VERIFICATION AND VALIDATION TOOLKIT

Determining Performance Characteristics of Qualitative Assays

To ensure correct diagnosis and treatment, clinical laboratory testing must be accurate and reliable. A key component of the quality assurance process is the verification or validation of new instruments and tests to confirm their ability to perform prior to implementation.

The Verification and Validation Toolkit walks users through this process and provides additional resources, templates and examples for use in the laboratory. This section of the toolkit provides information on the selection and frequency of quality controls and determining performance characteristics (accuracy, precision, sensitivity, specificity, and reportable and reference ranges) for qualitative assays. Find checklist examples in Additional Qualitative Assay Resources (page 5).

Find the complete toolkit at aphl.org/VV-Toolkit

The toolkit has eight sections:

1. Verification and Validation 101
2. Verification and Validation Process Checklist
3. Obtaining Appropriate Test Samples
4. Qualitative Assays
5. Quantitative Assays
6. Related Processes
7. Safety Considerations and Risk Assessments
8. Cost Analysis and Budget

Qualitative assays are methods that provide only two categorical results (i.e., positive or negative; present or absent; reactive or nonreactive; yes or no). Some qualitative assays have no numerical value associated with the result whereas other assays are labeled as qualitative because one of only two results is reported (i.e., positive or negative) even though a numerical value is derived. The overall objective of qualitative assays is to recognize the presence or absence of an analyte. In qualitative assays, the cutoff value is defined as the threshold above which the result is reported as positive and below which the result is reported as negative.

Clinical laboratory uses for qualitative assays are described as screening, diagnostic, confirmatory or monitoring.[1] The utility of a given assay is determined based on the sensitivity and specificity, predictive values, and the prevalence of disease or condition in the population tested.

- **Screening Methods:** Use to test a population subset for the presence or absence of an analyte or agent.
- **Diagnostic Methods:** Use clinical suspicion of a particular disease or condition to guide testing. Both screening and diagnostic assays should have high sensitivity; lower specificity is tolerated if a confirmatory test is available, and the results are low consequence.
- **Confirmatory Methods:** Follow screening or diagnostic test results and enable clinicians to establish a diagnosis with testing that is designed to be specific, sometimes at the expense of sensitivity, and have a high positive predictive value (PPV).

The type of verification or validation for qualitative assays is dependent on clearance or approval from a regulatory entity (i.e., FDA cleared or approved). In general, refer to the assay instructions for use (IFU) documentation to determine the number and type of samples to use for the verification or validation. Additional guidance recommendations are provided in this toolkit.

Table 1. Summary of Performance Characteristics Required Depending on Qualitative Test Type[2]

Test Type	Accuracy	Precision	Analytical Sensitivity	Analytical Specificity	Reportable Range	Reference Range
FDA Approved	Required	Required			Required	Required
FDA Cleared	Required	Required			Required	Required
FDA Modified or LDT	Required	Required	Required	Required	Required	Required
FDA Authorized (EUA)*	Required	Required			Required	Required

Controls

Positive and negative controls should be chosen such that they provide expected results when the test is functioning properly. Control design near the cutoff value can detect more errors; however, can lead to rejection of a test run that does not have significant errors. To determine an optimal set of controls, use the provided guidance by the manufacturer, stable commercial or clinical controls, or perform a precision experiment to understand the imprecision of the assay (refer to CLSI EP12[1] for performing an imprecision experiment). Per CLIA §493.1256,[2] a laboratory must not use control materials outside the patient reportable range. Control samples not containing the analytes or substances to be controlled are not acceptable as control material.

For most qualitative assays, it is acceptable to perform a negative and positive quality control daily, while other testing methods may require more frequent testing of controls on a per run basis (check the IFU per method). Verification or validation of an assay can help determine the performance of controls and ascertain the frequency required for addition of controls to a given assay. Depending on the assay, the laboratory could customize its QC plan using an Individualized Quality Control Plan (IQCP).[3,4] If controls fail to produce the expected results, the run must be rejected, and the failure should be investigated to identify the cause.

