Validation and Peer Review
of U.S. Environmental Protection Agency
Chemical Methods of Analysis

Prepared for:

The EPA Forum on Environmental Measurements (FEM)

Prepared by:

The FEM Method Validation Team

FEM Document Number 2005-01
October 14, 2005
REVISION: February 3, 2016

Principal Authors:

Elizabeth A. Mishalanie, Ph.D.
National Enforcement
Investigations Center
Denver, Colorado

Barry Lesnik
Office of Solid Waste
Washington, DC

Roy Araki
US EPA Region 10
Seattle, Washington

Robin Segall
Office of Air Quality Planning and
Standards
Research Triangle Park, North Carolina

Contributors:

David J. Munch
Office of Water
Office of Groundwater
and Drinking Water
Cincinnati, Ohio

Mike Wasko
US EPA Region 4
Athens, Georgia

Elizabeth Flynt
Office of Pesticide Programs
Stennis Space Center, Mississippi

Steven C. Wendelken, Ph.D.
Office of Groundwater
and Drinking Water
Cincinnati, Ohio

Jenny Scifres
US EPA Region 4
Athens, Georgia

John R. Mario
Suffolk County Crime Laboratory
Hauppauge, New York

Margo Hunt
Office of Environmental Information
Washington, DC

Diane Gregg
US EPA Region 6
Houston, Texas

Teri Schaeffer
Versar, Inc.
Springfield, Virginia

Meredith Clarage
US EPA Region 6
Houston, Texas

M. Susan Lumpkin
National Exposure Research Laboratory
Research Triangle Park, North Carolina

Disclaimer

This document contains direction developed solely to provide internal guidance to U.S. Environmental Protection Agency (EPA) personnel. EPA retains the discretion to adopt approaches that differ from this guidance on a case-by-case basis. The guidance set forth in this document does not create any rights, substantive or procedural, enforceable at law by a party to litigation with EPA or the United States. The use of mandatory language such as "must" and "require" in this guidance document reflects sound scientific practice and should not be construed to create any legal rights or requirements. The use of non-mandatory language such as "may," "can," or "should" in this guidance does not connote a requirement but does indicate EPA's strongly preferred approach for validating and peer reviewing EPA methods prior to publication for general use.

Foreword

The U.S. EPA Science Policy Council established the Forum on Environmental Measurements (FEM) in April 2003. The FEM is a standing committee of senior EPA managers who provide EPA and the public with a focal point for addressing measurement and methods issues with multiprogram impact. Action teams are commissioned by the FEM to address specific issues. The Method Validation Team was formed in October 2003, and tasked with developing Agency-wide, internal guidance for validating and peer reviewing EPA methods prior to publication for general use.

This document contains guidance for chemical methods of analysis. It is the first in a series of Agency-wide, method validation guidance documents. Subsequent guidance will be developed for other types of measurement methods (e.g., microbiology, biology, toxicology, radiochemistry, and field procedures).

Executive Summary

EPA program offices publish a wide variety of measurement methods for use by EPA personnel, other government agencies, and the private sector. These methods may originate from many sources such as EPA laboratories, EPA contractors, scientific organizations, other government laboratories, and from the private sector. Since these methods may be published as regulations, incorporated by reference in regulations, or published as guidance, they must be thoroughly tested and peer reviewed prior to publication as EPA methods.

This document provides Agency-wide guidance for EPA personnel who will evaluate the performance and suitability of new chemical methods of analysis before EPA publication. The method validation principles contained herein are based on current, international approaches and guidelines for intralaboratory (single laboratory) and interlaboratory (multiple laboratory) method validation studies. Peer review is required for EPA chemical methods of analysis. The EPA Science Policy Council Peer Review Handbook provides Agency-wide requirements and options for that process.

Table of Contents

1 Introduction

Method validation is the process of demonstrating that an analytical method is suitable for its intended use, and involves conducting a variety of studies to evaluate method performance under defined conditions (*1*). Method validation studies may involve a single laboratory (intralaboratory) or multiple laboratories (interlaboratory). The goal is to demonstrate that analytical results produced by the application of a particular method are fit for an intended purpose (*1*). Properly designed and successful method validation studies create confidence in the reliability of a test method. Method validation is one of several important quality system components that are designed to ensure the production of scientifically valid and useful analytical data (*2*).

EPA's chemical methods of analysis are designed for a wide variety of measurement objectives, such as monitoring compliance with environmental regulations, enforcement of environmental regulations, and for information-gathering scientific studies. Because measurement objectives and data quality requirements vary, the technical details and acceptance criteria for conducting method validation studies are contained in existing, program-specific EPA documents rather than in an Agency-wide guidance. Therefore, this guidance document describes fundamental method validation principles and general areas to be addressed when validating chemical methods of analysis. This information is based on current, international approaches and guidelines for single and multiple laboratory method validation studies.

