Public Health Laboratory Division
601 Robert Street North
St. Paul, Minnesota 55155 (651)
201-5200

Document Number:	DOC-452
Document Title:	Guidelines for Verification and Validation of Laboratory Methods
Document Notes:	LDT, accuracy, precision, reportable range, sensitivity, specificity, reference

Document Information

Revision:	3	Vault:	CLIN Public Published-release
Doc Type:	PHL-CLIN-13.3.20	Status:	Release

Date Information

Effective Date:	03 Jan 2017	Next Review Date:	29 Dec 2022
Release Date:	03 Jan 2017	Expiration Date:	

Control Information

Author:	BUSHM1	Previous Number:	G 1.17A
Owner:	PHL-r-CLIN-QAO	Change Number:	ADV-262

The MDH Public Health Laboratory Division posts approved documents in the MasterControl system. Controlled copies may only be created by authorized persons as allowed under applicable document control procedures. The signature manifest on the final page of each document displays the list of reviewers and approvers. A watermark appears on published versions of archived documents and documents protected under the Minnesota Data Practices Act.

MDH Public Health Laboratory Division	Document: DOC-452
Title: Guidelines for Verification and Validation of Laboratory Methods	Revision: 3

Title: Guidelines for Verification and Validation of Laboratory Methods

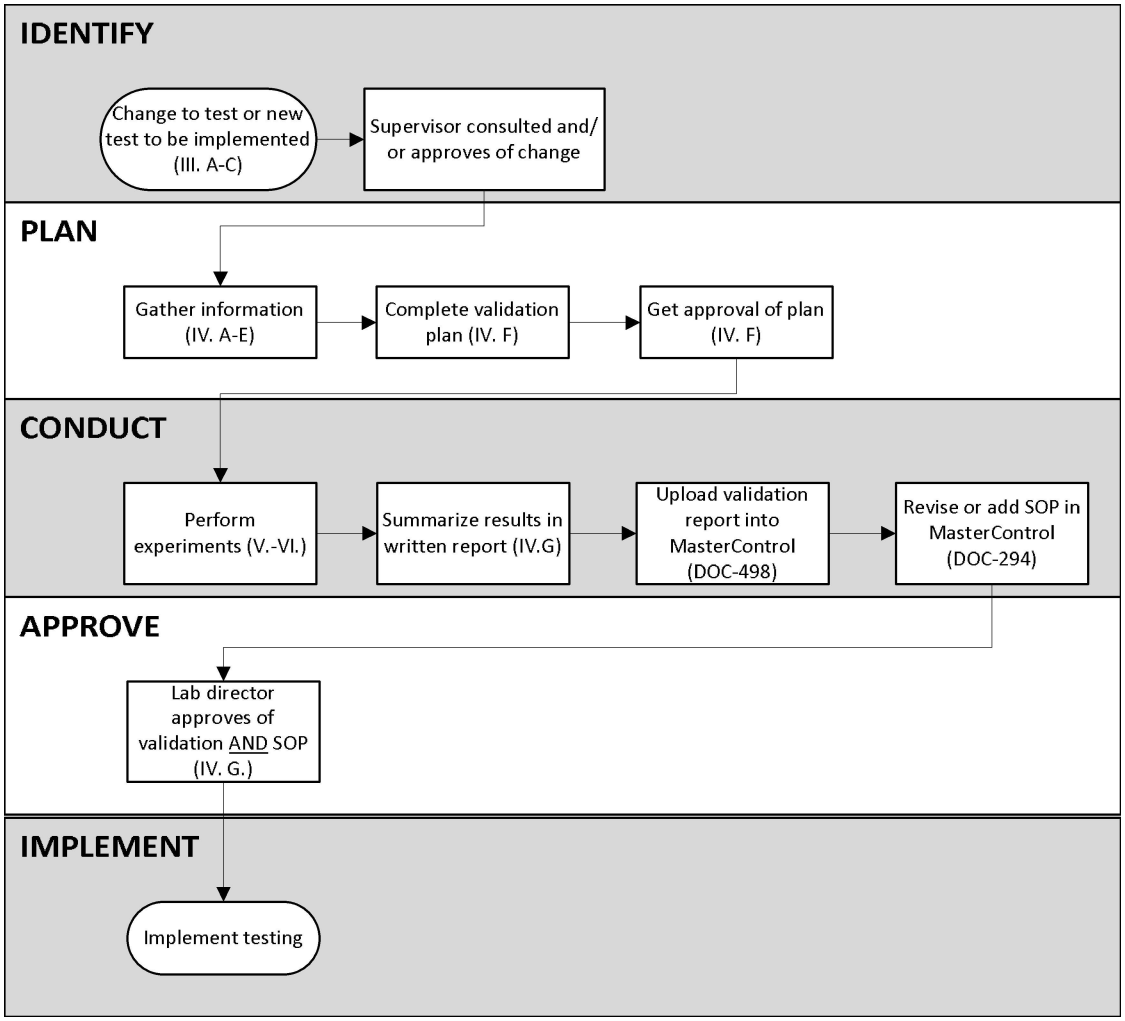
Table of Contents:

I. Purpose

The Clinical Laboratory Improvement Amendments of 1988 (CLIA '88) include specific recommendations for validation and verification of test methods used for purposes of disease diagnosis. Performance characteristics including analytical sensitivity, analytical specificity, accuracy, precision, reportable range, reference intervals and any other characteristic required to monitor test performance must be determined and documented prior to reporting patient test results. This validation needs to be done for in-house laboratory developed tests (LDT) or for tests that involve modifications of FDA-approved test methods. In addition, FDA approved kits must be verified prior to use in patient diagnostic testing, however only accuracy, precision, reference range, and reportable range must be determined for these test kits. Some manufacturers provide analyte specific reagents (ASRs), which are manufactured under GMP conditions and may be used in development of LDT's. These ASRs are not FDA-approved, however, and require the same level of validation as do other in-house developed tests. Not all performance characteristics listed above will apply in all validation and verifications, but as many of the characteristics as are reasonably applicable should be evaluated.

