NIST Special Publication
NIST SP 1500-28

Informational Scientific Primers for Officers of the Court: Intended to Strengthen Use of Forensic Science Evidence

Editors: John Paul Jones II, Laurel J. Farrell, Lisa M. Benson

This publication is available free of charge from:
https://doi.org/10.6028/NIST.SP.1500-28

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

NIST Special Publication
NIST SP 1500-28

Informational Scientific Primers for Officers of the Court: Intended to Strengthen Use of Forensic Science Evidence

Editors: John Paul Jones II, Special Programs Office
Laurel J. Farrell, Consultant
Lisa M. Benson, Consultant

This publication is available free of charge from:
https://doi.org/10.6028/NIST.SP.1500-28

March 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 SP 1500-28
March 2025

Publication History
Approved by the NIST Editorial Review Board on 2025-03-24

How to Cite this NIST Technical Series Publication
Jones, JP (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. https://doi.org/10.6028/NIST.SP.1500-28

Author ORCID iDs
John Paul Jones II: 0000-0002-3258-0770
Laurel J. Farrell: 0000-0002-5772-585X
Lisa M. Benson: 0009-0008-4430-6649

Contact Information
john.jones@nist.gov

Abstract

Officers of the Court including judges and attorneys can benefit from receiving information from neutral sources to inform their understanding of foundational scientific and forensic science principles. Inspired by a request from Congress and a National Research Council report, the National Institute of Standards and Technology (NIST) Forensic Science Program has prepared 17 short primers covering aspects of quality, statistics, and communication. This information is intended as an entry point for Officers of the Court on scientific topics that can enable more informed use of forensic science evidence in cases they encounter.

Keywords

accreditation; algorithms, certification; communications; documentary standards; error rates; frequency data; human factors; likelihood ratios; limitations; measurement uncertainty; method performance statistics; method validation and verification; metrological traceability; officers of the court; performance monitoring; population statistics; primers; probability; quality assurance; quality control; reports; statistics; and statistical sampling.

Table of Contents

Introduction

The National Academy of Sciences (NAS)/National Research Council published a report in 2009, entitled "Strengthening Forensic Science in the United States - A Path Forward," which highlighted many forensic science challenges in the United States from the actual practice of forensic science in laboratories to the understanding and use of forensic science information by judges and attorneys in court. The NAS specifically noted:

> "Lawyers and judges often have insufficient training and background in scientific methods, and they often fail to fully comprehend the approaches employed by different forensic science disciplines and the strengths and vulnerabilities of forensic science evidence offered during trials."[1]

In October 2021, inspired by a request from Congress[2] and the NAS report, the National Institute of Standards and Technology (NIST) Forensic Science Program launched an initiative to strengthen the understanding of foundational scientific and forensic science principles for judges and attorneys, referred to as "Officers of the Court." This project focused on the development of short primers designed to provide basic foundational knowledge that is critical for any forensic science discipline. In this context, foundational knowledge consists of understanding issues such as standards, measurement uncertainty, error rates, data interpretation, limitations of methods, weight of evidence, human factors, and communication of scientific results. These primers serve as an entry point for Officers of the Court to start their education on scientific topics and to enable them to become more informed consumers of forensic science information.

Forensic science is an applied science that is built upon the application of chemistry, biology, physics, measurement science, and statistics to address questions of identification of unknown substances (e.g., drugs, explosives, fibers, blood) and attributing evidence items (e.g., DNA, fingerprints) to a known individual. The NIST Forensic Science Program created simple briefs (in the form of 17 primers found here in) to help increase the knowledge of Officers of the Court on the components underlying the scientific basis of techniques used by forensic science service providers across the country and what results of those forensic analyses mean.

Primer Project Layout & Publication

This project resulted in the development of 17 primers. With one exception, all primers are two pages in length. All are written at a high level with the intention of making this initial exposure to scientific content more understandable by Officers of the Court. The primers are grouped into three main categories based on their content:

[1]National Research Council. Strengthening Forensic Science in the United States: A Path Forward, The National Academies Press. 2009 [https://doi.org/10.17226/12589.]
[2] Joint Explanatory Statement accompanying the Consolidated Appropriation Act of 2021 (P.L. 116-260), https://www.appropriations.senate.gov/imo/media/doc/Division%20B%20-%20CJS%20Statement%20FY21.pdf

I - Forensic Science: The Goal Is to Produce Quality Results
II - Forensic Science: Statistics Related to Results
III - Forensic Science: Communicating Results Using Reports

Each primer consists of an introduction, core content related to the primer topic, examples, key takeaways, related primer references, links with opportunities to learn more, and a link to the glossary. The 17 primers are consolidated in this NIST Publication enabling stakeholders to download the complete output in a single file. The primers will also be published as 17 stand-alone pdfs on a NIST website. It is anticipated that stakeholders may want easy reference and access to specific primers based on the case they are addressing or as topics become relevant to them. It is also anticipated that external entities that provide continuing legal education to Officers of the Court will want to provide links to specific primers enhancing the education being delivered by that provider.

Primer Production Process

The project launched in October 2021. The primer topics were originally selected by the NIST Forensic Science Program, and an external technical consultant and technical writer were retained through contracts to provide the focused support required to generate the documents. A NIST Program Manager worked with the technical consultant to identify subject matter experts at NIST who could contribute to and guide the production of each primer. Drafts of the primers were developed in a collaborative manner by the consultants and NIST staff. These drafts were then reviewed and edited by a team of NIST staff to further focus and upgrade the documents. For readability, the primers were then provided to several external reviewers who were Officers of the Court. Additional feedback and suggested upgrades were provided by members of the Organization of Scientific Area Committees (OSAC) for Forensic Science's Legal Task Group. Upgrades were made in response to this feedback and the resulting product was submitted through the NIST Editorial Review Board process for final publication.

Primer Contributors & Reviewers

The project was managed by John Paul Jones II, Forensic Science Standards Program Manager in the Special Programs Office at NIST. Laurel Farrell served as the primary technical consultant producing the initial content for each document in collaboration with NIST staff and managing content revisions after receiving feedback. Lisa Benson served as the technical writer consultant supporting the development of the documents and overall readability. The following NIST staff members contributed to the development of specific primers.

- John Butler, Special Programs Office
- Ruthmara Corzo, Materials Measurement Science Division
- Greg Fiumara, Information Access Division
- Katherine Gettings, Biomolecular Measurement Division
- Allison Getz, Special Programs Office
- Kristen Greene, Information Access Division

- Barbara Guttman, Software and Systems Division
- Hari Iyer, Statistical Engineering Division
- Kevin Kiesler, Biomolecular Measurement Division
- Melissa Phillips, Chemical Sciences Division
- Karen Reczek, Standards Coordination Office
- Christina Reed, Special Programs Office
- Kelly Sauerwein, Special Programs Office
- Edward Sisco, Materials Measurement Science Division
- Becky Steffen, Office of Reference Materials
- Melissa Taylor, Special Programs Office
- Robert Thompson, Special Programs Office
- Peter Vallone, Biomolecular Measurement Division

The NIST team that reviewed and provided additional upgrades on each primer after initial development consisted of:

- John Butler, Special Programs Office
- Allison Getz, Special Programs Office
- John Paul Jones II, Special Programs Office
- Karen Reczek, Standards Coordination Office
- Kelly Sauerwein, Special Programs Office

Henry Swofford and Vincent Desiderio, both in the Special Programs Office, provided additional helpful feedback on two primers that underwent further revisions.

The Officers of the Court who provided feedback on the readability of various primers include:

- Hon. Carlos F. Acosta, Associate Judge, Circuit Court for Montgomery County, MD
- Hon. Robert Taylor, Jr., Associate Judge, Circuit Court for Baltimore City, MD
- Amie Ely, Director, National Association of Attorneys General
- Hon. Eugene Gasiorkiewicz, Judge, Racine County Circuit Court, WI
- Hon. Ron Reinstein, Judge (Ret.), Superior Court of Arizona
- Hon. Kent E. Cattani, Judge, Arizona Court of Appeals
- Lynn Garcia, General Counsel, Texas Forensic Science Commission

Several members of the OSAC Legal Task Group also provided feedback on the primers that was considered prior to producing the final product that was routed through the NIST Editorial Review Board resulting in this final publication.

I. Forensic Science: The Goal Is to Produce Quality Results

Introduction:

Forensic science service providers (FSSPs) must make every effort to provide quality results to the legal system - law enforcement, investigators, attorneys, judges, juries, victims, and the accused - their customers. International Organization for Standardization (ISO) 9000:2015[1] defines quality as the "degree to which a set of inherent characteristics of an object fulfills requirements." Conformity assessment is the mechanism to evaluate either an organization or a person to determine whether they conform to specified requirements in a documentary standard published by a national or international Standards Developing Organization (SDO).

To consistently produce quality results, organizations need a culture and management system based on the following principles[1]:

- A focus on current and future customer requirements
- Leadership that is united in purpose and direction and that understands that everyone in the organization must be engaged in reaching the organization's quality objectives
- Staff that is competent, recognized, and empowered to assist in reaching the organization's quality objectives
- A process approach to the work performed
- An ongoing focus on improvement
- Decision-making based on the analysis and evaluation of data and information
- Relationship management of relevant interested parties

None of these principles individually, or even as a group, will guarantee that every result produced meets the quality expectations of the organization. A complete, interrelated quality assurance framework is necessary to establish and maintain an effective organization that can identify non-conforming results, facilitate an investigation into these occurrences to minimize recurrence and risk, maximize process improvement, and consistently produce quality results. Many aspects of a functioning quality assurance framework are the topics of primers in this category.

Primer Topics in This Category

A. Accreditation & Certification

Conformity assessment plays a critical role in many sectors, including the forensic sciences, to demonstrate an adherence by an organization, individual, product, process, or system to specific requirements within that sector. Accreditation of organizations and certification of people are two types of conformity assessments.

B. Documentary Standards

Documentary Standards convey an agreed-upon way of doing something. They are established by consensus and often approved by a recognized body such as a Standard Setting Organization (SSO) or a Standards Developing Organization (SDO). Documentary standards provide the basis for conformity assessment and are also used by manufacturers, purchasers, and others to verify that an object (e.g., person, product, test result, process, service) meets specific requirements and is fit for purpose.

Footnote: 1. ISO 9000:2015 Quality Management, International Organization for Standardization, Geneva, Switzerland

C. Quality Assurance versus Quality Control

Quality Assurance (QA) and quality control (QC) both play a vital role in an organization's overall quality management system. QA and QC are often used interchangeably to refer to actions performed to ensure a quality result. However, the terms have distinct definitions. QA is the framework put in place by an organization to provide confidence that quality requirements are being met. QC, a subset of QA, is put in place by an organization to focus specifically on the operational activities used to fulfill quality requirements. Both QA and QC reduce the risk of producing a result that does not meet an organization's quality expectations. Both QA and QC increase the probability of identifying a quality failure if one occurs, but neither is a guarantee of no quality failures.

D. Method Validation and Method Verification

In forensic science, methods are used to answer a question (e.g., how much does an item weigh? what is the concentration of a chemical in a substance? is there a DNA profile on the item?). Before methods can be used to examine an item they must be validated or verified to impart confidence in the output. Method validation establishes, through documented experimentation, that a scientific method is fit for purpose--in layman's terms, when used as specified it does what it is intended to do. Method verification involves a process to ensure that a previously validated method performs as expected when used by others.

E. Metrological Traceability

Metrology is the science of measurements. A measurement result must be tied to a globally accepted standard to ensure uniformity and consistency of measurement results worldwide. Metrological traceability ties a measurement result made by a FSSP to the International System of Units (SI).

F. Human Factors

High-reliability organizations understand and optimize how people interact with the system that is the foundation for the task they are performing. That system includes other people, facilities and equipment, and a management system. Optimizing the human factors related to each of these areas improves the quality of results produced by reducing the potential for error.

G. Algorithms

"Computational science is often the only way to process and understand a diverse set of artifacts that are available for analysis in criminal cases. Computational forensic science is built on algorithms and the software systems that execute those algorithms"[2]. Algorithms are used for a variety of purposes, including evaluating the quality of a sample, extracting features (also referred to as characteristics, attributes, or profiles), searching a database of features, and comparing an unknown sample to one or more known (reference) samples. They allow FSSPs to partially automate the examination process, thereby improving the speed and objectivity of the work performed.

H. Performance Monitoring: Methods, People, Organizations

Ongoing confidence in FSSP performance is essential for the FSSP, their customers (i.e., law enforcement, investigators, attorneys, judges, juries, victims, and the accused), regulators, certification bodies, accreditation bodies, and other interested parties. Performance monitoring, a type of quality assurance activity, provides this type of data but an understanding of the data is required to use it correctly.

I. Why Certain Items Are Selected for Examination

The initial recognition, recording, sampling, and possible examination of items of potential forensic value occurs at a scene. Generally, as the investigative questions become more refined, the number of items that proceed to examination at the FSSP facility narrows. The number of items introduced in a legal proceeding may be less than the number of items examined.

Footnote: 2. *Software & Algorithms Catalog*. NIST. (2022, January 28). Retrieved June 2, 2022, from https://forensicsoftware.nist.gov/index.php

A. Accreditation and Certification

Introduction

Conformity assessment plays a critical role in many industries, including the forensic sciences, to demonstrate an adherence by an organization, individual, product, process, or system to specific requirements within that sector. Accreditation of organizations and certification of people are two types of conformity assessments.

Accreditation and Certification Overview

Although accreditation and certification are granted following similar methodology, they differ in objectives and purposes. Accreditation is granted to an organization to recognize conformance with agreed-upon requirements, technical competence, and organizational effectiveness. Certification of people is granted to recognize an individual's competence such as obtaining required knowledge, skills, or abilities. An impartial third party conducts a review and attests to the requirements being met when they grant accreditation or certification.

Both are important and complementary components of conformity assessment. It should be noted that the impartiality, consistency, and competence of the accreditation or certification body impact the quality of the conveyed conformity assessment information. Similarly, the types of assessment activities and the adequacy and appropriateness of the standards used to evaluate conformity also impacts the quality of the conveyed assessment information.

A Comparison of Accreditation and Certification

	Accreditation of Organizations	Certification of People
Available For	• Testing and calibration laboratories • Inspection bodies (IBs) • Proficiency test providers (PTPs) • Reference material producers (RMPs) • Certification bodies (CBs)	• Individuals demonstrating competence in a specific area
Provided By	Accreditation bodies (ABs) – Several ABs exist in the forensic science sector, although one AB accredits most forensic science service providers (FSSP).	Certification bodies (CBs) – There are several CBs in the forensic science sector. CBs may be accredited.
Process	• The scope of work to be assessed is determined. • Accreditation requirements are agreed upon (e.g., International Organization for Standardization/International Electrotechnical Commission (ISO/IEC) standard, AB-specific requirements). • An audit of a sample of records, sufficient to determine conformance, is performed. • One or more reports are issued with audit findings. All non-conformance findings require acceptable remediation. • The AB decides to grant or renew accreditation for the scope of work assessed. • The AB makes the accreditation scope and certificate publicly available.	• CBs, often using certification boards with expertise in the certification area, determine the characteristics of competence to be evaluated. • CBs review applications against minimum qualifications (e.g., education, professional experience, training hours and/or professional development credits, or professional/character references). The applicant may be required to abide by a Code of Ethics. • The CB administers one or more exams (e.g., written, oral, practical, observational, or other means) to assess competence. Processes may allow for retesting. • The CB decides to grant or renew certification.
Surveillance	Assessment activities at a specified time frame (typically annually) ensure key elements of the requirements are still being met. This is less rigorous than the initial assessment or reassessment audits.	Support for ongoing attestation of certification may include employment, memberships, attending meetings/conferences, continuing education training, presentations, instruction of training workshops/course, service on committees, proctoring exams, other activism, or proficiency testing.
Reassessment	Reassessment must be made prior to the end of the accreditation cycle that cannot exceed five years.	Recertification varies by CB but is required at the certification's expiration that cannot exceed five years.
Additional Information	The AB can extend, reduce, suspend, or withdraw accreditation.	The CB can extend, reduce, suspend, or revoke certification.

The Conformity Assessment Method

Additional Considerations

	Accreditation of Organizations	Certification of People
Strength	Proponents see accreditation as a proven tool to improve the quality of an organization's work product.	Proponents see certification as a tool to standardize the evaluation of a person's competence.
Challenges	Critics raise concerns that it is an evaluation that an organization is doing what they say they will do rather than an evaluation of whether the work product (e.g., testing, certification, reference material) they provide is fit-for-purpose. Variability in the accreditation process is also raised as a concern.	Critics raise concerns about the criteria used to determine competence and how to best evaluate each on a large scale in a way that reflects day-to-day work and yet meets test design criteria.
Limitations	Accreditation is *not*: ▪ An assessment of every record produced by the accredited organization ▪ An attestation of an individual result produced by the organization ▪ Certification of personnel ▪ A first-party assessment (self-declaration) that an organization has implemented one or more standards	Certification is *not*: ▪ Available for all areas of forensic science ▪ An attestation of work by the individual on a specific case ▪ A mechanism to calculate an error rate for an individual or a method ▪ Licensing – however, licensing may require a certification
Voluntary vs. Mandatory	Accreditation of FSSPs has been primarily voluntary; however, some governmental organizations mandate accreditation. Mandatory accreditation was supported by both the National Academy of Sciences and the National Commission on Forensic Science. The National Commission on Forensic Science also supported the accreditation of certification bodies.	Certification for forensic science practitioners is primarily voluntary; however, some organizations require personnel certification when available. Mandatory certification of forensic science professionals was supported by both the National Academy of Sciences and the National Commission on Forensic Science.

Forensic Science Examples

The majority of United States (U.S.) FSSPs are accredited based on an ISO Standard (e.g., ISO/IEC 17025). Use this link to find U.S. ABs recognized for accrediting testing and calibration laboratories and inspection bodies. From the AB's website, you can search the FSSPs that they accredit and review their scope of accreditation. International Laboratory Accreditation Cooperation Mutual Recognition Arrangement (ILAC MRA)

Information on who a CB has certified is not publicly available, but this directory can be searched to identify CBs that have been granted accreditation. International Accreditation Forum Multilateral Recognition Arrangement (IAF MLA)

Key Takeaways

1. Stakeholders should review:
   - The scope of accreditation to determine whether it covers the work performed in a specific case.
   - The scope of a person's certification to determine if it is relevant for the specific case.
2. Stakeholders should remember:
   - Neither accreditation nor certification is an attestation related to a specific result.
   - The quality of the result still needs to be reviewed.

Related Primers

Documentary Standards

Learn More

For additional information on conformity assessment (accreditation and certification), see:

- ABCs of Conformity Assessment
- Why use an Accredited Laboratory (ILAC)
- NIST Standards Coordination Office, About Standards.gov

Glossary

B. Documentary Standards

Introduction

Documentary standards convey an agreed-upon way of doing something. They are established by consensus and often approved by a recognized body such as a Standard Setting Organization (SSO) or a Standards Developing Organization (SDO). Documentary standards provide the basis for conformity assessment and are also used by manufacturers, purchasers, and others to verify that an object (e.g., person, product, test result, process, service) meets specific requirements and is fit for purpose.

Standards Overview

Documentary standards are relied upon in just about every industry and play an essential role in protecting public health and safety, developing new technologies, facilitating national and international commerce, and more.

These standards come in many forms (e.g., best practices, guides, practices, technical reports, terminology, test methods). The type of standards published varies by SSO or SDO.

Standards are not unique to the forensic sector. In fact, they are so prevalent in society that it's estimated that 93 percent of global trade is affected by standards and technical regulations[1].

Use of standards is primarily voluntary. When incorporated by reference (IBR) into laws or regulations, voluntary standards become mandatory. An example of an IBR standard is the *Manual on Uniform Traffic Control Devices (MUTCD),* which designates requirements for road signs in the United States (U.S.).

In other instances, the marketplace drives compliance to voluntary standards, putting those companies or products that don't comply at a disadvantage. An example of market-driven compliance to voluntary standards is retailers demanding that electronic products have been UL Certified.

National and International Standards

Standards are created in response to the needs of an industry, government, or consumers. In the U.S., standards organizations are decentralized, separated by industry sectors. To avoid duplication of efforts, the American National Standards Institute (ANSI), a private, non-profit organization, coordinates standards development activities and represents the U.S. standards community internationally. ANSI is not an SDO but instead provides a framework for fair standards development and quality conformity assessment systems. In other countries, standards development is typically coordinated by a single government entity.

The three largest and most well-established international standards organizations are the International Organization for Standardization (ISO), the International Electrotechnical Commission (IEC), and the International Telecommunication Union (ITU). Standards developed by these organizations are consensus-based with a membership that includes government, private, and public-private entities.

Standards Development Process

In the U.S., documentary standards are developed in a coordinated, consensus-based process that involves various stakeholders, including individuals, companies, industry organizations, scientific and professional societies, or government agencies. SDOs subscribe to certain principles for the standard development process. These principles include:

- Balance – No one person, company, or organization dominates standard development activities, and input is valued from all participants.
- Consensus – Parties must compromise. Consensus does not mean everyone agrees with everything in the standard.
- Due process – Processes are transparent, and pathways exist to appeal decisions.
- Openness – The process is open to all interested parties.
- Transparency – Advanced public notice of standards activity is provided. The process generally allows for public comment on draft standards. Records are accessible.

Strengths of Standards

- They provide a foundation for consistency between processes related to testing, certification, management systems, and personnel certification.
- Use of generalized wording (e.g., appropriate, fit-for-purpose, meet the customer's needs) supports broad application.
- Wording that focuses on what is required and not on how it is to be done supports use by organizations of varying size, structure, location, budget, and various scopes of accreditation.
- Well-written standards have a positive impact on the quality of results produced.

Limitations of Standards

- They have little to no significance until adopted and used.
- Generalized wording is open to interpretation.
- Allowing organizations to determine how something is done means that multiple approaches can conform to the standard.
- Poorly written standards can reduce the quality of results produced if they do not sufficiently standardize the impactful attributes of the targeted activity.

Footnote:
1 ANSI Introduction. (2021). Retrieved from ANSI website: https://www ansi.org

Forensic Science Examples

ISO standards are used as the basis for accreditation of forensic science service providers (FSSPs), related service providers, and accrediting bodies within the forensic science sector. The ISO 17000 series of standards play an essential role in conformity assessment activities. Figure 1 depicts the standards from this series used for conformity assessment activities applicable to forensic science.

