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

Prepared for:

The EPA Forum on Environmental Measurements (FEM)

Prepared by:

The FEM Method Validation Team

FEM Document Number 2006-01
September 29, 2006

Principal Authors:

John G. Griggs, Ph.D.
U.S. EPA, National Air and Radiation
Environmental Laboratory
Montgomery, Alabama

Keith D. McCroan, Ph.D.
U.S. EPA, National Air and Radiation
Environmental Laboratory
Montgomery, Alabama

Contributors:

David E. McCurdy, Ph.D.
Environmental Management Support,
Inc.
Silver Spring, Maryland

Robert Litman, Ph.D.
Environmental Management Support, Inc.
Silver Spring, Maryland

Robert T. Shannon
Environmental Management Support, Inc.
Silver Spring, Maryland

Steven C. Wendelken, Ph.D.
U.S. EPA, Office of Groundwater and
Drinking Water, Technical Support Center
Cincinnati, Ohio

Mary F. Wisdom
U.S. EPA, National Air and Radiation
Environmental Laboratory
Montgomery, Alabama

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.

References within this document to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily imply its endorsement or recommendation by EPA. Neither EPA nor any of its employees, makes any warranty, expressed or implied, nor assumes any legal liability of responsibility for any third party's use, or the results of such use, of any information, apparatus, product, or process disclosed in this manual, nor represents that its use by such third party would not infringe on privately owned rights.

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 focus 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 radiochemical methods of analysis. It is the second in a series of Agency-wide method validation guidance documents and is based partly on the first document in this series, *Validation and Peer Review of U.S. Environmental Protection Agency Chemical Methods of Analysis*, which provided guidance for methods of chemical analysis (*16*). Subsequent guidance will be developed for other types of measurement methods (e.g., microbiology, biology, toxicology, 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. Because 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 radiochemical 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 radiochemical methods of analysis before they are published as EPA methods. The EPA Science Policy Council's *Peer Review Handbook* provides Agency-wide requirements and options for that process.

Contents

Table

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*). In addition, method validation for radiochemical methods is used to ensure that the uncertainty for each measurement can be properly estimated. 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 radiochemical 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 gathering scientific information. 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 radiochemical methods of analysis. This information is based on current, international approaches and guidelines for single- and multiple-laboratory method validation studies.

Before publication, all radiochemical methods of analysis also must be peer reviewed. The reader is referred to the most current version of the EPA Science Policy Council's *Peer Review Handbook* (*3*), an excellent source of information for both internal and external peer review processes.

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 radiochemical 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 radiochemical methods of analysis, as defined in Section 1.3 of this document, that are either to be published as serially numbered EPA methods, or published as or referenced in regulations.

1.3 Scope of Guidance

This guidance contains recommendations for validating new, rational, quantitative, radiochemical 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 radionuclides 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 radiochemical methods, which are used to estimate method-defined parameters. Examples of rational methods are "Method 903.1 Radium-226 in Drinking Water" and "Method 906.0 Tritium in Drinking Water" (*Prescribed Procedures for Measurement of Radioactivity in Drinking Water*, EPA 600/4-80-032, August 1980.)

- "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*).

- "Radiochemical methods of analysis, intended for use in analytical laboratories" means that the method is intended primarily for use in a laboratory to measure the specific decay emissions (e.g., alpha, beta, photon, etc.) from radionuclides or the radionuclide atoms if the half-life of the radionuclide is sufficiently long. Typically, radiochemical methods do not include sampling procedures but may have initial sample processing based on recommended sample preservation and collection practices. 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 8. 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 radiochemical 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 be revealed only 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 components of a method or measurement process requiring validation should be clearly described. Generally, a laboratory method includes all physical, chemical, and radiometric processes conducted at a laboratory in order to provide an analytical result (*ref. 5 chap. 6*). The processes for radiochemical methods may include initial sample preparation and dissolution, chemical separations, test-source preparation and mounting, nuclear counting, and analytical calculations. Method validation efforts should include an evaluation of all process components combined. Some radiochemical methods also may include procedures for sampling (e.g., methods for radon in air analysis and for volatile radioactive organic compounds in soils and other solid matrices), in which case the sampling procedures should be included in the validation tests. 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 and measurement quality objectives for the intended application (if applicable).

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, 6*). 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 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 true but not accurate, as the process is not precise.

