Molecular Diagnostic Assay Validation

Association for Molecular Pathology Clinical Practice Committee, October 2009

INTRODUCTION

The primary goal of method validation in the molecular diagnostics laboratory is to ensure the accuracy of the reported results. To reach that goal, each step of the testing process must be carefully evaluated and monitored to document the appropriateness of the method for the test being performed and to compare the results to another standard testing method. Such validation is relatively standardized in fields such as clinical chemistry that are highly automated. It is challenging, however, to apply those standardized practices to molecular diagnostics laboratories, which rely heavily on laboratory-developed tests and which are very labor-intensive. For this reason, laboratory directors may find it helpful to refer to a brief overview of the necessary components of assay validation as they can be applied in molecular diagnostics laboratories. Although excellent and comprehensive documents have recently been published on this topic (e.g., Jennings L, Van Deerlin VM, Gulley ML. Recommended principles and practices for validating clinical molecular pathology tests. *Arch Pathol Lab Med*. 2009;133:743-755; Marlowe, E.M. and Wolk, D.M. 2009. Molecular Method Verification. Diagnostic Molecular Microbiology, ASM Press, Washington, DC., in press), the goal of this overview is to provide a brief practical reference to be used during the validation process. Additionally, a sample summary checklist is provided as a template to allow the laboratory director to review and sign off on the validation process when new assays are developed or adopted.

ASSAY VALIDATION / EVALUATION

WHO definition: The process of proving that a procedure works as expected and consistently achieves the expected result.

CLIA (42 CFR 493.1253) and CAP (GEN 42020-42163) require that laboratories validate the performance of tests before reporting patient results

For UNMODIFIED FDA-approved or FDA-cleared tests, laboratories must verify that test(s) perform(s) as expected by obtaining data on:

- Accuracy
- Precision
- Reportable range (clinical reportable range and linearity)
- Reference intervals for laboratory patient population

For MODIFIED FDA-Approved tests or for non-FDA cleared tests (e.g., Laboratory Developed Test-LDT) laboratories must establish the test's:

- Accuracy
- Precision
- Analytical sensitivity (lower limit of target detection, as appropriate for each laboratory's protocol)
- Reportable range of test results
- Reference intervals (normal values)
- Efficiency or call rate for genotyping assays (for those assays in which a large number of samples are available)

I. Analytical Validation

A. Accuracy

1. Per the Standards for the Reporting of Diagnostic Accuracy (STARD): The amount of agreement between the information from the test under evaluation (the index test) and the reference standard (the best available method for establishing the presence or absence of the condition of interest)

Elements of accuracy that should be addressed during validation (of note, some of these statistical measurements, such as PPV and NPV, require a large number of samples in order to yield a meaningful result):

- Sensitivity
- Specificity
- Positive predictive value (PPV)
- Negative predictive value (NPV)
- False-positive rate
- False-negative rate

B. Analytical Sensitivity

1. The ability of a test to detect a mutation or disease when that mutation/disease is present

Sensitivity = True positive ÷ (True positive + False negative)

2. Also used to refer to the lower limit of detection for the analyte of interest (i.e., the lowest concentration of analyte that the assay can detect), For example 95% LLOD would be the lowest concentration of analyte (e.g. bacteria/ml or copies/ml) that would yield detection 95% of the time; measurements should be converted to analyte per ml of sample so that comparisons can be made. LOD listed in copies per reaction is of limited use as that format is not easily compared between studies.

C. Analytical Specificity

1. The ability of a test to give a normal (negative) result in specimens without the mutation or disease being tested

Specificity = True negative ÷ (True negative + False positive)

2. Also used to refer to the ability of a test to detect the analyte without cross-reacting with other substances or genetically or biologically similar microbes

D. Precision

1. Getting the same results with repetition of the assay (see Section III.B.2 below)

E. Reference and Testing Material

1. Reference materials (RM) should be used for a calibration of the measuring system, for assessment of a measurement procedure, for assigning values to materials, or for quality control.