Method verifications and validations are an integral requirement in assuring the quality of testing performed at the Public Health Lab. The purpose of a method validation is to assess the error of an assay or test system to be sure that the amount of error will not adversely affect the intended use of the test, for example patient care. A set of experiments to assess performance of the assay allows for the determination of error in the method. Method performance is judged acceptable when the observed error is less than or equal to the allowable error. Method verification and validation must be performed and signed off on by the laboratory CLIA director prior to placing laboratory tests into routine use for reporting of patient results. This document is meant as a guide to assist in this process and changes should be made in the experimental details where and when appropriate for the test method.

An outline of the process is demonstrated in the graphic below.



II. Definitions

A. **Accuracy.** Accuracy is the reliability of the test, i.e. the ability to return a correct result compared with a reference method.

B. **Analytical Sensitivity.** Analytical sensitivity is the lowest amount of target matrix that still returns a positive result, i.e. the limit of detection.

C. **Analytical Specificity**. Analytical specificity is the ability of an analytical method to measure only the sought-for analyte or target. Analytical specificity is typically demonstrated by testing the ability of the assay to discriminate between the target analyte and other related analytes. Analytical specificity determination may also include tests to demonstrate that substances present in clinical specimens do not interfere with assay performance.

D. **ASR.** Analyte specific reagent. The FDA classifies ASRs as "antibodies, both polyclonal and monoclonal, specific receptor proteins, ligands, nucleic acid sequences, and similar reagents which, through specific binding or chemical reaction with substances in a specimen, are intended for use in a diagnostic application for identification and quantification of an individual chemical substance or ligand in biological specimens." See 21 CFR 846.4020 for more information.

E. **Bias.** The difference between the averages of the values by the test method and a comparative method in a comparison of methods experiment. Bias indicates systematic error.

F. **Constant Error.** Systematic error in which the difference between the test method and the comparative method stays the same over a range of concentrations.

G. **Gold Standard.** This is the test method or procedure that is widely recognized as the best available to determine true disease state of a patient.

H. **Laboratory-Developed Test (LDT).** A LDT is any test, used in a single laboratory, that has not been approved or cleared by the FDA. This includes any FDA-approved or cleared kit that has been modified in any way by the laboratory. Modifications include, but are not limited to: changes in specimen stability, accepting different specimen types than what is listed in the package insert, any changes made to the procedures or reagents, etc.

I. **LoB.** Limit of blank. Highest measurement result that is likely to be observed with a stated probability for a blank sample.

J. **LoD.** Limit of detection. This is the lowest amount of target that can be detected by a test system with a stated probability.

K. **LoQ.** Limit of quantification. The lowest amount of analyte that can be quantitatively determined with a stated acceptable precision and trueness, under stated experimental conditions.

L. **Precision.** Precision is the reproducibility of the test, i.e. the ability to consistently produce the same result on repeated testing of the same sample over time.

M. **Proportional Error.** Systematic error in which the difference between the test method and the comparative method changes as concentration changes.

N. **Qualitative Assay.** A qualitative assay is one in which the results are reported out as either positive/negative or present/absent. The reported result may be based on a numerical value.

O. **Quantitative Assay.** A quantitative assay reports out a numerical value as a result.

P. **Random Error.** An error that can be either positive or negative, whose direction and exact magnitude cannot be predicted. Also known as imprecision.

Q. **Reportable Range.** The reportable range is the range of values that can be accurately measured by the assay. This may not apply for testing in which the analyte is either present or absent.

R. **Reference Range.** The reference range is the range of values found in healthy persons who do not have the condition for which the test is performed.

S. **Semi-quantitative Assay.** A semi-quantitative assay is one in which results are reported out as gradations such as 1+, 2+, 3+, etc. or as a titer.

T. **Systematic Error.** An error that is always in one direction and causes all the test results to be either higher or lower than the correct value. Also known as inaccuracy.

U. **Validation.** Although frequently used interchangeably with verification, validation actually refers to confirmation through objective evidence that the assay meets specified performance criteria to ensure generation of accurate, reliable, reproducible results that fulfill the intended use. Performance Characteristics required include:

1. Accuracy
2. Precision
3. Analytical sensitivity
4. Analytical specificity
5. Reportable range
6. Reference intervals (normal values).

V. **Verification.** This shall be performed for unmodified FDA-cleared tests where performance criteria have already been validated by the manufacturer for the intended use of the method. Performance criteria should meet the established manufacturer specifications. The performance characteristics to be evaluated include:

1. Accuracy
2. Precision
3. Reference range
4. Reportable range

III. Scope

A. Situations in which Verification/validation is required include but are not limited to:

1. Relocation or movement of an instrument or test system
2. If the same assay is performed on multiple instruments, each instrument must be verified/validated for that test.
3. When modifying an already validated/verified test system such as
   a. Changing reagent or specimen volumes used during the testing procedure
   b. Testing a non-validated/verified specimen type
   c. Adding an analyte to a test system
   d. Change in incubation times or temperatures
   e. Change in specimen handling instructions including specimen stability if specified by the manufacturer of the test
   f. Change in dilution of specimen or reagent
   g. Using a different calibration or reference material or changing the manufacturer's set points
   h. Introducing a different antibody (source, monoclonal-vs.-polyclonal, etc.)
   i. Change or elimination of a procedural step
   j. Change or addition of detector (conjugate) or substrate
   k. Change in solid phase

l. Change in cutoff or method of calculating the cutoff for semi-quantitative assays
m. Change in the endpoint or calculation of the endpoint
n. Addition of adsorbent
o. Change in the strain of antigen in serologic assays

B. Verification is required when:
1. The test system is FDA-cleared or approved. This includes ASR's made by another manufacturer that are FDA approved for use in the test system.

C. Validation is required when:
1. Modifications are made to FDA-cleared or approved test systems (see above)
2. Modifications to the intended use of FDA-cleared or approved test systems which includes but is not limited to:
   a. Using or promoting the test for another purpose (i.e. screening vs diagnostic)
   b. Changing the type of analysis (qualitative reported as quantitative)
3. "Home brew" laboratory developed tests (LDT's)
4. Test systems where the performance specifications are not provided such as text book procedures or non-FDA approved kits.

IV. Validation/Verification Considerations

A. Equipment Considerations

1. Determine the equipment, if any, used to run the test. If running the test on multiple analyzers, each analyzer will require a separate validation or verification. See Equipment Validation section for further information.
2. Ensure all environmental and operational requirements such as space, humidity, temperature, electrical and data ports, water, waste, or other requirements can be met.
3. Ensure that all function checks specified by the manufacturer have been performed prior to operation.

