Check for updates

NIST Interagency Report
NIST IR 8589

Validation in Forensic Science

Guiding Principles for the Collection and Use of Validation Data

Henry Swofford
Steven Lund
Melissa Taylor
Barbara Guttman
Erica L. Romsos
Johannes Soons
Hariharan Iyer
Peter M. Vallone
John Paul Jones II
Vincent Desiderio
John M. Butler
Sandra Koch
Robert M. Thompson
Gregory Fiumara
Edward Sisco
Kelly Sauerwein
Allison Getz
Robert Ramotowski

This publication is available free of charge from:
https://doi.org/10.6028/NIST.IR.8589

NIST NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY
U.S. DEPARTMENT OF COMMERCE

NIST Interagency Report
NIST IR 8589

Validation in Forensic Science

Guiding Principles for the Collection and Use of Validation Data

Henry Swofford,[1] Steven Lund,[2] Melissa Taylor,[1] Barbara Guttman,[2] Erica L. Romsos,[3] Johannes Soons,[4] Hariharan Iyer,[2] Peter M. Vallone,[3] John Paul Jones II,[1] Vincent Desiderio,[1] John M. Butler,[1] Sandra Koch,[1] Robert M. Thompson,[1*] Gregory Fiumara,[2] Edward Sisco,[3] Kelly Sauerwein,[1] Allison Getz,[1] Robert Ramotowski[1*]

[1] Laboratory Programs
[2] Information Technology Laboratory
[3] Material Measurement Laboratory
[4] Physical Measurement Laboratory

*Former NIST employee; all work for this publication was done while at NIST.

This publication is available free of charge from:
https://doi.org/10.6028/NIST.IR.8589

December 2025

U.S. Department of Commerce
Howard Lutnick, Secretary

National Institute of Standards and Technology
Craig Burkhardt, Acting Under Secretary of Commerce for Standards and Technology and Acting NIST Director

NIST IR 8589
December 2025

Certain equipment, instruments, software, or materials, commercial or non-commercial, are identified in this paper in order to specify the experimental procedure adequately. Such identification does not imply recommendation or endorsement of any product or service by NIST, nor does it imply that the materials or equipment identified are necessarily the best available for the purpose.

Information in this publication is not intended to imply criticism of any past, pending, or future legal proceedings involving forensic science or evidence based on forensic science.

NIST Technical Series Policies
Copyright, Use, and Licensing Statements
NIST Technical Series Publication Identifier Syntax

Publication History
Approved by the NIST Editorial Review Board on 2025-09-23

How to Cite this NIST Technical Series Publication
Henry Swofford, Steven Lund, Melissa Taylor, Barbara Guttman, Erica L. Romsos, Johannes Soons, Hariharan Iyer, Peter M. Vallone, John Paul Jones II, Vincent Desiderio, John M. Butler, Sandra Koch, Robert M. Thompson, Gregory Fiumara, Edward Sisco, Kelly Sauerwein, Allison Getz, Robert Ramotowski (2025) Validation in Forensic Science: Guiding Principles for the Collection and Use of Validation Data. (National Institute of Standards and Technology, Gaithersburg, MD), NIST Interagency Report (IR) NIST IR 8589. https://doi.org/10.6028/NIST.IR.8589

Author ORCID iDs
Henry Swofford: 0009-0003-3114-5958
Steven Lund: 0000-0002-5638-3309
Melissa Taylor: 0000-0002-4204-3864
Barbara Guttman: 0000-0001-9706-389X
Erica L. Romsos: 0000-0001-5738-2968
Johannes Soons: 0000-0002-9485-1636
Hariharan Iyer: 0000-0001-5198-9945
Peter M. Vallone: 0000-0002-8019-6204
John Paul Jones II: 0000-0002-3258-0770
Vincent Desiderio: 0009-0003-9598-2632
John M. Butler: 0000-0001-6472-9157
Sandra Koch: 0000-0001-5872-2266
Robert M. Thompson: 0000-0003-2578-8845
Gregory Fiumara: 0000-0002-4853-8315
Edward Sisco: 0000-0003-0252-1910
Kelly Sauerwein: 0000-0001-9855-3030
Allison Getz: 0000-0002-0648-9463
Robert Ramotowski: 0000-0002-4823-4506

Contact Information
Henry.Swofford@nist.gov

Abstract

This document provides guidance to the forensic science community on method validation. Building on the definition of validation in ISO/IEC 17025, it explores the specific components of validation within a forensic science context and introduces five principles for guiding the collection and use of validation data. These guiding principles are intended to improve how validation studies are conducted and promote collaborative, community-based approaches to validation testing, data sharing, and performance characterizations, leading to more robust validation practices among forensic science laboratories. Additionally, this document serves as a framework for other resources and materials to build upon, including the development of forensic science-specific documentary standards and guidelines; standardized templates for validation plans, reports, and data; and software tools and repositories for analyzing, interpreting, and sharing validation data.

Keywords

Forensic Science; Performance; Guiding Principles; Quality Assurance; Reliability; Standards; Validation; Verification.

Table of Contents

Preface

This document was developed by an interdisciplinary group of researchers at the National Institute of Standards and Technology (NIST) to support the forensic science community in strengthening the robustness, transparency, and reliability of validation practices. In developing this document, NIST subject matter experts participated in workshops and focus groups that brought together forensic science practitioners, researchers, legal experts, and statisticians. These engagements, along with a review of relevant literature and standards, guided the identification of five guiding principles for validation. The guiding principles emphasize the dynamic nature of validation, including the documentation of validation data, characterization of performance, and communication of results to improve confidence in forensic methods. This document is not intended to provide a step-by-step guide for *how* to conduct validation for any particular discipline or method (e.g., defining requirements, characterizing performance, establishing decision criteria). Rather, this document expands on the definition and specifications for validation as outlined by ISO/IEC 17025 *General requirements for the competence of testing and calibration laboratories* and the ISO 21043 *Forensic sciences* series, providing a general framework for the development of discipline- and method-specific standards, tools, templates, and resources that can help the forensic science community effectively approach validation and support collaborative and community-based approaches to validation. References provided in this document are intended to provide supplemental information and are not exhaustive.

1. Introduction

Validation is the process of assessing whether a method meets specified requirements for an intended use and is a prerequisite for any method used in scientific or technical practice, including forensic science [ISO/IEC 17025:2017; ISO 21043-1:2025]. Results of forensic analyses are often relied upon by law enforcement officers, attorneys, and triers of fact (i.e., courts) to make important decisions affecting life and liberty. Providing information about the validation of a forensic science method is fundamental to trusting scientific analyses and criminal justice outcomes [NRC 2009, pp. 111-116].

