VERIFICATION AND VALIDATION TOOLKIT

Related Processes

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.

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

Sample or Analyte Stability and Integrity

Integrity of the sample analyte should be scrutinized, especially regarding potential causes of instability. The stability of a sample can be assessed at the pre-analytical and analytical phases of testing with the acceptable parameters for sample stability determined during a validation and confirmed for a verification.

Creating an established written policy for sample stability is the end goal when selecting verification or validation samples for pre-analytical and analytical testing.

Pre-analytical sample criteria to consider are:

- Proper shipping and storage
- Appropriate collection devices
- Receipt within the acceptable testing timeframe.

Temperatures of acceptance for shipping, receipt and storage can be assessed during a validation by including samples within a temperature range.

Analytical sample criteria include:

- Assessing the expiration of samples (were the samples received after the acceptable timeframe for testing)
- Minimum volume requirements

- Effects of temperature fluctuations on sample integrity
- Specimen suitability requirements (pH, turbidity, specific gravity, etc.).

For example, molecular extracts or eluates often have limitations on freeze-thaw cycles; therefore, selecting samples that are nearing the allowable freeze-thaw limit is not advisable. Failure to fully assess sample integrity during a verification or validation may result in misguided result analysis and potential diagnostic testing errors.

Bridging Studies and Addendums

Modifications to an approved assay (FDA cleared or approved, EUA authorized, approved laboratory-developed tests (LDT)) must be validated or verified to demonstrate equivalent performance to the unmodified authorized method. These assay modifications may be referred to as bridging studies or addendums and are meant to demonstrate that the modifications do not negatively affect the result or clinical interpretation of the assay result.

Bridging Studies

Introduced by the FDA in 1998 in the context of drug and biological compound registrations, the concept of bridging studies was originally defined as a study to extrapolate the foreign efficacy data and/or safety data to a new region. The term bridging study has since been adopted in clinical laboratory validation testing to bridge a new component (i.e., reagent, equivalent compound) into a modified assay by establishing equivalent performance between parallel testing of samples with the new and original components. Bridging studies are generally more applicable with EUA assays where it has been deemed acceptable by the regulatory authority (i.e., FDA approval for SARS-CoV-2 molecular assays). The need for a verification or validation is dependent on the specific language in the EUA IFU and your laboratory director. For example, the FDA recommended bridging studies for SARS-CoV-2 molecular assays could consist of testing three-fold serial dilutions of viral material in a pooled respiratory sample matrix in triplicate until a positivity rate of <100% was achieved. If the resulting Level of Detection (LoD) was the same as the unmodified authorized test LoD, then the two methods are considered to have equivalent performance. Important note: some IFU documents will not accept any modifications; therefore, a bridging study would not be applicable and an LDT validation would be needed.

Addendums

CLSI defines an addendum as an "ancillary report with additional information that expands or clarifies the original final diagnosis but does not change it. Examples include adding information derived from additional diagnostic studies, recuts from specimen blocks, or consultations with experts." Suppose diagnostic real-time PCR testing demonstrated that the positive PCR control being used resulted in unclear result interpretation due to clinical samples demonstrating low relative fluorescence (RF) for a specific target gene. Confirmatory testing showed that the low RF is due to a specific point mutation where the probe anneals; parallel testing of a new PCR control containing that point mutation clarifies result interpretation. This study would be included as an addendum to the SOP.

Instrumentation Verification or Validation

New Instrument

When bringing online a new instrument or test system, reference the Qualitative Assays section for performance characteristics that must be considered for verification or validation. This applies even when the new instrument is the same as one already in use. The manufacturer's recommendations must be reviewed and either accepted or a justification for a deviation must be provided, noting this may lead to the FDA-modified path. Instrument maintenance and troubleshooting must be included in the protocols for establishment of the new method.