2. RM should be selected based on the needs of the assay, the methodology, and the availability of materials. Potential benefits and drawbacks should be evaluated with respect to each individually validated assay.

- Genomic DNA – mimics patient sample in term of complexity but can only control for one or a few alleles/genotypes or targets at a time
- Recombinant plasmids or synthesized oligonucleotides – do not resemble complexity of genomic DNA but can control for multiple alleles in a single reaction

Examples:

Human DNA, bacterial and viral genomic DNA, mitochondrial DNA, synthetic DNA, plasmids containing human DNA, amplicons, *in vitro* transcripts, synthetic oligonucleotides, recombinant DNA, phage and phage protein packaged nucleic acid, genetically modified cell lines

Resources:

a. Cell Repositories include:
   American Type Culture Collection (ATCC), Coriell Cell Repositories, European Collection of Cell Cultures, German Collection of Microorganisms and Cell Cultures, Japanese Collection of Research Bioresources, Riken Bioresource Center
b. Sources of Higher Level Controls (Certified Reference Materials (CRM), Standard Reference Materials (SRM)) include:
   Institute for Reference Materials and Measurements, National Institute of Biological Standards and Controls, National Institutes for Standards and Technology, National Genetics Reference Laboratory Manchester, World Health Organization-International Reference Materials
c. Commercial Sources of Reference Materials include:
   Acrometrix, ParagonDx, InVivoScribe, LGC Promochem, Maine Molecular Quality Controls, Molecular Controls, Asuragen, Ipsogen, Roche, Advanced Biotechnologies, Diagnostic Hybrids, Zeptometrix, Seracare Life Sciences, Third Wave Technologies
d. Other Sources include:
   National Genetics Reference Laboratory Wessex
e. Inter-laboratory Exchange of De-identified Residual Patient Samples

II. Clinical Validation

A. Clinical Validity: The ability of a test to detect or predict the associated disorder (phenotype) (http://www.cdc.gov/genomics/gTesting/ACCE.htm)

B. Clinical Utility: How useful the test is in the diagnosis or treatment of patients

C. Clinical Sensitivity:
   1. The proportion of patients with the mutation/disease who have a positive test result, or the likelihood that a positive result correctly determines that the patient has the condition being tested

   2. Positive Predictive value = True positive results ÷ (True pos + False pos)

D. Clinical Specificity
   1. The proportion of patients who lack the mutation/disease who have a negative test result, or the likelihood that a negative result correctly determines that the patient does not have the condition being tested

   2. Negative Predictive Value = True negative results ÷ (True neg + False neg)

(N.B.: Predictive values take into account the prevalence of the disease in the population being tested [e.g., the higher the prevalence, the higher the likelihood that a positive result is a true positive])

	CLINICAL CONDITION		
TEST OUTCOME	POSITIVE	NEGATIVE	
POSITIVE	True Positive	False Positive	**Positive Predictive Value TP / (TP + FP)**
NEGATIVE	False Negative	True Negative	**Negative Predictive Value TN / (TN + FN)**
	Sensitivity TP / (TP + FN)	**Specificity TN / (FP + TN)**	

III. Validation of Laboratory Developed Tests

A. Assay Optimization

1. Optimize all steps of the assay, individually and as a whole assay, including:
   a. Extraction
   b. Amplification, if this is one of the steps in the assay
   c. Detection
   d. Result interpretation and reporting

2. Evaluate following variables
   a. The "raw material" of the assay (e.g., the extracted DNA or RNA)
   b. Consider and evaluate: specimen source, storage, transport, stability, integrity
   c. Evaluate sufficient input amounts and volumes specimens to determine the correct dynamic range for each specimen type
   d. Assess quality (length, molecular weight, and purity - optical density at 260nm/280nm, 260/230, gel electrophoresis, bionanalyzer), and quantity (concentration) of extracted nucleic acids
   e. Matrix – related to specimen type
   f. Reagents, including stores and aliquots/working solutions
   g. Instrument calibration
   h. Evaluate well-to-well cross-contamination (for automated nucleic acid extraction)