Based on current trends in classifying method validation study components (*2*, *5*), the performance characteristics[2] that should be evaluated include, but are not limited to:

---

[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.*

- Method uncertainty
- Bias/Trueness
- Precision
- Quantification capability (minimum quantifiable concentration)
- Detection capability (minimum detectable concentration)
- Analyte concentration range
- Method specificity
- Method ruggedness (robustness)

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. The chemical state of a radionuclide may be changed through oxidation-reduction reactions within the matrix if not properly preserved. Some examples of these changes would be reduction to a volatile component, precipitation with other ionic components, adsorption onto container surfaces, polymerization, or chemical binding in a specific molecule. In the last case, if the analysis is based on the molecular form that the radionuclide is in, special precautions must address the chemical stability of that molecule and not necessarily the oxidation state of the radionuclide. Microorganisms have the potential to oxidize or reduce certain radionuclides, and can introduce by-products that will induce chemical changes. However, preservation to address microbial influences on chemical state is seldom implemented.

4.1 Method Uncertainty

Measurement uncertainty has been defined as a "parameter, associated with the result of a measurement that characterizes the dispersion of the values that could reasonably be attributed to the measurand" (*7, 8*). The *measurand* is the quantity being measured, which for radiochemical analyses is often the massic activity or volumic activity of a specified radionuclide, such as ^{137}Cs, or in some cases the combined activity of several radionuclides (e.g., total uranium). The uncertainty of a measurement result typically is expressed as an estimated standard deviation called the *standard uncertainty* or *combined standard uncertainty*. The *Guide to the Expression of Uncertainty in Measurement* (GUM) (*7*) and MARLAP (*5*) recommend that a combined standard uncertainty be calculated by propagating the standard uncertainties of the measured values used to calculate the result. The combined standard uncertainty may be multiplied by a *coverage factor* to obtain an *expanded uncertainty*, which describes a *coverage interval* about the result (similar to a confidence interval) that is believed to have a high probability of containing the true value of the measurand.

Measurement uncertainty as commonly defined is a property of each measurement result and varies with each measurement; so, it cannot be a method-performance characteristic. However, MARLAP (*5*) introduces the concept of "method uncertainty" when discussing the expected combined standard uncertainty of results produced by a radiochemical method at a specified analyte level. Method uncertainty is defined as the "predicted uncertainty of a result that would be measured if a method were applied to a hypothetical laboratory sample with a specified radionuclide activity or concentration." MARLAP recommends using the method uncertainty as a method performance

characteristic for the purpose of evaluating and selecting radiochemical methods for particular applications. When estimating the method uncertainty, the uncertainty of each input quantity (method parameter) that may contribute significantly to the total uncertainty is evaluated. For most radiological projects having regulatory limits for radioactive materials, a required method uncertainty at an action level (i.e., regulatory limit for the radionuclide) is considered essential to assure the quality of data used in statistical analyses and hypothesis testing.

4.2 Bias/Trueness

Bias refers to the overall magnitude of systematic errors associated with the use of an analytical method. ISO 5725-1 (*6*) defines bias as "[t]he difference between the expectation of the test results and an accepted reference value." The presence of systematic error, or bias, 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 (*6*). 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.

Bias is expressed typically as the absolute or relative deviation from a "true or known" analyte activity. Since it is a calculated estimate, a reported bias should be reported with an uncertainty. In radiochemical analyses, bias can arise if the measurement method:

- is improperly calibrated or calibrated with reference materials not traceable to the national standard,

- fails to quantitatively measure all of the radionuclide present,

- fails to correct properly for the chemical yield of the radionuclide or element,

- fails to correct properly for the contribution of instrument background or reagent activity to the sample result, or

- if the presence of another radionuclide interferes with the determination of the analyte.

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.

4.3 Precision

The general term "precision" is used to describe the magnitude of random 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 (*8*).

- *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 radiochemical method, 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. 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, sometimes called "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.

Both the standard deviation of repeated measurements at a given analyte level and the bias estimate should be compared to the method uncertainty at that analyte level. The standard deviation, and especially the repeatability standard deviation, should not be significantly larger than the method uncertainty. The standard deviation may in fact be somewhat smaller than the standard uncertainty, because uncertainty estimates account for the effects of both random error and systematic error (bias), whereas the repeatability standard deviation is generated by random error only. Any observed bias, if statistically significant, should be smaller than the method uncertainty.

4.4 Quantification Capability

For certain projects or programs where statistical analysis or data assessment of analytical results is needed, the quantification capability of a method is very important. MARLAP (*5*) expresses the quantification capability in terms of a method's minimum quantifiable concentration (MQC). A method's MQC is defined as the smallest concentration of an analyte in a sample that ensures a relative standard deviation of the measurement does not exceed a specified value, usually 10 %. Measurement results generated from a method meeting a MQC requirement ensures that data assessment can be performed on data of a defined quality.

The MQC is defined in terms of the relative standard deviation of measurement results, but MARLAP essentially equates this relative standard deviation with the expected relative standard uncertainty. The recommended approach for evaluating either the method uncertainty or the MQC requires that the laboratory be able to predict standard uncertainties for measurements at specified levels of activity. Such "uncertainty prediction" should be based on the same equations that are used to evaluate the uncertainty for each measurement result.

