Standardized Protocol for Method Validation/Verification

Standard Operating Procedure
Quality Assurance Unit
Laboratory Services Section - Austin

Table of Contents

I. Purpose 2

II. Scope 2

III. Definitions 2

IV. Reagents/Media/Standards 4

V. Equipment 4

A. Instrument to be used for method verification/validation 4

B.Method Validation/Verification Software 4

VI. Procedure 4

A.Qualitative Methods 5

1.FDA cleared or approved methods. 5

2.Non-FDA cleared 9

3.Validation Summary 10

B.Quantitative Methods 12

1.FDA cleared or approved methods: 12

2.Non-FDA cleared 18

3.Validation Summary 22

C.Instrument Validation 22

VII. Experiment Section 23

VIII. Decision on Method Performance 33

IX. Other Statistics 37

X. References 37

I. Purpose

This document provides instructions for a uniform process of validating methods in the laboratory. It is meant to be a guideline and help the laboratory meet applicable CAP/CLIA regulatory requirements.

The selection of a new or revised method is the responsibility of DSHS Laboratory management. Method selection should start with a clinical perspective to ensure sufficient analytical reproducibility and accuracy to meet the clinical requirements and other considerations such as space, equipment and personnel, efficiency and cost effectiveness.

Following selection of a method, the assessment of its suitability begins with the understanding of the sources of potential analytical error. With the properly planned experiments/studies the laboratory can measure the error produced in a method and determine if it is acceptable for use in the laboratory. The Validation/Verification study will document this process.

Total error is the sum of random and systemic error and is used to make the final judgment on the acceptability of a new or modified method in the laboratory. The laboratory will assess Random and Systemic error and document its findings.

II. Scope

All Laboratory tests must be validated or verified before being placed into routine use for testing and reporting of patient results. Method validations are required for all new tests as well as any modification of existing procedures. Equipment validation/verifications are required for all new instruments and instruments that have been moved. All validation/verifications must be approved, signed and dated by the Laboratory Services Section Director prior to use.

III. Definitions

A. Accuracy – How close is the measured value to the “true” value. The difference can be described as the **Systemic error** (**inaccuracy, bias**) in the method.

B. Analytic Measurement Range (AMR) - The range of analyte values that a method can directly measure on the specimen without any dilution, concentration, or other pretreatment not part of the usual assay process.

C. CAP – College of American Pathologists. Deemed to be an accreditation body by CLIA and currently directs the Laboratory Accreditation Program (LAP), established in 1961.

D. CLIA- Clinical Laboratory Improvement Amendments of 1988. Responsible under the Centers for Medicare & Medicaid Services (CMS), an agency within the US Department of Health and Human Services for the regulation of clinical laboratories in the United States.

E. Correlation Coefficient - A number between -1 and 1 which measures the degree to which two variables are linearly related. A perfect linear relationship will have a correlation coefficient of 1

F. FDA – US Food and Drug Administration

G. Precision – reproducibility. The ability of the laboratory to duplicate results time after time on different days and with different operators. Measures **Random error** the precision or **imprecision** can be expressed in CV% from the calculated standard deviation SD and mean. Repeat measurements of samples at varying concentrations, within-run and between run over a period of time should be performed.

H. Qualitative results – Test results that are not reported as numbers. They are reported as positive/negative or reactive/nonreactive, etc.

I. Quantitative results – Test results that are reported as numbers

J. Reportable Range – Same as Analytic Measurement Range (AMR). How high and low can test result values be and still be accurate? This can be determined by a linearity study for quantitative methods.

K. Reference Range – Normal values for your patient population.

L. Analytical Sensitivity – The smallest quantity of an analyte that can be reproducibly distinguished from background levels. Positive agreement as compared to reference method. For quantitative methods this includes determining the Limit of Detection (Can be described by the slope of the calibration curve).

M. Diagnostic Sensitivity – The percentage of subjects with the target condition whose test values are positive.

N. Analytical Specificity – the ability of a method to detect only the analyte it is designed to detect. Negative agreement as compared to reference method. Can be measured with interference and recovery experiments.

O. Diagnostic Specificity – the percentage of subjects without the target condition whose test values are negative

P. Validation – "…the process of assessing the assay and its performance characteristics to determine the optimal conditions that will generate a reliable, reproducible, and accurate…result for the intended application." The term is often used instead of Verification. This can be source of confusion. For non-FDA approved/cleared tests: the laboratory must establish the performance specifications.

Q. Verification – The one-time process performed to determine or to confirm a test's expected performance compared to actual results produced by the laboratory (CAP definition). For tests cleared or approved by FDA, verification is required

IV. Reagents/Media/Standards

1. The laboratory must have sufficient in-house supplies such as reagents and media to perform the validation/verification.

2. It is ideal if the same lot of reagents/media are used throughout the entire validation/verification study.
3. Expiration dates of reagents/media should be long enough to complete the validation/verification study.
4. Ensure that the media/reagents used are appropriate for the method
5. Communicate any needs or changes with the Media Prep Team and Consumer Micro QC related to the preparation of media and/or reagents
6. Ensure that a sufficient quantity of purchased materials such as standards, calibrators and controls are available prior to starting the validation/verification study.

V. Equipment

A. Instrument to be used for method verification/validation

1. Ensure that there is sufficient space and that the environmental requirements can be met. (Example: located out of direct sunlight, humidity, temperature, etc.)

2. Ensure that proper electrical requirements, data ports, water, waste, and other manufacturer requirements are met for the proper functioning of the instrument.

B. **Method Validation/Verification Software** - will be available to the laboratory. The use of validation/verification software is mandatory for quantitative and optional for qualitative methods. The Quality Assurance Officers (QAO) will train and provide assistance in its use.

VI. Procedure

Each method validation/ verification study is a collection of experiments to assess performance and error in order to judge a method's suitability for use in the laboratory. A validation/verification plan should be created and approved prior to starting the validation/verification experiments to prevent unnecessary testing and ensure that the study is complete.

Acceptability Criteria – the laboratory must establish acceptance criteria as part of the validation/verification plan. Parameters for accuracy, precision, sensitivity and specificity should include a confidence level of at least 90%, or meet the claims of the manufacturer.

A. Qualitative Methods – includes semi quantitative testing that use cut offs such as hepatitis testing and some molecular testing. No values/concentrations are included in the patient report. Test results are reported as positive/negative, normal/ borderline/abnormal, reactive/nonreactive, detected /not detected, etc.

1. FDA cleared or approved methods. According to the Standard CLIA:

CFR 42 § 493.1253: *Establishment and verification of performance specifications*: States that each laboratory that introduces an unmodified, FDA-cleared or approved test system must demonstrate that it can obtain performance specifications comparable to those established by the manufacturer for the following performance characteristics before reporting patient test results: Accuracy, Precision, Reportable Range of the test results and verification that the manufacturer's reference intervals (normal values) are appropriate for the laboratory's patient population.

a. Accuracy: Demonstrates how close to the “true” value the new method can achieve. Test material can include: calibrators/controls, reference material, proficiency testing material with known values, samples tested by another lab using the same or similar method, or by comparing results to an established comparative method. Test material matrix should match or be as close to the sample matrix as possible.

Most sources recommend comparing at least 40 patient specimens. CLIA current guidance suggests a minimum of 20 samples. Fewer than 20 samples will need to be approved by the QAO before proceeding. A larger number has a better chance to detect interferences. Depending on the test system and test volume the number used can vary. The actual number is less important than the quality of the samples. The estimate of systematic errors will depend more on obtaining a wide range of test results than on a large number of samples.

A method comparison experiment for accuracy is recommended to be done over a minimum of 5 days. Continue for another 5 days if discrepancies are observed. If side-by side testing is done samples should be tested within 2 hours of each other to ensure that sample stability will not affect results. If this is not possible, refrigerating or freezing samples between testing may preserve the sample. Please take into account any freeze/thaw cycle limitations your method may have. If the laboratory cannot perform the experiment for the 5 days due to lack of samples, resources or other reasons, consult with your QAO before proceeding.

Document the results of the new method comparing the known values from the reference sources, another certified laboratory’s results or with results from the current method. It is preferable to include both reference and patient samples, but priority will be given to patient samples.

Calculate the percent of positive, negative and total accuracy by dividing observed results over known results multiplied by 100.

Example: New method = 19 positives, 20 negatives. Current method or reference material with known values = 20 positive, 20 negatives

Percent positive accuracy 19/20 X 100 = 95%
Percent negative accuracy 20/20 X 100 = 100%
Total accuracy 39/40 X 100 = 98%

b. **Precision**: Also known as Reproducibility. Can the new method duplicate the same results? Use samples that have a matrix as close as possible to the real specimen. For clinical tests patient samples are the first choice followed by control material and reference solutions.