Before publication, all chemical methods of analysis also must be peer reviewed. The reader is referred to the most current version of the EPA Science Technology Policy Council Peer Review Handbook (*3*), an excellent source of information for both internal and external peer review processes. Emergency response situations (e.g., natural disaster, homeland security) are unique circumstances that require more rapid response. EPA developed an emergency response method validation policy (*Ensuring the Validity of Agency Methods Validation and Peer Review Guidelines: Methods of Analysis Developed for Emergency Response Situations*; Agency Policy Directive Number FEM-2010-01) that should be followed in these instances.

1.1 Purpose

The purpose of this guidance is to (a) describe scientific principles that should be addressed during method validation studies, and (b) harmonize terms and definitions related to analytical chemistry method validation studies and method performance characteristics.

1.2 Intended Audience

This guidance is intended for internal use by EPA personnel who are responsible for chemical methods of analysis, as defined in section 1.3 of this document, that either are to be published as serially numbered EPA methods, or published as, or incorporated by reference in, regulations.

1.3 Scope of Guidance

This guidance contains recommendations for validating new, rational, quantitative, chemical methods of analysis, intended for use in analytical laboratories. These terms are defined as follows:

- "New" means that the method is not currently published as an EPA method, or it is an update of a published EPA method that includes procedural changes that will affect method performance.

- "Rational," as opposed to empirical, means that the analytical results are *not* intended to be method dependent. A rational method of analysis is a method that determines one or more identifiable chemicals or analytes for which there may be several equivalent methods of analysis available. Empirical methods determine a value that can be arrived at only in terms of the method *per se* and serves, by definition, as the only method for establishing the measurand[1]. This guidance does not explicitly address empirical methods, which are used to estimate method-defined parameters. Examples of rational methods are Method 8260 (Volatile Organic Compounds by Gas Chromatography/Mass Spectrometry, SW-846), Method 200.8 (Determination of Trace Elements in Waters and Wastes by Inductively Coupled Plasma-Mass Spectrometry), and Method 26 (Determination of Hydrogen Chloride Emissions from Stationary Sources, 40 CFR Part 60, Appendix A). Examples of empirical methods are Method 1311 (Toxicity Characteristic Leaching Procedure, SW-846), Method 1664 (Oil and Grease), and Method 5 (Determination of Particulate Matter Emissions from Stationary Sources, 40 CFR Part 60, Appendix A).

- "Quantitative" means that the method is intended to produce numerical results (analyte concentration or other quantity) and an estimate of the associated measurement uncertainty (*4*).

- "Chemical methods of analysis, intended for use in analytical laboratories" means that the method primarily is intended for use in a laboratory to measure chemical properties. Although some methods may include sampling procedures (e.g., air Methods TO-15 and 26, solid waste Method 5035A), validation of field procedures is not explicitly addressed in this guidance.

    Much of the information provided may be useful for validating other types of analytical methods such as empirical, qualitative, or field procedures. However, these types of methods will have unique considerations that will require slightly different validation approaches from those presented in this document.

1.4 Terminology

    Scientific terms and meanings change with time. For many years, national and international standards organizations have attempted to harmonize terminology within scientific disciplines. For the purpose of this guidance, a glossary of terms and definitions is included in Section 9. The entries were compiled from current, authoritative sources, and references are included.

---

1   *Rational and empirical methods of analysis are defined and discussed in an IUPAC technical report by Thompson et al. The report was published in Pure & Applied Chemistry, Vol. 71, No. 2, pp. 337-348, 1999.*

2 Planning and Initiating Method Validation Studies

Proper planning is critical for successful method validation studies. The development of a chemical method of analysis may require months or years, and the formal validation phases should not be pursued until substantial evidence exists that a method is a good candidate for publication as an EPA method. EPA scientists are relied upon to determine which methods are good candidates, and it is not the intention of this guidance to provide acceptance criteria for new methods.

Method development and method validation studies generally are not viewed as completely separate processes. Method development, often a complex iterative process, is not explicitly addressed in this guidance. However, there are many similarities between the types of experiments conducted during method development and the types of tests performed during method validation.

Typically, an intralaboratory method validation study is initiated when a detailed procedure has been written, operational parameters are believed to be optimized and finalized, and no additional changes are anticipated. On occasion, difficulties are encountered after beginning an intralaboratory study, and it is necessary to troubleshoot and optimize a method. If procedural changes are made, and the changes may affect the method performance characteristics, then some or all of the performance tests may need to be repeated.

An interlaboratory method validation study is initiated only after a successful intralaboratory study has been completed. Unfortunately, some method performance difficulties may only be revealed after initiating a multiple laboratory study, and it may be necessary to troubleshoot and optimize a method after conducting such a study. If procedural changes are required, the changes may affect the method performance characteristics and some or all aspects of an interlaboratory study may need to be repeated.