B. Equipment Validation/Verification

1. The depth and scope of validating/verifying an assay on multiple analyzers will vary depending on the equipment. A full validation/verification should be performed on the analyzer designated as the primary for the purpose of the validation/verification experiments. The secondary equipment should have all components of the validation/verification including day-to-day precision but the reference range experiment can be omitted.
2. When analyzing data, results of the accuracy, precision, and reportable range (for verifications), and analytical sensitivity and specificity (for validations) should be

calculated separately for each piece of equipment and then compared to the designated primary equipment. Each piece of equipment must meet the performance requirements for accuracy, precision, reportable range, analytical sensitivity, and analytical specificity where applicable.

3. **Equipment of the same make/model:** Second, third, etc. pieces of equipment of the same make and model may have less robust experiments performed with fewer samples as compared to the main study. This is because instruments of the same make and model should have similar performance. Studies still need to be completed because there is still a risk that each piece of equipment could perform differently and the magnitude of the bias between the two instruments could be unacceptable.
4. **Equipment of a different make/model:** Secondary equipment of a different make and model will usually require a full validation. This is because it is more difficult to assess whether or not the bias and imprecision will be acceptable and thus must be more fully characterized. Some exceptions may exist where a different model is used that has very minor differences that would be unlikely to introduce bias and imprecision. For example, if there is no difference in sample or reagent volumes, pipetting, or incubation temperatures, then there is a lower likelihood of a significant difference in error and a less robust validation/verification can be performed. An explanation with justification for this should, however, be included in the write up.

C. Reagents, Media, and Other Consumables

1. Ideally, the same lot of reagents/media, should be used throughout the validation/verification study except where indicated in the experimental details outlined below. Expiration dates should be long enough to complete the study.
2. Ensure sufficient supply of reagents, media, calibrators, controls, etc. are available to complete the validation/verification. Communicate any needs to media prep when necessary and order a sufficient quantity of supplies prior to beginning the study.

D. Specimens

1. The types of specimens that will be received for testing in the laboratory should be taken into account when developing the validation plan. Each sample type should be validated/verified separately to verify that no bias exists between sample types during the performance of the assay.
2. It is good practice to determine the number of specimens to be tested based on statistical significance to ensure confidence in the results. Variability within the assay, if known, literature references, variability in the patient population, and availability of samples may help determine the number of samples to test. See the experimental sections for further recommendations.
3. Sample stability may be an issue, especially for remnant samples. If using an FDA-cleared test, any modification to the manufacturer's stated sample stability

for testing constitutes a modification of the assay and therefore is subject to the more rigorous validation experiment requirements. Sample stability should be evaluated for all LDT's.

4. For LDT's, it is recommended that sample/analyte stability and interferences should be investigated through a literature search to help design the proper specimen quality assurance for the assay.

E. Choosing a reference or comparative method

1. The analytical method used for comparison should be carefully selected because the interpretation of the experimental results will depend on the assumption that can be made about the correctness of results from the comparative method.
2. When possible, a "gold standard" should be chosen for the comparative method. A gold standard, also known as a reference method, is a high quality method whose results are considered to be the most accurate.
3. Any discrepancies between a test method and the gold standard method are assigned to the test method. If a gold standard is not available, then comparison may be made to a "comparative method" which does not imply that the correctness of the method has been documented. Many routine laboratory procedures fall into this category. Any differences between the test method and a routine method must be carefully interpreted. Samples with discrepant results between the test and comparative methods should be reanalyzed to confirm that the differences are real. If the differences are large, then it is necessary to identify which method is inaccurate. If necessary, samples may be split and sent to a reference laboratory to help resolve discrepancies.
4. In some cases no appropriate reference method exists. In these cases it will be necessary to correlate the presence of an analyte with other data such as phenotype, epidemiological data, patient diagnostic data, or other means.

F. Validation/Verification Planning

1. A draft SOP should be written (for new assays) or revised (for modified assays).
2. Assess whether or not IRB approval is required. If required, complete the IRB approval process prior to beginning a validation.
3. The Validation Plan Template is used to organize what will be needed to complete the validation or verification. The experimental sections below (V., VI.) should be reviewed for guidance. Enough detail should be included to offer a clear understanding of the steps needed to complete the validation/verification. The plan should include but is not limited to:
   a. which components of the verification or validation (accuracy, precision, etc.) need to be included in the experimental considerations
   b. The sample types, numbers and replicates of specimens to be analyzed
   c. Identify which equipment the testing will be performed on
   d. Steps that will be taken to troubleshoot discrepant results.
   e. Acceptance criteria for each component (accuracy, precision, etc.)

4. A meeting with the QAO and unit supervisor to discuss the validation/verification plan is recommended when obtaining the minimum number of samples is not possible. A meeting is also recommended if there is a question about which components of the plan template are applicable or for any other reasons that may require advisement prior to completing the validation/verification.
5. A plan may not be needed for retrospective studies leading to assay modifications (such as retrospective review of data to re-evaluate assay cutoff values, etc.).
6. The written validation plan must be reviewed and approved by the testing unit supervisor and section manager and the quality assurance officer before completing the testing experiments for the validation/verification.

G. Final Validation/Verification Write Up

1. Upon completion of all validation/verification experiments, the Clinical Validation Report Template can be used to summarize the results of the validation/verification. Refer to the template for sections that should be included in the write up.
2. The write up should read as a standalone document and should cover the purpose of the validation/verification, give a summarization of each applicable performance characteristic, and if any discrepancies were encountered, an explanation should be included of troubleshooting steps taken and the result of the investigation. A conclusion must be included about whether or not the test method should be implemented based on the acceptance criteria. Any extra control measures determined to be needed should also be explained. This may be necessary when a limited validation is performed due to lack of available materials or reference standards. Any questions about this should be directed to the QA Officer.
3. The final validation/verification should be uploaded into MasterControl. See DOC-498 How to Complete Validation/Verification Documents in MasterControl for more information.
4. The test method may not be used until both the validation/verification and accompanying SOP have been approved and signed by the laboratory director.