B. Analytic Validation

1. Accuracy

   a. Analyze a known sample (either of known concentration or result, or both) and compare the result with that obtained by a reference ("gold-standard") method
   - include specimens from the indicated patient population (specimens used for validation should reflect anticipated population with regards to sex, age, race etc.)
   - choose appropriate data analysis techniques
   - choose appropriate reference methods
   - choose appropriate comparative methods

   b. If disease is uncommon, it may be necessary to obtain commercially available reference materials, if they are available. Otherwise, controls may need to be obtained from another laboratory. For sequencing assays, however, the number of possible mutations essentially precludes the use of reference materials.

c. Compare results between new method and "reference" method or method already established in the lab

d. The appropriate number of specimens depends on many factors including, but not limited to:

- complexity of the assay
- relevance of targets/alleles in the intended use population
- data analysis schemes
- established accuracy of reference methods

EXAMPLES OF ANALYTIC VALIDATION

(Please note that these are only suggestions, not directives. The extent and type of validation that is needed for a particular test is left to the discretion of the laboratory director.)

For FDA-approved methods

i. Test 20-40 samples that span the entire reportable range for quantitative assays and different possible genotypes for genotyping assays

ii. Run samples by both new and comparative method.

iii. Evaluate bias (difference between new, comparative method) in one of the following ways:

(a) Calculate average bias and check CAP limits for passing proficiency testing.

(b) Calculate statistically significant difference in the mean by applying the t-test.

(c) Can compare using linear regression analysis:

i. Plot data (reference samples/method: x-axis; measured values/new method: y-axis)

ii. Calculate linear regression statistics (ideally: slope=1, intercept=0 and r=0.99)

iii. Establish criteria for accepting results (e.g. >95% confidence interval, +/- 2 SD)

(d) Comparison of results using the new method on certified reference materials. If no commercial materials are available, laboratories need to create their own reference materials:

FOR QUANTITATIVE METHODS:

(a) Select samples with high, middle and low quantitative results. If middle values are not available, high and low can be mixed to create mid-point samples.

(b) If the high result is outside of the upper limit of the analytical reference range, the control should be diluted so that it does fall in this range.

(c) Aliquots of the samples can be made and stored at -20^0C for DNA or -70^0C for RNA for future use (calibration verification every 6 months).

(d) For genotyping assays, samples can be characterized using a reference method (e.g. sequencing or restriction enzymes).

(e) Recovery study (spike samples with known amount of standard and measure)

2. Precision (getting the same results with repetition of the assay)
Closeness of agreement between independent (repeated) results of measurements obtained under stipulated conditions. Describes the degree of dispersion of the values of repeated measurements form each other.

Elements of precision:

- Repeatability – closeness of agreement between repeated tests of the same sample under the same operating conditions. Performed under the same conditions with multiple replicates, various sample concentrations (QUANTITATIVE), and various patters of variants (QUALITATIVE). This is referred to as intra-assay variation.
- Reproducibility – closeness of agreement between the results of measurements when operating conditions are varied. All potential sources of variability should be considered. This is referred to as inter-assay variation. Typically, the same material is analyzed by the same protocol on different days, by different operators and with different equipment.

Sources of variability:

- Instrument
- Laboratory
- Operator (most common source of variability)
- Sample concentration
- Sample source
- Reagent lot (second most common source of variability)
- Run
- Time of the day
- Laboratory environment

The laboratory should consider each source of variability that can occur during testing.

a. Within run: Run one sample several times in one run

b. Between runs: Run one or more samples in several different runs over several days

c. Determine the mean, standard deviation and coefficient of variation (standard deviation ÷ mean)

d. Precision studies should be carried out for the entire assay, as applicable, from extraction and amplification to detection

EXAMPLES OF PRECISION EVALUATION:
(Please note that these are only suggestions, not directives. The extent and type of validation that is needed for a particular test is left to the discretion of the laboratory director.)

