard to labeling of recorder tracings, such as identification with respect to item analyzed, axes, date, submitter, experimental conditions, and instrument settings.

Include in the report form a signature line for the analyst and lines for a printed or typed version of the name and address for correct acknowledgement.

Provide for a review by the laboratory supervisor. An example of a completed form is helpful. A questionnaire may be included or sent after completion of the analyses in which the questions can be designed to reveal if modifications have been made at critical steps in the method.

Request a copy of the calibration curve or other relationship between response and concentration or amount of analyte so that if discrepancies become apparent after examining all of the data, it can be determined whether the problem is in the calibration or in the analysis.

1.7 Familiarization or Practice Samples

If deemed necessary, supply as far ahead as practicable, familiarization samples, with instructions, before actual materials are sent. When familiarization samples have been submitted, supply forms for reporting progress toward satisfactory performance.

2. Design of the Collaborative Study

2.1 General Principles

The purpose of a collaborative study is to determine estimates of the attributes of a method, particularly the "precision" of the method that may be expected when the method is used in actual practice. The AOACI uses 2 terms to define the precision of a method under 2 circumstances of replication: repeatability and reproducibility. Repeatability is a measure of the variation, s_r^2, between replicate determinations by the same analyst. It defines how well an analyst can check himself using the same method on blind replicates of the same material or split levels (Youden pairs), under the same conditions (e.g., same laboratory, same apparatus, and same time). Reproducibility is a composite measure of variation, s_R^2, which includes the between-laboratory and within-laboratory variations. It measures how well an analyst in a given laboratory can check the results of another analyst in another laboratory using the same method to analyze the same test material under different conditions (e.g., different apparatus and different time). The between-laboratory variation represents a systematic error that reflects variation arising from environmental conditions (e.g., condition of reagent and instruments, variation in calibration factors, and interpretations of the steps of the method) associated with the laboratories used in the study. Therefore, it is important to identify the causes of the differences among laboratories so that they may be controlled. Otherwise they will be summed into s_R^2.

Present test samples sent for analysis as unknowns (blind) and coded in a random pattern. If necessary to conserve analyst time, an indication of the potential range of concentration or amount of analyte may be provided. If spiking solutions are used, provide one coded solution for each material. All spiking solutions should be identical in appearance and volume. Do not provide a single solution from which aliquots are to be removed for spiking. Any information with regard to concentration (e.g., utilizing factorial aliquots or serial dilutions of the same spiking solutions) or known replication is likely to lead to an underestimate of the variability.

The study must be extensive enough to assure sufficient data surviving in the face of possible loss of materials during shipment, inability of collaborators to participate after acceptance, and a maximum outlier rate of 2/9 and still maintain valid data from a minimum of 8 laboratories.

Improper preparation of reference standards and standard solutions can cause a significant portion of the analytical error. A decision must be made whether such error is to be considered separately or as part of the method, i.e., will the analysts procure their own standard solutions or will standards be provided by the Study Director. The decision depends primarily on the availability of the standard. If the standard is readily available, the analysts should prepare their own. If the standard is not readily available, the standard may be supplied, but physical constants, e.g., absorptivity of working standard solutions, should be incorporated into the description as a check on proper preparation of the solution.

Obtain the necessary administrative and operational approvals. Review by potential users of the method is also desirable.

2.2 Laboratories

Laboratories must realize the importance of the study. A large investment is being made in studying the method and this probably will be only collaborative study of the method that will performed.

Therefore, it is important to have a fair and thorough evaluation of the method.

Type

The most appropriate laboratory is one with a responsibility related to the analytical problem. Laboratory types may be representative (selection of laboratories that will be using the method in practice), reference (assumed to be "best"), or the entire population of laboratories (usually certified or accredited) that will be using the method. Final selection of participants should be based on a review with the General Referee and others of each laboratory's capabilities and past performance in collaborative studies, followed up, if possible, by telephone conversations or by personal visits. Selection may also be based on performance with familiarization samples. Sometimes only laboratories with dedicated or very specialized instruments must be used. If the study is intended for international consideration, laboratories from different countries should be invited to participate.

Number of Laboratories

Minimum of 8 laboratories submitting valid data (to avoid unduly large confidence bands about the estimated parameters). Only in special cases of very expensive equipment or specialized laboratories may the study be conducted with a minimum of 5 laboratories. Fewer laboratories widen the confidence limits of the mean and of the variance components (*see* design considerations). The optimum number of laboratories, balancing logistics and costs against information obtained, often is 8–10. However, larger studies are not discouraged.

For qualitative analyses, a minimum of 10 laboratories is needed; collaborative study must be designed to include 2 analyte levels per matrix, 6 test samples per level, and 6 negative controls per matrix. (*Note 1*: AOAC criteria for qualitative analyses are not part of the harmonized guidelines.)

Analysts

Most designs require only 1 analyst per laboratory. If analyst–within-laboratory variability is a desired variance component, multiple analysts should be requested from all participating laboratories. Ordinarily 2 analysts from the same laboratory cannot be substituted for different laboratories, unless standard solutions, reagents, chromatographic columns and/or materials, instrument calibrations, standard curves, etc., are prepared independently, and no consultation is permitted during the work. Different laboratories from the same organization may be used as separate laboratories if they operate independently with their own instruments, standards, reagents, and supervision.

2.3 Test Materials

Homogeneous Materials

Materials must be homogeneous; this is critical. Establish homogeneity by testing a representative number of laboratory samples taken at random before shipment. (A collaborator who reports an outlying value will frequently claim receipt of a defective laboratory sample.) The penalty for inhomogeneity is an increased variance in the analytical results that is not due to the intrinsic method variability.

Test Sample Coding

Code test samples at random so that there is no pre-selection from order of presentation.

Concentration Range

Choose analyte levels to cover concentration range of interest. If concentration range of interest is a tolerance limit or a specification level, bracket it and include it with materials of appropriate concentration. If design includes the determination of absence of analyte, include blank (not detectable) materials as part of range of interest.

Number of Materials

A minimum of 5 materials must be used in the collaborative study. Three materials are allowed but only when a single specification is involved for a single matrix.

Note 1: A material is an analyte (or test component)/matrix/concentration combination to which the method-performance parameters apply. This parameter determines the applicability of the method.

Note 2: The 2 test samples of blind or open duplicates are a single material (they are not independent).

The 2 test samples constituting a matched pair (called X and Y) are considered Youden matched pairs only if they are sufficiently close in composition. "Sufficiently close" would be considered as ≤5% difference in composition between X and Y. That is, given that the concentration of analyte in X (x_c) is higher than the concentration of the analyte in Y (y_c) then:

$$\frac{x_c - y_c}{x_c} \leq 0.05$$

or:

$$y_c \geq (x_c - 0.05x_c)$$

Note 3: The blank or negative control may or may not be a material, depending on the usual purpose of the analysis. For example, in trace analysis, where very low levels (near the limit of quantitation) are often sought, the blanks are considered as materials, and are necessary to determine certain statistical "limits of measurement;" however, if the blank is merely a procedural control, in macro-level analysis (e.g., fat in cheese), it would not be considered a material.

Nature of Materials

Materials should be representative of commodities usually analyzed, with customary and extreme values for the analyte.

Size of Test Samples

Furnish only enough test sample to provide the number of test portions specified in the instructions. If additional test portions are required, the collaborator must request them, with an explanation.

Interferences

If pertinent, some materials, but not all, should contain contaminants and interferences in concentrations likely to be encountered, unless they have been shown to be unimportant through within-laboratory study. The success of the method in handling interference on an intralaboratory basis will be demonstrated by passing systems suitability tests.

Familiarization Samples

With new, complex, or unfamiliar techniques, provide material(s) of stated composition for practice, on different days, if possible. The valuable collaborative materials should not be used until the analyst can reproduce the stated value of the familiarization samples within a given range. However, it should be pointed out that one of the assumptions of analysis of variance is that the underlying distribution of results is independent of time (i.e., there is no drift). The Study Director must be satisfied that this assumption is met.

2.4 Replication

When within-laboratory variability is also of interest, as is usually the case, independent replication can be ensured by applying at least one of the following procedures (listed in suggested order of desirability; the nature of the design should not be announced beforehand):

(*1*) *Split levels (Youden pairs).*—The 2 test materials, nearly identical but of slightly different composition (e.g., ≤5% difference in composition, *see 2.3 Number of Materials, Note 2*) are obtained either naturally or by diluting (or by fortifying) one portion of the material with a small amount of diluent (or of analyte). Both portions are supplied to the participating laboratories as test samples, each under a random code number, and each test sample should be analyzed only once; replication defeats the purpose of the design.

(*2*) *Split levels for some materials and blind duplicates for other materials in the same study.*—Obtain only single values from each test sample supplied.

(*3*) *Blind duplicate test samples, randomly coded.—Note*: Triplicate and higher replication are relatively inefficient when compared with duplicate test samples because replication provides additional information only on individual within-laboratory variability, which is usually the less important component of error. It is more effective to utilize resources for the analysis of more levels and/or materials rather than for increasing the number of replicates for the individual materials.

PRACTICAL PRINCIPLE: With respect to replication, the greatest net marginal gain is always obtained in going from 2 to 3 as compared to going from 3 to 4, 4 to 5, etc.

(*4*) *Independent materials*.—(*Note*: Unrelated independent materials may be used as a split level in the calculations of the precision parameters or for plotting. There should be ≤5% difference in composition for such materials (*see 2.3 Number of Materials, Note 2*). The more they differ in concentration, the less reliable the information they provide on within-laboratory variability.)

(*5*) *Known replicates.—Use of known replicates is a common practice*.—It is much preferable to use the same resources on blind replicates or split levels.

(*6*) *Quality control materials*.—Instead of obtaining repeatability parameters through the collaborative study, information can be obtained from use of quality control materials in each laboratory individually, for its own use, independent of the collaborative study, for a separate calculation of s_r, using 2 (or more) replicates from each quality control test, according to the pattern developed for each product.

2.5 Other Design Considerations

The design can be reduced in the direction of less work and less cost, but at the sacrifice of reduced confidence in the reliability of the developed information.

More work (values) is required if more confidence is needed, e.g., greater confidence is required to enforce a tolerance at 1.00 mg/kg than at 1.0 mg/kg. (The distinction is a precision requirement of the order of 1% rather than 10%.)

The estimate of the standard deviation or the corresponding relative standard deviation obtained from a collaborative study is a random variable that varies about its corresponding true value. For example, the standard deviation, s_r, which measures within laboratory or repeatability precision has associated with it a standard deviation (STD = s_r) describing its scatter about the true value σ_r. Therefore, s_r, whose STD (s_r) is a function of s_r^2, number of laboratories, and number of analyses per laboratory, will vary about σ_r from occasion-to-occasion even for the same test conditions and material. The STD s_R, which measures among laboratory or reproducibility precision, has a STD (s_R) that is a function of the random variables s_r^2 and s_L^2, number of laboratories, and number of analyses per laboratory. s_R will vary about its true value σ_R from occasion-to-occasion for the same test material.

The validity of extrapolating the use of a method beyond concentrations and components tested can be estimated only on the basis of the slope of the calibration curve (sensitivity) observed as a function of the nature and concentration of the matrix and contaminant components. If the signal is more or less independent of these variables, a reasonable amount of extrapolation may be utilized. The extrapolator assumes the burden of proof as to what is reasonable.

3. Preparation of Materials for Collaborative Studies

3.1 General Principles

Heterogeneity between test samples from a single test material must be negligible compared to analytical variability, as measured within the Study Director's laboratory.

The containers must not contribute extraneous analytes to the contents, and they must not adsorb or absorb analytes or other components from the matrix, e.g., water.

If necessary, the materials may be stabilized, preferably by physical means (freezing, dehydrating), or by chemical means (preservatives, antioxidants) which do not affect the performance of the method.

Composition changes must be avoided, where necessary, by the use of vapor-tight containers, refrigeration, flushing with an inert gas, or other protective packaging.

3.2 Materials Suitable for Collaborative Studies

Material and analyte stability: Ensure analyte and matrix stability over projected transport time and projected length of study.

Single batch of homogenous, stable product such as milk powder, peanut butter, vegetable oil, starch, etc., is the best type of material.

Reference materials supplied by standards organizations such as National Institute of Standards and Technology (NIST, Gaithersburg, MD) and EC's Joint Research Center and Institute on Reference Materials and Methods (IRMM, Belgium) are excellent, unless they have easily recognizable characteristics (e.g., odor and color of NIST Orchard Leaves). However, they are of limited availability, composition, and analyte level. If available, they are expensive. Sometimes the certification organization may be interested in making reference materials available for the analyte under study, in which case it may assist in providing the material for the study.

Synthetic materials may be especially formulated with known amounts of analytes by actual preparation for the study. This procedure is best used for macro-constituents such as drugs or pesticide formulations.

Spiked materials consisting of normal or blank materials to which a known amount of analyte has been added may be used. The amount of analyte added should not be excessive in relation to the amount present (e.g., about 2×), and the analyte added should be in the same chemical form as present in the commodities to be analyzed subsequently.

In drug and pesticide residue-type problems, it is often necessary to use spiked materials in order to assess recovery. However, because incurred residues are likely to present different problems from those of spiked residues, collaborative studies should include some test samples with incurred residues to ensure that the method is applicable under these conditions as well.

(*1*) *Preparation in bulk*.—This requires thorough and uniform incorporation of analyte, often by serial dilution of solids. The danger of segregation due to differences in densities always exists. Fluid materials susceptible to segregation should be prepared under constant agitation. Uniformity should be checked by direct analysis, with an internal standard, or by a marker compound (dye or radioactive label).

(*2*) *Test samples, individually prepared*.—A known amount of analyte is either weighed directly or added as an aliquot of a prepared solution to pre-measured portions of the matrix in individual containers. The collaborator is instructed to use each entire portion for the analysis, transferring the contents of the container quantitatively or a substantial weighed fraction of the portion. (This is the preferred alternative to spiked solid materials at trace [mg/kg] levels, at the expense of considerably more work.)

(*3*) *Concentrated unknown solutions for direct addition by collaborators to their own commodities*.—Should be used only as a last resort when instability of the analyte precludes distribution from a central point. To preclude direct analysis of the spiking solution, supply individual coded solutions to be added in their entirety to portions of the matrix for single analyses by each laboratory. All solutions should have the same volume and appearance. This type of material is analogous to that of test samples except for the source of matrix. This case should be used only for perishable commodities that are altered by all available preservation techniques.

Materials analyzed by another, presumably accurate, method, if available, in the Study Director's laboratory or by some or all the collaborators.

Only as an absolutely last resort (usually with unstable materials and preparation of material studies) should the collaborators be permitted to prepare their own materials for analysis. Since it is impossible to avoid the personal bias introduced by knowledge of the composition of the material, the materials should be prepared in each laboratory by an individual who will not be involved in the analyses.

3.3 Blanks

When the absence of a component is as important as its presence, when determinations must be corrected for the amount of the component or the presence of background in the matrix, or when recovery data are required, provision must be made for the inclusion of blank materials containing "none" (not detected) of the analyte. It is also important to know the variability of the blank and the tendency of the method to produce false positives. There are 2 types of blanks: matrix blanks and reagent blanks. Since laboratories often will utilize reagents from different sources, each laboratory should perform reagent blanks. Matrix blanks, when required, are an intrinsic part of the method, and the number of blanks needed depends on the combined variance of the material (s_M) and of the blank (s_B). Standard deviation reflecting the total variability of a blank corrected value will be $s = (s_M^2 + s_B^2)^{1/2}$.

3.4 Limit of Detection/Quantitation

If the limit of detection/quantitation is important, it is necessary to provide a design which gives special attention to the number of blanks, and to the necessity for interpreting false positives and false negatives. In all cases, the definition of limit of detection/quantitation used in the study must be given by the Study Director.

3.5 Controls

When separation from interferences is critical to the analysis, appropriate materials incorporating these interferences must be included.

PRACTICAL ADVICE: Always allow for contingencies and prepare more sets (e.g., 25% more) of laboratory samples than there are collaborators. Some packages may never arrive, some materials may spoil, and some may be lost or the container broken. New laboratories may have to be substituted for those which are unable to complete the promised work. Some sets may have to be analyzed at a later time for different purposes, such as to verify stability on storage.

4. Submission of Test Samples

4.1 Sending Collaborative Study Material

Notify collaborators of shipping arrangements, including waybill numbers, arrival time, and required storage conditions.

Label test samples legibly and without ambiguity.

Pack shipping cartons well and label properly to avoid transportation delays. If the containers are breakable, pack well to minimize possibility of breakage. If material is perishable, ship frozen with solid CO_2, sufficient to last several days longer than anticipated travel time. Use special transportation services, if necessary. For international delivery, mark as "Laboratory samples—no commercial value" or other designation as required by customs regulations of the country to which the package is being sent. Hazardous materials must be packed and labeled as required by transportation regulations. Animal and plant products sent across international borders may require special certification from health authorities.

Include a return slip, to confirm safe receipt, with each package. If not sent previously, include copy of method, instructions, and report forms.

Provide instructions for proper storage of test samples between unpacking and analysis. Note that analysts should not use thawed or decomposed test samples without consulting the Study Director.

When it is important to have instruments calibrated with the same reference material, supply reference material to collaborators. Provision for supplying reference standards is particularly important when commercial sources of standards have not yet been developed. The inclusion of a working standard solution as an unknown is useful to establish a consensus value for standardization of quality control parameters, such as absorptivity, retention time, and sensitivity (change in signal intensity divided by the change in concentration).

4.2 Obligations of Collaborators

Analyze test samples at times indicated, according to submitted protocol. With unstable materials (e.g., with microbial or decomposition problems), analyses must be started at specified times.

FOLLOW METHOD EXACTLY (*this is critical*). If method is unclear, contact Study Director. Any deviation, such as the necessity to substitute reagents, columns, apparatus, or instruments, must be recorded at the time and reported. If the collaborator has no intention of following the submitted method, he or she should not participate in the study. If the collaborator wishes to check another method on the same materials, additional test samples should be requested for that purpose, to be analyzed separately.

Conduct exactly the number of determinations stated in the instructions. Any other number complicates the statistical analysis. Too few determinations may require discarding the results from that laboratory for that material or inserting "missing values"; too many values may require discarding the contribution of that laboratory or at least some of the values. If a laboratory cannot follow instructions as to number of analyses to perform, it raises a question as to its ability to follow the method.

Report individual values, including blanks. Do not average or do other data manipulations unless required by the instructions. Undisclosed averaging distorts statistical measures. If blank is larger than determination, report the negative value; do not equate negative values to zero. Follow or request instructions with regard to reporting "traces" or "less than." Descriptive (i.e., nonquantitative) terms are not amenable to statistical analysis and should be avoided. When results are below the limit of determination, report actual calculated result, regardless of its value.

Supply raw data, graphs, recorder tracings, photographs, or other documentation as requested in the instructions.

Since collaborators may have no basis for judging whether a value is an outlier, the results should be communicated to the Study Director as soon as the protocol is complete and before time and equipment are reassigned, so that repeat assays may be performed at once, if necessary and if permitted by the protocol.

Note: The sooner an apparent outlier is investigated, the greater the likelihood of finding a reason for its occurrence.

The most frequent causes of correctable outliers are:

- Incorrect calculations and arithmetic errors.
- Errors in reporting, such as transposition of numbers, misplacement of the decimal point, or use of the wrong units.
- Incorrect standards due to weighing or volumetric errors (check physical constants or compare against freshly prepared standard solutions).
- Contamination of reagents, equipment, or test samples.

5. Statistical Analysis

5.1 Initial Review of Data (Data Audit)

The Study Director may first plot the collaborative study results, material by material (or one value against the other for a split level [Youden pair]), value vs laboratory, preferably in ascending or descending order of reported average concentration. Usually major discrepancies will be apparent: displaced means, unduly spread replicates, outlying values, differences between methods, consistently high or low laboratory rankings, etc.

Only valid data should be included in the statistical analysis. Valid data are values that the Study Director has no reason to suspect as being wrong. Invalid data may result when: (*1*) the method is not followed; (*2*) a nonlinear calibration curve is found although a linear curve is expected; (*3*) system suitability specifications were not met; (*4*) resolution is inadequate; (*5*) distorted absorption curves arise; (*6*) unexpected reactions occur; or (*7*) other atypical phenomena materialize. Other potential causes of invalid data are noted previously.

5.2 Outliers

Collaborative studies seem to have an inherent level of outliers, the number depending on the definition of outliers and the basis for calculation (analytes, materials, laboratories, or determinations). Rejection of more than 2/9 of the data from each material in a study, without an explanation (e.g., failure to follow the method), is ordinarily considered excessive. Study must maintain valid data from a minimum of 8 labs. For larger studies, a smaller acceptable percentage of rejections may be more appropriate. Determine the probability that the apparent aberrant value(s) is part of the main group of values considered as a normal population by applying the following tests in order:

(*1*) *Cochran test* for removal of laboratories (or indirectly for removal of extreme individual values from a set of laboratory values) showing significantly greater variability among replicate (within-laboratory) analyses than the other laboratories for a given material. Apply as a 1-tail test at a probability value of 2.5%.

To calculate the Cochran test statistic: Compute the within-laboratory variance for each laboratory and divide the largest of these by the sum of all of these variances. The resulting quotient is the Cochran statistic which indicates the presence of a removable outlier if this quotient exceeds the critical value listed in the Cochran table for P = 2.5% (1-tail) and L (number of laboratories), **Appendix 1**.

(*2*) Grubbs tests for removal of laboratories with extreme averages. Apply in the following order: single value test (2-tail; P = 2.5%); then if no outlier is found, apply pair value test (2 values at the highest end, 2 values at the lowest end, and 2 values, one at each end, at an overall P = 2.5%).

To calculate the single Grubbs test statistic: Compute the average for each laboratory and then calculate the standard deviation (SD) of these L averages (designate as the original s). Calculate the SD of the set of averages with the highest average removed (s_H); calculate the SD of the set averages with the lowest average removed (s_L). Then calculate the percentage decrease in SD as follows:

$$100 \times [1 - (s_L/s)] \text{ and } 100 \times [1 - (s_H/s)]$$

The higher of these 2 percentage decreases is the single Grubbs statistic, which signals the presence of an outlier to be omitted if it *exceeds* the critical value listed in the single Grubbs tables at the P = 2.5% level, 2-tail, for L laboratories, **Appendix 2**.

To calculate the Grubbs pair statistic, proceed in an analogous fashion, except calculate the standard deviations s_{2L}, s_{2H}, and s_{HL}, following removal of the 2 lowest, the 2 highest, and the highest and the lowest averages, respectively, from the original set of averages. Take the smallest of these 3 SD values and calculate the corresponding percentage decrease in SD from the original s. A Grubbs outlier pair is present if the selected value for the percentage decrease from the original s *exceeds* the critical value listed in the Grubbs pair value table at the P = 2.5% level, for L laboratories, **Appendix 2**.

(*3*) If the single value Grubbs test signals the need for outlier removal, remove the single Grubbs outlier and recycle back to the Cochran test as shown in the flow chart, **Appendix 3**.

If the single value Grubbs test is negative, check for masking by performing the pair value Grubbs test. If this second test is positive, remove the 2 values responsible for activating the test and recycle back to the Cochran test as shown in the flow chart, **Appendix 3**, and repeat the sequence of Cochran, single value Grubbs, and pair value Grubbs. Note, however, that outlier removal should stop before more than 2/9 laboratories are removed.

(*4*) If no outliers are removed for a given cycle (Cochran, single Grubbs, pair Grubbs), outlier removal is complete. Also, stop outlier removal whenever more than 2/9 of the laboratories are flagged for removal. With a higher removal rate, either the precision parameters must be taken without removal of all outliers or the method must be considered as suspect.

Note: The decision as to whether a value(s) should be removed as an outlier ultimately is not statistical in nature. The decision must be made by the Study Director on the basis of the indicated probability given by the outlier test and any other information that is pertinent. (However, for consistency with other organizations adhering to the harmonized outlier removal procedure, the estimate resulting from rigid adherence to the prescribed procedure should be reported.)

5.3 Bias (Systematic Deviation) of Individual Results

Bias is defined as follows:

(Estimated) bias =
mean amount found – amount added (or known or assigned value)

Single-value error and recovery are defined as follows:

Error of a single value =
the single value – amount added (true value)

There are 2 methods for defining percent recovery: marginal and total. The formulas used to estimate these percent recoveries are provided in the following:

$$\text{Marginal \%Rec} = 100R_M = 100((C_f - C_u)/C_A)$$

$$\text{Total \%Rec} = 100R_T = 100(C_f)/(C_u + C_A)$$

where C_f is the amount found for the fortified concentration, C_u is the amount present originally for the unfortified concentration, and C_A is the amount added for the added concentration. The amount added is known or fixed and should be a substantial fraction of, or more than, the amount present in the unfortified material; all other quantities are measured and are usually reported as means, all of which have variations or uncertainties. The variation associated with the marginal percent recovery is $\text{var}(100R_M) = (100^2/C_A^2)[\text{var}(C_f) + \text{var}(C_u)]$ is larger than the variation associated with the total percent recovery. The variation associated with total percent recovery is $\text{var}(100R_T) = [100^2/(C_u + C_A)^2][\text{var}(C_f) + (R_T^2)\text{var}(C_u)]$. In each formula var means variance and refers to the concentration variation for the defined concentrations.

A true or assigned value is known only in cases of spiked or fortified materials, certified reference materials, or by analysis by another (presumably unbiased) method. Concentration in the unfortified material is obtained by direct analysis by the method of additions. In other cases, there is no direct measure of bias, and consensus values derived from the collaborative study itself often must be used for the reference point.

Notes: (*1*) Youden equates "true" or "pure" between-laboratory variability (not including the within-laboratory variability) to the variability in bias (or variability in systematic error) of the individual laboratories. Technically, this definition refers to the average squared difference between individual laboratory biases and the mean bias of the assay.

(*2*) The presence of random error limits the ability to estimate the systematic error. To detect the systematic error of a single laboratory when the magnitude of such error is comparable to that laboratory's random error, at least 15 values are needed, under reasonable confidence limit assumptions.

5.4 Precision

The precision of analytical methods is usually characterized for 2 circumstances of replication: within laboratory or repeatability and among laboratories or reproducibility. Repeatability is a measure of how well an analyst in a given laboratory can check himself using the same analytical method to analyze the same test sample at the same time. Reproducibility is a measure of how well an analyst in one laboratory can check the results of another analyst in another laboratory using the same analytical method to analyze the same test sample at the same or different time. Given that test samples meet the criteria for a single material, the repeatability standard deviation (s_r) is:

$$s_r = (\Sigma d_i^2/2L)^{1/2}$$

where d_i is the difference between the individual values for the pair in laboratory i and L is the number of laboratories or number of pairs.

The reproducibility standard deviation (s_R) is computed as:

$$s_R = (1/2(s_d^2 + s_r^2))^{1/2}$$

where $s_d^2 = \Sigma(T_i - \bar{T})^2/(2(L-1))$, T_i is the sum of the individual values for the pair in laboratory i, $\bar{T}$ is the mean of the T_i across all laboratories or pairs, L is the number of laboratories or pairs, and s_r^2 is the square of $s_r = (\Sigma d_i^2/2L)^{1/2}$.

When the pairs of test samples meet the criteria for Youden matched pairs, i.e., when:

$$[(x_c - y_c)/x_c] \leq 0.05$$

or

$$y_c \geq (x_c - 0.05x_c),$$

s_r, a practical approximation for repeatability standard deviation, is calculated as:

$$s_r = [\Sigma(d_i - \bar{d})^2/(2(L-1))]^{1/2}$$

where d_i is the difference between the individual values for the pair in laboratory i, $\bar{d}$ is the mean of the d_i across all laboratories or pairs, and L is the number of laboratories or pairs. The reproducibility standard deviation, s_R, which reflects the square root of the average of the reproducibility variances for the individual materials (i.e., $s_R = [½(s_{Rx}^2 + s_{Ry}^2)]^{1/2}$), previously called X and Y, should be determined only if the individual variances are not significantly different from each other. To compare s_{Rx}^2 and s_{Ry}^2, the following formula may be used.

$$t = \frac{(s_{Rx}^2 - s_{Ry}^2)(L-2)^{1/2}}{2[(s_{Rx}^2)(s_{Ry}^2) - (cov_{xy})^2]^{1/2}}$$

where $s_{Rx}^2 = [1/(L-1)][\Sigma x_i^2 - (\Sigma x_i)^2/L]$, $s_{Ry}^2 = [1/(L-1)][\Sigma y_i^2 - (\Sigma y_i)^2/L]$, and $cov_{xy} = [1/(L-1)][\Sigma x_i y_i - (\Sigma x_i \Sigma y_i)/L]$. If t is greater than or equal to the tabular t-value for L – 2 degrees of freedom for a significance level of $\alpha = 0.05$, this may be taken to indicate that s_{Rx}^2 and s_{Ry}^2 are not equivalent and should not be pooled for a single estimate of s_R^2. That is, s_{Rx}^2 and s_{Ry}^2 should be taken as the reproducibility variance estimates for the individual test materials X and Y, respectively. This means that there is no rigorous basis for calculating s_r^2 because the within laboratory variability cannot be estimated directly.

Though s_r and s_R are the most important types of precision, it is the relative standard deviations ($RSD_r\% = 100s_r/\text{mean}$ and $RSD_R\% = 100s_R/\text{mean}$) that are the most useful measures of precision in chemical analytical work because the RSD values are usually independent of concentration. Therefore, the use of the RSD values facilitates comparison of variabilities at different concentrations. When the RSD increases rapidly with decreasing concentration or amount, the rise delineates the limit of usefulness of the method (limit of reliable measurement).

5.5 HorRat

HorRat value is the ratio of the reproducibility relative standard deviation, expressed as a percent (RSD_R, %) to the predicted reproducibility relative standard deviation, expressed as a percent ($PRSD_R$, %), i.e.,

$$\text{HorRat} = \frac{RSD_R,\%}{PRSD_R,\%}$$

where $PRSD_R$, % = $2C^{-0.1505}$ and C = the estimated mean concentration expressed as a decimal fraction (i.e., 100% = 1; 1% = 0.01; 1 ppm = 0.000001). HorRat values between 0.5 to 1.5 may be taken to indicate that the performance value for the method corresponds to historical performance. The limits for performance acceptability are 0.5–2.

The precision of a method must be presented in the collaborative study manuscript. The HorRat will be used as a guide to determine the acceptability of the precision of a method.

The HorRat is applicable to most chemical methods. HorRat is not applicable to physical properties (viscosity, RI, density, pH, absorbance, etc.) and empirical methods [e.g., fiber, enzymes, moisture, methods with indefinite analytes (e.g., polymers) and "quality" measurements, e.g., drained weight]. Deviations may also occur at both extremes of the concentration scale (near 100% and $\lesssim 10^{-8}$). In areas where there is a question if the HorRat is applicable, the General Referee will be the determining judge.

The following guidelines should be used to evaluate the assay precision:

- HorRat ≤ 0.5—Method reproducibility may be in question due to lack of study independence, unreported averaging, or consultations.
- 0.5 < HorRat ≤ 1.5—Method reproducibility as normally would be expected.
- HorRat > 1.5—Method reproducibility higher than normally expected: the Study Director should critically look into possible reasons for a "high" HorRat (e.g., were test samples sufficiently homogeneous, indefinite analyte or property?), and discuss this in the collaborative study report.
- HorRat > 2.0—Method reproducibility is problematic. A high HorRat may result in rejection of a method because it may indicate unacceptable weaknesses in the method or the study. Some organizations may use information about the HorRat as a criterion not to accept the method for official purposes (e.g., this is currently the case in the EU for aflatoxin methods for food analysis, where only methods officially allowed are those with HorRats ≤ 2).

5.6 Incorrect, Improper, or Illusory Values (False Positive and False Negative Values)

These results are not necessarily outliers (no *a priori* basis for decision), since there is a basis for determining their incorrectness (a positive value on a blank material, or a zero (not found) or negative value on a spiked material). There is a statistical basis for the presence of false negative values: In a series of materials with decreasing analyte concentration, as the RSD increases, the percent false negatives increases from an expected 2% at an RSD = 50% to 17% at an RSD = 100%, merely from normal distribution statistics alone.

When false positives and/or false negatives exceed about 10% of all values, analyses become uninterpretable from lack of confidence in the presence or absence of the analyte, unless all positive laboratory samples are re-analyzed by a more reliable (confirmatory) method with a lower limit of determination than the method under study. When the proportion of zeros (not necessarily

false negatives) becomes greater than approximately 30%, the distribution can become bimodal and even more uninterpretable (is the analyte present or absent?).

5.7 Final Collaborative Study Manuscript

The final manuscript should contain a description of the materials used, their preparation, any unusual features in their distribution, and a table of all *valid* data, including outliers. When replication is performed, the individual values, not just averages, must be given, unless the method requires averages (e.g., microbiological methods). Values not used for specified reasons, such as decomposition, failure to follow method, or contamination, should not be included in the table since they may be included erroneously in subsequent recalculations. AOAC INTERNATIONAL requires the calculation and reporting of mean, percent recovery (% Rec), HorRat, repeatability (within-laboratory, s_r) and reproducibility (interlaboratory, s_R) standard deviations, and repeatability and reproducibility relative standard deviations (RSD_r and RSD_R, respectively). The accuracy (bias, trueness) of a method measuring a specific, identifiable analyte should be presented in the collaborative study manuscript as a recovery of added (spiked) analyte, as the results of analysis of a reference material, or by comparison with results by a reference method. Methods that are unable to report accuracy because of the unavailability of an accepted "true" value, or because of the nature of the method (empirical, microbiological, quality factors) should mention the reason in the manuscript. Proofread tables very carefully since many errors are of typographical origin. Give the names of the participants and their organizations, including complete contact information (name, preliminary address, telephone and fax numbers, and e-mail address).

The final manuscript should be published in a generally accessible publication, or availability of the report from the organization sponsoring the method should be indicated in the published method. Without public documentation, the significance of the study is very limited.