At the top of the conformity assessment hierarchy (Figure 1) are accreditation bodies that provide accreditation to conformity assessment bodies. These accreditation bodies are subject to compliance with standard ISO/IEC 17011. To be accredited, conformity assessment bodies must meet specific standards. The middle level shows standards applicable to five types of conformity assessment bodies that can be accredited. The lower level is the object of assessment.

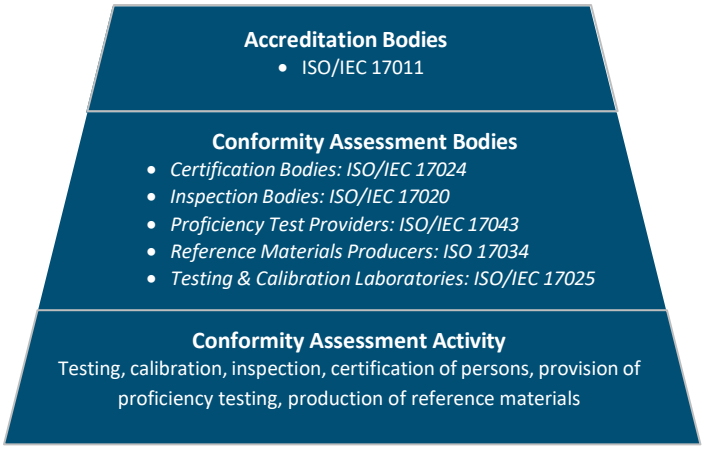


Figure 1 Conformity Assessment Hierarchy

Putting this altogether, a recognized accrediting body must conform to ISO/IEC 17011. They can assess an FSSP for conformity to ISO/IEC 17025. When the FSSP conforms to all applicable requirements for the testing services assessed, the accreditation body will issue a certificate and scope of accreditation to the FSSP.

There are a number of SSOs/SDOs that have active consensus bodies or working groups developing forensic science standards. ANSI provides a search function that can be used to find standards related to forensics.

In 2014, the Organization of Scientific Area Committees (OSAC) for Forensic Science was created to address a lack of discipline-specific forensic science standards[2]. The OSAC Registry is a repository of selected published and proposed standards for forensic science that have undergone a technical and quality review process. The types of forensic science standards on the OSAC Registry include terminology, training/competence, method validation, methods, quality assurance, reporting, opinions and testimony standards.

Use of non-consensus standards may result when these standards are developed to comply with a law or when developed by an organization for a profession. Examples of such standards used by FSSPs include:

- FBI Quality Assurance Standards for Forensic DNA Testing Laboratories
- FBI Quality Assurance Standards for DNA Databasing Laboratories
- NAME Inspection and Accreditation Checklist
- ABFT Forensic Toxicology Laboratory Accreditation Checklist

In 2021, the United Kingdom was the first country to establish government regulation of Forensic Science.

Key Takeaways

1. A standard is an agreed upon way of doing something.
2. Use of standards is primarily voluntary.
3. Standards allow for individual variability in implementation.
4. Standards used in conformity assessment have significant impact on the validity of the procedures, the value of the information conveyed, and the cost[3].
5. ISO/IEC 17025 is the most common standard used in the accreditation of FSSPs.

Related Primers

Accreditation and Certification

Learn More

Encyclopedia of Forensic Sciences, 3rd edition (2022), *Organization of Scientific Area Committees (OSAC) for Forensic Science*
Encyclopedia of Forensic Sciences, 3rd edition (2022), *Standards and Conformity Assessment*
The ABCs of Standards Activities
The ABC's of Conformity Assessment
NIST Standards Coordination Office, About Standards.gov

Glossary

Footnotes:

2. About OSAC, (2021), retrieved from National Institute of Standards and Technology website: https://www.nist.gov/osac
3. Carnahan, L. and Phelps, A. (2018), ABC's of Conformity Assessment, Special Publication (NIST SP), retrieved from National Institute of Standards and Technology website, https://doi.org/10.6028/NIST.SP.2000-01 (Accessed December 19, 2021)

C. Quality Assurance vs. Quality Control

Introduction

Quality assurance (QA) and quality control (QC) both play a vital role in an organization's overall quality management system. QA and QC are often used interchangeably to refer to actions performed to ensure the consistant quality of a result. However, the terms have distinct definitions. QA is the framework or program put in place by an organization to provide confidence that the requirements of a quality management system are being met. QC, a subset of QA, is a piece of that framework put in place by an organization that focuses specifically on the operational activities used to fulfill quality management requirements. Both QA and QC reduce the risk of producing a result that does not meet an organization's quality expectations. Both QA and QC increase the probability of identifying a quality failure if one occurs, but neither is a guarantee of zero quality failures.

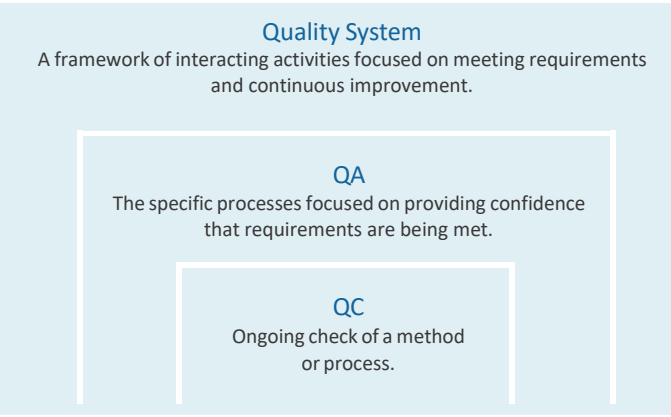
Quality Assurance and Quality Control Overview

QA and QC are used by organizations and industries, such as manufacturing, pharmaceutical, professional services, and information technology, to meet specific regulatory, industry, or customer requirements. The importance of QA and QC in forensic science was noted by both the National Academy of Sciences and the National Commission on Forensic Science.

As noted in the introduction to this category of primers (Forensic Science: The Goal is to Produce Quality Results), conformity assessment requires an organization to have a management system with elements that focus on producing quality results. QA pulls these elements together to increase confidence in the reported results. QC provides the ongoing operational 'check' of a method or process. In the most basic terms, QA focuses on preventing nonconforming products and correcting nonconforming processes, and QC focuses on detecting a nonconforming product if one should occur. See Figure 1 for the relationship between a quality system, QA, and QC.

A forensic science service provider (FSSP) granted accreditation based on an International Organization for Standardization (ISO) standard has met its quality objectives by implementing a quality system that includes QA and QC, as well as quality planning (i.e., identifying and defining requirements) and quality improvement.

Figure 1 Quality control (QC) is an element of quality assurance (QA). Both QA and QC are part of an organization's quality system.



Quality Assurance

QA is a large encompassing framework focused on preventing quality issues and continuously improving processes. It creates a foundation for assessing risks and opportunities within the quality system.

In some organizations, an activity listed below as QA may be viewed as QC.

In addition to including QC, QA encompasses the following aspects of the entire process from sample recognition through reporting of results:

- Sampling or guidance on how sampling should be performed to ensure the best sample for testing, calibration, and inspection
- Item handling from receipt through disposal, including how the item is submitted, chain of custody, appropriate storage, and handling through testing, calibration, inspection, and reporting
- Initial training and competency testing of staff
- Staff and organization performance monitoring
- Staff certification
- Method validation or method verification
- Establishing and maintaining metrological traceability
- Equipment function checks
- Statistical process control
- Replicate testing, calibration, or inspection
- Retesting, recalibration, or reinspection
- Blind testing, calibration, or inspection
- Technical and administrative review
- Internal audit
- Document control
- Record control
- Suggestions from staff
- Customer feedback
- Management review
- Accreditation

Quality Control

QC is designed to detect nonconforming work at an operational level. Using QC samples, also called controls, is one way to ensure this. A QC sample is a material of known composition that is either tested or inspected alongside unknown samples to ensure the method's reliability at that time or used to verify calibration of measurement equipment.

The following parameters for QC samples will be defined by method validation or method verification data:

- Frequency – A control sample may be included each time the method is performed or at some set interval.
- Matrix – A close representation of the sample being tested, calibrated, or inspected (e.g., the weight of a baggie of a known material, length of a reference firearm, blood matrix in a blood ethanol method).
- Sample number – The number of QC samples may be one or more.
- Type – A calibrated reference standard or certified reference material (e.g., mass reference standard, length reference standard, certified ethanol reference material) provides information on method precision and bias. Use of a material of known origin that is not a calibrated reference standard or certified reference material provides information on method precision.
- Value – The concentration or value of one or more QC samples may be tied to a legal requirement (e.g., 0.08% ethanol QC sample in an antemortem blood ethanol method or a 1 gram mass reference standard QC sample in a seized drug weight method).

When nonconforming work is identified through a QC or QA activity, the nonconforming work must be corrected and evaluated for significance, including impact on past work. A corrective action is taken, when applicable, to reduce the risk of a recurrence.

Forensic Science Examples

ISO/IEC 17025, General requirements for the competence of testing and calibration laboratories and *ISO/IEC 17020, Conformity assessment — Requirements for the operation of various types of bodies performing inspection* both serve as the foundation for accreditation of FSSPs and include requirements related to QA and QC.

Other Standards

Many of the standards listed on the Organization of Scientific Area Committees (OSAC) for Forensic Science Registry contain aspects of QA and QC. Here is an example from ANSI/ASB 054 *Standard for a Quality Control Program in Forensic Toxicology Laboratories*:

8.1 At a minimum, the following apply to the use of controls in all forensic toxicology analyses.

a) All controls shall be tested and treated the same as case samples.
b) Negative and positive controls shall be included with each analytical batch.
c) Process controls shall be included when a procedure includes a technique such as hydrolysis or oxidation.
d) The laboratory shall define parameters for accepting or rejecting controls (Section 8.3). Each control sample shall be checked for acceptability using these predefined criteria.

Other standards used by FSSPs that contain requirements related to QA and QC include:

- FBI Quality Assurance Standards for Forensic DNA Testing Laboratories
- NAME Inspection and Accreditation Checklist
- ABFT Forensic Toxicology Laboratory Accreditation Checklist

Key Takeaways

1. QA is the group of inter-related processes and procedures that focus on ensuring the production of quality results.
2. QC is a component of QA.
3. QC is put in place to confirm a method is still operating as validated.
4. QA is continuously improved to reduce risk and enhance opportunity.

Related Primers

Accreditation and Certification

Method Validation and Method Verification

Metrological Traceability

Performance Monitoring: Methods, People, Organizations

Learn More

For additional information on QA and QC, visit American Society for Quality (ASQ):

- Quality Assurance & Quality Control
- Learn About Quality
- Quality Glossary

Glossary

D. Method Validation and Method Verification

Introduction

In forensic science, methods are used to answer a question (e.g., how much does an item weigh? what is the concentration of a chemical in a substance? is there a DNA profile on the item?). Before methods can be used to examine an item they must be validated or verified to impart confidence in the output. Method validation establishes, through documented experimentation, that a scientific method is fit for purpose--in layman's terms, when used as specified it does what it is intended to do. Method verification involves a process to ensure that a previously validated method performs as expected when used by others.

Method Validation

Method validation involves three phases: development, optimization, and implementation. Depending on what has been done by other organizations, a forensic science service provider (FSSP) may not need to complete all of these phases (see Table 1 and Figure 1) prior to use.

Method validation:

- Phases can be performed by one or more individuals or teams
- Does not include opinion statements
- Establishes limitations of the method and reported results
- Aids in identifying what is required for ongoing quality assurance (QA) and quality control (QC)
- Aids in assessing measurement uncertainty or error rates

Table 1 The three phases of method validation

Phase One
Method development is typically performed by research scientists and published in a peer-reviewed journal
Phase Two
Method optimization commonly establishes: • "Fit-for-purpose" criteria • Equipment specifications and operating parameters • Metrological traceability • Sample preparation approach • Sample analysis approach • The observations, data, or calculations generated • Interpretation of observations, data, or calculations
Phase Three
Method implementation uses known materials (i.e., source, identity, concentration) that represent the range of anticipated work to evaluate "fit-for-purpose" criteria: • Method performance and limitations (i.e., precision, bias, sensitivity, specificity) • Determination of item suitability for examination

Method Verification

Verification provides objective evidence that the method performs at the stated performance level of the original validation. The validated method is used with no modification. The components of phase three that are susceptible to variation when the method is used by a different organization, another facility within a single organization, or a different analyst using different equipment are evaluated, generally with a smaller number of samples.

After Method Validation or Method Verification

Use of the method by multiple staff provides intermediate precision data. Use by other organizations or multiple facilities within a single organization after verification provides reproducibility data. Both, along with quality control (QC) and performance monitoring data, provide ongoing information to support that the method continues to perform as validated and continues to meet the fit-for-purpose criteria.

This additional data may allow refinements to measurement uncertainty or error rates and may identify opportunities for further method optimization. Additional optimization is considered method modification. Modifications return the method validation process to phase two, which must be followed by phase three to re-evaluate any criteria impacted by the modification.

Challenges

Limitations on the number of samples and types of samples required to adequately represent the range of anticipated work are not unique to forensic science. The number and types of samples used may impact method performance statistics. The samples used may increase the number or magnitude of limitations to observations, data, or calculations, which may also impact an interpretation or an opinion.

Consensus standards used for the accreditation of FSSPs use general terms such as "fit-for-purpose," "appropriate," "meeting the customer's requirements," and "needs of the customer" when characterizing a method. Performance criteria used to define these general terms may be included in consensus standards that are test methods. Generally, when performance criteria are not specifically defined in a standard, they are determined by the FSSP based on their understanding of the customer's needs.

These challenges generate differing opinions on the choices made related to the number and type of samples used in method validation and method verification and on the criteria used to establish that a method is fit-for-purpose.

Forensic Science Examples

In addition to Figure 2, see ANSI/ASB 018, Standard for Validation of Probabilistic Genotyping Systems on the OSAC Registry.

Figure 1 General flow of method validation

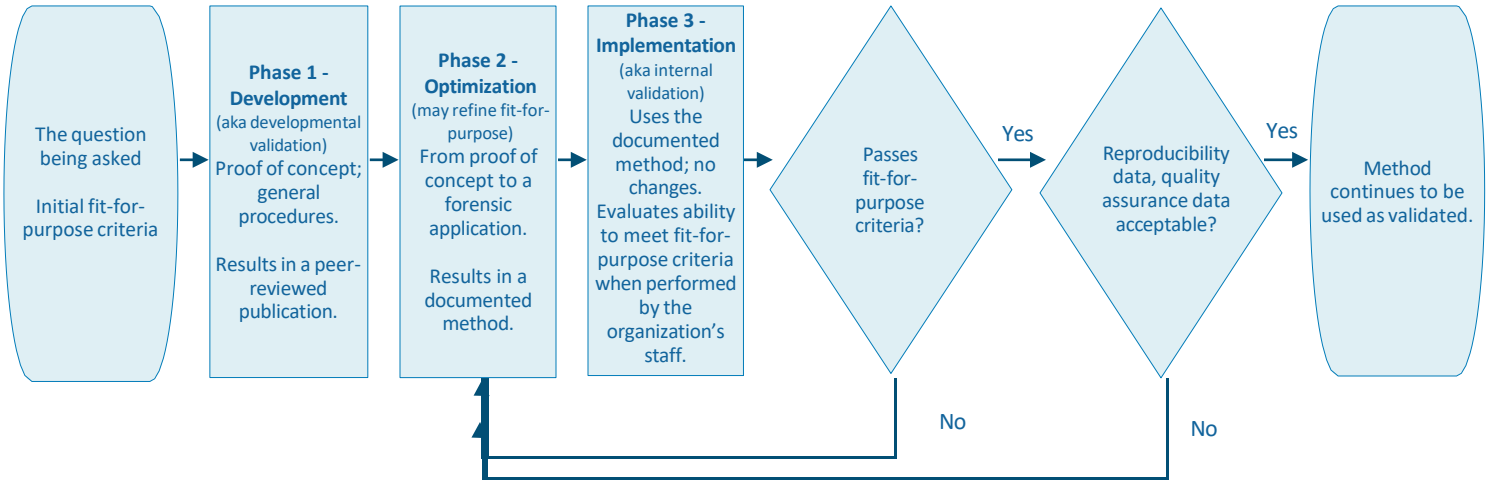
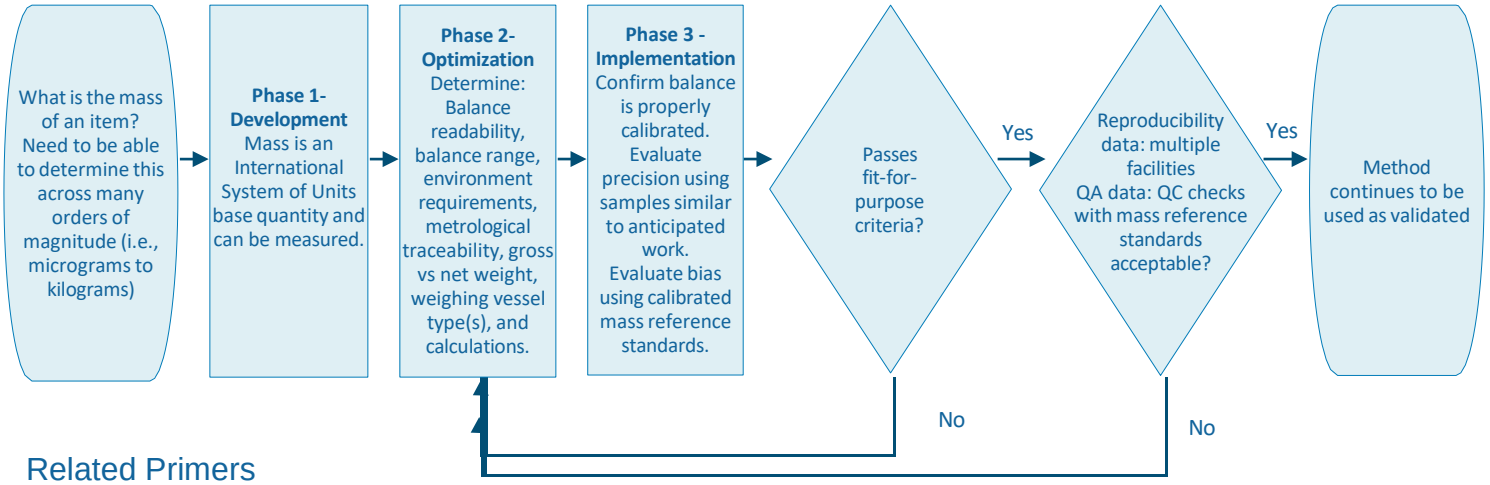


Figure 2 Forensic science example for determining the mass of an item



Related Primers

- Documentary Standards
- Method Performance Statistics
- Metrological Traceability
- Performance Monitoring: Methods, People, Organizations
- Quality Assurance versus Quality Control

Learn More

To learn more about validation and verification, see:

Eurachem's Planning and Reporting Method Validation Studies – Supplement to Eurachem Guide on the Fitness for Purpose of Analytical Methods

NIST's Validation Information to Aid Forensic DNA Laboratories

OSAC's Human Factors in Validation and Performance Testing of Forensic Science

Research article of Bradford et al., Accuracy and reliability of forensic latent print decisions

Key Takeaways

1. Method validation is performed prior to use in an examination.
2. Method validation consists of three phases - development, optimization, and implementation.
3. Method validation establishes the limitations of the method and reported results.
4. Method verification is appropriate when the validated method has not been modified.
5. Method validation and method verification are different from ongoing quality control.

Glossary

E. Metrological Traceability

Introduction:

Metrology is the science of measurements. A measurement result must be tied to a globally accepted standard to ensure uniformity and consistency of measurement results worldwide. Metrological traceability ties a measurement result made by a forensic science service provider (FSSP) to the International System of Units (SI).

Metrological Traceability Overview

Customers need work performed to be reliable and accurate. They also need to be assured that if the work is performed on a different day by the same provider or by another provider, the result will be similar and comparable within the stated measurement uncertainty. The need for reliability, uniformity, consistency, and comparability is the fundamental reason for establishing the traceability of a measurement result. The need for uniformity and consistency in measurements is not unique to forensic science but is also required in research, industry, and commerce.

Metrological traceability is only applicable to a measurement result. Measuring equipment and organizations do not have metrological traceability.

Sometimes the term metrological traceability is shortened to traceability. However, the entire term, metrological traceability, is preferred to avoid confusion with other concepts that use traceability to mean the history of an item, such as sample traceability, document traceability, equipment traceability, or material traceability.

Measurements

The measurement process may be a single step, such as a length measurement, or take many steps, such as quantifying the amount of a drug in blood. The measurement result may be reported, such as the weight of a seized drug, or the measurement may be part of a test, calibration, or inspection method where the reported result does not meet the definition of a measurement but instead meets the definition of a nominal property, such as a caliber of bullet (e.g., 9mm) or type of car (e.g., sedan/truck).

The International System of Units (SI)

The SI, commonly referred to as the metric system, is the international standard for measurement. The International Bureau of Weights and Measures (BIPM) is the organization that is tasked with ensuring worldwide uniformity of measurements traceable to the SI. Table 1 provides the seven base quantities in the SI, the base unit name, and the symbol.

Using these seven base units, users may make many more measurements from SI-derived units (e.g., kg·m/s2, which is a Newton), an SI prefix unit (e.g., giga, centi, milli), or by using a unit conversion (e.g., metric to U.S. customary units conversions).

Table 1: SI Base Units

BASE QUANTITY NAME	BASE UNIT NAME	SYMBOL
Time	second	s
Length	meter	m
Mass	kilogram	kg
Electric Current	ampere	A
Thermodynamic Temperature	kelvin	K
Amount Of Substance	mole	mol
Luminous Intensity	candela	cd

Source: NIST Special Publication 330:2019, Table 2

Metrological Traceability Chain

It is possible to tie a measurement result to the SI through a series of calibrations. This is called a metrological traceability chain. See Figure 1 for an example of a metrological traceability chain.

Provider Measurement Result
(e.g., weight, volume, length)

Provider Measuring Equipment
(e.g., balance, pipette, ruler)

Accredited Calibration Service Supplier
(1 or more between the provider and the NMI)

National Metrology Institute (NMI)
(e.g., National Institute of Standards and Technology (NIST)

International System of Units (SI)

Figure 1 Metrological Traceability Chain

In chemical analysis, metrological traceability is established through one or more calibration measurement standards called calibrators. If an FSSP prepares a calibrator, each piece of measuring equipment used in its preparation (i.e., pipette, volumetric flask, balance) will have been calibrated to establish metrological traceability of the measurement made using that measuring equipment. If the calibrator is a certified reference material, used exactly as provided by the reference material producer (RMP), the metrological traceability chain for the certified reference material will include the RMP and the calibration of the measuring equipment used in its production.