1. FOR QUANTITATIVE ASSAYS:

- Take at least 3 sample concentrations (plasma pools or control samples) that cover the clinically important decision levels (e.g. for HCV quantitative assays a “low” control of ~100 IU/mL, a “medium” of~5000 IU/mL, and a “high” of ~100,000 IU/mL). Depending on the assay the low concentration replicates can be two to four times the established Level of Detection (LoD), high concentrations should be close to the 99th percentile of tested concentrations/titers.
  Example: A single operator testing five replicates at three concentrations (low, medium, high) run for 3 days using a single lot of reagent measured on the same instrument (total of 15 replicates per concentration will be generated) (**CLSI, EP15-A2**). It may be

necessary to conduct multiple repeatability validation studies to cover all testing variables (see above).

- Use a spreadsheet to calculate mean value, standard deviation and %CV for within-run and between-run precision, and percent of agreement between tests performed under two different conditions (confidence interval should be calculated for the observed percent agreement). For FDA approved assays, compare to the manufacturer's claim; if higher %CV were obtained, then evaluate the cause.

- Compare precision to clinically acceptable variation (e.g. for HCV quantitative assays, changes beyond 0.5 log are considered true fluctuations and not intrinsic test variation).

- In general, when using plasma samples, make sure they were not frozen and thawed more than 3 times, as this may be a confounding factor.

2. FOR GENOTYPING ASSAYS:

- For assays that require a Tm: take control samples that cover the different genotypes (e.g. to identify prothrombin 20210G>A mutation, run at least one homozygous mutant, one heterozygous and one homozygous wild type sample, respectively) and run in triplicate per run for 5 days. Calculate the average Tm and SD of the Tm. The Tm values have to be within the range of the manufacturer's claim.

- For assays that require peak size to determine number of repeats: e.g., for Fragile X by PCR, take 3 control samples of known size within the normal range (≤ 49 repeats), 3 "low" pre-mutations (50-79 repeats), and 3 "high" pre-mutations (80-200 repeats). Run in duplicate per run for 3 days. Calculate the average size and SD. Acceptable range +/- 3 repeats.

- Include representation of rare alleles to ensure that their presence can be detected with precision.

- Test the ability of the system/assay to discriminate similar and adjacent mutations, where possible.

C. Assay Linearity (Quantitative assays)

1. Assay a series of standards or serial dilutions of a known standard or sample

2. Graph the results
   a. x-axis: Known values; y-axis: Measured values
   b. Draw best fit line through the data points or calculate regression
   c. Determine slope (ideal = 1.0) and intercept (ideal = 0)

D. Analytical Sensitivity

Defined as a change in response of a measuring instrument (analyte change) divided by the corresponding change in the stimulus (analyte).

1. The ability of a test to detect a mutation or disease when that mutation/disease is present

   Sensitivity = True positive ÷ (True positive + False negative)

2. Should be obtained by measuring samples number of times under different conditions.
3. Should be determined by titration. Excessive amount of analyte (eg. nucleic acid) can cause incorrect genotype calls, insufficient amount of analyte can cause allele drop-out.
4. False-negative rate can be determined by testing individuals who are known to have the condition being tested.
5. Compare results with those of a "gold-standard" reference method or of another validated method in use in the lab.
6. **Limit of Detection (LoD)** - the lowest amount of analyte in a sample that can be detected, but not quantified as an exact value (WHO). Can be expressed by weight, percentage, or calculated copy number, and should be related to independent dilution of specimens with, for example, known microscopic particles in the appropriate specimen matrix, as appropriate. It is a concentration of analyte that will be detected 95% of the time with a 95% confidence interval.
7. For multiplex assays LoD must be established for each target singly, as well as verified with other targets that could potentially be present in the same sample. If one target of multiplex assay is present at a high level, the detection of another target can present at low levels could be potentially impaired (e.g. fetal DNA in maternal blood, maternal cell contamination in fetal specimens, mosaics or minimal residual disease).
8. **Input range** (for genotyping assays) – acceptable range within which the multiplex assay yields accurate results for all variants tested.