International standards, including ISO/IEC 17025 *General requirements for the competence of testing and calibration laboratories* and the ISO 21043 *Forensic sciences* series, have emphasized the importance of methods supported by validation studies [ISO/IEC 17025:2017; ISO 21043-3:2025; ISO 21043-4:2025; ISO 21043-5:2025]. ISO/IEC 17025 is written to be industry- and sector-agnostic and generally applicable across different types of testing and calibration laboratories. While not specific to forensic science, it is the prevailing standard to which many forensic science laboratories conform and under which they are accredited [Swofford 2024, pp. 109-110]. The ISO 21043 series builds upon ISO/IEC 17025 by extending those specifications in a context that is specific to forensic science [Berger 2025]. While ISO/IEC 17025 and the ISO 21043 series provide a general framework for laboratories to demonstrate their competence, they do not guarantee that laboratories are generating valid results, nor do they provide detailed guidance on conducting validation studies.

Practitioners recognize the importance of validation but often struggle to apply it in practice. Designing and conducting robust validation studies can be complex and costly. Forensic science faces additional hurdles as evidentiary materials submitted for analysis may be limited in quantity and subject to contamination, degradation, and damage due to uncontrolled environmental conditions and other factors relating to evidence detection, collection, and handling. Accounting for a wide variety of **evidentiary circumstances**, such as the type, quality, or condition of materials, can be overwhelming and is often impractical in a single study. Furthermore, resources are often limited, and many laboratories are left to conduct validation studies in isolation [Wickenheiser & Farrell 2020].

This document is intended to help the forensic science community better understand how validation can be effectively approached. It expands on the definition of validation provided by ISO/IEC 17025, discusses the components of validation in the context of forensic science, and identifies five principles to guide validation activities in practice. It also encourages collaborative and community-based approaches to validation testing, including sharing plans, materials, methods, and data to: (1) reduce resource burdens on individual laboratories performing validation testing, (2) allow laboratories to make validation decisions more quickly, and (3) enable laboratories to convey information about the accuracy and reliability of their methods with greater confidence.

By expanding on the definition of validation provided by ISO/IEC 17025 and outlining principles guiding validation activities in practice, this document provides a framework for other resources and materials to build upon, including the development of:

- discipline- and method-specific documentary standards, guidelines, and best practice recommendations for how to conduct validation (e.g., defining requirements, characterizing performance, establishing decision-making criteria);
- standardized templates for validation plans, reports, and data; and
- software tools and repositories for analyzing, interpreting, and sharing validation materials and data.

Figure 1 illustrates how this document fits within the broader realm of guidance and resources for validation.

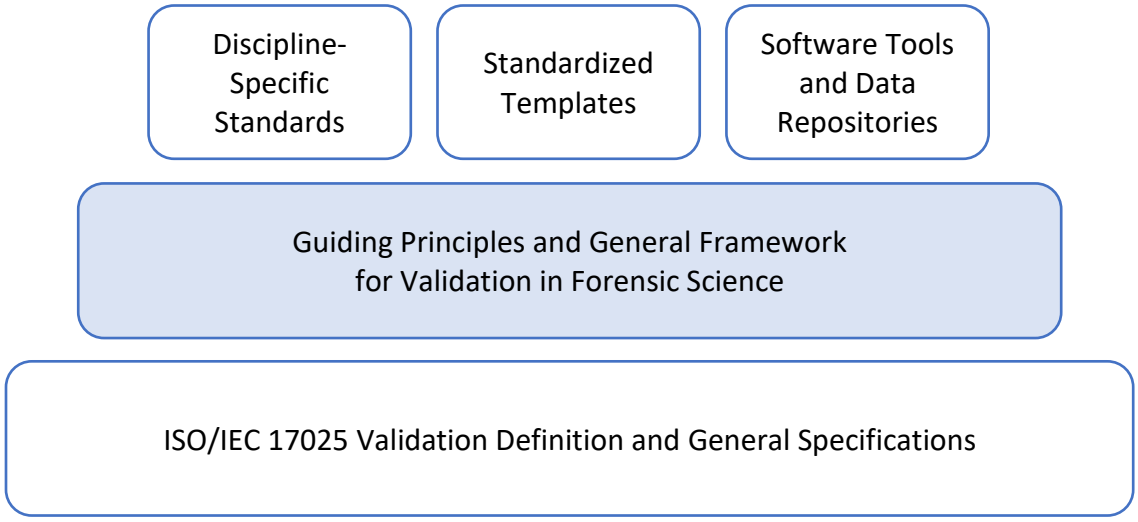

Figure 1: Diagram illustrating the purpose of this document (shaded portion) within the broader landscape of documentary standards and guidelines for forensic science.

2. Glossary

For purposes of this document, the following terms and definitions apply. Terms included in the Glossary appear in bold on their first use in the text.

Circumstance of use
A broad term that refers to any condition that may impact the performance of a method and includes evidentiary circumstances and implementation conditions.

Deterministic method
A method for which the output is certain and predictable by its initial state and input parameters—the same output is always produced for a given input.

Evidentiary circumstance
A broad term that refers to the type, quality, or condition of materials subject to forensic analysis, and includes physical items and digital artifacts.

Implementation conditions
A broad term that refers to the circumstances under which a method is applied and includes personnel (e.g., training, competency), facilities and environment (e.g., temperature, humidity, contamination), and equipment (e.g., function, calibration).

Performance characteristic
A parameter that describes an attribute of the method's effectiveness (e.g., accuracy, precision, discriminability, repeatability, reproducibility), expressed in terms of both the performance measurement and its associated uncertainty.

Performance characterization
A description of the performance of the method, to include performance measurements and their associated uncertainties, for all performance characteristics tested.

Performance measurement
A quantifiable expression of a performance characteristic.

Uncertainty
The degree of variability or confidence in the accuracy of a performance measurement.

Validation data
The individual test inputs and outputs of the method and associated metadata relevant to the testing and evaluation.

3. Validation Elements and Guiding Principles

Building on the definition of validation provided by ISO/IEC 17025 (i.e., provision of objective evidence that a given method fulfils specified requirements for an intended use), there are three elements fundamental to the process of validation, each of which is described in greater detail in Appendix A:

(1) *Defined scope and requirements* for an intended use

(2) *Characterization of method performance* within the bounds of a defined scope

(3) *Decision* whether the method is acceptable for use

Additionally, ISO/IEC 17025 specifies that laboratories must retain records relating to validation and that results be reported with all the information necessary for proper interpretation [ISO/IEC 17025:2017, pp. 11-12, 14].

The cornerstone of *validation* is ***validation data***. How the data is generated, documented, and conveyed all impact the **performance characterization** and ultimately how the performance of a method is assessed (see figure 2 for an infographic illustrating the definition and components of a performance characterization). These factors inform whether a method is acceptable for use and the interpretation of the results [ILAC-G19:2022, pp. 15-17].