The same forethought that a laboratory gives to estimating a method's minimum detectable concentration (MDC) for an analyte should be given to the MQC. The laboratory must consider the standard uncertainties of all input quantities (counting uncertainty, detection efficiency, chemical yields, interferences, decay corrections, sample size, etc.) when calculating the MQC of a method.

4.5 Detection Capability

The term "detection limit" in radiochemistry is used to describe the lowest analyte level that can be reliably and consistently detected (i.e., distinguished from zero). In principle, a detection limit is a property of a measurement process. However, there are many specific definitions for the term, and it is often used to describe the detection capabilities of detectors, instruments, and analytical methods.

For radiochemical methods, the *a priori* minimum detectable concentration (MDC) is used to express detection capability. MARLAP (*9*) defines the MDC as "an estimate of the true concentration of analyte [radionuclide] required to give a specified high probability that the measured response will be greater than the critical value." (The radionuclide is considered to be "detected" when the response exceeds the critical value.) Both ISO 11929 (*17*) and MARLAP provide equations that may be used to calculate the MDC for a measurement process. During method validation, the estimated MDC must be verified by testing materials containing analyte at the claimed detection level.

The minimum detectable concentration is the most commonly used measure of detection capability in the field of radiochemistry. However, for some purposes the method detection limit (MDL) or the Safe Drinking Water Act "detection limit" (DL) as defined in 40 CFR Part 17 may be more appropriate.

4.6 Analyte Concentration Range

The term "analyte concentration range" of a method is used to describe the span of radionuclide activity levels, as contained in a sample matrix, for which method performance has been tested, and data quality is deemed acceptable for its intended use. An analyte concentration range must include either a regulatory limit or a defined action level, typically near the midpoint of the range of radionuclide activity level (concentration) for a project.

A lower limit and an upper limit bound an analyte concentration range. An analyte concentration range may extend from lower bound of zero radionuclide activity to an upper bound that is a multiple of a project action level or orders of magnitude greater than a project action level. Radionuclide activity or concentration may have an effect on most method performance characteristics, including trueness and precision. At a minimum, method trueness and precision (as described in sections 4.2 and 4.3) should be evaluated at the extremes of the analyte concentration range.

4.7 Method Specificity

Method specificity refers to the ability to correctly identify and quantify the radionuclide(s) of interest in the presence of other radionuclide interferences in the sample. Method specificity may be evaluated by testing appropriate matrix blanks and fortified (spiked) matrix blanks. Matrix blanks should contain the chemical species and inherent radionuclides (e.g., naturally occurring radionuclides), other than the radionuclide(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 the inherent radionuclide 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 specificity 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 specificity is typically expressed qualitatively and quantitatively. A method specificity statement would include descriptions of parameters such as:

- Known radionuclide and chemical interferences.
- Effects of the interfering substances on the measurement process.
- Measurement information that substantiates the identity of the analyte (e.g., half-life, or decay emission and energy).
- Effects of oxidation or molecular state of the radionuclide.
- Chemical processes that can remove interfering materials (e.g., ion exchange, solvent extraction).
- Summary of results from analysis of standards, reference materials, and matrix blanks.

4.8 Method Ruggedness

Ruggedness, or robustness, 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] (*10, 11, 12*). 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.

5   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

---

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

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. Four such references are listed in section 10 of this document (5, *13, 14, 15*). For radiochemical methods, the method validation concepts provided in MARLAP (*5*) are recommended.

6   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/measurement quality objectives for the specific application.
Method Specificity	Specificity is a performance characteristic that demonstrates the ability of the method to yield useful data for the radionuclide(s) of interest, radionuclide concentration levels, and matrices in the presence of possible interfering radionuclides or elements defined within the scope of the method. Specificity 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, specificity is demonstrated by providing information that substantiates the identity of the analytes in the presence of expected matrix constituents.
Method Uncertainty	The method uncertainty is a performance characteristic that addresses all recognized sources of random and systematic error. The method uncertainty, expressed as a standard deviation, describes the expected uncertainty of measurement that should be attained when analyzing samples at a specified analyte level. Evaluating the method uncertainty generally requires an existing equation or procedure for evaluating the uncertainty of each measurement made using the method.
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 may vary with radionuclide, radionuclide 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	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). When practical, methods designed for demonstrating compliance with regulatory requirements should be evaluated for both repeatability (within-lab) and reproducibility (among labs).
Quantification Capability	The quantification capability is expressed in terms of a method's minimum quantifiable concentration (MQC), defined as the smallest concentration of an analyte in a sample that ensures a relative standard deviation of the measurement does not exceed a specified value, usually 10 percent. Radiochemical method validation studies may involve the estimation of the minimum quantifiable concentration and includes an explanation of how the MQC was estimated. MQC estimates (derived from mathematical definitions or statistics) must be verified by analyzing samples containing radionuclide at the claimed MQC level.
Detection Capability	This general term describes the lowest level of analyte that can be reliably detected. Radiochemical method validation studies may involve the estimation of the minimum detectable concentration and includes an explanation of how the MDC was estimated. MDC estimates (derived from mathematical definitions or statistics) must be verified by analyzing samples containing radionuclide 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.
Analyte Concentration Range	The analyte concentration range corresponds to the range of radionuclide concentrations (or other quantity) characterized for measurement accuracy (trueness and precision) during method validation. Analytical measurement uncertainty should be well characterized over an analyte concentration range.
Method 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. When practical, method ruggedness should be evaluated during method validation.