Accuracy

For qualitative assays, accuracy studies should validate if the test method detects the presence or absence of the analyte. Sources vary on recommended number of samples to test for accuracy. If there is no guidance from the IFU, CLSI EP09c[5] suggests 40 total samples (20 each of positive and negative value) should be tested. This number is a minimum suggestion, and an assessment should be conducted to determine if more samples are necessary to increase the statistical relevance. In addition, the clinical impact and repercussions or consequences to the patient of a false negative or false positive result is critical in determining the number of samples for testing. If fewer than 40 samples are all that can be obtained, prior approval by the laboratory director or quality assurance officer should be obtained.

Accuracy testing should be performed over a minimum course of five days to simulate a range of conditions over which samples would normally be run.

* Requirements may vary depending on EUA. Performance characteristics will be defined by the EUA IFU and the laboratory director. Instructions should include requirements for verification or validation.

2 x 2 Contingency Table

The table in **Figure 1** can be used to calculate the estimated sensitivity, specificity, total accuracy, PPV and negative predictive value (NPV). Results should correlate with an expected total accuracy of ≥95% agreement with the reference method. Westgard QC has a 2 x 2 contingency calculator.*

Figure 1. 2 x 2 Contingency Table

		Diagnostic Accuracy Criteria		
		Positive	**Negative**	**Total**
Method X	**Positive**	# true positive (TP)	# false positive (FP)	TP + FP
	Negative	# false negative (FN)	# true negative (TN)	FN + TN
	Total	TP + FN	FP + TN	N

Estimated Sensitivity

$$= 100 \times \frac{TP}{TP + FN}$$

Estimated Specificity

$$= 100 \times \frac{TN}{FP + TN}$$

Positive Predictive Value (PPV)

$$= 100 \times \frac{TP}{TP + FP}$$

Negative Predictive Value (NPV)

$$= 100 \times \frac{TN}{FN + TN}$$

Total Accuracy

$$= 100 \times \frac{TP + TN}{TP + FP + FN + TN}$$

Kappa Coefficient

If an imperfect standard is being used for the verification or validation, the overall agreement of the assay can be calculated using the Kappa coefficient. The Kappa coefficient is a measurement of the degree of agreement between the methods above what is expected by chance alone.

Use the formulas described below in **Figure 2** in association with the values from **Figure 1**. An online calculator is available at graphpad.com.**

Understanding Kappa:

- A Kappa of, or approaching, one indicates that there is very good agreement
- A Kappa approaching zero indicates that the agreement is no better than chance.
- A negative Kappa means that the agreement is worse than chance.

Figure 2. Calculating the Kappa coefficient

$$A = TP + \frac{FP}{N} \qquad B = TP + \frac{FN}{N}$$

Pr(e)

$$= (A \times B) + [(1-A) \times (1-B)]$$

Pr(a)

$$= 100 \times \frac{(TP + TN)}{N}$$

(overall percent agreement)

$$\text{Kappa (K)} = \frac{Pr(a) - Pr(e)}{1 - Pr(e)}$$

* www.westgard.com/qualitative-test-clinical-agreement.htm

** www.graphpad.com/quickcalcs/kappa1.cfm

Precision

Samples for precision should be near the high and low cutoff values to provide the best estimation of error at medically relevant decision levels. A minimum of one positive and one negative sample is recommended with a total of 10-30 measurements. Precision testing should be performed over a course of multiple days using more than one laboratorian to demonstrate reproducibility. The experimental design consists of the following precision measurements:

- **Intra-assay** (within run): Same samples run multiple times on the same run and day
- **Inter-assay** (between run): Same samples run in different runs on the same day or different days, and preferably by a different laboratorian.