Additional Instrument

For a test system or instrument in which there is already an approved method, and the aim is to add an additional test system or instrument of the same make or model, an additional verification or validation must be performed. It is up to the discretion of the laboratory director to determine the amount of testing to be involved in these studies, but the same criteria must be met as if bringing on the assay for the first time. Some criteria may be satisfied by the original verification or validation and a smaller study could be carried out to satisfy the requirement. If the intention is to bring on multiple new instruments, the studies may be divided across instruments as long as each instrument is involved in all parts of the study (accuracy, precision, reportable range) and the serial numbers are identified with the correlating data.

Instrument Move or Repairs

Any movement or repair of an instrument requires additional verification to ensure the alteration did not affect the test performance on that instrument. These studies are typically abbreviated when compared to a new method verification or validation and are up to the discretion of the laboratory director.

Instrument to Instrument

If the laboratory uses more than one instrument or method to test for a given analyte, the instruments or methods must be checked against each other at least twice a year for comparability of results. This is typically aimed to be performed every six months. A subset of samples (QC and/or patient samples) may be compared on all test systems for a given analyte.[1,2]

Reagent Lot-to-lot Comparison

It is important that each new reagent is verified upon receipt prior to being put into routine use to report patient results.[3] Reagents are validated by the manufacturer and must meet quality standards before release. However, there are several factors that could affect reagent performance once received by the laboratory that include:

- Changes in reagent component materials
- Instability of a component in the reagent
- Reagents compromised in transportation and/or storage
- Incorrect calibration of the new reagent lot.

Differences in performance between current and new or candidate reagent lots hold a potential risk for patient results, which needs to be assessed for all reagent lot changes. Assessment can occur before, or concurrent with, initial use of the candidate reagent lot for patient testing. It is important that verification of candidate reagent lots is not emergent; therefore, ensuring sufficient stock of the previously verified or current reagent lot is essential.

There are no universal acceptance criteria for defining the degree of difference in laboratory results between clinical samples that will influence the clinical decision. The laboratory director must determine the limits of acceptable differences with the primary objective that both the current and candidate reagent lots yield the same clinical interpretation of results.

Note: CLSI document EP26[1] Appendix A-D provides calculations and statistical considerations to help determine the number of samples needed for reagent lot comparisons and rejection limits.

Defining When to Perform Comparisons

Reagent Comparisons

Good laboratory practice would be to evaluate each new lot and shipment of reagents received. However, the following circumstances **do not** require a full evaluation of a new shipment or candidate reagent lot:

- A new shipment received in the laboratory that has already been evaluated and deemed acceptable for testing with clinical samples. Since the reagent lot was previously determined to not have a significant lot-to-lot difference with clinical samples, any potential shipping damage will be reliably detected using the QC samples. Therefore, each new shipment of the same reagent lot may be verified by testing QC samples only.
- Affiliated laboratories using the same reagent lots and vendor instruments do not need to repeat verification using clinical samples since an initial verification of consistency between current and candidate reagent lots would have been performed. However, all affiliated laboratories need to check each new shipment of reagents for shipping damage using QC samples for verification.
- If there are several of the same instruments used in a single laboratory, verification of the candidate reagent lot only needs to be performed with clinical samples using a single instrument, not all instruments used in the testing.

Methodology Comparisons

Careful attention should be given to laboratories who test the same analyte with different methodologies. In particular, if a laboratory employs more than one methodology to test for an analyte, verification studies should be conducted on both methods. This is true for quantitative and qualitative test procedures.

Examples of this would include the following scenarios.

- A laboratory routinely utilizes a moderate test kit on an analyte that is reported qualitatively. The laboratory decides to employ a back-up test kit due to shortages or other reasons. The laboratory validates and begins testing using the replacement test kit. The laboratory wishes to utilize both test kits when the original test kit is back in stock. In this case, a comparative validation would be warranted.
- The laboratory routinely conducts moderate qualitative kit testing using chromatographic immunoassay as well as moderate instrument testing on the same analyte using automated chemiluminescence. In this case, a comparative validation would be required. If the qualitative result is determined by a numerical value (i.e., concentration of the analyte in the specimen), the method comparison should seek to correlate with these quantitative values. The laboratory should also define the relationship between the two methods during the evaluation process.