The manuscript should be sent to all participants, preferably at the preliminary stage, so that clerical and typographical errors may be corrected before publication. If changes in values from the original submission are offered, they must be accompanied by an explanation.

Example of Table of Interlaboratory Study Results: *See* **Table 1**.

The summary table as it will appear in the *Official Methods of Analysis of AOAC INTERNATIONAL* is given in **Table 2**.

6. References

(1) W.J. Youden & E.H. Steiner (1975) *Statistical Manual of the AOAC*, AOAC INTERNATIONAL, 481 N. Frederick Ave, Suite 500, Gaithersburg, MD 20877-7077, USA. The fifth printing (1987) contains several explanatory footnotes.

(2) G.T. Wernimont (1985) *Use of Statistics to Develop and Evaluate Analytical Methods*, W. Spendley (Ed.) AOAC INTERNATIONAL, 481 N. Frederick Ave, Suite 500, Gaithersburg, MD 20877-7077, USA.

(3) T. Dols & B. Armbrecht (1976) *J. Assoc. Off. Anal. Chem.* **59**, 1204–1207.

(4) International Organization for Standardization Guide 18, ISO, Case Postale 56, CH-1211 Geneva, Switzerland, and other national standards organizations.

(5) International Organization for Standardization ISO 5725, ISO, Case Postale 56, CH-1211 Geneva, Switzerland, and other national standards organizations.

Appendix 1. Critical values for the Cochran maximum variance ratio at the 2.5% (1-tail) rejection level, expressed as the percentage the highest variance is of the total variance

L = number of laboratories at a given level (concentration)
r = number of replicates per laboratory

L	r = 2	r = 3	r = 4	r = 5	r = 6
4	94.3	81.0	72.5	65.4	62.5
5	88.6	72.6	64.6	58.1	53.9
6	83.2	65.8	58.3	52.2	47.3
7	78.2	60.2	52.2	47.3	42.3
8	73.6	55.6	47.4	43.0	38.5
9	69.3	51.8	43.3	39.3	35.3
10	65.5	48.6	39.9	36.2	32.6
11	62.2	45.8	37.2	33.6	30.3
12	59.2	43.1	35.0	31.3	28.3
13	56.4	40.5	33.2	29.2	26.5
14	53.8	38.3	31.5	27.3	25.0
15	51.5	36.4	29.9	25.7	23.7
16	49.5	34.7	28.4	24.4	22.0
17	47.8	33.2	27.1	23.3	21.2
18	46.0	31.8	25.9	22.4	20.4
19	44.3	30.5	24.8	21.5	19.5
20	42.8	29.3	23.8	20.7	18.7
21	41.5	28.2	22.9	19.9	18.0
22	40.3	27.2	22.0	19.2	17.3
23	39.1	26.3	21.2	18.5	16.6
24	37.9	25.5	20.5	17.8	16.0
25	36.7	24.8	19.9	17.2	15.5
26	35.5	24.1	19.3	16.6	15.0
27	34.5	23.4	18.7	16.1	14.5
28	33.7	22.7	18.1	15.7	14.1
29	33.1	22.1	17.5	15.3	13.7
30	32.5	21.6	16.9	14.9	13.3
35	29.3	19.5	15.3	12.9	11.6
40	26.0	17.0	13.5	11.6	10.2
50	21.6	14.3	11.4	9.7	8.6

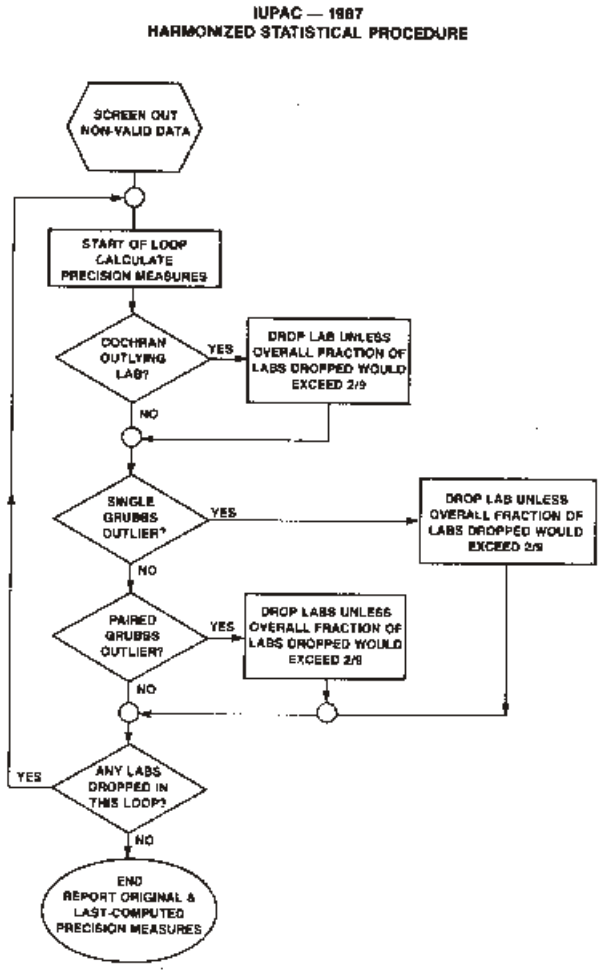
Cochran statistic = (largest individual within-laboratory variance)/(sum of all the within-laboratory variances).

Appendix 2. Critical values for the Grubbs extreme deviation outlier tests at the 2.5% (2-tail), 1.25% (1-tail) rejection level, expressed as the percent reduction in the standard deviations caused by removal of the suspect value(s) (*see* text for calculating the Grubbs statistics)

L = number of laboratories at a given level (concentration)

L	One highest or lowest	Two highest or two lowest	One highest and one lowest
4	86.1	98.9	99.1
5	73.5	90.3	92.7
6	64.0	81.3	84.0
7	57.0	73.1	76.2
8	51.4	66.5	69.6
9	46.8	61.0	64.1
10	42.8	56.4	59.5
11	39.3	52.5	55.5
12	36.1	48.5	51.6
13	33.8	46.1	49.1
14	31.7	43.5	46.5
15	29.9	41.2	44.1
16	28.3	39.2	42.0
17	26.9	37.4	40.1
18	25.7	35.9	38.4
19	24.6	34.5	36.9
20	23.6	33.2	35.4
21	22.7	31.9	34.0
22	21.9	30.7	32.8
23	21.2	29.7	31.8
24	20.5	28.8	30.8
25	19.8	28.0	29.8
26	19.1	27.1	28.9
27	18.4	26.2	28.1
28	17.8	25.4	27.3
29	17.4	24.7	26.6
30	17.1	24.1	26.0
40	13.3	19.1	20.5
50	11.1	16.2	17.3

Source: Both tables were calculated by R. Albert (October 1993) by computer simulation involving several runs of approximately 7000 cycles each for each value, and then smoothed. Although the table of **Appendix 1** is strictly applicable only to a balanced design (same number of replicates from all laboratories), it can be applied to an unbalanced design without too much error, if there are only a few deviations.



Appendix 3. Flowchart.

Table 1. [x] Collaborative tests carried out at the international level in [year(s)] by [organization(s)] in which [y and z] laboratories participated, each performing [k] replicates, gave the following statistical results [results expressed in (units)]:

Material [description and listed across the top in increasing order of magnitude of means]
Number of laboratories retained after eliminating outliers
Number of outlying laboratories removed

Mean ($\bar{x}$)
True or accepted value, if known

Repeatability standard deviation (s_r)
Repeatability relative standard deviation (RSD_r)
Repeatability value, r ($2.8 \times s_r$)
Total within laboratory standard deviation (s_e)—optional if s_r is not valid.

Reproducibility standard deviation (s_R)
Reproducibility relative standard deviation (RSD_R)
HorRat
Reproducibility value, R ($2.8 \times s_R$)
Percent recovery (% Rec), if applicable

The repeatability and reproducibility values may also be expressed as a relative value (as a percentage of the determined mean value), when the results so suggest.
If the recovery and precision values are more or less constant for all materials or for group of materials, an overall average value may be presented. Although such averaging may not have statistical validity, it does have practical value.

Table 2. Model table for presentation of chemistry results from AOAC Official Methods

Table 200X.XX Interlaboratory results for [analyte] by [technique]

Material		No. of labs[a(b)]	Mean (units)	Recovery, %	Repeatabiltiy RSD_r, %	Reproducibility	
Matrix	Level (units)					RSD_R, %	HorRat

[a(b)] a = Number of laboratories remaining after removal of the number of outliers indicated by (*b*).

Appendix F: Guidelines for Standard Method Performance Requirements

Contents

Introduction to Standard Method Performance Requirements 1

Annex A: Format of a Standard Method Performance Requirement 5

Annex B: Classification of Methods 11

Annex C: Understanding the POD Model 12

Annex D: Definitions and Calculations of HorRat Values from Intralaboratory Data 13

Annex E: AOAC Method Accuracy Review 15

Annex F: Development and Use of In-House Reference Materials 16

Introduction to Standard Method Performance Requirements

Standard method performance requirements (SMPRs) are a unique and novel concept for the analytical methods community. SMPRs are voluntary consensus standards, developed by stakeholders, that prescribe the minimum analytical performance requirements for classes of analytical methods. In the past, analytical methods were evaluated and the results compared to a "gold standard" method, or if a gold standard method did not exist, then reviewers would decide retrospectively if the analytical performance was acceptable. Frequently, method developers concentrated on the process of evaluating the performance parameters of a method, and rarely set acceptance criteria. However, as the *Eurachem Guide* points out: " . . . the judgment of method suitability for its intended use is equally important . . ." (1) to the evaluation process.

International Voluntary Consensus Standards

An SMPR is a form of an international, voluntary consensus standard. A standard is an agreed, repeatable way of doing something that is published as document that contains a technical specification or other precise criteria designed to be used consistently as a rule, guideline, or definition. SMPRs are a *consensus* standards developed by stakeholders in a very controlled process that ensures that users, research organizations, government departments, and consumers work together to create a standard that meets the demands of the analytical community and technology. SMPRs are also *voluntary* standards. AOAC cannot, and does not, impose the use of SMPRs. Users are free to use SMPRs as they see fit. AOAC is very careful to include participants from as many regions of the world as possible so that SMPRs are accepted as *international* standards.

Guidance for Standard Method Performance Requirements

Commonly known as the "SMPR Guidelines." The first version of the SMPR Guidelines were drafted in 2010 in response to the increasing use and popularity of SMPRs as a vehicle to describe the analytical requirements of a method. Several early "acceptance criteria" documents were prepared for publication in late 2009, but the format of the acceptance criteria documents diverged significantly from one another in basic format. AOAC realized that a guidance document was needed to promote uniformity.

An early version of the SMPR Guidelines were used for a project to define the analytical requirements for endocrine disruptors in potable water. The guidelines proved to be extremely useful in guiding the work of the experts and resulted in uniform SMPRs. Subsequent versions of the SMPR Guidelines were used in the Stakeholder Panel for Infant Formula and Adult Nutritionals (SPIFAN) project with very positive results. The SMPR Guidelines are now published for the first time in the *Journal of AOAC INTERNATIONAL* and *Official Methods of Analysis*.

Users of the guidelines are advised that they are: (*1*) a *guidance* document, not a statute that users must conform to; and (*2*) a "living" document that is regularly updated, so users should check the AOAC website for the latest version before using these guidelines.

The SMPR Guidelines are intended to provide basic information for working groups assigned to prepare SMPRs. The guidelines consist of the standard format of an SMPR, followed by a series of informative tables and annexes.

SMPR Format

The general format for an SMPR is provided in *Annex A*.

Each SMPR is identified by a unique SMPR number consisting of the year followed by a sequential identification number (YYYY.XXX). An SMPR number is assigned when the standard is approved. By convention, the SMPR number indicates the year a standard is approved (as opposed to the year the standard is initiated). For example, SMPR 2010.003 indicates the third SMPR adopted in 2010.

The SMPR number is followed by a method name that must include the analyte(s), matrix(es), and analytical technique (unless the SMPR is truly intended to be independent of the analytical technology). The method name may also refer to a "common" name (e.g., "Kjeldahl" method).

The SMPR number and method name are followed by the name of the stakeholder panel or expert review panel that approved the SMPR, and the approval and effective dates.

Information about method requirements is itemized into nine categories: (*1*) intended use; (*2*) applicability; (*3*) analytical technique; (*4*) definitions; (*5*) method performance requirements; (*6*) system suitability; (*7*) reference materials; (*8*) validation guidance; and (*9*) maximum time-to-determination.

An SMPR for qualitative and/or identification methods may include up to three additional annexes: (*1*) inclusivity/selectivity panel; (*2*) exclusivity/cross-reactivity panel; and (*3*) environmental material panels. These annexes not required.

Informative tables.—The SMPR Guidelines contain seven informative tables that represent the distilled knowledge of many years of method evaluation, and are intended as guidance for SMPR working groups. The informative tables are not necessarily AOAC

policy. SMPR working groups are expected to apply their expertise in the development of SMPRs.

Table A1: Performance Requirements. Provides recommended performance parameters to be included into an SMPR. Table A1 is organized by five method classifications: (*1*) main component quantitative methods; (*2*) trace or contaminant quantitative methods; (*3*) main component qualitative methods; (*4*) trace or contaminant quantitative methods; and (*5*) identification methods. The table is designed to accommodate both microbiological and chemical methods. Alternate microbiological/chemical terms are provided for equivalent concepts.

Table A2: Recommended Definitions. Provides definitions for standard terms in the SMPR Guidelines. AOAC relies on *The International Vocabulary of Metrology Basic and General Concepts and Associated Terms* (VIM) and the International Organization for Standadization (ISO) for definition of terms not included in Table A2.

Table A3: Recommendations for Evaluation. Provides general guidance for evaluation of performance parameters. More detailed evaluation guidance can be found in *Appendix D, Guidelines for Collaborative Study Procedures to Validate Characteristics of a Method of Analysis* (2); *Appendix I, Guidelines for Validation of Biological Threat Agent Methods and/or Procedures* (3); *Appendix K, AOAC Guidelines for Single-Laboratory Validation of Chemical Methods for Dietary Supplements and Botanicals* (4); Codex Alimentarius Codex Procedure Manual (5); and ISO Standard 5725-1-1994 (6).

Table A4: Expected Precision (Repeatability) as a Function of Analyte Concentration. The precision of a method is the closeness of agreement between independent test results obtained under stipulated conditions. Precision is usually expressed in terms of imprecision and computed as a relative standard deviation (RSD) of the test results. The imprecision of a method increases as the concentration of the analyte decreases. This table provides target RSDs for a range of analyte concentrations.

Table A5: Expected Recovery as a Function of Analyte Concentration. Recovery is defined as the ratio of the observed mean test result to the true value. The range of the acceptable mean recovery expands as the concentration of the analyte decreases. This table provides target mean recovery ranges for analyte concentrations from 1 ppb to 100%.

Table A6: Predicted Relative Standard Deviation of Reproducibility ($PRSD_R$). This table provides the calculated $PRSD_R$ using the Horwitz formula:

$$PRSD_R = 2C^{-0.15}$$

where C is expressed as a mass fraction.

Table A7: POD and Number of Test Portions. This table provides the calculated probability of detection (POD) for given sample sizes and events (detections). A method developer can use this table to determine the number of analyses required to obtain a specific POD.

Informative annexes.—The SMPR Guidelines contain informative annexes on the topics of classification of methods, POD model, HorRat values, reference materials, and method accuracy and review. As with the informative tables, these annexes are intended to provide guidance and information to the working groups.

Initiation of an SMPR

See Figure 1 for a schematic flowchart diagram of the SMPR development process.

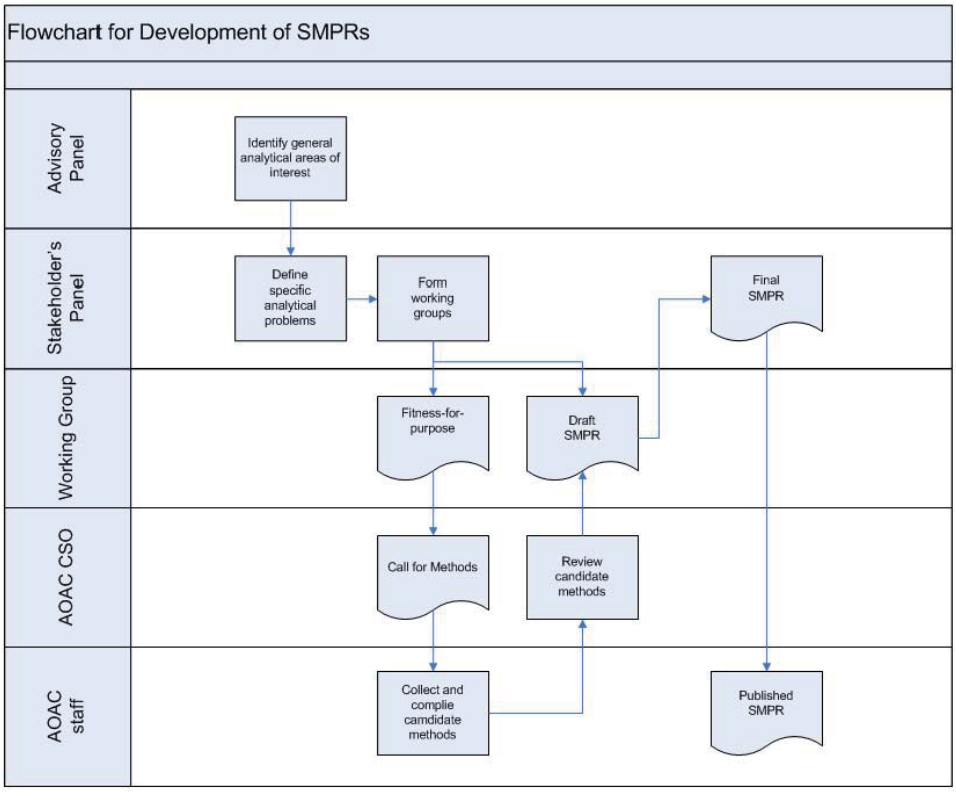


Figure 1. Schematic flowchart diagram of the SMPR development process.

Advisory panels.—Most commonly, an SMPR is created in response to an analytical need identified by an advisory panel. Advisory panels normally consist of sponsors and key stakeholders who have organized to address analytical problems. Usually, the advisory panel identifies general analytical problems, such as the need to update analytical methods for determination of nutrients in infant formula. An advisory panel, with the input of appropriate subject matter experts, also prioritizes the specific analytical problems within the general topic. This panel is critical in planning for the stakeholder panel meeting.

Stakeholder panels.—After an advisory panel has identified a general analytical problem, AOAC announces the standards development activity, identifies stakeholders, and organizes a stakeholder panel. Membership on a stakeholder panel is open to anyone materially affected by the proposed standard. AOAC recruits scientists to participate on stakeholder panels on the basis of their expertise with the analytical problem identified by the advisory panel. Experts are recruited from academia, government, nongovernmental organizations (such as ISO), industry, contract research organizations, method developers, and instrument/equipment manufacturers. AOAC employs a representative voting panel model to ensure balance with regards to stakeholder perspective, and to ensure that no particular stakeholder perspective dominates the proceedings of the stakeholder panel. All stakeholder candidates are reviewed by the AOAC Chief Scientific Officer (CSO) for relevant qualifications, and again by the Official Methods Board to ensure that the stakeholder panel is balanced and all stakeholders are fairly represented.

Stakeholder panels are extremely important as they serve several functions: (*1*) identify specific analytical topics within the general analytical problem described by the advisory panel; (*2*) form working groups to address the specific analytical topics; (*3*) identify additional subject matter experts needed for the working groups; (*4*) provide oversight of the SMPR development; and (*5*) formally adopt SMPRs originally drafted by working groups.

Working groups.—Working groups are formed by the stakeholder panel when a specific analytical topic has been identified. The primary purpose of a working group is to draft an SMPR. Working groups may also be formed to make general recommendations, such as developing a common definition to be used by multiple working groups. For example, SPIFAN formed a working group to create a definition for "infant formula" that could be shared and used by all of the SPIFAN working groups.

The process of drafting an SMPR usually requires several months, and several meetings and conference calls. An SMPR drafted by a working group is presented to a stakeholder panel. A stakeholder panel may revise, amend, or adopt a proposed SMPR on behalf of AOAC.

Fitness-for-Purpose Statement and Call for Methods

One of the first steps in organizing a project is creating a fitness-for-purpose statement. In AOAC, the fitness-for-purpose statement is a very general description of the methods needed. It is the responsibility of a working group chair to draft a fitness-for-purpose statement. A working group chair is also asked to prepare a presentation with background information about the analyte, matrix, and the nature of the analytical problem. A working group chair presents the background information and proposes a draft fitness-for-purpose statement to the presiding stakeholder panel. The stakeholder panel is asked to endorse the fitness-for-purpose statement.

The AOAC CSO prepares a call for methods based on the stakeholder panel-approved fitness-for-purpose statement. The call for methods is posted on the AOAC website and/or e-mailed to the AOAC membership and other known interested parties. AOAC staff collects and compiles candidate methods submitted in response to the call for methods. The CSO reviews and categorizes the methods.

Creating an SMPR

Starting the process of developing an SMPR can be a daunting challenge. In fact, drafting an SMPR should be a daunting challenge because the advisory panel has specifically identified an analytical problem that has yet to be resolved. Completing an SMPR can be a very rewarding experience because working group members will have worked with their colleagues through a tangle of problems and reached a consensus where before there were only questions.

It is advisable to have some representative candidate methods available for reference when a working group starts to develop an SMPR. These methods may have been submitted in response to the call for methods, or may be known to a working group member. In any case, whatever the origin of the method, candidate methods may assist working group members to determine reasonable performance requirements to be specified in the SMPR. The performance capabilities of exisiting analytical methodologies is a common question facing a working group.

Normally, a working chair and/or the AOAC CSO prepares a draft SMPR. A draft SMPR greatly facilitates the process and provides the working group with a structure from which to work.

Working group members are advised to first consider the "intended use" and "maximum time-to-determination" sections as this will greatly affect expectations for candidate methods. For example, methods intended to be used for surveillance probably need to be quick but do not require a great deal of precision, and false-positive results might be more tolerable. Whereas methods intended to be used for dispute resolution will require better accuracy, precision, and reproducibility, but time to determination is not as important.

Once a working group has agreed on the intended use of candidate methods, then it can begin to define the applicability of candidate methods. The applicability section of the SMPR is one of the most important, and sometimes most difficult, sections of the SMPR. The analyte(s) and matrixes must be explicitly identified. For chemical analytes, International Union of Pure and Applied Chemistry (IUPAC) nomenclature and/or Chemical Abstracts Service (CAS) registry numbers should be specified. Matrixes should be clearly identified including the form of the matrix such as raw, cooked, tablets, powders, etc. The nature of the matrix may affect the specific analyte. It may be advantageous to fully identify and describe the matrix before determining the specific analyte(s). It is not uncommon for working groups to revise the initial definition of the analyte(s) after the matrix(es) has been better defined.

Table 1. Example of method performance table for a single analyte

Analytical range	7.0–382.6 µg/mL	
Limit of quantitation (LOQ)	≤7.0 µg/mL	
Repeatability (RSD_r)	<10 µg/mL	≤8%
	≥10 µg/mL	≤6%

Table 2. Example of method performance table for multiple analytes

	Analyte 1		Analyte 2		Analyte 3	
Analytical range	10–20 µg/mL		100–200 µg/mL		200–500 µg/mL	
Limit of quantitation (LOQ)	≤10 µg/mL		≤100 µg/mL		≤200 µg/mL	
Repeatability (RSD_r)	<10 µg/mL	≤8%	<10 µg/mL	≤8%	<200 µg/mL	≤10%
	≥10 µg/mL	≤6%	≥10 µg/mL	≤6%	≥200 µg/mL	≤8%

For projects with multiple analytes, for example, vitamins A, D, E, and K in infant formula, it may be useful to organize a separate working group to fully describe the matrix(es) so that a common description of the matrix(es) can be applied to all of the analytes.

For single analyte SMPRs, it is most common to organize the method performance requirements into a table with 2–3 columns as illustrated in Table 1. For multiple analyte SMPRs, it is often convenient to present the requirements in an expanded table with analytes forming additional columns as illustrated in Table 2.

Once the intended use, analytical techniques, and method performance requirements have been determined, then a working group can proceed to consider the quality control parameters, such as the minimum validation requirements, system suitability procedures, and reference materials (if available). It is not uncommon that an appropriate reference material is not available. *Annex F* of the SMPR Guidelines provides comprehensive guidance for the development and use of in-house reference materials.

Most working groups are able to prepare a consensus SMPR in about 3 months.

Open Comment Period

Once a working group has produced a draft standard, AOAC opens a comment period for the standard. The comment period provides an opportunity for other stakeholders to state their perspective on the draft SMPR. All collected comments are reviewed by the AOAC CSO and the working group chair, and the comments are reconciled. If there are significant changes required to the draft standard as a result of the comments, the working group is convened to discuss and any unresolved issues will be presented for discussion at the stakeholder panel meeting.

Submission of Draft SMPRs to the Stakeholder Panel

Stakeholder panels meet several times a year at various locations. The working group chair (or designee) presents a draft SMPR to the stakeholder panel for review and discussion. A working group chair is expected to be able to explain the conclusions of the working group, discuss comments received, and to answer questions from the stakeholder panel. The members of the stakeholder panel may revise, amend, approve, or defer a decision on the proposed SMPR. A super majority of 2/3 or more of those voting is required to adopt an SMPR as an AOAC voluntary consensus standard.

Publication

Adopted SMPRs are prepared for publication by AOAC staff, and are published in the *Journal of AOAC INTERNATIONAL* and in the AOAC *Official Methods of Analysis*℠ compendium. Often, the AOAC CSO and working group chair prepare a companion article to introduce an SMPR and describe the analytical issues considered and resolved by the SMPR. An SMPR is usually published within 6 months of adoption.

Conclusion

SMPRs are a unique and novel concept for the analytical methods community. SMPRs are voluntary, consensus standards developed by stakeholders that prescribe the minimum analytical performance requirements for classes of analytical methods. The SMPR Guidelines provide a structure for working groups to use as they develop an SMPR. The guidelines have been employed in several AOAC projects and have been proven to be very useful. The guidelines are not a statute that users must conform to; they are a "living" document that is regularly updated, so users should check the AOAC website for the latest version before using the guidelines.

References

(1) Eurachem, *The Fitness for Purpose of Analytical Methods: A Laboratory Guide to Method Validation and Related Topics, Validation*, http://www.eurachem.org/guides/pdf/valid.pdf, posted December 1998, accessed March 2012

(2) *Guidelines for Collaborative Study Procedures to Validate Characteristics of a Method of Analysis* (2012) *Official Methods of Analysis, Appendix D*, AOAC INTERNATIONAL, Gaithersburg, MD

(3) *AOAC INTERNATIONAL Methods Committee Guidelines for Validation of Biological Threat Agent Methods and/or Procedures* (2012) *Official Methods of Analysis,* 19th Ed., *Appendix I, Calculation of CPOD and dCPOD Values from Qualitative Method Collaborative Study Data*, AOAC INTERNATIONAL, Gaithersburg, MD

(4) *AOAC Guidelines for Single-Laboratory Validation of Chemical Methods for Dietary Supplements and Botanicals* (2012) *Official Methods of Analysis*, 19th Ed., *Appendix K*, AOAC INTERNATIONAL, Gaithersburg, MD

(5) Codex Alimentarius Codex Procedure Manual

(6) International Organization for Standardization, Geneva, Switzlerland

ANNEX A
Format of a
Standard Method Performance Requirement

AOAC SMPR YYYY.XXX
(YYYY = Year; XXX = sequential identification number)

Method Name: Must include the analyte(s), matrix(es), and analytical technique [unless the standard method performance requirement (SMPR) is truly intended to be independent of the analytical technology]. The method name may refer to a "common" name (e.g., "Kjeldahl" method).

Approved By: Name of stakeholder panel or expert review panel

Final Version Date: Date

Effective Date: Date

1. Intended Use: Additional information about the method and conditions for use.

2. Applicability: List matrixes if more than one. Provide details on matrix such as specific species for biological analytes, or International Union of Pure and Applied Chemistry (IUPAC) nomenclature and Chemical Abstracts Service (CAS) registry number for chemical analytes. Specify the form of the matrix such as raw, cooked, tablets, powders, etc.

3. Analytical Technique: Provide a detailed description of the analytical technique if the SMPR is to apply to a specific analytical technique; or state that the SMPR applies to any method that meets the method performance requirements.

4. Definitions: List and define terms used in the performance parameter table (*see* Table A2 for list of standard terms).

5. Method Performance Requirements: List the performance parameters and acceptance criteria appropriate for each method/analyte/matrix. *See* Table A1 for appropriate performance requirements.

If more than one analyte/matrix, and if acceptance criteria differ for analyte/matrix combinations then organize a table listing each analyte/matrix combination and its minimum acceptance criteria for each performance criteria.

6. System Suitability Tests and/or Analytical Quality Control: Describe minimum system controls and QC procedures.

7. Reference Material(s): Identify the appropriate reference materials if they exist, or state that reference materials are not available. Refer to *Annex E* (*AOAC Method Accuracy Review*) for instructions on the use of reference materials in evaluations.

8. Validation Guidance: Recommendations for type of evaluation or validation program such as single-laboratory validation (SLV), *Official Methods of Analysis*℠ (OMA), or *Performance Tested Methods*℠ (PTM).

9. Maximum Time-to-Determination: Maximum allowable time to complete an analysis starting from the test portion preparation to final determination or measurement.

Annex I: Inclusivity/Selectivity Panel. Recommended for qualitative and identification method SMPRs.

Annex II: Exclusivity/Cross-Reactivity Panel. Recommended for qualitative and identification method SMPRs.

Annex III: Environmental Materials Panel. Recommended for qualitative and identification method SMPRs.

Table A1. Performance requirements

Classifications of methods[a]				
Quantitative method		Qualitative method		
Main component[b]	Trace or contaminant[c]	Main component[b]	Trace or contaminant[c]	Identification method
Parameter				
Single-laboratory validation				
Applicable range	Applicable range	Inclusivity/selectivity	Inclusivity/selectivity	Inclusivity/selectivity
Bias[d]	Bias[d]	Exclusivity/cross-reactivity	Exclusivity/cross-reactivity	Exclusivity/cross-reactivity
Precision	Precision	Environmental interference	Environmental interference	Environmental interference
Recovery	Recovery	Laboratory variance	Laboratory variance	
Limit of quantitation (LOQ)	LOQ			
		Probability of detection (POD)[e]	POD at AMDL[f]	Probability of identification (POI)
Reproducibility				
RSD_R or target measurement uncertainty	RSD_R or target measurement uncertainty	POD (0)	POD (0)	POI (c)
		POD (c)	POD (c)	
		Laboratory POD[g]	Laboratory POD[g]	Laboratory POI

[a] *See Annex B* for additional information on classification of methods.

[b] ≥100 g/kg.

[c] <100 g/kg.

[d] If a reference material is available.

[e] At a critical level.

[f] AMDL = Acceptable minimum detection level.

[g] LPOD = CPOD.

Table A2. Recommended definitions

Bias	Difference between the expectation of the test results and an accepted reference value. Bias is the total systematic error as contrasted to random error. There may be one or more systematic error components contributing to the bias.
Environmental interference	Ability of the assay to detect target organism in the presence of environmental substances and to be free of cross reaction from environmental substances.
Exclusivity	Strains or isolates or variants of the target agent(s) that the method must not detect.
Inclusivity	Strains or isolates or variants of the target agent(s) that the method can detect.
Laboratory probability of detection (POD)	Overall fractional response (mean POD = CPOD) for the method calculated from the pooled POD_j responses of the individual laboratories (j = 1, 2, ..., L).[a] *See Annex C*.
Limit of quantitation (LOQ)	Minimum concentration or mass of analyte in a given matrix that can be reported as a quantitative result.
POD (0)	Probability of the method giving a (+) response when the sample is truly without analyte.
POD (c)	Probability of the method giving a (–) response when the sample is truly without analyte.
POD	Proportion of positive analytical outcomes for a qualitative method for a given matrix at a given analyte level or concentration. Consult *Annex C* for a full explanation.
Probability of identification (POI)	Expected or observed fraction of test portions at a given concentration that gives positive result when tested at a given concentration. Consult *Probability of Identification (POI): A Statistical Model for the Validation of Qualitative Botanical Identification Methods*.[c]
Precision (repeatability)	Closeness of agreement between independent test results obtained under stipulated conditions. The measure of precision is usually expressed in terms of imprecision and computed as a standard deviation of the test results.[d]
Recovery	Fraction or percentage of the analyte that is recovered when the test sample is analyzed using the entire method. There are two types of recovery: (*1*) Total recovery based on recovery of the native plus added analyte, and (*2*) marginal recovery based only on the added analyte (the native analyte is subtracted from both the numerator and denominator).[e]
Repeatability	Precision under repeatability conditions.
Repeatability conditions	Conditions where independent test results are obtained with the same method on identical test items in the same laboratory by the same operator using the same equipment within short intervals of time.
Reproducibility	Precision under reproducibility conditions.
Reproducibility conditions	Conditions where independent test results are obtained with the same method on identical test items in different laboratories with different operators using different equipment.
Relative standard deviation (RSD)	$RSD = s_i \times 100/\bar{x}$
Standard deviation (s_i)	$s_i = [\Sigma(x_i - \bar{x})^2/n]^{0.5}$

[a] *AOAC INTERNATIONAL Methods Committee Guidelines for Validation of Biological Threat Agent Methods and/or Procedures* (*Calculation of CPOD and dCPOD Values from Qualitative Method Collaborative Study Data*), *J. AOAC Int*. **94**, 1359(2011) and *Official Methods of Analysis of AOAC INTERNATIONAL* (2012) 19th Ed., *Appendix I*.

[b] *International Vocabulary of Metrology (VIM)—Basic and General Concepts and Associated Terms* (2008) JCGM 200:2008, Joint Committee for Guides in Metrology (JCGM), www.bipm.org

[c] LaBudde, R.A., & Harnly, J.M. (2012) *J. AOAC Int*. **95**, 273–285.