Most sources agree that a minimum of 2 negative samples and 2 positive samples run in triplicate for 5 days will provide data for within-run and between-run components to estimate precision. Having different operators perform the precision experiment must be done for methods that are operator dependent.

Calculate the percent within-run (intra), between-run (inter) and total precision by dividing observed results over known results multiplied by 100.

Example:

ID	Day 1			Day 2			Day 3		
Pos sample	Pos	Pos	Pos	Pos	Pos	Pos	Pos	Pos	Pos
Pos sample	Pos	Pos	Pos	**Neg**	Pos	Pos	Pos	Pos	Pos
Neg sample	Neg	Neg	Neg	Neg	Neg	Neg	Neg	Neg	Neg
Neg sample	Neg	Neg	Neg	Neg	Neg	Neg	Neg	Neg	Neg
Within run %	12/12/x 100 = 100%			11/12 x 100 = 92%			12/12/x 100 = 100%		

ID	Day 4			Day 5			Between run %
Pos sample	Pos	Pos	Pos	Pos	Pos	Pos	15/15x100 = 100%
Pos sample	Pos	Pos	Pos	Pos	Pos	Pos	14/15x100 =93%
Neg sample	Neg	Neg	Neg	Neg	Neg	**Pos**	14/15x100 =93%
Neg sample	Neg	Neg	Neg	Neg	Neg	Neg	15/15x100 = 100%
Within run %	12/12/x 100 = 100%			11/12 x 100 = 92%			

Total Precision: 58/60 x 100 = 96.7%

c. **Reportable Range:** CLIA defines this as the highest and lowest test values that can be analyzed while maintaining accuracy. For tests without high or low values, define method criteria for a positive result.

To verify reportable range, test at least 3-5 low and high positive samples once. These samples can be combined with the accuracy/precision experiments. Include both weak and strong positive samples.

For methods depending on a cut-off value to determine positive results, testing positive specimens near the cut-off can serve as the cut-off validation. This is required by CAP:

IMM.33905 Qualitative Cut-Off Phase II
For qualitative tests that use a cut-off value to distinguish positive from negative, the cut-off value is established initially, and verified every 6 months thereafter.

NOTE: This requirement does not apply to FDA-cleared/approved in vitro diagnostic assays that report the qualitative result based on a predefined cut-off value.

This requirement applies only to certain tests that report qualitative results based on a quantitative measurement using a laboratory established threshold (cut-off value) to discriminate between a positive and negative clinical interpretation. The cut-off value that distinguishes a positive from a negative result should be established when the test is initially placed in service, and verified every 6 months thereafter. If the value of a calibrator or calibration verification material is near that of the cut-off, then the process of calibration or calibration verification satisfied this checklist requirement. If the laboratory is not able to access the actual numerical value from the instrument, this checklist requirement does not apply.

d. **Reference Range (Normal Values):** Provided by the manufacturer and verified by running known normal patients. If the lab has a similar patient population then the manufacturer's ranges or even published reference ranges from textbooks or scientific articles may be used.

The Reference Range can be verified by testing 20 known normal samples; if no more than 2 results fall outside the manufacturer/published range then that reference range can be considered to be verified. (CLSI guideline C28-A3c)

If the laboratory cannot reference the normal values, then the reference range will need to be established. This involves a selection of at least 120 reference samples for each group or subgroup that needs to be characterized. See your QAO to discuss options.

e. **Sensitivity & Specificity:** CLIA does not require that these parameters to be verified. CAP All Common Checklist 07.29.2013 says:

COM.40400 Analytic Sensitivity Phase II
The laboratory verifies or establishes the analytic sensitivity (lower detection limit) of each assay, as applicable.

> *NOTE: For laboratories subject to US regulations, documentation for FDA cleared/approved tests may consist of data from manufacturers or the published literature.*

CAP does not spell out what to do with FDA-cleared tests for Specificity but it is recommended that the laboratory reference literature or manufacturer documentation for the specificity of the method.

COM.40450 Analytical Specificity/Interfering Substances Phase II
For modified FDA-cleared/approved tests or LDT's, the results of each validation study include a sufficient number of samples to establish the test's analytical specificity.

f. **Validation Summary:** Once the method experiments are complete, summarize the results in a Method Validation/Verification Summary. Clearly state the purpose of the verification, what platform/method and the number of samples for each experiment. Any discrepant results should be investigated and explained in the Summary. Test results that show sample problems such as contamination and degradation should not be used in the assessment but still listed with an explanation.

The Summary should also contain a Conclusion stating weather the study met the acceptance criteria or not and its suitability for us in the laboratory.

Add the CAP Validation cover sheet (see attached form) and submit to the QA Officer for approval.

When parameters are just outside acceptance criteria, additional testing can be performed (add more samples to the study), but do not delete data. If the results show poor performance, check the instrument set-up, reagents, and procedures. Perform corrective actions and repeat the entire validation/verification study. Any discrepant results should be investigated and explained in the Summary.

If the study results fail to meet pre-established criteria, the test cannot be implemented for use in the laboratory

2. Non-FDA Cleared tests

Qualitative methods developed in-House, non-FDA cleared methods and FDA-cleared methods modified by the laboratory. According to CAP/CLIA:

Establishment of performance specifications: 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), 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:

Accuracy, Precision, Analytical sensitivity, Analytical specificity to include interfering substances, Reportable range, Reference intervals, Any other performance characteristics required for test performance, Determine calibration and control procedures and document all of the above.

If the calibrators, controls or standards are significantly changed or modified during the validation study, everything must be repeated.

a. For Qualitative methods follow the instructions above for **Accuracy, Precision, Reportable Range** and **Reference Range**.

b. **Sensitivity -** Due to the lack of quantitative data, Qualitative sensitivity validation is not addressed by:

Analytical Sensitivity- (Detection limit) has also been defined as "the lowest concentration of the analyte which the test can reliably detect as positive in the given matrix".

But rather by:

Diagnostic Sensitivity – The percent of subjects with the target condition whose test values are positive. Calculate by dividing the number of true positives by the sum of the number of true positives plus the number of false negatives and multiplying by 100. [TP ÷ (TP + FN)] x 100 = Estimated Diagnostic Sensitivity. Use the data from the comparison study to calculate Diagnostic Sensitivity.

Methods such as Molecular may have alternative guidelines, consult with the QAO before proceeding.

c. Specificity - Due to the lack of quantitative data, Qualitative sensitivity validation is not addressed by:

Analytical Specificity – the ability of a method to detect only the analyte that it was designed to detect.

But rather by:

Diagnostic Specificity: the percent of subjects without the target condition whose test values are negative. Calculate by dividing the number of true negatives by the sum of the number of true negatives plus the number of false positives and multiplying by 100. [TN ÷ (TN + FP)] x 100 = Estimated Diagnostic Specificity. Use the data from the comparison study to calculate Diagnostic Specificity.

Methods such as Molecular may have alternative guidelines, consult with the QAO before proceeding.

Interference: The laboratory must be aware of common interfering substances by referencing studies performed elsewhere (manufacturer or literature) or by performing studies.

Interference Study: may be required when reference interference information is not available. Consult with your QAO for more information.

Substances to be included in the interference study can be selected from, scientific articles, literature references, etc. Common blood interferences are; hemolysis, bilirubin, lipemia, preservatives and anticoagulants used in specimen collection.

See Experiment section for details on performing an Interference Study.

3. Validation Summary: Follow the same instructions as were given in A.1.e. Summary. In addition, summarize the results of the interference study if applicable. The specimen acceptance criteria may need to be adjusted depending on interference study results.

The validation study should include any other performance characteristics concerning your test method. For example, if a different sample type will be analyzed, then it must be included in the validation study.

For Laboratory Developed tests refer to CAP ALL Common checklist 07.29.2013 item:

COM.40630 LDT Reporting Phase I
Reports for laboratory-developed tests (LDT) contain a description of the method, a statement that the assay was developed by the laboratory and appropriate performance characteristics.

NOTE: General guidelines for reports are given in the Results Reporting sections of the checklists. Laboratories often include an LDT disclaimer as follows: "This test was developed and its performance characteristics determined by <insert laboratory/company name>. It has not been cleared or approved by the FDA. The laboratory is regulated under CLIA as qualified to perform high-complexity testing. This test is used for clinical purposes. It should not be regarded as investigational or for research."

