General Checklist for Establishment of Performance Specifications for Tests Not FDA-Cleared or Approved

Compiled by the APHL/CDC STD Steering Committee
October 15, 2009

General Checklist for Establishment of Performance Specifications
Tests Not FDA-Cleared or Approved

According to the Clinical Laboratory Improvement Amendments (CLIA)[1], each laboratory that modifies an FDA-cleared or approved test system, or introduces a test system not subject to FDA clearance or approval (including methods developed in-house and standardized methods such as text book procedures, Gram stain, or potassium hydroxide preparations), or uses a test system in which performance specifications are not provided by the manufacturer must, before reporting patient test results, establish for each test system the performance specifications for the following performance characteristics, as applicable:

(i) Accuracy.
(ii) Precision.
(iii) Analytical sensitivity.
(iv) Analytical specificity to include interfering substances.
(v) Reportable range of test results for the test system.
(vi) Reference intervals (normal values).
(vii) Any other performance characteristic required for test performance.
(viii) Based upon the performance specifications verified or established, the laboratory must determine the test system's calibration and control procedures for patient testing.

Although it is preferable that FDA cleared or approved tests be used in clinical laboratory testing, there are instances where these tests do not exist, or need to be used in a manner not covered by the product insert, and there is a critical need for the testing. This checklist has been developed as a guide to aid laboratories in verifying and documenting the performance characteristics of tests that are not FDA-cleared or approved, or for verifying and documenting off-label testing (using an FDA cleared or approved test in a manner not covered in the product insert).

This checklist is intended to be used as a general guideline for the verification of all tests that are not FDA-cleared or approved, however, some of the information may be more specifically geared to molecular testing for infectious disease agents.

This checklist has been reviewed by laboratory professionals at the Centers for Medicare and Medicaid (CMS). However, the local CLIA inspector approves the final verification of performance characteristics. In some instances, it may be beneficial to clear your verification study design with your local CLIA inspector prior to the initiation of verification testing.

☐ **Protocol to Establish Performance Characteristics**

The verification protocol for establishing performance characteristics should include the elements listed here. Examples for two STD related tests are available as additional guidance.

- **Objective**
  This statement will include the purpose of the protocol, stating exactly which testing platform and methodology is being used to establish performance characteristics.

- **Study Design**
  The study design will be an overview of the verification procedure that will be used to establish the performance characteristics of the assay.

- **Materials and Methods**
  This section will include the actual procedure(s) for establishing the performance characteristics, including how specimens will be obtained, the number of specimens that will be tested, and the actual test procedure itself. Included in the Materials and Methods section should be the

[1] CLIA Sec. 493.1253 Standard: Establishment and verification of performance specifications, B (2) Establishment of performance specifications.

procedure for resolving discrepant results, or for addressing additional means to meet criteria, such as increasing the sample size, should this be necessary.

- **Acceptance Criteria**
  Prior to the initiation of the study, acceptance criteria will be established for each performance characteristic. The Laboratory Director or designee should approve these criteria before proceeding.

  If comparing two like methodologies, accuracy may achieve 100% - all specimens should agree. However, if comparing two different methodologies, it would be unrealistic to set your accuracy performance characteristics at 100%. For example, culture may be the gold standard, but real time PCR will be more sensitive, so 100% correlation would not be expected.

  A Coefficient of Variation (CV) of 10% or less could be an acceptable range for precision, but the actual numeric values that are measured must be taken into account, as very small variations in very low numbers can result in large CVs.

- **Evaluation and Conclusions**
  This section will be a summary of the results of the protocol. It will document how well the assay performed in each of the performance characteristics, and make a recommendation for either accepting or not accepting the new or modified assay.

☐ **Performance Characteristics**:

- **Accuracy**
  Validations using authentic clinical specimens are preferred, but spiked clinical specimens are acceptable. When spiking specimens, the matrices should be true specimens (cervical swab, sputum, urine, CSF).

  When documenting accuracy, a quantitative measurement should be used rather than qualitative measurements whenever possible. Examples are measurements in Relative Light Units (RLUs), or Crossing Thresholds. A minimum of 10 positive and 10 negative previously characterized specimens should be tested. Ideally, a total of 30 specimens should be tested, and results should be within the pre-established acceptance criteria set forth in the protocol.

  These are minimal numbers for sample size. It may be necessary to either increase or reduce the sample size depending on the type of test being verified and the availability of appropriate specimens. There may be instances when it is difficult to obtain a sufficient number of positive specimens for every specimen source being verified. In this case, the sample size may need to be reduced. Random statistical error may require that a larger sample size be required in order to meet acceptance criteria. If this problem occurs, it is recommended that another 30 specimens be tested. If acceptance criteria still cannot be met, additional discrepant analysis may be required or the pre-determined acceptance criteria may need to be re-established.