[d] ISO 5725-1-1994.

[e] *Official Methods of Analysis* (2012) *Appendix D* (*Guidelines for Collaborative Study Procedures to Validate Characteristics of a Method of Analysis*), AOAC INTERNATIONAL, Gaithersburg, MD.

Table A3. Recommendations for evaluation

Bias (if a reference material is available)	A minimum of five replicate analyses of a Certified Reference Material.[a]
Environmental interference	Analyze test portions containing a specified concentration of one environmental materials panel member. Materials may be pooled. Consult with AOAC statistician.
Exclusivity/cross-reactivity	Analyze one test portion containing a specified concentration of one exclusivity panel member. More replicates can be used. Consult with AOAC statistician.
Inclusivity/selectivity	Analyze one test portion containing a specified concentration of one inclusivity panel member. More replicates can be used. Consult with AOAC statistician.
Limit of quantitation (LOQ)	Estimate the LOQ = average (blank) + 10 × s_0 (blank). Measure blank samples with analyte at the estimated LOQ. Calculate the mean average and standard deviation of the results. Guidance[b]: For ML ≥ 100 ppm (0.1 mg/kg): LOD = ML × 1/5. For ML < 100 ppm (0.1 mg/kg): LOD = ML × 2/5.
Measurement uncertainty	Use ISO 21748: *Guidance for the use of repeatability, reproducibility, and trueness estimates in measurement uncertainty estimation to analyze data collected for bias, repeatability, and intermediate precision to estimate measurement uncertainty*.
POD(0)	Use data from collaborative study.
POD (c)	
Repeatability	Prepare and homogenize three unknown samples at different concentrations to represent the full, claimed range of the method. Analyze each unknown sample by the candidate method seven times, beginning each analysis from weighing out the test portion through to final result with no additional replication (unless stated to do so in the method). All of the analyses for one unknown sample should be performed within as short a period of time as is allowed by the method. The second and third unknowns may be analyzed in another short time period. Repeat for each claimed matrix.
Probability of detection (POD)	Determine the desired POD at a critical concentration. Consult with Table A7 to determine the number of test portions required to demonstrate the desired POD.
Probability of identification (POI)	Consult *Probability of Identification (POI): A Statistical Model for the Validation of Qualitative Botanical Identification Methods*[c].
Recovery	Determined from spiked blanks or samples with at least seven independent analyses per concentration level at a minimum of three concentration levels covering the analytical range. Independent means at least at different times. If no confirmed (natural) blank is available, the average inherent (naturally containing) level of the analyte should be determined on at least seven independent replicates. Marginal % recovery = $(C_f - C_u) \times 100/C_A$ Total % recovery = $100(C_f)/(C_u + C_A)$ where C_f = concentration of fortified samples, C_u = concentration of unfortified samples, and C_A = concentration of analyte added to the test sample.[d] Usually total recovery is used unless the native analyte is present in amounts greater than about 10% of the amount added, in which case use the method of addition.[e]
Reproducibility (collaborative or interlaboratory study)	Quantitative methods: Recruit 10–12 collaborators; must have eight valid data sets; two blind duplicate replicates at five concentrations for each analyte/matrix combination to each collaborator.
	Qualitative methods: Recruit 12–15 collaborators; must have 10 valid data sets; six replicates at five concentrations for each analyte/matrix combination to each collaborator.

[a] *Guidance for Industry for Bioanalytical Method Validation* (May 2001) U.S. Department of Health and Human Services, U.S. Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Center for Veterinary Medicine (CVM).

[b] Codex Alimentarius Codex Procedure Manual.

[c] LaBudde, R.A., & Harnly, J.M. (2012) *J. AOAC Int.* **95**, 273–285.

[d] *Guidelines for Collaborative Study Procedures to Validate Characteristics of a Method of Analysis* (2012) *Official Methods of Analysis*, 19th Ed., *Appendix D,* AOAC INTERNATIONAL, Gaithersburg, MD.

[e] *AOAC Guidelines for Single-Laboratory Validation of Chemical Methods for Dietary Supplements and Botanicals* (2012) *Official Methods of Analysis*, 19th Ed., *Appendix K*, AOAC INTERNATIONAL, Gaithersburg, MD.

Table A4. Expected precision (repeatability) as a function of analyte concentration[a]

Analyte, %	Analyte ratio	Unit	RSD, %
100	1	100%	1.3
10	10^{-1}	10%	1.9
1	10^{-2}	1%	2.7
0.01	10^{-3}	0.1%	3.7
0.001	10^{-4}	100 ppm (mg/kg)	5.3
0.0001	10^{-5}	10 ppm (mg/kg)	7.3
0.00001	10^{-6}	1 ppm (mg/kg)	11
0.000001	10^{-7}	100 ppb (µg/kg)	15
0.0000001	10^{-8}	10 ppb (µg/kg)	21
0.00000001	10^{-9}	1 ppb (µg/kg)	30

[a] Table excerpted from *AOAC Peer-Verified Methods Program, Manual on Policies and Procedures* (1998) AOAC INTERNATIONAL, Gaithersburg, MD.

The precision of a method is the closeness of agreement between independent test results obtained under stipulated conditions. Precision is usually expressed in terms of imprecision and computed as a relative standard deviation of the test results. The imprecision of a method increases as the concentration of the analyte decreases. This table provides targets RSDs for a range of analyte concentrations.

Table A5. Expected recovery as a function of analyte concentration[a]

Analyte, %	Analyte ratio	Unit	Mean recovery, %
100	1	100%	98–102
10	10^{-1}	10%	98–102
1	10^{-2}	1%	97–103
0.01	10^{-3}	0.1%	95–105
0.001	10^{-4}	100 ppm	90–107
0.0001	10^{-5}	10 ppm	80–110
0.00001	10^{-6}	1 ppm	80–110
0.000001	10^{-7}	100 ppb	80–110
0.0000001	10^{-8}	10 ppb	60–115
0.00000001	10^{-9}	1 ppb	40–120

[a] Table excerpted from *AOAC Peer-Verified Methods Program, Manual on Policies and Procedures* (1998) AOAC INTERNATIONAL, Gaithersburg, MD.

Recovery is defined as the ratio of the observed mean test result to the true value. The range of the acceptable mean recovery expands as the concentration of the analyte decreases. This table provides target mean recovery ranges for analyte concentrations from 100% to 1 ppb.

Table A6. Predicted relative standard deviation of reproducibility ($PRSD_R$)[a]

Concentration (C)	Mass fraction (C)	$PRSD_R$, %
100%	1.0	2
1%	0.01	4
0.01%	0.0001	8
1 ppm	0.000001	16
10 ppb	0.00000001	32
1 ppb	0.000000001	45

[a] Table excerpted from *Definitions and Calculations of HorRat Values from Intralaboratory Data*, HorRat for SLV.doc, 2004-01-18, AOAC INTERNATIONAL, Gaithersburg, MD.

Predicted relative standard deviation = $PRSD_R$. Reproducibility relative standard deviation calculated from the Horwitz formula:

$$PRSD_R = 2C^{-0.15}, \text{ where C is expressed as a mass fraction}$$

This table provides the calculated $PRSD_R$ for a range of concentrations. *See Annex D* for additional information.

Table A7. POD and number of test portions[a,b]

	Sample size required for proportion						
Assume	1. Binary outcome (occur/not occur). 2. Constant probability rho of event occurring. 3. Independent trials (e.g., simple random sample). 4. Fixed number of trials (*N*)						
Inference	95% Confidence interval lies entirely at or above specified minimum rho						
Desired	Sample size *N* needed						
Minimum probability rho, %	Sample size (*N*)	Minimum No. events (*x*)	Maximum No. nonevents (*y*)	1-Sided lower confidence limit on rho[c], %	Expected lower confidence limit on rho, %	Expected upper confidence limit on rho, %	Effective AOQL[d] rho, %
50	3	3	0	52.6	43.8	100.0	71.9
50	10	8	2	54.1	49.0	94.3	71.7
50	20	14	6	51.6	48.1	85.5	66.8
50	40	26	14	52.0	49.5	77.9	63.7
50	80	48	32	50.8	49.0	70.0	59.5
55	4	4	0	59.7	51.0	100.0	75.5
55	10	9	1	65.2	59.6	100.0	79.8
55	20	15	5	56.8	53.1	88.8	71.0
55	40	28	12	57.1	54.6	81.9	68.2
55	80	52	28	55.9	54.1	74.5	64.3
60	5	5	0	64.9	56.5	100.0	78.3
60	10	9	1	65.2	59.6	100.0	79.8
60	20	16	4	62.2	58.4	91.9	75.2
60	40	30	10	62.4	59.8	85.8	72.8
60	80	56	24	61.0	59.2	78.9	69.1
65	6	6	0	68.9	61.0	100.0	80.5
65	10	9	1	65.2	59.6	100.0	79.8
65	20	17	3	67.8	64.0	94.8	79.4
65	40	31	9	65.1	62.5	87.7	75.1
65	80	59	21	65.0	63.2	82.1	72.7
70	7	7	0	72.1	64.6	100.0	82.3
70	10	10	0	78.7	72.2	100.0	86.1
70	20	18	2	73.8	69.9	97.2	83.6
70	40	33	7	70.7	68.0	91.3	79.7
70	80	63	17	70.4	68.6	86.3	77.4
75	9	9	0	76.9	70.1	100.0	85.0
75	10	10	0	78.7	72.2	100.0	86.1
75	20	19	1	80.4	76.4	100.0	88.2
75	40	35	5	76.5	73.9	94.5	84.2
75	80	67	13	75.9	74.2	90.3	82.2
80	11	11	0	80.3	74.1	100.0	87.1
80	20	19	1	80.4	76.4	100.0	88.2
80	40	37	3	82.7	80.1	97.4	88.8
80	80	70	10	80.2	78.5	93.1	85.8
85	20	20	0	88.1	83.9	100.0	91.9
85	40	38	2	86.0	83.5	98.6	91.1
85	80	74	6	86.1	84.6	96.5	90.6
90	40	40	0	93.7	91.2	100.0	95.6
90	60	58	2	90.4	88.6	99.1	93.9
90	80	77	3	91.0	89.5	98.7	94.1
95	60	60	0	95.7	94.0	100.0	97.0
95	80	80	0	96.7	95.4	100.0	97.7
95	90	89	1	95.2	94.0	100.0	97.0
95	96	95	1	95.5	94.3	100.0	97.2
98	130	130	0	98.0	97.1	100.0	98.6
98	240	239	1	98.2	97.7	100.0	98.8
99	280	280	0	99.0	98.6	100.0	99.3
99	480	479	1	99.1	98.8	100.0	99.4

[a] Table excerpted from Technical Report TR308, *Sampling plans to verify the proportion of an event exceeds or falls below a specified value*, LaBudde, R. (June 4, 2010) (not published). The table was produced as part of an informative report for the Working Group for Validation of Identity Methods for Botanical Raw Materials commissioned by the AOAC INTERNATIONAL Presidential Task Force on Dietary Supplements. The project was funded by the Office of Dietary Supplements, National Institutes of Health.

[b]

[c] Based on modified Wilson score 1-sided confidence interval.

[d] AOQL = Average outgoing quality level.

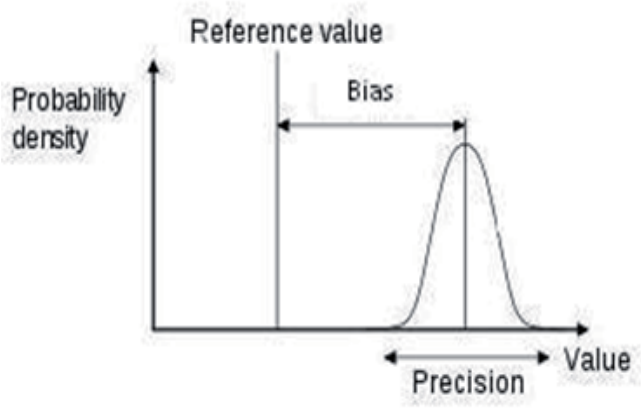


Figure A1. Relationship between precision versus bias (trueness). Trueness is reported as bias. Bias is defined as the difference between the test results and an accepted reference value.

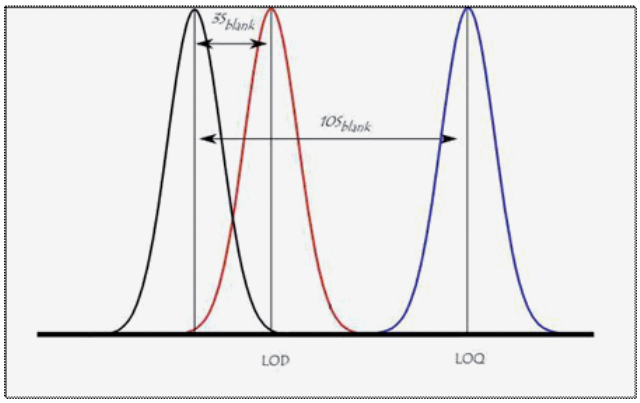


Figure A2. Relationship between LOD and LOQ. LOD is defined as the lowest quantity of a substance that can be distinguished from the absence of that substance (a blank value) within a stated confidence limit. LOQ is the level above which quantitative results may be obtained with a stated degree of confidence.

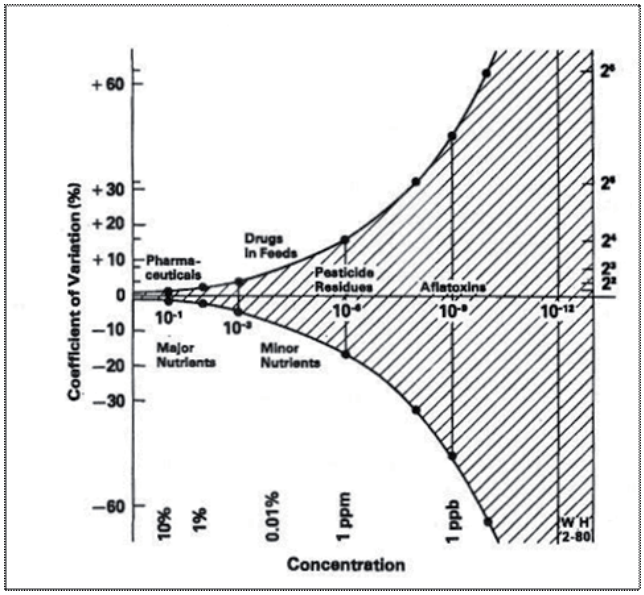


Figure A3. Horwitz Curve, illustrating the exponential increase in the coefficient of variation as the concentration of the analyte decreases [*J. AOAC Int*. 89, 1095(2006)].

ANNEX B
Classification of Methods

The following guidance may be used to determine which performance parameters in Table A1 apply to different classifications of methods. AOAC INTERNATIONAL does not recognize the term "semiquantitative" as a method classification. Methods that have been self-identified as semiquantitative will be classified into one of the following five types:

Type I: Quantitative Methods

Characteristics: Generates a continuous number as a result.

Recommendation: Use performance requirements specified for quantitative method (main or trace component). Use recovery range and maximum precision variation in Tables A4 and A5.

In some cases and for some purposes, methods with less accuracy and precision than recommended in Tables A4 and A5 may be acceptable. Method developers should consult with the appropriate method committee to determine if the recommendations in Tables A4 and A5 do or do not apply to their method.

Type II: Methods that Report Ranges

Characteristics: Generates a "range" indicator such as 0, low, moderate, and high.

Recommendation: Use performance requirements specified for qualitative methods (main component). Specify a range of POD for each range "range" indicator.

Type III: Methods with Cutoff Values

Characteristics: Method may generate a continuous number as an interim result (such as a CT value for a PCR method), which is not reported but converted to a qualitative result (presence/ absence) with the use of a cutoff value.

Recommendation: Use performance requirements specified for qualitative methods.

Type IV: Qualitative Methods

Characteristics: Method of analysis whose response is either the presence or absence of the analyte detected either directly or indirectly in a specified test portion.

Recommendation: Use performance requirements specified for qualitative methods.

Type V: Identification Methods

Characteristics: Method of analysis whose purpose is to determine the identity of an analyte.

Recommendation: Use performance requirements specified for identification methods.

ANNEX C
Understanding the POD Model

Excerpted from AOAC INTERNATIONAL Methods Committee Guidelines for Validation of Biological Threat Agent Methods and/or Procedures, J. AOAC Int. ***94****, 1359(2011) and Official Methods of Analysis of AOAC INTERNATIONAL (2012) 19th Ed., Appendix I.*

The Probability of Detection (POD) model is a way of characterizing the performance of a qualitative (binary) method. A binary qualitative method is one that gives a result as one of two possible outcomes, either positive or negative, presence/absence, or +/–.

The single parameter of interest is the POD, which is defined as the probability at a given concentration of obtaining a positive response by the detection method. POD is assumed to be dependent on concentration, and generally, the probability of a positive response will increase as concentration increases.

For example, at very low concentration, the expectation is that the method will not be sensitive to the analyte, and at very high concentration, a high probability of obtaining a positive response is desired. The goal of method validation is to characterize how method response transitions from low concentration/low response to high concentration/high response.

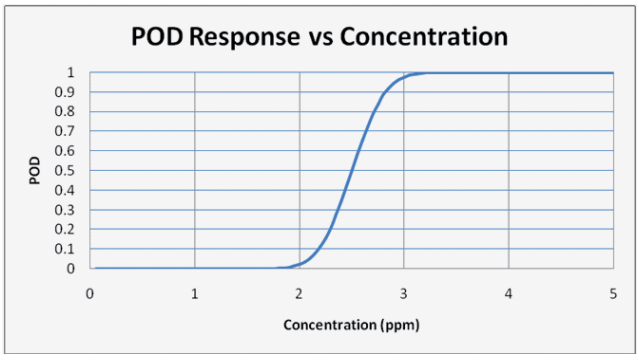


Figure C1. Theoretical POD curve for a qualitative detection method.

POD is always considered to be dependent upon analyte concentration. The POD curve is a graphical representation of method performance, where the probability is plotted as a function of concentration (*see*, for example, Figure C1).

The POD model is designed to allow an objective description of method response without consideration to an a priori expectation of the probabilities at given concentrations. The model is general enough to allow comparisons to any theoretical probability function.

The POD model is also designed to allow for an independent description of method response without consideration to the response of a reference method. The model is general enough to allow for comparisons between reference and candidate method responses, if desired.

Older validation models have used the terms "sensitivity," "specificity," "false positive," and "false negative" to describe method performance. The POD model incorporates all of the performance concepts of these systems into a single parameter, POD.

For example, false positive has been defined by some models as the probability of a positive response, given the sample is truly negative (concentration = 0). The equivalent point on the POD curve for this performance characteristic is the value of the curve at Conc = 0.

Similarly, false negative has sometimes been defined as the probability of a negative response when the sample is truly positive (concentration >0). In the POD curve, this would always be specific to a given sample concentration, but would be represented as the distance from the POD curve to the POD = 1 horizontal top axis at all concentrations except C = 0.

The POD model incorporates all these method characteristics into a single parameter, which is always assumed to vary by concentration. In other models, the terms "false positive," "false negative," "sensitivity," and "specificity" have been defined in a variety of ways, usually not conditional on concentration. For these reasons, these terms are obsolete under this model (*see* Table C1).

The terms "sensitivity," "specificity," "false positive," and "false negative" are obsolete under the POD model (*see* Figure C2).

Table C1. Terminology

Traditional terminology	Concept	POD equivalent	Comment
False positive	Probability of the method giving a (+) response when the sample is truly without analyte	POD(0) POD at conc = 0	POD curve value at conc = 0; "*Y*-intercept" of the POD curve
Specificity	Probability of the method giving a (-) response when the sample is truly without analyte	1-POD(0)	Distance along the POD axis from POD = 1 to the POD curve value
False negative (at a given concentration)	Probability of a (–) response at a given concentration	1-POD(c)	Distance from the POD curve to the POD = 1 "top axis" in the vertical direction
Sensitivity (at a given concentration)	Probability of a (+) response at a given concentration	POD(c)	Value of the POD curve at any given concentration
True negative	A sample that contains no analyte	C = 0	Point on concentration axis where c = 0
True positive	A sample that contains analyte at some positive concentration	C > 0	Range of concentration where c > 0

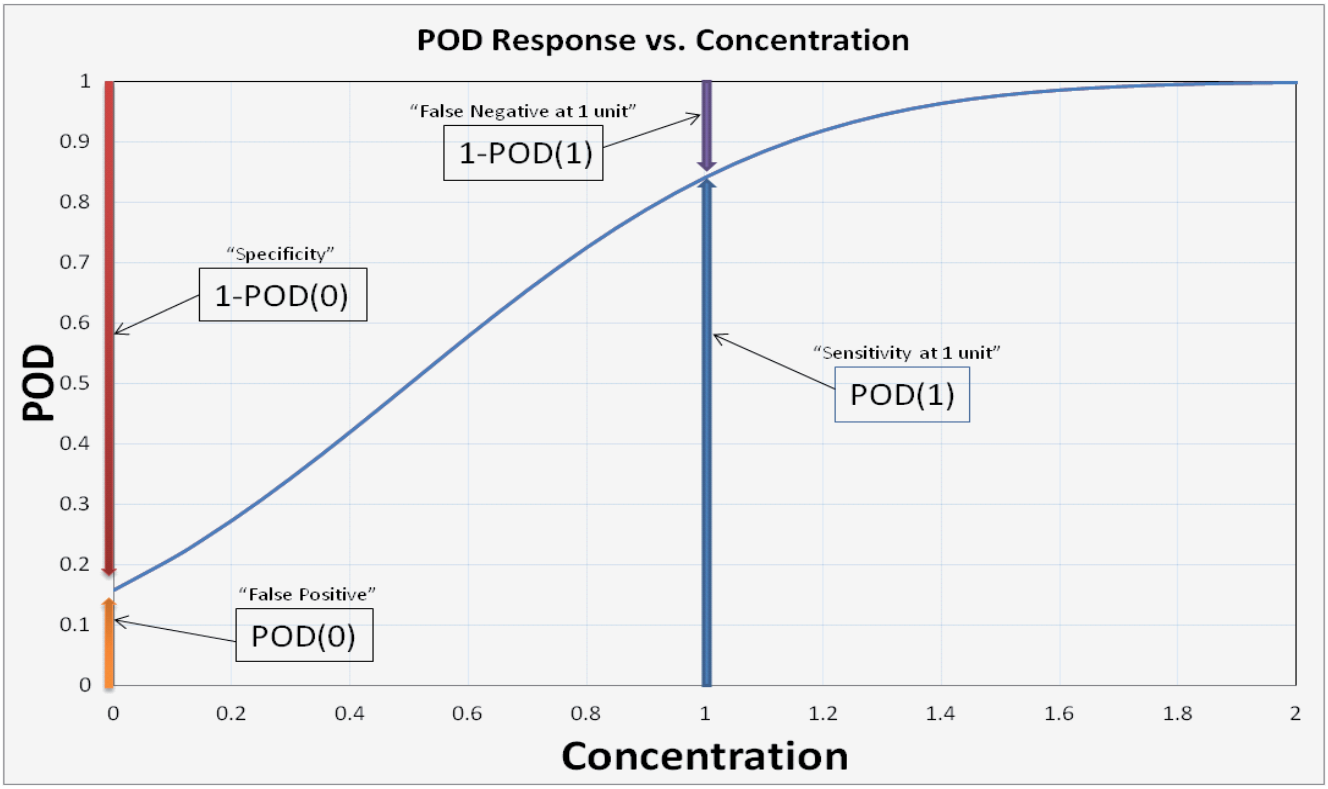


Figure C2. Comparison of POD model terminology to other obsolete terms.

ANNEX D
Definitions and Calculations of HorRat Values from Intralaboratory Data

Excerpted from Definitions and Calculations of HorRat Values from Intralaboratory Data, AOAC INTERNATIONAL, HorRat for SLV.doc, 2004-01-18.

1. Definitions

1.1 Replicate Data

Data developed under common conditions in the same laboratory: simultaneous performance, or, if necessary to obtain sufficient values, same series, same analyst, same day. Such data provides "repeatability statistical parameters."

1.2 Pooled Data

Replicate data developed in the same laboratory under different conditions but considered sufficiently similar that, for the purpose of statistical analysis, they may be considered together. These may include different runs, different instruments, different analysts, and different days.

1.3 Average

0 = Sum of the individual values, x_i, divided by the number of individual values, *n*.

$$0 = (\Sigma\, x_i)/n$$

1.4 Standard Deviation

$$s_i = [\Sigma(x_i - (\bar{x})^2/n]^{0.5}$$

1.5 Relative Standard Deviation

$$RSD = s_i \times 100/\bar{x}$$

1.5.1 Repeatability Relative Standard Deviation [RSD(r) or RSD_r]

The relative standard deviation calculated from within-laboratory data.

1.5.2 Reproducibility Relative Standard Deviation [RSD(R) or RSD_R]

The relative standard deviation calculated from among-laboratory data.

Table D1. Predicted relative standard deviations

Concentration (C)	Mass fraction (C)	$PRSD_R$, %
100%	1.0	2
1%	0.01	4
0.01%	0.0001	8
1 ppm	0.000001	16
10 ppb	0.00000001	32
1 ppb	0.000000001	45

1.6 Mass Fraction

Concentration, C, expressed as a decimal fraction. For calculating and reporting statistical parameters, data may be expressed in any convenient units (e.g., %, ppm, ppb, mg/g, µg/g; µg/kg; µg/L, µg/µL, etc.). For reporting HorRat values, data must be reported as a mass fraction where the units of the numerator and denominator are the same: e.g., for 100% (pure materials), the mass fraction C = 1.00; for 1 µg/g (ppm), C = 0.000001 = (E-6). *See* Table D1 for other examples.

1.7 Predicted Relative Standard Deviation [PRSD(R) or $PRSD_R$]

The reproducibility relative standard deviation calculated from the Horwitz formula:

$$PRSD(R) = 2C^{-0.15}$$

where C is expressed as a mass fraction. *See* Table D1.

In spreadsheet notation: PRSD(R) = 2 * C ^(–0.15).

1.8 HorRat Value

The ratio of the reproducibility relative standard deviation calculated from the data to the PRSD(R) calculated from the Horwitz formula:

$$HorRat = RSD(R)/PRSD(R)$$

To differentiate the usual HorRat value calculated from reproducibility data from the HorRat value calculated from repeatability data, attach an R for the former and an r for the latter. But note that the denominator always uses the PRSD(R) calculated from reproducibility data because this parameter is more predictable than the parameter calculated from repeatability data:

$$HorRat(R) = RSD_R/PRSD(R)$$

$$HorRat(r) = RSD_r/PRSD(R)$$

Some expected, predicted relative standard deviations are given in Table D1.

2 *Acceptable HorRat Values*

2.1 For Interlaboratory Studies

HorRat(R): The original data developed from interlaboratory (among-laboratory) studies assigned a HorRat value of 1.0 with limits of acceptability of 0.5 to 2.0. The corresponding within-laboratory relative standard deviations were found to be typically 1/2 to 2/3 the among-laboratory relative standard deviations.

Table D2. Predicted relative standard deviations

Concentration (C)	$PRSD_R$, %	$PRSD_r$, %
100%	2	1
1%	4	2
0.01%	8	4
1 ppm	16	8
10 ppb	32	16
1 ppb	45	22

2.1.1 Limitations

HorRat values do not apply to method-defined (empirical) analytes (moisture, ash, fiber, carbohydrates by difference, etc.), physical properties or physical methods (pH, viscosity, drained weight, etc.), and ill-defined analytes (polymers, products of enzyme reactions).

2.2 For Intralaboratory Studies

2.2.1 Repeatability

Within-laboratory acceptable predicted target values for repeatability are given in Table D2 at 1/2 of PRSD(R), which represents the best case.

2.2.2 HorRat(r)

Based on experience and for the purpose of exploring the extrapolation of HorRat values to SLV studies, take as the minimum acceptability 1/2 of the lower limit ($0.5 \times 0.5 \approx 0.3$) and as the maximum acceptability 2/3 of the upper limit ($0.67 \times 2.0 \approx 1.3$).

Calculate HorRat(r) from the SLV data:

$$HorRat(r) = RSD(r)/PRSD(R)$$

Acceptable HorRat(r) values are 0.3–1.3. Values at the extremes must be interpreted with caution. With a series of low values, check for unreported averaging or prior knowledge of the analyte content; with a series of high values, check for method deficiencies such as unrestricted times, temperatures, masses, volumes, and concentrations; unrecognized impurities (detergent residues on glassware, peroxides in ether); incomplete extractions and transfers and uncontrolled parameters in specific instrumental techniques.

2.3 Other Limitations and Extrapolations

The HorRat value is a very rough but useful summary of the precision in analytical chemistry. It overestimates the precision at the extremes, predicting more variability than observed at the high end of the scale (C > ca 0.1; i.e., >10%) and at the low end of the scale (C < E-8; i.e., 10 ng/g; 10 ppb).

ANNEX E
AOAC Method Accuracy Review

Accuracy of Method Based on Reference Material

Reference material (RM) used.—The use of RMs should be seen as integral to the process of method development, validation, and performance evaluation. RMs are not the only component of a quality system, but correct use of RMs is essential to appropriate quality management. RMs with or without assigned quantity values can be used for measurement precision control, whereas only RMs with assigned quantity values can be used for calibration or measurement trueness control. Method development and validation for matrices within the scope of the method is done to characterize attributes such as recovery, selectivity, "trueness" (accuracy, bias), precision (repeatability and reproducibility), uncertainty estimation, ruggedness, LOQ or LOD, and dynamic range. RMs should be chosen that are fit-for-purpose. When certified reference materials (CRMs) are available with matrices that match the method scope, much of the work involved in method development has already been completed, and that work is documented through the certificate. RMs with analyte values in the range of test samples, as well as "blank" matrix RMs, with values below or near detection limits, are needed.

Availability of RM.—Consideration needs to be given to the future availability of the chosen RM. Well-documented methods that cannot be verified in the future due to lack of material may lose credibility or be seen as inferior.

Fit to method scope.—Natural matrix CRMs provide the greatest assurance that the method is capable of producing accurate results for that matrix. When selecting an RM to perform a method validation, analysts should consider the method to material fit. An example of a good fit would be a method for specified organic molecules in infant formula and using an infant formula or powder milk RM. A poor fit would be a method for specified organic molecules in infant formula and using a sediment material.

Stability.—Providing a stable RM can be challenging where analytes are biologically active, easily oxidized, or interactive with other components of the matrix. CRM producers provide assurance of material stability, as well as homogeneity.CRMs are accompanied by a certificate that includes the following key criteria:

(*1*) Assigned values with measurement uncertainty and metrological traceability
(*2*) Homogeneity
(*3*) Stability, with the expiration date for the certificate
(*4*) Storage requirements
(*5*) Information on intended use
(*6*) Identity of matrix

For some RMs, such as botanical RMs, the source and/or authenticity can be a very important piece of information that should be included with the certificate. Even under ideal storage conditions, many analytes have some rate of change. Recertification may be done by the supplier, and a certificate reissued with a different expiration date and with certain analyte data updated or removed.

Definition of CRM.—Refer to the AOAC TDRM document for definitions from ISO Guide 30, Amd. 1 (2008), http://www.aoac.org/divisions/References.pdf.

Information on source of RM is available.—It is the responsibility of the material producer to provide reliable authentication of the RM and make a clear statement in the accompanying documentation. This should be an as detailed listing as possible, including handling of ingredients, identification of plant materials as completely as feasible (species, type, subtype, growing region), etc. This is comparable to other required information on an RM for judging its suitability for a specific application purpose (e.g., containing how much of the targeted analyte, stabilized by adding acid—therefore not suited for certain parameters/procedures, etc.).

Separate RM used for calibration and validation.—A single RM cannot be used for both calibration and validation of results in the same measurement procedure.

Blank RM used where appropriate.—Blank matrix RMs are useful for ensuring performance at or near the detection limits. These are particularly useful for routine quality control in methods measuring, for instance, trace levels of allergens, mycotoxins, or drug residues.

Storage requirements were maintained.—Method developers should maintain good documentation showing that the RM producer's recommended storage conditions were followed.

Cost.—The cost of ongoing method checks should be considered. Daily use of CRMs can be cost prohibitive. Monthly or quarterly analysis of these materials may be an option.

Concentration of analyte fits intended method.—Concentration of the analyte of interest is appropriate for standard method performance requirements (SMPRs).

Uncertainty available.—Every measurement result has an uncertainty associated with it, and the individual contributions toward the combined uncertainty arise from multiple sources. Achieving the target measurement uncertainty set by the customer for his/her problem of interest is often one of the criteria used in selecting a method for a given application. Estimation of measurement uncertainty can be accomplished by different approaches, but the use of RMs greatly facilitates this part of a method validation.

Demonstration of Method Accuracy when No Reference Material Is Available

If an RM is not available, how is accuracy demonstrated?

There are many analytes for which a CRM with a suitable matrix is not available. This leaves the analyst with few options. For some methods, there may be proficiency testing programs that include a matrix of interest for the analyte. Proficiency testing allows an analyst to compare results with results from other laboratories, which may or may not be using similar methods. Spiking is another technique that may be used. When alternative methods are available, results may be compared between the different methods. These alternatives do not provide the same level of assurance that is gained through the use of a CRM.

Spike recovery.—In the absence of an available CRM, one technique that is sometimes used for assessing performance is the spiking of a matrix RM with a known quantity of the analyte. When this method is used, it cannot be assumed that the analyte is bound in the same way as it would be in a natural matrix. Nevertheless, a certified blank RM would be the preferred choice for constructing a spiked material.

When preparing reference solutions, the pure standards must be completely soluble in the solvent. For insoluble materials in a liquid suspension or for powdered forms of dry materials, validation is required to demonstrate that the analyte is homogeneously distributed and that the response of the detection system to the analyte is not affected by the matrix or preparation technique. When a matrix material is selected for spiking, it should be reasonably

The document, *AOAC Method Accuracy Review*, was prepared by the AOAC Technical Division on Reference Materials (TDRM) and approved by the AOAC Official Methods Board in June 2012.

characterized to determine that it is sufficiently representative of the matrix of interest. Spiked samples must be carried through all steps of the method. Many analytes are bound in a natural matrix and whether the spiked analyte will behave the same as the analyte in a natural matrix is unknown.