Summary chart for CAP Accreditation requirements for validating laboratory tests

	FDA approved/cleared	LDTs & modified FDA tests
Accuracy & Precision (COM.40300, ph II)	Verify mfger's results	Establish (= validation)
Analytic sensitivity (LOD) (COM.40400, ph II)	Verify: manufacturer or literature documentation OK	Establish
Analytic specificity (interferences) (COM.40500, ph II)	Reference literature or manufacturer documentation	Establish; studies by manufacturer or in literature OK
Reportable range (AMR)(COM.40600, ph II)	Verify*	Establish*
Reference range (COM.50000, ph II)	Verify **	Establish**

*Reportable range (AMR, generally) is the range of values that the method can directly measure without dilution or concentration, while meeting specifications for accuracy & precision

--Details on establishing & validating AMR are in other checklists (ex. CHM, HEM, MOL)

**In some cases laboratories may use manufacturer or literature data when verification/establishment of a reference range is not practical: ex. pediatric blood cell count / index parameters; therapeutic drug levels.

B. Quantitative Methods – includes laboratory methods that report numbers. QA will provide Validation software to assist in statistical analysis.

1. FDA cleared or approved methods:

According to the Standard CAP/CLIA:
CFR 42 § 493.1253: *Establishment and verification of performance specifications*: States that each laboratory that introduces an unmodified, FDA-cleared or approved test system must demonstrate that it can obtain performance specifications comparable to those established by the manufacturer for the following performance characteristics before reporting patient test results: Accuracy, Precision, Reportable Range of the test results and verification that the manufacturer's reference intervals (normal values) are appropriate for the laboratory's patient population.

The same requirements apply to the Quantitative methods that were stated above with the qualitative methods. The approach to method validation is to perform a series of experiments designed to estimate certain types of errors:

Accuracy (systematic error or bias): comparison of method experiment. Perform the Recovery experiment as needed. Recovery studies assess proportional systemic error due to competitive reactions from substances within the sample including matrix effects.
Precision (random error): replication experiment, calculation of standard deviation
Reportable Range: linearity experiment
Reference Range: Reference range experiment

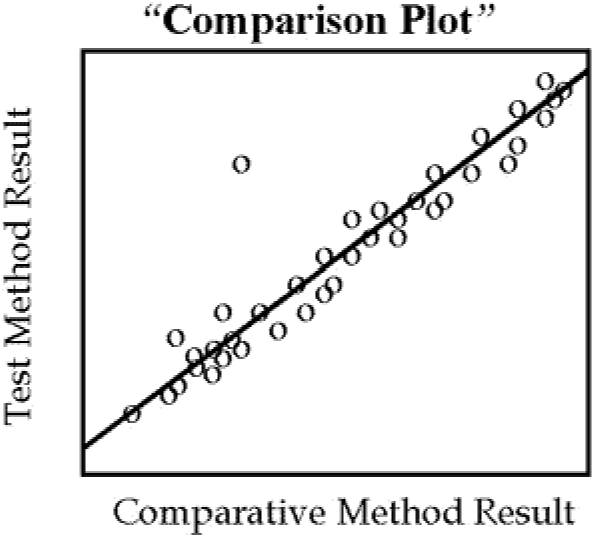
Acceptance criteria: For FDA-cleared or approved methods, observed results must compare or exceed the manufacturer's data. When performance specifications are not provided by the manufacturer refer to the Experiment section for information on Allowable total error.

a. Accuracy – Demonstrates how close to the "true" value the new method can achieve. A method comparison experiment is used to estimate inaccuracy or systematic error. Test material can include: calibrators/controls, reference material, proficiency testing material with known values, samples tested against a reference standard, high-quality method or another lab using the same method or by comparing results to an established in-house method.

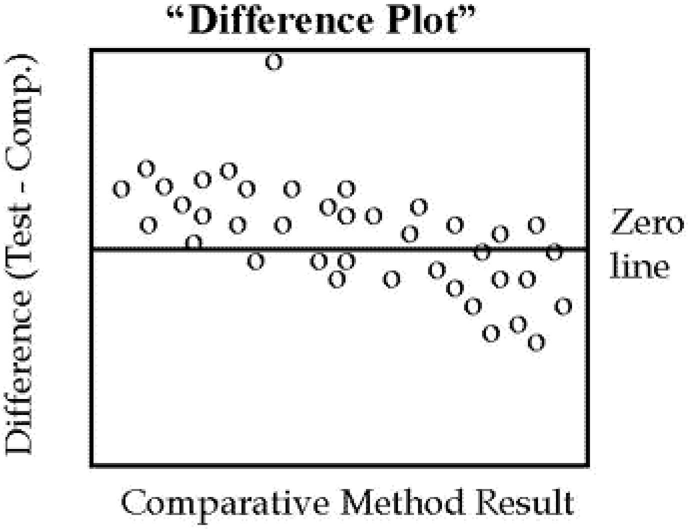
Most sources recommend comparing at least 20-40 patient specimens for a FDA-cleared or approved method. Using less than 20 samples will need to be approved by the QAO. A larger number has a better chance to detect interferences. Depending on the test system and test volume the number used can vary. The actual number is less important than the quality of the samples. The estimate of systematic error is more dependent on wide range of test results than on a large number of samples.

The method comparison experiment for accuracy is recommended to be done over a minimum of 5 days. Continue for another 5 days if discrepancies are observed. If the laboratory cannot perform the experiment for the 5 days due to lack of samples, resources or other reasons, consult with the QAO.

Prepare a comparison plot of all the data to assess the range, outliers, and linearity.



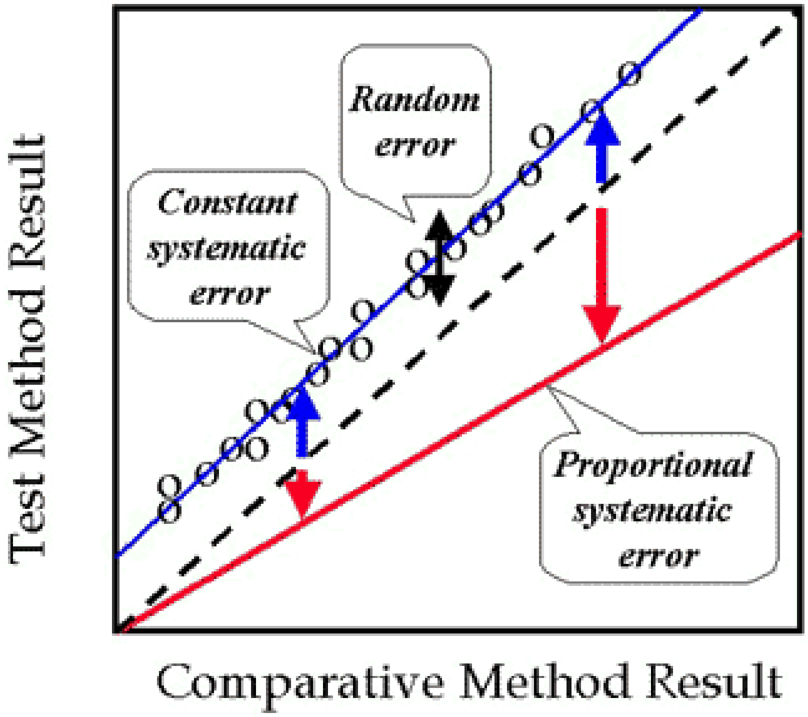
For methods that are not expected to show one-to-one agreement, for example enzyme analyses having different reaction conditions, the graph should be a "comparison plot" that displays the test result on the y-axis versus the comparison result on the x-axis, as shown by the figure above. As points are accumulated, a visual line of best fit should be drawn to show the general relationship between the methods and help identify discrepant results



If the two methods are expected to show one-to-one agreement, the initial graph may be a "difference plot" or "bias plot" that displays the difference between the test method results minus the comparative results on the y-axis versus the comparative result on the x-axis, such as shown in the figure above. The differences should scatter around the line of zero differences, half being above and half being below the line. Any large differences will stand out and draw attention to those specimens whose results need to be confirmed by repeat measurements. Review the data and graphs for any outlying points that do not fall within the general pattern of the other data points. For example, in the figure above there is one suspicious point in the difference plot. In addition, there are points that tend to scatter above the line at low concentrations and below the line at high concentrations, suggesting possibility of some constant and/or proportional systematic errors.

Precision or imprecision = Random error, Acurracy/Bias = Systematic Error, can be of two types: **constant systematic error** or **proportional systematic error.**

Constant and proportional systematic error can be seen on a Comparision plot.


Random error
Constant systematic error
Proportional systematic error
Test Method Result
Comparative Method Result

If the Comparison Plot shows a significant Proportional error, a Recovery Experiment may need to be performed. Consult the QAO for guidance.