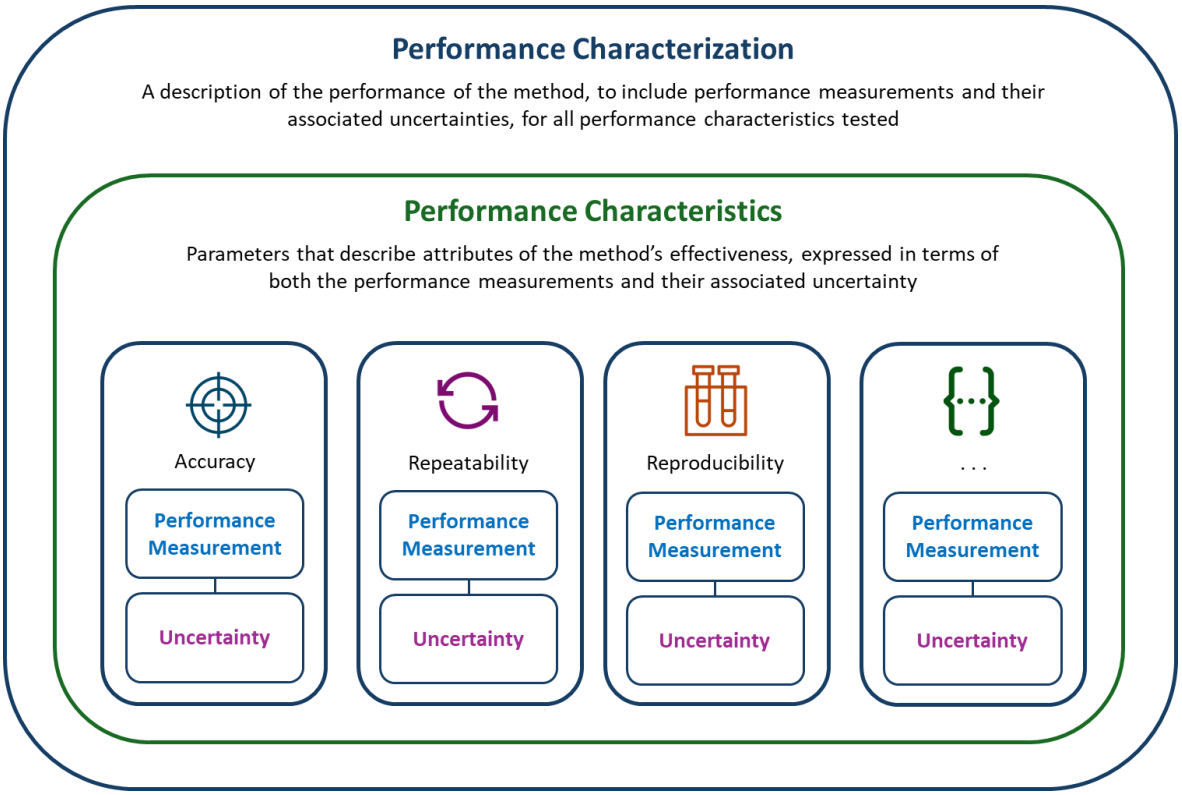

Figure 2: Components of a performance characterization. In this example, accuracy, repeatability, and reproducibility have been deemed as relevant characteristics to describe the effectiveness of the method. The bracketed ellipsis refers to additional performance characteristics that may be deemed relevant.

Engagement with forensic science practitioners, researchers, legal experts, statistical experts, and NIST subject matter experts, along with a review of relevant literature and standards, led to the development of five principles for guiding validation activities (Box 1). These principles are intended to provide an interdisciplinary framework for conducting validation activities to improve the robustness, transparency, and access to data supporting validation decisions. Although these principles are industry- and sector-agnostic, they are discussed in the context of forensic science.

Box 1: Guiding Principles for Validation

The guiding principles outlined below provide an interdisciplinary framework for conducting validation activities.

1. Validation is an ongoing process, and each element of validation—scope and requirements, performance characterization, and validation decisions—should be routinely reassessed as data accumulates and evidentiary circumstances evolve.
2. Validation data should be collected and documented in a manner that allows the user to distinguish among different circumstances of intended use.
3. Validation data should cover the range of circumstances of intended use.
4. Performance characterizations should be comprehensive, including characteristics relevant to describing the method's effectiveness and uncertainties associated with measurements of those characteristics.
5. The suitability of a method should be assessed within the context of each analysis, and performance characterizations should be conveyed in parallel with the result of the analysis, including the underlying data and any limitations that may affect the interpretation of the result.

3.1. Guiding Principle #1

Validation is an ongoing process, and each element of validation—scope and requirements, performance characterization, and validation decisions—should be routinely reassessed as data accumulates and evidentiary circumstances evolve.

The elements of validation are interrelated, and validation activities are dynamic. The scope, requirements, performance characterizations, and decisions related to acceptable use can change; therefore, performance should be monitored and validation elements should be revisited periodically as more data accumulates [Eurachem 2025].

The initial scope and requirements inform the approach for characterizing method performance among the different potential **circumstances of use**. The results of the performance characterization, in turn, define those circumstances for which the method is acceptable [ANZPAA 2025, pp. 13]. Additional testing may improve the performance characterization(s) or inform refinements to the scope. Figure 3 illustrates the process of validation and the interrelatedness among the different elements.

Even within the bounds of a scope of use, evidentiary circumstances can vary and evolve. This is especially important in rapidly changing fields (e.g., emergence of 3D printed guns, novel psychoactive substances, digital evidence) or for methods that rely on software or web-based services that are routinely updated.

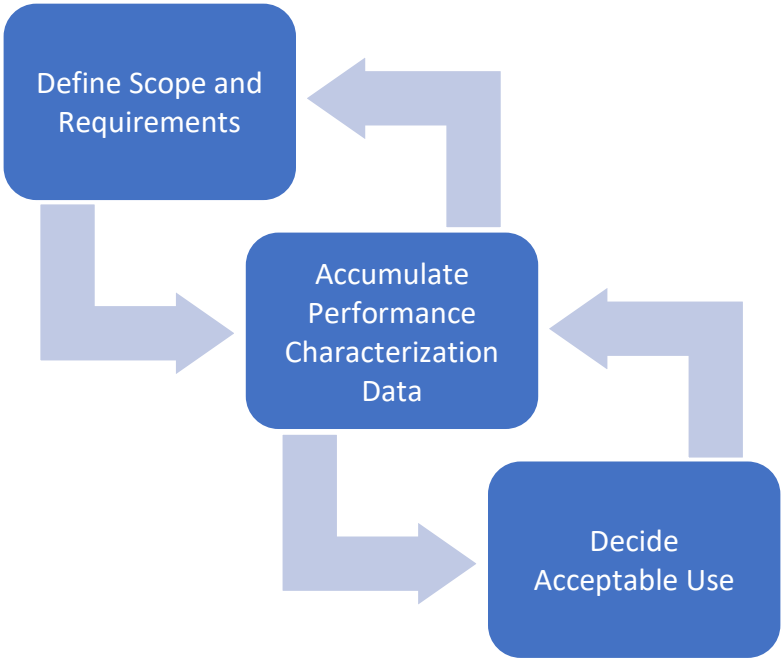

Figure 3: The validation process, reflecting the interrelatedness among the different validation elements (see Appendix A for further details).