Establishing Metrological Traceability

Metrological traceability is established by considering and then ensuring the following[1]:

a) The measurand, the quantity to be measured, is defined.
b) A documented unbroken chain of calibrations is established between the measurement being made back to stated and appropriate references (appropriate references include national or international standards and intrinsic standards).
c) Measurement uncertainty for each step in the traceability chain is evaluated according to agreed methods.
d) Each step of the chain is performed in accordance with appropriate methods, with the measurement results and with associated, recorded measurement uncertainties.
e) The organizations performing one or more steps in the chain supply evidence of their technical competence.

Maintaining Metrological Traceability

It is not enough to merely establish metrological traceability for a measurement process. Maintenance of metrological traceability (i.e., continued metrological confirmation and statistical control of the measurement process) is necessary for subsequent measurement results to have this property.

A proper measurement assurance program, part of overall quality assurance, must be established to ensure the ongoing validity of the measurement process and the accuracy of the reference standard used at the time of the measurement.

Forensic Science Examples:

Table 2: Examples of Measuring Equipment Used by FSSPs and the Corresponding SI Base Quantity for the Measurement Made

MEASURING EQUIPMENT	SI BASE QUANTITY
Balances	Mass
Calipers	Length
Pipettes	Derived from Mass
Rulers/Tapes	Length
Thermometers	Temperature
Trigger-Pull Devices	Derived from Mass
Voltmeters	Derived from Electric Current
Volumetric Glassware	Derived from Mass

Measurements made by FSSPs include, but are not limited to, determinations of the following:

- Concentration (e.g., 0.187 g/dL ethanol in blood)
- Length (or height or size) (e.g., firearm barrel length of 16 inches)
- Refractive index (e.g., glass refractive index of 1.522409)
- Temperature (e.g., ethanol and water wet bath simulator temperature of 34.2^0 C)
- Volume (e.g., volume of liquid 67.4 mL)
- Weight (e.g., net weight 29.4 grams)

Key Takeaways

1. If the reported result is a measurement result, or if the measurement will affect the validity of the reported result, ensure:
   - Metrological traceability was established.
   - The measuring equipment was calibrated or certified reference material was provided by an appropriately accredited service provider.

Related Primers

Accreditation and Certification

Quality Assurance versus Quality Control

Measurement Uncertainty, Frequency Data and Error Rates

Learn More

Eurachem Education and Training Working Group's Terminology in Analytical Measurement: Introduction to VIM 3

International Bureau of Weights and Measures (BIPM), SI Brochure: The International System of Units (SI)

International Vocabulary of Metrology's (VIM), JCGM 200:2012

NIST Resources:
- SI Units
- Metrological Traceability: Frequently Asked Questions and NIST Policy
- GMP 13 – Good Measurement Practice for Ensuring Metrological Traceability

Glossary

Footnote

1. International Organization for Standardization. (2017). General requirements for the competence of testing and calibration laboratories, Annex A. ISO/IEC 17025:2017. Geneva Switzerland.

F. Human Factors

Introduction

High-reliability organizations understand and optimize how people interact with the system that is the foundation for the task they are performing. That system includes other people, facilities and equipment, and a management system. Optimizing the human factors related to each of these areas improves the quality of results produced by reducing the potential for error.

Human Factors Summary

The field of human factors is a combination of numerous disciplines, such as psychology, sociology, engineering, industrial design, physiology, anthropometry, and user experience, with a focus on designing systems that make it hard for people to do the wrong thing and easy for people to do the right thing. Systems are made up of policies, practices, facilities, equipment, and people that operate according to a set of rules to form a unified whole. "Cars are designed so that drivers cannot start them while in reverse because that prevents accidents. Work schedules for pilots are designed so they don't fly too many consecutive hours without rest because alertness and performance are compromised."[1] What is true of drivers and pilots is also true for persons in forensic science and the legal community. The study of human factors examines the interactions between humans and other elements of a system with the goal of optimizing those interactions.

Human Factors and Quality Assurance

To quote Alexander Pope, "to err is human." With system optimization, the impact of human factors issues can be mitigated and errors can be minimized, leading to reliable results. Different from errors, violations are a deliberate deviation from a rule or procedure that can also be identified and addressed with a focus on human factors.

It is important to have a framework of quality assurance activities that can effectively detect, track, and correct deviations, both errors and violations, from desired practice. Blaming individuals for an error is counterproductive. Errors indicate that there may be an opportunity to improve interactions with the system. Not evaluating the root causes of violations raises the potential to miss how deviations from rules or procedures can become routine, what the situation was that resulted in the violation, or why someone would think that the violation was the right thing to do. For both errors and violations, cause analysis completes the feedback loop that supports continuous improvement of the human/system interaction, building in necessary redundancy to improve detection and tracking.

Human Factors - A Model

The mnemonic P.E.A.R. is a tool to help identify elements that could impact system performance. P.E.A.R. stands for People who do the job, Environment in which they work, Actions they perform, Resources needed for job completion. An example of how this is used is seen in Table 1.

Decision Making

Decision-making is a complex process that is impacted by all aspects of P.E.A.R. One aspect of decision making that is frequently discussed in forensic science is bias. Bias can be further broken down into specific types of bias that include cognitive bias, confirmation bias, contextual bias, expectation bias, or motivational bias. A bias can impact sampling, testing, or inspection strategies; the actual observations made of an item; how observations, data and calculations are interpreted; and how opinions are reached. The need to understand the impact of bias on forensic decision making was discussed by both the National Academy of Sciences and the National Commission on Forensic Science. Awareness alone of one's own bias does not prevent its impact on decision-making unless system level policies are in place to detect, understand, and address errors.

Forensic Science Examples

P.E.A.R. in Action

In March of 2004, the FBI Laboratory erroneously identified Brandon Mayfield as the source of a fingerprint found on a bag of detonators in Madrid that was connected to the 2004 bombing of several commuter trains. The FBI Laboratory "attributed its mistake to 'practitioner error' as distinguished from a failure of the science"[2]; however, a deeper review identified systemic issues. Illustrative findings and solutions from the 2006 and 2011 Office of Inspector General (OIG) reports have been listed in the P.E.A.R. model in Table 1.

Additional applications of human factors in forensic science include:

- The Organization of Scientific Area Committees (OSAC) for Forensic Science task group with expertise in human factors and completed process maps: including Footwear & Tire, Friction Ridge, Human Forensic DNA Analysis and Speaker Recognition
- NIST published research: Latent Print, Forensic Handwriting and Forensic DNA Interpretation
- Forensic Technology Center of Excellence: Just Human Factors & Human Factor in Forensic Science Practice
- Cognitive Consultants/ Forensic Identification tab

Footnotes:
1 Institute of Medicine (US) Committee on Quality of Health Care in America, Kohn, L. T., Corrigan, J. M., & Donaldson, M. S. (Eds.). (2000). To Err is Human: Building a Safer Health System. National Academies Press (US).

2 U.S Department of Justice website. (2006). A review of the FBIs Handling of the Brandon Mayfield Case, Retrieved from https://oig.justice.gov/sites/default/files/archive/special/s0601/final.pdf

Table 1. Illustrative findings and solutions from the 2006 and 2011 OIG reports have been organized in the P.E.A.R model

ASPECT OF P.E.A.R.	PEOPLE WHO DO THE JOB	ENVIRONMENT IN WHICH THEY WORK	ACTIONS THEY PERFORM	RESOURCES NECESSARY TO COMPLETE THE JOB
HUMAN FACTORS CONSIDERED	Physical, physiological, and psychological factors	Physical and socio-technical factors	Interaction with the system when performing the task	Tangible and intangible resources needed to complete the task
EXAMPLES OF HUMAN FACTORS	Size, age, sex, fatigue, health, workload, experience, and training	Lighting, temperature, distractions/interruptions, organizational culture, supervision, morale, and job security	Sequence of steps, documentation, and communication requirements	People, time, equipment, technology, continuing training, and communication between people
P.E.A.R. IN ACTION: ELEMENTS IN OIG REPORT OF THE FBI'S HANDLING OF THE BRANDON MAYFIELD CASE	Staff had many years of experience Staff had appropriate motivation Bias was shown in the examination process after looking at a known print Bias potential was elevated due to an awareness of the initial result during the verification process	International high-profile case that involved mass casualties Work at times was done on the weekend or through the night Staff faced a deadline for completion of work Reluctance to think that an error could have been made	Staff followed the approved procedure for examination, including verification No specific examination minimum criteria for acceptable item and image quality Requirements for examination documentation were lacking Documentation of and basis of support for explanation of differences was inadequate Requirements for documentation of verification review were lacking	System prevented task-irrelevant information (i.e., suspect demographic information) from being disclosed to the examiner during examination Verbal communication of results was not clear, and results were interpreted in different ways Lack of thorough root cause analysis completed by the FBI when notified of a conflicting result by another forensic provider
P.E.A.R. IN ACTION: SOLUTIONS THAT WERE IMPLEMENTED BY THE FBI	Major upgrade to the training program for latent print examiners Revised procedures and training to mitigate bias by analyzing the unknown before looking at any known fingerprint Use of blind verification in cases with greatest risk	Culture shift in approach to examinations with differing conclusions	Support for research projects Increased the level of detail in the procedures Increased requirements for documentation for each phase of the Analysis, Comparison, Evaluation and Verification (ACE-V) methodology Clear separation of the analysis and comparison phases Limited acceptance of any difference Limited use of Level 3 detail Verifier required to separately complete ACE Conflict resolution process revised including documentation expectations Increased structure for a technical review	Review of past cases with similar aspects (e.g., digital image rather than the original evidence and single prints) Increased rigor in corrective action process

Key Takeaways

1. Errors happen
2. A systematic approach to human factors considers People, Environment, Actions, Resources (remember P.E.A.R.)
3. The potential for errors, including bias, can be identified and mitigated or removed
4. Quality assurance activities are the main avenue to detect, correct and prevent errors and violations

Related Primers

Documentary Standards

Quality Assurance versus Quality Control

Learn More

Reason, J. (2000). Human error: models and management. BMJ, 320:768

Dr. Johnson, W., & Dr. Maddox, M. (2007). A Model to explain Human Factors in Aviation Maintenance. Avionics News.
Transport Canada. (2003). Human Performance Factors for Elementary Work and Servicing

Glossary

G. Algorithms

Introduction

"Computational science is often the only way to process and understand a diverse set of artifacts that are available for analysis in criminal cases. Computational forensic science is built on algorithms and the software systems that execute those algorithms[1]". Algorithms are used for a variety of purposes, including evaluating the quality of a sample, extracting features (also referred to as characteristics, attributes, or profiles), searching a database of features, and comparing an unknown sample to one or more known (reference) samples. They allow forensic science service providers (FSSPs) to partially automate the examination process, thereby improving the speed and objectivity of the work performed.

Overview of Algorithms

An algorithm is a sequence of steps for solving a problem or accomplishing a task. Algorithms are often implemented as a computer program. One type of algorithm that is particularly useful in forensic science is search algorithms, but there are many other types (see Table 1).

Algorithm Development

Algorithms are developed to solve a specific problem utilizing a specific type of input data. In some cases, algorithms are written by programmers, but others are developed using a type of artificial intelligence (AI) called machine learning. Machine learning allows software applications to become more accurate without being explicitly programmed to do so. Machine learning algorithms use historical data as input to make decisions and predict new output values. Many algorithms will have tradeoffs in performance characteristics based on their purpose. Screening algorithms may be optimized to find any potential candidate, whereas confirmation algorithms may be optimized to reduce false positives.

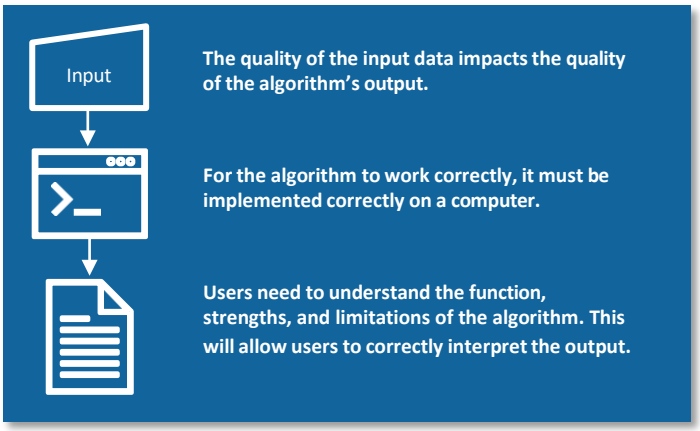
Algorithm Testing

After an algorithm is developed, it is checked to ensure that it meets its intended purpose. AI-based algorithms will depend on the data set used to train the AI. If the training data set is not relevant to how the algorithm will be applied, the algorithm could produce misleading outputs (see Figure 1). The algorithm's implementation is tested by running a variety of data through it. The testing of algorithms used to perform steps in a forensic method becomes part of method validation.

Algorithm as Part of a System

An algorithm is always part of something bigger. At a minimum, software includes the algorithm, but the software may be part of a larger system (e.g., in chemical analysis, more than one algorithm is part of the data processing system for a gas chromatography mass spectrometry system, which itself also includes sample handling/delivery). Aspects of the system impact the function of the algorithm. People are an important part of the system and must understand what the algorithm is designed to do, how different types of input may impact the output, and how to interpret the output. Method validation must consider both the algorithm testing and how the algorithm is used as part of a system for a specific type of input data.

Figure 1. Factors impacting an algorithm's output



Understanding Limitations and Error Rates of Algorithms and the Systems that Implement them

The algorithm, and the systems that implement the algorithm must be understood. Both can be the source of an error that has the potential to impact the output. These types of errors are systematic and not random. Algorithm performance can sometimes, but not always, be described by a false positive and false negative rate.

Different algorithms and implementations optimize the scores and the rankings for different situations based on the needs of the user. This will impact the output. A search of the same sample in the same database by different algorithms is likely to have different outputs. Similarly, a search of different databases using the same sample and algorithm is likely to have different outputs.

Types of Forensic Science Algorithms

There are many types of algorithms used in forensic science (see Table 1 and Figure 2 for examples). With one exception noted, the function can be performed by either a human or an algorithm. The limitations listed in Table 1 will apply whether the function is performed by a human or an algorithm.

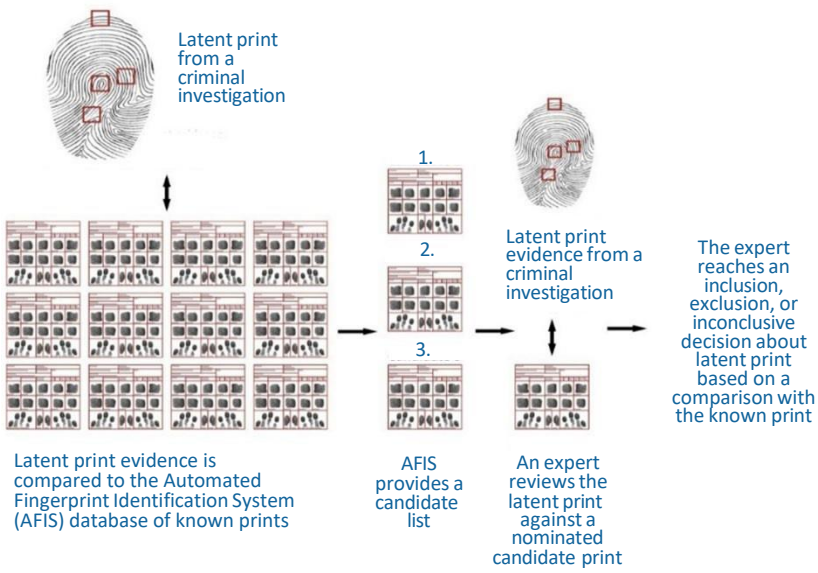
Footnote: 1. *Software & Algorithms Catalog*. NIST. (2022, January 28). Retrieved June 2, 2022, from https://forensicsoftware.nist.gov/index.php

Forensic Science Examples

Table 1: Types of algorithms used in forensic science

TYPE OF ALGORITHM	PURPOSE	STRENGTHS	LIMITATIONS	FORENSIC SCIENCE EXAMPLE
QUALITY OF ITEM EVALUATION	Ensure an item meets specified criteria for examination	Objective, standardized evaluation	Only as good as the quality criteria specified May miss unusual features	Primarily done manually at this time DFIQI and LQMetric – friction ridge
FEATURE EXTRACTION/ PROFILE GENERATION	Identify, extract, and annotate if applicable	Speed Scaleability Consistency	Output is dependent on the quality of input Only as good as the feature criteria specified May miss unusual features	DNA electropherogram, drug or toxicology mass spectrum, latent print minutia, bullet striations
SEARCH (1:N)	Search a database of known samples for similar features/profiles	This function is often not feasible for a human to perform	Risk of incorrect candidate or incorrect ranking on candidate list Output will be dependent on the relevance and size of the database	Databases: • Mass spectral library • National DNA Index System (NDIS) • Next Generation Identification (NGI) • National Integrated Ballistic Information (NIBIN) • Solemate
1:1 COMPARISON	Compare the features from an unknown sample to one known sample	Speed Consistency	Output will be dependent on the quality of the sample and feature extraction Risk of incorrect outcome May miss unusual features	Unknown DNA profile compared to a DNA reference sample Unknown drug sample compared to a reference material
SPECIAL PURPOSE	Need- and algorithm-specific	Algorithm-specific	Dependent on the purpose of the algorithm but may be impacted by the quality of the criteria, steps of the algorithm, and databases used	Probabilistic genotyping software (PGS) Digital and Multimedia – (e.g., recovery of deleted files, hashing algorithm, searching for strings or file headers)

Figure 2. Latent Print Example (Source: GAO \ GAO-20-479SP)



In this example, the output is a candidate list of known samples with similar features to the unknown latent fingerprint. This same scenario occurs when a DNA profile or chemical structure is searched using an algorithm. Each candidate (e.g., person, DNA profile, or chemical compound) on the list is generally accompanied by a similarity score. Candidates are then compared to the unknown sample by a forensic science practitioner.

Key Takeaways

1. An algorithm utilizes a specifc type of input data to solve a specific problem.
2. Machine learning algorithms are only as good as the training data used.
3. Two different algorithms may yield different outputs for the same data based on the strategy for the algorithm.
4. Algorithm performance can sometimes be described by a false positive and false negative rate that may be impacted by the components within the system.

Related Primers

Method Performance Statistics
Method Validation and Method Verification
Quality Assurance versus Quality Control

Learn More

NIST Software & Algorithms Catalog

U.S. Government Accountability Office Report: Forensic Technology: Algorithms Used in Federal Law Enforcement

Next Generation Identification (NGI)

Glossary

H. Performance Monitoring: Methods, People, Organizations

Introduction

Ongoing confidence in forensic science service provider (FSSP) performance is essential for the FSSP, their customers (i.e, law enforcement, investigators, attorneys, judges, juries, victims, and the accused), regulators, certification bodies, accreditation bodies, and other interested parties. Performance monitoring, a type of quality assurance activity, provides this type of data but an understanding of the data is required to use it correctly.

Performance Monitoring Overview

Interlaboratory or intralaboratory comparison data provides information on a laboratory's performance over time.

Performance monitoring operates on the premise that the monitoring test's attribute of interest (e.g., concentration, source of the fingerprint) is known to the provider but not to the participant. In some instances, the participant may be unaware that they are participating in a performance monitoring activity.

The purposes and benefits of performance monitoring include[1]:

- Evaluation of an individual practitioner's performance;
- Evaluation of the FSSP performance;
- Identification of unexpected results that initiates correction and corrective action;
- Support for evaluations of measurement uncertainty;
- Identification of a method limitation that had not been identified during method validation;
- Identification of intra - and inter-FSSP differences; and
- Establishment of method effectiveness and comparability.

Types of Performance Monitoring

Intralaboratory Comparison

Performance monitoring conducted within an organization is an intralaboratory comparison. Random case reanalysis by an FSSP is an example of an intralaboratory comparison. In this scenario, the "ground truth" or correct answer is not known for the item(s), but the results from two or more forensic science practitioners can be evaluated for consistency.

Interlaboratory Comparison

An interlaboratory comparison involves two or more organizations. Also referred to as a round-robin or collaborative trial, the interlaboratory comparison can be structured to focus on a specific aspect of interest (e.g., matrix, quality of comparison test item) and is often used for education with the goal to improve comparability and competence. *Forensic Science Example:* *NIST Cannabis Quality Assurance Program*

Proficiency Testing

Proficiency testing (PT), a specific type of interlaboratory comparison, is coordinated by a Proficiency Test Provider (PTP).

PT schemes vary in the frequency and number of test items provided over a specified timeframe. PTPs strive to provide test items that align with the type of items routinely encountered.

When homogeneous test items cannot be created on a large scale, such as with latent print lifts, handwriting, shoes, and tires, then reproductions are used.

Primary Relevance of the Performance Monitoring

Depending on how the performance monitoring activity is used by the FSSP, the primary relevance of the information gained will vary (see Table 1).

Table 1. Primary relevance of performance monitoring

Performance Monitoring Activity Scenario	Primary relevance of the performance monitoring		
	Organization	Method	Individual
Single practitioner completes all examinations on the comparison test item. Technical review (by a separate practitioner) *is not* performed.			✓
Single practitioner completes all examinations on the comparison test item. Technical review *is* performed.	✓		✓
Multiple practitioners are involved in different steps of the examination of the comparison test item. Technical review *is* performed.	✓		
Single practitioner uses three different validated methods to examine the comparison test item.		✓	

Benefits and Challenges of Performance Monitoring

The primary benefit is to compare the performance of one FSSP to other FSSPs providing the same service.

Numerous challenges include:

- A lack of monitoring for all services provided by FSSPs
- Alignment with routine items but possibly not representative of the most challenging items received
- Ensuring homogeneity of items provided to participants
- Stability, handling, storage, and environmental conditions during distribution
- Cost

Footnote:

1. ISO 17043:2010 Conformity Assessment – General Requirements for Proficiency Testing, International Organization for Standardization, Geneva, Switzerland

Performance Monitoring Reports

The report from a non-PT interlaboratory comparison does not contain pass/fail grades but is provided to the participants to see how their results compare to other participants and may be published in the scientific literature to disseminate the findings.

PTPs provide both PT item summary statistics and individual participant performance data. Summary data will include staff who have participated in known or blind proficiency testing, who are both certified and non-certified individuals, who work in both accredited and non-accredited FSSPs, and who may be in training or involved with research and development. A PT report may also provide demographic data, method information, and report language.