Recovery Experiment: In the absence of a reliable comparison method, recovery studies can take on more importance however it is preferred to identify another more reliable, closer to a 'gold standard' method for use in a method comparison study. Consult with the QAO prior to performing.

The recovery experiment is performed to estimate proportional systematic error. Proportional Systematic error is observed when the difference of error increases as the concentration of the analyte increases. This type of error is often caused by a substance in the sample matrix that reacts with the sought for analyte and therefore competes with the analytical reagent. A recovery experiment may also be helpful for investigating calibration solutions whose assigned values are used to establish instrument set points. See the Experiment section for details.

Statistics: Accuracy / Bias (= systematic error):

Run comparison of methods study (test method vs. reference method, laboratory's previous method, or manufacturer's results, etc.). The line of best fit (calculated using a statistics program) provides the linear regression equation Y = a + bX

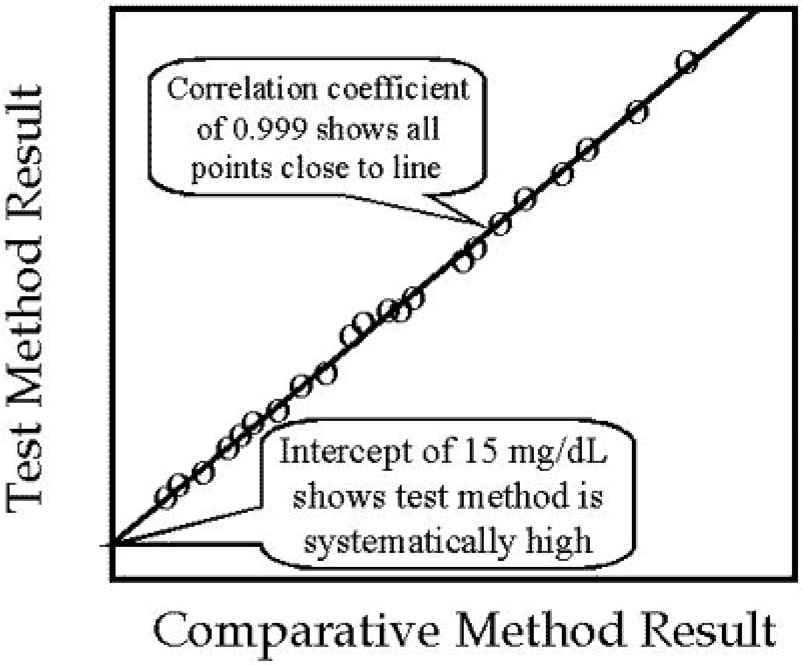
Calculate correlation coefficient "r". See VII. Experiment Section for more information

If “r” is high (≥ 0.99), use the regression line to find the bias at analyte concentrations that correspond to critical decision points (ex. glucose: 126 mg/dL).

If “r” < 0.975, the regression equation will not be reliable; use paired t-test to determine if a bias is present at the mean of the data. See Experiment section for details on t-test.

Analytes with a wide range (cholesterol, glucose, enzymes, etc.) tend to have a high “r” in comparison studies; analytes with a narrow range (electrolytes) tend to have low “r”

- **“r” should not be used to determine the acceptability of a new method. “r” measures how well the results from the 2 methods correlate (change together).**



b. Precision - Also known as Reproducibility. Can the new method duplicate the same results? It is important to test samples that have a matrix as close as possible to the real specimens. For clinical tests, patient samples are the first choice followed by control material and reference solutions.

Most sources agree that a minimum of 2-3 samples near each medical decision levels run for 3-5 replicates over 5 days will provide sufficient data for within-run and between-run components to estimate precision. Having different operators perform the precision experiment is important for methods that are operator dependent.

Statistics - Precision (= random error) – *Imprecision* is measured, when determined within a run = *repeatability*; imprecision across multiple runs across multiple days = *reproducibility*. The latter is most reflective of actual lab practice.

Random error is described quantitatively by calculating (use a statistics program) the mean (x), standard deviation (s), and coefficient of variation (CV). Compare these calculations to the manufacturer's data.

CLIA says that the laboratory should verify the manufacturer's claim for precision. This can be done with the F-test, as follows:

Use F test to see if variance (=SD^2) of test method is statistically different from old method, or claim of manufacturer

Example of how to use the F test:

Obtain the expected SD and number of measurements used in the replication experiment from the manufacturer's claims (usually included in the instrument documentation), e.g., SD =3 mg/dL based on 31 measurements.

Obtain the SD and number of measurements from the replication experiment, e.g., SD = 4 mg/dL based on 21 measurements.

Calculate the F-value, larger SD squared divided by smaller SD squared, i.e., $(4)^2/(3)^2 = 16/9 = 1.78$.

Look up the critical F-value for 20 degrees of freedom (df=N-1) in the numerator and 30 df in the denominator in the F-table (see Experiment Section), where the value found should be 1.93.

In this case, the calculated-F is less than the critical-F, which indicates there is no real difference between the SD observed in the laboratory and the SD claimed by the manufacturer.

Conclusion – the manufacturer's claim is verified when the calculated F value is less than the critical F value. See Experiment section for more information on F-Test.

c. **Reportable Range:** CAP Reportable range (analytic measurement range= AMR), is the range of values that the method can directly measure without dilution or concentration

For FDA-cleared tests with established parameters, Reportable Range (AMR) can be verified by running 3 points near low end, midpoint, and high end using calibration/control/reference matrix appropriate materials.

The AMR must be reverified at least every 6 months, and following changes in major system components or lots of analytically critical reagents (unless the laboratory can demonstrate that changing reagent lot numbers does not affect the range used to report patient test results, and control values are not adversely affected)

Data must be within the laboratory's acceptance criteria or within the manufacturer's stated range to be acceptable.

d. **Reference Range (Normal Values):** Provided by the manufacturer and verified by running known normal patients. If the laboratory has a similar patient population then the manufacturer's ranges or even published reference ranges from textbooks or scientific articles may be used.

The Reference Range can be verified by testing 20 known normal samples; if no more than 2 results fall outside the manufacturer/published range then that reference range can be considered to be verified. (CLSI guideline C28-A3c)

If the laboratory cannot reference the normal values, then the reference range will need to be established. This involves a selection of at least 120 reference samples for each group or subgroup that needs to be characterized. See your QAO to discuss options

Example: If there were no published or manufacturer normal ranges on total serum Cholesterol levels on women or children (define age range), then the laboratory would have to test at least 120 normal samples for each group to determine the reference range for women and the reference range for children (define age range).

2. Non-FDA Cleared

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), 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:

Accuracy, Precision, Analytical sensitivity, Analytical specificity to include interfering substances, Reportable range, Reference intervals, Any other performance characteristics required for test performance, Determine calibration and control procedures and document all of the above.

Accuracy (systematic error or bias): comparison of method experiment. Perform the Recovery experiment as needed. Recovery studies assess proportional systemic error due to competitive reactions from substances within the sample including matrix effects.
Precision (random error): replication experiment, calculation of standard deviation
Reportable Range: linearity experiment
Specificity (Systemic errors due to other materials present in samples): Interference experiment
Sensitivity: Detection limit experiment

Reference Range: Reference range experiment

Acceptance Criteria: When performance specifications are not provided by the manufacturer refer to the Experiment section for information on Allowable total error.

a. **Accuracy/ Bias (= systematic error):** Same as above in FDA-cleared tests except: Most sources recommend comparing at least 40 patient specimens for a Laboratory Developed Test (LDT). Using less than 40 samples will need to be approved by the QAO. A larger number has a better chance to detect interferences. Depending on the test system and test volume the number used can vary. The actual number is less important than the quality of the samples. The estimate of systematic error is more dependent on wide range of test results than on a large number of samples.

b. **Precision (= random error):** Select at least 2 different control materials that represent low and high medical decision concentrations. Analyze the low control and high control at least 20 times each within a run to obtain short term imprecision. Calculate mean, standard deviation and coefficient of variation for each material. Determine if short term imprecision is acceptable before proceeding to the long term imprecision experiment.

   Long-term imprecision experiment: Analyze 1 sample of each of the 2 materials on 20 different days to estimate long-term imprecision. Calculate the mean, standard deviation, and coefficient of variation for each material. Determine whether long-term imprecision is acceptable. Having different operators perform the precision experiment must be done for methods that are operator dependent.

   Using fewer results will have to be approved by the QAO. Compare to manufacturer's data or if there is none, compare to the allowable total error. See Experiment Section for information on Allowable total error.