Key Takeaway 3.1:
Validation is a dynamic, iterative process involving interrelated components.

3.2. Guiding Principle #2

Validation data should be collected and documented in a manner that allows the user to distinguish among different circumstances of intended use.

Validation data and related performance characterizations serve as an important basis for deciding whether a method is acceptable for use. They also provide the information necessary to properly interpret the output of a method (i.e., the extent to which a result can be relied upon). The validation data should be structured in a way that allows the user to distinguish among individual circumstances of use (i.e., evidentiary circumstances and factors relating to the particular implementation of the method encountered in casework) [ANZPAA 2025, p. 25]. This can be accomplished by documenting the output from each test used to assess the method, along with descriptors of the circumstances represented by each test. Table 1 provides an example format for capturing validation data.

Structuring the validation data in this way is important because **performance measurements** and their associated uncertainties might vary when a method is applied to different evidentiary circumstances (e.g., quantity, quality, condition of material or analyte available) or under different **implementation conditions** (e.g., personnel, facilities and environmental conditions, equipment). Examinations involving partial, degraded, or low-quality materials may yield worse performance when compared to those involving high-quality materials [OSAC 2020, pp. 7-8]. Aggregate summaries of a method's performance across varying circumstances of use can be misleading with respect to the performance of the method when used in a particular circumstance.

Structuring the validation data in this way is also useful when determining the appropriateness of using a method in a particular circumstance. Performance characterizations should be provided in a way that is relevant to the specific conditions of each analysis (see Guiding Principle #5). For example, subsets of the validation data could be labeled as originating from testing conditions that are more difficult, equally difficult, or less difficult than the evidentiary circumstances for the analysis within the case at hand, with performance characterizations offered for each subset. Regardless of how data are labeled or aggregated when evaluating performance, the individual outputs produced by the method during testing, along with descriptors of the circumstances represented by each test, should be documented such that the underlying data can be considered from multiple perspectives [ANZPAA 2025, p. 28].

Validation Data	Type of material	Quality of material	Condition of material	. . .	Method Outputs
Test Input #1					
Test Input #2					
Test Input #3					
Test Input #4					
Test Input #5					
. . .					

Table 1: Example format for documenting validation data (including metadata) that distinguishes among the individual circumstances for which the method is being tested. Each output produced by the method should be documented along with descriptors of the circumstances represented by each test. The ellipsis indicates a continuation of descriptors, as applicable.

Key Takeaway 3.2:

Relying on performance characterizations that combine results across circumstances of use—where performance is likely to vary—can be misleading about the performance of a method when used in a particular circumstance.

3.3. Guiding Principle #3

Validation data should cover the range of circumstances of intended use.

Performance characterizations should be based on validation data that (1) covers the range of circumstances expected to be encountered in casework [ILAC-G19:2022, pp. 15-17; ANZPAA 2025, p. 21] and (2) reflects the performance of a method as implemented within a particular laboratory. It can be challenging to test the performance of a method against *all* possible types, qualities, and conditions of materials that forensic science laboratories may encounter. However, the burden of such comprehensive testing does not necessarily need to be borne by each individual laboratory [Wickenheiser & Farrell 2020].

To ease this burden, the performance characterization can be based on data from the laboratory implementing the method[1] as well as data shared by other sources, such as researchers, developers, and other laboratories, when those data are applicable to the specific implementation[2] [Eurachem 2019a; Eurachem 2019b; Eurachem 2025; FSR 2024; FSR 2025; ANZPAA 2025, p. 7]. The extent to which data from other sources is applicable to a particular laboratory's performance characterization depends on the extent to which the outputs are predictive of each other. This is typically assessed by testing the reproducibility of outputs among different implementations of the method using the same or similar samples, such as through well-designed interlaboratory comparison studies. For deterministic methods, such as computer algorithms with fixed logic, this can be achieved by referencing well-established theories and principles [Lyle et al. 2022, p. 43].

Sharing validation data helps practitioners, researchers, and developers to evaluate the extent to which existing data covers the range of evidentiary circumstances, implementation conditions, or other factors encountered by forensic laboratories that can impact performance. Sharing data also enables evaluation of the robustness of a method and how existing data can inform performance characterizations by considering the extent to which:

(1) differences in circumstances of use impact performance;

(2) results from different methods (or individual components of the methods) are comparable;

(3) results from one laboratory are predictive of the results of the same analysis from another laboratory; and

(4) aspects of data format, quality, provenance, and integrity are acceptable for use.

[1] Validation data obtained from within the laboratory implementing the method might also be referred to as "internal validation data" or "single-laboratory validation data."
[2] Validation data obtained from other sources might also be referred to as "developmental validation data" or "multi-laboratory validation data."

Box 2 provides guidance that can increase the likelihood of validation data being used by laboratories across the community.

Box 2: Guidance for strengthening validation through community collaboration

Laboratories can increase the breadth of validation data available and the usefulness of performance characterizations across the community if they:

(a) Establish policies and procedures to allow for the ethical collection, sharing, and use of data and materials, such as de-identified biological material or biometric information, for use in validation.

(b) Adopt a common set of method procedures so that the same methods are applied across laboratories.

(c) Document validation data using standardized templates so that the data are in a consistent format and data structure that can be readily compared.

(d) Share data used to support validation decisions, including but not limited to:

i. validation plans and method procedures;

ii. materials tested, including applicable informed consent and privacy protections;

iii. validation data, including descriptions of the evidentiary circumstances and implementation conditions represented by each test; and

iv. description of the performance of the method, to include all performance measurements and their associated uncertainties (i.e., performance characterization).

Key Takeaway 3.3:

Approaching validation as a community-based effort reduces the burden on individual laboratories and increases the information known about a method's performance.

3.4. Guiding Principle #4

Performance characterizations should be comprehensive, including characteristics relevant to describing the method's effectiveness and uncertainties associated with measurements of those characteristics.

For performance characterizations to be useful, the **performance characteristics** measured, such as accuracy, precision, discriminability, repeatability, and reproducibility, must be sufficient to permit a comprehensive assessment of performance. This entails identifying the performance characteristics relevant to describing the effectiveness of the method within the context of its intended application and assessing the uncertainties associated with the measurements of those performance characteristics.

Choosing the relevant performance characteristics depends on the intended use of the method, the environment in which it will be applied, and the requirements that need to be met [ANZPAA 2025, p.17]. In forensic science, error rates (e.g., false positive rates and false negative rates) are often relied upon as standalone performance characteristics. While error rates are informative, describing the effectiveness of a method using error rates alone is an oversimplification when the method can produce more than two outcomes [Swofford et al. 2024]. This is often the situation when using categorical interpretation scales, where an inconclusive result is possible in addition to inclusionary or exclusionary results, or when using numeric scales, such as a likelihood ratio. In these situations, relying solely on error rates as the basis for a performance characterization can be incomplete and misleading.[3] Similarly, errors stemming from systematic issues, such as software bugs, can be informative about the technical limitations of a software application, but are not well described in terms of error rates. For example, in computer software that relies on deterministic code, the same result (correct or incorrect) will occur every time for a given input [Lyle et al. 2022, p. 43]. Careful consideration should be given to ensure that the measured performance characteristics convey the information needed to assess the suitability of the method and interpret the results produced.

