2F. 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

Part of NIST SP 1500-28
https://doi.org/10.6028/NIST.SP.1500-28

Glossary Primer 2F Measurement Uncertainty, Frequence Data, and Error Rates

Term	Definition	Reference	Primer Nos.
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
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
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
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

Term	**Definition**	**Reference**	**Primer Nos.**
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