A PTP determines the assigned value for a proficiency test item in several ways (increasing in uncertainty)[2]:

- Known value based on PT test item formulation
- Certified reference value (for a quantitative test) - metrological traceability established
- Reference values - metrological traceability established
- Consensus value from expert participants
- Consensus value from participants

The final determination of successful completion of a proficiency test must be made by the FSSP and, if applicable, the accrediting or certification body.

Relevancy of a performance monitoring activity to a specific legal case

Performance monitoring is one piece of data that depending on relevancy to a specific case, can demonstrate the performance of an individual, a method, or an organization over time. Not all performance monitoring activities will be representative of the complexity in a specific legal case (see Table 2).

It is not appropriate to take performance data (pass/fail) on a performance monitoring activity and use it as the error rate of a specific legal case, for the individual testifying, or to calculate the error rate of the discipline for that method of examination.

Learn More

Peterson JL, Markham PN. Crime laboratory proficiency testing results, 1978-1991, I: Identification and classification of physical evidence and II: Resolving questions of common origin. J Forensic Sci. 1995 Nov;40(6):994 - 1029.

Inter-laboratory Studies in Analytical Chemistry

Implementation of a Blind Quality Control Program in a Forensic Laboratory

Implementing Blind Proficiency Testing in Forensic Laboratories: Motivation, Obstacles, and Recommendations

Solving *Daubert's* Dilemma for the Forensic Sciences Through Blind Testing

Table 2. Relevancy of performance monitoring activity reports

Specific Legal Case and Performance Monitoring Activity Scenario	Relevancy of Performance Monitoring Activity Report to a Specific Legal Case	
	More Relevant	Less Relevant
Same method is used in both	✓	
The method was used by a majority of the monitoring participants	✓	
The practitioner performed the same steps of the examination (e.g., extraction, quantification) in both	✓	
A different method was used in the case than used in the monitoring		✓
The method was used by only a small number of monitoring participants		✓
The practitioner did not perform the examination but performed a technical review		✓
The FSSP quantitative result +/- the measurement uncertainty includes the monitoring activity quantitative value	✓	

Key Takeaways

1. Performance monitoring activities provide different information related to an organization, a method, or an individual.
2. Not all performance monitoring activities will be representative of the complexity of a specific legal case.
3. Successful completion may represent evidence of competence for that performance monitoring activity but may not reflect ongoing competence of the FSSP, and unsuccessful completion may reflect only a random departure from an FSSP's normal state of competence[2].

Related Primers

Accreditation and Certification
Documentary Standards
General Statistics
Measurement Uncertainty, Frequence Data & Error Rates
Method Validation and Method Verification
Method Performance Statistics
Metrological Traceability
Quality Assurance versus Quality Control

Glossary

Footnote:

2. ISO 17043:2010 Conformity Assessment – General Requirements for Proficiency Testing, International Organization for Standardization,

I. Why Certain Items Are Selected for Examination

Introduction:

The initial recognition, recording, sampling, and possible examination of items of potential forensic value occurs at a scene. Generally, as the investigative questions become more refined, the number of items that proceed to examination at the forensic science service provider's (FSSP) facility narrows. The number of items introduced in a legal proceeding may be less than the number of items examined.

Topic Summary

The initial decision of whether an item will be sampled for examination (testing, inspection, or analysis) occurs at a scene. A forensic science practitioner may make these decisions, but this is not always the case (e.g., detective). Decisions are made based on the case information available, the items available at the scene, what items may best be transported to the FSSP facility for examination, and the questions being considered at the time of the scene investigation (see Figure 1).

At this stage, the questions may be quite broad, such as how did someone die, was someone assaulted, or is a material a controlled substance? There may be examinations performed at a scene (e.g., a presumptive test of a body fluid, enhancement of ridge detail to locate fingerprints, bullet trajectory) that assist in determining what items will be collected for additional examination.

All work performed at a scene must be documented. A chain-of-custody, which ensures the integrity of each item, is initiated at the scene and maintained through any examination performed at the scene, packaging, transportation and storage, continuing through submission to and examination by an FSSP.

Process Optimization

Decisions of what items will be examined are conducted by humans and can be negatively influenced by human factors. Optimization of this process ensures resources are used effectively, items most likely to answer proposed questions will be collected and examined, and appropriate known samples will be obtained. Process optimization may include using a case management team, which usually consists of representatives from the FSSP, the responsible law enforcement agency, and the prosecuting attorney.

If this type of approach is not employed, the FSSP may not be aware of all the items in the custody of law enforcement, law enforcement or the prosecuting attorney may not understand the capability of the FSSP, or the FSSP may examine items that are not relevant to the questions being considered by the investigators or attorney.

Item Examination Strategy

Once submitted to a FSSP, a detailed examination strategy may have to be developed for individual items to ensure proper sequencing of the examinations performed. An examination strategy for an item is often required when more than one forensic discipline or method will be used to examine an item (e.g., fingerprints, firearms, and DNA) or when the amount, volume, or size of a sample is limited. During examination, recognition of additional items (e.g., hair, fibers) may occur and therefore refine or change the examination strategy for that item or even the case.

The questions being considered will continue to be refined even after items have been submitted to a FSSP. This can result in the examination strategy being changed or stopped. Examination of an item may also end if the item is insufficient or unsuitable

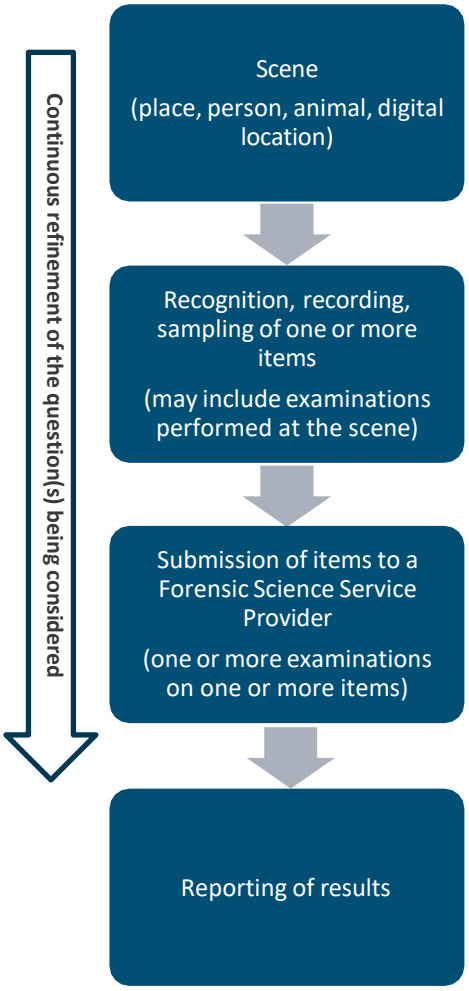


Figure 1. Process for why certain items are selected for examination. It should be noted that initial submission of items to the FSSP may not include all items sampled. Based on new information or results of initial examinations, additional items may be submitted at later dates. However, all items submitted to a FSSP may not be examined.

Legal counsel representing the defendant may also request an item be examined if relevance can be shown.

for examination, or if the appropriate methodology is not available. If the necessary methodology is instead available from another FSSP, the case management team would be best suited to decide whether to transport the item.

Forensic Science Example:

Hypothetical attempted homicide case

Scene response:
Paramedics arrive at a residence where a toddler was found in the kitchen with a serious gunshot wound. The father told paramedics that it was an accident. The baby was stabilized and recovered in the hospital. The firearm was recovered at the scene.

The father left the scene before officers arrived and was not located until months later.

The bullet that passed through the toddler and through the kitchen ceiling was recovered in an upstairs bathroom.

Initial question: Was the firearm the source of the recovered bullet?

> The initial request for the FSSP was to determine whether the expended bullet was fired from the suspect firearm.

Father told paramedics that he dropped the firearm in the kitchen, and it discharged.

Refinement: Was the shooting event described by the father consistent with the victim's wound track?

Refinement: Could the firearm have discharged when dropped?

> A review of the shooting reconstruction and medical reports supported that the trajectory was from the kitchen floor, through the toddler from the lower torso exiting from the upper torso, and passed through the kitchen ceiling.
>
> The examination strategy was refined. Results reported were that the firearm would consistently fire when dropped, consistent with the father's statements of an accident.

POSSIBLE ITEMS AT THE SCENE	ITEMS COLLECTED	ITEMS SUBMITTED TO FSSP	ITEMS EXAMINED BY FSSP
PORTION OF THE KITCHEN CEILING WITH HOLE	✓		
SECTION OF FLOORING WITH HOLE	✓		
VICTIM CLOTHING (PACKAGED SEPARATELY)	Not recovered from hospital		
FIREARM	✓	✓	✓
EXPENDED BULLET	✓	✓	✓
EXPENDED CARTRIDGE CASE	✓	✓	

Key Takeaways

1. Understand what was collected but was not examined and why.
2. Ensure that the report is clear as to what was and was not examined.
3. Strive for a high level of real time communication during the examination process to refine the questions that are considered.
4. Examination strategies are used to determine what methods will be used and in what order to obtain the most information available for an item.

Related Primers

Human Factors

Statistical Sampling

Learn More

For additional information on why certain items get examined, see:

Crime Scene Investigation: A Guide for Law Enforcement

FBI Handbook of Forensic Services

Houck, M. & J. Siegel. 2015. *Fundamentals of Forensic Science, Third Edition*. Cambridge, MA: Academic Press.

Houck, M., Crispino, F., & McAdam. 2018. *The Science of Crime Scenes, Second Edition.* Cambridge, MA. Academic Press

NIST Biological Evidence Guidance

Saferstein, R. 2018. *Criminalistics: Introduction to Forensic Science.* Boston: Pearson Education.

Glossary

II. Forensic Science: Statistics Related to Results

Introduction:

Is the result accurate? Often this is the question that is asked. One can quickly list over 25 synonyms for "accurate" or "accuracy" that show the variability in perspectives when this term is used. Individuals often have their own definition of what "accurate" equates to.

Accuracy, defined as the closeness of agreement between a result and the "true value," requires information on both precision and bias to describe the closeness of a result to a known or reference value, the "true value." [1] Encompassing both precision and bias, at times referred to as variability and trueness, involves a combination of random and systematic error components.

Results, the product of the forensic science service provider, is a broad term that includes observations, data, calculations, interpretations, and opinions.[2] Some observations, data, and calculations generated by a forensic science method make sense as generated. Weight data is understandable as it is; no interpretation is required. It is more common for observations, data, and calculations to require interpretation prior to reporting to be useful. Latent print minutia observations are an example of data that require interpretation prior to reporting. Interpretations are explanations for the observations, data, or calculations from an examination.[2] If interpretation is necessary, it is incorporated in method validation and would be part of any statistics describing the performance of the method. Opinions take into consideration other information in addition to observations, data, calculations, and interpretations.[2]

Statistics related to results can be generated for different types of methods (i.e., qualitative, quantitative, pattern-based), although some statistics are more commonly used for certain types of methods. Understanding the strengths and limitations of these statistics is critical to their correct use.

The primers in this category cover statistics that support a method being used to produce results, how statistical sampling and population statistics are used, and approaches for interpreting results and expressing opinions.

Primer Topics in This Category

A. General Statistics

Statistics is the science concerned with developing and studying methods for collecting, analyzing, interpreting, and presenting empirical data. The data are either described or used to make an inference about a population.

Footnotes: 1. ISO 3534-2:2006 Statistics — Vocabulary and symbols — Part 2: Applied statistics, International Organization for Standardization, Geneva, Switzerland

2. OSAC Preferred Terms, Organization of Scientific Area Committees for Forensic Sciences. (2022). Retrieved from https://www.nist.gov/system/files/documents/2021/12/14/OSAC%20Preferred%20Terms_December%202021.pdf

B. Method Performance Statistics

Empirical data from method validation and method verification can be used to calculate statistics that describe a method's performance. The statistics used may vary by the type of method, but all will fall into one or more of these four areas: precision, bias, sensitivity, and specificity. These statistics establish method limitations and, therefore, provide limitations for the results produced from using that method. These statistics can also assist in identifying what is required for ongoing quality assurance and aid in assessing measurement uncertainty or error rates.

C. Statistical Sampling

Statistical sampling methods are used in forensic science when all items in a population will not be examined and there is a need to make an inference to that population. The use of statistical sampling methods do not guarantee that a sample is representative of the population but do support representativeness of the sample. If a sample is not representative, the usefulness of the inference being made about the population may be negatively impacted. The relevancy of the population to the question being asked is the topic of the primer on Population Statistics

D. Population Statistics

Population statistics provide information about attributes, such as variability, in a population. Population statistics are almost always estimates based on one or more population samples. When inferring from a sample to a population, the value of the inference depends on how the sample was selected, sample size, number of samplings, and population's relevance to the inference topic.

E. Probability & Likelihood Ratios

Statistics and probability are both useful tools for interpreting data and observations. They have some overlapping features, but they are conceptually different. Those who support the use of probability and likelihood ratios (LRs) in forensic science opinions see them as an application of logical reasoning with uncertainty. Opposition to using probability and LRs in forensic science opinions focuses on the subjectiveness (variability) in the LRs that are provided, the use of LRs that do not align with performance data and misunderstanding of the expression.

F. Measurement Uncertainty, Frequency Data & Error Rates

All results produced by a forensic science service provider have an associated uncertainty. When the result is a measurement, that uncertainty is expressed as a quantitative value. When the result is not a measurement, the uncertainty can be expressed as frequency data or as an error rate for different possible results. All assist the user of the information in understanding the limitations of the reported results.

A. General Statistics

Introduction

Statistics is the science concerned with developing and studying methods for collecting, analyzing, interpreting, and presenting empirical data. The data are either described or used to make an inference about a population.

General Statistics Overview

Populations Versus Samples

The science of statistics can be focused on data from either a population or a sample. A population is an entire group about which specific information will be obtained. A population can be finite (e.g., the weight of every student in the 2nd-grade class at Hypothetical Elementary on 3/1/2022), or it can be very large, infinite, or continuing to emerge (e.g., the weight of every 2nd-grade student in the world), making it impractical or impossible to have values for all items in the population. If less than the entire population is in the data set, then the data set is a subset of the population and is referred to as a sample.

Descriptive and Inferential Statistics

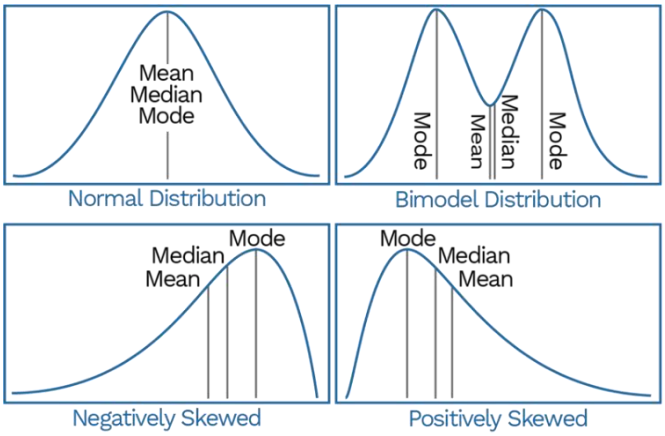
Descriptive statistics reduces a large number of values to a manageable few or presents the values in a way to facilitate visualization. Descriptive statistics are calculated using data (from a sample or from a population) and can be presented in graphical, tabular, or numerical ways. Examples of graphical and tabular summaries include histograms, probability plots, and frequency distributions. Examples of numerical descriptive statistics include mean, variance, and standard deviation.

Unlike descriptive statistics, inferential statistics uses a sample to generalize about a population. The sample must be representative of the population and the process used to select the sample must be repeatable. This may be referred to as random or probability sampling.

Measure of Central Tendency

A measure of central tendency is a method to represent a population or data set by a single point at the center of its data values. The mean (average value) is the most widely used statistic to describe the central tendency of a data set. The median (middle value) is the next most widely used statistic to express the central tendency of a data set. For symmetric data sets, the mean and median will be the same. Other statistics that are used to express the central tendency of a data set are midrange (halfway between the minimum and maximum values) and mode (the value occurring most often). A data set will have no mode if each value occurs only once in the data set, or the data set may be bimodal or multimodal if values occur with the same frequency (see figure 1).

Figure 1. Examples of Distributions and Skews



Measure of Dispersion

Similarly, there is often a desire to represent a population or a sample by a single number that expresses the data's dispersion (variability, scatter, or spread) around the center of the distribution. The range (the difference between highest and lowest values) is the simplest statistic to calculate. Due to the sensitivity of a range statistic to extreme values, it may be preferable to provide the maximum and minimum values or the interquartile range that focuses on the middle 50% of the values in the data set. A large interquartile range means that there is a large spread, greater variation, between these values.

Standard deviation is the most used measure of dispersion. It is appropriate to use standard deviation as a measure of dispersion when the mean is used to measure central tendency. Standard deviation compares each data point to the mean of all data points. Variance also measures the deviation of the data from the mean and is the square of the standard deviation.

Distributions

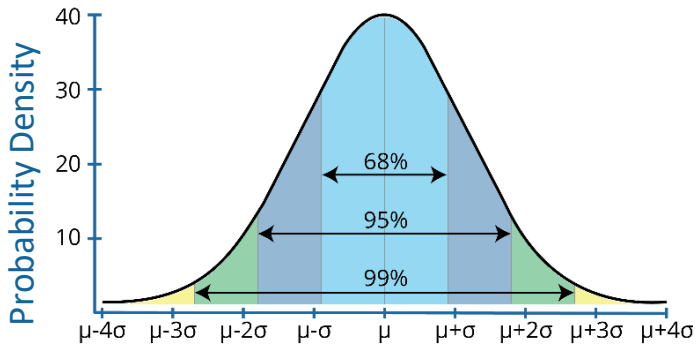
The frequency of each value can be displayed graphically with the values on the x-axis and the frequency of occurrence of each value on the y-axis. If the data are limited to distinct, separate values, known as a discrete random variable, the data will be displayed as the different values that are possible over the interval (e.g., quality control data from a balance with a readability to 0.01 gram would be displayed in bins of 0.01 g each). If the entries in the data set across the stated interval can take on any value with an infinite number of decimal places possible, a continuous random variable, then the distribution, or density curve, becomes smooth (e.g., quality control data for a quantitative method).

The Normal Distribution

Many people are familiar with normal distribution, also referred to as a Gaussian distribution or a bell curve. Data from many natural phenomena as well as measurement processes can be approximated by a normal distribution. This is a symmetric distribution where the majority of the data values center around the mean, and data values at the outside margins of the distribution are less likely to be observed. The mean, median, and mode are the same value in a normal distribution.

If data from a process approaches a normal distribution with a population mean μ and population standard deviation σ, then the probability that the next value from that same process will be within one standard deviation of the mean is approximately 68%. The probability that the next value will be within two standard deviations of the mean is approximately 95%, and the probability that the next value will be within three standard deviations of the mean is approximately 99% (see figure 2). If data are not normally distributed, then statements regarding "the next value" can be made but the percentages may differ.

Figure 2. Normal Distribution (μ = mean, σ = standard deviation)



It is possible for two data sets to have the same mean or the same standard deviation; therefore, when describing a data set, it is preferable to provide both a descriptor of central tendency (e.g., mean) and of dispersion (e.g., standard deviation).

Related Primers

Statistical Sampling

Probability and Likelihood Ratios

Forensic Science Example

Table 1: Example calculations for ongoing quality control statistics of central tendency and dispersion using control data for a quantitative method for an attribute where there is a legal specification at "20". Decimal places have been truncated.

Data Set	Describes	Sample Statistic	Calculation and/or Value for this data
20.000 19.997 19.998 19.999 19.997 20.003 20.003 20.001 20.000 20.002 19.998 19.999 20.000 20.003 19.997 20.001 20.000 20.001 20.003 20.001 19.997 20.000 20.002 20.003 20.000	Size	Sample: *n*	= 25
	Central Tendency	Mean: $\bar{x}$	= 20.000
	Central Tendency	Median	= 20.000
	Central Tendency	Midrange	= 20.000
	Central Tendency	Mode	= 20.003
	Dispersion	Range	= 0.006
	Dispersion	Interquartile Range	= 0.003
	Dispersion	Standard Deviation: s	= 0.002
	Dispersion	Variance: s^2	= 0.000004

Key Takeaways

1. A sample of one or more items is a subset of a population that contains all items under consideration.
2. Mean, median, mode, and midrange are measures of central tendency.
3. Standard deviation, variance, range, and interquartile range are measures of dispersion.
4. Both a descriptor of central tendency and of dispersion are required to adequately describe a data set.

Learn More

Khan Academy – search for statistics

ASQ's Introduction to Statistical Concepts

ANAB's Uncertainty, Sampling and Data Analysis: Understanding Statistical Calculations (course)

ASTM E2586-19e1 Standard Practice for Calculating and Using Basic Statistics

Glossary

B. Method Performance Statistics

Introduction

Empirical data from method validation and method verification can be used to calculate statistics that describe a method's performance. The statistics used may vary by the type of method, but all will fall into one or more of these four areas: precision, bias, sensitivity, and specificity. These statistics establish method limitations and, therefore, provide limitations for the results produced from using that method. These statistics can also assist in identifying what is required for ongoing quality assurance and aid in assessing measurement uncertainty or error rates.

Precision and Bias

Statistics provide information on precision (variability) and bias (trueness). Any repeated method will show variability if the mechanism for gathering data is sufficiently discriminating. Evaluation of bias requires a known or reference value. This may not be available in all circumstances.