3 Method Description

3.1 Procedures Composing a Method

The measurement process components subjected to validation should be clearly described. For example, a method may include procedures for laboratory subsampling, subsample preparation, and instrumental analysis. In this case, the method validation effort should include an evaluation of all three measurement process components combined. Some chemical methods of analysis also include procedures for sampling (e.g., methods for air analysis and for volatile organic compounds in soils and other solid matrices), in which case the sampling procedures should be included in the validation tests. In contrast, many SW-846 methods (EPA publication "Test Methods for Evaluating Solid Waste: Physical/Chemical Methods") are separate procedures for sample preparation, sample cleanup, or instrumental analysis. The SW-846 methods may be validated and used in various combinations to obtain the final analytical results. In any case, the measurement process components validated, and the combination of procedures composing a method, must be clearly and completely stated.

3.2 Method Purpose, Scope, and Applicability

The purpose of a method (i.e., measurement objectives) and the intended use of the data must be clearly defined. In addition, method scope and applicability must be well defined and clearly described. This helps minimize misapplications by the user. Method scope and applicability includes the following:

- The measurement process components validated
- The nature of the analytes and matrices studied
- The range of analyte levels for which the method is claimed to be suitable
- A description of any known limitations and any assumptions upon which a method is based
- A description of how the method and analytical parameters chosen meet the data quality objectives for the specific application

4 Method Performance Characteristics

The purpose, scope, and applicability determine the method performance characteristics chosen for study. The accuracy of a method, with respect to the materials and conditions studied during validation, may be evaluated from the performance characteristics. The meaning of the term "accuracy" has changed over the years, and accuracy should be viewed as a qualitative concept rather than a synonym for bias (*4, 5*). Accuracy is defined as the closeness of agreement between a measured value and either a true or accepted reference value. The following excerpt from reference 4 (pages 167-168) provides a clear explanation of the current use of the term accuracy.

> Part 1 of the international standard ISO 5725-1 on the accuracy of measurement methods and results defines accuracy as the closeness of agreement between a test result and the accepted reference value. This definition is supplemented by a note which states that the term accuracy when applied to a set of test results involves a combination of random components and a common systematic error or bias component. Accuracy is thus viewed as a characteristic of a measurement process consisting of precision as well as bias components. A process is considered to be accurate only if it is precise as well as unbiased. A measurement process that generates repeat observations of a characteristic which are not close enough to one another cannot generate an accurate result. In fact, a new word has been coined to replace the word *bias*. Bias is now characterized by the attribute "trueness." *Trueness* indicates the positive aspect of a measurement process, in contrast to bias, which indicated a negative aspect. The old concept of accuracy has been changed to mean trueness. The greater the bias of a measurement process, the less true it is. The term *accuracy* is now used in a broader context which includes both precision and trueness. Many times a measurement result is derived as an average of repeat measurement observations rather than a single observation. When the result is an average, it is likely that it is close to the conventional true value but the repeat measurements are not close enough to one another. Such a measurement process may be considered to be true but not accurate, as the process is not precise.

Based on current trends in classifying method validation study components (*2*), the performance characteristics[2] that should be evaluated include, but are not limited to:

- Selectivity
- Instrument calibration
- Bias/Trueness
- Precision
- Quantitation limits and range
- Detection limits

"Sensitivity" is not listed in this guidance as a method performance characteristic. The scientific literature reflects different meanings and uses for this term. It is used to describe a performance characteristic of a detector, instrument, or an analytical method. However, it is possible to evaluate any of these sensitivity parameters from data and information obtained by studying the performance characteristics listed above.

---

2 *Most performance characteristics listed in this section are also studied during method development. However, experiments designed for method development purposes are typically not performed under final operating conditions.*

Before validating an analytical method, it is necessary to ensure that proper sample preservation and storage stability studies were performed during method development. Storage stability must investigate the stability of the analyte(s) from the time of sampling through the time of analysis. If an extraction is performed, the extract stability must also be investigated. Analytes may be lost through volatilization, sorption, chemical degradation (abiotic reactions), and microbial degradation. Traditionally, preservation to prevent chemical degradation is occasionally evaluated but preservation to prevent microbial degradation is generally ignored. Microorganisms have the potential to degrade target analytes and represent a significant pathway for the fate and destruction of organic compounds.

4.1 Selectivity

Selectivity refers to the ability to correctly identify the analyte(s) of interest in the presence of expected chemical/physical interferences. Selectivity may be evaluated by testing appropriate matrix blanks and fortified (spiked) matrix blanks. Matrix blanks should contain the chemical species, other than the analyte(s) of interest, which are reasonably expected to be present in an actual sample. A worst-case sample matrix should be tested, as defined in the scope of a method, during both intra- and interlaboratory validation studies.