   For within-run the acceptable SD is ¼ or less than the defined total error. For between-run studies the SD should be 1/3 or less than the defined total error.

c. **Reportable Range (analytic measurement range= AMR):** Same as above in FDA-cleared tests. The AMR must be reverified at least every 6 months. If the range has not been established, a linearity experiment will have to be performed.

   Linearity Experiment: Test a series of known dilutions of a highly elevated specimen or patient pool. The measured or reported test values are compared to the assigned values or to the dilution values, typically by plotting the measured values on the y-axis and the assigned or dilution values on the x-axis.

The Clinical Laboratory Standards Institute (CLSI) recommends a minimum of at least 4, preferably 5 different concentration levels. Less than 5 levels will have to be approved by the QAO. More than 5 may be used, particularly when the upper limit of the reportable range needs to be maximized.

Dilute the elevated sample into a series of dilutions, at least 5 levels. Run each level in triplicate. Plot the mean of the measured values on the y-axis versus the assigned values, relative values or dilution values on the x-axis. First draw a line point-to-point through the entire analytical range. Then manually draw the best straight line through as many points as possible, making sure that the line adheres to the lower points or lower standards or dilution values. At concentrations where the straight line no longer adheres to the points, estimate the systematic error due to non-linearity. Compare that systematic error plus the expected random error at the concentration (2 SDs) to the allowable total error for the test. See Experiment Section for details.

d. **Reference Range (Normal Values):** Same as above in FDA-cleared tests. When there are no well-established reference intervals are available, additional samples will be required.

The Clinical Laboratory Standards Institute (CLSI) recommends the use of carefully selected reference sample groups to establish reference intervals. These protocols typically use a minimum of 120 reference individuals for each group (or subgroup) that needs to be characterized.

When collecting 120 samples is not possible, an experimental validation may be performed by collecting and analyzed specimens from 40-60 individuals who represent the reference sample population.

Use of 40-60 specimens to make estimates of reference interval when the reference interval information from the manufacturer is not adequate, when the new test method is based on a different measurement principle and different measurement specificity, or when the test is being applied to a different patient population. Consult with your QAO if sufficient samples are unavailable.

e. **Analytical Sensitivity:** (Detection limit) is defined as “the lowest concentration of the analyte which the test can reliably detect as positive in the given matrix”.

US laboratory regulations require that detection limit (or analytical sensitivity) be established only for non-waived methods that have been modified by the laboratory and test systems not subject to FDA clearance, such as methods developed in-house. Good laboratory practice also dictates that detection limit be verified, when relevant, e.g., all forensic and therapeutic drug tests; TSH and similar immunoassay tests; some cardiac markers such the troponins; PSA and other cancer markers.

Two different kinds of samples are generally analyzed. One sample is a “blank” that has a zero concentration of the analyte of interest. The second is a “spiked” sample that has a low concentration of the analyte of interest. In some situations, several spiked samples may need to be prepared at progressively higher analyte concentrations. The blank and spiked samples are measured 20 times each, the means and SDs are calculated from the values observed, and the estimate of detection limit is calculated from. See Experiment section for details

f. **Analytic Specificity:** the ability of a method to detect only the analyte that it is designed to detect.

CAP All Common Checklist 07.29.2013:

COM.40450 Analytical Specificity/Interfering Substances Phase II
For modified FDA-cleared/approved tests or LDT's, the results of each validation study include a sufficient number of samples to establish the test's analytical specificity.

NOTE: The analytical specificity refers to the ability of a test or procedure to correctly identify or quantify an entity in the presence of interfering or cross-reactive substances that might be expected to be present. Laboratories are encouraged to review the cited references for guidance and provided confidence intervals to estimated performance characteristics.

The interference experiment is performed to estimate the systematic error caused by other materials that may be present in the specimen being analyzed. This error is defined as constant systematic errors since a given concentration of interfering material will generally cause a constant amount of error, regardless of the concentration of the sought for analyte in the specimen being tested. As the concentration of interfering material changes, however, the size of the error is expected to change.

A pair of test samples are prepared for analysis by the method under study. The first test sample is prepared by adding a solution of the suspected interfering material (called "interferer,") to a patient specimen that contains the sought-for analyte. A second test sample is prepared by diluting (with the same quantity of solution as used in the first specimen) another aliquot of the same patient specimen with pure solvent or a diluting solution that doesn't contain the suspected interference. Both test samples are analyzed by the method of interest to see if there is any difference in values due to the addition of the suspected interferer.

The substances to be tested are selected from the manufacturer's performance claims, literature reports, and summary articles on interfering materials, and data tabulations or databases. See Experiment section for details.

3. **Validation Summary**: Once the method experiments are complete, summarize the results in a Method Validation/Verification Summary. Clearly state the purpose of the validation/verification, platform/method and the number of samples for each experiment. Any discrepant results should be investigated and explained in the Summary. Test results that show sample problems such as contamination and degradation should not be used in the assessment but still listed with an explanation

   The Summary should also contain a Conclusion stating weather the study met the acceptance criteria or not and its suitability for use in the laboratory.

   Add the CAP Validation cover sheet and submit to your QA Officer for approval.

   If some parameters are just outside acceptance criteria, additional testing can be performed (add more samples to the study), but do not delete data. If the results show poor performance, check the instrument set-up, reagents, and procedures. Perform corrective actions and repeat the entire validation/verification study. Any discrepant results should be investigated and explained in the Summary

   If the study results fail to meet pre-established criteria, the test may not be implemented for use in the laboratory

C. **Instrument Validation** – New instruments as well as instruments that have been moved in the laboratory must be validated/verified prior to use.

CAP Requirements: METHOD PERFORMANCE SPECIFICATIONS

NOTE: Sound laboratory practice requires full characterization of each test/method/instrument system before its use in patient testing, without regard to when the test was first introduced by a given laboratory. For each test performed on blood, the laboratory must have data on accuracy, precision, analytic sensitivity, interferences and reportable range (i.e. analytic measurement range (AMR) as applicable).
The method performance specifications must be performed in the location in which patient testing will be performed. If an instrument is moved, the laboratory must verify the method performance specifications (i.e. accuracy, precision, reportable range) after the move to ensure that the test system was not affected by the relocation process or any changes due to the new environment (e.g. temperature, humidity, reagent storage conditions, etc.). The laboratory must follow manufacturer's instructions for instrument set up, maintenance, and system verification. Each instrument is considered a separate test system, including instruments of the same make and model. The laboratory must verify the performance specifications of each instrument.

1. **New Instrument of a different make or model of current instrument** – Must be validated for all method performance specifications including: accuracy, precision, analytic sensitivity, specificity and reportable range.

2. **Additional Instruments of same make & model as the current instrument**- Each instrument must be validated separately. If several instruments are validated/verified

at the same time only one validation is needed. Each instrument must be validated for method performance specifications including: accuracy, precision, reference range and reportable range (AMR).

a. Accuracy may be verified for the additional instrument by comparison study with the instrument currently in-use(15-20 samples).
b. No separate reference range study is needed for 2nd instrument, assuming comparison study showed absence of significant bias.

3. **Instruments that have been moved from one location to another in the laboratory** - Must be validated for method performance specifications including: accuracy, precision and reportable range (AMR).

4. **Validation Summary** - Once the method experiments are complete, summarize the results in a Method Validation/Verification Summary. Clearly state the purpose of the verification, what platform/method and the number of samples for each experiment. Any discrepant results should be investigated and explained in the Summary. Test results that show sample problems such as contamination and degradation should not be used in the assessment but still listed with an explanation.

   The Summary should also contain a Conclusion stating weather the instrument study met the acceptance criteria or not and its suitability for use in the laboratory.

   Add the CAP Validation cover sheet (see attached) and submit to the QA Officer for approval.

   Note: CAP requirement: If the laboratory uses more than one instrument to test for a given analyte, the instruments are checked against each other at least twice a year for correlation of results.

   Contact the QAO for more information

VII. Experiment Section

A. Detection Limit Experiment for Analytical Sensitivity

The detection limit experiment is intended to estimate the lowest concentration of an analyte that can be measured. This low concentration limit is obviously of interest in forensic drug testing, where the presence or absence of a drug may be the critical information from the test.