Sample Type, Number and Volume for Comparisons

The optimal samples for reagent comparison testing are native clinical samples appropriately collected using validated methods, properly stored, and transported under appropriate conditions. Reagent lot-to-lot comparisons should not be performed using QC samples only since observed differences between current and candidate lots could be attributed to a matrix effect (influence of sample property independent of analyte presence) with no change in patient sample measurements. Inversely, it is also possible to have no change in the QC result, but patient results are significantly impacted by reagent lot-to-lot differences. In addition, manufacturer supplied QC material may be optimized to perform correctly with each new reagent lot underscoring any lot-to-lot reagent effect on clinical samples. Therefore, performance of new reagent lots should be evaluated using both patient and QC samples with few exceptions.

Sample Type

- Native clinical samples collected using validated methods and sources.
- Concentrations of the clinical samples should span the measurement interval, or medical decision of interest. If a range cannot be obtained, sample concentrations at two key medical decision points (true positive and true negative) is sufficient. Typically, concentrations near the QC concentrations are clinically relevant.

- Residual clinical samples can be used. It is important to re-measure the clinical samples using both the current and candidate reagent lots, as close in time as possible, to determine any alterations in concentration or activity that are affected by time or storage conditions.
- QC samples are included with clinical samples in each reagent lot verification.

The use of sample material other than native or pooled clinical samples (described in **Sample Volume** below) should only be considered when the aforementioned sample types are unavailable. This could include samples with stability limitations. In these instances, the material used for external evaluation of performance (proficiency testing [PT], external quality assessment, QC material, internal challenge samples) may be considered.

Sample Number

A minimum of three clinical samples is recommended to reduce any potential bias of a single sample. The number of replicate measurements per sample increases the power and reduces error. However, increasing replicate measurements alone, rather than testing more samples, increases the risk that reagent lot or patient sample interactions will not be evaluated.

Sample Volume

- Estimate twice the combined measurement procedure plus dead volume. It is recommended to have sufficient sample volume to allow for repeat testing.
- Samples can be pooled if it is not feasible or practical to obtain sufficient volume per sample. Pools need to be tested using both current and candidate reagent lots, and previous individual sample results cannot be used. A minimum number of unique comparisons should be three (same as individual native samples).

Reagent Comparison Failures

If there are any clinically significant differences in patient results between the current and candidate reagent lots, the candidate reagent lot should not be used until the discrepancy is resolved. Initially, the instrument should be calibrated using controls to minimize any differences that may be attributed to calibration-to-calibration variability. More robust evaluation of the current and candidate reagent lots can be performed using a minimum of 40 clinical samples that span the measurement interval. However, this approach may not be feasible, and the manufacturer should be contacted for follow-up investigation.

If the candidate reagent demonstrates a significant shift in the QC values only, and acceptable performance with the clinical samples, then the QC target values should be updated to reflect the changed performance with the candidate reagent lot. QC material performance may change with different reagent lots, which are typically artifacts of the interaction of the reagent with an altered sample matrix of the QC material. Note: Some local regulatory requirements may restrict any changes to QC target values, which would require a more extensive reagent comparison.

Software Upgrades

Software upgrades for clinical laboratory instruments, or process and reporting management at all levels, is a complex process.[4] Any software upgrades that will negatively affect the institution or patient results should be carefully and thoroughly considered before proceeding. In particular, the level and extent of software upgrade testing necessary is largely dependent on the total risk associated with failure: the greater the potential harm associated with failure, the more extensive planning and testing required.

Prior to implementation of software upgrades (whether using new software programs or simply updating to a new version of the current software), the following factors should be considered at minimum:

- Features and functions
- Cost
- Service and upgrade terms and policies
- Hardware requirements
- Performance characteristics (database access and size)
- Compatibility with other institutional systems (i.e., laboratory information system [LIS])
- Data interchange formats supported (e.g., third party applications, Health Level 7 [HL7] standards)
- Data security.