V. Qualitative Methods

A. FDA-Cleared/Approved

1. Accuracy.

Accuracy is determined by a comparison of methods experiment that involves testing samples with known target values against a reference method. Samples may be compared against one or more reference methods. The purpose of evaluating accuracy is to determine systematic error, or bias, of the method.

a. **Testing Materials.** Testing materials should match or be as close to the sample matrix as possible. The preferred test materials are patient samples with known values tested against the reference method(s). Other sources of test materials include proficiency samples, quality control materials, calibration materials, and

samples of the appropriate matrix spiked with a known level of the analyte. Best practice is to include samples over a range of values such that some will be near the cutoff and others will have varying strengths of positivity and negativity.

b. **Number of samples.** Sources vary on the recommended number of samples to test but guidance by CLIA suggests 20 samples of each positive and negative value should be tested. This number is a minimum and an assessment should be conducted to determine if more samples may be necessary for a more statistically relevant evaluation. HIV assays, for example, would require a greater number of samples to validate accuracy since the repercussions to the patient of a false negative or false positive result are much greater. If obtaining 20 samples is not possible, testing fewer than 20 samples must be approved by the QAO or laboratory director.

c. **Experimental Design.** Per CLSI standards, it is recommended that the chosen number of samples be run over a minimum of five days to simulate a range of conditions over which samples would normally be run. The comparison of methods experiment (accuracy study) can also be combined in this fashion with the reproducibility (precision) study.

d. **Performance Criteria.** Manufacturer statements of accuracy should be used as the minimum performance requirements.

e. Data Interpretation.

i. If the reference method is a "gold standard", then calculation of sensitivity and specificity using a 2 x 2 table is appropriate as comparison with gold standard methods indicates diagnostic rather than analytical accuracy. Shown below is the standard format of a 2 x 2 table.

		Gold Standard	
		Positive	Negative
New Method	Positive	a	b
	Negative	c	d

Where:

Sensitivity = 100%*a/(a+c)

Specificity = 100%*d/(b+d)

Total Accuracy = 100% *[(a+d)/(a+b+c+d)]

ii. Where the comparative method is not a gold standard, then sensitivity and specificity cannot be determined and instead percent positive agreement

(PPA) and percent negative agreement (NPA) can be calculated instead. In this case, the same calculations apply where positive percent agreement and negative percent agreement are calculated the same as sensitivity and specificity respectively.

iii. If the comparator is an imperfect standard that may be inaccurate due to lack of sensitivity or specificity, such as culture or clinical diagnosis, then it may be appropriate to calculate a measure of overall agreement to ensure that the percent agreement is not due to chance alone. In this case a statistical measure of agreement, such as a Kappa coefficient may be calculated.

1. Kappa coefficient is a measurement of the degree of agreement between the methods above what is expected by chance alone. A kappa of or approaching one indicates that there is very good agreement. Kappas approaching zero indicate that the agreement is no better than chance. Negative kappa means that agreement is worse than chance.
2. The formula for Kappa is:

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

3. To calculate Kappa:

   From the 2 x 2 chart above, first calculate A and B
   A= (a+b)/(a+b+c+d)
   B= (a+c)/(a+b+c+d)

   Now, Pr(e) = (A*B) + [(1-A)(1-B)]

   And Pr(a) = overall percent agreement = 100%*(a+d)/(a+b+c+d)

   Now input Pr(a) and Pr(e) as calculated above into the formula for Kappa

4. There are many online calculators that will also calculate Kappa from a data set. One may be found at GraphPad's website: http://graphpad.com/quickcalcs/kappa1.cfm

2. Precision.

Precision can be determined by a replication experiment in which replicate testing of patient samples or controls are analyzed within and across runs over several days. Multiple testing personnel must be included in the experimental design for methods that are not completely automated, including automation of pipetting or other methods of sample introduction into the testing system. Precision evaluates the amount of random error of the method.

a. **Test Materials.** Test materials can be any target positive and negative sample

including QC, calibrator, proficiency test material, and patient specimens. Enough test material should be prepared to complete the study and consideration should be given to replicating the patient matrix as closely as possible. Materials should be near the cut off value as evaluating imprecision near the cut off will give the best estimation of error at medically relevant decision levels.

b. **Number of Samples.** At minimum, at least one positive and one negative specimen should be run although two positives and two negatives are recommended. See experimental details below.

c. **Experimental Design.**

i. Types of precision measurements: (All three should be performed.)

a) Within run (same sample run multiple times on the same day)

b) Between run (same sample run in different runs on the same day or different days)

c) Day-to-day (same sample run on different days)

ii. The number of measurements should be established and may be dependent to some extent on the availability of test material. Thirty tests for each precision measurement have been recommended, although 10 may be acceptable. A recommended experimental design would be to run 2 negative and 2 positive samples in triplicate over 5 days, thus obtaining 30 test results for each precision measurement requirement.

d. **Performance Criteria.** Manufacturer statements of precision should be used as the minimum performance requirements.

e. **Data Interpretation.** Results may be expressed as a percentage agreement (number of repeats in agreement/ total number of results x 100). Calculate the percent agreement within run, between runs, and total. If numerical results are available, calculate standard deviation (SD) and coefficient of variation (CV).

Standard Deviation $\delta = \sqrt{\sum(\bar{x} - x)^2 / N}$

Coefficient of Variation $CV = (\delta / \bar{x}) * 100\%$

Where δ is the standard deviation and $\bar{x}$ is the mean of the result of a quantitative test

3. Reportable Range.

When a numerical cutoff value is used to determine whether a test is positive or negative, running positive specimens near the cutoff can verify the cutoff, or limit of detection (LoD). This is typically accomplished by running one or several concentrations of samples near the lower limit of the assay in replicate. For this reason, the data acquired from the accuracy and precision studies may be enough to satisfy this requirement. If the assay has quantitated calibrators, rerunning the

calibrators as unknowns and comparing them with the expected result may also help meet this requirement. The manufacturer claim for the cutoff is verified when 95% or better of the replicates are positive for the stated concentration. If no LoD is given by the manufacturer, the accuracy and precision claims of the manufacturer at the cutoff should be used as the acceptance criteria.

4. **Reference Range (normal values).**

CLIA requires that the normal values stated by the manufacturer are appropriate for the laboratory's patient population.