Other.—Use of a substitute RM involves the replacement of the CRM with an alternative matrix RM matching the matrix of interest as close as possible based on technical knowledge.

ANNEX F
Development and Use of In-House Reference Materials

The use of reference materials is a vital part of any analytical quality assurance program. However, you may have questions about their creation and use. The purpose of this document is to help answer many of these questions.

- What is a reference material?
- Why use reference materials?
- What certified reference materials are currently available?
- Why use an in-house reference material?
- How do I create an in-house reference material?
- How do I use the data from an in-house reference material?

What Is a Reference Material?

The International Organization for Standardization (ISO) defines a reference material as a "material or substance one or more of whose property values are sufficiently homogeneous and well established to be used for the calibration of an apparatus, the assessment of a measurement method, or for assigning values to materials" (1). In plain English, natural-matrix reference materials, such as those you might prepare for use in-house, can be used to validate an analytical method or for quality assurance while you're using your method to analyze your samples. (Natural-matrix materials are not generally used as calibrants because of the increased uncertainty that this would add to an analysis.) The assigned values for the target analytes of an in-house reference material can be used to establish the precision of your analytical method and, if used in conjunction with a CRM, to establish the accuracy of your method.

ISO defines a certified reference material (CRM) as a "reference material, accompanied by a certificate, one or more of whose property values are certified by a procedure which establishes traceability to an accurate realization of the unit in which the property values are expressed, and for which each certified value is accompanied by an uncertainty at a stated level of confidence" (1).

Why Use Reference Materials?

Certified reference materials can be used across the entire scope of an analytical method and can provide traceability of results to the International System of Units (SI). During method development, CRMs can be used to optimize your method. During method validation, they can be used to ensure that your method is capable of producing the "right" answer, and to determine how close your result is to that answer. During routine use, they can be used to determine within-day and between-day repeatability, and so demonstrate that your method is in control and is producing accurate results every time it is used.

Natural-matrix reference materials should mimic the real samples that will be analyzed with a method. They should behave just as your samples would during a procedure, so if you obtain accurate and precise values for your reference material, you should obtain accurate and precise values for your samples as well.

What Certified Reference Materials Are Currently Available?

CRMs are available from a number of sources, including (but not limited to):

- American Association of Cereal Chemists (AACC)
- American Oil Chemists Society (AOCS)
- International Atomic Energy Agency (IAEA)
- Institute for Reference Materials and Measurements (IRMM)
- LGC Promochem
- National Institute of Standards and Technology (NIST)
- National Research Council Canada (NRC Canada)
- UK Food Analysis Proficiency Assessment Program (FAPAS)

A number of websites provide general overviews and catalogs of producers' and distributors' reference materials:

http://www.aocs.org/tech/crm/
http://www.comar.bam.de
http://www.erm-crm.org
http://www.iaea.org/oregrammeslaqcs
http://www.aaccnet.org/checksample
http://www.irmm·ire.be/mrm.html
http://www.lgcpromochem.com
http://www.naweb.iaea.org/nahu/nmrm/
http://www.nist.gov/srm
http://www.fapas.com/index. cfm
http://www.virm.net.

Because new reference materials are produced regularly, it is important to check these websites to determine what is currently available.

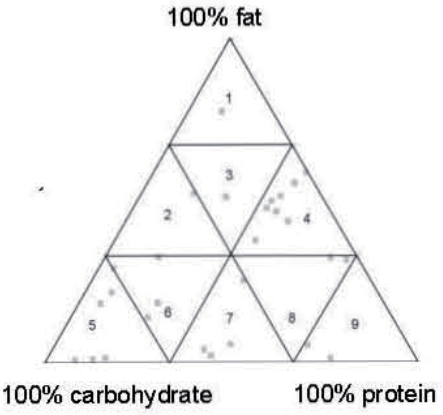
Why Use an In-House Reference Material?

There are many benefits to the use of a CRM. CRMs have been prepared to be homogeneous and, if stored under the proper conditions, stable. You are provided with a certified value as well as the statistical data for theconcentration of your analyte; this is about as close as you can come to knowing the true value of the concentration of the analyte. The material has been tested by experienced analysts in leading laboratories, so you have the security of knowing that your method is generating values similar to those generated in other competent laboratories. The CRMs from the sources mentioned above are nationally and/or internationally recognized, so when you obtain acceptable results for a CRM using your analytical method, you give credibility to your methodology and traceability to your results.

But there are some drawbacks associated with CRMs. Unfortunately, many analyte/matrix combinations are not currently available. When testing food products for nutrient content, for example, a laboratory can be asked to analyze anything that might be found in a kitchen or grocery store. Reference materials that represent all of the types of foods that need to be tested are not available, and most CRMs are certified for a limited number of analytes. It is important to match the reference material matrix to your sample matrix. (Food examples dominate the discussion below, but the same processes apply to the development of in-house RMs in other areas of analytical chemistry.)

To demonstrate the applicability of an analytical method to a wide variety of food matrices, AOAC INTERNATIONAL's Task

Excerpted from *Development and Use of In-House Reference Materials*, Rev. 2, 2009. Copyright 2005 by the AOAC Technical Division on Reference Materials (TDRM).



Force on Methods for Nutrition Labeling developed a triangle partitioned into sectors in which foods are placed based on their protein, fat, and carbohydrate content (2, 3). Since ash does not have a great impact on the performance of an analytical method for organic-material foods, and water can be added or removed, it can be assumed that the behavior of an analytical method is determined to large extent by the relative proportions of these proximates. AOAC INTERNATIONAL anticipated that one or two foods in a given sector would be representative of other foods in that sector and therefore would be useful for method assessment. Similarly, one or two reference materials in a given sector (or near each other in adjacent sectors) should be useful for quality assurance for analyses involving the other foods in the sector. The positions of many of the food-matrix CRMs from the sources listed above are shown in the triangle and are provided in the list.

These food-matrix reference materials are spread through all sectors of the triangle, thereby making it likely that you can find an appropriate CRM to match to your samples. Ultimately, however, the routine use of a CRM can be cost prohibitive, and is not really the purpose of CRMs. For example, in order to use NIST's Standard Reference Material (SRM) 2387 Peanut Butter for all mandatory nutrition labeling analyses, you could buy one sales unit (three jars, each containing 170 g material) for $649 (2009 price). If you charge your customer about $1000 for analysis of all mandatory nutrients in a test material, the control material would account for more than 60% of your fees. Therefore, many laboratories have found it more cost-effective to create in-house reference materials for routine quality control and characterize them in conjunction with the analysis of a CRM (4). You can prepare larger quantities of a reference material by preparing it in-house, and you have more flexibility in the types of matrices you can use. There are not many limitations on what can be purchased.

How Do I Create an In-House Reference Material?

There are basically three steps to preparing an in-house reference material: selection (including consideration of homogeneity and stability), preparation, and characterization. Additional guidance through these steps can be provided from TDRM as well as in ISO Guides 34 (5) and 35 (6).

References

(1) JCGM 200:2008, *International vocabulary of metrology—Basic and general concepts and associated terms (VIM)*, International Bureau of Weights and Measures (www.bipm.org)

Sector	RM No.	Matrix
	NIST 1563	Coconut oil
1	NIST 3274	Fatty acids in botanical oils
1	NIST 3276	Carrot extract in oil
1	LGC 7104	Sterilized cream
2	NIST 2384	Baking chocolate
3	NIST 2387	Peanut butter
4	NIST 1546	Meat homogenate
4	LGC 7106	Processed cheese
4	LGC 7000	Beef/pork meat
4	LGC 7150	Processed meat
4	LGC 7151	Processed meat
4	LGC 7152	Processed meat
4	SMRD 2000	Fresh meat
4	LGC 7101	Mackerel paste
4	LGC QC1001	Meat paste 1
4	LGC QC1004	Fish paste 1
5	BCR-382	Wleat flour
5	BCR-381	Rye flour
5	LGC 7103	Sweet digestive biscuit
5	LGC 7107	Madeira cake
5	LGC QC1002	Flour 1
6	NIST 1544	Fatty acids
6	NIST 1548a	Typical diet
6	NIST 1849	Infant/adult nutritional formula
6	LGC 7105	Rice pudding
7	LGC 7001	Pork meat
7	NIST 1566b	Oyster tissue
7	NIST 1570a	Spinach leaves
7	NIST 2385	Spinach
8	NIST 1946	Lake trout
8	LGC 7176	Canned pet food
9	NIST 1974a	Mussel tissue
9	NIST 3244	Protein powder

(2) Wolf, W.R., & Andrews, K.W. (1995) *Fresenius' J. Anal. Chem.* **352**, 73–76

(3) Wolf, W.R. (1993) *Methods of Analysis for Nutrition Labeling*, D.R. Sullivan & D.E. Carpenter (Eds), AOAC INTERNATIONAL, Gaithersburg, MD

(4) European Reference Materials (2005) *Comparison of a Measurement Result with the Certified Value*, Application Note 1

(5) *ISO Guide 34 General Requirements for the Competence of Reference Material Producers* (2009) 2nd, International Organization for Standardization, Geneva, Switzerland

(6) *Guide 35 Certification of Reference Materials—General and Statistical Principles* (2006) International Organization for Standardization, Geneva, Switzerland

For more information about the AOAC Technical Division on Reference Materials, visit http://aoac.org/divisions/tdrm.

Appendix G: Procedures and Guidelines for the Use of AOAC Voluntary Consensus Standards to Evaluate Characteristics of a Method of Analysis

Expert Review Panels, Official Methods Board, First and Final Action *Official Methods*℠

In early 2011, an AOAC Presidential Task Force recommended that AOAC use Expert review panels (ERPs) to assess candidate methods against standard method performance requirements (SMPRs) to ensure that adopted First Action Official Methods℠ are fit for purpose.

Formation of an ERP

AOAC ERPs are authorized to adopt candidate methods as First Action *Official Methods* and to recommend adoption of these methods to Final Action *Official Methods* status. Scientists are recruited to serve on ERPs by a variety of ways. Normally, a call for experts is published at the same time as a call for methods is posted. Interested scientists are invited to submit their *curriculum vitae* (CV) for consideration. Advisory panel, stakeholder panel, and working group members may make recommendations to AOAC for ERP members. All CVs are reviewed and evaluated for expertise by the AOAC Chief Scientific Officer (CSO). The CVs and CSO evaluations are forwarded to the OMB for formal review. Both the CSO and OMB strive to ensure that the composition of a proposed ERP is both qualified and represent the various stakeholder groups. The recommended ERP members are submitted to the AOAC president who then appoints the ERP members.

Review of Methods

Methods submitted to AOAC in response to a call for methods are collected and compiled by AOAC staff. The AOAC CSO and working group chair perform a preliminary review of the methods and classify them into three categories: (*1*) fully developed and written methods that appear to meet SMPRs; (*2*) fully developed and written methods that may or may not meet SMPRs; and (*3*) incomplete methods with no performance data. Method submitters are apprised of the evaluation of their methods. Method developers with submissions that are classified as Category 2 or 3 are encouraged to provide additional information if available. A list of all the submitted methods and their classifications are posted for public review.

Usually, two ERP members (sometimes more) are assigned to lead the review of each Category 1 method. An ERP meeting is convened to review the methods. ERP meetings are open to all interested parties, and are usually well-attended events with about 50–60 attendees common. Each Category 1 method is reviewed and discussed by the ERP. If stakeholders have designated the method to be a dispute resolution method (as stated in the SMPR), then the ERP is asked to identify the single best candidate method to be adopted as a First Action *Official Method*. If the SMPR does not specify the need for a dispute resolution method, then the ERP may choose to adopt all methods that meet the SMPRs, or may choose to adopt the single best method in their collective, expert opinion.

In addition, an ERP may choose to require changes to a candidate method as part of its First Action adoption and/or identify issues that are required to be resolved prior to adoption as a Final Action *Official Method.*

Methods adopted by an ERP as First Action *Official Methods* may not be in AOAC *Official Methods* format. Method developers/authors are asked to assist AOAC to rewrite the method and accompanying manuscript into an AOAC-acceptable format.

Two-Year First Action Evaluation Period

Under the new pathway, a method may be designated as a First Action *Official Method* based on the collective judgment of an ERP. *Official Methods* remain as First Action for a period of about 2 years. During the First Action period, the method will be used in laboratories, and method users will be asked to provide feedback on the performance of the method.

As previously described, two (or more) ERP members are assigned to lead the review of candidate methods for adoption as First Action *Official Methods*. After a method has been adopted as First Action, these lead reviewers are expected to keep track of the use of and experience with the First Action *Official Method*. At the conclusion of the 2-year evaluation period, one or both of the lead reviewers will report back to the ERP on the experience of the First Action *Official Method.*

The presiding ERP will monitor the performance of the method, and, at the completion of the 2-year First Action evaluation period, determine whether the method should be recommended to the OMB for adoption as an AOAC Final Action *Official Method*.

It is also possible that First Action *Official Methods* are not recommended for Final Action. There are two possibilities for an ERP to decide not to proceed with a First Action method: (*1*) feedback from method users indicates that a First Action method is not performing as well in the field as was expected; or (*2*) another method with better performance characteristics has been developed and reviewed. In either case, the ERP may choose to repeal the First Action status of a method.

OMB Review

The OMB will review all methods recommended for Final Action or repeal by the ERP, and will consider a number of factors in their decision. A guidance document for factors to consider is provided on the AOAC website at http://www.aoac.org/vmeth/OMB_ERP_Guidance.pdf. Some of the factors identified by the guidance document for OMB consideration are (*1*) feedback from method users, (*2*) comparison to the appropriate SMPR, (*3*) results from single-laboratory validation, (*4*) reproducibility/uncertainty and probability of detection, (*5*) availability of reference materials, and (*6*) safety concerns.

Conclusion

The new pathway to *Official Methods*℠ is deliberately designed to avoid creation of elaborate review systems. The intent of the model is for method experts to use their scientific knowledge, experience, and good judgment to identify and adopt the best methods possible for the analytical need.

These methods are then published as First Action *Official Methods,* and used by analysts while additional information about the method is collected.

Method reviewers may consider other forms of information in lieu of the traditional collaborative study to demonstrate method reproducibility.

Additional Information

Coates, S. (2012) "Alternative Pathway," *Inside Laboratory Management* **16**(3), pp 10–12

Expert Review Panels, Policies and Procedures, AOAC INTERNATIONAL, http://www.aoac.org/News/EXPERT%20REVIEW%20PANELS%20final%20revision.pdf

Standard Format and Guidance for AOAC Standard Method Performance Requirement (SMPR) Documents, AOAC INTERNATIONAL, http://www.aoac.org/ISPAM/pdf/3.5%20SMPR%20Guideline%20v12.1.pdf

Guidance Documents

Requirements for First Action Official Methods℠ Status

See Figure 1 for process flowchart.

Expert Review Panels

(*1*) Supported by relevant stakeholders.

(*2*) Constituted solely for the ERP purpose, not for SMPR purposes or as an extension of an SMPR.

(*3*) Consist of a minimum of seven members representing a balance of key stakeholders. A quorum is the presence of seven members or 2/3 of total vetted ERP membership, whichever is greater.

(*4*) ERP constituency must be approved by the OMB.

(*5*) Hold transparent public meetings only.

(*6*) Remain in force as long as method in First Action status.

First Action Official Method℠ Status Decision

(*1*) Must be made by an ERP constituted or reinstated post March 28, 2011 for First Action *Official Method℠* status approval.

(*2*) Must be made by an ERP vetted for First Action *Official Method℠* status purposes by OMB post March 28, 2011.

(*3*) Method adopted by ERP must perform adequately against the SMPR set forth by the stakeholders.

(*4*) Method must be adopted by unanimous decision of ERP on first ballot. If not unanimous, negative votes must delineate scientific reasons.

(*5*) Negative voter(s) can be overridden by 2/3 of voting ERP members after due consideration.

(*6*) Method becomes Official First Action on date when ERP decision is made.

(*7*) Methods to be drafted into AOAC format by a knowledgeable AOAC staff member or designee in collaboration with the ERP and method author.

(*8*) Report of First Action *Official Method℠* status decision complete with ERP report regarding decision, including scientific background (references, etc.), to be published concurrently with method in traditional AOAC publication venues.

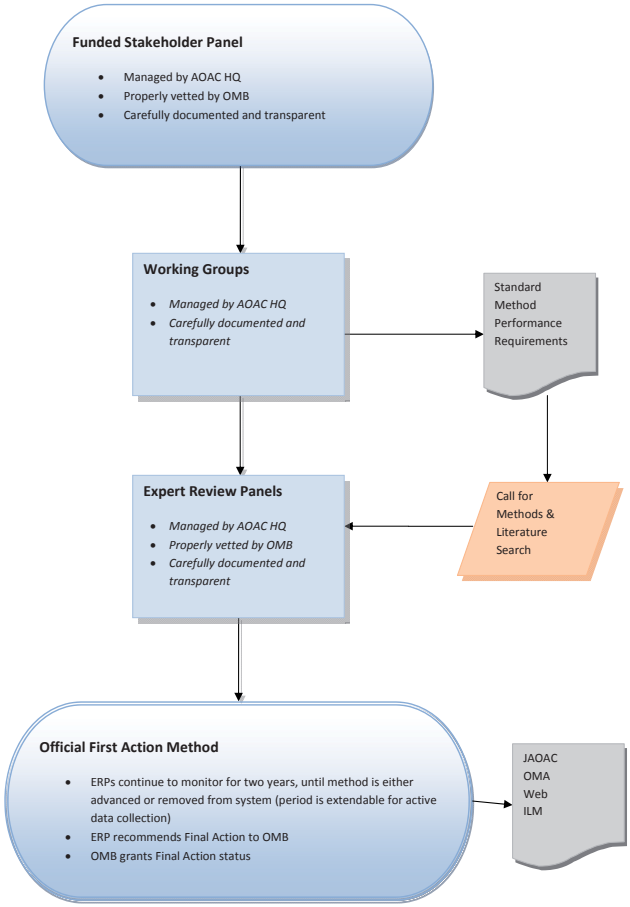


Figure 1. Summary of standards development through *Official Methods of Analysis*.

Method in First Action Status and Transitioning to Final Action Status

(*1*) Further data indicative of adequate method reproducibility (between laboratory) performance to be collected. Data may be collected via a collaborative study or by proficiency or other testing data of similar magnitude.

(*2*) Two years maximum transition time [additional year(s) if ERP determines a relevant collaborative study or proficiency or other data collection is in progress].

(*3*) Method removed from Official First Action and OMA if no evidence of method use available at the end of the transition time.

(*4*) Method removed from Official First Action and OMA if no data indicative of adequate method reproducibility is forthcoming as outlined above at the end of the transition time.

(*5*) ERP to recommend method to Final Action Official status to the OMB.

(*6*) OMB decision on First to Final Action status.

These guidance documents were approved by the AOAC Board of Directors on May 25, 2011. Revised in February 2014 to include the definition of a quorum under the section *Expert Review Panels*, item (*3*).

First Action to Final Action Methods: Guidance for AOAC Expert Review Panels

In December 2011, the Official Methods Board (OMB) approved a guidance document for ERPs to support their work as they deliberate on methods, adopt methods as Official First Action, and, subsequently, track method usage and performance between First Action status and Final Action consideration. The guideline is based on parameters of a method that the OMB will consider when deliberating on methods recommended for Final Action status. ERPs are to use this guideline in their deliberations.

ERPs working within the AOAC process may recommend a First Action status method be elevated to Final Action status. Such a recommendation leverages the ERP's high level of expertise supported by data from the initial evaluation, and results from the subsequent 2-year method performance evaluation period.

The OMB receives the recommendation with supporting documentation, and determines if Final Action status is warranted. OMB's review verifies the method process was conducted in compliance with the guidelines and protocols of the Association.

For transparency and to expedite the review process, the main areas OMB will review when evaluating ERP recommendations to promote methods to Final Action are listed below. Documentation of the areas listed below will also increase confidence in method performance and assist users to properly and safely perform the methods at their locations.

A. Method Applicability

(**a**) A method's applicability to the identified stakeholder needs is best assessed by the stakeholder panel and should be a part of the process from the onset. OMB liaisons will remind stakeholder panels to maintain this focus point.

(**b**) OMB may ask ERPs and stakeholder panels for feedback to improve the applicability of the method, such as potential method scope expansions and potential points of concern.

B. Safety Concerns

(**a**) A safety review must be performed for a method to be recognized as First Action.

(**b**) All safety concerns identified during the 2-year evaluation period must be addressed.

(**c**) Guidance and support can be obtained from the AOAC Safety Committee.

C. Reference Materials

(**a**) Document efforts undertaken to locate reference materials. Methods may still progress to Final Action even if reference materials are not available.

(**b**) Guidance and support can be obtained from the AOAC Technical Division on Reference Materials.

D. Single-Laboratory Validation

(**a**) Data demonstrating response linearity, accuracy, repeatability, LOD/LOQ, and matrix scope must be present. Experimental designs to collect this data may vary with the method protocol and the intended use of the method.

(**b**) Resources can be identified by the AOAC Statistics Committee.

E. Reproducibility/Uncertainty and Probability of Detection

(**a**) For quantitative methods, data demonstrating reproducibility and uncertainty must be present. Experimental designs to collect this data may vary with the method protocol, available laboratories, and the intended use of the method (i.e., collaborative studies, proficiency testing, etc.).

(**b**) For qualitative methods, data must be present demonstrating the probability of detection at specified concentration levels as defined by the SMPR. Experimental designs to collect this data may vary with the method protocol, available laboratories, and the intended use of the method.

(**c**) Guidance and support can be obtained from the AOAC Statistics Committee.

F. Comparison to SMPR

(**a**) Document method performance versus SMPR criteria. Note which SMPR criteria are met. For SMPR criteria not met, the ERP documents the reasoning why the method is still acceptable.

(**b**) Data is present to assure the matrix and analyte scopes are covered. This is critical for methods used for dispute resolutions.

G. Feedback from Users of Method

(**a**) Document positive and negative feedback from users of the method during the trial period.

(**b**) Feedback from users demonstrating method ruggedness should be documented.

(**c**) Assess the future availability of vital equipment, reference materials, and supplies.

H. ERP Recommendations to Repeal First Action Methods

Recommendations to repeal First Action methods shall be accompanied with detailed reasons for the decision.

The First to Final Action guidance for ERPs was approved by the OMB in December 2011 and effective as of February 1, 2012.

Appendix K: Guidelines for Dietary Supplements and Botanicals

This appendix contains three complementary documents for the validation of dietary supplements and botanical methods:

Part I: AOAC Guidelines for Single-Laboratory Validation of Chemical Methods for Dietary Supplements and Botanicals

Part II: AOAC Guidelines for Validation of Botanical Identification Methods

Part III: Probability of Identification: A Statistical Model for the Validation of Qualitative Botanical Identification Methods

PART I
AOAC Guidelines for Single-Laboratory Validation of Chemical Methods for Dietary Supplements and Botanicals

Contents

1 Introduction

1.1 Definitions

1.1.1 Validation

1.1.2 Method of Analysis

1.1.3 Performance Characteristics of a Method of Analysis

2 Single-Laboratory Validation Work

2.1 Preparation of the Laboratory Sample

2.2 Identification

2.3 Method of Analysis or Protocol

2.3.1 Optimization

2.3.2 Reference Standard

2.3.3 Ruggedness Trial

2.3.4 Specific Variables
- a. Analyte Addition
- b. Reextraction of the Extracted Residue
- c. Comparison with Different Solvents
- d. Comparison with Results from a Different Procedure
- e. System Suitability Checks

3 Performance Characteristics

3.1 Applicability (Scope)

3.2 Selectivity

3.3 Calibration

3.3.1 External Standard Method

3.3.2 Internal Standard Method

3.3.3 Standard Addition Method

3.4 Reliability Characteristics

3.4.1 Accuracy

3.4.2 Repeatability Precision (s_r, RSD_r)

3.4.3 Measurement Uncertainty

3.4.4 Reproducibility Precision (s_R, RSD_R)

3.4.5 Intermediate Precision

3.4.6 Limit of Determination

3.4.7 Reporting Low-Level Values

3.4.8 Dichotomous Reporting

3.5 Controls

3.5.1 Control Charts

3.5.2 Injection Controls

3.5.3 Duplicate Controls

3.6 Confirmation of Analyte

3.7 Stability of the Analyte

4 Report (as applicable)

4.1 Title

4.2 Applicability (Scope)

4.3 Principle

4.4 Reagents

4.5 Apparatus

4.6 Calibration

4.7 Procedure

4.8 Calculations

4.9 Controls

4.10 Results of Validation

4.10.1 Identification Data

4.10.2 Performance Data

Under a 5-year contract (2003–2008) with the National Institutes of Health-Office of Dietary Supplements, through the U.S. Food and Drug Administration, AOAC undertook an effort to validate methods for dietary supplement ingredients of interest. As part of the initiative, AOAC adapted and revised the traditional *Official Methods*℠ process to include single-laboratory validation (SLV). Methods were first validated within a single laboratory to test their suitability and ruggedness without the complications of a multilaboratory collaborative study. SLVs proved to be an excellent debugging tool for complex methods; problems found within one laboratory could be dealt with so that a stronger method went on to the collaborative study. The SLV process, thus, became a step in preparation for the collaborative study.

The SLV guidelines were approved by the AOAC Official Methods Board and Board of Directors in December 2002.

4.10.3 Low-Level Data

4.10.4 Stability Data

Annex A: Abbreviations and Symbols Used

Annex B: Example of a Ruggedness Trial

Because of the time and expense required for the determination of modern analytes such as pesticide residues, industrial contaminants, veterinary drugs, allergens, botanicals, dietary supplements, and alternative medicines in complex matrices, there is considerable interest in obtaining acceptable methods of analysis faster and cheaper. It has been suggested that accreditation of laboratories, internal quality control, and external proficiency exercises can improve laboratory performance to the point where interlaboratory validation is no longer an absolute necessity. To this end AOAC INTERNATIONAL has been exploring alternatives to the full interlaboratory study design that requires the examination of a minimum of five matrices by eight laboratories (*see* www.aoac.org under method validation programs). These have included "minicollaborative" studies that reduced the required number of matrices and laboratories, the "Peer-Verified Methods Program," which merely required verification of the analytical parameters by a second laboratory, "*Performance Tested Methods*^{SM}" for test kits, the developing e-CAM compiling program (www.AOAC.org/AOAC_e-CAM.pdf), and the International Union of Pure and Applied Chemistry (IUPAC) sanctioned single-laboratory validation (SLV) protocol [*Pure & Appl. Chem*. **74**(5), 835–855(2002)].

The IUPAC single-laboratory protocol necessarily deals in generalities and specifically points out, "The total cost to the analytical community of validating a specific method through a collaborative trial and then verifying its performance attributes in the laboratories wishing to use it, is frequently less than when many laboratories all independently undertake SLV of the same method." The protocol also indicates that the degree of validation depends upon the status of the method in the analytical structure. At one extreme is the initial application of a well-established method in a laboratory that merely requires verification of the capability of that laboratory to achieve the published performance characteristics. The opposite extreme is the initial presentation of a new method or the initial application of an established method to a new matrix or application. Methods that are developed in response to a continued need for compliance, surveillance, and enforcement of laws and contracts involving a number of laboratories are expected to proceed to a multilaboratory validated status.

This AOAC document is intended to present guidelines for the evaluation of the initial use of a new or old method in a laboratory. It assumes that a proposed or available method is fairly well developed, optimized, and stabilized, that it has been applied to some practical test samples with acceptable results, and that a description of the method and its initial performance results are available in some kind of document. The initiating or another laboratory must then decide if the demonstrated performance appears to be satisfactory for the same or for another purpose.

Although the output from method development is the input to method validation, method developers cannot expect much input from method validators. Although method validators may have had considerable experience in the analysis of practical analytical samples, they are not expected to have the basic knowledge to recommend improvement in methods, such as certain solvents as useful for extraction of certain classes of analytes or column-solvent combinations as useful for optimization of separations. Method developers are expected to bring methods to the point where they satisfy validation requirements.

By definition, SLV does not provide any information on what values would be expected on examination of identical test samples by other laboratories. Therefore such methods probably would be used by regulatory agencies only for monitoring purposes—to explore compliance with laws and regulations unless the statutes under which they operate assign correctness to their results. Ordinarily such methods would not be used to bring a legal action or to settle a commercial dispute until their properties had been further explored in an environment provided by an interlaboratory collaborative study or a proficiency study utilizing that method. As stated in the FDA Center for Drug Evaluation and Research (CDER) "Reviewer Guidance/Validation of Chromatographic Methods" (November 1994), "Methods should be reproducible when used by other analysts, on other equivalent equipment, on other days and locations, and throughout the life of the drug product."

1 Introduction

The primary purpose of validating a method of analysis is to show that the method is fit for its intended purpose. Some purposes are:

(*1*) Determine how much of a valuable, necessary, or characteristic ingredient is present in a product.

(*2*) Determine if a product meets specifications.

(*3*) Determine if a product meets regulatory requirements.

(*4*) Survey an environment to determine the presence and amount of a component, contaminant, or a nutrient.

(*5*) Identify a product and/or its components.

The purposes usually answer the questions, "What is this product?" in the sense of its common or usual name, chemical identity, or components, and "How much of something [an analyte] is in this product [matrix]?"

At least at the initial stages of a problem, only a single or at most a very few laboratories require validation of a method of analysis. These circumstances include situations similar to the following:

(*1*) Methods for research.

(*2*) Only a few test samples are anticipated.

(*3*) For quality control of a manufacturing process of a single item by a single producer,

(*4*) Checking the reliability of a method imported from another source.

(*5*) Rechecking the reliability of a previously used method after a period of disuse.

(*6*) Situations where there is a lack of interest by other laboratories in participating in an interlaboratory validation exercise.

(*7*) Multi-analyte, multi-matrix methods where a conventional interlaboratory validation exercise is impractical.

For the present purpose we assume:

(*1*) We know or can assume the chemical identity of the material we are dealing with.

(*2*) We have a specimen of the material that can be used as a reference to compare the signal produced by the analyte isolated from the product we are examining with the same signal produced by a known amount of the reference analyte (traceable to a stated reference).

If either or both of these requirement are not met, much useful information can still be obtained, but our information will be "floating" in the same sense as a ship at sea does not know where it is without landmarks to determine its position. If the identity of an analyte must

be determined, not merely verified, a whole new dimension is added to the problem. This involves bringing in a laboratory or an individual with skill in determining chemical structure, a highly specialized, expensive, and time-consuming exercise.

It is often found during the initial experience with application or validation of a method that deficiencies appear, unexpected interferences emerge, reagents and equipment are no longer available, instruments must be modified, and other unanticipated problems require returning the method to a development phase. Frequently a method that functions satisfactorily in one laboratory fails to operate in the same manner in another. Often there is no clear-cut differentiation between development and validation and the two procedures constitute an iterative process. For that reason some aspects of method development that provide an insight into method performance, such as ruggedness, are included in this document.

In some cases it is impossible to set specific requirements because of unknown factors or incomplete knowledge. In such cases it is best to accept whatever information is generated during development and validation and rely upon the "improvements" that are usually forthcoming to asymptotically approach performance parameters developed for other analytes in the same or in a similar class.

1.1 Definitions

1.1.1 Validation

Validation is the process of demonstrating or confirming the performance characteristics of a method of analysis.

This process of validation is separate from the question of acceptability or the magnitude of the limits of the characteristics examined, which are determined by the purpose of the application. Validation applies to a specific operator, laboratory, and equipment utilizing the method over a reasonable concentration range and period of time.

Typically the validation of a chemical method of analysis results in the specification of various aspects of reliability and applicability. Validation is a time-consuming process and should be performed only after the method has been optimized and stabilized because subsequent changes will require revalidation. The stability of the validation must also be verified by periodic examination of a stable reference material.

1.1.2 Method of Analysis

The method of analysis is the detailed set of directions, from the preparation of the test sample to the reporting of the results, that must be followed exactly for the results to be accepted for the stated purpose.

The term "method of analysis" is sometimes assigned to the technique, e.g., liquid chromatography or atomic absorption spectrometry, in which case the set of specific directions is referred to as the "protocol."

1.1.3 Performance Characteristics of a Method of Analysis

The performance characteristics of a method of analysis are the functional qualities and the statistical measures of the degree of reliability exhibited by the method under specified operating conditions.

The functional qualities are the selectivity (specificity), as the ability to distinguish the analyte from other substances; applicability, as the matrices and concentration range of acceptable operation; and degree of reliability, usually expressed in terms of bias as recovery, and variability as the standard deviation or equivalent terms (relative standard deviation and variance).

Measurements are never exact and the "performance characteristics of a method of analysis" usually reflect the degree to which replicate measurements made under the same or different conditions can be expected or required to approach the "true" or assigned values of the items or parameters being measured. For analytical chemistry, the item being measured is usually the concentration, with a statement of its uncertainty, and sometimes the identity of an analyte.

For abbreviations and symbols used in this guideline, *see Annex A*.

2 Single-Laboratory Validation Work

2.1 Preparation of the Laboratory Sample

Product and laboratory sampling are frequently overlooked aspects of analytical work because very often product sampling is not under the control of the laboratory but the sample is supplied by the customer. In this case, the customer assumes the responsibility of extrapolating from the analytical result to the original lot. If the laboratory is requested to sample the lot, then it must determine the purpose of the analysis and provide for random or directed sampling accordingly.

The laboratory is responsible for handling the sample in the laboratory to assure proper preparation with respect to composition and homogeneity and to assure a suitable analytical sample. The laboratory sample is the material received by the laboratory and it usually must be reduced in bulk and fineness to an analytical sample from which the test portions are removed for analysis.