  If two different methods are being compared (e.g., culture and NAAT) with the object of validating one of the methods, then all discordant results need to be resolved. Attempts to resolve these discrepancies could include retesting by one or both methods or referral to another facility for testing.

  Specimens for accuracy testing can be obtained by:

1. Collaboration with another laboratory who has already validated the assay
2. Spiking samples, preferably blinded
3. Commercially purchased validation specimens

- **Precision**
  Both Intra-run and Inter-run reproducibility should be documented using both a negative specimen and a positive specimen.

  For Intra-run reproducibility, the same negative specimen is repeatedly tested in one run 10 times, and the same positive specimen is repeatedly tested in one run 10 times. The CV and Standard Deviation of these 10 specimens are determined and compared to acceptance criteria. Alternatively, if specimens are in short supply or the values obtained are only qualitative, the test specimens can be run in duplicate on one run.

  For Inter-run reproducibility, the same negative specimen from the intra-run study is run 5 additional times on 2 subsequent runs. The first 5 observances from the intra-run study are included with the additional 10 observances resulting in 15 results, tested on 3 different test runs. The CV and Standard Deviation of these 15 specimens are determined and compared to acceptance criteria. This same procedure is performed using the positive specimen. Alternatively, if specimens are in short supply or the values obtained are only qualitative, the specimens need only be tested on a subsequent run, in singlet.

- **Analytical Sensitivity**
  Analytical sensitivity to determine the limit of detection can be determined using dilution studies. The lowest amount of target that can be detected should be tested at least 3 times to ensure the reliability of the limit of detection. When the assay can be performed on multiple specimen types (matrices), each matrix (cervical swab, sputum, urine, CSF) must be evaluated.

- **Analytical Specificity**
  To demonstrate that cross-reactions do not occur, and that there are no interfering substances, specimens positive for related organisms, or specimens containing inherent substances known to cause interference, need to be tested. Normal flora commonly found at the test site must also be shown not to cross-react with the test system.

  To demonstrate that the test system detects all known strains, species or subtypes of the target, a representative sample of these strains must be tested.

- **Reportable Range of Test Results**
  Reportable range is defined as the upper and lower limits over which the laboratory can establish or verify the accuracy of the instrument or test system measurement response. Establishment or verification of the reportable range of patient test results may be accomplished by assaying low and high calibration materials or control materials, or by evaluating known samples of abnormal high and abnormal low values. This may not pertain to an assay that is determining presence/absence of a target. However, if an equivocal range will be reported, you will need to verify the response of the test system by using QC or reference materials known to be in that range.

- **Reference Intervals**
  A reference “range” or “interval” is defined as the lower and upper limit expected for a designated population of healthy (normal) individuals. Establishment or verification of the reportable range may be accomplished by evaluation of an appropriate number of specimens from presumed

healthy or uninfected individuals. This testing verifies published reference ranges or package insert data. The reference range for any laboratory-developed tests must be appropriate for the laboratory's patient population and reflect the type of specimen and demographic variables such as age and gender, as applicable.

- **Other Applicable Performance Characteristics**
  There could be instances when other performance characteristics need to be documented in the course of verification.

- **Calibration and Control Procedures**
  When establishing the calibration and/or quality control frequency, the laboratory must consider the stability of the instrument and reagents, the frequency in which testing is performed, how robust the test method is, how often quality control fails, and the training, experience, and competency of the laboratorians performing the assay. In the early establishment of performance characteristics, it may be prudent to restrict testing to only one well trained laboratorian, and perform QC more frequently.

☐ **Standard Operating Procedure, including Quality Assurance measures**
A standard operating procedure, written in CLSI recommended format[2], should be in place prior to a test being put into practice. This SOP should include the principle, the supplies needed, the actual performance of the procedure, quality control measures, interpretation, results reporting, limitations of the procedure and references.

☐ **Training Documentation**
A training checklist or some other form of documentation is needed to demonstrate that staff, in addition to the person performing the verification, have been trained in the new procedure.

☐ **Documentation Review and Approval by Laboratory Director/Designee**
Verification studies should be approved and signed by the Laboratory Director prior to a test being put into practice and prior to any results being reported.

[2] Clinical and Laboratory Standards Institute. (2006). Clinical Laboratory Technical Procedure Manuals; approved guideline-fifth edition. CLSI Document GP02-A5.Wayne, PA: CLSI.