Performance measurements are *estimated* values of performance characteristics that carry some degree of **uncertainty**, which can vary based on the amount of data available from testing, among other factors [ANZPAA 2025, pp. 26-27]. Uncertainty is inherent in any estimated value and reflects the degree of variability or confidence in the accuracy of a measured outcome due to factors such as limited data, measurement errors, model

[3] A similar example could be constructed to show that the alternative metrics of sensitivity (true positive rate) and specificity (true negative rate) also do not adequately characterize method performance. Such examples illustrate the perils of trying to summarize the performance of a method with a non-binary range of conclusions with the same number of parameters as a method with a binary range of conclusions. Two additional independent parameters or rates are required to fully characterize method performance for each element added to a binary conclusion scale.

assumptions, or inherent randomness. Performance measurements without their associated uncertainties cannot be meaningfully interpreted [Possolo 2015].

There is no general rule for a minimum or maximum amount of data needed to characterize method performance for a particular evidentiary circumstance; more validation data is often informative but not always necessary or practical. Performance measurements based on limited amounts of validation data tend to have greater uncertainty, whereas measurements based on more extensive validation data often have less uncertainty. Methods which have measurements that reflect poor performance (e.g., low accuracy, repeatability, reproducibility) with limited validation data are unlikely to reflect good performance when additional validation data is collected. The choice of how much validation data to collect for a given evidentiary circumstance depends on the requirements that need to be met relating to the performance characterization (e.g., to demonstrate that the lower bound for a performance measurement at a given confidence level exceeds a specified minimum threshold) [ANZPAA 2025, pp. 21-22]. Figure 4 illustrates a performance measurement with and without its associated uncertainty intervals.

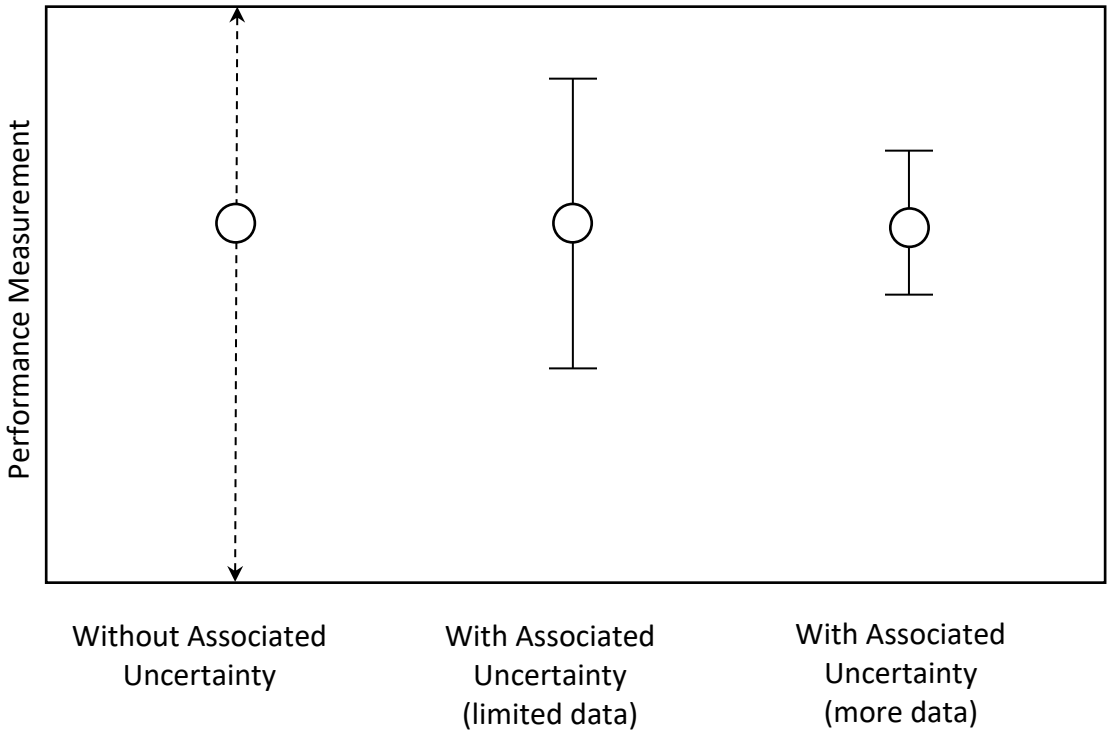

Figure 4: Illustration of the importance of ensuring performance measurements include their associated uncertainty. The image reflects three identical performance measurements without (left) and with (middle and right) their associated uncertainties. The circle reflects the estimated value of the performance characteristic. The vertical line represents the degree of variability or confidence in the accuracy of the estimated value of the performance characteristic due to factors such as limited data, measurement errors, or inherent randomness in the method. It may represent a confidence interval or posterior credible interval, for example, and what the error bars represent must be stated. Performance measurements without an associated uncertainty (left) cannot be meaningfully interpreted as the degree of variability is not defined. Performance measurements based on limited amounts of validation data tend to have greater uncertainty (middle) than those based on more extensive amounts of validation data, which tend to have less uncertainty (right).

Key Takeaway 3.4:

Performance characterizations are incomplete without all relevant performance characteristics, performance measurements of those characteristics, and an expression of their uncertainty.

3.5. Guiding Principle #5

The suitability of a method should be assessed within the context of each analysis, and performance characterizations should be conveyed in parallel with the result of the analysis, including the underlying data and any limitations that may affect the interpretation of the result.

For a user to understand how to interpret the results of a method, performance characterizations should be made available [ANZPAA 2025, p. 30]. These performance characterizations should reflect the performance of the method both (1) in general across the range of circumstances of use and (2) at or near the specific circumstance of use for a particular analysis. The underlying data on which the characterizations are based, as reflected in Table 1, and any limitations that may affect the interpretation of the result (e.g., uncertainties associated with the performance measurements due to limited validation data relevant to the specific circumstance) should also be made available [ISO/IEC 17025:2017, p. 14].

Validation decisions can vary among laboratories depending on their requirements, ability to perform the method, case-specific evidentiary circumstances, or needs of the user(s). While the decision that a method is acceptable for use is influenced by the performance characterization of the method, the assertion that a method has been deemed acceptable does not convey any of the information needed for others to assess that decision or properly interpret the results in a particular circumstance [Swofford et al. 2024].