Excellent instructions for this purpose will be found in the "Guidelines for Preparing Laboratory Samples" prepared by the American Association of Feed Control Officials, Laboratory Methods and Service Committee, Sample Preparation Working Group (2000) (AAFCO, Oxford, IN) that cover the preparation of particularly difficult mineral and biological material. The improper or incomplete preparation of the analytical sample is an often overlooked reason for the nonreproducibility of analytical results.

If a laboratory prepares test samples for the purpose of validating a method, it should take precautions that the analyst who will be doing the validation is not aware of the composition of the test samples. Analysts have a bias, conscious or unconscious, of permitting knowledge of the identity or composition of a test sample to influence the result [*J. AOAC Int.* **83**, 399–406(2000)].

2.2 Identification

Identification is the characterization of the substance being analyzed, including its chemical, mineral, or biological classification, as applicable. In many investigations the identity of the analyte is assumed and the correctness of the assumption is merely confirmed. With some products of natural origin, complete identification and characterization is not possible. In these cases identification often may be fixed by chemical, chromatographic, or spectrophotometric fingerprinting—producing a reproducible pattern of reactions or characteristic output signals (peaks) with respect to position and intensity.

For botanical products, provide:

- Common or usual name of the item
- Synonyms by which it is known
- Botanical classification (variety, species, genus, family)

- Active or characteristic ingredient(s) (name and Chemical Abstracts Registry number or Merck Index number) and its chemical class. If the activity is ascribable to a mixture, provide the spectral or chromatographic fingerprint and the identity of the identifiable signals.

2.3 Method of Analysis or Protocol

The protocol or method of analysis is the set of permanent instructions for the conduct of the method of analysis. The method of analysis that is finally used should be the same as the one that was studied and revised as a result of research, optimization, and ruggedness trials and edited to conform with principles and practices for the production of *Official Methods of Analysis of AOAC INTERNATIONAL* (OMA). At this point the text is regarded as fixed. Substantive changes (those other than typographical and editorial) can only be made by formal public announcement and approval.

This text should be in ISO-compatible format where the major heads follow in a logical progression [e.g., Title, Applicability (Scope), Equipment, Reagents, Text, Calculations, with the addition of any special sections required by the technique, e.g., chromatography, spectroscopy]. Conventions with respect to reagents and laboratory operations should follow those given in the section "Definition of Terms and Explanatory Notes," which explains that "water is distilled water," reagents are of a purity and strength defined by the American Chemical Society (note that these may differ from standards set in other parts of the world), alcohol is the 95% aqueous mixture, and similar frequently used working definitions.

AOAC-approved methods may be considered as "well-recognized test methods" as used by ISO 17025. This document requires that those method properties, which may be major sources of uncertainties of measurements, be identified and controlled. In AOAC methods the following operations or conditions, which may be major contributors to uncertainties, should be understood to be within the following limits, unless otherwise specified more strictly or more loosely:

- Weights: Within ±10% (but use actual weight for calculations)
- Volumes: Volumetric flasks, graduates, and transfer pipets (stated capacity with negligible uncertainty)
- Burets: Stated capacity except in titrations
- Graduated pipets: Use volumes >10% of capacity
- Temperatures: Set to within ±2°
- pH: Within ±0.05 unit
- Time: Within ±5%

If the operational settings are within these specifications, together with any others derived from the supporting studies, the standard deviation obtained from these supporting studies in the same units as the reported result with the proper number of significant figures, usually 2 or 3, may be used as the standard measurement uncertainty.

2.3.1 Optimization

Prior to determining the performance parameters, the method should be optimized so that it is fairly certain that the properties of the "final method" are being tested. Validation is not a substitute for method development or for method optimization. If, however, some of the validation requirements have already been performed during the development phase, there is no need to repeat them for the validation phase. A helpful introduction is the AOAC publication "Use of Statistics to Develop and Evaluate Analytical Methods" by Grant T. Wernimont. This volume has only three major chapters: the measurement process, intralaboratory studies, and interlaboratory studies. No simpler explanation in understandable chemical terms exists of the analysis of variance than that given in pages 28–31. It supplements, explaining in greater detail, the concepts exemplified in the popular "Statistical Manual of AOAC" by W.J. Youden. Other useful references are *Appendices D* and *E* of OMA.

2.3.2 Reference Standard

All chemical measurements require a reference point. Classical gravimetric methods depend on standard weights and measures, which are eventually traceable to internationally recognized (SI) units. But modern analytical chemistry depends on other physical properties in addition to mass and length, usually optical or electrical, and their magnitude is based upon an instrumental comparison to a corresponding physical signal produced from a known mass or concentration of the "pure" analyte. If the analyte is a mixture, the signals or components must be separated and the signal from each compound compared to the signal from a known mass or concentration of the pure material or expressed in terms of a single reference compound of constant composition.

All instrumental methods require a reference material, even those that measure an empirical analyte. An "empirical analyte" is an analyte or property whose value is not fixed as in stoichiometric chemical compounds but which is the result of the application of the procedure used to determine it; examples are moisture, ash, fat, carbohydrate (by difference), and fiber. It is a "method-dependent analyte." Usually the reference material or "standard," which are specific chemical compounds, can be purchased from a supplier of chemicals and occasionally from a national metrological institute. When used for reference purposes, a statement should accompany the material certifying the identity, the purity and its uncertainty, how this was measured (usually by spectroscopy or chromatography), and its stability and storage conditions. If no reference material is available, as with many isolates from botanical specimens, an available compound with similar properties may serve as a surrogate standard—a compound that is stable and which behaves like the analyte but which is well resolved from it. Sometimes an impure specimen of the analyte must serve temporarily as the reference material until a purer specimen becomes available. The measured values assigned to empirical analytes are determined by strict adherence to all the details of the method of analysis. Even so, their bias and variability are usually larger (poorer) than chemically specified analytes. In some cases, as in determining the composition of milk by instrumental methods, the reference values for fat, protein, and lactose are established by use of reference methods. In routine operation, the bias and uncertainty of the final values are the combination of the uncertainties and bias correction arising from the routine operation with that of the reference values used for the calibration.

Modern instrumentation is complicated and its operation requires training and experience not only to recognize acceptable performance but also to distinguish unacceptable performance, drift, and deterioration on the part of the components. Continuous instruction and testing of the instruments and operators with in-house and external standards and proficiency exercises are necessary.

The records and report must describe the reference material, the source, and the basis for the purity statement (certification by the supplier is often satisfactory). If the reference material is hygroscopic, it should be dried before use either in a 100°C oven, if stable, or over a drying agent in a desiccator if not. The conversion factor of the analyte to the reference material, if different, and its

uncertainty must be established, often through spectrophotometric or chromatographic properties such as absorptivity or peak height or area ratios.

For recovery experiments the reference standard should be the highest purity available. In the macro concentration range (defined as about 0.1–100%) the standard ordinarily approaches 100%; in the micro or trace (defined as μg/g to 0.1%) and ultramicro or ultratrace range (μg/g and below) the standard should be at least 95% pure. The purity of rare or expensive standards is often established, referenced, and transferred through an absorptivity measurement in a specific solvent. The impurities present should not interfere significantly with the assay.

2.3.3 Ruggedness Trial

Although the major factors contributing to variability of a method may be explored by the classical, one variable at a time procedure, examining the effect of less important factors can be accomplished by a simpler Youden Ruggedness Trial [Youden, W.J., & Steiner, E.H. (1975) *Statistical Manual of the Association of Official Analytical Chemists*, pp 50–55]. This design permits exploring the effect of 7 factors in a single experiment requiring only eight determinations. It also permits an approximation of the expected standard deviation from the variability of those factors that are "in control." An example of exploring the extraction step of the determination of the active ingredient in a botanical is detailed in *Annex B*.

2.3.4 Specific Variables

If a variable is found to have an influence on the results, further method development is required to overcome the deficiency. For example, extraction of botanicals is likely to be incomplete and there are no reference materials available to serve as a standard for complete extraction. Therefore various techniques must be applied to determine when extraction is complete; reextraction with fresh solvent is the most common. Considerable experimentation also may be necessary to find the optimum conditions, column, and solvents for chromatographic isolation of the active ingredient(s).

(**a**) *Analyte addition*.—Addition of a solution of the active ingredient to the test sample and conducting the analysis is generally uninformative because the added analyte is already in an easily extractable form. The same is true for varying the volume of the extracting solvent. These procedures do not test the extractability of the analyte embedded in the cell structure. For this purpose, other variables must be tried, such as changing the solvent polarity or the extraction temperature.

(**b**) *Reextraction of the extracted residue*.—Reextraction after an original extraction will test for complete extraction by the original procedure. It will not test for complete extraction from intractable (unextractable) plant material. For this purpose a reagent that will destroy fibrous cellular material without damaging the active ingredient is required. If the analytes will not be destroyed or interfered with by cell wall disrupting or crude fiber reagents (1.25% H_2SO_4 and 1.25% NaOH) and are water soluble, use these solutions as extractives. But since the active ingredients are likely to contain compounds hydrolysable by these reagents, mechanical grinding to a very fine mesh will be the more likely choice.

The efficiency of extraction is checked by application of the extract to TLC, GLC, or HPLC chromatography. Higher total extractables is not necessarily an indicator of better extraction. The quantification of the active ingredient(s) is the indicator of extraction. Many natural compounds are sensitive to light and the decrease of a component suggests that the effect of this variable should be investigated.

(**c**) *Comparison with different solvents*.—Solvents with different polarities and boiling points will extract different amounts of extractives, but the amount of active ingredient(s) must be pursued by chromatographic separation or by specific reactions.

(**d**) *Comparison with results from a different procedure*.—A number of analyte groups, e.g., pesticide residues, have several different standard methods available based on different principles to provide targets for comparison.

(**e**) *System suitability checks*.—Chromatographic systems of columns, solvents (particularly gradients), and detectors are extremely sensitive to changes in conditions. Chromatographic properties of columns change as columns age and changes in polarity of solvents or temperature must be made to compensate. Therefore the specified properties of chromatographic systems in standard methods such as column temperatures and solvent compositions are permitted to be altered in order to optimize and stabilize the chromatographic output—peak height or area, peak resolutions, and peak shape. Similarly optical filters, electrical components of circuits, and mechanical components of instruments deteriorate with age and adjustments must be made to compensate. Specifications for instruments, and their calibration and operation must be sufficiently broad to accommodate these variations.

3 Performance Characteristics

The performance characteristics are required to determine if the method can be used for its intended purpose. The number of significant figures attached to the value of the characteristic generally indicates the reliability of these indices. They are generally limited by the repeatability standard deviation, sr. In most analytical work requiring calibration the best relative sr that can be achieved is about 1%. This is equivalent to the use of 2 significant figures. However, in order to avoid loss of "accuracy" in averaging operations, carry one additional figure with all reported values, i.e., use at most 3 significant figures in reporting. This statement, however, does not apply to recorded raw data, such as weighing or instrument readings, calibration, and standardization, which should utilize the full reading capacity of the measurement scales. This exception is limited by the measurement scale with the least reading capacity.

The purpose of the analysis determines which attributes are important and which may be less so.

3.1 Applicability (Scope)

A method must demonstrate acceptable recovery and repeatability with representative matrices and concentrations to which it is intended to be applied. For single materials, use at least three typical specimens, at least in duplicate, with different attributes (appearance, maturity, varieties, age). Repeat the analyses at least one day later. The means should not differ significantly and the repeatability should approximate those listed in *Section 3.4.2* for the appropriate concentration. If the method is intended to be applied to a single commodity, e.g., fruits, cereals, fats, use several representative items of the commodity with a range of expected analyte concentrations. If the method is intended to apply to "foods" in general, select representative items from the food triangle [Sullivan, D.M., & Carpenter, D.E. (1993) "Methods of Analysis for Nutrition Labeling," AOAC INTERNATIONAL, Gaithersburg, MD, pp 115–120]. In the case of residues, the matrices are generalized into categories such as "fatty foods" and "nonfatty foods" that require different preliminary treatments

to remove the bulk of the "inert" carrier. In all cases, select test materials that will fairly represent the range of composition and attributes that will be encountered in actual practice. Applicability may be inferred to products included within tested extremes but cannot be extrapolated to products outside the tested limits.

Similarly the range of expected concentrations should be tested in a number of typical matrices, spiking if necessary, to ensure that there is no interaction of analyte with matrix.

Semipermanent "house standards" for nutrients often can be prepared from a homogeneous breakfast cereal for polar analytes and from liquid monounsaturated oil like olive oil for nonpolar analytes for use as concurrent controls or for fortification.

The authority for the authenticity of botanical specimens and their source and the origin or history of the test materials must be given.

The determination of freedom from the effects of interfering materials is tested under selectivity, *Section 3.2*, and properties related to the range of quantification of the target analyte are tested under the reliability characteristics, *Section 3.4*.

3.2 Selectivity

The term selectivity is now generally preferred by IUPAC over specificity.

Selectivity is the degree to which the method can quantify the target analyte in the presence of other analytes, matrices, or other potentially interfering materials. This is usually achieved by isolation of the analyte through selective solvent extraction, chromatographic or other phase separations, or by application of analyte-specific techniques such as biochemical reactions (enzymes, antibodies) or instrumentation [nuclear magnetic resonance (NMR), infrared, or mass spectrometry (MS)].

Methods must be tested in the presence of accompanying analytes or matrices most likely to interfere. Matrix interference is usually eliminated by extraction procedures and the desired analyte is then separated from other extractives by chromatography or solid-phase extraction. Nevertheless, many methods for low-level analytes still require a matrix blank because of the presence of persistent, nonselective background.

The most useful separation technique is chromatography and the most important requirement is resolution of the desired peak from accompanying peaks. Resolution, R_s, is expressed as a function of both the absolute separation distance expressed as retention times (minutes) of the two peaks, t_1 and t_2, and the baseline widths, W_1 and W_2, of the analyte and nearest peak, also expressed in terms of times, as

$$R_s = 2 (t_2 - t_1) / (W_1 + W_2)$$

Baseline widths are measured by constructing tangents to the two sides of the peak band and measuring the distance between the intersection of these tangents with the baseline or at another convenient position such as half-height. A resolution of at least 1.5 is usually sought and one of 1.0 is the minimum usable separation. The U.S. Food and Drug Administration (FDA) suggests an R_s of at least 2 for all compounds accompanying active drug dosage forms, including hydrolytic, photolytic, and oxidative degradation products. In addition, the isolated analyte should show no evidence of other compounds when chromatographed on other systems consisting of different columns and solvents, or when examined by techniques utilized for specificity (infrared, NMR, or MS). These requirements were developed for synthetic drug substances, and must be relaxed for the families of compounds commonly encountered in foods and botanical specimens to a resolution of 1.5 from adjacent nontarget peaks.

If the product is mixed with other substances, the added substances must be tested to ensure that they do not contain any material that will interfere with the identification and determination of the analyte sought. If the active constituent is a mixture, the necessity for separation of the ingredients is a decision related to the complexity of the potential separation, the constancy of the relationship of the components, and the relative biological activity of the constituents.

3.3 Calibration

Modern instrumental methods depend upon the comparison of a signal from the unknown concentration of an analyte to that from a known concentration of the same or similar analyte. This requires the availability of a reference standard, *Section 2.2.2*. The simplest calibration procedure requires preparation of a series of standard solutions from the reference material, by dilution of a stock solution, covering a reasonable range of signal response from the instrument. Six to 8 points, approximately equally spaced over the concentration range of interest, performed in duplicate but measured at random (to avoid confusing nonlinearity with drift) is a suitable calibration pattern. Fit the calibration line (manually or numerous statistical and spreadsheet programs are available) and plot the residuals (the difference of the experimental points from the fitted line) as a function of concentration. An acceptable fit produces a random pattern of residuals with a 0 mean. For checking linearity, prepare the individual solutions by dilution from a common stock solution to avoid the random errors likely to be introduced from weighing small (mg) quantities for individual standards.

As long as the purity of the reference material is 95% or greater, as determined by evaluating secondary peaks or spots in gas, liquid, or thin-layer chromatography or other quantitative technique, the impurities contributes little to the final variance at micro- and ultramicro concentrations and may be neglected. (Recovery trials, however, require greater purity or correction for the impurities.) The identity of the material used as the reference material, however, is critical. Any suggestion of nonhomogenity such as multiple or distorted peaks or spots, insoluble residue, or appearance of new peaks on standing requires further investigation of the identity of the standard.

Similarly, certified volumetric glassware may also be used after initial verification of their stated capacity by weighing the indicated volume of water for flasks and the delivered volume for pipets and burets and converting the weight to the volume delivered.

Do not use serological pipets at less than 10% of their graduated capacity. Check the stability of the stock and initial diluted solutions, stored at room or lower temperatures, by repeating their measurements several days or weeks later. Prepare the most dilute solutions fresh as needed from more concentrated, stable solutions in most cases. Bring solutions stored at refrigerator or lower temperatures to room temperature before opening and using them.

Plot the signal response against the concentration. A linear response is desirable as it simplifies the calculations, but it is not necessary nor should it be regarded as a required performance characteristic. If the curve covers several orders of magnitude, weighted regression, easily handled by computer programs, may be useful. Responses from electrochemical and immunological methods are exponential functions, which often may be linearized by using logarithms. Some instruments perform signal-to-concentration calculations automatically using disclosed or undisclosed algorithms. If the method is not used routinely, several standards should accompany

the test runs. If the method is used routinely, the standard curve should be repeated daily or weekly, depending on its stability. Repeat the standard curve as frequently as necessary with those instruments where drift is a significant factor.

A high correlation coefficient (e.g., >0.99) is often recommended as evidence of goodness of fit. Such use of the correlation coefficient as a test for linearity is incorrect [Analytical Methods Committee, *Analyst* **113**, 1469–1471(1988); **119**, 2363(1994)]. Visual examination is usually sufficient to indicate linearity or nonlinarity, or use the residual test, *Section 3.3*.

If a single (parent or associated) compound is used as the reference material for a series of related compounds, give their relationship in structure and response factors.

Note that the calibration is performed directly with the analyte reference solutions. If these reference solutions are carried through the entire procedure, losses in various steps of the procedure cannot be explored but are automatically compensated for. Some procedures require correction of the final result for recovery. When this is necessary, use a certified reference material, a "house" standard, or analyte added to a blank matrix conducted through the entire method for this purpose. If several values are available from different runs, the average is usually the best estimate of recovery. Differences of calibration curves from day to day may be confused with matrix effects because they are often of the same magnitude.

3.3.1 External Standard Method

The most common calibration procedure utilizes a separately prepared calibration curve because of its simplicity. If there is a constant loss in the procedure, this is handled by a correction factor, as determined by conducting a known amount of analyte through the entire procedure. The calculation is based on the ratio of the response of equal amounts of the standard or reference compound to the test analyte. This correction procedure is time consuming and is used as a last resort since it only improves accuracy at the expense of precision. Alternatives are the internal standard procedure, blank matrix process, and the method of standard addition.

If the method is intended to cover a substantial range of concentrations, prepare the curve from a blank and five or seven approximately equally spaced concentration levels and repeat on a second day. Repeat occasionally as a check for drift. If an analyte is examined at substantially different concentration levels, such as pesticide residues and formulations, prepare separate calibration curves covering the appropriate range to avoid excessive dilutions. In such cases, take care to avoid cross contamination. However, if the analyte always occurs at or near a single level as in a pharmaceutical, a 2-point curve may be used to bracket the expected level, or even a single standard point, if the response over the range of interest is approximately linear. By substituting an analyte-free matrix preparation for the blank, as might be available from pesticide or veterinary drug residue studies or the excipients from a pharmaceutical, a calibration curve that automatically compensates for matrix interferences can be prepared.

3.3.2 Internal Standard Method

The internal standard method requires the addition of a known amount of a compound that is easily distinguished from the analyte but which exhibits similar chemical properties. The response ratio of the internal standard to a known amount of the reference standard of the analyte of interest is determined beforehand. An amount of internal standard similar to that expected for the analyte is added at an early stage of the method. This method is particularly useful for addition to the eluate from an HPLC separation when the fractions are held in an autosampler that is run overnight, where it compensates for any losses of solvent by evaporation. An internal standard is also frequently used in GLC residue methods where many analytes with similar properties are frequently encountered.

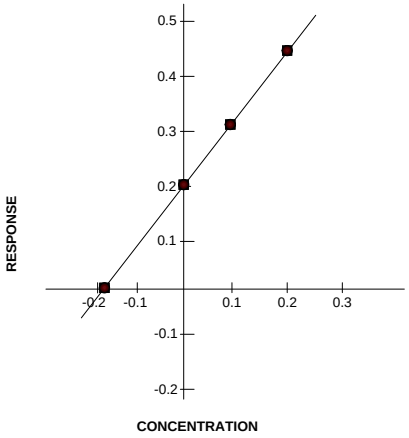


Figure 1

3.3.3 Standard Addition Method

When the matrix effect on an analyte is unknown or variable, the method of standard additions is useful. Make measurements on the isolated analyte solution and add a known amount of the standard analyte at the same level and at twice or three (or known fractions) times the original level. Plot the signal against the concentration with the initial unknown concentration set at 0. Extrapolate the line connecting the measured responses back to 0 response and read the concentration value off the (negative) *x*-axis. The main assumption is that the response is linear in the working region. This method is used most frequently with emission spectroscopy, electrochemistry, and radiolabeled isotopes in mass spectrometric methods.

See Figure 1 for example [from Rubinson, K.A. (1987) "Chemical Analysis," Little, Brown and Co., Boston, MA, p. 205].

Concn Cu added, μg	Instrument response
0.0	0.200
0.10	0.320
0.20	0.440
Concn Cu found by extrapolation to 0.00 response	(–)0.18

3.4 Reliability Characteristics

These are the statistical measures of how good the method is. Different organizations use different terms for the same concept. The important questions are:

- How close is the reported value to the true, reference, or accepted value?
- How close are repeated values to each other as determined in the same or different laboratories?
- What is the smallest amount or concentration that can be recognized or measured?

Recently accreditation organizations have been requesting the calculation of the parameter "Measurement Uncertainty" (MU). This is a term indicative of the reliability of the particular series of

measurements being reported. The standard uncertainty is equal to the standard deviation of the series of measurements of the analyte. The expanded uncertainty is two times the standard uncertainty and is expected to encompass about 95% of similar future measurements. If too few values are available in a measurement series to calculate a stable MU, the standard deviation obtained from the validation study within the laboratory, s_r, may be substituted, if it covered the same or similar analyte/matrix/concentration range. If a collaboratively studied method is being validated for use within a laboratory, the standard deviation among-laboratories, s_R, reported for the method from the study should be used to determine if the anticipated measurement uncertainty will be satisfactory for the intended purpose, assuming satisfactory repeatability as demonstrated by control charts or proficiency testing. In fact, the determination of the reliability characteristics in the validation study should not be undertaken until the developmental work demonstrates that the data are repeatable and in statistical control.

The Codex Alimentarius, an international body organized by the Food and Agricultural Organization (FAO) and the World Health Organization (WHO) of the United Nations (UN) to recommend international food standards to governments, suggests the following "Guidelines for the Assessment of the Competence of Testing Laboratories Involved in the Import and Export Control of Food" (FAO, Rome, Italy, CAC/GL 27-1997) for laboratories:

- Comply with the general competence criteria of ISO 17025
- Participate in proficiency testing schemes for food analysis
- Utilize validated methods
- Utilize internal quality control procedures

3.4.1 Accuracy

The term "accuracy" has been given so many meanings that it is better to use a more specific term. Ordinarily it means closeness of the test result to the "true" or accepted value. But the test result can be an individual value, the average of a set of values, or the average of many sets of values. Therefore, whenever the term is used, the number of values it represents and their relationship must always be stated, e.g., as an individual result, as the average of duplicates or n replicates, or as the average of a set of a number of trials. The difference of the reported value from the accepted value, whether it is an individual value, an average of a set of values, or the average of a number of averages, or an assigned value, is the bias under the reported conditions. The frequently used term for bias or "accuracy" when the average of a set of values is reported is "trueness."

The fraction or percentage of the analyte that is recovered when the test sample is conducted through the entire method is the recovery. The best reference materials for determining recovery are analyte-certified reference materials (CRMs) distributed by national metrological laboratories, but in most cases material certified by a commercial supplier must be accepted. Occasionally standards are available from a government agency, such as pesticides from the Environmental Protection Agency (EPA). They are rarely, if ever, available in the matrix of interest but rather as a solution in a convenient solvent with a stated concentration and uncertainty. Such reference materials must then be tested in the matrix of interest. Even rarer is an isotopically labeled analyte that can be easily followed by isotopic analytical techniques.

The available certified or commercial analyte standard, diluted if necessary, is added to typical analyte-free matrices at levels about 1x or 2x the expected concentration. Analyte-free matrices for residues are obtained from growers who certify that the chemical is not used in their cultivation, growth, or feeding and verified analytically. They may also be obtained from the residues of previously extracted materials or from test samples shown to be negative for the analyte.

If an analyte-free matrix is not available, the analyte standard is added to separate test portions and the recovery is calculated from the base determined by the method of addition, *Section 3.3.3*. Run the set of such controls with each set of test samples. If a sufficient number of batches are expected to be run (at least 20–30), the % recovery can be plotted against the run number as the basis for a control chart. Recovery also can be obtained as a byproduct of the precision determinations, *Sections 3.4.2* and *3.4.4*.

Acceptable recovery is a function of the concentration and the purpose of the analysis. Some acceptable recovery requirements for individual assays are as follows:

Concentration	Recovery limits, %
100%	98–101
10%	95–102
1%	92–105
0.1%	90–108
0.01%	85–110
10 μg/g (ppm)	80–115
1 μg/g	75–120
10 μg/kg (ppb)	70–125

The Codex Alimentarius "Residues of Veterinary Drugs in Foods" [2nd Ed., Vol. 3 (1993) Joint FAO/WHO Food Standards Program, FAO, Rome, Italy, p. 59] suggests the following limits for residues of veterinary drugs in foods:

Concentration, μg/kg	Acceptable range
≤1	50–120
≥1 < 10	60–120
≥10 < 100	70–110
≥100	80–110

These limits may be modified as needed in view of the variability of individual results or which set of regulatory requirements are referenced. (As a rough guide to typical performance, about 95% of normally distributed typical results in a single laboratory at 1 μg/g will fall within 80–120% of the mean.) In the case of the examination of the general USDA pesticide residue proficiency study, limits of 50–150% were applied; the USFDA acceptability criterion for recovery of drug residues at the 10 ppb level is 70–120%. Generally, however, recoveries less than 60–70% should be subject to investigations leading to improvement and average recoveries greater than 110% suggest the need for better separations. Most important, recoveries greater than 100% must not be discarded as impossible. They are the expected positive side from a typical distribution of analytical results from analytes present at or near 100% that are balanced by equivalent results on the negative side of the mean.

If an extraction of active ingredient from a matrix with a solvent is used, test extraction efficiency by reextracting the (air-dried) residue and determining the active ingredient(s) in the residue by the method.

The number of units to be used to establish bias is arbitrary, but the general rule is the more independent "accuracy" trials, the better. The improvement, as measured by the width of the confidence interval for the mean, follows the square root of the number of trials. Once past 8–10 values, improvement comes slowly. To fully contribute, the values must be conducted independently, i.e., nonsimultaneously, throwing in as many environmental or spontaneous differences as possible, such as different analysts, instruments, sources of reagents, time of day,

temperature, barometric pressure, humidity, power supply voltage, etc. Each value also contributes to the within-laboratory precision as well. A reasonable compromise is to obtain 10 values from a reference material, a spiked matrix, or by the method of standard addition scattered over several days or in different runs as the basis for checking bias or recovery. By performing replicates, precision is obtained simultaneously. Precision obtained in such a manner is often termed "intermediate precision" because its value is between within-laboratory and among-laboratory precision. When reported, the conditions that were held constant and those that were varied must be reported as well.

Note that the series of determinations conducted for the method of addition are not independent because they are probably prepared from the same standard calibration solution, same pipets, and are usually conducted almost simultaneously. This is satisfactory for their intended purpose of providing an interrelated function, but it is not satisfactory for a precision function estimation intended for future use.

Related to recovery is the matter of reporting the mean corrected or not corrected for recovery. Unless specifically stated in the method to correct or not, this question is usually considered a "policy" matter and is settled administratively outside the laboratory by a regulatory pronouncement, informal or formal agreement, or by contract. If for some reason a value closest to theory is needed, correction is usually applied. If a limit or tolerance has been established on the basis of analytical work with the same method correlated with "no effect" levels, no correction should be applied because it has already been used in setting the specification. Corrections improve "accuracy" at the expense of impairing precision because the variability of both the determination and the recovery are involved.

When it is impossible to obtain an analyte-free matrix to serve as a base for reporting recovery, two ways of calculating recovery must be distinguished: (*1*) Total recovery based on recovery of the native plus added analyte, and (*2*) marginal recovery based only on the added analyte (the native analyte is subtracted from both the numerator and denominator). Usually total recovery is used unless the native analyte is present in amounts greater than about 10% of the amount added, in which case use the method of addition, *Section 3.3.3*.

When the same analytical method is used to determine both the concentration of the fortified, C_f, and unfortified, C_u, test samples, the % recovery is calculated as

$$\text{Recovery, \%} = (C_f - C_u) \times 100/C_a$$

where C_a is the calculated (not analyzed) concentration of analyte added to the test sample. The concentration of added analyte should be no less that the concentration initially present and the response of the fortified test sample must not exceed the highest point of the calibration curve. Both fortified and unfortified test samples must be treated identically in the analysis.

3.4.2 Repeatability Precision (s_r, RSD_r)

Repeatability refers to the degree of agreement of results when conditions are maintained as constant as possible with the same analyst, reagents, equipment, and instruments performed within a short period of time. It usually refers to the standard deviation of simultaneous duplicates or replicates, s_r. It is the best precision that will be exhibited by a laboratory but it is not necessarily the laboratory's typical precision. Theoretically the individual determinations should be independent but this condition is practically impossible to maintain when determinations are conducted simultaneously and therefore this requirement is generally ignored.

To obtain a more representative value for the repeatability precision perform the simultaneous replicates at different times (but the same day), on different matrices, at different concentrations. Calculate the standard deviation of repeatability from at least five pairs of values obtained from at least one pair of replicates analyzed with each batch of analyses for each pertinent concentration level that differs by approximately an order of magnitude and conducted at different times. The object is to obtain representative values, not the "best value," for how closely replicates will check each other in routine performance of the method. Therefore these sets of replicate analyses should be conducted at least in separate runs and preferably on different days. The repeatability standard deviation varies with concentration, C expressed as a mass fraction. Acceptable values approximate the values in the following table or calculated by the formula:

$$RSD_r, \% = 2C^{-0.15}$$

unless there are reasons for using tighter requirements.

Concentration	Repeatability (RSD_r), %
100%	1
10%	1.5
1%	2
0.1%	3
0.01%	4
10 μg/g (ppm)	6
1 μg/g	8
10 μg/kg (ppb)	15

Acceptable values for repeatability are between ½ and 2 times the calculated values. Alternatively a ratio can be calculated of the found value for RSD_r to that calculated from the formula designated as $HorRat_r$. Acceptable values for this ratio are typically 0.5 to 2:

$$HorRat_r = RSD_r \text{ (found, \%)}/RSD_r \text{ (calculated, \%)}$$

The term "repeatability" is applied to parameters calculated from simultaneous replicates and this term representing minimum variability is equated to the "within-laboratory" parameter (standard deviation, variance, coefficient of variation, relative standard deviation) of the precision model equation. It should be distinguished from a somewhat larger within-laboratory variability that would be induced by non-simultaneous replicates conducted in the same laboratory on identical test samples on different days, by different analysts, with different instruments and calibration curves, and with different sources of reagents, solvents, and columns. When such an "intermediate" within-laboratory precision (standard deviation, variance, coefficient of variation, relative standard deviation) is used, a statement of the conditions that were not constant must accompany it. These within-laboratory conditions have also been called within-laboratory reproducibility, an obvious misnomer.

3.4.3 Measurement Uncertainty

Accreditation organizations have been requesting laboratories to have a parameter designated as "measurement uncertainty" associated with methods that the laboratory utilizes. The official metrological definition of measurement uncertainty is "a parameter

associated with the result of a measurement that characterizes the dispersion of values that could reasonably be attributed to the measurand." A note indicates, "the parameter may be, for example, a standard deviation (or a given multiple of it), or the width of a confidence interval."

Of particular pertinence is the fact that the parameter applies to a measurement and not to a method (*see Section 3.4*). Therefore "standard" measurement uncertainty is the standard deviation or relative standard deviation from a series of simultaneous measurements. "Expanded" uncertainty is typically twice the standard uncertainty and is considered to encompass approximately 95% of future measurements. This is the value customarily used in determining if the method is satisfactory for its intended purpose although it is only an approximation because theoretically it applies to the unknown "true" concentration.

Since the laboratory wants to know beforehand if the method will be satisfactory for the intended purpose, it must use the parameters gathered in the validation exercises for this purpose, substituting the measurement values for the method values after the fact. As pointed out by M. Thompson [*Analyst* **125**, 2020–2025 (2000); *see Inside Lab. Mgmt*. **5**(2), 5(2001)], a ladder of errors exist for this purpose.

- Duplicate error (a pair of tests conducted simultaneously)
- Replicate or run error (a series of tests conducted in the same group)
- Within-laboratory error (all tests conducted by a laboratory)
- Between-laboratory error (all tests by all laboratories)

As we go down the series, the possibility of more errors being included is increased until a maximum is reached with the all inclusive reproducibility parameters. Thompson estimates the relative magnitude of the contribution of the primary sources of error as follows

Level of variation	Separate	Cumulative
Repeatability	1.0	1.0
Runs	0.8	1.3
Laboratories	1.0	1.6
Methods	1.5	2.2

Ordinarily only one method exists or is being validated so we can ignore the last line. Equating duplicates to replicability, runs to within-laboratory repeatability, and laboratories to among-laboratories reproducibility, Thompson points out that the three sources of error are roughly equal and not much improvement in uncertainty would result from improvement in any of these sources. In any case, the last column gives an approximate relative relationship of using the standard deviation at any point of the ladder as the basis for the uncertainty estimate prior to the actual analytical measurements.