Figure 1. Precision and Bias

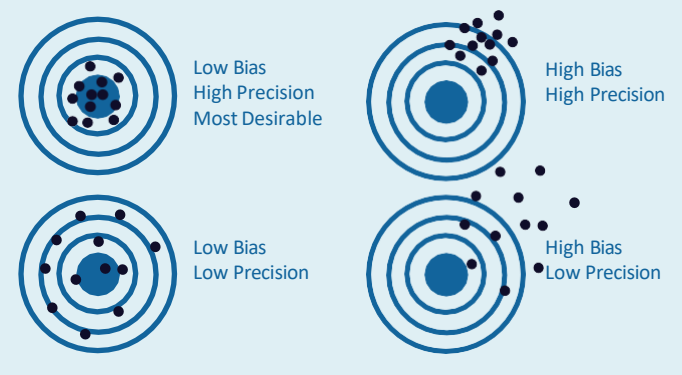


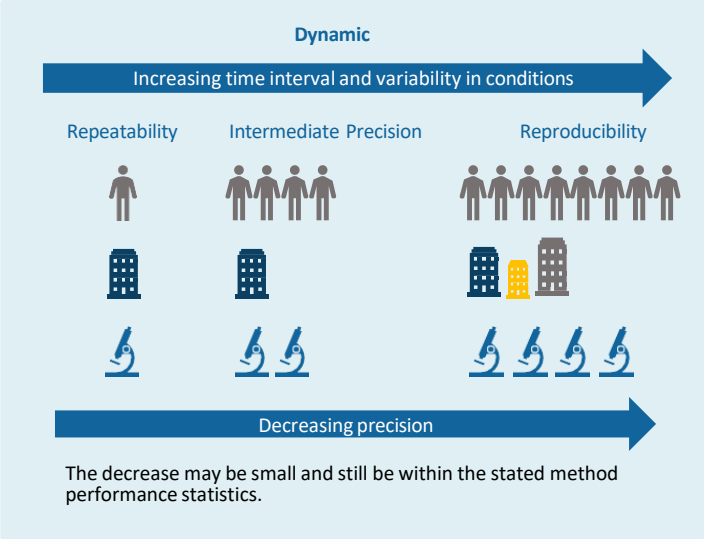
Figure 1 depicts the difference between precision and bias. The bull's eye on the target is the known or reference value. The scatter of the many attempts to hit the bull's eye shows the variability and trueness in this method.

Precision depends only on the distribution of random errors and does not relate to the known or reference value. The measure of precision is usually expressed in terms of imprecision and is commonly computed as a standard deviation or relative standard deviation. A small standard deviation is desirable, as a larger standard deviation reflects less precision, indicating more variability.

Repeatability, intermediate precision, and reproducibility are all terms that relate to precision. They differ by how many "conditions" (e.g., people, equipment, facilities, time) vary in the use of the method (see Figure 2).

To evaluate repeatability, one organization could use the same method, same person, same equipment, etc., to complete method validation. In evaluating reproducibility, more than one organization examines the same samples using the same method with its stated parameters and limitations.

Figure 2. Example of change in precision when the method and samples remain the same



Bias is the difference between the result and the known or reference value. In Figure 1, the difference between each individual attempt and the bull's eye would be calculated, and those values would be averaged to arrive at a value for the method's bias. Low bias is desirable. As depicted in the graph, it is possible to have low bias with either low or high precision.

Selectivity

Selectivity is the ability of a method to discriminate between the attribute of interest and the interferents that are present. If it is documented that the method is selective in the worst-case scenario (e.g., similar colors of fibers, similar chemical compounds, glass fragments of similar refractive index), then it is assumed to perform well in less challenging scenarios. Understanding the samples chosen to evaluate selectivity is important.

Often more than one measurement principle is combined in a single method (e.g., gas chromatography and mass spectrometry), or multiple methods are combined into an examination scheme. The combination increases the overall selectivity (the discriminating power) of the examinations performed.

Sensitivity and Specificity

The selectivity of a method, optimized in method validation, impacts sensitivity and specificity statistics and error rate calculations. Sensitivity measures how well a method can identify true positive samples (the true positive rate), and specificity measures how well a method can identify true negative samples (the true negative rate). Typically, false-positive and false-negative rates are also calculated. Data is often displayed as shown in Table 1.

Table 1: Example data and resulting statistic calculations for a method that produces a binary result (e.g., positive or negative)

		True Value	
		# Known positive cases (pc)	# Known negative cases (nc)
Result (Observed)	# Positive results (p)	# True positive results (tp) **186**	# False positive results (fp) **1**
	# Negative results (n)	# False negative results (fn) **4**	# True negative results (tn) **189**
True positive rate, TP (Sensitivity, SS)		tp/pc = tp/ (tp + fn) = 1-FN = 186/190 x 100 = 97.9%	
False positive rate, FP (Type I error)		fp/nc = fp/ (tn + fp) = 1 – TN = 1/190 X 100 = 0.5%	
True negative rate, TN (Specificity, SP)		tn/nc = tn/ (tn + fp) = 1 – FP = 189/190 x 100 = 99.5%	
False negative rate, FN (Type II error)		fn/pc = fn/ (tp + fn) = 1 – TP = 4/190 x 100 = 2.1%	

Note: This type of table can be generated for qualitative methods (e.g., pattern disciplines) based on a variety of principles.

It is important to note:

- For sensitivity, it does not matter how many false positives were encountered or how many known negative cases were evaluated because they are not part of the calculation. With either 5 or 500 known negative cases, the calculation for sensitivity would be the same and result in the same statistic. The opposite is also true; the calculation of specificity is only impacted by the number of, and results for, the known negative cases and is not impacted by the number of known positive cases.
- The estimates for specificity and sensitivity depend on the number (more or less) and complexity (easy, medium, difficult) of samples tested.
- When there are three or more possible results, a table of frequency data with no calculation of FP and FN may be the most helpful. If "inconclusive" is a possible result and TP, FP, TN, and FN are calculated, it is important to know whether an inconclusive result was defined to be correct, defined to be an error or if the inconclusive result was not included. The calculation and value of TP, FP, TN, FN will be impacted by how inconclusive results are defined.
- When looking at sensitivity and specificity data, it is important to have all four statistics (i.e., TP, FP, TN, FN).
- It is important to understand how the samples used to calculate method performance statistics represent the range of samples allowed to be examined by the method.
- The limit of detection (LOD) or limit of quantitation (LOQ) may be used as a decision point to determine a positive and negative result, or the decision point may be set at some concentration above the LOD or LOQ.
- For multiple methods in a scheme, the methods may be set up so that the method has a higher false-positive rate (lower specificity) to pre-screen samples, thereby forcing more samples to the scheme's second method that will be set-up to minimize the number of false-positive results (high specificity).

Key Takeaways

1. Method validation and method verification require the method performance criteria to be evaluated and known.
2. Method performance statistics are specific to the method. If method parameters are changed, method performance statistics will need to be recalculated.
3. Method performance statistics provide information on a method's precision/bias (variability/trueness).
4. Method performance statistics establish method limitations and, therefore, result limitations from using that method. These statistics can also assist in identifying what is required for ongoing quality assurance and aid in assessing measurement uncertainty or error rates.

Related Primers

General Statistics

Measurement Uncertainty, Frequency Data and Error Rates

Method Validation and Method Verification

Learn More

Eurachem Guide: The Fitness for Purpose of Analytical Methods

Eurachem/CITAC Guide: Assessment of performance and uncertainty in qualitative chemical analysis

Chapter 9, S. L. R. Ellison, V. J. Barwick, T. J. Duguid Farrant, Practical statistics for the analytical scientist, A Bench Guide.

Glossary

C. Statistical Sampling

Introduction

Statistical sampling methods are used in forensic science when all items in a population will not be examined and there is a need to make an inference to that population. The use of statistical sampling methods do not guarantee that a sample is representative of the population but do support representativeness of the sample. If a sample is not representative, the usefulness of the inference being made about the population may be negatively impacted. The relevancy of the population to the question being asked is the topic of the primer on Population Statistics.

Sampling

The initial decision of whether an item will be sampled for examination (testing, inspection, or analysis) occurs at a scene. Decisions are made based on the case information available, the items available at the scene, what items may best be transported to the facility for examination, and the questions considered at the scene investigation. Further sampling may occur once an item is submitted to a forensic science service provider (FSSP). Whether sampling will have a non-statistical or statistical basis is determined by the anticipated use of the results.

Non-Statistical Sampling

Forensic science has many instances where examination results from a non-statistical sampling can answer the question being asked. Examples include:
Item 1 = 38 baggies of white powder

Question: Does Item 1 contain cocaine?

One baggie is examined and found to contain cocaine. That test result is reported. The report makes no inference about the contents of the 37 baggies that were not examined.

Item 1a = 15 fragments of material were collected from Item 1, a jacket

Question: Are any of the fragments glass?

Two fragments were examined and found to be glass. Those test results were reported. The report makes no inference to the other thirteen fragments of material not examined.

Statistical Sampling

There are other instances where the number of items in a finite population can be large, and it would be burdensome to examine all items. Statistical sampling allows less than all items to be examined and for an inference to be made about all items in the population or a large number of items in the population. Example scenarios include:

Item 1 = 200 small baggies containing a white powder

Questions:

- Is cocaine present in at least 100 of the 200 baggies?
- Is cocaine present in all 200 baggies?
- What is the net weight of the powder in all 200 baggies?

Risks and Opportunities

The decision to use statistical sampling has both risks and opportunities (see Table 1).

Table 1. Comparison of risks and opportunities when using a statistical sample

Risks When Using Statistical Sampling	Opportunities When Using Statistical Sampling
• Not all items in the population will be examined • The sample may not be representative of the population; there may be a bias in the sample chosen • A poor or incorrect decision may be made	• Resource savings (e.g., personnel time and consumables) • Reduced handling potential hazardous materials • More timely reporting • Can report on both the items examined and an inference to the population

How a Sample Will be Selected

Non-statistical sampling: The number of sampling units that will answer the question being asked are selected. The sampling units may or may not be selected at random.

Statistical sampling methods available, include:

Simple random sampling – The units selected are chosen at random and each has an equal chance of being selected.

Stratified sampling – The population to be sampled is first divided into mutually exclusive (homogenous) subsets or strata, and independent units are selected from each stratum.

Cluster sampling – The units available for sampling are separated into clusters, groups, that may or may not be homogenous. Clusters included in the sample are identified, and all units in those clusters are examined, or further sub-sampling is performed within the clusters.

Area sampling – A map serves as the sampling frame, and small land plot units are selected from the larger area.

Composite sampling – Combine a large group of units and conduct a single examination.

Random selection of the sampling units is required for statistical sampling (see Figure 1). Statistical sampling risk can be minimized with a proper method choice. For example, if the units available for sampling are heterogeneous, stratified sampling, if done properly, will lead to a valid inference about the population.

Figure 1. Comparison of statistical vs non-statistical population sampling

Non-statistical – Can only provide information about the sample

Statistical – Can make an inference about the population

Determining Sample Size

Choosing a statistical sampling method provides *how the units included in the sample will be chosen.* A statistical approach must also be taken when determining *how many units will be included in the sample* to reduce the statistical sampling risk. The number of units selected from the population is called the sample size. The sample size directly impacts the inference that can be made about the population, so more is generally better. Still, there is a point of diminishing returns where examining more units from the population does not meaningfully improve the inference.

Decisions must be made on the level of confidence in the required inference. A challenge with statistical sampling is that people may differ in their opinion on what level of confidence is required to answer the legal question being asked. Either a frequentist or Bayesian approach can be used to determine the sample size at a stated level of confidence.

Table 2: Example using a hypergeometric approach, frequentist perspective, to determine the sample size that is necessary to support an inference that at least 500 packets from a finite population of 800 packets contain cocaine.

Confidence level	Sample size if all packets examined are positive for cocaine	Sample size if two units examined are negative for cocaine
68%	3	9
90%	5	13
95%	7	15
99%	10	19

Note: Increasing the confidence level that 500 out of 800 packets are positive for cocaine increases the required sample size.

Table 3: The scenario has been revised. Now the sample size must support an inference that at least 700 packets out of the 800 packets contain cocaine.

Confidence level	Sample size if all packets examined are positive for cocaine	Sample size if two units examined are negative for cocaine
68%	9	27
90%	17	40
95%	22	47
99%	34	62

Note: Increasing the proportion of the population that is positive for cocaine from 500 out of 800 to 700 out of 800 increased the sample size required at each level of confidence.

A Bayesian approach considers the forensic science practitioner's belief, their confidence, that the units in the population have the characteristic of interest (e.g., the units do contain cocaine, or the units do contain child pornography). The information that is considered is unique to the forensic science practitioner based on their experience and case-specific information. There is no one correct way to quantify belief, confidence. Other people may not share the same belief and would calculate a different sample size. Generally, using a Bayesian approach will yield a smaller required sample size than the hypergeometric approach.

Uncertainty in Statistical Sampling

When a sample is used to estimate a population statistic, it will differ from the true population statistic (e.g., $\bar{x}$ to estimate μ). How much it differs can be expressed as an uncertainty, also referred to as sampling variability or sampling error, that can be calculated to estimate the impact of bias and variability in the sample selected. Many different samples can be drawn from the population, and each would lead to a different estimate of the population statistic. By looking at the variability in the statistic based on a number of different samples, the uncertainty, or standard error, can be calculated and communicated along with the statistic. If sampling is multi-stage, then any estimate of uncertainty will need to look at all stages of sampling.

When a hypergeometric approach is used, as in the example provided, uncertainty is expressed by the confidence level and accounted for in spreadsheets using formulas designed for this purpose.

Key Takeaways

1. Statistical sampling methods support representativeness but do not guarantee that a sample is representative of the population. If the samples are not representative, the usefulness of the inference being made may be negatively impacted.
2. A "good sample" to answer one question may not be a good sample to answer a different question.
3. Statistical sampling allows for an inference, with a frequentist confidence statement, to be made about a population.

Related Primers

General Statistics
Population Statistics

Learn More

ASTM 2548-16 - Standard Guide for Sampling Seized Drugs for Qualitative and Quantitative Analysis

CSAFE Sampling for Forensic Practitioners Short Course

CSAFE: Statistical Thinking for Forensic Practitioners

ENSFI DWG Qualitative Sampling Calculator

UNODC Guidelines on Representative Drug Sampling

Glossary

D. Population Statistics

Introduction

Population statistics provide information about attributes, such as variability, in a population. Population statistics are almost always estimates based on one or more population samples. When inferring from a sample to a population, the value of the inference depends on how the sample was selected, sample size, number of samplings, and population's relevance to the inference topic.

Population and Samples

A population is an entire group about which specific information will be obtained. A population can be finite, but more commonly, the population is very large, infinite, or continuing to emerge, making it impractical or impossible to obtain values for all items in the population. This forces the use of a sample, a subset of the population, to make an inference about the population.

Choice of Population

Relevant, related, representative, and appropriate are all terms used to evaluate a choice of population. A relevant population for one question may be less relevant for another, even if the questions are very similar. Representative samples may be incomplete or non-existent. In these situations, the sample may be limited to an individual forensic science service provider's data or even an individual forensic science practitioner's data. People may have different thoughts on whether a sample is representative or when a population no longer provides value, is no longer relevant, in evaluating a question.

Factors considered when evaluating the relevance of a population may include, but are not limited to, geographical and item type considerations.

Example: STRidER (STRs for Identity ENFSI Reference), launched in 2017, is a freely accessible functional forensic online curated DNA short tandem repeat (STR) allele frequency population database that provides STR genotype rarity estimation and quality control of autosomal STR data. Queries can be done for 1 to 41 loci, limited to a specific kit, and for varying geographical regions.

Table 1: The impact to the relative frequency estimate when using a "closed population" of a single country, the more open population for a continent, or the entire database.

Profile Query Submitted: D3S1358:15/17, D8S1179:12/13, D16S539:9/13, D18S51:12/18, D21S11:28/30, FGA:23/27, THO1:9.3/9/3, VWA:15/16	
Origin	Actual Matching Probability
Entire Database	8.9826e-13
Asia	5.0408e-15
Europe	2.4598e-12
Belgium	1.2908e-11

STR databases that are used in the United States include:

- Corrigendum to U.S. Population data for 29 Autosomal STR Loci
- Erratum Population data on the thirteen CODIS core short tandem repeat loci in African Americans, US Caucasians, Hispanics, Bahamians, Jamaicans, and Trinidadians

Example: Paint Data Query (PDQ) is an international original equipment manufacturer (OEM) automotive paint database used to determine the make, model, and year range of a paint sample from a scene based on the chemical compositions of the paint system (clear coat, base coat, undercoat(s)). Data fields (properties) can be searched individually or in combination.

Geographical considerations:

- Contains entries for most domestic and foreign vehicles marketed in North America
- Automotive industry is a global market - Mitsubishi, Toyota, and Nissan vehicles assembled in Australia have been found to have comparable paint systems to those same or corresponding vehicles marketed in North America

Item type considerations:

- Updated yearly
- Year of manufacture (1973 to present)
- Representation of chemistry diversity
- Representation of every automotive paint system

Population Statistics Symbols

Population statistics, often called parameters, have different symbols than the same statistic for a sample. It is important to understand the symbols used to ensure you understand the meaning of the statistic. Population and sample statistics are given Greek and Roman letters, respectively, to aid in distinguishing them.

Table 2: Symbols for population and sample statistics

Meaning	Symbol
Population size	N
Sample size	n
Sampling fraction, f = n/N	f
Average of measurement in a population	μ
Average of measurement in a sample	$\bar{x}$
Standard deviation of measurement in a population	σ
Standard deviation of measurement in a sample	s
Variance of a population	σ^2
Variance of a sample	s^2

To make an inference from a sample to a population, one must choose the sample statistically. When a sample is used to estimate a population statistic, it will differ from the true population statistic (e.g., $\bar{x}$ to estimate μ). Many different samples can be drawn from the population, and each would lead to a different estimate of the population statistic. Generally, more data from a larger sample size or multiple samples will lead to a better estimate of the population statistic and a more certain inference. However, if the population is not relevant to the question being asked, then more data will not improve the inference.

Making an Inference

Can a sample of 1000 be used to make an inference (estimate) of the relative frequency for a population of thousands, hundreds of thousands, or millions or even trillions? The short answer is – it depends.

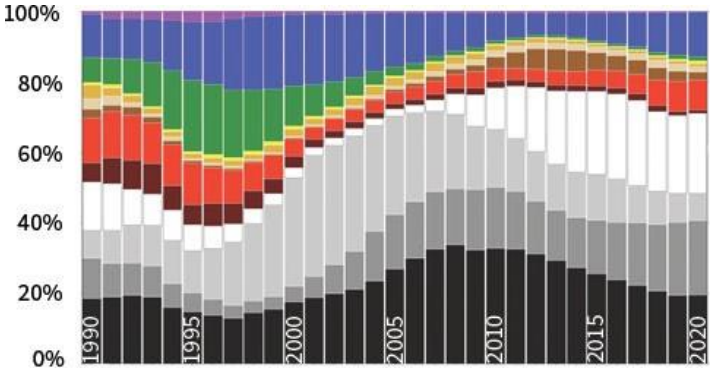
To evaluate the estimate of a population statistic, an inference, one must know the sample size, how the sample was collected, and how the population relates to the question being asked. An inference may be acceptable (e.g., acceptable limitations, acceptable risk) in one scenario and not acceptable in another. "Assessing the adequacy of an inference is never a purely statistical matter in the final analysis because the adequacy of an inference is relative to its purpose and what is at stake in any particular context in relying on it." [1]

Example: The color of cars in an airport parking lot was used to estimate the color of cars in all U.S. parking lots

To evaluate the value of this inference to a specific scenario, one might look at the following:

Car color over time

Figure 3: Car Colors by Year from 1990–2020, The color band reflects the percentage of all sales each year. (Source: thedrive.com)



Due to the significant shifts in car color sales data over time, any inference may need to be made from a recent sample (i.e., yellow is more stable than white over time).

Car color by location

The percentage of non-white or black cars ranges from 30% to 17.1% across the 50 states.

Therefore, the sample may or may not reflect the proper state of interest in the U.S.

How the sample was collected

Sales data indicates that the number of cars by state in the U.S. ranged in 2020 from 14.2 million cars in California to 171 thousand in Alaska.

Which state, time of year, time of day, and how the random sample was collected will all impact the population estimate

"As forensic tests become more sensitive, and the amount and complexity of scientific evidence increases, there will be a need for more sophisticated models and statistics to obtain meaningful inferences and interpretation of the evidence given the specific circumstances of a case. This includes the generation of ground truth datasets (where the provenance of the data is known) relating to trace evidence and pattern evidence, as well as an understanding of the transfer, persistence, recovery, and background abundance of materials in general and in case-specific circumstances." [2]

Key Takeaways

1. **Populations are dynamic.**
2. **Any population statistic based on a sample is an estimate.**
3. **Understanding how well the sample represents the population is necessary to determine the value of any inference being made.**

Related Primers

General Statistics
Statistical Sampling

Learn More

Advanced Topics in Forensic DNA Typing: Interpretation; John M. Butler; 2015; Elsevier Inc.

The Art of Statistics; David Spiegelhalter; 2021; Basic Books

The Royal Society – The use of statistics in legal proceedings

The Royal Statistical Society Guides – Fundamentals of Probability and Statistical Evidence in Criminal Proceedings and Statistics and Statistics and Probability for Advocates: Understanding the use of statistical evidence in courts and tribunals

Glossary

Footnote:
1. The Royal Statistical Society; Fundamentals of Probability and Statistical Evidence in Criminal Proceedings - section 2.6
2. The Royal Society – The use of statistics in legal proceedings; Section 5

E. Probability and Likelihood Ratios

Introduction

Statistics and probability are both useful tools for interpreting data and observations. They have some overlapping features, but they are conceptually different. Those who support the use of probability and likelihood ratios (LRs) in forensic science opinions see them as an application of logical reasoning with uncertainty. Opposition to using probability and LRs in forensic science opinions focuses on the subjectiveness (variability) in the LRs that are provided, the use of LRs that do not align with performance data and misunderstanding of the expression.

Statistics Versus Probability

Statistics are derived and extracted from the empirical world, whereas probability projects itself onto the empirical world.[1]

Statistics
- Applied Science
- Frequency of past events
- Collects and summarizes empirical data from observed events
- Estimates a characteristic of a population based on a sample
- Given an observation, what processes/models would explain the observation
- Inductive
- Uncertainty is expressed by a level of confidence

Probability
- Theoretical Science
- Assessing the probability of future events or past events in doubt
- Begins with general assumptions and quantifies outcomes
- Given a process/model with stated assumptions that align with the individual's beliefs, what are the probabilities of possible outcomes
- Deductive
- An individual's uncertainty is expressed in the probability value itself on the continuum between zero and one

Probabilities and Likelihood Ratios

Although probability is often thought to be an objective measure because it is expressed as a number, it is in fact subjective. Probability measures the strength of an individual's belief that the event will occur or did occur and is subjective because it relies on differing information available and assumptions made to arrive at the probability number. There is no single, correct probability for the possibility that an event will occur. In theory, an infinite number of combinations of assumptions can be made that would result in an infinite number of probability values. No probability is right or wrong, although some probability assignments may be considered more reasonable than others.

A likelihood ratio (LR), a ratio of two mutually exclusive probabilities, is also subjective.

Figure 1. Calculation for likelihood ratios

$$Likelihood\ Ratio = \frac{Probability\ 1}{Probability\ 2}$$

In forensic science, the LR is typically the ratio of the probabilities of the evidence under two alternative propositions.