If any of the aforementioned precision measurements are not applicable to a given assay, discuss with the quality assurance officer (QAO) or laboratory directory to determine feasibility or requirements of testing. The manufacturers' statements of precision should be used as a minimum performance requirement. Alternatively, if numerical data are available and standard deviation can be calculated, the coefficient of variation (CV) can be used to express the precision and repeatability of an assay.

CV Calculation:[6] The CV is the ratio of the standard deviation to the mean. CV is expressed as a percentage. The ideal CV is <15%, and generally should not exceed 20%. Use the formula in **Figure 3** to calculate the CV.

Figure 3. Calculating the Coefficient of Variation

$$CV\ \% = 100 \times \frac{\text{Standard Deviation}}{\text{Mean}}$$

Imprecision of the method can be analyzed using concentrations near the cutoff. However, it is not appropriate to measure the imprecision of qualitative assays with low-negative or high-positive samples since these values are usually too far away in analyte concentration from the medical decision point. Details regarding how to perform a qualitative method precision experiment to understand imprecision of an assay can be found in CLSI EP12 section 8.3.[1]

Analytical Sensitivity

Analytical sensitivity is referred to as 'limit of detection studies.' Limit of detection (LoD) seeks to define the lowest concentration of an analyte in a matrix that can be consistently detected. For LDT's, this measurement must be established during the method validation. For other assays, the manufacturer has completed LoD studies.

Analytical sensitivity for qualitative assays can be challenging to complete and will not always provide a definitive quantity as the LoD. Some qualitative methods will have a measurable value (i.e., cycle threshold, optical density, titer, colony forming unit, etc.), or measurand, that is subsequently used to determine the qualitative test result. In these cases, CLSI EP17:A2[7] recommends making serial dilutions of a sample with a known measurand content in replicate. The samples should be run in duplicate or triplicate over three days with a recommended minimum of 20 measurements for each sample concentration to verify a manufacturer claim, and 60 measurements to establish the LoD. However, the exact number of samples to use should be determined on a case-by-case basis with input from the QAO, supervisors, or the laboratory director. If the laboratory wants to establish a precise LoD, the laboratorian calculates and plots the hit rate for each dilution, which is defined as the total number of positive results divided by the total number of replicates, using regression modeling with hit rate on the y-axis and measurand dose on the x-axis. The laboratorian then selects the hit rate that corresponds to detection of the analyte in the majority of samples (i.e., 95%) as the LoD. A probit fit analysis using computer software can be used to easily perform this calculation. A detailed description of probit analysis is found on the Westgard QC website.[8]

Westgard offers an alternative and simpler approach to address LoD in qualitative assays that have a measurand that is assumed to be continuous along a range of concentrations. If the assay cutoff is known, the laboratory can test samples that are expected to fall below and above the cutoff in replicates of 20. Next, the number of positive results are evaluated. No more than 5% of replicate results below the cutoff value should be positive. Conversely, at least 95% of replicate results above the cutoff value should be positive. These calculations correspond to the 95% confidence interval.

Analytical Specificity

Analytical specificity is the evaluation of cross-reactivity by testing a panel of similar, potentially interfering organisms, substances, or analytes to assess constant systematic error. For LDT's, this measurement must be established during the method validation. For other assays, the manufacturer has completed analytical specificity studies.

When determining analytical specificity, the test agents or substances should include as many organisms or analytes as possible that may be found in the relevant test sample or that cause the same symptoms as the target agent. Consider potential sources of variability that could affect the assay (i.e., matrix composition, lot-to-lot variability, temperature, etc.) and include them in the design of the verification or validation. The recommended number of samples to use is between three to five, containing each of the potentially interfering or cross-reactive test organisms, analytes or substances to test. Results should correlate with an expected value ≥95%. If cross-reactivity is observed, assay conditions may need to be adjusted or reevaluated. In instances where cross-reactivity cannot be eliminated, it must be noted as a limitation of the assay. An inhibition control may need to be included in assay runs where inhibition is prone and needs to be monitored (i.e., molecular assays direct from specimens).