The process of software testing may include boundary testing, input testing and stress testing:

- **Boundary testing** is the use of test cases generated using the extremes of the input value range (i.e., maximum, minimum, within the limits, expected or typical values, and out-of-range [error] values).
- **Input testing** assesses the entry of data input function values including valid, invalid, null and zero, and incorrect data types.
- **Stress testing** is a more general type of focused, thorough testing used to evaluate the stability of a given system, including testing beyond normal operational expectations of a system in order to observe how failure conditions are handled.

Software version upgrades that improve the overall functionality of the system (i.e., enable better visualization, more detailed report summaries, improve the operator experience, etc.) may justify a verification whereas version upgrades that add significant features that could affect the quality or accuracy (e.g., MALDI-TOF database updates including new reference library organisms) necessitate more comprehensive validation of the new system. The software vendor may release technical bulletins or alerts that recommend or require a software version upgrade. Follow instructions from the vendor to correctly upgrade the software package(s). It is important to validate or verify the software upgrade prior to implementation in the laboratory workflow: ensure the previous software version is available for comparison testing until the software upgrade or new software has been approved.

Verification of software updates uses previous data or known controls to run through the system verifying all the potential result outcomes. The verification should ensure that boundary and input testing are performed to ensure that there is a range of valid input and output values obtained. An example is a qualitative molecular test with Ct value cutoffs of <35 (positive), 36-40 (indeterminate), >40 (negative), and control failure (invalid). The verification should ensure that Ct values within each respective range yield the correct end result, which should match the previous software version. If the end results do not match, contact the software vendor to troubleshoot and/or perform a more extensive validation of the system.

Validation of the software upgrade should follow the key elements described in Qualitative Assays using the manufacturer or vendor data or documentation as the comparator. A validation will likely encompass an entire workflow focus, which would include both wet-lab and dry-lab testing. For example, a database upgrade for the MALDI-TOF would require culturing standardized microorganisms to spot onto the MALDI target plate followed by system comparison of the

microorganism mass spectrum with the database library of reference mass spectra. If the end results do not match the vendor specifications, contact the software vendor to troubleshoot.

The implementation of a new LIS would be another example of a software upgrade requiring a validation. The clinical implication of improper LIS functionality has high risk associated with failure. Therefore, extensive boundary, stress, and input testing methods should be applied to test cases performed in an off-line static testing environment and during parallel testing once the system is 'live.' The length of 'live' parallel testing should be determined based on risk associated with failure and decided on by the QAO and laboratory director.

The system upgrade validations or verifications are deemed acceptable once all relevant measures of performance can be demonstrated. The major features and functions of the system should have been sufficiently challenged and all test cases or samples should have passed. Failed test cases should be repeated and reviewed to determine if the error is obvious and can be rectified. Any failed cases or unexplained conditions need to be analyzed and determined to be noncritical for the system upgrade to pass.

Additional Related Processes Resources

Templates

- Bridging, Addendum and Extension Studies Template
- Reagent Comparison QC Studies Template
- Software Update SOP and Template

Example

- Analytical Instrument Verification Example

References

1 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: www.ecfr.gov/current/title-42/chapter-IV/subchapter-G/part-493/subpart-K/subject-group-ECFRc96daead380f6ed/section-493.1253

2 CMS CLIA. Calibration and Calibration Verification Brochure. Accessed Jan 14, 2024 from: www.cms.gov/files/document/clia-brochure-calibration-and-calibration-verification-april-2006.pdf

3 CLSI. User Evaluation of Acceptability of a Reagent Lot Change. 2nd ed. CLSI Guideline EP26. Clinical & Laboratory Standards Institute. 2022. Available from: https://clsi.org/standards/products/method-evaluation/documents/ep26/

4 CLSI. Laboratory Instruments and Data Management Systems: Design of Software User Interfaces and End-User Software Systems Validation, Operation, and Monitoring, 2nd Edition. CLSI document AUTO13-A2. Wayne, PA: Clinical and Laboratory Standards Institute; 2003. Available from: https://clsi.org/standards/products/automation-and-informatics/documents/auto13/

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.