In some cases, adequate representations of matrix blanks cannot be obtained or prepared. In such situations, actual samples of material containing an undetermined amount of native analyte may need to be studied to evaluate method selectivity. This usually requires the use of multiple analytical procedures (e.g., sample preparation, wet-chemistry, and/or instrumental analysis procedures) during an intralaboratory validation study to substantiate the selectivity of the method being evaluated. A combination of procedures or techniques that are based on different measurement principles should be used. This approach increases the likelihood that interferences will be identified.

Method selectivity is typically expressed qualitatively. A qualitative selectivity statement includes a description of known interferences, interference effects, and the nature of the analytical data and information that substantiates the identity of the analyte(s) in the matrix of concern (e.g., elemental or molecular structure data, retention times from chromatographic separations, selective reaction chemistry, and results from reference standards, reference materials, matrix blanks, other blanks, or matrix fortifications). Quantitative measures of selectivity may also be used, although there is no generally accepted approach for the quantitative treatment of selectivity data (*2, 6*). Therefore, the basis for quantitative selectivity measures should be thoroughly described.

4.2 Instrument Calibration

Instrument calibration refers to the procedures used for correlating instrument response to an amount of analyte (concentration or other quantity). The characteristics of a calibration function and justification for a selected calibration model should be demonstrated during an intralaboratory method validation study.

The performance of a calibration technique and the choice of calibration model (e.g., first-order linear, curvilinear, or nonlinear) are critical for minimizing sources of instrument bias and

optimizing precision. A calibration model is a mathematical function that relates composition to instrument response. The parameters of the model are usually estimated from the responses of known, pure analytes. Calibration errors can result from failure to identify the best calibration model; inaccurate estimates of the parameters of the model; or inadequately studied, systematic effects from matrix components.

During method development and validation, calibration models are typically evaluated by analyzing multiple levels of calibration standards over a selected working range. Measurements should be replicated at each calibrant level to evaluate the random error associated with instrument response, and to evaluate goodness-of-fit for various linear, curvilinear, or non-linear models. For example, if a method states that an average response factor may be used for calibration over a specified analyte range, then the instrument response should be demonstrated, during intralaboratory method validation, to be a first-order linear equation, and that the y-intercept is zero within error. The calibration study results should be included in a method validation report. Access to information about calibration performance characteristics assists the user in implementing new methods and verifying that a laboratory's instrument performance is acceptable.

As mentioned above, the calibration approach stated in a method should be thoroughly tested and justified during an intralaboratory validation study, and tested according to the required/recommended calibration procedure during an interlaboratory validation study. For example, an intralaboratory study may include several calibration standards and replicate readings of each standard. If the calibration performance characteristics are deemed acceptable, the required/recommended calibration approach may include the use of only three calibration standards, single readings of each standard, and the calculation of an average response factor. In this case, the interlaboratory study would involve the use of an average response factor as opposed to the full calibration approach used during the intralaboratory study. In any case, the instrument calibration procedure used in the final EPA method should be thoroughly studied, justified, and documented in a method validation report.

4.3 Bias/Trueness

Bias refers to the overall magnitude of known systematic (determinate) errors associated with the use of an analytical method. The presence of systematic errors can only be determined by comparison of the average of many results with a reliable, accepted reference value. Method bias may be estimated by measuring materials whose composition is reasonably well known, such as reference materials, by comparing results to those from at least one alternate method or procedure, or by analyzing spiked materials.

The relationship between the terms bias and trueness is clearly described in ISO 5725-1 (*5*). Two excerpts from this document are as follows:

> The "trueness" of a measurement method is of interest when it is possible to conceive of a true value for the property being measured. Although, for some measurement methods, the true value cannot be known exactly, it may be possible to have an accepted reference value for the property being measured; for example, if suitable reference materials are available, or if the accepted reference value can be established by reference to another measurement method or by preparation

of a known sample. The trueness of the measurement method can be investigated by comparing the accepted reference value with the level of the results given by the measurement method. Trueness is normally expressed in terms of bias. Bias can arise, for example, in chemical analysis if the measurement method fails to extract all of an element, or if the presence of one element interferes with the determination of another.

The term bias has been in use for statistical matters for a very long time, but because it caused certain philosophical objections among members of some professions (such as medical and legal practitioners), the positive aspect has been emphasized by the invention of the term trueness.

It should be noted that all approaches for evaluating trueness have limitations. Examples of these limitations are as follows:

- The reference materials approach introduces the important uncertainty of matrix matching, and the reliability of a bias estimate depends upon the relationship between the composition of the reference material and the samples.

- When using the alternate method approach, the reliability of a bias estimate is dependent upon how much is known about the performance characteristics of the alternate method.

- The matrix spiking approach introduces uncertainty regarding the behavior of spiked materials, compared to materials containing native analyte. This may be particularly problematic for solid materials.