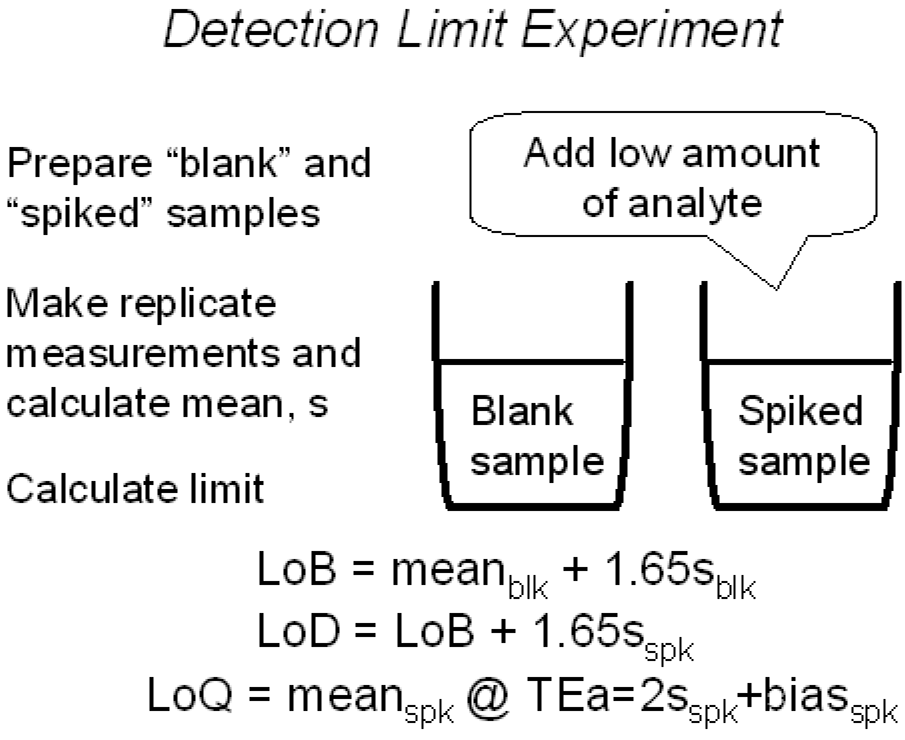
US laboratory regulations require that detection limit (or analytical sensitivity) be established for non-waived methods that have been modified by the laboratory and test systems not subject to FDA clearance, such as in-house developed methods.

Limit of Blank (LoB): Highest measurement result that is likely to be observed (with a stated probability) for a blank sample; typically estimated as a 95% one-side confidence limit by the mean value of the blank plus 1.65 times the SD of the blank.

Limit of Quantification (LoQ): Lowest amount of analyte that can be quantitatively determined with acceptable precision and accuracy.

Limit of Detection (LoD): Lowest amount of analyte in a sample that can be detected with (stated) probability, although perhaps not quantified as an exact value; Estimated as a 95% one-sided confidence limit by the mean of the blank plus 1.65 time the SD of the blank plus 1.65 times the SD of a low concentration sample.

Functional Sensitivity (FS): The analyte concentration at which the method CV is 20%



Detection Limit Experiment

Prepare “blank” and “spiked” samples

Make replicate measurements and calculate mean, s

Calculate limit

$$LoB = mean_{blk} + 1.65s_{blk}$$

$$LoD = LoB + 1.65s_{spk}$$

$$LoQ = mean_{spk} \text{ @ } TEa=2s_{spk}+bias_{spk}$$

A general description of the experimental procedure is provided in the accompanying figure. Two different kinds of samples are generally analyzed. One sample is a “blank” that has a zero concentration of the analyte of interest. The second is a “spiked” sample that has a low concentration of the analyte of interest. In some situations, particularly the estimation of FS and LoQ, several spiked samples may need to be prepared at progressively higher analyte concentrations. Both the blank and spiked samples are measured repeatedly in a replication type of experiment, then the means and SDs are calculated from the values observed, and the estimate of detection limit is calculated.

Blank solution. One aliquot of the blank solution is typically used for the "blank" and another aliquot is used to prepare a spiked sample. Ideally, the blank solution should have the same matrix as the regular patient samples. However, it is also common to use the "zero standard" from a series of calibrators as the blank and the lowest standard as the "spiked" sample.

Spiked sample. In verifying a claim for the detection limit of a method, the amount of analyte added to the blank solution should represent the detection concentration claimed by the manufacturer. To establish a detection limit, it may be necessary to prepare several spiked samples whose concentrations are in the analytical range of the expected detection limit. For some tests, it may be of interest to use samples from patients who are free of disease following treatment (i.e., PSA sera from patients treated for prostate cancer).
Number of replicate measurements. Generally 20 replicate measurements are recommended in the literature. This number is reasonable given that the detection limit experiment is a special case of the replication experiment, where 20 measurements are generally accepted as the minimum. The CLSI guideline suggests 20 replicates be made by a laboratory to verify a claim, but recommends a minimum of 60 by a manufacturer to establish a claim.

Time period of study. A within-run or short term study is often carried out when the main focus is the method performance on a blank solution. A longer time period, representing day-to-day performance, is recommended when the focus is on a "spiked" sample. The CLSI guideline recommends that LoD be estimated from data obtained over a period of "several days" and LoQ from data obtained over at least 5 runs, assumedly over a 5 day period.

For LoD, the claim is verified if no more than 1 of the 20 results on a spiked sample is below the LoB.

B. The Linearity or Reportable Range Experiment

It is important to determine the reportable range of a laboratory method, i.e., the lowest and highest test results that are reliable and can be reported. Manufacturers make claims for reportable range by stating the lower and upper limits of the range. It is critical to check those claims, particularly when a method is assumed to be linear and "two-point calibration" is used.

The Clinical Laboratory Standards Institute (CLSI) recommends the use of a minimum of at least 4 – preferably 5 – different concentration levels. More than 5 levels may be used,

particularly when the upper limit of the reportable range needs to be maximized. Often 5 levels are convenient and almost always sufficient.
It is convenient to use two pools – one near the zero level or close to the detection limit and the other near or slightly above the expected upper limit of the reportable range. Determine the total volume needed for the analyses, select appropriate volumetric pipettes and follow the steps below:

1. Label the low pool "Pool 1" and the high pool "Pool 5."
2. Prepare Mixture 2 (75/25) with 3 parts Pool 1 + 1 part Pool 5.
3. Prepare Mixture 3 (50/50) with 2 parts Pool 1 + 2 parts Pool 5.
4. Prepare Mixture 4 (25/75) with a part Pool 1 + 3 parts Pool 5.

If more levels are desired, this dilution protocol can be modified, e.g., the two pools could be mixed 4 to 1, 3 to 2, 2 to 3, and 1 to 4 to give four intermediate levels for a total of six levels for the experiment.

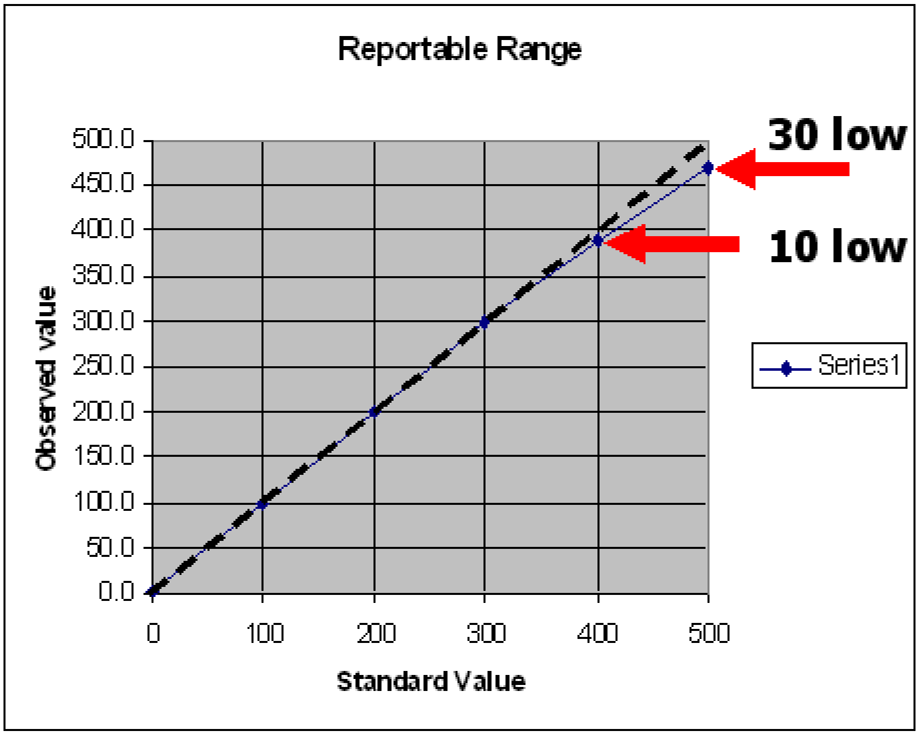
Number of replicate measurements
CLSI recommends making 4 measurements on each specimen or pool. However, 3 replicates are generally sufficient, including triplicate measurements on the original high and low pools.