In the discussion of uncertainty it must be noted that bias as measured by recovery is not a component of uncertainty. Bias (a constant) should be removed by subtraction before calculating standard deviations. Differences in bias as exhibited by individual laboratories become a component of uncertainty through the among-laboratory reproducibility. The magnitude of the uncertainty depends on how it is used—comparisons within a laboratory, with other laboratories, and even with other methods. Each component adds uncertainty. Furthermore, uncertainty stops at the laboratory's edge. If only a single laboratory sample has been submitted and analyzed, there is no basis for estimating sampling uncertainty. Multiple independent samples are required for this purpose.

3.4.4 Reproducibility Precision (s_R, RSD_R)

Reproducibility precision refers to the degree of agreement of results when operating conditions are as different as possible. It usually refers to the standard deviation (s_R) or the relative standard deviation (RSD_R) of results on the same test samples by different laboratories and therefore is often referred to as "between-laboratory precision" or the more grammatically correct "among-laboratory precision." It is expected to involve different instruments, different analysts, different days, and different laboratory environments and therefore it should reflect the maximum expected precision exhibited by a method. Theoretically it consists of two terms: the repeatability precision (within-laboratory precision, s_r) and the "true" between-laboratory precision, s_L. The "true" between-laboratory precision, sL, is actually the pooled constant bias of each individual laboratory, which when examined as a group is treated as a random variable. The between-laboratory precision too is a function of concentration and is approximated by the Horwitz equation, $s_R = 0.02C^{0.85}$. The AOAC/IUPAC protocol for interlaboratory studies requires the use of a minimum of eight laboratories examining at least five materials to obtain a reasonable estimate of this variability parameter, which has been shown to be more or less independent of analyte, method, and matrix.

By definition s_R does not enter into single-laboratory validation. However, as soon as a second (or more) laboratory considers the data, the first question that arises involves reanalysis by that second laboratory: "If I had to examine this or similar materials, what would I get?" As a first approximation, in order to answer the fundamental question of validation—fit for the intended purpose—assume that the recovery and limit of determination are of the same magnitude as the initial effort. But the variability, now involving more than one laboratory, should be doubled because variance, which is the square of differences, is involved, which magnifies the effect of this parameter. Therefore we have to anticipate what another laboratory would obtain if it had to validate the same method. If the second laboratory on the basis of the doubled variance concludes the method is not suitable for its intended purpose, it has saved itself the effort of revalidating the method.

In the absence of such an interlaboratory study, the interlaboratory precision may be estimated from the concentration as indicated in the following table or by the formula (unless there are reasons for using tighter requirements):

$$RSD_R = 2C^{-0.15}$$

or

$$S_R = 0.02C^{0.85}$$

Concentration, C	Reproducibility (RSD_R), %
100%	2
10%	3
1%	4
0.1%	6
0.01%	8
10 μg/g (ppm)	11
1 μg/g	16
10 μg/kg (ppb)	32

Acceptable values for reproducibility are between ½ and 2 times the calculated values. Alternatively a ratio can be calculated

of the found value for RSD_R to that calculated from the formula designated as $HorRat_R$. Acceptable values for this ratio are typically 0.5 to 2:

$$HorRat_R = RSD_R \text{ (found, \%)}/RSD_R \text{ (calculated, \%)}$$

As stated by Thompson and Lowthian ("The Horwitz Function Revisited," (1997) *J. AOAC Int.* **80**, 676–679), "Indeed, a precision falling within this 'Horwitz Band' is now regarded as a criterion for a successful collaborative trial."

The typical limits for HorRat values may not apply to indefinite analytes (enzymes, polymers), physical properties, or to the results from empirical methods expressed in arbitrary units. Better than expected results are often reported at both the high (>10%) and low (<E-8) ends of the concentration scale. Better than predicted results can also be attained if extraordinary effort or resources are invested in education and training of analysts and in quality control.

3.4.5 Intermediate Precision

The precision determined from replicate determinations conducted within a single laboratory not simultaneously, i.e., on different days, with different calibration curves, with different instruments, by different analysts, etc. is called intermediate precision. It lies between the within- and among-laboratories precision, depending on the conditions that are varied. If the analysis will be conducted by different analysts, on different days, on different instruments, conduct at least five sets of replicate analyses on the same test materials under these different conditions for each concentration level that differs by approximately an order of magnitude.

3.4.6 Limit of Determination

The limit of determination is a very simple concept: It is the smallest amount or concentration of an analyte that can be estimated with acceptable reliability. But this statement contains an inherent contradiction: the smaller the amount of analyte measured, the greater the unreliability of the estimate. As we go down the concentration scale, the standard deviation increases to the point where a substantial fraction of values of the distribution of results overlaps 0 and false negatives appear. Therefore the definition of the limit comes down to a question of what fraction of values are we willing to tolerate as false negatives.

Thompson and Lowthian (loc. cit.) consider the point defined by $RSD_R = 33\%$ as the upper bound for useful data, derived from the fact that $3RSD_R$ should contain 100% of the data from a normal distribution. This is equivalent to a concentration of about 8×10^{-9} (as a mass fraction) or 8 ng/g (ppb). Below this level false negatives appear and the data goes "out of control." From the formula, this value is also equivalent to an $RSD_r \approx 20\%$. The penalty for operating below the equivalent concentration level is the generation of false negative values. Such signals are generally accepted as negative and are not repeated.

An alternative definition of the limit of detection and limit of determination is based upon the variability of the blank. The blank value, x_{Bl}, plus 3 times the standard deviation of the blank ($x_{Bl} + 3s_{Bl}$) is taken as the detection limit and the blank value plus 10 times the standard deviation of the blank ($x_{Bl} + 10s_{Bl}$) is taken as the determination limit. The problem with this approach is that the blank is often difficult to measure or is highly variable. Furthermore, the value determined in this manner is independent of the analyte. If blank values are accumulated over a period of time, the average is likely to be fairly representative as a basis for the limits and will probably provide a value of the same magnitude as that derived from the relative standard deviation formulae.

The detection limit is only useful for control of undesirable impurities that are specified as "not more than" a specified low level and for low-level contaminants. Useful ingredients must be present at high enough concentrations to be functional. The specification level must be set high enough in the working range that acceptable materials do not produce more than 5% false-positive values, the default statistical acceptance level. Limits are often at the mercy of instrument performance, which can be checked by use of pure standard compounds. Limits of detection and determination are unnecessary for composition specifications although the statistical problem of whether or not a limit is violated is the same near zero as it is at a finite value.

Blank values must be monitored continuously as a control of reagents, cleaning of glassware, and instrument operation. The necessity for a matrix blank would be characteristic of the matrix. Abrupt changes require investigation of the source and correction. Taylor [J.K. Taylor (1987) "Quality Assurance of Chemical Measurements," Lewis Publishers, Chelsea, MI, p. 127] provides two empirical rules for applying a correction in trace analysis: (*1*) The blank should be no more than 10% of the "limit of error of the measurement", and (*2*) it should not exceed the concentration level.

3.4.7 Reporting Low-Level Values

Although on an absolute scale low level values are miniscule, they become important in three situations:

(*1*) When legislation or specifications decrees the absence of an analyte (zero tolerance situation).

(*2*) When very low regulatory or guideline limits have been established in a region of high uncertainty (e.g., a tolerance of 0.005 μg/kg aflatoxin M_1 in milk).

(*3*) When dietary intakes of low-level nutrients or contaminants must be determined to permit establishment of minimum recommended levels for nutrients and maximum limits for contaminants.

Analytical work in such situations not only strains the limits of instrumentation but also the ability of the analyst to interpret and report the findings. Consider a blank that is truly 0 and that the 10% point of the calibration curve corresponds to a concentration of 1 μg/kg (E-9). By the Horwitz formula this leads to an expected RSD_r in a single laboratory of about 23%. If we assume a normal distribution and we are willing to be wrong 5% of the time, what concentration levels would be expected to appear? From 2-tail normal distribution tables (the errant value could appear at either end), 2.5% of the values will be below 0.72 μg/kg and 2.5% will be above 1.6 μg/kg. Note the asymmetry of the potential results, from 0.7 to 1.6 μg/kg for a nominal 1.0 μg/kg value from the nature of the multiplicative scale when the RSD is relatively large.

But what does the distribution look like at zero? Mathematically it is intractable because it collapses to zero. Practically, we can assume the distribution looks like the previous one but this time we will assume it is symmetrical to avoid complications. The point to be made will be the same. For a distribution to have a mean equal to 0, it must have negative as well as positive values. But negative concentration values per se are forbidden but here they are merely an artifact of transforming measured signals. Negative signals are typical in electromotive force and absorbance measurements.

Analysts have an aversion to reporting a zero concentration value because of the possibility that the analyte might be present, but below the detection limit. Likewise, analysts avoid reporting

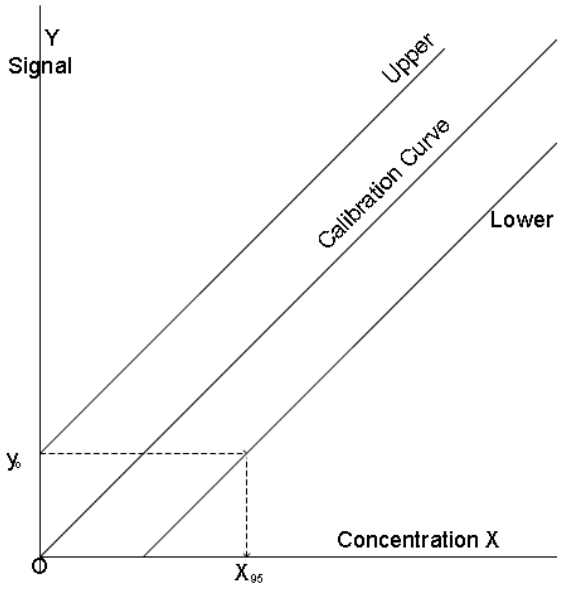


Figure 2. The statistical situation at the zero concentration level: A signal as high as y_0 could be measured at a 0 concentration, which corresponds to a "true" concentration value as high as x_{95}, but with only a 5% probability.

negative values as physical impossibilities although they are required by arithmetic averaging of random fluctuations to attain a real zero. Analysts avoid the issue by linguistic subterfuges such as "less than the detection limit" or by substituting an arbitrary fractional value such as one half the detection limit. Statisticians must discard such values as useless and consequently much effort is simply wasted by such reports.

Therefore the recommendation for handling low level values for validation purposes is to report whatever value is returned by converting the recorded instrument reading to a concentration through the calibration chart: positive, negative, or zero and rely on the power of averaging to produce the best estimate. As stated by the (UK) Analytical Methods Committee (*Anal. Tech. Brief No. 5*, April 2001), "analytical results are not concentrations but error-prone estimates of concentrations."

Such advice is impractical for reporting to a nontechnical or even a technical reviewer unfamiliar with the statistical problem of reporting results near zero. In such cases, the simplest solution is to report "zero" or "none found" for all signal values within the region of (blank value + 3 x (standard deviation of the blank signal)). This can be supplemented by a statement that the variability of results in the region of zero is such that it would permit as much as *x* µg/kg to be present with not more than a 5% probability, where *x* is roughly 5. If the laboratory can calculate the confidence interval of the calibration curve, a better estimate is obtained by drawing a line parallel to the *x*-axis from the *y* (signal) value where the upper confidence line intersects the *y*-axis (y_0) until it intersects the lower confidence line and reading the *x* (concentration) value (x_{95}) of the line parallel to the *y*-axis where it intersects the *x*-axis (*see* Figure 2). This curve can be used to supply a statement that any signal less than y_0 can be reported as "zero" or "none found" with only a 5% chance of being wrong.

3.4.8 Dichotomous Reporting (Qualitative Analysis)

In an effort to bypass the laborious effort to develop and validate a method of analysis, a request is often made to obtain a test that will merely verify the presence or absence of an analyte. Such a request assumes correctly that it is simpler to divide a continuum of measurements of a property into two parts than into more than two parts. This concept assigns all values on one side of the division as acceptable, positive, or present and all values on the other side as unacceptable, absent, or negative. Even assuming that it is easy to set a dividing value through an external specification, tolerance, or limit-setting procedure, we cannot escape the statistical problem of interpretation of a measured value because of the accompanying distribution or halo of uncertainty.

This problem was discussed many years ago in connection with the interpretation of very simple spot tests by Feigl, the developer of this technique [Feigl, F. (1943) "Laboratory Manual of Spot Tests," Academic Press, New York, NY]. "If the sensitivity of a spot reaction is checked by progressively diluting a given standard solution, and then at each dilution, one drop is taken for the test, different observers will practically never agree absolutely in their determinations of the identification limit, even though the same experimental conditions have been closely maintained by all. Almost always there will be a certain range of variation." (p. 4)

We now understand the reason for the "range of variation." It arises from the statistical distribution of any physical measurement characterized by a location parameter (mean) and a distribution parameter (standard deviation). Any single observation removed from the distribution at the dividing value could have been anywhere within the envelope of that distribution. Half of the observations will be above and half below even though the "true value" of the property is a fixed number. The property may be fixed, but the measurements are variable.

A qualitative test has been defined in terms of indicating if an analyte is present or absent, above or below a limit value, and as a test with "poorer" precision than a quantitative method. But all of these definitions degenerate into the single test of whether a measured value is significantly different (in a statistical sense) from a fixed value.

Consequently when a test is used in a qualitative manner, any anticipated gain in the number of test samples examined at the expense of reliability, is illusionary. The test is fundamentally no different from determining if a found value is above or below a quantitative specification value. When the concentration drops into a region of high measurement variability the signal degenerates from real measurements into false positives for the blanks and false negatives for the measurements.

Nevertheless, the Codex Alimentarius "Residues of Veterinary Drugs in Foods" [Vol. 3, 2nd Ed. (1993) Joint FAO/WHO Food Standards Program, FAO, Rome, Italy, pp 55–59] recognizes such methods as a Level III method to determine the presence or absence of a compound "at some designated level of interest." It anticipates that such methods involve microbiological or immunological principles and they "should produce less than 5% false negatives and less than 10% false positives when analysis is performed on the test sample." It is doubtful if the statistical properties (e.g., power) of this recommendation have been examined and if such requirements are achievable with a reasonable number of examinations. A rough calculation indicates that to achieve the required specification more than 200 independent tests on the same test sample would have to be made, a requirement that would probably exhaust the analytical sample before a dozen tests were made.

3.5 Controls

3.5.1 Control Charts

Control charts are only useful for large volume or continuous work. They require starting with at least 20–30 values to calculate a mean and a standard deviation, which form the basis for control values equivalent to the mean ± 2 s_r (warning limits) and the mean ± 3 s_r (rejection limits). At least replicate test portions of a stable house reference material and a blank are run with every batch of multiple test samples and the mean and standard deviations (or range of replicates) of the controls and blank are plotted separately. The analytical process is "in control" if not more than 5% of the values fall in the warning zone. Any value falling above the rejection limit or two consecutive values in the warning region requires investigation and corrective action.

3.5.2 Injection Controls

A limit of 1 or 2% is often placed on the range of values of the peak heights or areas or instrument response of repeated injections of the final isolated analyte solution. Such controls are good for checking stability of the instrument during the time of checking but give no information as to the suitability of the isolation part of the method. Such a limit is sometimes erroneously quoted as a relative standard deviation when range is meant.

3.5.3 Duplicate Controls

Chemists will frequently perform their analyses in duplicate in the mistaken belief that if duplicates check, the analysis must have been conducted satisfactorily. ISO methods often require that the determinations be performed in duplicate. Simultaneous replicates are not independent—they are expected to check because the conditions are identical. The test portions are weighed out using the same weights, aliquots are taken with the same pipets, the same reagents are used, operations are performed within the same time frame, instruments are operated with the same parameters, and the same operations are performed identically. Under such restraints, duplicates that do not check would be considered as outliers. Nevertheless, the parameter calculated from duplicates within a laboratory is frequently quoted as the repeatability limit, r, as equal to $2*\sqrt{2}*s_r$ and is expected to encompass 95% of future analyses conducted similarly. The corresponding parameter comparing two values in different laboratories is the reproducibility limit, $R = 2*\sqrt{2}*s_R$. This parameter is expected to reflect more independent operations. Note the considerable difference between the standard deviations, s_r and s_R, an average-type parameter, and the repeatability and reproducibility limits, r and R, which are 2.8 times larger. If duplicates do not check within the r value, look for a problem—methodological, laboratory, or sample in origin. Note that these limits ($2*\sqrt{2} = 2.8$) are very close to the limits used for rejection in control charts $3*s_r$. Therefore they are most useful for large volume routine work rather than for validation of methods. Note the considerable difference between the standard deviations, s_r and s_R, an average-type parameter, and the repeatability and reproducibility limits, r and R, which are 2.8 times larger.

3.6 Confirmation of Analyte

Because of the existence of numerous chemical compounds, some of which have chemical properties very close to analytes of interest, particularly in chromatographic separations, but different biological, clinical, or toxicological properties, regulatory decisions require that the identity of the analyte of interest be confirmed by an independent procedure. This confirmation of chemical identity is in addition to a quantitative "check analysis," often performed independently by a second analyst to confirm that the quantity of analyte found in both analyses exceeds the action limit.

Confirmation provides unequivocal evidence that the chemical structure of the analyte of interest is the same as that identified in the regulation. The most specific method for this purpose is mass spectrometry following a chromatographic separation with a full mass scan or identification of three or four fragments that are characteristic of the analyte sought or the use of multiple mass spectrometric (MS^n) examination. Characteristic bands in the infrared can also serve for identification but this technique usually requires considerably more isolated analyte than is available from chromatographic separations unless special examination techniques are utilized. Visible and ultraviolet spectra are too subject to interferences to be useful, although characteristic peaks can suggest structural characteristics.

Other techniques that can be used for identification, particularly in combination, in approximate order of specificity, include:

(*1*) Co-chromatography, where the analyte, when mixed with a standard and then chromatographed by HPLC, GLC, or TLC, exhibits a single entity, a peak or spot with enhanced intensity.

(*2*) Characteristic fluorescence (absorption and emission) of the native compound or derivatives.

(*3*) Identical chromatographic and spectral properties after isolation from columns of different polarities or with different solvents.

Identical full-scan visible or ultra-violet spectra, with matching peak(s).

Furthermore, no additional peaks should appear when chromatographic conditions are changed, e.g., different solvents, columns, gradients, temperature, etc.

3.7 Stability of the Analyte

The product should be held under typical or exaggerated storage conditions and the active ingredient(s) assayed periodically for a period of time judged to reasonably exceed the shelf life of the product. In addition, the appearance of new analytes from deterioration should be explored, most easily by a fingerprinting technique, *Section 2.1*.

4 Report (as applicable)

4.1 Title

- Single-Laboratory Validation of the Determination of [Analyte] in [Matrix] by [Nature of Determination]
- Author, Affiliation
- Other Participants

4.2 Applicability (Scope)

- Analytes (common and chemical name; CAS registry number or Merck index number)
- Matrices used
- In presence of
- In absence of
- Safety statements applicable to product

4.3 Principle

- Preparation of test portion
- Extraction
- Purification

- Separation
- Measurement
- Alternatives
- Interferences

4.4 Reagents

(Reagents usually present in a laboratory need not be listed.)

- Reference standards, identity, source, purity
- Calibration standard solutions, preparation, storage, stability
- Solvents (special requirements)
- Buffers
- Others

4.5 Apparatus

(Equipment usually present in a laboratory need not be listed; provide source, Web address, and catalog numbers of special items.)

- Chromatographic equipment (operating conditions; system suitability conditions; expected retention times, separation times, peak or area relations)
- Temperature-controlled equipment
- Separation equipment (centrifuges, filters)
- Measurement instruments

4.6 Calibration

- Range, number and distribution of standards, replication, stability

4.7. Procedure

- List all steps of method, including any preparation of the test sample.
- Critical points
- Stopping points

4.8 Calculations

- Formulae, symbols, significant figures

4.9 Controls

4.10 Results of Validation

4.10.1 Identification Data

- Analytes measured and properties utilized (matrices tested; reference standard, source, identity, purity)

4.10.2 Performance Data

- Recovery of control material
- Repeatability (by replication of entire procedure on same test sample)
- Limit of determination]concentration where RSD_r = 20% or (blank + 10 * s_{blank})]
- Expanded measurement uncertainty $2*s_r$

4.10.3 Low-Level Data

Report instrument reading converted to a concentration through the calibration curve: positive, negative, or zero. Do not equate to 0, do not truncate data, or report "less than."

Interpretation: Concentrations less than 5 µg/kg may be reported as "zero" or "less than 5 µg/kg" with a 95% probability (5% chance of being incorrect).

4.10.4 Stability Data

ANNEX A
Abbreviations and Symbols Used

CAS	Chemical Abstracts Service (Registry Number)
CRM	Certified Reference Material
FDA	U.S. Food and Drug Administration
EPA	U.S. Environmental Protection Agency
GLC	Gas-liquid chromatography
HPLC	High-performance liquid chromatography
i	(as a subscript) Intermediate in precision terms
ISO	International Organization for Standardization
MU	Measurement Uncertainty
MS	Mass Spectrometry
MS^n	Multiple mass spectrometry
NMR	Nuclear magnetic resonance
r, R	Repeatability, reproducibility limits: The value less than or equal to the absolute difference between two test results obtained under repeatability (reproducibility) conditions is expected to be with a probability of 95% = $2*\sqrt{2}*s_r(s_R)$
RSD_r	Repeatability relative standard deviation = $s_r \times 100$
RSD_R	Reproducibility relative standard deviation = $s_R \times 100$
s_r	Repeatability standard deviation (within-laboratories)
s_R	Reproducibility standard deviation (among-laboratories)
$\bar{x}$	Mean, average

ANNEX B
Example of a Ruggedness Trial

Choose seven factors that may affect the outcome of the extraction and assign reasonable high and low values to them as follows:

Factor	High value	Low value
Weight of test portion	A = 1.00 g	a = 0.50 g
Extraction temperature	B = 30°	b = 20°
Volume of solvent	C = 100 mL	c = 50 mL
Solvent	D = Alcohol	d = Ethyl acetate
Extraction time	E = 60 min	e = 30 min
Stirring	F = Magnetically	f = Swirl 10 min intervals
Irradiation	G = Light	g = Dark

Conduct eight runs (a single analysis that reflects a specified set of factor levels) utilizing the specific combinations of high and low values for the factors as follows, and record the result obtained for each combination. (It is essential that the factors be combined exactly as specified or erroneous conclusions will be drawn.)

Run No.	Factor combinations	Measurement obtained
1	A B C D E F G	x1
2	A B c D e f g	x2
3	A b C d E f g	x3
4	A b c d e F G	x4
5	a B C d e F g	x5
6	a B c d E f G	x6
7	a b C D e f G	x7
8	a b c D E F g	x8

To obtain the effect of each of the factors, set up the differences of the measurements containing the subgroups of the capital letters and the small letters from column 2 thus:

Effect of A and a

$$[(x1 + x2 + x3 + x4)/4] - [(x5 + x6 + x7 + x8)/4] = J$$

$$4A/4 - 4a/4 = J$$

Note that the effect of each level of each chosen factor is the average of four values and that the effects of the 7 other factors

cancel out. (The Youden ruggedness trial or fractional factorial experiment was designed for this outcome.) Similarly,

Effect of B and b
$$[(x1 + x2 + x5 + x6)/4] – [(x3 + x4 + x7 + x8)/4] = K$$
$$4B/4 – 4b/4 = K$$

Effect of C and c
$$[(x1 + x3 + x5 + x7)/4] – [(x2 + x4 + x6 + x8)/4] = L$$
$$4C/4 – 4c/4 = L$$

Effect of D and d
$$[(x1 + x2 + x7 + x8)/4] – [(x3 + x4 + x5 + x6)/4] = M$$
$$4D/4 – 4d/4 = M$$

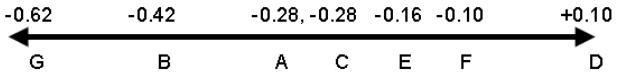
Effect of E and e
$$[(x1 + x3 + x6 + x8)/4] – [(x2 + x4 + x5 + x7)/4] = N$$
$$4E/4 – 4e/4 = N$$

Effect of F and f
$$[(x1 + x4 + x5 + x8)/4] – [(x2 + x3 + x6 + x7)/4] = O$$
$$4F/4 – 4f/4 = O$$

Effect of G and g
$$[(x1 + x4 + x6 + x7)/4] – [(x2 + x3 + x5 + x8)/4] = P$$
$$4G/4 – 4g/4 = P$$

Perform the eight determinations or runs carefully using the assigned factor level combinations and tabulate the values found. Then unscramble the 7 factors and obtain the effect of the assigned factor as the last number. It is important to use the combination of subscripts as assigned for proper interpretation.

Expt.	Found, %	Factors
x1	1.03	J (A) = 4A/4 – 4a/4 = 4.86 – 5.14 = –0.28
x2	1.32	K (B) = 4B/4 – 4b/4 = 4.79 – 5.21 = –0.42
x3	1.29	L (C) = 4C/4 – 4c/4 = 4.86 – 5.14 = –0.28
x4	1.22	M (D) = 4D/4 – 4d/4 = 5.05 – 4.95 = +0.10
x5	1.27	N (E) = 4E/4 – 4e/4 = 4.92 – 5.08 = –0.16
x6	1.17	O (F) = 4F/4 – 4f/4 = 4.95 – 5.05 = –0.10
x7	1.27	P (G) = 4G/4 – 4g/4 = 4.69 – 5.31 = –0.62
x8	1.43	



These values are plotted on a line. In this case they are more or less uniformly scattered along the line, but some attention should be paid to the extremes. Factor D, the highest positive value represents a difference in solvent, as expected, and this factor has to be investigated further to determine if the high values represents impurities or additional active ingredient. The extreme value of factor G suggests that the extraction should be conducted in the dark. As discussed by Youden, considerably more information can be obtained by utilizing several different materials and several independent replications in different laboratories, so as to obtain an estimate of the standard deviation to be expected between laboratories. Although the ruggedness trial is primarily a method development technique, validation of the application of a method to different matrices and related analytes can be explored simultaneously by this procedure.

Comments not used (may be added later):

3.3 Calibration: Run standards from low to high to compensate for any carryover. [Run in random order to compensate for drift is more important than allowing for carryover which should not occur.]

Independently made standards results in considerable random error in the calibration curve and is in fact the major source of random error in spectrophotometry. [Therefore a common stock solution is the preferred way of preparing the individual standards.]

Version 54 contains revisions as a result of comments from levanseler@nsf.org and McClure. Outline:

I. Types and benefits of each method validation study without reproducibility

II. Preparing for a Single-Laboratory Method Validation Study

III. Review of Performance Characteristics of a Method

IV. Errors

V. Calibration and Types

VI. Bias and Precision Estimations (no reference standard; no reproducibility)

VII. Detection and Quantification Limits

VIII. Ruggedness

PART II
AOAC Guidelines for Validation of Botanical Identification Methods

Contents

1 Scope

2 Applicability

3 Terms and Definitions

3.1 Botanical

3.2 Botanical Identification Method (BIM)

3.3 Candidate Method

3.4 Exclusivity

3.5 Exclusivity Sampling Frame (ESF)

3.6 Exclusivity Panel

3.7 Identity Specification (IS)

3.8 Inclusivity

3.9 Inclusivity Sampling Frame (ISF)

3.10 Inclusivity Panel

3.11 Laboratory Sample

3.12 Nontarget Botanical Material

3.13 Physical Form

3.14 Probability of Identification (POI)

3.15 Sample

3.16 Specified Inferior Test Material (SITM)

3.17 Specified Superior Test Material (SSTM)

3.18 Standard Method Performance Requirements (SMPRs)

3.19 Target Botanical Material

3.20 Test Portion

4 Validation Study Guidelines

4.1 SMPRs

4.2 SLV Study

4.3 Independent Validation Study

4.4 Collaborative Study

Annex A: Candidate Method (or Prevalidation Study)

Annex B: Understanding the POI Model

Annex C: Number of Test Portions

This document provides technical protocol guidelines for the AOAC validation of botanical identification methods and/or procedures, and covers terms and their definitions associated with the *Performance Tested Methods*SM and *Official Methods of Analysis*SM programs.

The guidelines working group consisted of James Harnly (Chair, USDA, ARS), Wendy Applequist (Missouri Botanical Garden), Paula Brown (British Columbia Institute of Technology), Steven Caspar (FDA/CFSAN), Peter Harrington (Ohio University), Danica Harbaugh-Reynaud (AuthenTechnologies, LLC), Norma Hill (Alcohol and Tobacco Tax and Trade Bureau Compliance Laboratory), Robert LaBudde (Least Cost Formulations and Old Dominion University), James Neal-Kababick (Flora Research Laboratories), Mark Roman (Tampa Bay Analytical Research), Shauna Roman (Schiff Nutrition International, Inc.), Darryl Sullivan (Covance Laboratories), Barry Titlow (Compound Solutions), and Paul Wehling (General Mills/Medallion Laboratories).

The guidelines were approved by the AOAC Official Methods Board on October 13, 2011.

This work was funded by the National Institutes of Health, Office of Dietary Supplements.

Reference: *J. AOAC Int*. **95**, 268–272(2012); DOI: 10.5740/jaoacint.11-447

1 Scope

The purpose of this document is to provide comprehensive technical guidance for conducting AOAC INTERNATIONAL (AOAC) validation studies for botanical identification methods submitted for AOAC *Official Methods of Analysis*SM (OMA) status and/or for *Performance Tested Methods*SM (PTM) status. The requirements for single-laboratory validation (SLV) studies, independent validation studies, and collaborative validation studies for those methods are described.

2 Applicability

These guidelines are intended to be applicable to the validation of all candidate botanical identification methods (*Annex A*) submitted to AOAC for (*1*) OMA status through either a collaborative study or an alternative pathway study or (*2*) PTM certification.

3 Terms and Definitions

3.1 Botanical

Of, or relating to, plants or botany. May also include algae and fungi. May refer to the whole plant, a part of the plant (e.g., bark, woods, leaves, stems, roots, rhizomes, flowers, fruits, seeds, etc.), or an extract of the parts.

3.2 Botanical Identification Method (BIM)

A method that establishes identity specifications for a botanical material and determines, within a specified statistical limit, a binary test result: YES, the test material is a true example of the target botanical material and meets the identity specifications, or NO, it is not the target botanical. Thus, a BIM answers the question, "Is the test material the same as the target material?" not "What is this material?" In most cases, the method will achieve this goal by comparison of the test material with materials from the inclusivity panel and will return a YES/NO (or, in some cases, a consistent/nonconsistent) answer.

3.3 Candidate Method

The method to be validated or submitted for validation (*Annex A*).

3.4 Exclusivity

Ability of a BIM to correctly reject nontarget botanical materials.

3.5 Exclusivity Sampling Frame (ESF)

A list of practically obtainable nontarget botanical materials that have taxonomic, physical, or chemical composition characteristics similar to the target botanical and must give a negative result when tested by the BIM.

3.6 Exclusivity Panel

A subset of the ESF that is selected for the validation study. The identity of these materials should be verified by an appropriate method or process.

3.7 Identity Specification (IS)

The morphological, genetic, chemical, or other characteristics that define a target botanical material. Specifications may include, but are not limited to, data from macroscopic, microscopic, genetic (e.g., DNA sequencing), chromatographic fingerprinting (e.g., capillary electrophoresis, gas chromatography, liquid chromatography, or thin-layer chromatography), and spectral fingerprinting (e.g., infrared, near-infrared, nuclear magnetic resonance, ultraviolet/visible absorbance, or mass spectrometry) methods.

3.8 Inclusivity

Ability of a BIM to correctly identify variants of the target material that meet the identity specification.

3.9 Inclusivity Sampling Frame (ISF)

A list of practically obtainable botanical materials that are expected to give a positive result when tested by the BIM. The inclusivity frame should be sufficiently large that the botanical variation is adequately represented. Sources of variation may include, but are not limited to, species, subspecies, cultivar, growing location, growing conditions, growing season, and post-harvest processing.

3.10 Inclusivity Panel

A subset of the ISF that is selected for the validation study. These materials should be authenticated by an appropriate method.

3.11 Laboratory Sample

Sample as prepared for sending to the laboratory intended for inspection or testing.

3.12 Nontarget Botanical Material

Any botanical material that does not meet the identity specification.

3.13 Physical Form

Botanical materials exist in a number of physical forms. The form(s) will be specified by the standard method performance requirements (SMPRs).

3.14 Probability of Identification (POI)

The expected or observed fraction of test portions at a given concentration that give a positive result when tested by the BIM. A general description is provided in *Annex B*.

3.15 Sample

A small portion or quantity, taken from a population or lot that is ideally a representative selection of the whole. Sample homogeneity is usually determined with multiple samples.

3.16 Specified Inferior Test Material (SITM)

A botanical material mixture that has the maximum concentration of target material that is considered unacceptable, as specified by the SMPRs. The BIM must reject this material with a specified minimum level of (1 – POI) with 95% confidence. The ideal BIM would reject the SITM 100% of the time (i.e., accept 0% of the time). The SITM will typically be high-quality target material mixed with the worst-case (for identification) nontarget material.

3.17 Specified Superior Test Material (SSTM)

A botanical material mixture that has the minimum acceptable concentration of the target material, as specified by the SMPR. The BIM must identify this material with a specified minimum level of POI with 95% confidence. The ideal BIM would accept the SSTM 100% of the time. The SSTM will typically be high-quality target material mixed with a small amount of worst-case (for identification) nontarget material.

3.18 Standard Method Performance Requirements (SMPRs)

Performance requirements based on the fitness-for-purpose statement for each method. For BIMs, the SMPRs should include the physical form of the sample, the ISF, the ESF, the SSTM, the SITM, the number of samples for the inclusivity/exclusivity panels, and the desired probability and confidence limits for the method.

3.19 Target Botanical Material

The botanical material of interest as described in the identity specification.

3.20 Test Portion

The portion of the laboratory sample that is subjected to analysis by the method.