Conveying relevant information about the performance of a method is important so that users of the result can understand the extent to which the output of a method can be relied upon. Without information reflecting the performance of the method—both in general and specific to a particular analysis—there is a risk that the weight assigned to the method's result could be under- or over-valued, impacting judicial outcomes [Swofford et al. 2024].

Figure 5 illustrates the steps that laboratory analysts should consider for each analysis in which a method is used. First, the evidentiary circumstance(s) for a particular analysis need to be described. Second, the performance of the method needs to be evaluated against the performance requirements using validation data that is relevant to the evidentiary circumstances in the case at hand. Third, the laboratory analyst must determine whether the method is acceptable for use in that circumstance and consider any limitations that may affect the interpretation of the result. Fourth, the performance characterizations (including the underlying data) and limitations affecting the interpretation of the result should be communicated alongside the result of the analysis.

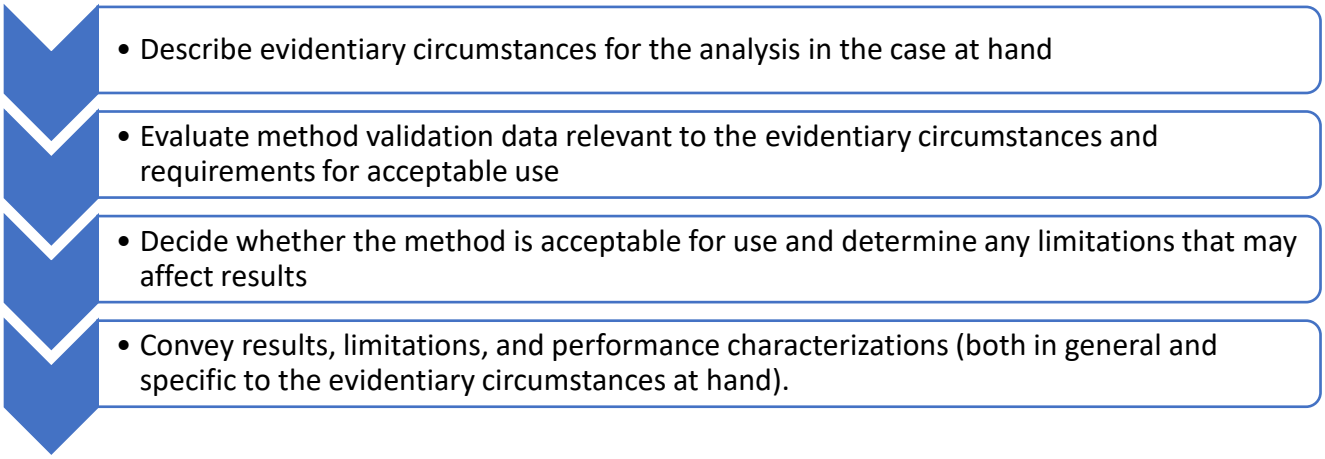

Figure 5: Workflow illustrating considerations made by the laboratory analyst for each analysis in which a method is used.

Key Takeaway 3.5:

Information about the performance of a method—both in general and relevant to the individual circumstance(s) for which the method was used—is important for validation decisions to be assessed and the result of the method to be properly interpreted.

4. Conclusion

Validation is a critical process in forensic science, enabling users to evaluate whether methods are (or are not) acceptable for their intended purpose and capable of routinely producing accurate and reliable results. Given the importance of forensic evidence to the criminal justice system, robust validation practices are essential for ensuring the system operates effectively. International standards such as ISO/IEC 17025, provide a general framework for demonstrating competence, yet forensic laboratories face challenges in conducting validation studies because of resource constraints and the wide variation and complexity of evidentiary materials encountered in casework. This document outlines five principles for guiding the collection and use of validation data, emphasizing the need for continuous assessment, relevant performance characteristics, and data sharing. By adopting these principles and fostering community-based approaches to validation, forensic laboratories can deploy new methods more quickly, with greater confidence and lower costs, strengthening the quality of information relied upon by law enforcement and courts.

References

References provided in this document are intended to provide supplemental information and are not exhaustive. References noted with an asterisk are cited within this document.

[1] *[ANZPAA 2025] Australia New Zealand Policing Advisory Agency (ANZPAA) National Institute of Forensic Science (NIFS) Validation Working Group (2025) Guideline for the Validation of Forensic Science Methods. Available at https://www.anzpaa.org.au/ArticleDocuments/357/ANZPAA%20NIFS%20Guideline%20for%20the%20Validation%20of%20Forensic%20Science%20Methods.PDF.aspx.

[2] *Berger, C.E.H. (2025) Finally a Really Forensic Worldwide Standard: ISO 21043 Forensic sciences, Part 4, Interpretation. *Forensic Science International: Synergy* 10:100589. Available at https://doi.org/10.1016/j.fsisyn.2025.100589.

[3] BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, and OIML. Evaluation of Measurement Data — Guide to the Expression of Uncertainty in Measurement. Joint Committee for Guides in Metrology, JCGM 100:2008. Available at https://doi.org/10.59161/JCGM100-2008E.

[4] Butler J.M. (2006) Debunking Some Urban Legends Surrounding Validation within the Forensic DNA Community. *Profiles in DNA* 9(2):3-6. Available at https://www.promega.com/-/media/files/resources/profiles-in-dna/902/debunking-some-urban-legends-surrounding-validation-within-the-forensic-dna-community.pdf.

[5] Cavanagh, R.R., Berge, A., Coute, A., Fazio, R., Graham, T., Marshall, C., Miller, J., Mullen, L., Sozer, A., Word, C. (2019) Notes from the NIST Research Innovation to Implementation in Forensic Science Symposium (RI2I). (National Institute of Standards and Technology, Gaithersburg, MD), NIST Special Publication (SP) NIST 2100-02. Available at https://doi.org/10.6028/NIST.SP.2100-02.

[6] *[Eurachem 2019a] Barwick, V. (ed.) (2019) Planning and Reporting Method Validation Studies: Supplement to Eurachem Guide on the Fitness for Purpose of Analytical Methods (First edition). Available at https://www.eurachem.org/index.php/publications/guides/planning-validation-studies.

[7] *[Eurachem 2019b] Cantwell, H. (ed.) (2019) Blanks in Method Validation: Supplement to Eurachem Guide The Fitness for Purpose of Analytical Methods (First edition). Available at https://www.eurachem.org/index.php/publications/guides/blanks-in-method-validation.

[8] *[Eurachem 2025] Cantwell, H. (ed.) (2025) Eurachem Guide – The Fitness for Purpose of Analytical Methods: A Laboratory Guide to Method Validation and Related Topics (Third edition). Available at https://www.eurachem.org/index.php/publications/guides/mv.

[9] Federal Bureau of Investigation (2025) Quality Assurance Standards for Forensic DNA Testing Laboratories. Available at https://le.fbi.gov/file-repository/forensic-qas-070125.pdf/view.

[10] Food and Drug Administration and International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH) (2024) Q2(R2) Validation of Analytical Procedures: Guidance for Industry. Available at https://www.fda.gov/media/161201/download.