Figure 2. Calculation

$$R = \frac{Probability\ of\ the\ Evidence\ |Prosecution\ Hypothesis\ (H_1)}{Probability\ of\ the\ Evidence|Defense\ Hypothesis\ (H_2)}$$

A probability is always a non-negative number between zero (a certainty of *non*occurrence) and one (a certainty of occurrence). Values between zero and one represent the uncertainty in the individual's belief of a particular event occurring. Probability can also be expressed as a percentage between 0% and 100%.

Since probabilities are a non-negative number, an LR will also be a non-negative number. An LR will have a value between zero and infinity (see Figures 1 and 2).
- If Probability 1 approaches zero, the LR will approach zero
- If Probability 2 approaches zero, the LR will approach infinity
- If Probability 1 and 2 are equal, the LR will equal 1

LR – Frequentist and Bayesian Approaches

Although sometimes referred to as the Bayes factor, an LR can be assigned using either a frequentist or Bayesian approach. Regardless of how it is assigned, an officer of the court will need to make their own evaluation of any LR provided.

A frequentist approach relates probability to the frequency of observing an event in a large number of experiments where the event may or may not occur. Inferences are based on random sampling. Frequentist statistics focus on the probability (Pr) of the evidence (E) given a hypothesis (H), and is expressed as Pr(E|H).

Under a Bayesian approach, the probability of an event occurring is defined as the degree of one's belief in the truth of the proposition that asserts that the event will occur. Inferences can be made even if the information is incomplete using probability distribution models to describe a belief. The Bayesian approach helps an individual quantify their personal probability assessment of a hypothesis given the evidence, Pr(H|E).

For both, any single LR is looking at two mutually exclusive hypotheses (H_1 and H_2) but to be exhaustive, within the context of the case, additional LRs (H_1 vs H_3, H_1 vs H_4, and so on) may need to be assessed.

What Adds Subjectivity?

Combinations of different assumptions or different levels of belief can result in the same probability or LR (e.g., one person's probability of 80% may not mean the same as another person's probability of 80%). Conversely, two people with the exact same information may not provide the same probability or LR.

Footnote:
1. RSS: Fundamentals of Probability and Statistical Evidence in Criminal Proceedings, section 1.5.

In forensic science, some aspects can vary between forensic science practitioners (FSPs) within a single forensic science service provider (FSSP) or between FSSPs that will result in variation in the LR produced (see Table 1). Standardization may help reduce the variability of the LRs produced.

Table 1 Aspects that add variation to a probability, to an LR

Aspect	***Possible Variability***
Data	• Type of data (e.g., feature, attribute, characteristic, profile, spectrum) • Source of data
Method of examination	• Different technology • Different operating parameters
Method of data interpretation	• Degree of standardization between FSPs • Degree of standardization between FSSPs
Calculations performed	• Done without calculation • Calculated by a human or an algorithm • Distribution chosen to represent the data • Model used
Additional information	• Case Information available, used, not used • FSP knowledge & experience • FSSP data

Figure 3 demonstrates that multiple models can meet a set of criteria for reasonableness and produce an LR. When the evaluation of the model includes an assessment of reasonableness based on the LRs produced for known samples, the number of models that produce reasonable LRs may diminish. Models that "pass" will still produce different LRs for the same data.

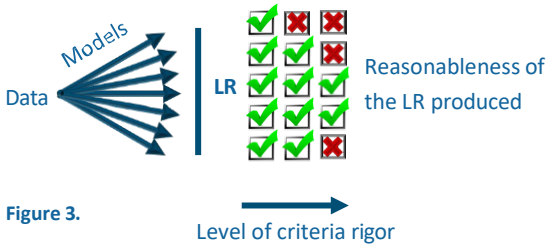


Figure 3.

Example 1: Interlaboratory study to evaluate the forensic analysis and interpretation of glass evidence: Standardized technology (LA-ICP-MS), method (ASTM E2927), case information, model, and samples. LRs: Supported the correct hypothesis with some indicating varying strength for an association.
Example 2: GHEP-ISFG collaborative exercise on mixture profiles (GHEP-MIX06): Two mixture profiles in a paper format. The data interpretation method was specific to the participant. With the same LR model, LR varied by up to 12 orders of magnitude.
Example 3: Collaborative Testing Service, Inc. DNA Interpretation Test No. 22-5881 Summary Report – LRs provided ranged from 10^9 to 10^{28} for Item 3 and 10^7 to 10^{26} for Item 4.

LR as a Numerical or Verbal Scale

Since the numerical value of an LR can range from zero to infinity, an LR is at times translated to, and accompanied by, a strong support, extremely strong support) to convey the magnitude of the LR. The inverse of the scale is used to represent the alternative proposition.

Strengths and Limitations of an LR

Strengths:

- Is a mechanism to express the relative likelihood of two competing hypotheses
- Can be assigned with or without empirical data; therefore, there is no limitation to when an LR can be provided
- Can be evaluated and calibrated, for consistency (over or understatement) with known samples
- Can be limited (e.g., a lower and upper limit) to that supported by available data
- Can be evaluated for usefulness to a specific case when context information and performance data demonstrating how often this level of LR has occurred in both H_1 and H_2 true cases is provided

Limitations:

- There is no "true value" or "correct value"
- Is subjective because of personal or organizational choices
- Can have values that vary by orders of magnitude
- Other models may be as "good" or even "better"
- Is difficult to have sufficient data from known samples to evaluate the reasonableness of the LR values produced
- An officer of the court will need to determine if an LR provided by an FSP aligns with their own evaluation of the performance data and their personal beliefs

Key Takeaways

1. Probability will have a value between 0 and 1; an LR will have a value between 0 and infinity.
2. Probabilities and LRs are personal and will vary depending on the data (empirical or personal belief), assumptions made, and calculation model used.
3. An officer of the court or the trier of fact will need to determine, either informally or formally with calculations, their own value for a probability or LR.

Related Primers

Algorithms
General Statistics
Population Statistics

Learn More

Likelihood Ratio as Weight of Forensic Evidence: A Closer Look & Response & Bayesian Reasoning and Evidence Communication

The Royal Society – The use of statistics in legal proceedings

The Royal Statistical Society Guides

Glossary

F. Measurement Uncertainty, Frequency Data, and Error Rates

Introduction

All results produced by a forensic science service provider have an associated uncertainty. When the result is a measurement, that uncertainty is expressed as a quantitative value. When the result is not a measurement, the uncertainty can be expressed as frequency data or as an error rate for different possible results. All assist the user of the information in understanding the limitations of the reported results.

Measurement Uncertainty (MU)

Measurements

Examinations may include one or more measurements that determine the numerical magnitude of a characteristic of interest (e.g., weight of a substance, diameter of a cartridge case, concentration of ethanol), referred to as the measurand.

If a measuring process has sufficient resolution, repeated measurements may produce different quantity values. In theory, increased resolution to an infinite number of decimal places is possible, making it impossible to ever know the "true value" of a measurement.

An evaluation or estimation of MU quantifies what is known about a specific measurement process and the quantity being measured at a stated level of confidence. A complete measurement result consists of the quantity value and the MU.

Components of a Measurement Process that Impact Uncertainty

Regardless of the measurement process, the significant contributors to MU are common for all types of measurement processes and include potential sources of both random and systematic error:

- Equipment calibration or certified reference materials used to establish metrological traceability
- Sampling, if performed during the measurement process
- Precision data: Includes human factors, sample preparation, sample analysis, interpretation of data, or calculations generated
- Bias data and the uncertainty in the reference used to evaluate bias. Assumes that all significant bias has been identified and resolved during method validation

Evaluation or Estimation of MU

An evaluation or estimation of MU quantifies the variation from different contributors in a measurement process. Information about contributors may come from the measurement process itself (e.g., precision data) or outside sources (e.g., calibration certificate). Contributors can be quantified individually (e.g., each aspect of sample analysis), or a single statistic may cover many individual contributors (e.g., long-term precision data). Contributors are expressed as the standard deviation of the probability distribution (e.g., normal, rectangular) that the contributor is based on.

The uncertainties of each contributor are combined according to well-established procedures. The National Institute of Standards and Technology (NIST) has developed an 8-step process for evaluating and reporting MU. This framework conforms to the principles set forth in the Joint Committee for Guides in Metrology (JCGM) Evaluation of Measurement Data-Guide to the Expression of Uncertainty in Measurement (GUM).

Combining quantity values for all contributors yields a standard uncertainty, expressed as a standard deviation, for the entire measurement process. This is commonly "expanded" by a coverage factor, *k*, to define a range that encompasses a large fraction of the values within which the quantity being measured is expected to lie at a chosen confidence level (e.g., 95%, 99%).

Table 1 Generalized Evaluation of MU

Measurement:			Measurement being made				
Range of measurement:			If applicable (e.g., weight, concentration)				
Measurement process:			Process identifiers (e.g., method name, revision)				
Evaluation prepared by:			Name/Section/Discipline				
Uncertainty Contributor	**Value**	**Units**	**Distribution**	**Type**	**Divisor**	**Degrees Freedom (n-1)**	**Standard Uncertainty**
Equipment Calibration Certified Reference Material Certificate Uncertainty	0.005	unit	normal	B	2.00	∞	0.0025
Measurement Process Reproducibility	0.092	unit	normal	A	1.00	>100	0.092
Uncertainty of Reference used to Evaluate Bias	0.005	unit	rectangular	B	1.73	∞	0.0029
Combined Standard Unc	u_c *(using root sum of squares calculation)*						0.0921
Expanded Unc at 95.45%	U (k = 2.000)						0.1842
Expanded Unc at 99.73%	U (k = 3.000)						0.2762

Frequency Data and Error Rates

When the results are not numerical, the data can be summarized using frequency of occurrence data for each possible result.

When there are only two possible results, error rates can be unambiguously defined as the portion of samples in which the examination led to an incorrect (false) result. The false-positive rate (FP) estimates the frequency or probability of obtaining an erroneous positive result when the known result is negative. The false-negative rate (FN) represents the probability of obtaining an erroneous negative result when the known result is positive. This frequency data is commonly presented in a 2x2 table (see Table 2).

When three or more results are possible from a method (e.g., same source, different source, inconclusive), an error rate is not unambiguously definable. When an error rate is provided, someone has decided how to handle the additional result types. There is no single approach used for this decision; therefore,

multiple calculations of error rate can arise from the same set of data. Some may decide to not include the additional type of results in the calculations. Some may decide to view them as false positives; others will decide to view them as false negatives.

When there are three or more possible results, a table of the frequency data with no calculation of FP and FN may be the most helpful (see Tables 3 and 4).

Example data and calculation of false positive and false negative rates with a varying number of possible result types

Table 2. Two Possible Result Types

	Same Source	Different Source	Total Cases
Known Same Source	302	0	302
Known Different Source	0	215	215

FP Rate = 0.0%; FN Rate = 0.0%;

Table 3. Three Possible Result Types

	Same Source	Inconclusive	Different Source	Total Cases
Known Same Source	302	13	0	315
Known Different Source	0	10	215	225

With Inconclusive not included: FP Rate = 0.0%; FN Rate = 0.0%
With inconclusive results looked at as incorrect = FP Rate = 4.4%; FN Rate = 4.1%

Table 4. Nine Possible Result Types

	VSI	SI	MI	LI	N	LE	ME	SE	VSE	Total Cases
Known Same Source	302	76	75	38	13	4	0	0	0	508
Known Different Source	0	0	1	9	10	99	107	67	215	508

Available result types:
Very Strong Inclusion through Strong, Medium, Low Inclusion; Noninformative; Low Exclusion, Medium, Strong and Very Strong Exclusion

If only VSI and SI are viewed as "correct," then there are no false positives, but there are 130 false negatives. (FPR=0%; FNR=25.6%). If VSI, SI, MI, and LI are all viewed as correct inclusions, then there are 10 false positives and 17 false negatives. (FPR=2%; FNR=3.3%).

Determination of error rates has several challenges and limitations:

- An average error rate is seldom correct for any single situation (e.g., individual types (difficulty) of samples encountered, examiner, or both).
- An error rate can only be applied to the question that the method is answering (e.g., classification versus individualization).
- The data available that is relevant to the sample type and to the examiner, and therefore to a specific case, may be limited. There will be significant uncertainty to any error rate provided based on limited data.

Relationship to Quality Control

Having a MU, frequency data, or an error rate does not preclude the need to review the examination performed in a specific case and perform the appropriate quality control functions to ensure that the method continues to function as demonstrated in validation studies.

Relationship to Accreditation Requirements

ISO/IEC 17025, the standard used most frequently for accreditation of forensic science service providers, does contain requirements related to the evaluation of MU and the determination of method performance characteristics. There is no mention of or requirement to calculate an error rate.

Forensic Science Examples

Documentary standards related to the evaluation and estimation of MU for forensic science methods are available on the OSAC Registry.

Key Takeaways

1. Measurement uncertainty quantifies the variation in a measurement process at a stated level of confidence.
2. Calculation of an error rate is applicable only when you have two possible result types.
3. An average error rate may not apply to a case-specific sample type or examiner.
4. Frequency data allows for evaluation of method performance for each possible result type.

Related Primers

General Statistics
Method Performance Statistics
Metrological Traceability
Communication of Limitations

Learn More

NIST SOP 29-Standard Operating Procedure for the Assignment of Uncertainty (part of **NISTIR 6969**)
Joint Committee for Guides in Metrology (JCGM) Evaluation of Measurement Data-Guide to the Expression of Uncertainty in Measurement (GUM)
Strengthening Forensic Science in the United States: A Path Forward, Chapter 4
Report to the President: Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods
Swofford et al., (2024). Inconclusive decisions and error rates in forensic science. FSI Synergy, Vol. 8, 100472

Glossary

III. Forensic Science: Communicating Results Using Reports

Introduction

Communication is the process by which information is exchanged between individuals. The process uses a common language and terminology specific to the particular application.

Communication is a two-way activity consisting of seven major elements: sender, encoding, message, channel (with or without noise), receiver, decoding, and feedback (figure 1).

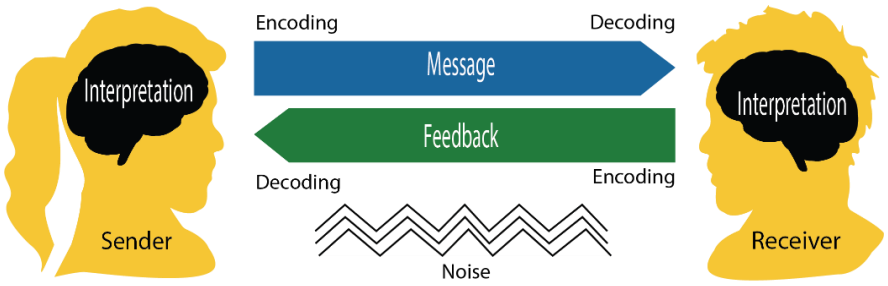


Figure 1. Encoding/Decoding Information Flow

The responsibility for clear communication usually falls on the sender, but the receiver is also responsible for confirming a clear understanding of the message. In the communication process, getting feedback is viewed to be as important as sending the message. It is through feedback that the sender ensures that encoding (word selection) and decoding (understanding) happen correctly. Until this is achieved, the communication process is dynamic and cyclical. It is after feedback that the process of communication is viewed as complete.

The communication process for reporting forensic science results faces a basic process challenge. The process becomes a one-direction linear flow because there is no feedback loop to ensure that each individual reader of a report has correctly understood the message of the sender, the forensic science service provider (FSSP). The readers of the reports are diverse, varying in age, culture, background knowledge, education, and experience, and there may be multiple reports related to the same item or the same examination. Much of the time, an officer of the court will take their understanding of the message, the report, and move forward.

The risk associated with a lack of feedback is significant. The risk is reduced when documentary standards are used as the basis for accreditation with requirements for accurate, clear, unambiguous, and objective reports and when report elements are standardized. The risk is further reduced when receiver feedback is considered and incorporated into a quality assurance (QA) system. This is true whether the feedback is provided on an individual report or a standardized report format used by the FSSP.

The primers in this category cover aspects of how the "message" is encoded and sent in the reporting of forensic science results.

Primer Topics In This Category

A. Forensic Science Reports

FSSPs are responsible for ensuring their reports are understandable. This is necessary because people with varying levels of topic knowledge asking different questions may read the reports near the time they are produced or many years later. Reports are not standardized, but they often share common elements based on accreditation requirements. Understanding the types of reports and categories that may be provided can increase the recipient's comprehension of the report.

B. Communicating Limitations

Results of forensic science examinations are used for making decisions that have significant consequences. Therefore, it is important to understand that all methods of examination have limitations (e.g., measurement uncertainty, variability, possible sources of error). Science is a "work in progress." What is known today may only be applicable to a particular context or population, leaving unanswered questions. Science will continue to evolve using ever-increasing sophisticated technology, and this may change the nature of the limitations, but limitations will still exist. Understanding the limitations of reported results allows the decision-maker to assess the value of the information provided in a specific case. Information on the limitations of a method or a result will allow the comparison of results from different FSSPs, the comparison of a result to a legal specification, and the comparison of differing opinions, as well as help the decision maker identify what information or knowledge is not available to them.

A. Forensic Science Reports

Introduction

Forensic science service providers (FSSPs) are responsible for ensuring their reports are understandable. This is necessary because people with varying levels of topic knowledge asking different questions may read the reports near the time they are produced or many years later. Reports are not standardized, but they often share common elements based on accreditation requirements. Understanding the types of reports and categories that may be provided can increase the recipient's comprehension of the report.

Communication

If the effectiveness of preparing the report (encoding) is maximized, then the effort that the report recipient must make in interpreting (decoding) the report can be minimized, thereby reducing the risk of misunderstanding by recipients (figure 1).

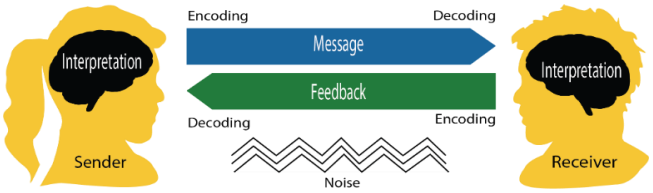


Figure 1. Encoding/Decoding Information Flow

Putting a Report Together

There are accreditation requirements related to the reporting of results. "Result" is a broad, encompassing word that includes the observations, data, and calculations from examining an item (i.e., testing, calibration, inspection), the interpretation of the same, and any opinion that is reported.[1] ISO /IEC (International Organization for Standardization/International Electrotechnical Commission) 17025 General requirements for the competence of testing and calibration laboratories, clause 7.8.1.2 provides the overarching expectation:

> "The results shall be provided accurately, clearly, unambiguously, and objectively, usually in a report (e.g., a test report or a calibration certificate or report of sampling) and shall include all the information agreed with the customer and necessary for the interpretation of the results and all information required by the method used."

ISO/IEC 17020 Conformity assessment — Requirements for the operation of various types of bodies performing inspection, clause 7.4.4 similarly mentions correctness, accuracy, and clarity.

Report Elements

Having required elements supports accurate, clear, unambiguous, and objective reporting of results. Both ISO/IEC 17025 (figure 2) and ISO/IEC 17020, which serve as the basis for accreditation of FSSPs, contain required report elements when providing results.

Figure 2. ISO/IEC 17025 report elements:

- Title
- Name & address of FSSP
- Location where work was performed
- Customer contact information
- Unique identification of the report
- Dates of receipt of item, performance of work, report issuance
- Description and identification of the item
- Identification of results provided by external providers
- Information on sampling performed
- Method used as well as additions to, deviations, or exclusion from the method
- Results
- Identification of who authorized the report

An important caveat to note: ISO/IEC 17025 allows for an agreement with the customer to issue a simplified report that will not include all required elements. ISO/IEC 17020 has several of these as optional report elements. Therefore, the elements included in a report from an accredited FSSP can vary by discipline within an organization or between FSSPs.

In addition to these required report elements, there are also requirements for reporting interpretations and opinions. An accrediting body may specify other requirements related to reporting for their customers, or a documentary standard can be published that adds detail to these more general report elements. ANSI/ASB 053 Standard Report Content in Forensic Toxicology, a standard on the Organization of Scientific Area Committees (OSAC) for Forensic Science Registry, and the FBI Quality Assurance Standards for Forensic DNA Testing Laboratories are examples of documentary standards.

Both the National Commission on Forensic Science in recommendations and views documents and the National Academy of Sciences address report content. The NAS supports "complete and thorough" reports but acknowledges that there are limits to the amount of detail reports can reasonably provide. FSSPs have the challenge of determining how much information to include in a report. If something is not clear or

Footnote:

1. OSAC Preferred Terms, Organization of Scientific Area Committees for Forensic Sciences. (2022). Retrieved from https://www.nist.gov/system/files/documents/2021/12/14/OSAC%20Preferred%20Terms_December%202021.pdf

not included, report recipients can request and review supporting records (i.e., chain-of-custody, observations, data, calculations, interpretation), commonly referred to as the case record or case file. The FSSP may or may not include all the contents of the case record with the report. If not provided, the National Commission on Forensic Science recommends the report state that it does not contain all the documentation associated with the work performed.

Report Wording

The terminology used in forensic science often differs from that used in other sectors. Differences in definitions of the same term can also occur between forensic science disciplines. Both the National Academy of Sciences and the National Commission on Forensic Science recommended that the forensic science community establish standard terminology to be used in reporting.

Using standardized wording to report results within an FSSP is becoming more common with the increased use of laboratory information management systems to generate reports. Acknowledging that common terminology and report wording does reduce the risk of misunderstanding, applicable documentary standards are being developed, with many posted on the OSAC Registry. Implementing these voluntary standards is occurring, but until that is universal, terminology and report wording may or may not be standardized between forensic science practitioners (FSPs) within and between FSSPs.

Types and Categories of Reports

Reports can be provided in either oral or written format. The different options within each format are known as channels of communication. One written channel is referred to as a report, but the same content can be provided using other channels as well. The chosen channel can impact the interpretation of the message positively or negatively. Some channels are more prone to "noise" that can interfere with the communication process and influence how people understand the message (figure 3).