Rigorous evaluations of trueness should be included in EPA method validation studies. Minimally, trueness should be evaluated at the extremes of the quantitation range stated in the scope of a method, and at regulatory levels.

4.4 Precision

The general term "precision" is used to describe the magnitude of random (indeterminate) errors associated with the use of an analytical method. The sources of random error evaluated depend upon the range of conditions over which the data are collected. The following definitions are recommended (*7*).

- *Repeatability* (of results of measurements) - closeness of agreement between the results of successive measurements of the same measurand carried out under the same conditions of measurement. These conditions are called repeatability conditions. Repeatability conditions include the same measurement procedure, observer, measuring instrument (used under the same conditions), location, (and) repetition over a short period of time. Repeatability may be expressed quantitatively in terms of the dispersion characteristics of the results.

- *Reproducibility* (of results of measurements) - closeness of agreement between the results of measurements of the same measurand carried out under changed conditions of measurement. A valid statement of reproducibility requires specification of the conditions changed. The

changed conditions may include principle of measurement, method of measurement, observer, measuring instrument, reference standard, location, conditions of use, (and) time. Reproducibility may be expressed quantitatively in terms of the dispersion characteristics of the results.

Minimally, method repeatability and reproducibility should be evaluated at the extremes of the quantitation range stated in the scope of a method, and must be evaluated at regulatory levels. (The quantitation range must include regulatory levels; see section 4.5.) Common measures of dispersion are the standard deviation and relative standard deviation of repeated measurements. The repeatability and reproducibility conditions should be clearly stated so that the measures of dispersion can be properly interpreted and evaluated.

Precision and trueness evaluations, known as accuracy studies, are often performed concurrently. A variety of test materials may be used for these studies, and the effects of laboratory subsampling should be included when appropriate.

4.5 Quantitation Limits and Range

The term "quantitation range" is used to describe the span of analyte levels, as contained in a sample matrix, for which method performance has been tested, and data quality is deemed acceptable for its intended use. A quantitation range must include either a regulatory or other type of action level.

A lower limit and an upper limit bound a quantitation range. A quantitation range may be wider than an instrument's calibration range because of dilution or concentration steps performed during sample preparation. Dilution factors or concentration factors are used to relate the calibration range to the quantitation range. Analyte concentration will have an effect on most method performance characteristics, including trueness and precision. At a minimum, method trueness and precision (as described in sections 4.3 and 4.4) should be evaluated at the extremes of the quantitation range.

The lower limit of the quantitation range is commonly referred to as the "limit of quantitation" or LOQ. There are numerous definitions in the scientific literature for this limit, and no single definition is appropriate for all EPA chemical methods of analysis. Defined limits, such as those based on specifications for precision, should be verified by testing materials containing analyte at the claimed LOQ.

4.6 Detection Limit(s)

The term "detection limit" is used to describe the lowest analyte level that can be confidently identified. There are many specific definitions for this term, and it is used to describe the detection capabilities of detectors, instruments, and analytical methods. The term "detection limit" must be defined, and a description of how it was evaluated during method validation must be provided. Limits derived from mathematical definitions or statistical models must be verified by testing materials containing analyte at the claimed detection level.

5 Ruggedness Testing

Ruggedness refers to the extent to which an analytical method remains unaffected by minor variations in operating conditions. Ruggedness testing involves experimental designs for examining method performance when minor changes are made in operating or environmental conditions. The changes should reflect expected, reasonable variations that are likely to be encountered in different laboratories.

A commonly used, efficient approach for a ruggedness test is based on a fractional factorial experimental design[3] (*8, 9, 10*). This approach is more efficient than varying conditions one at a time. Various types of precision studies may also reveal the effects of some influential factors. Analyst observations are also important, and may reveal critical control points that should be documented in a method.

Ruggedness testing may be conducted at the end of a method development effort, or during an intralaboratory study. Regardless of timing, ruggedness testing should be completed before initiating an interlaboratory method validation study, and critical control points should be identified in the documented method.

---

3   *The fractional factorial experimental design referenced herein is not intended for method optimization purposes.*

6 Interlaboratory Studies

Interlaboratory studies determine whether an analytical method can be transferred for use in other laboratories and used for regulatory testing. A method that proved rugged for use in one laboratory may lose that characteristic when tried in several other laboratories. Therefore, an interlaboratory method validation study should be completed prior to publishing an EPA method for general use. The extent of the study will vary depending on available resources and the intended use and applicability of the method. These studies are resource intensive, and should not be initiated until a successful intralaboratory study has been completed and the issuing entity has demonstrated a pressing programmatic need for publishing a national method.

An interlaboratory method validation study is a practical testing of the written method on identical materials, usually derived from split samples, by a number of laboratories. The study is not intended to evaluate laboratories; it is intended to evaluate method reproducibility among laboratories. Deviations from the interlaboratory study protocol should be strongly discouraged, and any deviations that occur should be documented.