4 Validation Study Guidelines

A validated BIM requires a method validation study that demonstrates its acceptability according to the SMPRs. The guidelines presented here are intended to be applied to any qualitative BIM that returns a binary, YES/NO test result (*Annex A*). The guidelines provide technical guidance in validating the method based on the POI model (*Annex B*).

4.1 SMPRs

The SMPRs will be prepared by the appropriate AOAC body as per AOAC policy. The SMPRs will specify (*1*) the target botanical material, (*2*) the physical form of the material, (*3*) a list of botanical materials for the ISF/ESF, (*4*) composition of the SSTM and SITM, (*5*) maximum POI for the SITM and minimum POI for the SSTM, and (*6*) the desired probability and confidence limits for the inclusivity/exclusivity and SSTM/SITM measurements.

The SMPRs will consider the nature of the material being tested and determine the necessary breadth and depth of the inclusivity and exclusivity panels. In some cases, a few, very similar exclusivity panel materials may require in-depth testing (more test portions of a smaller group of materials). Conversely, the nature of the material may require greater breadth (fewer test portions of a greater number of materials).

The number of test portions needed should be determined on sound statistical grounds (*Annex C*) and subject matter expertise.

4.2 SLV Study

4.2.1 Scope

An SLV study is intended to determine the performance of a candidate method (*Annex A*). For validation purposes, the candidate BIM may be regarded as a black box providing a binary, YES/NO test result. The study is designed to evaluate performance parameters for the candidate method including (*1*) inclusivity/exclusivity, (*2*) POI for the SSTM and the SITM, and (*3*) POI as a function of the concentration of the target material (analytical response curve). This last parameter may be optional as specified by the SMPRs.

4.2.2 Inclusivity/Exclusivity Study

The purpose of this study is to confirm the ability of the candidate method to provide positive results (YES answers) for botanical materials on the inclusivity panel and negative results (NO answers) for materials on the exclusivity panel.

4.2.2.1 Inclusivity/Exclusivity Panel Selection

Botanical materials selected from the ISF/ESF will comprise the inclusivity/exclusivity panels. If the ISF/ESF specified by the SMPRs are sufficiently large, a representative subgroup will be selected for the panels by the method validator. Primary requirements for the panel materials are their availability and identity verification by an appropriate method or process. All test portions should be as uniform and homogeneous as possible. The level of replication of the inclusivity/exclusivity panels will be specified in the SMPRs.

4.2.2.2 Study Design

Prepare the test samples in a form appropriate for the candidate method. All test samples will be blinded and randomized so that the analyst(s) cannot know the identity of the samples. Analyze the test samples following the instructions of the candidate method.

4.2.2.3 Data Analysis and Reporting

The data will be analyzed for positive and negative responses. Unexpected results will be investigated, evaluated, and resolved prior to continuing the validation. The data is reported for individual inclusivity/exclusivity material as the number correctly identified. For example, "Of the 30 specific botanical materials of the inclusivity panel that were tested, 28 were identified correctly (gave a positive result) and two were not identified correctly (gave a negative result). Those materials not identified correctly were the following: …" or "Of the 30 specific botanical materials of the exclusivity panel that were tested, 27 were identified correctly (gave a negative result) and three were not identified correctly (gave a positive result). Those not identified correctly were the following: …" The study report should include a table titled "Inclusivity/Exclusivity Panel Results," which lists all materials tested, their source, origin, and essential characteristics and testing outcome. The implications of each unexpected result should be discussed and evaluated.

4.2.3 SSTM/SITM Study

The purpose of this study is to demonstrate method performance at two concentrations, the SSTM and the SITM.

4.2.3.1 Test Samples

The appropriate amount of a target material is selected from the inclusivity panel and is mixed with an appropriate amount of a nontarget material from the exclusivity panel to produce the SSTM and SITM as specified by the SMPRs. The test materials may be prepared using individual botanical materials from the inclusivity/exclusivity panels or composites of materials from the two panels as specified by the SMPRs.

All test portions should be as uniform and homogeneous as possible. The level of replication of the SSTM and SITM will be specified in the SMPR.

4.2.3.2 Study Design

Prepare the test samples in a form appropriate for the candidate method. All test samples will be blinded and randomized so that the analyst(s) cannot know the identity of the samples. Analyze the test samples following the instructions of the candidate method.

4.2.3.3 Data Analysis and Reporting

The data will be analyzed for positive and negative responses. For the SSTM and the SITM, report the POI results with 95% confidence intervals and the total number tested and the total number correctly identified. Comparison to SMPRs should be made and discussed.

4.2.4 Analytical Response Curve

This study will characterize the POI curve for mixtures of SSTM and SITM.

4.2.4.1 Test Samples

The appropriate amount of a target material is selected from the inclusivity panel and is mixed with an appropriate amount of a nontarget material from the exclusivity panel to produce mixtures with concentrations intermediate between the SSTM and SITM. The test materials shall be prepared using the same target and nontarget botanical material samples used in the SSTM and SITM study. The test materials may also be prepared by mixing appropriate ratios of the SSTM and SITM.

4.2.4.2 Study Design

Prepare the test samples in a form appropriate for the candidate method. All test samples will be blinded and randomized so that the analyst(s) cannot know the identity of the samples. Analyze the test samples following the instructions of the candidate method.

4.2.4.3 Data Analysis and Reporting

The data will be analyzed for positive and negative responses. For each mixture, report the POI results with 95% confidence intervals, the total number of samples tested, and the total number of positive responses. Plot the POI curve and confidence intervals.

4.3 Independent Validation Study

This study is identical to the SLV Study in Section 4.2.

4.4 Collaborative Study

The collaborative study is a route to an *Official Method*℠. The purpose of the collaborative study is to estimate the reproducibility and determine the performance of the candidate method among collaborators.

4.4.1 Number of Collaborators

A minimum of 10 independent laboratories reporting valid data is required. The study director should plan on including additional laboratories in the case of invalid data sets.

4.4.2 Number of Tests

Each collaborator receives 12 replicates of each material to be studied. At a minimum these materials will include the SSTM and SITM. Prepare the test samples in a form appropriate for the candidate method. All test samples will be blinded and randomized so that the analyst(s) cannot know the identity of the samples. Analyze the test samples following the instructions of the candidate method.

4.4.3 Data Analysis and Reporting

The data will be analyzed by the laboratory for positive and negative responses. For the SSTM and the SITM, report the POI results with confidence intervals for each laboratory, and for the

combined results. Estimate reproducibility as in *Annex C* and evaluate compared to the SMPRs.

ANNEX A
Candidate Method (or Prevalidation Study)

1 Scope

The candidate method must measure appropriate characteristics that are suitable to the question being asked and that will meet predetermined SMPRs. The method may be based on new principles or modifications of an existing method. The identity specifications will be based on morphological, genetic, and/or chemical characteristics, or any other defining feature of the botanical material. The candidate method may use visual inspection, DNA sequencing, instrumental analysis, or any other appropriate measurement. The measured characteristics will collectively provide a single analytical parameter that will be used to determine the final YES or NO result. The analytical parameter may be based on the degree of similarity or the degree of difference of the test sample and the reference material.

2 Inclusivity/Exclusivity Panel Selection

The method developer will select representative botanical materials from the ISF and ESF for use as target and nontarget botanical materials, respectively, in development of the method. These materials must be authenticated by an appropriate method.

3 Analytical Parameter

The method developer will prepare all the botanical samples in a form appropriate for the candidate method. The developer will analyze the target and nontarget botanical materials using the candidate method and develop an analytical parameter that is suitable for distinguishing between the two sets of materials.

4 Probability of Identification (POI)

Target materials will be mixed with systematically increasing amounts of nontarget materials to produce a series of target materials whose concentrations range from 100% to a concentration below the minimum acceptable concentration specified by the SMPRs. The developer will analyze the target and diluted target materials using the candidate method and determine the analytical parameter for each concentration.

5 Specific Superior/Inferior Test Materials

Based on the analytical parameters measured for the diluted target materials, a threshold value will be established that will permit positive identification of the minimum acceptable concentration of the target material with the specified confidence (e.g. 95%). The developer will use the threshold to determine a POI for each concentration (*Annex B*). The POIs measured for each concentration will be used to construct the POI curve.

6 Data Analysis and Reporting

The method developer will document the candidate method and the POI results.

ANNEX B
Understanding the POI Model

[*See Official Methods of Analysis* (2012) *Appendix K*, Part III, "Probability of Identification: A Statistical Model for the Validation of Qualitative Botanical Identification Methods," by Robert LaBudde and James M. Harnly, *J. AOAC Int.* **95**, 273–285 (2012). http://dx.doi.org/10.5740/jaoacint.11-266]

ANNEX C
Number of Test Portions

See Table C1.

Notes: (*1*) Enter the first column with the maximum error fraction tolerated by the SMPR, e.g., 10%.

(*2*) Select the sample size required by the number of misclassifications to be allowed, e.g., one erroneous result gives a sample size of $n = 48$ for a maximum error probability of 10%.

(*3*) Allowing more erroneous results increases the sample size required.

(*4*) The last (AOQL) column indicates the maximum error probability of a method which passes the SMPR for the test. For the example sampling plan indicated, this is 5.4%, approximately ½ of the maximum error probability in the SMPR. Typically the AOQL must be only 50–60% of the SMPR value to reliably pass the validation test. Method developers should take this into account.

Sample Size Required for Proportion

ASSUME: 1. Binary outcome (occur / not occur).
2. Constant probability rho of event occurring.
3. Independent trials (e.g., simple random sample).
4. Fixed number of trials N.

INFERENCE: 95% confidence interval lies entirely at or BELOW specified maximum rho.
DESIRED: Sample size N needed.
NOTES: 1. Based on modified Wilson score 1-sided confidence interval.
2. AOQL = Average Outgoing Quality Level

Maximum Probability rho	Sample Size N	**Maximum** Number Events x	**Minimum** Number Non-events y	1-sided **Upper** Confidence Limit on rho	Expected Lower Confidence Limit on rho	Expected Upper Confidence Limit on rho	Effective AOQL rho
50%	3	0	3	47.4%	0.0%	56.1%	28.1%
50%	10	2	8	45.9%	5.7%	51.0%	28.3%
50%	20	6	14	48.4%	14.5%	51.9%	33.2%
50%	40	14	26	48.0%	22.1%	50.5%	36.3%
50%	80	32	48	49.2%	30.0%	51.0%	40.5%
45%	2	0	2	57.5%	0.0%	65.8%	32.9%
45%	10	1	9	34.8%	0.0%	40.4%	20.2%
45%	20	5	15	43.2%	11.2%	46.9%	29.0%
45%	40	12	28	42.9%	18.1%	45.4%	31.8%
45%	80	28	52	44.1%	25.5%	45.9%	35.7%
40%	5	0	5	35.1%	0.0%	43.4%	21.7%
40%	10	1	9	34.8%	0.0%	40.4%	20.2%
40%	20	4	16	37.8%	8.1%	41.6%	24.8%
40%	40	10	30	37.6%	14.2%	40.2%	27.2%
40%	80	24	56	39.0%	21.1%	40.8%	30.9%
35%	6	0	6	31.1%	0.0%	39.0%	19.5%
35%	10	1	9	34.8%	0.0%	40.4%	20.2%
35%	20	3	17	32.2%	5.2%	36.0%	20.6%
35%	40	9	31	34.9%	12.3%	37.5%	24.9%
35%	80	21	59	35.0%	17.9%	36.8%	27.3%
30%	7	0	7	27.9%	0.0%	35.4%	17.7%
30%	10	0	10	21.3%	0.0%	27.8%	13.9%
30%	20	2	18	26.2%	2.8%	30.1%	16.4%
30%	40	7	33	29.3%	8.7%	31.9%	20.3%
30%	80	17	63	29.6%	13.7%	31.4%	22.6%
25%	9	0	9	23.1%	0.0%	29.9%	15.0%
25%	10	0	10	21.3%	0.0%	27.8%	13.9%
25%	20	1	19	19.6%	0.0%	23.6%	11.8%
25%	40	5	35	23.5%	5.5%	26.1%	15.8%
25%	80	13	67	24.1%	9.7%	25.8%	17.8%
20%	11	0	11	19.7%	0.0%	25.9%	12.9%
20%	20	1	19	19.6%	0.0%	23.6%	11.8%
20%	24	1	23	16.7%	0.0%	20.2%	10.1%
20%	36	3	33	19.1%	2.9%	21.8%	12.4%
20%	40	3	37	17.3%	2.6%	19.9%	11.2%
20%	48	5	43	19.9%	4.5%	22.2%	13.3%
20%	60	6	54	18.2%	4.7%	20.1%	12.4%
20%	72	8	64	18.7%	5.7%	20.4%	13.1%
20%	80	10	70	19.8%	6.9%	21.5%	14.2%
15%	20	0	20	11.9%	0.0%	16.1%	8.1%
15%	24	0	24	10.1%	0.0%	13.8%	6.9%
15%	36	1	35	11.5%	0.0%	14.2%	7.1%
15%	40	2	38	14.0%	1.4%	16.5%	8.9%
15%	48	3	45	14.6%	2.1%	16.8%	9.5%
15%	60	4	56	14.0%	2.6%	15.9%	9.3%
15%	72	5	67	13.6%	3.0%	15.2%	9.1%
15%	80	6	74	13.9%	3.5%	15.4%	9.4%
10%	40	0	40	6.3%	0.0%	8.8%	4.4%
10%	48	1	47	8.8%	0.0%	10.9%	5.4%
10%	60	2	58	9.6%	0.9%	11.4%	6.1%
10%	72	3	69	10.0%	1.4%	11.5%	6.5%
10%	80	3	77	9.0%	1.3%	10.5%	5.9%
5%	60	0	60	4.3%	0.0%	6.0%	3.0%
5%	72	0	72	3.6%	0.0%	5.1%	2.5%
5%	80	0	80	3.3%	0.0%	4.6%	2.3%
5%	90	1	89	4.8%	0.0%	6.0%	3.0%

Table C1

PART III
Probability of Identification: A Statistical Model for the Validation of Qualitative Botanical Identification Methods

A botanical is an herbal material that is frequently used as an ingredient in a dietary supplement regulated in the United States under the Federal Food, Drug, and Cosmetic Act of 1938, as amended by the Dietary Supplement Health and Education Act of 1994 (1). More recently, current Good Manufacturing Practices for foods and dietary supplements (2) issued by the U.S. Food and Drug Administration has tasked manufacturers with establishing specifications and developing a QA program for all botanical ingredients. As a consequence, both processors of botanicals and regulators are interested in the verification of the identity of botanical materials. Thus, the development of reliable methods for the identification of botanical materials and minimum acceptable levels of contamination are critical.

A botanical identification method (BIM) is any qualitative method that reliably identifies a botanical material and returns a binary result of either 1 = "identified" or 0 = "not identified." The actual method used can be presumed unknown and a "black box" with respect to the protocols involved in the validation studies. The BIM must be validated in terms of inclusivity, exclusivity, probability of identification, robustness, reproducibility, repeatability, and other criteria.

The heart of the BIM is the probability of identification (POI) model. The POI model has been developed as a means of characterizing and validating the performance of a qualitative method based on simple statistics and associated confidence intervals (3, 4). Figure 1 (modified from ref. 3) shows a plot where the concentration of the target material increases towards the right while the concentration of a nontarget material increases to the left. The parameter of interest is the POI (the vertical axis), which is defined as the probability, at a given percentage of target material, of getting a positive response by the detection method. The positive response of the BIM indicates that the test material matches the target botanical material. While the plot in Figure 1 is symmetrical, POI plots are usually asymmetrical. The POI model is based on the probability of detection model which was developed for binary qualitative methods (3, 4).

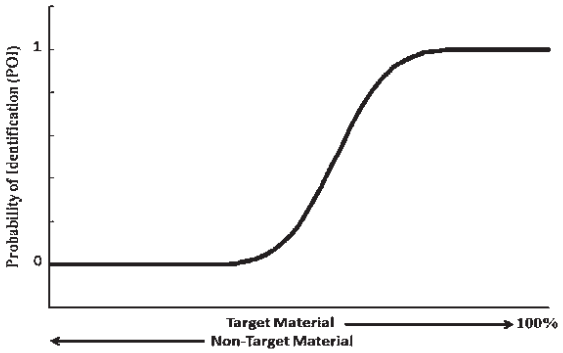


Figure 1. Probability of identification for botanical identification.

The POI, as illustrated in Figure 1, is dependent on the concentration of the target botanical material. The probability of a positive response increases as the concentration of the target botanical increases and decreases as the concentration of the nontarget material increases. The goal of method development and validation is primarily to determine if the method meets method performance requirements (MPRs), and secondarily to characterize how the method makes the transition from a negative to a positive response.

The MPRs, as established by the developer, will specify the target botanical materials (inclusivity sampling frame; ISF), the nontarget materials (exclusivity sampling frame; ESF), the physical form of the materials, the minimum concentration of target material that is acceptable in the presence of nontarget material, and the maximum concentration target material that is unacceptable. These latter materials are the specific superior and specific inferior test materials (SSTM and SITM, respectively). The idealized goal of the BIM is to discriminate (with a specified degree of confidence, e.g., 95%) between the SSTM (for which the POI is high) and the SITM (for which the POI is low). Additionally, samples of the SSTM and SITM may be mixed to obtain the intermediate test concentrations that are used to characterize the POI curve in its transitional range.

In some studies, full characterization of the transition of the POI curve may be of lesser importance and the intermediate concentrations omitted. In this care the only concentrations used are those for which the performance requirements are applied, typically the SITM and SSTM (0% and 100% SSTM, respectively). Two factors are important to method development: industrial-regulatory requirements, and the technological limit (state of the measurement art). If the technological limit exceeds the industry-regulatory requirement, then the industrial-regulatory requirement can be set at a value reasonably attainable by existing technology. In this case, the cost of the analysis may be the major factor governing validation study design. If the technological limit cannot meet the industrial-regulatory requirement, then improved technology must be developed before a BIM fit for the purpose intended can be found.

Glossary

Analytical parameter (AP).—A measured or computed analytical value used to determine whether the test material matches the target material. The analytical parameter may be based on morphological

---

A qualitative botanical identification method (BIM) is an analytical procedure that returns a binary result (1 = identified, 0 = not identified). A BIM may be used by a buyer, manufacturer, or regulator to determine whether a botanical material being tested is the same as the target (desired) material, or whether it contains excessive nontarget (undesirable) material. The report describes the development and validation of studies for a BIM based on the proportion of replicates identified, or probability of identification (POI), as the basic observed statistic. The statistical procedures proposed for data analysis follow closely those of the probability of detection (POD), and harmonize the statistical concepts and parameters between quantitative and qualitative method validation. Use of POI statistics also harmonizes statistical concepts for botanical, microbiological, toxin, and other analyte identification methods that produce binary results. The POI statistical model provides a tool for graphical representation of response curves for qualitative methods, reporting of descriptive statistics, and application of performance requirements. Single collaborator and multicollaborative study examples are given.

Reference: LaBudde, R.A., & Harnly, J.M. (2012) *J. AOAC Int.* **95**, 273–285. http://dx.doi.org/10.5740/jaoacint.11-266

The POI statistical model was approved by the AOAC Official Methods Board on October 13, 2011.

features, genetic sequences, chromatographic patterns, spectral patterns, or any other metric appropriate for the target material.

Botanical.—Of or relating to plants or botany. May also include algae and fungi. May refer to the whole plant, a part of the plant (e.g., bark, woods, leaves, stems, roots, rhizomes, flowers, fruits, seeds, extracts, etc.), or an extract of the plant.

BIM.—A method that establishes identity specifications for a botanical material and determines, within a specified statistical limit, a binary result: yes, the test material is a true example of the target botanical material and meets the identity specifications; or no, it is not the target botanical. Thus, a BIM answers the question, "*Is the test material the same as the target material?*" not "*What is this material?*" In most cases, the method will achieve this goal by comparison of the test material with materials from the inclusivity panel and will return a yes/no (or, in some cases, a consistent/nonconsistent) answer.

Candidate method.—The method to be validated.

Exclusivity.—Ability of a BIM to correctly reject nontarget botanical materials.

ESF.—A list of practically obtainable nontarget botanical materials that have similar taxonomic, physical, or chemical composition characteristics that are expected to give a negative result when tested by the BIM.

Exclusivity panel.—A subset of the ESF that is selected for the validation study. These materials should be authenticated by an appropriate method.

False-negative fraction (FNF).—1–POI for 100% SSTM. Not defined for other concentrations.

False-positive fraction (FPF).—POI for 100% SITM. Not defined for other concentrations.

Identity specification.—The morphological, genetic, chemical, or other characteristics that define a target botanical material. Specifications may include, but are not limited to, data from macroscopic, microscopic, genetic (e.g., DNA sequencing, barcoding), chromatographic fingerprinting (e.g., CE, GC, LC, TLC), and spectral fingerprinting (e.g., IR, NIR, NMR, MS, UV-Vis) methods.

Inclusivity.—Ability of a BIM to correctly identify variants of the target material that meet the identity specification.

ISF.—A list of practically obtainable botanical materials that are expected to give a positive result when tested by the BIM. The inclusivity sampling frame should be sufficiently large that the botanical variation is adequately represented. Sources of variation may include, but are not limited to, species, subspecies, cultivar, growing location, growing conditions, growing season, and post-harvest processing.

Inclusivity panel.—A subset of the ISF that is selected for the validation study. These materials should be authenticated by an appropriate method.

Laboratory sample.—Sample as prepared for sending to the laboratory intended for inspection or testing.

MPRs.—Performance requirements based on the fitness-for-purpose statement for each method. For BIMs, the MPRs should minimally include the physical form of the sample, the ISF, the ESF, the SSTM, and the SITM.

Nontarget botanical material.—Any botanical material that does not meet the identity specification.

Physical form.—Botanical materials exist in a number of physical forms. The form(s) to be analyzed by the method will be specified by the MPRs.

POI.—The expected or the observed fraction of test portions that provide a positive result at a given concentration when tested by the BIM.

Sample.—A small quantity, taken from a population or lot that is a representative selection of the whole.

SITM.—A mixture of botanical materials that contains the maximum concentration of target material that is considered unacceptable, as specified by the MPRs. The BIM must reject this material with a specified minimum level of (1–POI) with 95% confidence. The ideal BIM would reject the SITM 100% of the time (i.e., identify 0% of the time). The SITM will typically be high-quality target material mixed with worst-case (for identification) nontarget material.

SSTM.—A mixture of botanical material that contains the minimum acceptable concentration of the target material, as specified by the MPR. The BIM must identify this material with a specified minimum level of POI with 95% confidence. The ideal BIM would identify the SSTM 100% of the time. The SSTM will typically be high-quality target material mixed with a small amount of worst-case (for identification) nontarget material.

Target botanical material.—The botanical material of interest as described in the identity specification.

Target material concentration.—The percentage, by weight, of the target botanical material in the sample.

Test portion.—The portion of the laboratory sample that is subjected to analysis by the method.

Inclusivity Panel

When a botanical material is identified for development of a BIM, a target material is usually specified. Biological materials, however, are complex. While the genotype of a species or subspecies may be relatively stable, the phenotype (metabolite composition) will vary with location, season, weather, and many other variables. Thus, "target material" becomes "target materials." Ideally, the target materials will encompass the expected botanical variation.

An inclusive list of all the variations for a target material can be quite extensive and impractical. For example, the list for a specific botanical might ideally include samples from the last 10 years from eight international locations (80 samples). In reality, only 25 of the desired samples may be practically obtainable. These 25 obtainable samples comprise the ISF. Of these 25 samples, only 10 may be selected for method development/validation. These 10 samples comprise the inclusivity panel.

For each candidate BIM, the MPRs must provide a list of all necessary botanical variants that should provide a positive identification. This should include species, varieties, geographic or seasonal variants, and other variants that are believed to possibly associate with BIM identification performance. The information tabulated should include variety, season, locality, source from which the variant is obtainable, species, variety or subclass, and whether or not it is essential that the variant be tested. The age of the plant may also be a factor of importance. The subset of this list, which is practically obtainable for a validation study, is the ISF.

The MPRs should identify the minimum number of materials in the ISF that must be tested to verify identifiability (inclusivity panel), as well as the number of replicates needed. If at all possible, any exchangeability (choice among variants which MPRs do not discriminate) should result in random selection from the ISF.

Generally, the inclusivity panel of target variants should include all of the ISF if the number of variants is small. Otherwise, all

necessary variants plus additional ones randomly selected should comprise the inclusivity panel. More randomized replicate variants may allow a quantitative statistical inference to be made concerning inclusivity. An inclusivity panel with no randomization, only subjective selection, does not permit statistical statements of inference with respect to inclusivity.

Exclusivity Panel

The list of nontarget materials can be quite extensive, theoretically including all the botanicals not on the inclusivity list. However, of prime interest are those materials that might accidentally or intentionally be used to replace or augment the target materials. The exclusivity list should include botanical materials that are closely related taxonomically, morphologically, or phenotypically. Again, this list may be extensive and impractical. The ESF will comprise those botanical materials that are practically obtainable. The exclusivity panel will comprise those samples used for method development and validation.

The MPRs must provide a list of all necessary or commonly encountered nontarget botanical materials and variants. This list should include botanical materials that are believed to accidentally or intentionally alter the composition of the target material. The information tabulated should include variety, season, locality, source from which the variant is obtainable, species, variety or subclass, and whether or not it is essential that the nontarget material be tested. The subset of this list, which is practically obtainable for a validation study, should then be identified as the ESF.

The MPRs should identify the minimum number of nontarget materials of the ESF that should be included on the exclusivity panel and be tested to verify non-identifiability, as well as the number of replicates needed. If at all possible, any exchangeability (choice among variants which expertise does not discriminate) should result in random selection from the ESF.

Generally, the exclusivity panel of authentic variants should include all of the ESF if the number of variants is small. Otherwise, all necessary variants, plus optional ones randomly selected, should comprise a set as specified by the ERP. More replicates and randomization may allow a quantitative statistical inference to be made concerning exclusivity.

Inclusivity and Exclusivity Testing

The purpose of inclusivity/exclusivity testing is to verify that the BIM correctly identifies all of the botanical materials listed in the ISF and correctly rejects all nontarget materials listed in the ESF. The BIM should clearly and unequivocally discriminate between the target and nontarget materials. Testing materials from the inclusivity/exclusivity panels should provide sufficient confidence that this is the case. The number of samples tested and the number of replicates is specified by the MPRs.

Typically, inclusivity/exclusivity panel results are verified during method development. Any unexpected results should be followed up with a minimum number of additional replications (determined by the MPRs) to characterize the POI on the variant quantitatively. If the variant fails to meet minimum acceptable performance requirements as set by the MPRs, the exception should be noted in the study report and reviewed for acceptability by the relevant method reviewers.

If the method development results are acceptable, inclusivity and exclusivity should be verified in an independent laboratory, although possibly on a less-intensive (fewer replicates or randomly selected variants) basis, as the objective is verification, not validation. If no randomization is used, all that can be reported are the actual results obtained, but without suggestive quantitative statistics. For example, without randomization, the use of percentages or other quantitative measures is inappropriate.

Performance Requirements and the Specification and Preparation of the SITM and SSTM

After inclusivity and exclusivity studies have been completed, target and nontarget material(s) are chosen to verify that the method can discriminate between the SSTM and the SITM. Either the worst-case nontarget materials, or perhaps the most common nontarget materials, would typically be chosen. In addition, a combination of target and nontarget materials should be selected to challenge method performance (worst-case, most common, etc.). The number of samples tested and the number of replicates is specified by the MPRs.

The MPRs should identify the composition and the minimum POI acceptable (with 95% confidence) for the SSTM and SITM. The SSTM and SITM would be made of the target material(s) mixed with the combination of nontarget material(s).

Application of the POI to an Analytical Method

Analytically, a BIM will be based on a series of measured values. These values may be derived from morphological features, genetic sequences, chromatographic patterns, spectral patterns, or any other metric appropriate for the target material. These values will be combined to provide a single AP that will be used to determine whether the test sample does or does not match the materials from the inclusivity panel. This decision is made by comparing the AP of the test material to a threshold value that provides the level of identification specified by the MPRs.

The first step in the development of the method is the selection of the analytical approach and the analysis of samples from the ISF and ESF. Multiple replicates of multiple samples should, ideally, give results similar to those in Figure 2. Here, the AP, not the POI, is plotted on the vertical axis. The standard deviations (SDs) are shown as sample distribution functions, rather than as error bars. Ideally, the separation of the ISF and ESF samples should be as large as possible. For the data in Figure 2, the threshold to distinguish between the ISF and ESF can be placed at almost any value of the AP.

The width of the sample distribution function will depend on the number of samples analyzed from the ISF and ESF. If replicates

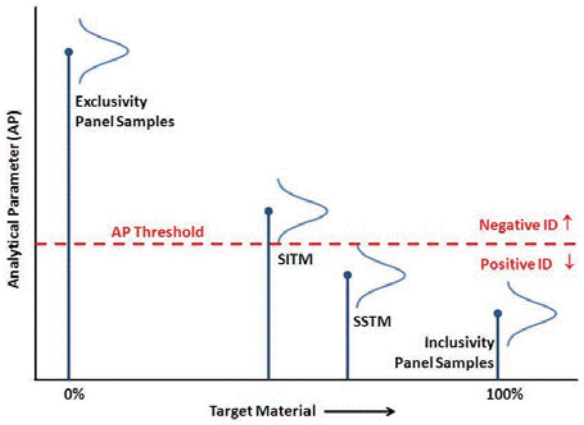


Figure 2. Inclusivity/exclusivity and SSTM/SITM characterization.

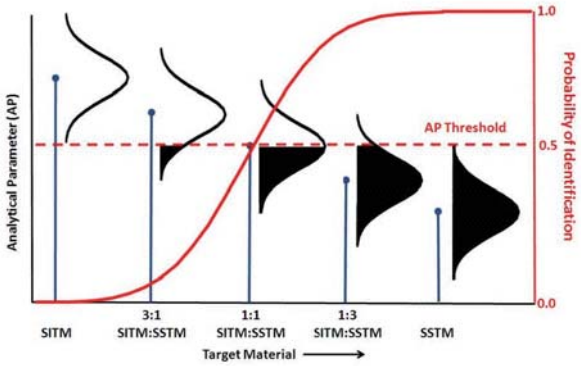


Figure 3. Conversion of SSTM, SITM, and intermediate concentrations to POI.

of a single sample are analyzed, then the width of the distribution will be narrow (a smaller SD), and only reflect the instrumental variance. As more samples are analyzed from the ISF and ESF, the distribution functions will broaden, reflecting the increasing biological variance.

The next step is to determine whether the method can distinguish between the SSTM and the SITM. The concentrations of the SSTM and the SITM are specified by the MPRs. Figure 2 illustrates an arbitrary specification. It can be seen that the distributions of the SSTM and SITM are completely resolved and the threshold must be located exactly between the two distributions to provide 100% identification of the SSTM (POI = 1) and 100% rejection of the SITM (POI = 0). If the concentration of target material in the SSTM was lower, or the concentration in the SITM higher, the distribution functions would overlap and 100% identification or rejection would not be possible. In this case, the confidence limit would have to be lowered or another method selected.

Finally, the shape of the POI curve can be determined. As shown in Figure 3, concentrations of the target materials that fall between the SSTM and SITM must be prepared. In each case, the threshold will intersect each peak and determine the POI. As the SSTM:SITM values change from 1:0 to 3:1 to 1:1 to 1:3 to 0:1, the POI decreases from 1.0 to 0.9 to 0.5 to 0.1 to 0.0.

The models in Figures 2 and 3 assume that the SITM and SSTM have the same, symmetrical distribution function and width. This is not a reasonable assumption for real samples. However, the POI model is valid regardless of the shape of the distribution functions involved.

A Specific Example: American Ginseng Mixed with Asian Ginseng

The data set presented here illustrates the analytical measurements discussed in the previous section. The target botanical material is American ginseng (AG) and the nontarget material is Asian ginseng (CG). The inclusivity panel consists of 43 AG samples grown in the United States (harvested over 3 years from 20 different farms in Wisconsin), and the exclusivity panel consists of eight CG samples grown in China (Table 1).

The AG and CG samples were analyzed by direct injection MS, and yielded spectra with approximately 1000 ions. The SSTM and SITM were generated synthetically by combining different percentages of the AG and CG mass spectra. For example, the spectra for 98% AG mixed with 2% CG was computed as 0.98 of an AG spectra added to 0.02 of a CG spectra. In all, 344 SSTM spectra were generated (43 AG × 8 CG).

The multivariate data set (395 samples × 1000 variables) was analyzed using soft independent modeling of class analogy (SIMCA; *Annex A*). SIMCA fit a principal component model to the data for the inclusivity panel (100% AG) and produced a goodness-of-fit value, the Q residual, for every sample analyzed. The Q residual was used to compare the test (100% CG, SSTM, and SITM) and the target (100% AG) materials. In every case, the SIMCA model was based on 100% AG and a single principal component. The Q residual describes how far a sample falls outside the model (*Annex A*).

Figure 4 (A) shows the inclusivity/exclusivity study. The Q residual is plotted for individual samples. With 100% AG

Table 1. Panax samples analyzed in this study

No.	Label	Provider	Source
	Inclusivity panel (American ginseng)		
26	American ginseng		USA
13	American ginseng		USA
4	American ginseng		USA
	Exclusivity panel (Chinese ginseng)		
3	Asian ginseng, red	American Herbal Pharmacopoeia 2	China
1	Kirin Red No. 1	Internet retailer	China
1	Kirin Red No. 3	Internet retailer	China
1	Kirin Red No. 5	Internet retailer	China
1	Shih Chu No. 25	Internet retailer	China
1	Shih Chu No. 80	Internet retailer	China
	SSTM/SITM[a]		
344	SSTM[a]	0.98 American ginseng + 0.02 Asian ginseng	
344	SITM[a]	0.90 American ginseng + 0.10 Asian ginseng	

[a] In each case, each of the 43 American ginseng samples were mixed with each of the eight Asian ginseng samples (43 × 8 = 344).