Oral	Written
Channels: Phone call, voicemail message, in-person conversation	Channels: Report, memo, email, text message

Figure 3. Oral vs Written Channels

Depending on the work performed by an FSSP, a single item for examination could have multiple reports issued. The FSSP can choose to provide a separate report for each examination performed on the item, or they can choose to group all examinations within a discipline into a single report. Multiple disciplines may also examine the item, thereby generating multiple reports. After the initial reporting of an examination result, these additional types of reports may be issued (figure 4):

Report Type	Purpose
Supplemental Report	To report on additional work
Amended Report	To change a report
Withdrawn Report	To be replaced by a new report

Figure 4. Report Type and Purpose

At times, reports of results are broken into categories that are used to describe the purpose of the report (figure 5):

Report Category	Purpose
Intelligence Report	Provides information to link cases, events, and situations to help design strategies.
Investigative Report	Sometimes referred to as a preliminary or interim progress report, with the intent to assist further inquiry, interview, or strategy.
Technical Report	Reports observations, data, calculations, and interpretations that may or may not include an opinion.
Evaluative Report	Includes an opinion. This opinion is the evaluation of two specified competing propositions.

Figure 5. Report Category and Purpose

Key Takeaways

1. The elements included in a report can vary between disciplines of an FSSP and between FSSPs.
2. Terminology may not be standardized; therefore, it is important to understand how an individual FSSP uses a term.
3. There may be multiple reports (types or categories) for an item selected for examination.
4. Supporting records that serve as the basis of the report may or may not be provided; the receiver of a report may need to request these from the FSSP.

Related Primers

Accreditation and Certification

Documentary Standards

Learn More

FSSPs do not make reports publicly available but may provide examples when requested. Reviewing the format(s) used will aid in understanding and reduce the risk of misunderstanding the information provided.

Glossary

B. Communicating Limitations

Introduction

Results of forensic science examinations are used for making decisions that have significant consequences. Therefore, it is important to understand that all methods of examination have limitations (e.g., measurement uncertainty, variability, possible sources of error). Science is a "work in progress." What is known today may only be applicable to a particular context or population, leaving unanswered questions. Science will continue to evolve using ever-increasing sophisticated technology, and this may change the nature of the limitations, but limitations will still exist. Understanding the limitations of reported results allows the decision-maker to assess the value of the information provided in a specific case. Information on the limitations of a method or a result will allow the comparison of results from different forensic science service providers (FSSPs), the comparison of a result to a legal specification, and the comparison of differing opinions, as well as help the decision maker identify what information or knowledge is not available to them.

Limitations

Although stated slightly differently, both the National Academy of Sciences and the National Commission on Forensic Science support identifying the limitations of methods and results in forensic science reports. Table 1 uses previous primer topics to provide examples of limitations communicated in forensic science reports.

Table 1: Examples of limitations communicated in forensic science reports

TOPIC	LIMITATION	EXAMPLE STATEMENT OR SCENARIO
Accreditation	A reference to accredited status can only be included when the examinations performed are within the scope of accreditation. An FSSP cannot give the impression that they are accredited for work outside the scope of accreditation.	Reference to accredited status in a report **Scenarios** Discipline: Accredited in the disciplines of biology, seized drugs, friction ridge, toxicology, and firearms but not accredited in the discipline of scene investigation Service: Accredited in the discipline of firearms but not accredited to perform trigger pull examination Technology: Accredited in the discipline of toxicology and the use of gas chromatography-mass spectrometry (GC-MS) but not accredited to use liquid chromatography-mass spectrometry (LC-MS)
Sample	Sample acceptance criteria for the method were not met.	**Scenarios** DNA: Quantity of sample is insufficient Latent Print: No print of value for comparison Toxicology: The volume of the sample is insufficient
Method	Quantitative: Limitations on lower and upper limits of quantitation. The compound of interest is present but at a concentration below the method's limit of quantitation (LOQ) or above the method's upper limit of quantitation (ULOQ). Qualitative: The compound of interest is present but at a concentration below the method's limit of detection (LOD) or the FSSP's reporting limit.	**Example** Item 1: Post-mortem blood sample Drug A: None detected at a reporting limit of 5 ng/mL. *Understanding: When tested, the instrument response for Item 1 was less than the instrument response for the calibrator at the decision point of 5 ng/mL.* *There is uncertainty in the instrument response. For a full understanding of the result, the range of possible instrument response values is necessary.* **Example** Item 2: Fire debris sample No ignitable liquid was identified. *Understanding: When tested, there was no instrument response above the method's limit of detection (LOD) for the ignitable liquids included in the method.* *Information on what ignitable liquids are included in the method should be available from the FSSP but may not be provided with the report.* *If there was no instrument response, the instrumentation used in the method may not be sensitive enough to identify an ignitable liquid that is present.*

TOPIC	LIMITATION	EXAMPLE STATEMENT OR SCENARIO
		If there was an instrument response, there is uncertainty in the instrument response. For a full understanding of the result, the range of possible values of instrument response is necessary.
	The method lacks selectivity; it is unable to resolve chemical isomers.	**Example** Item 3: Electronic cigarette cartridge containing light brown liquid Tetrahydrocannabinol (THC) was identified, but the method used in the examination of Item 3 cannot separate delta-8-THC from delta-10-THC. *Understanding: The instrument response for THC could be:* • *100% due to the presence of delta-8* • *100% due to the presence of delta-10* • *Any combination of delta-8 and delta-10*
Statistical Sampling	The confidence level and the proportion of the population to which the interference applies as well as the number of "negative" results that are acceptable (0 – 2) are used to determine the number of items in the population that will be selected and examined by an FSSP. There is a risk of misunderstanding the report statement and mixing up the confidence level and the inference level. Neither the confidence level nor the inference level can be confirmed without examining the entire population.	**Example** Item 1: 100 packages containing white powder Twenty-three packages were examined and found to contain Drug A. The 23 packages were selected using a hypergeometric sampling plan which allows an inference with a 95% confidence level that at least 90% of the 100 packages in Item 1 contain Drug A. *Understanding: There is an estimated 5% risk (95% confidence level) that fewer than 90 (90% inference level) of the packages in Item 1 will contain Drug A.*
Measurement Uncertainty (MU)	There is always MU. The MU may not always be included in a report of a measurement (i.e., weight, length, concentration, force). Accreditation requirements require the reporting of MU when it affects the evaluation of conformity with a specification. If the MU was not included in the report, obtaining the MU will be required to be able to calculate the range of possible results at a stated level of confidence. Different terminology may be used in the report statement: expanded uncertainty, coverage probability, coverage level, level of confidence, confidence interval, and confidence level. The terms confidence interval and confidence level are only appropriate to use when all uncertainty contributors are quantified using data from the measurement process. There is a risk of misunderstanding the report statement.	**Example** Item 1: Ante-mortem blood specimen The specimen was found to have an ethanol concentration of 0.082 g/dL +/- 0.004 g/dL at a 99.7% coverage probability. *Understanding: If one were to repeatedly perform this measurement 1000 times using this method, one would expect the true value to be included in 997 out of 1000 ranges of the measurement result +/- 0.004 g/dL. For this example measurement result, the range is 0.078 to 0.086 g/dL.* **Example** Item 2: Contains white powder The net weight of Item 2 is 28.55 g +/- 0.03 g at a level of confidence of approximately 95% (95.45%). *Understanding: If one were to repeatedly perform this measurement 100 times using this method, one would expect the true value to be included in 95 out of 100 ranges of the measurement result +/- 0.03 g. For this example measurement result, the range is 28.52 g to 28.58 g.*

TOPIC	LIMITATION	EXAMPLE STATEMENT OR SCENARIO
Frequency Data and Error Rate	It is not common for a report to include method performance frequency data or an error rate. Different terminology may be used: false-positive rate and false-negative rate, error rate, or average error rate. There is a risk of misunderstanding the applicability of frequency data or error rate provided.	**Example** This method has been shown to have an overall false positive rate (FPR) of 0.1% and an overall false negative rate (FNR) of 5.8%. *Understanding: If three or more results are possible from a method (e.g., same source, different source, inconclusive), more information will need to be provided on how each result type was handled in the calculation of FPR and FNR.* • *An average error rate may not apply to a case specific sample type or examiner. Understanding the method performance for that type of sample or for that examiner, will require obtaining detailed method performance frequency data.* • *Did the study protocol align with the quality assurance measures used by the FSSP (e.g., technical review, verification, or retesting)?* • *Accuracy can be affected by the type of item being tested. Did the study protocol include sufficient items similar to the case item type to support the calculation of an error rate?* • *Accuracy has been shown to vary between examiners. Did the study protocol include sufficient examiners with similar level of training and experience to the examiner in the case to support calculation of an error rate that applies to that case specific examiner?*
Likelihood Ratio (LR)	An LR evaluates the evidence from two perspectives (H_1 and H_2 propositions). The selection of these propositions is a limitation. The chosen propositions may not be correct. There may be multiple or changing propositions to evaluate different scenarios. The default defense proposition when pleading not guilty is that the prosecution proposition is not true. An LR may not always be provided in a report. An LR is provided in a report when an inclusionary statement is made. Statistical analysis is not required for exclusionary associations, comparisons between multiple questioned samples without a comparison to a known sample, nor application to inconclusive/uninterpretable results. Obtaining an LR in exclusionary associations may assist a decision maker to appropriately decide the weight to be given to the evidence. When an LR is provided, it is personal to the person providing it. It is based on their beliefs of what is important to consider. This basis may be robust data or experience and knowledge. An officer of the court will need to determine, either informally or formally with calculations, their own value for an LR. There is a risk that the LR number can lead to an incorrect interpretation of the evidence or misleading or inappropriate weight given to the evidence if the LR is misunderstood or misstated (prosecutor's or defense fallacy).	**Example** Item 1: DNA reference sample from (victim) Item 2: DNA reference sample from (suspect) Item 4: DNA sample from scene The genetic profile obtained from Item 4 is interpreted as a mixture of DNA from 3 contributors. Given this genetic profile, assuming Item 1 (victim) is a contributor, it is 1 million times more likely to observe this genetic profile if Item 1 (victim), Item 2 (suspect), and one unknown, unrelated individual are the contributors than if Item 1 (victim) and two unknown, unrelated individuals are the contributors. *Understanding:* *This example report statement explains the probability (Pr) of the evidence (E) and then provides the hypothesis (H). Pr (E|H).* $$LR = 1\ Million = \frac{(H_1)\ the\ contributors\ are\ the\ victim,\ the\ suspect,\ and\ one\ unknown,\ unrelated\ individual}{(H_2)\ the\ contributors\ are\ the\ victim,\ and\ two\ unknown,\ unrelated\ individuals}$$ *There is no single, correct LR. This provided LR of 1 million can vary between FSPs, FSSPs, or others by orders of magnitude depending on the information available, the assumptions made, and the model used for the calculation.* *The officer of the court will need to determine their own LR that aligns with their beliefs and the information that they have available.* *This statement does not (prosecutor's fallacy) mean H1 is 1 million times more likely to be true than H2 is.* *This statement does not mean (defense fallacy) that there are 323 people in the US (population 323 M) and 8000 people in the world (population 8 billion) that could be a contributor to the mixture, and this is reasonable doubt.*

Key Takeaways

1. Readers of reports should be aware that limitations exist both within the methods used and the results provided
2. Readers of reports must understand the report statements provided to ensure that they are applying any statistics correctly

Related Primers

Accreditation and Certification
Method Validation and Method Verification
Method Performance Statistics
Statistical Sampling
Population Statistics
Probability and Likelihood Ratios
Measurement Uncertainty, Frequency Data and Error Rates

Learn More

Likelihood Ratios for Lawyers

The Prosecutor's Fallacy and the Defense Attorney's Fallacy

Royal Statistical Society – Statistics and probability for advocates – Section 1.7

Glossary

Appendix A. Glossary

Term	Definition	Reference	Primer Nos.
Accreditation	Third-party attestation related to a conformity assessment body, conveying formal demonstration of its competence, impartiality, and consistent operation in performing specific conformity assessment activities	ISO 17000:2020 Conformity assessment — Vocabulary and general principles, International Organization for Standardization, Geneva, Switzerland	1A Accreditation and Certification
Accreditation body	An authoritative body that performs accreditation	ISO 17000:2020 Conformity assessment — Vocabulary and general principles, International Organization for Standardization, Geneva, Switzerland	1A Accreditation and Certification
Algorithm	A sequence of steps for solving a problem or accomplishing a task	Lyle, J. , Guttman, B. , Butler, J. , Sauerwein, K. , Reed, C. and Lloyd, C. (2022), Digital Investigation Techniques: A NIST Scientific Foundation Review, NIST Interagency/Internal Report (NISTIR), National Institute of Standards and Technology, Gaithersburg, MD,	1G Algorithms
Artificial Intelligence (AI)	A branch of computer science devoted to developing data processing systems that perform functions normally associated with human intelligence, such as reasoning, learning, and self-improvement	ISO/IEC 2382:2015 Information technology – Vocabulary, International Organization for Standardization, Geneva, Switzerland	1G Algorithms
Assigned value	Value attributed to a particular property of a proficiency test item (ISO/IEC 17043:2010)	ISO/IEC 17043:2010, Conformity Assessment – General Requirements for Proficiency Testing, International Organization for Standardization, Geneva, Switzerland	1H Performance Monitoring: Methods, People, Organizations
Attestation	An issue of a statement, based on a decision, that fulfillment of specified requirements has been demonstrated	ISO 17000:2020 Conformity assessment — Vocabulary and general principles, International Organization for Standardization, Geneva, Switzerland	1A Accreditation and Certification

Term	Definition	Reference	Primer Nos.
Audit	A process for obtaining relevant information about an object of conformity assessment and evaluating it objectively to determine the extent to which specified requirements are fulfilled	ISO 17000:2020 Conformity assessment — Vocabulary and general principles, International Organization for Standardization, Geneva, Switzerland	1A Accreditation and Certification
Audit findings	The results of the evaluation of the collected audit evidence against audit criteria	ISO 9000:2015: Quality management systems — Fundamentals and vocabulary, International Organization for Standardization, Geneva, Switzerland	1A Accreditation and Certification
Bias (Human Factors)	A systematic pattern of deviation in human judgment	Taylor, M, Bird, C, Bishop, B., Burkes, T., Caligiuri, M., Found, B, Grose, W., Logan, L., Melson, K., Merlino, M., Miller, L., Mohammed, L., Morris, J., J., Osborne, N., Ostrum, B., Saunders, C., Shappell, S., H., Srihari, S., Forensic Handwriting Examination and Human Factors: Improving the Practice Through a Systems Approach. NIST Interagency/Internal Report (NISTIR). Gaithersburg, MD: National Institute of Standards and Technology, 2020. [Online], https://doi.org/10.6028/NIST.IR.8282	1F Human Factors
Bias (Method Performance Statistics)	Difference between the expectation of a test result or measurement result and a true value	ISO 3534-2:2006 Statistics — Vocabulary and symbols — Part 2: Applied statistics, International Organization for Standardization, Geneva, Switzerland	2B Method Performance Statistics
Blind proficiency testing	One special application of proficiency testing where the proficiency test item is indistinguishable from normal customer items or samples received	ISO/IEC 17043:2010, Conformity Assessment – General Requirements for Proficiency Testing, International Organization for Standardization, Geneva, Switzerland	1H Performance Monitoring: Methods, People, Organizations

Term	Definition	Reference	Primer Nos.
Calibration	The operation that, under specified conditions, in a first step, establishes a relation between the quantity values with measurement uncertainties provided by measurement standards and corresponding indications with associated measurement uncertainties and, in a second step, uses this information to establish a relation for obtaining a measurement result from an indication	JCGM 200:2012, International vocabulary of metrology – Basic and general concepts and associated terms (VIM), Joint Committee for Guides in Metrology (JCGM)	1E Metrological Traceability
Calibrator	Measurement standard used in calibration	JCGM 200:2012, International vocabulary of metrology – Basic and general concepts and associated terms (VIM), Joint Committee for Guides in Metrology (JCGM)	1E Metrological Traceability
Candidate list	Set of zero, one or more candidates (modified)	ISO/IEC 2382-37:2022 Information technology – Vocabulary – Part 37: Biometrics, International Organization for Standardization, Geneva, Switzerland	1G Algorithms
Certification	Third-party attestation related to an object of conformity assessment, with the exception of accreditation	ISO 17000:2020 Conformity assessment — Vocabulary and general principles, International Organization for Standardization, Geneva, Switzerland	1A Accreditation and Certification
Certified reference material (CRM)	Reference material, accompanied by documentation issued by an authoritative body and providing one or more specified property values with associated uncertainties and traceabilities, using valid procedures	JCGM 200:2012, International vocabulary of metrology – Basic and general concepts and associated terms (VIM), Joint Committee for Guides in Metrology (JCGM)	1E Metrological Traceability

Term	Definition	Reference	Primer Nos.
Chain of custody	Chronological record of the handling, and storage of an item from its point of collection to its final return or disposal	ISO 21043-1:2018 Forensic Sciences Part1: Terms and definitions, International Organization for Standardization, Geneva, Switzerland	1I Why Certain Items are Selected for Examination
Characteristic	Distinguishing feature	ISO 3534-2:2006 Statistics – Vocabulary and symbols – Part 2: Applied statistics, International Organization for Standardization, Geneva, Switzerland	1G Algorithms; 2C Statistical Sampling
Cluster	Part of a population divided into mutually exclusive groups of sampling units related in a certain manner	ISO 3534-2:2006 Statistics — Vocabulary and symbols — Part 2: Applied statistics, International Organization for Standardization, Geneva, Switzerland	2C Statistical Sampling
Cognitive bias	A set of influences that may affect the reliability and validity of one's observations and conclusions	Organization of Scientific Area Committees for Forensic Sciences. (2022). Retrieved from OSAC Lexicon: https://www.nist.gov/glossary/osac-lexicon	1F Human Factors
Comparison	Estimation, calculation or measurement of similarity or dissimilarity between a sample and a reference(s) (modified)	ISO/IEC 2382-37:2022 Information technology – Vocabulary – Part 37: Biometrics, International Organization for Standardization, Geneva, Switzerland	1G Algorithms
Comparison test item	Sample, product, artefact, reference material, piece of equipment, measurement standard, data set or other information used for proficiency testing, interlaboratory comparison, or intralaboratory comparison	ISO/IEC 17043:2010, Conformity Assessment – General Requirements for Proficiency Testing, International Organization for Standardization, Geneva, Switzerland	1H Performance Monitoring: Methods, People, Organizations
Confirmation bias	The tendency to search for data or interpret information in a manner that supports one's preconceptions, expectations or desires	Organization of Scientific Area Committees for Forensic Sciences. (2022). Retrieved from OSAC Lexicon: https://www.nist.gov/glossary/osac-lexicon	1F Human Factors

Term	Definition	Reference	Primer Nos.
Conformity assessment	Demonstration that specified requirements are fulfilled	ISO 17000:2020 Conformity assessment — Vocabulary and general principles, International Organization for Standardization, Geneva, Switzerland	1A Accreditation and Certification
Contextual bias	A deviation in human judgment caused by exposure to information that is either irrelevant to the judgmental task or inappropriate for consideration	Organization of Scientific Area Committees for Forensic Sciences. (2022). Retrieved from OSAC Lexicon: https://www.nist.gov/glossary/osac-lexicon	1F Human Factors
Decision point	An administratively defined cutoff or concentration that is at or above the method's limit of detection or limit of quantitation and is used to discriminate between positive and negative results	ANSI/ASB Standard 036, 2019, Standard Practices for Method Validation in Forensic Toxicology, ANSI, Colorado Springs, CO.	2B Method Performance Statistics
Descriptive statistics	Graphical, numerical or other summary depiction of observed values	ISO 3534-1:2006 Statistics — Vocabulary and symbols — Part 1: General statistical terms and terms used in probability, International Organization for Standardization, Geneva, Switzerland	2A General Statistics
Error	Action or decision which is not intended	Organization of Scientific Area Committees for Forensic Sciences. (2022). Retrieved from OSAC Lexicon: https://www.nist.gov/glossary/osac-lexicon	1F Human Factors; 2F Measurement Uncertainty, Frequency Data and Error Rates
Examination	Act or process of observing, searching, detecting, recording, prioritizing, collecting, analysing, measuring, comparing and/or interpreting	ISO 21043-1:2018 Forensic Sciences Part1: Terms and definitions, International Organization for Standardization, Geneva, Switzerland	1I Why Certain Items are Selected for Examination
Examination strategy	Plan developed to specify the requirements and activities for the examination phase of a forensic process	ISO 21043-1:2018 Forensic Sciences Part1: Terms and definitions, International Organization for Standardization, Geneva, Switzerland	1I Why Certain Items are Selected for Examination

Term	Definition	Reference	Primer Nos.
Expectation bias	The expectation of what an individual will find affects what is actually found	Forensic Sci. Regulator, Guidance on Cognitive Bias Effects Relevant to Forensic Science Examinations FSR-G-217, Guidance on Cognitive Bias (publishing.service.gov.uk)	1F Human Factors
Facility	Physical environment used to protect the item integrity, conduct testing, or support any other aspect of the forensic process	ISO 21043-1:2018 Forensic Sciences Part1: Terms and definitions, International Organization for Standardization, Geneva, Switzerland	1I Why Certain Items are Selected for Examination
False negative	Operational usage: A response indicating that something is not true or not present when it is true or present. Scientific usage: Type II error	Organization of Scientific Area Committees for Forensic Sciences. (2022). Retrieved from OSAC Lexicon: https://www.nist.gov/glossary/osac-lexicon	1G Algorithms; 2F Measurement Uncertainty, Frequency Data and Error Rates
False positive	Operational usage: A response indicating that something is true or present when it is not true or absent; Scientific usage: Type I error	Organization of Scientific Area Committees for Forensic Sciences. (2022). Retrieved from OSAC Lexicon: https://www.nist.gov/glossary/osac-lexicon	1G Algorithms; 2F Measurement Uncertainty, Frequency Data and Error Rates
Frequency	Number of occurrences or observed values in a specified class	ISO 3534-1:2006 Statistics — Vocabulary and symbols — Part 1: General statistical terms and terms used in probability, International Organization for Standardization, Geneva, Switzerland	2D Population Statistics
Human factors	The scientific discipline concerned with the understanding of interactions among humans and other elements of a system	Adapted from Human Factors and Ergonomic Society, 2022 accessed 4/02/2022, HFES | Human Factors and Ergonomics Society	1F Human Factors
Inspection	Examination of an object of conformity assessment and determination of its conformity with detailed requirements or, on the basis of professional judgement, with general requirements	ISO/IEC 17000:2020 Conformity assessment — Vocabulary and general principles, International Organization for Standardization, Geneva, Switzerland	1I Why Certain Items are Selected for Examination