Data analysis
Plot the mean of each measured concentration level on the y-axis versus the assigned values, relative values or dilution values on the x-axis. Draw a line point-to-point through the entire analytical range. Manually draw the best straight line through as many points as possible, making sure that the line adheres to the lower points or lower standards or dilution values. At concentrations where the straight line no longer adheres to the points, estimate the systematic error due to non-linearity. Compare that systematic error plus the expected random error at the concentration (2 SDs) to the allowable total error for the test.

Cholesterol Example:
The data are as follows:

0 assigned, observed 0, 5, 10, average 5.0;
100 assigned, observed 95, 100, 105, average 100;
200 assigned, observed 200, 195, 205, average 200;
assigned 300, observed 310, 300, 290, average 300;
assigned 400, observed 380, 390, 400, average 390;
assigned 500, observed 470, 460, 480, average 470.

The figure below shows the average values plotted on the y-axis against the assigned values on the x-axis.



The solid line represents the line drawn point-to-point and the dashed line represents the straight line fitted to the points in the low to middle part of the range. Systematic differences are estimated to be 0 mg/dL at 300 mg/dL, 10 mg/dL at 400 mg/dL, and 30 mg/dL at 500 mg/dL. The reportable range clearly extends to 300 mg/dL, but does it extend to 400 mg/dL or 500 mg/dL?

At 500 mg/dL, given a method with a CV of 3.0%, the SD would be 15 mg/dL and the 2SD estimate of random error would be 30 mg/dL. This means that a sample with a true value of 500 would, on average, be observed to be 470 mg/dL due to the systematic error from non-linearity. In addition, that value could be ±30 mg/dL due to random error, i.e., the expected value would be in the range from 440 to 500 mg/dL for a sample with a true value of 500 mg/dL. Given that the CLIA criterion for the allowable total error is 10%, (see page 29 for allowable error table)which is 50 mg/dL at a level of 500 mg/dL, the errors that would be observed at 500 mg/dL could be larger than the allowable error, thus the reportable range should be restricted to a lower concentration.

At 400 mg/dL, the SD would be 12 mg/dL, giving a 2SD estimate of random error as 24 mg/dL. A sample with a true value of 400 mg/dL would, on average, be observed to be 390 mg/dL due to the systematic error from non-linearity. Addition of the random error gives an expected range from 366 to 414 mg/dL, which means a result might be in error by as much as 34 mg/dL. The CLIA criterion of 10% provides an allowable total error of 40 mg/dL at 400 mg/dL, thus those expected results are correct with the allowable total error (34 mg/dL < 40 mg/dL), thus the reportable range does extend to 400 mg/dL.

C. **Regression Statistics for Comparison Experiment -** Statistical software will be available to assist in calculating parameters needed to evaluate method performance.

The regression statistics that should be calculated are the slope (b) and y-intercept of the line (a), the standard deviation of the points about that line ($s_{y/x}$), and the correlation coefficient (r, the Pearson product moment correlation coefficient). You may also see the slope designated as m, the y-intercept as b, and the standard deviation as $s_{residuals}$, respectively. The correlation coefficient is included to help you decide whether the linear regression statistics or the t-test statistics will provide the most reliable estimates of systematic error.

correlation coefficient "**r**", is a number between -1 and 1 and describes how well the results between the methods change together. If there is perfect linear relationship with positive slope between the two variables, we have a correlation coefficient of 1; if there is positive correlation, whenever one variable has a high (low) value, so does the other. If there is a perfect linear relationship with negative slope between the two variables, we have a correlation coefficient of -1; if there is negative correlation, whenever one variable has a high (low) value, the other has a low (high) value. A correlation coefficient of 0 means that there is no linear relationship between the variables.

A comparison plot should be used to display the data from the comparison of methods experiment (plotting the comparison method value on the x-axis and the test method value on the y-axis). This plot is then used to visually inspect the data to identify possible outliers and to assess the range of linear agreement

Statistical tests such as the t-test and the F-test can be used to determine whether a difference exists between two quantities which are estimates of performance parameters. These tests are called *tests of significance* and they test whether the experimental data are adequate to support a conclusion that a difference has been observed. The hypothesis being tested is called the *null hypothesis*, which states that there is no difference between the two quantities. When the test statistic (t or F) is large, the null hypothesis is disproved. The conclusion is that the difference is *statistically significant*. In practical terms, this means that a real difference has been observed. When the test statistic is small, the conclusion is that the null hypothesis stands and there is *no statistically significant difference* between the two quantities. No real difference has been observed.

t-Test – A t-test can be used to test two means and determine whether a difference exists between them. There are both paired and unpaired forms of the t-test. This refers to whether the two means being compared come from the same statistical samples or from different statistical samples. For example, the paired t-test is used when there are pairs of measurements on one set of samples such as in the comparison of methods experiment in which every sample is analyzed by both the test and comparative method. The unpaired form is used when testing the difference between means in two separate sets of samples, such as the mean of the reference values for females versus the mean for males.

$$\boxed{\mathbf{t = \frac{bias}{s_d \,/\, \sqrt{N}}}}$$

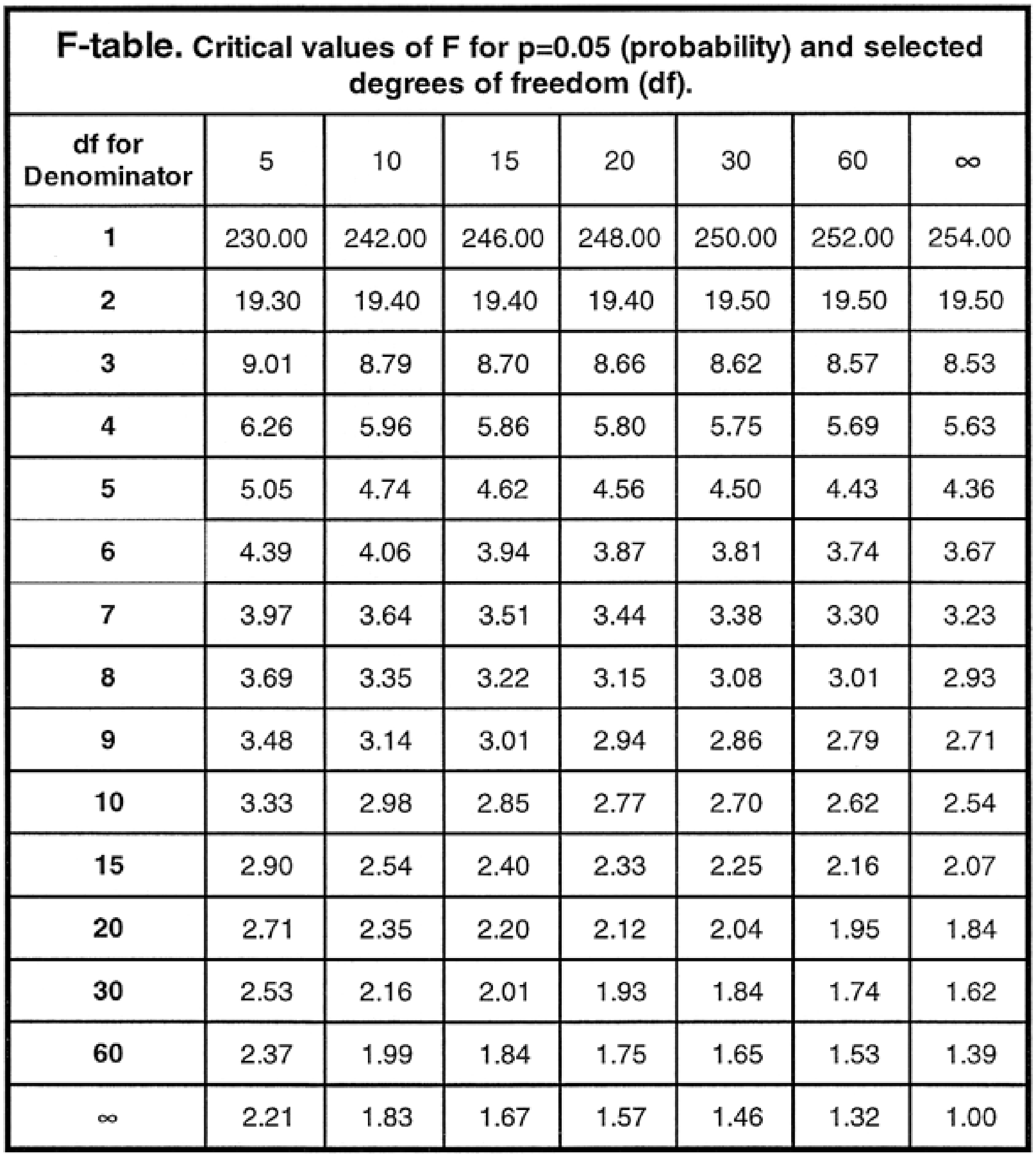
It is a ratio of two terms, one that represents a systematic difference or error (bias) and another that represents a random error ($SD_{diff}/N^{1/2}$; in this case it has the form of a standard error of a mean because mean values are being tested). The value of t expresses the magnitude of the systematic error in multiples of random error. For example, a t-value of six would indicate that the systematic error term is six times larger than the random error term. This amount of systematic error is much larger than the amount that might be observable just due to the uncertainty in the experimental data. Ratios greater than two or three are not expected

The interpretation of the t-test does not address the acceptability of the method's performance, but only whether there is systematic error present.