- For PCR and microbiological assays for infectious disease agents, the reference range is typically negative as defined by testing samples known to be free of the disease agent. Results of the accuracy study may be used to demonstrate that the assay detects the target in infected patients and not in uninfected patients.
- Since many qualitative tests detect for the presence or absence of an organism, the normal state may simply be that the specimen is expected to be negative. In this case, no specimen testing may be necessary and documentation of a literature search or the manufacturer's package insert will suffice.
- Should it be necessary to verify that the normal values stated by the manufacturer apply to the patient population of the labortory, 20 normal patient samples should be run and no more than 90% of patient samples should fall outside the manufacturer's stated range. If the manufacturer does not provide a normal range and one needs to be established, guidelines recommend testing samples from a minimum of 40 individuals although 120 is recommended but may be difficult to obtain.

B. Laboratory Developed Test or modified FDA-Cleared Test.

1. **Accuracy.**

a. **Test Materials.** Follow the guidelines above for FDA-Cleared/Approved assays.
b. **Number of Samples.** At minimum, follow the guidelines above for FDA Cleared/Approved assays. In addition, it may be advisable to include additional positive and negative samples. Many reference documents suggest a minimum of 40 or more of each positive and negative samples for non-FDA-Cleared assay validations, but the quality of the specimens is more important than the quantity.
c. **Experimental Design.** Follow the guidelines above for FDA-Cleared/Approved assays.
d. **Performance Criteria.** Non-FDA-Cleared/Approved assays are less likely to contain information regarding the performance specifications of the test unless the assay is a modification of an FDA-Cleared test. Ideally, the test being evaluated should show 95% agreement with the reference method. However, some assays such as molecular methods frequently are more sensitive than reference methods, such as culture, so allowance may be made for an imperfect standard and the expected percent agreement may be lowered.

e. **Data Interpretation.** Follow the guidelines above for FDA-Cleared/Approved assays.

2. Precision.

a. **Test Materials.** Follow the guidelines above for FDA-Cleared/Approved assays.
b. **Number of Samples.** At minimum, follow the guidelines above for FDA-Cleared/Approved assays. In addition, it may be advisable to include additional samples.
c. **Experimental Design.** Follow the guidelines above for FDA-Cleared/Approved assays.
d. **Performance Criteria.** Non-FDA-Cleared/Approved assays are less likely to contain information regarding the performance specifications of the test unless the assay is a modification of an FDA-Cleared test. Literature searches may be helpful as well. If numerical data is available and SD and CV can be calculated, CV should generally not exceed 20%.
e. **Data Interpretation.** Follow the guidelines above for FDA-Cleared/Approved assays.

3. **Reportable Range.**

Follow the guidelines above for FDA-Cleared/Approved assays.

4. **Analytical Sensitivity.**

Analytical sensitivity is determined by running dilutions of samples around the limit of detection (i.e. **Limit of Detection Experiment)**. For microbiological samples, this may translate to the number of CFU's or lowest amount of antigen an assay can detect. For other qualitative assays such as molecular or some immunological methods, the LoD is often near the cutoff. Understanding the LoD of these methods is particularly important for monitoring new lots of reagents and making sure their use does not change the limit of detection.

a. **Test Materials.** Ideally, the background matrix should be similar or the same as that of patient samples the laboratory intends to assay. Test materials should be quantitated samples at the expected lower limit of the assay which may include patient samples, calibrators, control materials, or the target analyte spiked into an appropriate matrix. Dilutions of these samples may be made in order to reach the desired concentration and several dilutions of each sample may need to be included in the experiment.
b. **Number of Samples.** As few as one specimen may be tested, but the recommended number is four low positive. The number of specimens needed should be determined on a case by case basis with input from the supervisor. Often analytical sensitivity can be evaluated at the same time as precision, accuracy, and reportable range minimizing the number or runs necessary.
c. **Experimental Design.** Samples should be run in duplicate over a minimum of

three days. Literature and CLSI guidelines recommend a minimum of 20 measurements for each sample concentration to verify a manufacturer claim and 60 measurements to establish the LoD. Number of samples, days and replicates may be adjusted to meet these goals.

d. **Performance Criteria.** The limit of detection should be defined as the concentration in which the assay can distinguish positive specimens from negative specimens 95% of the time (i.e. 95% or better of samples are positive at the stated concentration).
e. **Data Interpretation.** See performance criteria above.

5. **Analytical Specificity.**

Analytical specificity, or cross-reactivity, is evaluated by testing a panel of similar, potentially interfering organisms, substances, or analytes in interference experiments to assess constant systematic error. Recovery experiments may also be used to assess proportional systematic error due to inhibition, an important consideration in molecular assays.

a. **Test Materials.** Test agents should include as many organisms as possible that may be found in the relevant test sample or that cause the same symptoms as the target agent. For example, a method to detect *Bordetella pertussis* in respiratory samples should be tested against organisms commonly found in the respiratory tract such as *Chlamydia pneumoniae*, *Streptococcus pneumoniae*, etc. Other materials may include targets that are structurally similar or otherwise closely related to the target being tested or other substances known to interfere. Consult the literature to determine relevant potential interferants. For molecular assays where the sequences of the primers and probes are known, consulting a genomic database for potentially cross reacting organisms may be helpful. All specimen types intended to be tested with the assay should be included to assess the potential effect of interfering substances in each matrix.
b. **Number of Samples.** For the interference experiment, the recommended minimum is to test 3-5 samples containing each of the potentially interfering or cross-reacting test organisms, analytes, or substances. If performing the recovery experiment, multiple concentrations of the potential interferant should be tested.
c. **Experimental Design.** For a recovery experiment, samples containing the analyte should be prepared in pairs. Ideally, the analyte concentration should be close to the cutoff. One of the paired samples should contain a defined concentration of the potential interferant and the other sample should contain the same amount of water or other appropriate diluent. Samples should be run in duplicate. For determining cross reactivity, analyte negative samples with potentially interfering organisms or analytes should be tested. Samples should be run in duplicate.
d. **Performance Criteria.** For the recovery experiment, positive samples at each concentration tested should remain positive at least 95% of the time. For cross reactivity, results of each organism or analyte tested are expected to be negative at least 95% of the time. Additionally, a review of the literature may be used to set alternative performance standards to those listed above.

e. **Data Interpretation.** Record the results of testing and note any cross-reactivity with non-target agents or inhibition of amplification. If cross-reactivity is encountered, the assay conditions may need to be changed and the assay must be reevaluated. If cross-reactivity cannot be eliminated, then it must be noted as a limitation of the assay procedure. If inhibition is found in molecular methods, an inhibition control must be included in each assay run.