Term	Definition	Reference	Primer Nos.
Interlaboratory comparison	Organization, performance and evaluation of measurements or tests on the same or similar items by two or more laboratories in accordance with predetermined conditions	ISO/IEC 17043:2010, Conformity Assessment – General Requirements for Proficiency Testing, International Organization for Standardization, Geneva, Switzerland	1H Performance Monitoring: Methods, People, Organizations
Intermediate precision	Precision of results from examinations conducted on identical material by the same method in a single laboratory at the same or various times with one or more known sources of variability controlled at multiple levels	ASTM E456-13a (Reapproved in 2017) (modified), Standard Terminology Relating to Quality and Statistics, ASTM International, West Conshohocken, PA	2B Method Performance Statistics
Interpretation	Explanations for the observations, data, and calculations	Organization of Scientific Area Committees for Forensic Sciences. (2022). Retrieved from OSAC Lexicon: https://www.nist.gov/glossary/osac-lexicon	1D Method Validation & Method Verification
Interquartile range	IQR, n—the 75th percentile (0.75 quantile) minus the 25th percentile (0.25 quantile), for a data set. IQR=Q3-Q1	ASTM E2586-19e1 Standard Practice for Calculating and Using Basic Statistics, ASTM International, West Conshohocken, PA, 2006	2A General Statistics
Intralaboratory comparison	Organization, performance and evaluation of measurements or tests on the same or similar items within the same laboratory (3.6) in accordance with predetermined conditions	ISO/IEC 17025:2017 General Requirements for the Competence of Testing and Calibration Laboratories, Geneva, Switzerland	1H Performance Monitoring: Methods, People, Organizations
Item	Object, substance or material that is collected, derived, or sampled as part of the forensic process	ISO 21043-1:2018 Forensic Sciences Part1: Terms and definitions, International Organization for Standardization, Geneva, Switzerland	1I Why Certain Items are Selected for Examination

Term	Definition	Reference	Primer Nos.
Likelihood ratio	A comparison of the probabilities of the evidence under two alternative propositions	Butler, J. Advanced Topics in Forensic DNA Typing: Interpretation, 1st Edition, 2014; p226	2E Probability & Likelihood Ratios
Limit of detection (LOD)	Lowest amount of an analyte that is detectable with a given confidence level	ISO 18158:2016/ ASTM E1732:2022	2B Method Performance Statistics
Limit of quantitation (LOQ)	*also* Lower limit of quantitation (LLOQ) - An estimate of the lowest concentration of an analyte in a sample that can be reliably measured with acceptable bias and precision	Organization of Scientific Area Committees for Forensic Sciences. (2022). Retrieved from OSAC Lexicon: https://www.nist.gov/glossary/osac-lexicon	2B Method Performance Statistics
Machine Learning (ML)	Process by which a functional unit improves its performance by acquiring new knowledge or skills, or by reorganizing existing knowledge or skills	ISO/IEC 2382:2015 – Information technology – Vocabulary, International Organization for Standardization, Geneva, Switzerland	1G Algorithms
Mean (arithmetic mean, average)	n—of a population, μ, average or expected value of a characteristic in a population – of a sample, $\bar{x}$, sum of the observed values in the sample divided by the sample size.	ASTM E2586-19e1 Standard Practice for Calculating and Using Basic Statistics, ASTM International, West Conshohocken, PA, 2006	2A General Statistics
Measurand	Quantity intended to be measured	JCGM 200:2012, International vocabulary of metrology – Basic and general concepts and associated terms (VIM), Joint Committee for Guides in Metrology (JCGM)	2F Measurement Uncertainty, Frequency Data and Error Rates
Measurement	Process of experimentally obtaining one or more quantity values that can reasonably be attributed to a quantity	JCGM 200:2012, International vocabulary of metrology – Basic and general concepts and associated terms (VIM), Joint Committee for Guides in Metrology (JCGM)	1E Metrological Traceability; 2F Measurement Uncertainty, Frequency Data and Error Rates

Term	Definition	Reference	Primer Nos.
Measurement result	Set of quantity values being attributed to a measurand together with any other available relevant information	JCGM 200:2012, International vocabulary of metrology – Basic and general concepts and associated terms (VIM), Joint Committee for Guides in Metrology (JCGM)	1E Metrological Traceability
Measurement uncertainty	Non-negative parameter characterizing the dispersion of the quantity values being attributed to a measurand, based on the information used	JCGM 200:2012, International vocabulary of metrology – Basic and general concepts and associated terms (VIM), Joint Committee for Guides in Metrology (JCGM)	2B Method Performance Statistics; 2F Measurement Uncertainty, Frequency Data and Error Rates
Median	n—the 50th percentile in a population or sample	ASTM E2586-19e1 Standard Practice for Calculating and Using Basic Statistics, ASTM International, West Conshohocken, PA, 2006	2A General Statistics
Method	A combination of procedural steps used to perform a specific technical process	FBI Quality Assurance Standards for DNA Testing Laboratories, 2020	1D Method Validation & Method Verification
Metrological traceability	Property of a measurement result whereby the result can be related to a reference through a documented unbroken chain of calibrations, each contributing to the measurement uncertainty	JCGM 200:2012, International vocabulary of metrology – Basic and general concepts and associated terms (VIM), Joint Committee for Guides in Metrology (JCGM)	1E Metrological Traceability
Metrological traceability chain (traceability chain)	Sequence of measurement standards and calibrations that are used to relate a measurement result to a reference	JCGM 200:2012, International vocabulary of metrology – Basic and general concepts and associated terms (VIM), Joint Committee for Guides in Metrology (JCGM)	1E Metrological Traceability
Midrange	n—average of the minimum and maximum values in a sample	ASTM E2586-19e1 Standard Practice for Calculating and Using Basic Statistics, ASTM International, West Conshohocken, PA, 2006	2A General Statistics
Mode	The value occurring most frequently in a data set	ASQ Quality Glossary	2A General Statistics

Term	Definition	Reference	Primer Nos.
Motivational bias	Influence on decision-making results consistent with a favored conclusion tending to be subject to a lower level of scrutiny than information that may support a less favored outcome	Forensic Sci. Regulator, Guidance on Cognitive Bias Effects Relevant to Forensic Science Examinations FSR-G-217, Guidance on Cognitive Bias (publishing.service.gov.uk)	1F Human Factors
Multi-stage sampling	Sampling in which the sample is selected by stages, the sampling units at each stage being selected from subunits of the larger sampling units chosen at the previous stage	ISO 3534-2:2006 Statistics — Vocabulary and symbols — Part 2: Applied statistics, International Organization for Standardization, Geneva, Switzerland	2C Statistical Sampling
Nominal property	Property of a phenomenon, body, or substance, where the property has no magnitude	JCGM 200:2012, International vocabulary of metrology – Basic and general concepts and associated terms (VIM), Joint Committee for Guides in Metrology (JCGM)	1E Metrological Traceability
One-to-many search (1:N)	A process in which a sample or feature of one subject is searched against the references of more than one subject to return a candidate list (modified)	ISO/IEC 2382-37:2022 Information technology – Vocabulary – Part 37: Biometrics, International Organization for Standardization, Geneva, Switzerland	1G Algorithms
One-to-one comparison (1:1)	A process in which a sample or feature from one subject is compared to a reference(s) from one subject to produce a similarity score (modified)	ISO/IEC 2382-37:2022 Information technology – Vocabulary – Part 37: Biometrics, International Organization for Standardization, Geneva, Switzerland	1G Algorithms
Opinion	View, judgment, belief – takes into consideration other information in addition to observations, data, calculations, and interpretations	Organization of Scientific Area Committees for Forensic Sciences. (2022). Retrieved from OSAC Lexicon: https://www.nist.gov/glossary/osac-lexicon	1D Method Validation & Method Verification

Term	Definition	Reference	Primer Nos.
Participant	Laboratory, organization or individual that receives comparison test items and submits results for review by the proficiency test provider or comparison coordinator	ISO/IEC 17043:2010, Conformity Assessment – General Requirements for Proficiency Testing, International Organization for Standardization, Geneva, Switzerland	1H Performance Monitoring: Methods, People, Organizations
Population	Totality of items under consideration	ISO 3534-1:2006 Statistics — Vocabulary and symbols — Part 1: General statistical terms and terms used in probability, International Organization for Standardization, Geneva, Switzerland	2A General Statistics; 2C Statistical Sampling; 2D Population Statistics
Population parameter	Summary measure of the values of some characteristic of a population	ISO 3534-2:2006 Statistics — Vocabulary and symbols — Part 2: Applied statistics, International Organization for Standardization, Geneva, Switzerland	2A General Statistics
Precision	(ISO 3534-2:2006) - Closeness of agreement between independent test/measurement results obtained under stipulated conditions	ISO 3534-2:2006 Statistics — Vocabulary and symbols — Part 2: Applied statistics, International Organization for Standardization, Geneva, Switzerland	2B Method Performance Statistics
Probability	A quantified measure between zero and one indicating how probable or likely it is that an event will or has occurred. In the frequentist interpretation, probability is based on the rates at which events occur. In the Bayesian interpretation, probability reflects a degree of belief. On this scale, zero indicates impossibility and one indicates absolute certainty	Organization of Scientific Area Committees for Forensic Sciences. (2022). Retrieved from OSAC Lexicon: https://www.nist.gov/glossary/osac-lexicon	2E Probability & Likelihood Ratios
Process map	A tool that graphically shows the inputs, actions and outputs for all steps of a process	Organization of Scientific Area Committees for Forensic Sciences. (2022). Retrieved from OSAC Lexicon: https://www.nist.gov/glossary/osac-lexicon	1F Human Factors

Term	Definition	Reference	Primer Nos.
Product	Output of an organization that can be produced without any transaction taking place between the organization and the customer	ISO 9000:2015 Quality management systems — Fundamentals and vocabulary, International Organization for Standardization, Geneva, Switzerland	1C Quality Assurance vs. Quality Control
Proficiency test provider	Organization which takes responsibility for all tasks in the development and operation of a proficiency testing scheme	ISO/IEC 17043:2010, Conformity Assessment – General Requirements for Proficiency Testing, International Organization for Standardization, Geneva, Switzerland	1H Performance Monitoring: Methods, People, Organizations
Proficiency testing	Evaluation of participant performance against pre-established criteria by means of interlaboratory comparisons	ISO/IEC 17043:2010, Conformity Assessment – General Requirements for Proficiency Testing, International Organization for Standardization, Geneva, Switzerland	1H Performance Monitoring: Methods, People, Organizations
Quality	The degree to which a set of inherent characteristics of an object fulfills requirements	ISO 9000:2015 Quality management systems — Fundamentals and vocabulary, International Organization for Standardization, Geneva, Switzerland	1C Quality Assurance vs. Quality Control; 1D Method Validation & Method Verification
Quality assurance	A part of quality management focused on providing confidence that quality requirements will be fulfilled	ISO 9000:2015 Quality management systems — Fundamentals and vocabulary, International Organization for Standardization, Geneva, Switzerland	1C Quality Assurance vs. Quality Control
Quality control	A part of quality management focused on fulfilling quality requirements	ISO 9000:2015 Quality management systems — Fundamentals and vocabulary, International Organization for Standardization, Geneva, Switzerland	1C Quality Assurance vs. Quality Control
Quality management	Management with regard to quality	ISO 9000:2015 Quality management systems — Fundamentals and vocabulary, International Organization for Standardization, Geneva, Switzerland	1C Quality Assurance vs. Quality Control

Term	Definition	Reference	Primer Nos.
Quality requirement	A requirement related to quality	ISO 9000:2015 Quality management systems — Fundamentals and vocabulary, International Organization for Standardization, Geneva, Switzerland	1C Quality Assurance vs. Quality Control
Quantity	Property of a phenomenon, body, or substance, where the property has a magnitude that can be expressed as a number and a reference	JCGM 200:2012, International vocabulary of metrology – Basic and general concepts and associated terms (VIM), Joint Committee for Guides in Metrology (JCGM)	2F Measurement Uncertainty, Frequency Data and Error Rates
Quantity value	Number and reference together expressing magnitude of a quantity	JCGM 200:2012, International vocabulary of metrology – Basic and general concepts and associated terms (VIM), Joint Committee for Guides in Metrology (JCGM)	2F Measurement Uncertainty, Frequency Data and Error Rates
Random sampling	Sampling where a sample of n sampling units is taken from a population in such a way that each of the possible combinations of n sampling units has a particular probability of being taken	ISO 3534-1:2006 Statistics — Vocabulary and symbols — Part 1: General statistical terms and terms used in probability, International Organization for Standardization, Geneva, Switzerland	2C Statistical Sampling
Range	R, n—maximum value minus the minimum value in a sample	ASTM E2586-19e1 Standard Practice for Calculating and Using Basic Statistics, ASTM International, West Conshohocken, PA, 2006	2A General Statistics
Recognition	Identification of items or characteristics that can have potential forensic value	ISO 21043-1:2018 Forensic Sciences Part1: Terms and definitions, International Organization for Standardization, Geneva, Switzerland	1I Why Certain Items are Selected for Examination
Reference measurement standard (reference standard)	Measurement standard designated for the calibration of other measurement standards for quantities of a given kind in a given organization or at a given location	JCGM 200:2012, International vocabulary of metrology – Basic and general concepts and associated terms (VIM), Joint Committee for Guides in Metrology (JCGM)	1E Metrological Traceability
Relative frequency	Frequency divided by the total number of occurrences or observed values	ISO 3534-1:2006 Statistics — Vocabulary and symbols — Part 1: General statistical terms and terms used in probability, International Organization for Standardization, Geneva, Switzerland	2D Population Statistics

Term	Definition	Reference	Primer Nos.
Relative standard deviation	relsd = 100*(standard deviation/|mean|)	NIST SEMATECH	2B Method Performance Statistics
Repeatability	Extent of agreement between more than one result determined in the same place, by the same person, on the same equipment, in the same way, at similar times	OSAC Preferred Terms, Organization of Scientific Area Committees for Forensic Sciences. (2022). Retrieved from https://www.nist.gov/glossary/osac-lexicon?k=&name=&committee=All&standard=&items_per_page=50&preferred=1#top	2B Method Performance Statistics
Reproducibility	Extent of agreement between more than one result determined under any combination of different conditions	OSAC Preferred Terms, Organization of Scientific Area Committees for Forensic Sciences. (2022). Retrieved from https://www.nist.gov/glossary/osac-lexicon?k=&name=&committee=All&standard=&items_per_page=50&preferred=1#top	2B Method Performance Statistics
Requirement	A need or expectation that is stated, generally implied, or obligatory	ISO 9000:2015 Quality management systems — Fundamentals and vocabulary, International Organization for Standardization, Geneva, Switzerland	1C Quality Assurance vs. Quality Control
Result	The product of the forensic service provider. This term is broad and includes observations, data, calculations, interpretations, and opinions	Organization of Scientific Area Committees for Forensic Sciences. (2022). Retrieved from OSAC Lexicon: https://www.nist.gov/glossary/osac-lexicon	1C Quality Assurance vs. Quality Control; 1D Method Validation & Method Verification
Sample (Primer 1I)	Portion drawn from a whole or population for the purpose of examination/testing, not necessarily representative of the whole	ISO 21043-1:2018 Forensic Sciences Part1: Terms and definitions, International Organization for Standardization, Geneva, Switzerland	1I Why Certain Items are Selected for Examination
Sample (General Statistics)	Subset of a population made up of one or more sampling units	ISO 3534-1:2006 Statistics — Vocabulary and symbols — Part 1: General statistical terms and terms used in probability, International Organization for Standardization, Geneva, Switzerland	2A General Statistics; 2C Statistical Sampling; 2D Population Statistics

Term	Definition	Reference	Primer Nos.
Sample size	Number of sampling units in a sample.	ISO 3534-1:2006 Statistics — Vocabulary and symbols — Part 1: General statistical terms and terms used in probability, International Organization for Standardization, Geneva, Switzerland	2C Statistical Sampling
Sample statistic	Summary measure of the observed values of a sample	ASTM E2586-19e1 Standard Practice for Calculating and Using Basic Statistics, ASTM International, West Conshohocken, PA, 2006	2A General Statistics
Sampling (Primer 1I)	Selection and/or collection of material or data regarding an object of conformity assessment	ISO/IEC 17000:2020 Conformity assessment — Vocabulary and general principles, International Organization for Standardization, Geneva, Switzerland	1I Why Certain Items are Selected for Examination
Sampling (Statistical Sampling)	Act of drawing or constituting a sample	ISO 3534-1:2006 Statistics — Vocabulary and symbols — Part 1: General statistical terms and terms used in probability, International Organization for Standardization, Geneva, Switzerland	2C Statistical Sampling
Sampling unit	An item, group of items, or segment of material, that can be selected as part of a probability sampling plan	ASTM E456-13a(2022) Standard Terminology Relating to Quality and Statistics, ASTM International, West Conshohocken, PA, 2006	2C Statistical Sampling
Scene	Place or object that is subject to and/or requires forensic examination	ISO 21043-1:2018 Forensic Sciences Part1: Terms and definitions, International Organization for Standardization, Geneva, Switzerland	1I Why Certain Items are Selected for Examination
Selectivity	The extent to which other substances interfere with the determination of a substance according to a given procedure	International Union of Pure Chemistry (IUPAC)	2B Method Performance Statistics
Similarity score	Comparison score that increases with similarity	ISO/IEC 2382-37:2022 Information technology – Vocabulary – Part 37: Biometrics, International Organization for Standardization, Geneva, Switzerland	1G Algorithms

Term	Definition	Reference	Primer Nos.
Simple random sampling	Sampling where a sample of n sampling units is taken from a population in such a way that all the possible combinations of n sampling units have the same probability of being taken	ISO 3534-1:2006 Statistics — Vocabulary and symbols — Part 1: General statistical terms and terms used in probability, International Organization for Standardization, Geneva, Switzerland	2C Statistical Sampling
Standard	A document, established by consensus and approved by a recognized body, that provides, for common and repeated use, rules, guidelines, or characteristics for activities or their results, aimed at the achievement of the optimum degree of order in a given context	ISO/IEC Guide 2:2004	1B Documentary Standards
Standard Developing Organization	A body or organization that carries out recognized activities in the area of standardization, such as the development, adoption, and publication of standards and whose membership may be composed of national, regional, or other standardizing bodies, or of organizations such as companies, governmental, academic or other institutions and individuals	ISO/IEC Guide 2:2004	1B Documentary Standards

Term	Definition	Reference	Primer Nos.
Standard deviation	Of a population, σ, the square root of the average or expected value of the squared deviation of a variable from its mean; —of a sample, s, the square root of the sum of the squared deviations of the observed values in the sample from their mean divided by the sample size minus 1	ASTM E2586-19e1 Standard Practice for Calculating and Using Basic Statistics, ASTM International, West Conshohocken, PA, 2006	2A General Statistics; 2B Method Performance Statistics
Standard Setting Organization	Incorporates all variants of groups that develop standards, including special interest groups (SIGs), standards-development organizations (SDOs), consortia, and other entities. The acronym SSO is often used interchangeably with SDO but, in principle, the former term covers the activities of both setting and managing standards, including associated intellectual property issues	National Research Council. 2013. Patent Challenges for Standard-Setting in the Global Economy: Lessons from Information and Communications Technology. Washington, DC: The National Academies Press.	1B Documentary Standards
Statistic	See sample statistic	ASTM E2586-19e1 Standard Practice for Calculating and Using Basic Statistics, ASTM International, West Conshohocken, PA, 2006	2A General Statistics
Statistical process control (SPC)	the use of statistical techniques to control a process or production method	ASQ Quality Glossary	2A General Statistics
Stratified sampling	Sampling in which the population to be sampled is first divided into mutually exclusive subsets or strata, and independent samples taken within each stratum	ASTM E456-13a(2022) Standard Terminology Relating to Quality and Statistics, ASTM International, West Conshohocken, PA, 2006	2C Statistical Sampling

Term	Definition	Reference	Primer Nos.
Stratum	Mutually exclusive and exhaustive sub-population considered to be more homogeneous with respect to characteristics investigated than the total population	ISO 3534-1:2006 Statistics — Vocabulary and symbols — Part 1: General statistical terms and terms used in probability, International Organization for Standardization, Geneva, Switzerland	2C Statistical Sampling
Surveillance	A systematic iteration of conformity assessment activities as a basis for maintaining the validity of the statement of conformity	ISO 17000:2020 Conformity assessment — Vocabulary and general principles, International Organization for Standardization, Geneva, Switzerland	1A Accreditation and Certification
System	A group of interacting or interrelated elements that act according to a set of rules to form a unified whole	Organization of Scientific Area Committees for Forensic Sciences. (2022). Retrieved from OSAC Lexicon: https://www.nist.gov/glossary/osac-lexicon	1F Human Factors
Task	Work to be undertaken	Organization of Scientific Area Committees for Forensic Sciences. (2022). Retrieved from OSAC Lexicon: https://www.nist.gov/glossary/osac-lexicon	1F Human Factors
Testing	Determination of one or more characteristics of an object of conformity assessment, according to a procedure	ISO/IEC 17000:2020 Conformity assessment — Vocabulary and general principles, International Organization for Standardization, Geneva, Switzerland	1I Why Certain Items are Selected for Examination
True value	Value which characterizes a quantity or quantitative characteristic perfectly defined in the conditions which exist when that quantity or quantitative characteristic is considered	ISO 3534-2:2006 Statistics — Vocabulary and symbols — Part 2: Applied statistics, International Organization for Standardization, Geneva, Switzerland	2B Method Performance Statistics

Term	Definition	Reference	Primer Nos.
Trueness	Closeness of agreement between the expectation of a test result or a measurement result and a true value	ISO 3534-2:2006 Statistics — Vocabulary and symbols — Part 2: Applied statistics, International Organization for Standardization, Geneva, Switzerland	2B Method Performance Statistics
Validation	Verification, where the specified requirements are adequate for an intended use	ISO/IEC 17025:2017 General Requirements For The Competence Of Testing And Calibration Laboratories, Geneva, Switzerland	1D Method Validation & Method Verification
Variance	σ2, s2, n—square of the standard deviation of the population or sample	ASTM E2586-19e1 Standard Practice for Calculating and Using Basic Statistics, ASTM International, West Conshohocken, PA, 2006	2A General Statistics
Verification	Provision of objective evidence that a given item fulfills specified requirements	ISO/IEC 17025:2017 General Requirements For The Competence Of Testing And Calibration Laboratories, Geneva, Switzerland	1D Method Validation & Method Verification
Violation	Deliberate deviation from a rule or procedure	Organization of Scientific Area Committees for Forensic Sciences. (2022). Retrieved from OSAC Lexicon: https://www.nist.gov/glossary/osac-lexicon	1F Human Factors