



2B. 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

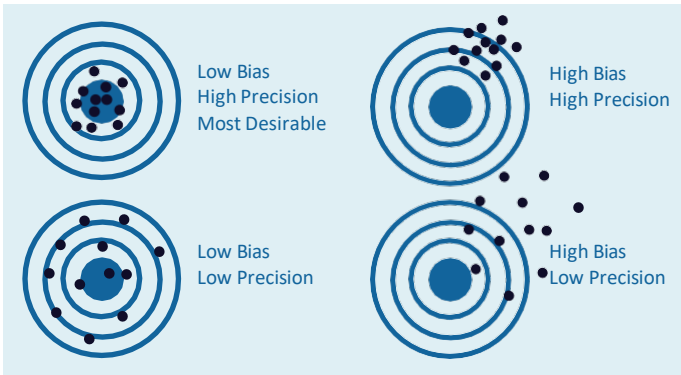


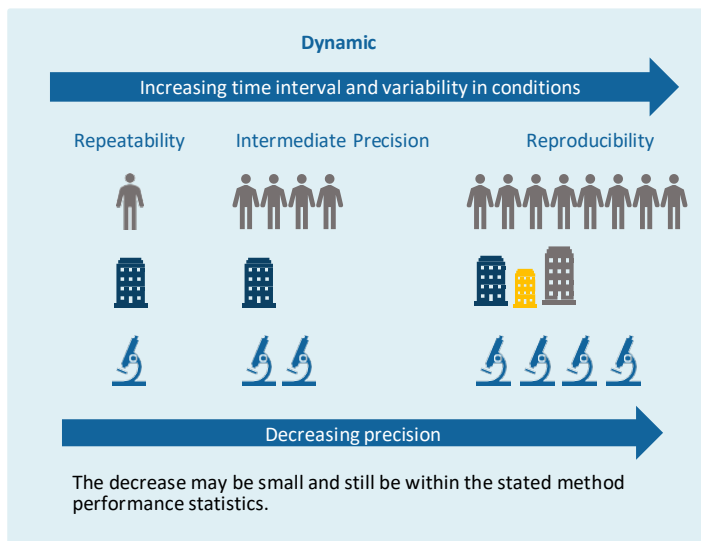
Figure 1 depicts the difference between precision and bias. The bull's eye on the target is the known or reference value. The scatter of the many attempts to hit the bull's eye shows the variability and trueness in this method.

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 interferences 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)		$tp/pc = tp / (tp + fn) = 1 - FN$ $= 186/190 \times 100 = 97.9\%$	
False positive rate, FP (Type I error)		$fp/nc = fp / (tn + fp) = 1 - TN$ $= 1/190 \times 100 = 0.5\%$	
True negative rate, TN (Specificity, SP)		$tn/nc = tn / (tn + fp) = 1 - FP$ $= 189/190 \times 100 = 99.5\%$	
False negative rate, FN (Type II error)		$fn/pc = fn / (tp + fn) = 1 - TP$ $= 4/190 \times 100 = 2.1\%$	

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

- Method validation and method verification require the method performance criteria to be evaluated and known.
- Method performance statistics are specific to the method. If method parameters are changed, method performance statistics will need to be recalculated.
- Method performance statistics provide information on a method's precision/bias (variability/trueness).
- 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.](#)

Part of NIST SP 1500-28

<https://doi.org/10.6028/NIST.SP.1500-28>



Glossary: Primer 2B Method Performance Statistics

Term	Definition	Reference	Primer Nos.
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
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
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
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
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
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
Relative standard deviation	$\text{relsd} = 100 * (\text{standard deviation} / \text{mean})$	NIST SEMATECH	2B Method Performance Statistics



Term	Definition	Reference	Primer Nos.
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
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
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
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
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