6. **Reference Intervals.**

Follow the guidelines above for FDA-Cleared/Approved assays.

VI. Quantitative Methods

A. FDA-Cleared/Approved

1. **Accuracy.**

Accuracy is determined by a **comparison of methods experiment** that involves testing samples with known target values against a reference method. Samples may be compared against one or more reference methods. The purpose of evaluating accuracy is to determine systematic error, or bias, of the method.

a. Testing Materials. Testing materials should match or be as close to the sample matrix as possible. The preferred test materials are patient samples with known values tested against the reference method(s). Other sources of test materials include proficiency samples, quality control materials, calibration materials, and samples of the appropriate matrix spiked with a known level of the analyte such as a reference standard. Samples should be at concentrations over the range of the assay. At least three concentrations at low, middle, and high values should be tested as well as concentrations near any relevant medical decision limit. In addition, if samples with concentrations above the reportable range will be diluted to obtain a valid result, samples above the reportable range should also be included in the study and dilution protocols validated.

b. Number of Samples. A minimum of 5 samples per concentration is required, but 20-40 samples are recommended. The actual number of specimens is less important than the quality of specimens, but consideration should be given to the purpose of the assay and potential outcome for the patient if an incorrect result were to be reported. A more robust evaluation of accuracy may be warranted for assays such as HIV where an incorrect result would have a greater impact to the patient.

c. Experimental Design. Per CLSI standards, it is recommended that the chosen number of samples be run over a minimum of five days to simulate a range of conditions over which samples would normally be run. The comparison of methods experiment can also be combined in this fashion with the reproducibility (precision) study.

d. **Performance Criteria.** Manufacturer statements of accuracy should be used as the minimum performance requirements.
e. **Data Interpretation.** Data should be plotted in a comparison plot to visually identify any outliers. Use linear regression statistics to determine the correlation coefficient and regression equation (Y = a + bx). Constant systematic error is indicated by the y intercept, a. Proportional systematic error is indicated by the slope, b. The correlation coefficient, r, is used to determine whether or not the range of data gives a good estimate of the slope and intercept. Generally, an r >0.99 shows a good range of data and linear regression statistics are appropriate. For an r < 0.975, a bias plot with t-test statistics may be more appropriate. The correlation coefficient should not be used to judge acceptability of the error of the method, rather what statistics are appropriate for determining error. If the manufacturer does not provide an acceptable range for the slope and intercept, calculate the 95% confidence interval (CI) of the manufacturer's data to determine if the slope and intercept obtained is within that range.

2. **Precision.**

Precision is measured with a replication experiment. Within run, between run, and day to day precision must all be evaluated. Operator variation must be included unless differences in operators will not affect the result outcome (i.e. a fully automated, including performance of sampling, test system).

a. **Testing Materials.** Test materials can be any sample of known concentration including QC, calibrator, proficiency test material, and patient specimens. Enough test material should be prepared to complete the study and consideration should be given to replicating the patient matrix as closely as possible. Material concentrations should be near medically relevant decision levels if known.
b. **Number of Samples.** A minimum of two or three samples at each medically relevant concentration should be used. When medically relevant decision levels are not known, at least three concentrations should be run; one at the low end of the analytical range, one in the middle, and one at the high end of the range.
c. **Experimental Design.** Most sources recommend running each sample in duplicate or triplicate over five days. Different operators should participate in the replication experiment for methods that are operator dependent.
d. **Performance Criteria.** Manufacturer statements of precision should be used as the minimum performance requirements.
e. **Data Interpretation.** Calculate the mean, standard deviation, and coefficient of variation (CV) for each concentration. Use the F test to determine if the variance of the method is statistically different from the manufacturer's claim. For questions on how to perform the F test, contact the QAO.

3. **Reportable Range.**

Also known as the linear or analytical range, reportable range must be determined in quantitative assays to determine the lowest and highest values that can be reliably reported. Often this is based on the lowest and highest calibrator. The range may be

extended, however, either at higher concentrations of the analyte through dilutions of the specimen, or on the lower end by determining the lower limit of quantitation (LoQ) that may be below the lowest calibrator. The LoQ determination experiment is covered by the analytical sensitivity section below.

a. **Testing Materials.** CLSI guidelines recommend a minimum of at least 4 but preferably 5 concentrations. Materials should be as close to the sample matrix as possible, especially when verifying different specimen types for the test method. Calibrators, purchased linearity materials, quality controls, or patient specimens are all possible materials, though the concentration of analyte must be known no matter what material is used. Pools of patient specimens may be used with standards spiked into them where available as well.
b. **Number of Samples.** One set of 4-5 concentrations of the analyte is recommended.
c. **Experimental Design.** A dilution series may be used. Alternatively, if the method has enough calibrators, each calibrator may be run as would a patient sample. Each concentration should be run in triplicate.
d. **Performance Criteria.** Results should meet accuracy and precision statements by the manufacturer at each concentration.
e. **Data Interpretation.** See accuracy and precision sections.

4. **Reference Range (normal values).**

Reference range refers to the test values typically observed in a healthy population and this is why it is often called the "normal" range. This may be assessed subjectively by reviewing the reference range from the manufacturer and judging whether or not it is applicable based on PHLD's patient population. Published literature may be reviewed and taken into account as well. This approach is known as transference and does not require any experiments to be carried out by PHLD. Alternatively, obtaining specimens from 20 individuals from the appropriate reference patient population and testing them may also be used to verify the manufacturer's stated reference range. 90% of results should fall within the manufacturer's normal range. If <90% of results fall within the normal range, a more extensive study with a larger number of samples will have to be performed, typically with a minimum of 40 individuals although 120 is recommended but may be difficult to obtain.