There are several excellent references for the design and conduct of interlaboratory method validation studies. Three such references are listed in section 10 of this document (*11, 12, 13*).

7 Method Validation Reports

A method validation report, suitable for placing in the public docket, should be prepared. The report should address the method validation topics outlined in this guidance document (summarized in Table 1), and (a) background information on method development, (b) a description of the actual method validation techniques, (c) any changes made to the method as a result of the validation studies, and (d) any recommendations for future work

Table 1
Method Validation Report Topic Areas

Topic	Explanation
Purpose	The purpose of the method is a clear and concise description of the measurement objectives and the intended use of the data.
Scope and Applicability	The method scope and applicability serve to define the range of method performance studies. The scope and applicability includes (a) the measurement process components validated, (b) the nature of the analytes and matrices studied, (c) the range of analyte levels for which the method is claimed to be suitable, (d) a description of any known limitations and any assumptions upon which a method is based, and (e) a description of how the method and analytical parameters chosen meet the data quality objectives for the specific application.
Selectivity	Selectivity is a performance characteristic that demonstrates the ability of the method to yield useful data for the analytes, analyte levels, and matrices defined within the scope of the method. Selectivity is especially important at action levels of concern, and must be demonstrated in the presence of species known or predicted to present analytical challenges to the method. Typically, selectivity is demonstrated by providing information that substantiates the identity of the analytes in the presence of expected matrix constituents.
Instrument Calibration	This performance characteristic is sometimes referred to as "instrument linearity." In a broader sense, it addresses instrument calibration and the nature of the resulting calibration model. During method validation, sufficient data should be generated to demonstrate and justify the method-recommended calibration approach.
Bias/Trueness	Trueness is a performance characteristic that addresses sources of known systematic error, and bias is a measure of trueness. The extent of bias in a measurement process typically varies with analyte, analyte level, and matrix. Therefore, trueness should be evaluated using samples that span the degree of analytical difficulty encompassed by the specified scope and applicability. Examples of approaches used to assess trueness include the use of reference materials, alternate methods or techniques, and matrix fortification.
Precision (Repeatability and Reproducibility)	Precision is a performance characteristic that reflects sources of random error in a measurement process. There are several different types of precision studies (i.e., repeatability and reproducibility studies). Methods designed for demonstrating compliance with regulatory requirements should be evaluated for both repeatability (within-lab) and reproducibility (among labs).
Quantitation Range and Limits	A quantitation range corresponds to the range of analyte concentrations (or other quantity) characterized for measurement accuracy (trueness and precision) during method validation. Analytical measurement uncertainty should be well characterized over a quantitation range.
Detection Limits	This general term describes the lowest level of analyte that can be identified with confidence. A method validation study typically involves a specific detection limit definition, and includes an explanation of how the defined detection limit was estimated. Detection limit estimates derived from mathematical definitions or statistics must be verified by analyzing samples containing analyte at the claimed detection level. This performance characteristic may not be important for all analytical applications, and its relevance depends on the method scope and applicability.
Ruggedness	Ruggedness refers to the capacity of an analytical method to remain unaffected by small variations in operating conditions. Ruggedness testing involves experimental designs for examining method performance when small changes are made in operating or environmental conditions. The changes should reflect expected, reasonable variations that are likely to be encountered in different laboratories.
Interlaboratory Study	Interlaboratory studies determine whether an analytical method can be transferred for use in other laboratories and used for regulatory testing. The design and results of the interlaboratory study should be included in the method validation summary report. Data from the interlaboratory study should be reported in tabular form. There should also be a discussion included describing the details of, and rationale for, any changes made to the method resulting from the interlaboratory study.

8 Peer Review

Before publication, EPA methods of analysis are either peer reviewed in accordance with the information provided in the most current version of the EPA Science Technology Policy Council Peer Review Handbook (*3*) or reviewed by a documented alternate procedure that ensures the robustness and technical viability of the method while considering stakeholder and expert input. The Peer Review Handbook provides Agency-wide guidance for the consistent implementation of peer review, and Program Offices have the flexibility to design peer reviews to suit their specific needs. The Handbook is the source of detailed information about which Agency work products are subject to peer review. In addition, there is information about selecting peer review mechanisms (internal and external), planning a peer review process, conducting a peer review, and preparing a peer review record. Each EPA office has a Peer Review Coordinator available for assistance.