[11] Found, B. and Edmond, G. (2012) Reporting on the Comparison and Interpretation of Pattern Evidence: Recommendations for Forensic Specialists. *Australian Journal of Forensic Sciences* 44(2):193–196. https://doi.org/10.1080/00450618.2011.644260.

[12] *[FSR 2024] Forensic Science Regulator (2024) Guidance - Forensic Science Providers: Validation. FSR-G-201, Issue 2. Available at https://www.gov.uk/government/publications/forensic-science-providers-validation/forensic-science-providers-validation-accessible.

[13] *[FSR 2025] Forensic Science Regulator (2025) Draft Code of Practice 2025 (Version 2). United Kingdom Home Office. Available at https://assets.publishing.service.gov.uk/media/67daba1e594182179fe0883b/E03313596+-+CoP+Forensic+Science+Regulator+2025_A4_v02_Web+Accessible.pdf.

[14] *[ILAC-G19:2022] International Laboratory Accreditation Cooperation (2022) Modules in a Forensic Science Process. ILAC-G19:06/2022. Available at https://ilac.org/publications-and-resources/ilac-guidance-series/.

[15] *[ISO 21043-1:2025] International Organization for Standardization Technical Committee (ISO/TC 272) (2025) Forensic sciences – Part 1: Vocabulary. ISO 21043-1, second edition. Available for a fee at https://www.iso.org/standard/87128.html.

[16] *[ISO 21043-2:2018] International Organization for Standardization Technical Committee (ISO/TC 272) (2018) Forensic sciences – Part 2: Recognition, Recording, Collecting, Transport and Storage of Items. ISO 21043-2. Available for a fee at https://www.iso.org/standard/72041.html.

[17] *[ISO 21043-3:2025] International Organization for Standardization Technical Committee (ISO/TC 272) (2025) Forensic sciences – Part 3: Analysis. ISO 21043-3, first edition. Available for a fee at https://www.iso.org/standard/72040.html.

[18] *[ISO 21043-4:2025] International Organization for Standardization Technical Committee (ISO/TC 272) (2025) Forensic sciences – Part 4: Interpretation. ISO 21043-4, first edition. Available for a fee at https://www.iso.org/standard/72039.html.

[19] *[ISO 21043-5:2025] International Organization for Standardization Technical Committee (ISO/TC 272) (2025) Forensic sciences – Part 5: Reporting. ISO 21043-5, first edition. Available for a fee at https://www.iso.org/standard/73896.html.

[20] *[ISO/IEC 17025:2017] International Organization for Standardization/International Electrotechnical Commission (2017) General Requirements for the Competence of Testing and Calibration Laboratories. ISO/IEC 17025, third edition. Available for a fee at https://www.iso.org/standard/66912.html.

[21] Jones, J.P., Farrell, L.J., Benson, L.M. (eds.) (2025) Informational Scientific Primers for Officers of the Court: Intended to Strengthen Use of Forensic Science Evidence. National Institute of Standards and Technology, Gaithersburg, MD, NIST Special Publication (SP) NIST SP 1500-28. Available at https://doi.org/10.6028/NIST.SP.1500-28 [see Section 1D "Method Validation and Method Verification"; https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1500-28pt1D.pdf].

[22] *Lyle, J.R., Guttman, B., Butler, J., Sauerwein, K., Reed, C., Lloyd, C. (2022) Digital Investigation Techniques: A NIST Scientific Foundation Review. National Institute of Standards and Technology, Gaithersburg, MD, NIST Internal Report (IR) NIST IR 8354. Available at https://doi.org/10.6028/NIST.IR.8354.

[23] Martire, K.A. & Kemp, R.I. (2018) Considerations when designing human performance tests in the forensic sciences. *Australian Journal of Forensic Sciences* 50(2):166-182. Available at https://doi.org/10.1080/00450618.2016.1229815.

[24] Mnookin, J.L., Cole, S.A., Dror, I.E., Fisher, B.A.J., Houck, M., Inman, K., Kaye, D.H., Koehler, J.J., Langenburg, G., Risinger, D.M., Rudin, N., Siegel, J., Stoney, D.A. (2011) The Need for a Research Culture in the Forensic Sciences. *UCLA Law Review* 58(3):725-779. Available at https://doi.org/10.2139/ssrn.1755722.

[25] *[NRC 2009] Committee on Identifying the Needs of the Forensic Sciences Community, Strengthening Forensic Science in the United States: A Path Forward, National Research Council, National Academies of Science, 2009.

[26] *[OSAC 2020] Organization of Scientific Area Committees (OSAC) for Forensic Science (2020) Human Factors in Validation and Performance Testing of Forensic Science. OSAC Technical Series 0004. Available at: https://doi.org/10.29325/OSAC.TS.0004.

[27] *Possolo, A. (2015) *Simple Guide for Evaluating and Expressing the Uncertainty of NIST Measurement Results.* NIST Technical Note 1900. Available at https://nvlpubs.nist.gov/nistpubs/TechnicalNotes/NIST.TN.1900.pdf.

[28] *[Swofford 2024] Swofford, H. (2024) Forensic Science Environmental Scan 2023, NIST Interagency/Internal Report (NISTIR) 8515, National Institute of Standards and Technology, Gaithersburg, MD, pp. 109-110. Available at https://doi.org/10.6028/NIST.IR.8515.

[29] *[Swofford et al. 2024] Swofford, H., Lund, S., Iyer, H., Butler, J., Soons, J., Thompson, R., Desiderio, V., Jones J.P., Ramotowski, R. (2024) Inconclusive Decisions and Error Rates in Forensic Science. *Forensic Science International: Synergy* 8:100472. Available at https://doi.org/10.1016/j.fsisyn.2024.100472.

[30] Taylor, B. & Kuyatt, C. (1994) Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results: NIST Technical Note 1297. Available at https://dx.doi.org/10.6028/NIST.TN.1297.

[31] *Wickenheiser, R. and Farrell, L. (2020) Collaborative Versus Traditional Method Validation Approach: Discussion and Business Case. *Forensic Science International: Synergy* 2:230-237. Available at https://doi.org/10.1016/j.fsisyn.2020.08.003.

[32] NIST Consensus Builder. National Institute of Standards and Technology. Available at https://consensus.nist.gov/app/nicob.

[33] NIST/SEMATECH e-Handbook of Statistical Methods, April 2012. National Institute of Standards and Technology. Available at https://doi.org/10.18434/M32189.

[34] NIST Uncertainty Machine. National Institute of Standards and Technology. Available at https://uncertainty.nist.gov/.