B. Non-FDA-Cleared/Approved

1. **Accuracy.**

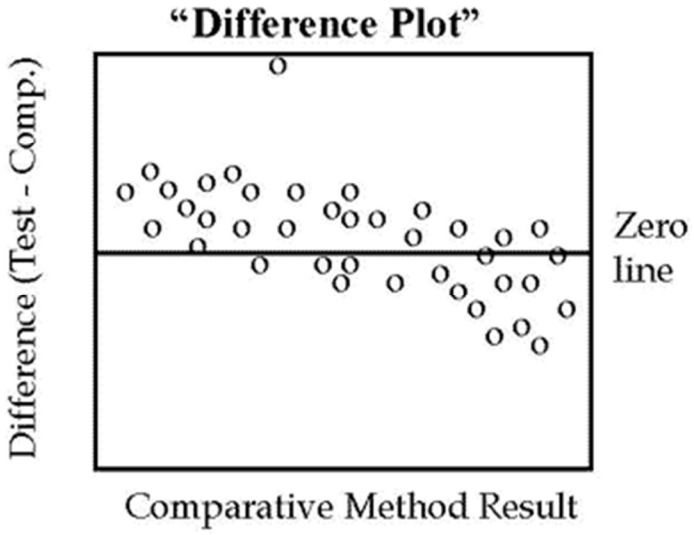
a. **Testing Materials.** Follow the guidelines above for quantitative FDA-Cleared/Approved assays.
b. **Number of Samples.** At minimum, follow the guidelines above for quantitative FDA-Cleared/Approved assays. In addition, it may be advisable to include additional samples at each concentration. Many reference documents suggest a

minimum of 40 or more samples for non-FDA-Cleared assay validations, but the quality of the specimens is more important than the quantity.

c. **Experimental Design.** Follow the guidelines above for quantitative FDA-Cleared/Approved assays.

d. **Performance Criteria.** Performance goals can come from a variety of sources. Some sources of these are medical decision points, published professional recommendations, and performance goals set forth by regulatory agencies such as under subpart I of the CLIA regulations. Assays obtained from CDC may have minimum quality requirements as well. FDA guidance for chemistry assays indicates that the mean value not deviate from the expected value by more than 15% except at the LoQ where it should not deviate more than 20%.

e. **Data Interpretation.** Follow the guidelines above for quantitative FDA-Cleared/Approved assays. In addition, bias can be calculated by taking the difference of the averages between the test method and the reference method. Plot data into a difference plot. If bias appears to be constant across the range of concentrations (i.e. an even scatter of points with no obvious bunching), an overall bias can be calculated. If bias does not appear to be constant and a difference in scatter is noted in one range of concentrations over another (i.e. higher scatter at lower concentration), then it may be appropriate to divide up the data into subsections to calculate bias at differing sets of concentrations. Data that has a high bias beyond what the performance criteria deems acceptable may indicate that the reportable range needs to be tightened. See CLSI document EP09 for more information.



2. **Precision.**

Precision is measured with a replication experiment. Within run, between run, and day to day precision must all be evaluated. Operator variation must be included unless differences in operators will not affect the result outcome.

a. **Testing Materials.** Follow the guidelines above for quantitative FDA-Cleared/Approved assays.
b. **Number of Samples.** At minimum, follow the guidelines above for FDA-Cleared/Approved assays.
c. **Experimental Design.** At minimum, follow the guidelines above for FDA-Cleared/Approved assays.
f. **Performance Criteria.** Performance goals can come from a variety of sources. Some sources of these are medical decision points, published professional recommendations, and performance goals set forth by regulatory agencies such as under subpart I of the CLIA regulations. Assays obtained from CDC may have minimum quality requirements as well. FDA guidance for chemistry assays indicates that the precision of each concentration level should not exceed 15% CV except at the LoQ where it should not exceed 20% CV.
d. **Data Interpretation.** Calculate the mean, standard deviation, and coefficient of variation (CV) for each concentration. Use the F test to determine if the variance of the method is statistically different from the reference method. For questions on how to perform the F test, contact the QAO.

3. Reportable Range.

The reportable range study should be performed as per the guidelines above for FDA-Cleared/Approved assays.

4. **Analytical Sensitivity.**

Analytical sensitivity involves a limit of detection experiment where the lowest concentration of an analyte that can be measured is determined. Many different terms exist for this value but CLSI guideline EP-17 recommends the following three terms: Limit of Blank (LoB), Limit of Detection (LoD), and Limit of Quantification/lower limit of quantification (LoQ). The LoB is less than the LoD which is less than or equal to the LoQ. See the definitions section above for clarification and distinction between each of these terms.

a. **Testing Materials.** Ideally, samples should be from actual patient specimens, although use of the zero calibrator is often used as the blank sample and the lowest standard calibrator as the low level "spiked" sample. Samples that do not contain the analyte or samples containing low level of the analyte should be used depending on whether the LoB or LoD and/or the LoQ are being determined. Blank (i.e. analyte negative) samples would be required for LoB experiments whereas low level analyte specimens would be required for LoD and LoQ experiments. Patient samples may be created that are spiked with low levels of the analyte. Non-patient samples such as water, saline, or other matrix should be used with caution and first tested to ensure there are no matrix effects as compared to

patient samples. Multiple independent samples or pools of samples should be used.

b. **Number of Samples.** The minimum recommended number of is 4-5 independent samples or pools of samples.
c. **Experimental Design.** The samples should be run on multiple lots of reagents and calibration curves (where relevant) to get the best determination of error. Ideally, the detection limit experiment will contain both determination of LoB and LoD, although depending on the method, both may not be possible. To determine LoB, perform a replication experiment using blank samples (see above information about testing materials and number). To determine LoD and LoQ, spike the blank sample with a low level of the analyte or use samples containing known low level concentrations of the analyte. Several different low level concentrations may be necessary. According to CLSI guideline EP-17, the minimum experimental design should include samples be run in replicates of 4 over 3 days. In this way, the limit of detection experiment can often be combined with the precision study.
d. **Performance Criteria.** LoB should not be higher than any medically relevant concentrations. Depending on the assay, the LoD may be equal to or less than the LoQ. LoQ (and LoD if LoD = LoQ) must meet accuracy and precision requirements for the assay. The following must be true: LoB < LoD ≤ LoQ.
e. **Data Interpretation.** When multiple lots of reagents are used, calculate the mean and SD of the blank and low level sample replicates separately. For the LoB, if the distribution of data is normal and no data censoring from the system software exists (i.e. reporting negative values obtained during the analysis as zero), then the following equations can be used to determine the LoB and LoD:

$$LoB = mean + 1.645 * SD_{(of\ the\ blank\ samples)}$$

$$LoD = LoB + 1.645 * SD_{(of\ the\ low\ concentration\ samples)}$$

For LoQ, calculate the mean and SD and compare to the accuracy and precision requirements. If data censoring occurs (i.e. values do not make up a Gaussian, or normal, distribution), consult CLSI guideline EP17 for more information on calculations.