9 Terms and Definitions

Terms and definitions were obtained from the following authoritative sources:

- American National Standards Specification and Guidelines for Quality Systems for Environmental Technology Programs, Annex A (2004) (ANSI/ASQ)
- International Vocabulary of Basic and General Terms in Metrology, 2nd Edition, International Organization for Standardization (1993) (VIM)
- Accuracy (trueness and precision) of measurement methods and results – Part 1: General principles and definitions, International Organization for Standardization, ISO 5725-1, 1994-12-15 (ISO 5725-1)

For each term and definition, the specific reference is indicated in parentheses. Numbers identifying VIM terms and definitions are also included. The numbering of the notes from ISO 5725-1 is the same as the numbering in the original ISO document.

accepted reference value: A value that serves as an agreed up reference for comparison, and which is derived as:

a) a theoretical or established value, based on scientific principles;
b) an assigned or certified value, based on experimental work of some national or international organization;
c) a consensus or certified value, based on collaborative experimental work under the auspices of a scientific or engineering group;
d) when a), b) and c) are not available, the expectation of the (measurable) quantity, i.e. the mean of a specified population of measurements. (ISO 5725-1)

accuracy: The closeness of agreement between a test result and the accepted reference value. (ISO 5725-1)

> NOTE 2   The term accuracy, when applied to a set of test results, involves a combination of random components and a common systematic error or bias component.

accuracy of measurement: closeness of the agreement between the result of a measurement and a true value of the measurand (VIM 3.5)

> NOTES
> 1   "Accuracy" is a qualitative concept.
> 2   The term **precision** should not be used for "accuracy".

bias: The difference between the expectation of the test results and an accepted reference value. (ISO 5725-1)

> NOTE 5 Bias is the total systematic error as contrasted to random error. There may be one or more systematic error components contributing to the bias. A larger systematic difference from the accepted reference value is reflected by a larger bias value.

calibration: set of operations that establishes, under specified conditions, the relationship between values of quantities indicated by a measuring instrument or measuring system, or values represented by a material measure or a reference material, and the corresponding values realized by standards (VIM 6.11)

>   NOTES
>   1   The result of a calibration permits either the assignment of values of measurands to the indications or the determination of corrections with respect to indications.
>   2   A calibration may also determine other metrological properties such as the effect of influence quantities.
>   3   The result of a calibration may be recorded in a document, sometimes called a **calibration certificate** or a **calibration report**.

measurand: particular quantity subject to measurement (VIM 2.6)

>   EXAMPLE   vapour pressure of a given sample of water at 20 °C.
>   NOTE   The specification of a measurand may require statements about quantities such as time, temperature and pressure.

measurement procedure: set of operations, described specifically, used in the performance of particular measurements according to a given method (VIM 2.5)

>   NOTE   A measurement procedure is usually recorded in a document that is sometimes itself called a "measurement procedure" (or a **measurement method**) and is usually in sufficient detail to enable an operator to carry out a measurement without additional information.

method: A body of procedures and techniques for performing an activity (e.g., sampling, chemical analysis, quantification) systematically presented in the order in which they are to be executed. (ANSI/ASQ)

precision: The closeness of agreement between independent test results obtained under stipulated conditions. (ISO 5725-1)

>   NOTES
>   9   Precision depends only on the distribution of random errors and does not relate to the true value or the specified value.
>   10   The measure of precision is usually expressed in terms of imprecision and computed as a standard deviation of the test results. Less precision is reflected by a larger standard deviation.
>   11   "Independent test results" means results obtained in a manner not influenced by any previous result on the same or similar test object. Quantitative measures of precision depend critically on the stipulated conditions. Repeatability and reproducibility conditions are particular sets of extreme conditions.

procedure: A specified way to carry out an activity or process. (ANSI/ASQ)

random error: result of a measurement minus the mean that would result from an infinite number

of measurements of the same measurand carried out under repeatability conditions (VIM 3.13)
> NOTES
> 1   Random error is equal to error minus systematic error.
> 2   Because only a finite number of measurements can be made, it is possible to determine only an estimate of random error.

repeatability (of results of measurements): closeness of the agreement between the results of successive measurements of the same measurand carried out under the same conditions of measurement (VIM 3.6)

> NOTES
> 1   These conditions are called **repeatability conditions.**
> 2   Repeatability conditions include:
>> - the same measurement procedure
>> - the same observer
>> - the same measuring instrument, used under the same conditions
>> - the same location
>> - repetition over a short period of time.
> 3   Repeatability may be expressed quantitatively in terms of the dispersion characteristics of the results.

reproducibility (of results of measurements): closeness of the agreement between the results of measurements of the same measurand carried out under changed conditions of measurement. (VIM 3.7)

> NOTES
> 1   A valid statement of reproducibility requires specification of the conditions changed.
> 2   The changed conditions may include:
>> - principal of measurement
>> - method of measurement
>> - observer
>> - measuring instrument
>> - reference standard
>> - location
>> - conditions of use
>> - time.
> 3   Reproducibility may be expressed quantitatively in terms of the dispersion characteristics of the results.
> 4   Results are here usually understood to be corrected results.

result of a measurement: value attributed to a measurand, obtained by measurement (VIM 3.1)

> NOTES
> 1   When a result is given, it should be made clear whether it refers to:
>> - the indication
>> - the uncorrected result
>> - the corrected result
> and whether several values are averaged.