Reportable Range

Reportable range refers to the range of diagnostic results that will be reported. Depending on the assay method used, the reportable range could be a binary result (positive or negative), non-binary result (positive, negative, indeterminate or invalid), the LoD, cutoff value or the 95% confidence internal. The result outcomes are stated in the verification or validation plan and may be adjusted for the final report based on data from the verification or validation. The reportable range should be included in the final SOP.

Reference Range

Reference range is the typical result expected in a healthy population that does not have the condition for which the test is performed. No samples are tested to determine the reference range. Instead, the expected result for a healthy population is stated within the verification or validation plan, report and final SOP. The laboratory may use the manufacturer's reference range provided it is appropriate for the laboratory's patient population. If the manufacturer has not provided reference ranges appropriate for the laboratory's patient population, the laboratory may use published reference range(s).

Additional Qualitative Assay Resources

- Example of Microbiology MALDI-TOF Validation Supplemental Checklist
- Example of Microbiology NAAT Checklist
- NGS Method Validation Plan Template
- NGS Method Validation Summary Report Template

References

1 CLSI. Evaluation of Qualitative, Binary Output Examination Performance, 3rd ed. CLSI guideline EP12. Clinical and Laboratory Standards Institute, 2023. Available from: clsi.org/standards/products/method-evaluation/documents/ep12/

2 Code of Federal Regulations. 42 CFR 493.1253 Standard: Establishment and verification of performance specifications. National Archives and Records Administration; 2003. Accessed Oct 1, 2023 from: www.ecfr.gov/current/title-42/chapter-IV/subchapter-G/part-493/subpart-K/subject-group-ECFRc96daead380f6ed/section-493.1253

3 CMS CLIA. State Operations Manual Appendix C - Survey Procedures and Interpretive Guidelines for Laboratories and Laboratory Services, Rev. 166, 02-03-17. IQCP is on page 198. Accessed Jan 15, 2024 from: www.cms.gov/regulations-and-guidance/guidance/manuals/downloads/som107ap_c_labpdf.pdf

4 CDC Laboratory Quality Tools and Resources. Individualized Quality Control Plan (IQCP). Accessed Oct 15, 2023 from: www.cdc.gov/labquality/iqcp.html

5 CLSI. Measurement Procedure Comparison and Bias Estimation Using Patient Samples, 3rd ed., CLSI guideline EP09c, Wayne, PA: Clinical and Laboratory Standards Institute; 2018. Available from: clsi.org/standards/products/method-evaluation/documents/ep09/

6 Bio-Rad UnityWeb. Coefficient of Variation. Accessed October 15, 2023 from: unityweb.qcnet.com/Documentation/Help/UnityWeb/402.htm

7 CLSI. Evaluation of Detection Capability for Clinical Laboratory Measurement Procedures; Approved Guideline-Second Edition. CLSI document EP17-A2. Wayne, PA: Clinical and Laboratory Standards Institute; 2012. Available from: clsi.org/standards/products/method-evaluation/documents/ep17/

8 Westgard QC. James O. Westgard, Sten A. Westgard. Probit Analysis 1: Practical Application to Determine Limit of Detection. August 2020. Accessed October 1, 2023 from: www.westgard.com/probit-part-one.htm

Association of Public Health Laboratories

The Association of Public Health Laboratories (APHL) works to strengthen laboratory systems serving the public's health in the US and globally. APHL's member laboratories protect the public's health by monitoring and detecting infectious and foodborne diseases, environmental contaminants, terrorist agents, genetic disorders in newborns and other diverse health threats.

7700 Wisconsin Avenue, Suite 1000 Bethesda, MD 20814 | 240.485.2745 | www.aphl.org

This publication was supported by Cooperative Agreements #NU60OE000104, 100% funded by the US Centers for Disease Control and Prevention (CDC). Its contents are solely the responsibility of the authors and do not necessarily represent the official views of CDC or the US Department of Health and Human Services.