Topic	Explanation
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.

7 Peer Review

Before publication, EPA methods of analysis are peer reviewed in accordance with the information provided in the most current version of the EPA Science Policy Council's *Peer Review Handbook* (*3*). 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.

8 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)
- Multi-Agency Radiological Laboratory Protocols Manual (MARLAP) (2004). EPA 402-B-04-001A

For each term and definition, the specific reference is indicated in parentheses. Numbers identifying definitions taken from the VIM or the GUM 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-upon 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: 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: 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**.

combined standard uncertainty: standard uncertainty of the result of a measurement when that result is obtained from the values of a number of other quantities, equal to the positive square root of a sum of terms, the terms being the variances or covariances of these other quantities weighted according to how the measurement result varies with changes in these quantities (GUM 2.3.4)

expanded uncertainty: quantity defining an interval about the result of a measurement that may be expected to encompass a large fraction of the distribution of values that could reasonably be attributed to the measurand (GUM 2.3.5)

measurand: particular quantity subject to measurement (VIM 2.6)

EXAMPLE   Vapor pressure of a given sample of water at 20 °C [or the massic activity of ^{137}Cs in a given sample of soil on a given date].

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.

measurement quality objectives: The analytical data requirements of the data quality objectives. The measurement quality objectives, either quantitative or qualitative, serve as measurement performance criteria or objectives for the radioanalytical process. (MARLAP, Chapter 3)

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

method uncertainty: predicted uncertainty of a result that would be measured if a method were applied to a hypothetical laboratory sample with a specified radionuclide activity or concentration (MARLAP, Chapter 3)

minimum detectable concentration: an estimate of the true concentration of analyte [radionuclide] required to give a specified high probability that the measured response will be greater than the critical value (MARLAP, Chapters 6 and 20)

minimum quantifiable concentration: smallest concentration of an analyte whose presence in a laboratory sample ensures the relative standard deviation of the result does not exceed a specified value, usually 10 %. (MARLAP, Chapters 6 and 20)

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

NOTES
1   Precision depends only on the distribution of random errors and does not relate to the true value or the specified value.
2   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.
3   "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
>   - analyst/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
- whether several values are averaged.

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

standard uncertainty: uncertainty of the result of a measurement expressed as a standard deviation (GUM 2.3.1)

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
1 The measure of trueness is usually expressed in terms of bias.
2 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, GUM 2.2.3)

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 evaluated from the statistical distribution of the results of series of measurements and can be characterized by experimental standard deviations. The other components, which also 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.

9   References

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

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

3. U.S. Environmental Protection Agency (EPA). 2006. *Peer Review Handbook, 3rd ed.* Science Policy Council, Office of the Science Advisor, Washington, DC. EPA/100/B-06/002. Available at http://www.epa.gov/peerreview.

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

5. *Multi-Agency Radiological Laboratory Protocols Manual* (MARLAP) (2004). Chapters 3 and 6. EPA 402-B-04-001A. Available at: http://www.epa.gov/radiation/marlap/manual.htm.

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

7. International Organization for Standardization (ISO). 1995. *Guide to the Expression of Uncertainty in Measurement.* ISO, Geneva, Switzerland.

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

9. *Multi-Agency Radiological Laboratory Protocols Manual* (MARLAP). 2004. Chapter 20. EPA 402-B-04-001A.

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

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

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

13. AOAC International. 2002. 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.

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

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

16. U.S. Environmental Protection Agency (EPA). 2005. *Validation and Peer Review of U.S. Environmental Protection Agency Chemical Methods of Analysis*. Available at: http://www.epa. gov/OSA/fem/pdfs/FEM_MV_doc_final_10-14-2005.pdf.

17. International Organization for Standardization (ISO). 2000. *Determination of the detection limit and decision threshold for ionizing radiation measurements*. ISO 11929. Geneva, Switzerland.