2   A complete statement of the result of a measurement includes information about the uncertainty of measurement.

systematic error: mean that would result from an infinite number of measurements of the same measurand carried out under repeatability conditions minus a true value of the measurand (VIM 3.14)

NOTES
1   Systematic error is equal to error minus random error.
2   Like true value, systematic error and its causes cannot be completely known.
3   For a measuring instrument, see "bias" (VIM 5.25).

true value (of a quantity): value consistent with the definition of a particular quantity (VIM 1.19)

NOTES
1   This is a value that would be obtained by a perfect measurement.
2   True values are by nature indeterminate.
3   The indefinite article "a", rather than definite article "the", is used in conjunction with "true value" because there may be many values consistent with the definition of a given particular quantity.

trueness: The closeness of agreement between the average value obtained from a large series of test results and an accepted reference value. (ISO 5725-1)

NOTES
3   The measure of trueness is usually expressed in terms of bias.
4   Trueness has been referred to as "accuracy of the mean". This usage is not recommended.

uncertainty of measurement: parameter, associated with the result of a measurement, that characterizes the dispersion of the values that could reasonably be attributed to the measurand (VIM 3.9)

NOTES
1   The parameter may be, for example, a standard deviation (or a given multiple of it), or the half-width of an interval having a stated level of confidence.
2   Uncertainty of measurement comprises, in general, many components. Some of these components may be elevated from the statistical distribution of the results of series of measurements and can be characterized by experimental standard deviations. The other components, which can be characterized by standard deviations, are evaluated from assumed probability distributions based on experience or other information.
3   It is understood that the results of the measurement is the best estimate of the value of the measurand, and that all components of uncertainty, including those arising from systematic effects, such as components associated with corrections and reference standards, contribute to the dispersion.

10 References

1.   Eurachem. *The Fitness for Purpose of Analytical Methods: A Laboratory Guide to Method Validation and Related Topics*; LGC (Teddington) Ltd.: Middlesex, United Kingdom, 1998. (ISBN 0-94892-612-0)

2.   Thompson, M.; Ellison, S.L.R.; Wood, R. Harmonized guidelines for single-laboratory validation of methods of analysis, IUPAC Technical Report. *Pure Appl. Chem.* **2002**, *74 (5)*, 835-855.

3.   *EPA Science and Technology Policy Council, Peer Review Handbook,* 4th ed.; EPA/100/B-15/001; Science and Technology Policy Council, US Environmental Protection Agency, Washington, DC, October 2015.

4.   Kimothi, S.K. *The Uncertainty of Measurements;* ASQ Quality Press: Milwaukee, WI, 2002. (ISBN 0-87389-535-5)

5.   International Organization for Standardization (ISO). *Accuracy (trueness and precision) of measurement methods and results - Part 1: General principles and definitions;* ISO 5725-1: 1994(E), Geneva, Switzerland, 1994.

6.   Vessman, J.; Stefan, R.; van Standen, J.; Danzer, K.; Lindner, W.; Burns, D.; Fajgelj, A.; Muller, H. Selectivity in Analytical Chemistry, *Pure Appl. Chem.* **2001**, *73(8)*, 1381-1386.

7.   International Organization for Standardization (ISO). *International Vocabulary of Basic and General Terms in Metrology (VIM)*, 2nd ed., Geneva, Switzerland, 1993.

8.   ASTM E 1169-02. Standard Guide for Conducting Ruggedness Tests; *Annu. Book of ASTM Stand.* **2003**, *(14.02),* 318-324.

9.   Youden, W.J.; Steiner, E.H. *Statistical Manual of the AOAC;* AOAC: Arlington, VA, 1975. (Fifth printing 1987; ISBN 0-935584-15-3)

10.   Wernimont, G. *Use Of Statistics To Develop And Evaluate Analytical Methods;* AOAC: Arlington, VA, 1985. (Fourth printing 1993; ISBN 0-935584-31-5)

11.   AOAC International. *AOAC® Official Methods^SM Program Manual (OMA Program Manual, Part 6, Appendix D: Guidelines for Collaborative Study Procedures to Validate Characteristics of a Method of Analysis*; Gaithersburg, MD, 2002.

12.   Horwitz, W. Protocol for the design, conduct and interpretation of method performance studies: Revised 1994 (Technical Report), *Pure Appl. Chem.* **1995***, (67)*, 332-343.

13.   ASTM E 691-99. Standard Practice for Conducting an Interlaboratory Study to Determine the Precision of a Test Method, *Annu. Book of ASTM Stand.* **2003**, *(14.02),* 203-224.