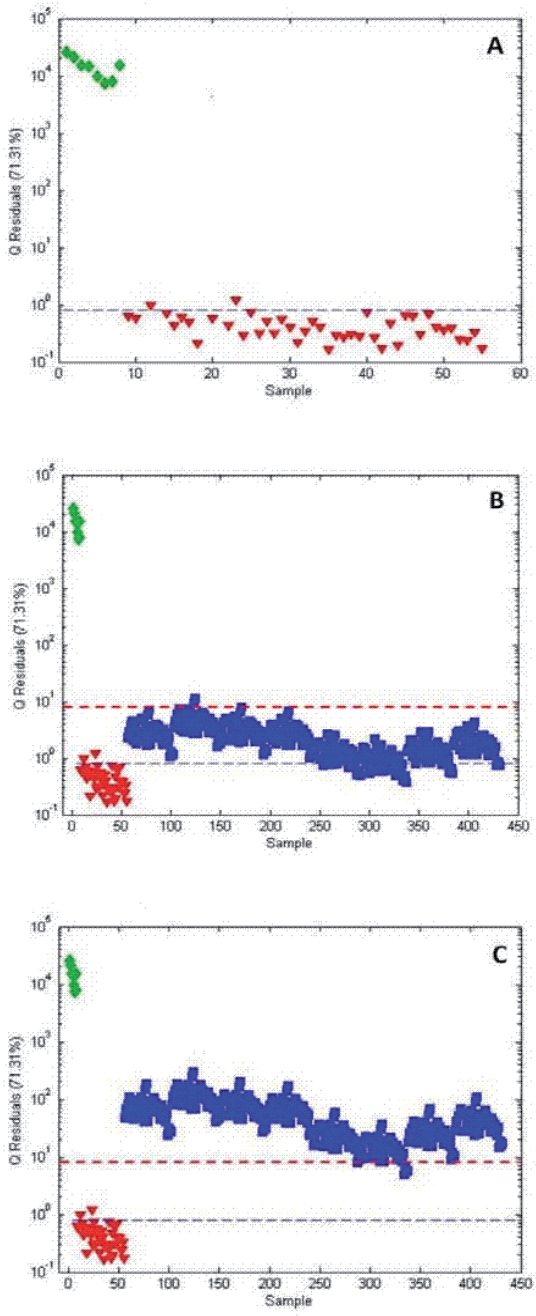


Figure 4. SIMCA plots for (A) 100% American ginseng (AG; ▼) and 100% Asian ginseng (CG; ◆); (B) SSTM (■), 100% AG, and 100% CG; and (C) SITM (■), 100% AG, and 100% CC.

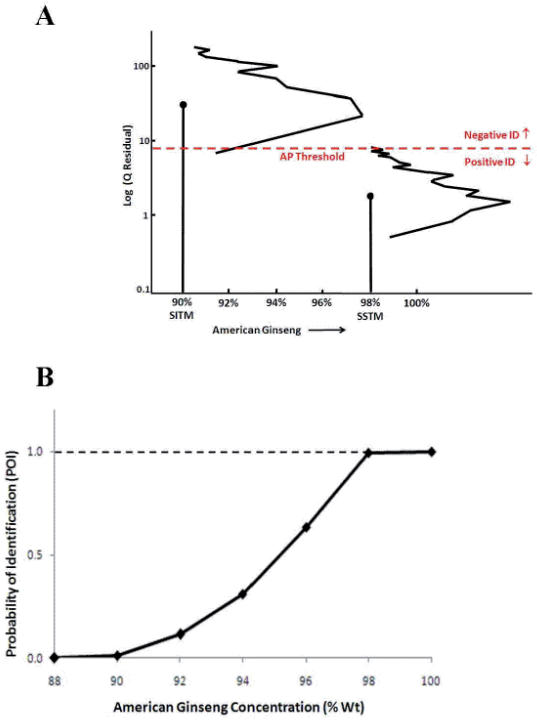


Figure 5. Target material AG, nontarget material CG: (A) SITM and SSTM, and (B) POI.

(inclusivity panel samples) as the model, the CG (exclusivity panel samples) falls well above the 95% confidence limit (dashed line). Both the AG and CG show considerable variation on the vertical axis, which reflects biological variation. Two of the AG samples fall above the 95% confidence limit, which is 4.6% for 43 samples and is to be expected.

For the SSTM/SITM study, 98 and 90% AG were arbitrarily selected as the MPRs for this model. Figure 4 (B) shows the SSTM samples (98% AG), as well as 100% AG and 100% CG samples. The pattern of eight groupings for the SSTM samples reflects that all 43 AG samples were diluted by each of the eight CG samples in sequence. A threshold of a Q residual value of 9.0 was selected arbitrarily and provides 99.4% positive identification (342 out of 344).

Figure 4 (C) shows the SITM at 90% AG. The threshold provides negative identification of the SITM for 99.1% of the samples (341 out of 344). The distribution of the SSTM and SITM are plotted in Figure 5 (A). The distributions appear to be roughly symmetrical. However, since the vertical axis is a logarithmic scale, the distributions are badly skewed on a linear scale and have dramatically different widths. If the SSTM were specified at a lower concentration of AG, or the SITM at a higher concentration, the method would not be appropriate unless lower confidence limits were chosen.

Based on the AP threshold shown in Figures 4 (B, C) and 5, the POI in Figure 5 (B) was computed. Synthetic samples of 96, 94, and 92% were generated and analyzed. The curve shape for the POI is very non-symmetric.

For our example, the SSTM corresponds to 98% AG mixed with 2% CG. The required minimum POI is 0.90, with 95% confidence for 100% SSTM (Table 2). The SITM corresponds to 90% AG mixed with 10% CG. The required maximum POI is 0.10,

Table 2. Example performance requirements

Requirement	SSTM, %	Measure	Limit	No. of replicates to be tested	No. of failures allowed[a]
POI	100	95% 1-sided LCL	0.90 (FNF<0.10)	60	2
POI	0	95% 1-sided UCL	0.10 (FPF<0.10)	60	2

[a] In each case, no more than two failures are allowed.

with 95% confidence. Table 2 shows that, for these performance requirements, 60 replicates must be tested at each level with no more than two failures. More stringent requirements (i.e., 0.95 and 0.05, with 95% confidence) would require more replicates and/or fewer failures. Conversely, less-stringent requirements would require fewer replicates. Depending upon the desired performance requirement for SSTM or SITM, alternative test plans (confidence levels) may be selected from Table 3. For more plans, *see* LaBudde (5).

Single-Laboratory Validation

Consider an example of a BIM being evaluated with respect to the performance requirements of Table 2. The internal operating methodology of the BIM is possibly a trade-secret of the method developer, and may not be known at the time of validation. All that is known for sure is that a test portion is utilized by the method, and binary result of yes = Identified or no = Not Identified is returned.

Consider testing in a single independent laboratory, or an SLV. With respect to the performance requirements of Table 2, the SITM and SSTM are used to prepare mixtures in the proportions 0:100%, 33:67%, 67:33%, and 100:0%. From each of these mixtures, 60

Table 3. Alternative test plans to obtain 1-sided upper 95% modified Wilson confidence limit at or below specified maximum value for FNF or FPF[a]

Specified maximum[b]	No. of replicates to be tested	No. of failures allowed[c]	1-sided 95% UCL[d]	2-sided 95% LCL[e]	2-sided 95% UCL[e]	AOQL[f]
0.20	11	0	0.197	0.000	0.259	0.129
0.20	20	1	0.196	0.000	0.236	0.118
0.20	24	1	0.167	0.000	0.202	0.101
0.20	36	3	0.191	0.029	0.218	0.124
0.20	48	5	0.199	0.045	0.222	0.133
0.20	72	8	0.187	0.057	0.204	0.131
0.15	20	0	0.119	0.000	0.161	0.081
0.15	24	0	0.101	0.000	0.138	0.069
0.15	36	1	0.115	0.000	0.142	0.071
0.15	48	3	0.146	0.021	0.168	0.095
0.15	72	5	0.136	0.030	0.152	0.091
0.10	40	0	0.063	0.000	0.088	0.044
0.10	48	1	0.088	0.000	0.109	0.054
0.10	60	2	0.096	0.009	0.114	0.061
0.10	72	3	0.100	0.014	0.115	0.065
0.05	60	0	0.043	0.000	0.060	0.030
0.05	72	0	0.036	0.000	0.051	0.025
0.05	96	1	0.045	0.000	0.057	0.028
0.02	130	0	0.020	0.000	0.029	0.014
0.02	240	1	0.018	0.000	0.023	0.012
0.01	280	0	0.010	0.000	0.014	0.007

[a] Excerpted from LaBudde (5).

[b] Desired maximum level of FNF or FPF to attain with 95% confidence.

[c] Maximum number of failures that can occur in the replicates tested and still meet specification.

[d] Worst-case 1-sided 95% modified Wilson upper confidence limit on FNF or FPF if maximum failures are observed.

[e] 95% modified Wilson 2-sided confidence interval on FNF or FPF if maximum failures are observed.

[f] Observed FNF or FPF corresponding to maximum failures allowed.

Table 4. Observed SLV results for example BIM

SSTM, %	No. of test portions	No. identified	No. not identified	POI
0.0	60	1	59	0.0167
33.3	60	7	53	0.1167
66.7	60	27	33	0.4500
100.0	60	60	0	1.0000

test portions are prepared, randomized, and labeled in a masked way. The test portions are measured by the BIM, each with a result of 0 or 1. Suppose example results are as shown in Table 4. Note the FPF performance requirement succeeds at 0% SSTM, because no more than two test portions reported identification. Also, the FNF performance requirement at 100% SSTM succeeds because, in both cases, fewer than two test portions were not identified.

Using the methods of Wehling et al. (3) and LaBudde (6, 7), the reported 1-sided and 2-sided 95% confidence intervals on the POI would be as shown in Table 5. Note that the 1-sided 95% confidence limit for the POI falls below 10% at 0% SSTM, and above 90% at 100% SSTM, indicating performance requirement success. The results in Table 5 are plotted in Figure 6.

Because the concentrations (% SSTM) are known with certainty here, one of several regression models might be fit to possibly obtain more precise estimates of POI and its confidence limits (although this is not guaranteed), but at the expense of some additional assumptions (*see Annex B*).

Collaborative Study

The primary purpose of a collaborative study is to establish that performance is reproducible among different collaborators (laboratories). A secondary purpose might be to compare the candidate method to another (possibly gold standard) method to establish differential performance (e.g., equivalency) across laboratories.

The primary purpose requires a minimum number of collaborators whose data persist (i.e., not excluded for cause) until the final results of the study. Rules of thumb in statistical mixed modeling (treating the collaborator effect as random) suggest that fewer than six collaborators does not allow inference with respect to the general collaborator population, eight collaborators allows reasonable estimation, and 10 collaborators is desirable. More than 10 collaborators is useful, but not necessary. For fewer than six collaborators, the collaborator effect should be regarded as fixed, and any inferences are applicable only to that particular set of collaborators, not some hypothetical general population of collaborators. The recommendation, therefore, is that 12 or more collaborators should be enrolled in the study, with a desired 8 to 10 remaining after removal for cause, and an absolute limit of no fewer than six remaining until the study end. Studies with this minimum number of collaborators can hope to provide a measure of collaborator effect or collaborator-method interaction, if one of reasonably large size exists.

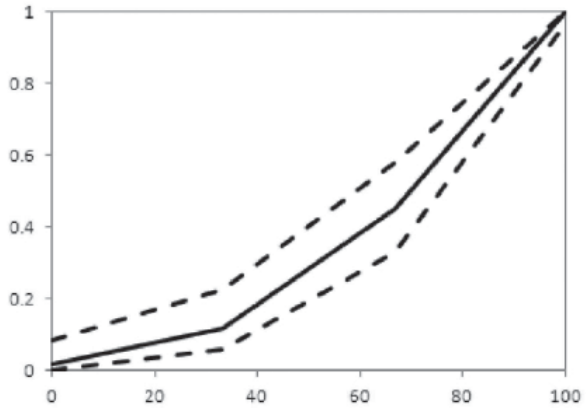


Figure 6. Expected POI versus %SSTM for an example BIM showing POI (solid line), lower 95% confidence limit (dashed line below the POI), and upper 95% confidence limit (dashed line above the POI). Note the POI at 0% is the false-positive fraction and 1-POI at 100% is the false-negative fraction.

Concentration levels (i.e., percentage of SSTM in a SSTM:SITM mixture) must include 0% SSTM (100% SITM) and 100% SSTM (0% SITM) in order to establish performance requirements (Figure 2). In addition, it is sometimes beneficial to provide for two intermediate concentrations (e.g., 33 and 67%) in order to provide information about identification performance across the range where the POI changes.

In order to isolate a collaborator effect in the presence of quantal noise (repeatability error), 12 replicates per collaborator is the suggested minimum. Therefore, the smallest acceptable collaborative study final data would be six collaborators × 12 replicates = 72 test portions.

It should be noted that due to the intercollaborator variation, a performance requirement imposed on a collaborative study will be more difficult for a candidate BIM to achieve than that imposed on an SLV study with the same number of total replicates. The performance requirements imposed on a single laboratory study and a collaborative study should be logically and statistically consistent.

The study director could, for example, prepare batches of SITM and SSTM, then prepare samples of mixtures at the 0:100%, 33:67%, 67:33%, and 100:0% proportions. From each of the well-mixed sample aliquots, test portions would be selected, such that each participating collaborator would receive the requisite number

Table 5. Reported SLV results

SSTM, %	*n*	ID	Not ID	POI	1-sided 95%	LCL 95%	UCL 95%
0.0	60	1	59	0.0167	0.0713	0.0000	0.0886
33.3	60	7	53	0.1167		0.0577	0.2218
66.7	60	27	33	0.4500		0.3309	0.5751
100.0	60	60	0	1.0000	0.9568	0.9398	1.0000

of replicates (*see* section on SLV). All test portions for each collaborator would be randomly assigned IDs before distribution. The study is masked so that collaborators cannot visually identify the composition of the test portions. Additional unmasked test portions may be provided for proficiency training purposes. Each collaborator would use the BIM according to instructions to analyze each test portion provided, and report results by test portion number and 1 = Identified or 0 = Not Identified.

Suppose a collaborative study is to be evaluated with respect to the performance requirements of Table 2. The primary goal is to validate that performance is sufficiently homogeneous across collaborators and that the performance requirements are met. As mentioned before, the number of replicate test portions for each collaborator should be 12 or more to control the quantal repeatability error sufficiently to allow detection of an intercollaborator effect. Suppose the plan was to enroll 12 collaborators, with the expectation that on or two might have to be removed for cause (spoilage of test portions, failing to follow instructions, cross-contamination, etc.) Consequently 144 test portions are prepared for each of the four % SSTM values (0, 33.3, 66.7, and 100%).

After completion of the study, two collaborators are removed for cause, and the results shown in Table 6 are obtained. For the 0% SSTM concentration, the statistical analysis of the data gives the results in Table 7. There is no detected intercollaborator effect (P-value = 0.43, point estimate = 0.00, confidence interval includes 0.000 and has an upper limit of 0.040), and the upper 2-sided confidence limit for combined POI is 0.0457, well below the performance requirement of 0.10. There is little evidence that the method is irreproducible, and the method meets the POI (or FPF) performance requirement.

For the 33% SSTM concentration, the statistical analysis of the data gives the results in Table 8. Again, there is no detected intercollaborator effect (P-value = 0.66), so there is little evidence that the method is irreproducible.

For the 67% SSTM concentration, the statistical analysis of the data gives the results in Table 9. Once again, there is no detected intercollaborator effect (P-value = 0.18), so there is little evidence that the method is irreproducible.

Finally, for the 100% SSTM concentration, the statistical analysis of the data gives the results in Table 10. There is no detected intercollaborator effect (P-value = 0.25, point estimate = 0.027, confidence interval includes 0.000 and has an upper limit of 0.093), and the lower 2-sided confidence limit for combined POI is 0.917, well above the performance requirement of 0.90. There is little evidence that the method is irreproducible, and the method meets the POI (or FNF) performance requirement.

Lot-Lot Variability, Time Stability, and Robustness Studies

The SLV and collaborative studies discussed above do not represent worst-case, end-of-life conditions with respect to method materials and parameters. For this reason, it is customary to augment these studies with additional studies to verify proper results despite reasonable variations among method materials, equipment, and parameters.

A lot-lot variability study is meant to verify results across different lots of method materials (supplies used) and sets of equipment. Each lot would consist of a different manufactured or prepared batch of materials (reagents, supplies, etc.), and possibly a different set of measurement equipment. Date of manufacture is not an issue in this study, only variation among lots, so ideally, the lots tested should have been produced at near the same times.

Table 6. Collaborative study results

SSTM, %	Collaborator	Replicates	No. identified
0	1	12	1
0	2	12	0
0	3	12	0
0	4	12	0
0	5	12	0
0	6	12	0
0	7	12	0
0	8	12	0
0	9	12	0
0	10	12	0
33.33	1	12	2
33.33	2	12	2
33.33	3	12	2
33.33	4	12	2
33.33	5	12	0
33.33	6	12	1
33.33	7	12	1
33.33	8	12	4
33.33	9	12	2
33.33	10	12	3
66.67	1	12	4
66.67	2	12	9
66.67	3	12	5
66.67	4	12	8
66.67	5	12	7
66.67	6	12	4
66.67	7	12	7
66.67	8	12	3
66.67	9	12	8
66.67	10	12	5
100	1	12	12
100	2	12	10
100	3	12	11
100	4	12	12
100	5	12	12
100	6	12	11
100	7	12	12
100	8	12	12
100	9	12	12
100	10	12	12

Table 7. Collaborative study results for 0% SSTM concentration

AOAC Binary Data Interlaboratory Study Workbook Study Reported Values, Version 2.2					
Sample ID 0% SSTM					
Sequence	Item	Symbol	Value	Approximately 95% LCL[a]	Approximately 95% UCL[b]
1	Total number of laboratories	p	10		
2	Total number of replicates	Sum(n(L))	120		
3	Overall mean of all data (grand mean)	LPOI or LPOD	0.0083	0.0015	0.0457
4	Repeatability SD	s(r)	0.0913	0.0807	0.1713
5	Among-laboratories SD	s(L)	0.0000	0.0000	0.0402
6	Homogeneity test of laboratory PODs	P-value	0.4303		
7	Reproducibility SD	s(R)	0.0913	0.0814	0.1064
8	Intraclass correlation coefficient for repeatability	I(r)	1.0000	0.8335	1.0000

[a] LCL = Lower confidence level.

[b] UCL = Upper confidence level.

Just as with collaborators in a collaborative study, estimation of the lot random effect requires that at least six different lots be involved in the study. Each lot should result in attainment of any BIM performance requirements, and the variation in performance among lots should be immaterial in size.

A time stability study is meant to verify that there is no material degradation in performance over the life of lots of materials and equipment. This may be accomplished by determination of the parametric aging effect by use of time-staggered lots, or simply verifying performance on end-of-life lots.

Note that the lot-lot variability and time-stability studies cannot be merged into a single study unless there are sufficient replicate lots at or near the same time point(s) to allow separation of the lot-lot and time effects. If lot-lot and time effects are negatively correlated, one factor may mask the effect of the other in an inadequate combined study (e.g., a different single lot at each different time point). Testing only end-of-life lots would be a satisfactory combined study, even though time and lot effects could not be resolved.

A robustness study (also denoted a sensitivity study) is meant to verify performance under worst-case conditions of method critical parameter (e.g., times, temperatures, concentrations) variation. Disturbances of method parameters should reflect maximum excursions to be expected in practical use. Performance requirements should be met at each of these excursions. The statistical design should be capable of measuring at least main effects.

Conclusions

The purpose of a qualitative BIM is to discriminate between acceptable target material and target material with an unacceptable concentration of nontarget material. This concept was particularized to discrimination between the SSTM and SITM for the purpose of method validation. A general overview of the application of the POI model and analysis was given, which allows validation and/or characterization of qualitative BIMs. Examples are given for both SLV and collaborative studies with MPRs. The use of POI statistics harmonizes statistical concepts among botanical, microbiological, toxin, and other analyte identification or detection methods for which binary results are obtained. The POI statistical model provides a tool for graphical representation of response curves for qualitative methods, reporting of descriptive statistics, and application of performance requirements.

Table 8. Collaborative study results for 33.33% SSTM concentration

AOAC Binary Data Interlaboratory Study Workbook Study Reported Values, Version 2.2					
Sample ID 33.33% SSTM					
Sequence	Item	Symbol	Value	Approximately 95% LCL	Approximately 95% UCL
1	Total number of laboratories	p	10		
2	Total number of replicates	Sum(n(L))	120		
3	Overall mean of all data (grand mean)	LPOI or LPOD	0.1583	0.0913	0.2253
4	Repeatability SD	s(r)	0.3703	0.3272	0.4266
5	Among-laboratories SD	s(L)	0.0000	0.0000	0.1400
6	Homogeneity test of laboratory PODs	P-value	0.6563		
7	Reproducibility SD	s(R)	0.3703	0.3304	0.4275
8	Intraclass correlation coefficient for repeatability	I(r)	1.0000	0.8889	1.0000

Table 9. Collaborative study results for 66.67% SSTM concentration

AOAC Binary Data Interlaboratory Study Workbook Study Reported Values, Version 2.2					
Sample ID 66.67% SSTM					
Sequence	Item	Symbol	Value	Approximately 95% LCL	Approximately 95% UCL
1	Total number of laboratories	p	10		
2	Total number of replicates	Sum(n(L))	120		
3	Overall mean of all data (grand mean)	LPOI or LPOD	0.5000	0.3919	0.6081
4	Repeatability SD	s(r)	0.4939	0.4364	0.5222
5	Among-laboratories SD	s(L)	0.0948	0.0000	0.2779
6	Homogeneity test of laboratory PODs	P-value	0.1783		
7	Reproducibility SD	s(R)	0.5029	0.4489	0.5222
8	Intraclass correlation coefficient for repeatability	I(r)	0.9644	0.7547	1.0000

Table 10. Collaborative study results for 100.0% SSTM concentration

AOAC Binary Data Interlaboratory Study Workbook Study Reported Values, Version 2.2					
Sample ID 100% SSTM					
Sequence	Item	Symbol	Value	Approximately 95% LCL	Approximately 95% UCL
1	Total number of laboratories	p	10		
2	Total number of replicates	Sum(n(L))	120		
3	Overall mean of all data (grand mean)	LPOI or LPOD	0.9667	0.9174	0.9870
4	Repeatability SD	s(r)	0.1784	0.1576	0.2055
5	Among-laboratories SD	s(L)	0.0273	0.0000	0.0930
6	Homogeneity test of laboratory PODs	*P*-value	0.2506		
7	Reproducibility SD	s(R)	0.1804	0.1610	0.2121
8	Intraclass correlation coefficient for repeatability	I(r)	0.9772	0.7818	1.0000

Acknowledgments

We wish to thank the Expert Review Panel for Botanical Identification Methods for kindly reviewing this article and supplying numerous comments for improvement. In particular, we wish to thank Paul Wehling of General Mills/Medallion Laboratories and Danica Reynaud of AuthenTechnologies for the extraordinary amount of time they spent both in reviewing and providing constructive criticism.

References

(1) U.S. Food and Drug Administration (1994) *Dietary Supplement Health and Education Act of 1994,* Washington, DC
(2) U.S. Food and Drug Administration (2007) *Current Good Manufacturing Practice in Manufacturing, Packaging, Labeling, or Holding Operations for Dietary Supplements, Code of Federal Regulations, Title 21, Part III,* U.S. Government Printing Office, Washington, DC
(3) Wehling, P., LaBudde, R.A., Brunelle, S.L., & Nelson, M.T. (2011) *J. AOAC Int.* **94**, 335–347
(4) LaBudde, R.A. (2008) *Statistical Analysis of Interlaboratory Studies, XX, Measuring the Performance of a Qualitative Test Method,* TR290, Least Cost Formulations, Ltd, Virginia Beach, VA
(5) LaBudde, R.A. (2010) *Sampling Plans to Verify the Proportion of an Event Exceeds or Falls Below a Specified Value,* TR308, Least Cost Formulations, Ltd, Virginia Beach, VA
(6) LaBudde, R.A. (2009) *Coverage Accuracy for Binomial Proportion 95% Confidence Intervals for 12 to 100 Replicates,* TR297, Least Cost Formulations, Ltd, Virginia Beach, VA
(7) LaBudde, R.A. (2009) *Statistical Analysis of Interlaboratory Studies, XXII, Statistical Analysis of a Qualitative Collaborative Study as a Quantitative Study Under the Large Sample Approximation,* TR296, Least Cost Formulations, Ltd, Virginia Beach, VA

ANNEX A
SIMCA

Principal component analysis (PCA) is a mathematical procedure used to convert observations for samples with a large number of possibly correlated variables (ions, wavelength, or wavenumbers) into a set of uncorrelated variables called principal components (1). The transformation takes place in manner that assigns the maximum variance to the first principal component with less variance being accounted for by each successive principal component. PCA is applied to the entire data set to determine what groupings of the samples can be seen without any prior decisions (i.e., it is unsupervised). The first two or three principal components (displayed as two- or three-dimensional plots) can be used to demonstrate general patterns in the data.

SIMCA is a supervised approach that builds a PCA model for each specified category of samples (2). Distances between the models are then used to determine the independence of each category of samples. New samples can be assigned to one of the categories or classified as not fitting in any of them.

SIMCA is used for BIMs because predetermined categories of samples are established and modeled. For a BIM, however, only a single PCA model is constructed, and that is for the samples in the inclusivity panel. All other samples are then evaluated using the PCA model to determine whether it is described by the inclusivity PCA model or whether it lies a significant distance from the model, i.e., it does not belong to the inclusivity panel category of samples.

Two statistics used to evaluate whether a sample fits the PCA model are the Q residual and the Hotelling T^2 statistic. The Hotelling T^2 statistic is the multivariate analog of the univariate Students' *t* statistic. It describes how a sample fits in the model. The Q residual, also called the squared prediction error, is more commonly used for process control applications. It describes how far a sample falls outside the model. Some chemometric programs provide both of these statistics as a means of evaluating the fit of a PCA model to the data (1).

Figure A1 provides a simplified illustration of the relationship of the two statistics. In this case, a PCA model is fit to one category of samples. Since only the first principal component was used for this model, the model is a straight line. The data have been mean-centered, so they are centered around the origin, i.e., the intersection of the *x* and *y*-axis. The distribution of each sample with respect to the model is determined by dropping a line from the sample point perpendicular to the model line. The distance from the point where the perpendicular of a sample intersects the model line to the origin provides the Hotelling T^2 value for that point. With sufficient data and a normal distribution, the data distribution should appear as a bell-shaped function centered at the origin. Using this distribution, it can be determined whether a sample is well-fit by the model, i.e., falls inside the 95% confidence limits.

The variance of the sample data with respect to the model is the variance computed along the straight line. In this case, it would be analogous the Students' *t* calculation, i.e., the sum of square of the distance for each sample. In Figure A1, the first principal component for the modeled category passes through the sample data in a manner that provides the maximum variance. A second principal component, perpendicular to the first, would account for the distance of the points from the line and, in this case, provide far less variance than the first principal component. For a model based just on the first principal component, the variance associated with the distance of the sample points from the line is accounted for by the Q residual.

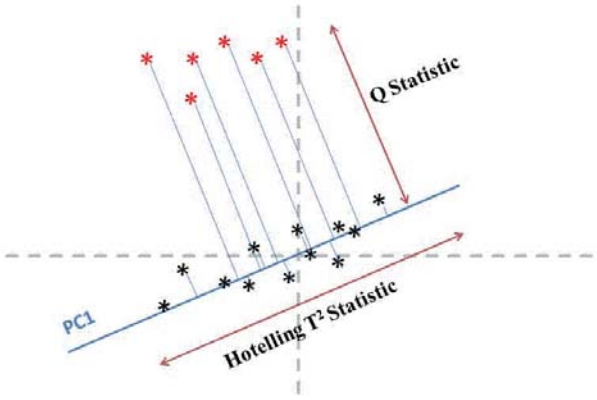


Figure A1. Illustration of Hotelling T^2 and Q statistic: (*) modeled samples and (*) unknown samples.

The distribution of unmodeled data from a second category of samples can be evaluated using the model for the first category of samples. As shown in Figure A1, the distribution of the second category of samples on the first model is very reasonable. Perpendicular lines from the samples in the second category intercept the model line at reasonable distances from the origin. If this were real data, and a 95% confidence limit had been computed, the second category of samples would undoubtedly be within that limit. However, for the second category of samples, a much larger fraction of the total variance is incorporated in the distance from the model line. The second category samples will fall well outside the 95% confidence limit for the Q residual established by the first category samples.

SIMCA can be applied to a BIM by constructing a PCA model using the data from the inclusivity panel botanical materials. New samples are fit to the model and the Q residual is determined. If the Q residual for a sample falls outside the 95% confidence limit, the new sample is not the same as the target materials. Conversely, if the new sample falls within the 95% confidence limit, it would be classified as a target material.

References

(1) Wold, S., & Sjostrom, M. (1977) in *Chemometrics Theory and Application*, American Chemical Society Symposium Series 52, American Chemical Society, Washington, DC, pp 243–282
(2) Wold, S. (1987) *Chemom. Intel. Sys*. **2**, 37–52

ANNEX B
Modeling of the POI Using Logistic Regression

The models in common use for this kind of problem include, among many others: (*1*) discriminant analysis; (*2*) logistic regression; or (*3*) normit regression. There is also a choice of metamer x (i.e., transform of %SSTM). Common choices include x = % SSTM, or x = $\log_{10}$ (%SSTM + 0.5). Logistic and normit regression assume the POI versus x curve is symmetrical, which that of Figure 4 obviously is not.

Suppose we choose logistic regression with an identity metamer (x = % SSTM), which implies the model:

```
Call:
glm(formula = cbind(id, notid) ~ x, family =
binomial("logit"),
    data = dat)
Deviance Residuals:
      1       2       3       4
 0.8314  0.9386 -1.5687  2.6222
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -5.04711    0.67021  -7.531 5.05e-14 ***
x            0.07878    0.01001   7.869 3.57e-15 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.'
0.1 ' ' 1
(Dispersion parameter for binomial family taken to be
1)
    Null deviance: 186.241  on 3  degrees of freedom
Residual deviance:  10.908  on 2  degrees of freedom
AIC: 25.12
Number of Fisher Scoring iterations: 5
```

Figure B1. Fit of Equation 1 to the sample data.

$$\text{logit(POI)} = \ln\{\text{POI}/(1 - \text{POI})\} = \alpha + \beta x = \alpha + \beta\ (\%\ \text{SSTM}) \quad \text{(Equation 1)}$$

For the sample data, the fit is as shown in Figure B1.

The model fits poorly and is highly overdispersed (dispersion = 10.908 / 2 = 5.454). Consequently, the standard errors found in the fit should be multiplied by 2.34 = √5.454. (Note that this overdispersion suggests that the logistic regression model with specified link is a poor choice for the data.)

An estimate of the point at which POI = 0.5000 is given by the negative ratio of the intercept by the slope, or x = 64.1% SSTM. This would be denoted "Effective Concentration at POI = 0.50" or "EC50." (It should be noted that EC50 depends upon the definitions of the SSTM and SITM.)

From the logistic regression fit, we get the results shown in Table B1 and Figure B2. The logistic regression does not do as well as the direct POI descriptive statistics of Table 6, because of serious failure of the model assumptions. (It turns out that *none* of the usual generalized model forms fits the asymmetrical POI versus % SSTM curve very well for this example. So it should be noted that the standard error of POI is *not* always reduced by fitting across the combination of concentrations used.) Note that, based on the logistic model, the BIM continues to pass the 0% SSTM performance requirement, but fails the 100% SSTM requirement.

It is generally recommended that the methods of Table 6 be used for evaluating performance requirements, rather than those of unvalidated regression models. One of the advantages, however, of fitting such a model is that continuous curves may be obtained, as shown in Figure B3.

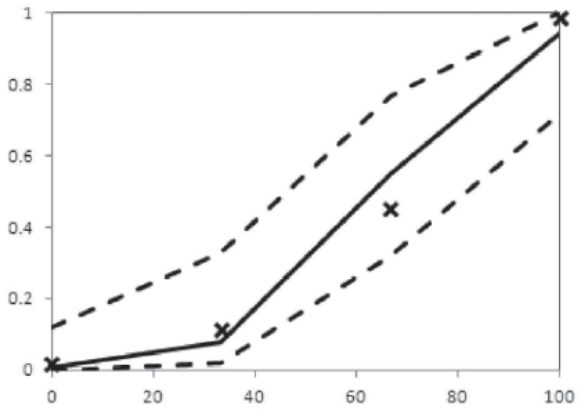


Figure B2. Example SLV results from a logistic regression fit showing POI (solid line), lower 95% confidence limit (dashed line below the POI), and upper 95% confidence limit (dashed line above the POI), and measured POI values (X).

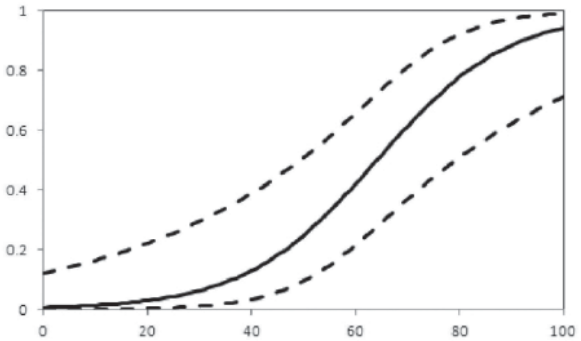


Figure B3. Continuous curves from SLV logistic regression fit showing POI (solid line), lower 95% confidence limit (dashed line below the POI), and upper 95% confidence limit (dashed line above the POI).

Table B1. SLV results (logistic regression fit)

	Fitted	Obs.	1-sided	LCL	UCL
% SSTM	POI	POI	95%	95%	95%
0.0	0.0064	0.0167	0.0778	0.0003	0.1214
33.3	0.0816	0.1167		0.0162	0.3239
66.7	0.5511	0.4500		0.3181	0.7636
100.0	0.9443	1.0000	0.7715	0.7126	0.9915