E. Analytical Specificity

The ability of a test to give a normal (negative) result in specimens without the mutation or disease being tested. The proportion of biological samples that have a negative test result or no identifies mutation (being tested for) and that are correctly classified as negative.

Specificity = True negative ÷ (True negative + False positive)

1. The ability of the procedure to distinguish the target sequence(s), allele, or mutations(s), from other sequences/alleles/mutations in the genome. Also used to refer to the ability of a test to detect the analyte without cross-reacting with other substances.

Aspects of analytical specificity:
- Interference
- Cross reactivity

Sources of introduction of contamination, admixture, and substances that might interfere with downstream application to the sample
- Poor sampling
- Lack of sample stabilizer
- Maternal cell contamination in prenatal specimens
- Cross contamination during sample processing
- Inclusion of normal, nondiseased tissue with the diseased tissue of interest
- Tissue from source addition to the desired sample (e.g., maternal cells obtained during fetal specimen collection)
- Bacteria
- Endogenous substances (e.g. hemoglobin, cholesterol, triglycerides, medications, anticoagulants, residual sample processing, stabilization reagents)

2. Testing of potential interference should be conducted at the highest concentration that would occur in a clinical setting. The maximum amount of a potentially interfering substance the assay can tolerate without causing actual interference or adversely affecting the rest results should be determined. Interference studies can be conducted by spiking the specimens with

interfering agents (spiked vs. unspiked). Studies should be performed for each specimen matrix used in the assay.

3. RNA copurified with varying levels of DNA and vice versa.

4. Residual reagents such as organic solvents, wash buffers, residual high salt concentrations, heparin, and others.

5. Cross reactivity – determine if analytes other than the ones it designed to measure, a panel of closely related organisms/alleles should be assessed.

False-positive rate can be determined by testing individuals known to have conditions different than that being tested, but that is similar enough to be in the differential diagnosis.

False-positive rate can also be determined by testing individuals known to be negative for the condition being tested

Compare results with those of a "gold-standard" reference method or of another validated method in use in the lab

F. Reference Range

1. Normal range: Perform assay on normal controls (i.e., patients without the condition being tested). The number of samples to be tested is at the discretion of the laboratory director and will be dependent in part on the condition being tested and on the availability of appropriate control specimens.

a. If clinically indicated for the condition being tested, consider including in the normal control range individuals of both sexes, a representative ethnic background to the population being tested, pregnancy status, etc.

b. May also consider including individuals with other medical conditions, if such conditions represent the clinical status of the patients who will be tested

2. Abnormal range: Include in the abnormal control group individuals who are heterozygous and individuals who are homozygous, if possible (at least 20 samples if possible, preferably more)

IV. Ongoing Quality Control Measures

A. Controls should be included in each assay (both positive and negative controls and a blank), when appropriate. Exceptions would include an assay in which the majority of patients are negative for the mutation being tested; in this case, the patients can serve as a negative control. Also, normal controls need not be run for sequencing assays, because the reference sequence is known).

1. The positive control should have a clinically relevant amount of analyte.

2. The negative control should contain other nucleic acid targets that would be expected to be present in the patient sample.

3. Failure of any of the controls invalidates the run; repeating the entire run should be considered in this case.

4. The blank should contain the complete reaction mixture EXCEPT nucleic acids.

5. Amplification of an internal nucleic acid target (either endogenous or spiked) is recommended to ensure that inhibitors are not interfering with the assay.

B. Standards (e.g., size markers) should be used in each run
- Of note, it may also be prudent to perform an external positive control for assays such as viral loads, to permit analysis of trends and to assess variation of non-kit based controls over time

C. Proper calibration of the instruments used in the assay should be maintained.

D. Routine preventive maintenance of all instruments (e.g., thermocyclers, pipettes, etc)

E. Testing of new lots of reagents

New lots of all reagents should be tested side-by-side with the old reagent, prior to implementation

F. Proficiency Testing

1. Labs should participate in national testing programs (e.g., CAP) for all tests they perform

2. If no proficiency test is offered, labs should participate in an alternative assessment at least twice per year (per CLSI). An interlaboratory exchange is preferable. CAP has a sample exchange registry for assays not offered in its commercially available proficiency testing.