Appendix A. Validation According to ISO/IEC 17025

ISO/IEC 17025 defines validation as "*verification*, where the specified requirements are adequate for an intended use" [ISO/IEC 17025:2017, p. 2]. Verification is defined as the "provision of objective evidence that a given item fulfils specified requirements" [ISO/IEC 17025:2017, p. 2]. ISO/IEC 17025 clarifies that the term "item" in the definition can refer to a "measurement procedure" [ISO/IEC 17025:2017, p. 2] which is considered synonymous with the term "method" [ISO/IEC 17025:2017, p. 10]. Thus, within the context of ISO/IEC 17025, validation is described as the provision of objective evidence that a given method meets specified requirements, including a performance that is deemed to be acceptable for its intended use. ISO/IEC 17025 also specifies that laboratories must retain records relating to validation and that results be reported with all the information necessary for proper interpretation [ISO/IEC 17025:2017, pp. 11-12, 14].

There are three fundamental elements of validation. Each of these elements and its relation to one another is described in turn below.

(1) *Defined scope and requirements* for an intended use

(2) *Characterization of method performance* within the bounds of a defined scope

(3) *Decision* whether the method is acceptable for use

A.1. Defined Scope and Requirements

The scope and requirements for an intended application provide the context in which the method's performance will be characterized and the basis for deciding that the performance characterization is acceptable for use.

The scope describes the purpose of the method and the circumstances (i.e., conditions, parameters, or other factors that can impact performance) for which the performance will be characterized. The requirements describe the criteria for deciding whether the method is acceptable for the intended purpose and circumstances. Requirements include, but are not limited to, specifications relating to measures of performance characteristics, such as measurement range, accuracy, limit of detection, limit of quantification, selectivity of the method, linearity, repeatability, reproducibility, robustness against external influences or cross-sensitivity against interference from the matrix of the sample or test object, and bias [ISO/IEC 17025:2017, p. 11]. Requirements relating to performance measurements should include specifications of confidence level(s) or uncertainty interval(s), as applicable. Requirements may also include specifications relating to deployment, throughput, ease of use, and health and safety. Requirements of a method for an intended application can vary depending on the context of the application, evidentiary circumstances, or needs of the users.

A.2. Characterization of Method Performance

A performance characterization is based on performance measurements and their associated uncertainties for all relevant performance characteristics. Performance measurements, including their associated uncertainties, describe the effectiveness of a method under different evidentiary circumstances, such as type, quality, or conditions of materials encountered within the scope of a given application. The performance characterization is an important basis upon which validation decisions are made (i.e., that the method is acceptable for use). The performance characterization is also essential for users of the method's result to properly interpret the output (i.e., the extent to which a particular result can be relied upon in that circumstance).

A performance characterization requires that the method and its components be subjected to an empirical evaluation. This includes empirical testing but may also include other types of analyses such as expert review. Empirical testing provides the data used to estimate the performance measurements (including their associated uncertainties) for various performance characteristics such as accuracy, precision, discriminability, repeatability, and reproducibility within the bounds of a defined scope. The choice of what performance characteristics to measure depends on the requirements specified for the intended application and the needs of the users of the method's results. Likewise, the testing should be as extensive as necessary to meet the needs of the intended application [ISO/IEC 17025:2017, p. 11].

Performance measurements should be based on data that covers the range of varying circumstances relating to evidentiary material, implementation conditions, and other factors that can impact performance within the scope of the intended application. This is important because performance measurements vary when a method is subjected to different evidentiary circumstances (e.g., quantity, quality, conditions of material or analyte available). Performance measurements may also vary among different implementations of the method or auxiliary factors, such as personnel (e.g., training, competency), facilities and environmental conditions (e.g., temperature, humidity, contamination), and equipment (e.g., function, calibration). Thus, a method could produce strong performance measurements in some evidentiary circumstances or implementation conditions, but not in others. As more validation data accumulates for a particular evidentiary circumstance and across a range of evidentiary circumstances, the more confident one can be about the performance characterization of the method and results produced.

In situations where some individual circumstances are not represented by the validation data either at all or in significant quantity, related uncertainty values will be larger. Additional validation data may be needed to adequately characterize the performance for those circumstances. These data can be obtained from testing within the laboratory or sometimes from other sources (see Guiding Principle #3). Alternatively, performance measurements for a particular circumstance (or defined subset of circumstances) for which data is limited or nonexistent can be interpolated from nearby values represented within the range of circumstances accounted for in the validation data. The extent to which such interpolation is appropriate depends on factors such as the degree to which the circumstances differ, the

amount of validation data available, and the adequacy of the performance characterization within those bounds.[4]

A.3. Validation Decision

A validation decision is the assessment of whether a method is acceptable for use. This decision is based on data demonstrating the method satisfies applicable requirements, including an acceptable performance characterization and that the method, as implemented, can be properly applied and achieve the required performance. A decision that a performance characterization is acceptable may relate to either or both: (1) a demonstration that the measures of performance and their associated uncertainties satisfy a particular specification that is appropriate for the intended use, such as minimum or maximum threshold value(s); or (2) a demonstration that the performance characterization accounts for relevant sources of error when evaluating the uncertainty of the measurement result, to an extent that is appropriate for the intended use. The decision of whether a method is acceptable for use is two-pronged.

First, a laboratory administrator must (a) evaluate the performance characterization along with other applicable requirements, (b) verify the method as implemented can be properly performed, and (c) decide whether the method is acceptable for use within a defined scope of application. Such decisions effectively *authorize* the use of the method within the laboratory.[5] However, a method that has been authorized for use does not mean it is appropriate for all evidentiary circumstances that may be encountered. The extent to which a method is appropriate for a particular evidentiary circumstance depends on the degree to which data exists to characterize its performance in a particular circumstance or similar circumstances.

Second, a laboratory analyst must consider the specific evidentiary circumstance(s) represented for a particular analysis and decide whether a previously authorized method is acceptable for use in that case and any limitations that may affect the interpretation of the result (e.g., uncertainties associated with the performance measurements due to limited validation data relevant to the specific circumstance).

Validation decisions are the responsibility of individual laboratories based on their requirements, their ability to perform the method, case-specific evidentiary circumstances, or needs of the user(s), given the context of the application. Therefore, validation decisions made

[4] Interpolation is contrasted with extrapolation. Interpolation involves estimating performance measurements for circumstances that lie *within* a range of circumstances for which data exists. Extrapolation involves estimating performance measurements for circumstances that lie *outside* a range of circumstances for which data exists. Extrapolation is often not appropriate because the estimated performance measurements would extend beyond the boundaries of the existing data, particularly when performance is likely to be worse than the circumstances represented by the existing data.

[5] A method that has been authorized for use might also be referred to as a "validated" method. However, doing so can lead to misinterpretations because not all methods that have been authorized for use within the laboratory are appropriate for all evidentiary circumstances that might be encountered.

by one laboratory are not necessarily applicable across all other laboratories or cases. A method that has been authorized by one laboratory does not obligate or constrain other laboratories to accept that decision. Requirements for “acceptable use” may differ, or the ability to properly perform the method could be limited. Likewise, a contradictory decision by one laboratory does not necessarily negate validation decisions by other laboratories.