5. **Analytical Specificity.**

Analytical specificity, or cross-reactivity, is evaluated by testing specimens with potentially interfering organisms, substances, or analytes in interference experiments to assess constant systematic error. Recovery experiments may also be used to assess proportional systematic error due to assay interferences that may be present due to various materials in the specimen. For analysis of blood specimens, this often includes bilirubin, lipids, and hemolysis in plasma and serum samples. It may also include testing of structurally similar analytes.

a. **Testing Materials.** Follow the guidelines above for qualitative non-FDA-Cleared/Approved assays.
b. **Number of Samples.** Follow the guidelines above for qualitative non-FDA-Cleared/Approved assays.
c. **Experimental Design.** Follow the guidelines above for qualitative non-FDA-Cleared/Approved assays.
d. **Performance Criteria.** Follow the guidelines above for qualitative non-FDA-Cleared/Approved assays.
e. **Data Interpretation.** Follow the guidelines above for qualitative non-FDA-Cleared/Approved assays.

6. **Reference Range (normal values).**

   See section above for FDA-Cleared/approved quantitative methods.

VII. Disclaimers and Reporting of Results

A. General Requirements

1. For genetic tests including DNA or nucleic acid based genetic tests, the test method(s) employed and any mutations should be included on the test report.
2. It may be desirable to add additional comments where testing is used for screening purposes and will be sent out for confirmation or to assist submitters in interpreting of results.

B. FDA-Cleared/Approved Assays

1. No additional disclaimers are necessary for regulatory purposes. Follow manufacturer requirements for reporting.

C. Non-FDA-Cleared/Approved Assays

1. Tests that were developed using an ASR (analyte specific reagent) may include the following statement on the test report: "*This test was developed and its performance characteristics determined by the Minnesota Department of Health Public Health Laboratory. It has not been cleared or approved by the U.S. Food and Drug Administration.*" This statement is required by the FDA under 21 CFR, Part 809.30(e).
2. All other non-FDA-Cleared/approved assays must include the following disclaimer: "*The performance characteristics of this test were determined by the Minnesota Department of Health Public Health Laboratory. It has not been cleared or approved by the U.S. Food and Drug Administration.*"

VIII. Related Documents

A. DOC-563 Clinical Validation Report Template

B. DOC-562 Clinical Validation Plan Template
C. DOC-498 How to Complete Validation/Verification Documents in MasterControl
D. DOC-294 MasterControl Quick Reference Guide – Uploading Documents

IX. References

A. Westgard, James O. *Basic Method Validation, 3rd Edition*. Westgard QC, Inc. 2008.

B. CLSI. *Evaluation of Detection Capability for Clinical Laboratory Measurement Procedures; Approved Guideline – Second Edition*. CLSI document EP17-A2. Wayne, PA: Clinical Laboratory Standards Institute; 2012.

C. CLSI. *Measurement Procedure Comparison and Bias Estimation Using Patient Samples; Approved Guideline – Third Edition*. CLSI document EP09-A3. Wayne, PA: Clinical Laboratory Standards Institute; 2013.

D. CLSI. *User Protocol for Evaluation of Qualitative Test Performance; Approved Guideline – Second Edition*. CLSI document EP12-A2. Wayne, PA: Clinical Laboratory Standards Institute; 2008.

E. CLSI. *Assessment of the Diagnostic Accuracy of Laboratory Tests Using Receiver Operating Characteristic Curves; Approved Guideline – Second Edition*. CLSI document EP24-A2. Wayne, PA: Clinical Laboratory Standards Institute; 2011.

F. U.S. Department of Health and Human Services. Medicare, Medicaid and CLIA programs: Regulations implementing the Clinical Laboratory Improvement Amendments of 1988 (CLIA). Final rule. Federal Register. 2003; 42 CFR Part 493.

G. Cohen's kappa. Wikipedia. May 9, 2014. Retrieved from http://en.wikipedia.org/wiki/Cohen%27s_kappaGraphPad Software. Visited May 15, 2014. http://graphpad.com/quickcalcs/kappa1.cfm

H. Burd Eileen M. *Validation of Laboratory-Developed Molecular Assays for Infectious Diseases*. Clin Microbiol Rev. 2010;23(3):550–76. doi: 10.1128/CMR.00074-09.

I. Wolk, Donna M. and Olson, Daniel. *Verification of Qualitative Real-time PCR Methods.* Cepheid. January, 2012. On-Demand Winter 2012.

J. U.S. Department of Health and Human Services. Food and Drug Administration. *Guidance for Industry: Bioanalytical Method Validation.* May 2001. BP.

Signature Manifest

Document Number: DOC-452 **Revision:** 3
Title: Guidelines for Verification and Validation of Laboratory Methods
Effective Date: 03 Jan 2017

All dates and times are in Central Time.

Validation SOP upload

Director Approval

Name/Signature	Title	Date	Meaning/Reason
Joanne Bartkus (BARTKJ1)	PUBLIC HEALTH LAB DIV DIR	03 Jan 2017, 11:43:58 AM	Approved

Review: DOC-452 3 Guidelines for Verification and Validation of Laboratory Methods

Review

Name/Signature	Title	Date	Meaning/Reason
Joanne Bartkus (BARTKJ1)	PUBLIC HEALTH LAB DIV DIR	25 Oct 2018, 01:00:29 PM	Reviewed - Under Revision

Review: DOC-452 3 Guidelines for Verification and Validation of Laboratory Methods

Review

Name/Signature	Title	Date	Meaning/Reason
Jacqueline Benjamin (BENJAJ1)	State Program Admin Prinicipal	19 Aug 2020, 09:42:41 AM	Reviewed
Sara Vetter (VETTES1)	CLIA Director and ADD	29 Dec 2020, 11:51:37 AM	Reviewed