G. Assessment of continued technologist competency

H. Documentation

The laboratory must keep results of validation studies as long as the protocol is in use, and for at least two years after the protocol is retired

I. Reporting/Laboratory Information System

The test report should reflect the reason for the patient's referral. Additionally, if pre-written interpretations ("canned comments") are used, the laboratory director should ensure that the correct test interpretation is issued.

V. Ongoing Quality Assurance Measures

A. Regular review of normal and abnormal results
- e.g., review of percentage of positive reactions, by virus or other analyte

B. Review of turnaround times

C. Review of rejected specimens

D. Annual Quality Improvement project(s)

REFERENCES

1. Prence EM. A practical guide for the validation of genetic tests. *Genetic Testing*. 1999;3:201-205.

2. Ferreira-Gonzalez A, Wilkinson DS, Garrett CT. Establishing a molecular diagnostics laboratory. From: *McPherson & Pincus: Henry's Clinical Diagnosis and Management by Laboratory Methods, 21st ed*. WB Saunders. 2006.

3. Ferriera-Gonzalez A, Garrett CT. Laboratory-developed tests in molecular diagnostics. From: *Molecular Diagnostics: For the Clinical Laboratorian, 2nd ed.* WB Coleman and GJ Tsongalis. Humana Press.

4. Rennerg H and Leonard DGB. Quality control and quality assessment programs. From: *Molecular Pathology in Clinical Practice,* pp. 568-574. DGB Leonard, ed. Springer. 2007.

5. Association for Molecular Pathology. Association for Molecular Pathology Statement: Recommendations for in-house development and operation of molecular diagnostic tests. *Am J Clin Pathol*. 1999;111:449-463.

6. Jennings L, Van Deerlin VM, Gulley ML. Recommended principles and practices for validating clinical molecular pathology tests. *Arch Pathol Lab Med*. 2009;133:743-755.

7. Marlowe, E.M. and Wolk, D.M. 2009. Molecular Method Verification. Diagnostic Molecular Microbiology, ASM Press, Washington, DC. In press.

8. MM17-A, Verification and Validation of Qualitative Nucleic Acid Tests: A Review of the Approved CLSI Guideline

9. McPherson & Pincus: Henry's Clinical Diagnosis and Management by Laboratory Methods, 21st ed.

VALIDATION SUMMARY

ASSAY _____

NEW / MODIFIED

FDA-APPROVED or CLEARED / LDT

DOCUMENTATION OF VALIDATION STUDIES

1. Analytical validation

   a) Accuracy (getting the right results)
      -- Ex: Parallel testing with an alternate method

   b) Analytical sensitivity (the lowest amount of target nucleic acid that can be detected by the assay)
      -- Ex: Testing various dilutions of a known target in a relevant matrix

   c) Analytical specificity (the ability of a test to give a negative result when the target nucleic acid is not present)

   d) Precision (the ability of a test to produce the same result on repeated testing of the same target)
      -- Ex: Replicate testing within and/or between runs

2. Clinical validation
   -- Documentation of the clinical utility and efficacy of the test in light of clinical characteristics of the disease/marker being tested (e.g., prevalence of the disease in a relevant population, prevalence of SNPs or other copy number variants)

VALIDATION CHECKLIST

☐ Optimization of assay on an adequate number of each of the expected specimen types

☐ Comparison of results with a reference method

☐ Determination of reference and/or reportable ranges

☐ Procedure written and signed off by Laboratory Director

☐ Reporting criteria established and final report form written

☐ Ongoing QC procedures established and documented

APPROVED: Medical Director _____ **Date:** _____