F-Test - In method validation studies, the *F-test* is sometimes used to compare the variance of the test method with the variance of the comparative method. *Variance* is simply the square of the standard deviation. Whereas the t-test tells whether the difference between two mean values is statistically significant, the F-test tells whether the difference in variances is statistically significant. In short, the t-test is used for systematic error or inaccuracy, and the F-test is used for random error or imprecision.
To perform the F-test, the standard deviations of the test and comparative methods are squared and the larger variance is divided by the smaller variance, as shown below:

$$\mathbf{F = \frac{(s_1)^2}{(s_2)^2}}$$

where s_1 is the larger s (or less precise method) and s_2 is the smaller s (or more precise method).



F-table. Critical values of F for p=0.05 (probability) and selected degrees of freedom (df).

df for Denominator	5	10	15	20	30	60	∞
1	230.00	242.00	246.00	248.00	250.00	252.00	254.00
2	19.30	19.40	19.40	19.40	19.50	19.50	19.50
3	9.01	8.79	8.70	8.66	8.62	8.57	8.53
4	6.26	5.96	5.86	5.80	5.75	5.69	5.63
5	5.05	4.74	4.62	4.56	4.50	4.43	4.36
6	4.39	4.06	3.94	3.87	3.81	3.74	3.67
7	3.97	3.64	3.51	3.44	3.38	3.30	3.23
8	3.69	3.35	3.22	3.15	3.08	3.01	2.93
9	3.48	3.14	3.01	2.94	2.86	2.79	2.71
10	3.33	2.98	2.85	2.77	2.70	2.62	2.54
15	2.90	2.54	2.40	2.33	2.25	2.16	2.07
20	2.71	2.35	2.20	2.12	2.04	1.95	1.84
30	2.53	2.16	2.01	1.93	1.84	1.74	1.62
60	2.37	1.99	1.84	1.75	1.65	1.53	1.39
∞	2.21	1.83	1.67	1.57	1.46	1.32	1.00

The F-test is interpreted by comparing the calculated F-value with a critical F-value, which is obtained from the statistical table above. The null hypothesis being tested is that there is no difference between the variances of the two methods. The null hypothesis is rejected when the observed F-value is greater than the critical F-value, and at that point, the difference in variances or random errors is said to be statistically significant.

Observe that the F-test interpretation says nothing about whether the random error of the test method is acceptable, but only whether it is different from that of the comparative method. This test is good for comparing the test method’s random error against manufacturer’s data.

If the test method is being compared to a different method, then acceptability depends on the size of the random error, regardless of whether it is less than or greater than the random error of the comparative method.

D. **Allowable Total Error-** is used to determine if data is good or bad. If no information is available from the manufacturer, use other sources for data on allowable error/acceptable performance. A number of organizations have published parameters for Acceptable performance.

The table below contains information on CLIA proficiency testing criteria for acceptable analytical performance, as printed in the Federal Register February 28, 1992;57(40):7002-186. These guidelines for acceptable performance can be used as Analytical Quality Requirements.

Total Allowable Error Table

Test or Analyte	Acceptable Performance
Cholesterol, total	Target value ± 10%
Cholesterol, high dens. lipoprotein	Target value ± 30%
Glucose	Target value ± 6 mg/dL or ± 10% (greater)
Triglycerides	Target value ± 25%
Blood lead	Target value ± 10% or ± 4 mcg/dL (greater)
Hemoglobin	Target ± 7%
Rubella	Target value ± 2 dilution or (pos. or neg.)

For information on analytes not included in the table, consult with the QAO.

E. Interference Experiment

For Qualitative testing:

Collect 1 - 2 negative samples and 1 - 2 positive samples. For each sample aliquot the same volume into 2 samples (A&B). For sample A add the amount of interferer that is near the maximum concentration expected in the patient population. For sample B add the same amount of saline, water or a solvent that matches the sample matrix. The amount of interferer substance should be small relative to the original test volume to minimize dilution effects. Precision is more important because it is essential to maintain the exact same volumes in the pair of test samples. Run both A & B in duplicate and compare results.

Results

Sample ID	First result	Second result
Pos A (I added)	Pos	Pos
Pos A (blank added)	Pos	Pos
Pos B (I added)	Pos	Pos
Pos B (blank added)	Pos	Pos
Neg C (I added)	Pos	Neg
Neg C (blank added)	Neg	Neg
Neg D (I added)	Neg	Neg
Neg D (blank added)	Neg	Neg

Calculate the % of correct values

Observed Results/Correct results x 100 or 15/16 x 100 = 93.8%

Since the acceptable criteria for qualitative testing is 90%, this would be acceptable. If the results do not meet the criteria, additional samples may be tested and included in the data base. Review the sample acceptance criteria.

For Quantitative Testing:

Perform the Interference Experiment for each interference substance tested (interferer). Collect a minimum of 1 - 2 samples that will achieve a distinctly elevated level. For each sample aliquot the same volume of interfering substance or blank into 2 samples (A&B). For sample A add the amount of interferer. For sample B add the same amount of saline, water or a solvent that matches the sample matrix (blank). The amount of interfering substance should be small relative to the original test volume to minimize the effects of dilution. Precision is more important because it is essential to maintain the exact same volumes in the pair of test samples. Run both A & B in duplicate and compare results.

Data Calculation: Example Glucose Test (mg/dL)

1. Tabulate results.
   Sample A (with I added) = 110, 112 mg/dL
   Sample A (with blank added) = 98, 102 mg/dL
   Sample B (with I added) = 106,108 mg/dL
   Sample B (with blank added) = 93, 95 mg/dL

2. Calculate the averages.
   Sample A (with I added) = 111 mg/dL
   Sample A (with blank added) = 100 mg/dL
   Sample B (with I added) =107 mg/dL
   Sample B (with blank added) = 94 mg/dL

3. Calculate differences of sample averages.

Sample A difference = 11 mg/dL
Sample B difference = 13 mg/dL

4. <u>Average the differences of sample averages.</u>
   Average difference = 12 mg/dL

Criteria for acceptable performance: The judgment on acceptability is made by comparing the observed systematic error (Interference) with the amount of error allowable for the test. The example above used a glucose test. CLIA states that glucose testing should be correct within 10%. At the upper end of the reference range (110mg/dl), the allowable error would be 11 mg/dl. Because the observed interference of 12 mg/dl is greater than the allowable error, the performance of this method is not acceptable. Samples would have to be free of this particular interfering substance to be acceptable for testing in the laboratory.

See VII Experimental Section D: Allowable Total Error for information on the allowable error for other analytes.

VIII. Decision on Method Performance

Acceptance Criteria for most CAP/CLIA validation/verifications will meet the following:

> Accuracy, precision, sensitivity and specificity parameters will match or exceed 90% as compared to the current/reference method or with reference materials with known values. AND/OR
>
> Meet or exceed the claims of the manufacturer.

The following section is to be used when there is no documented information on acceptable performance for a method. In addition, for methods with marginal performance, the Method Performance Chart can help assess the acceptability of the procedure for the laboratory.

The decision about the acceptability of method performance depends on the size of the observed errors relative to a "standard" or quality requirement that defines the total allowable error. Method performance is acceptable when the observed errors are smaller than or equal to the total allowable error. Method performance is NOT acceptable when the observed errors are larger than the total allowable error.

See section VII. D. Experiment Section, Total allowable Error Table for information on acceptable analytical performance and the Total Allowable Error (TEa)

Systematic Error (SE) + Random Error (RE) = Total Error (TE)

Literature provides four different recommendations on how to combine random error (RE) and systematic errors (SE):

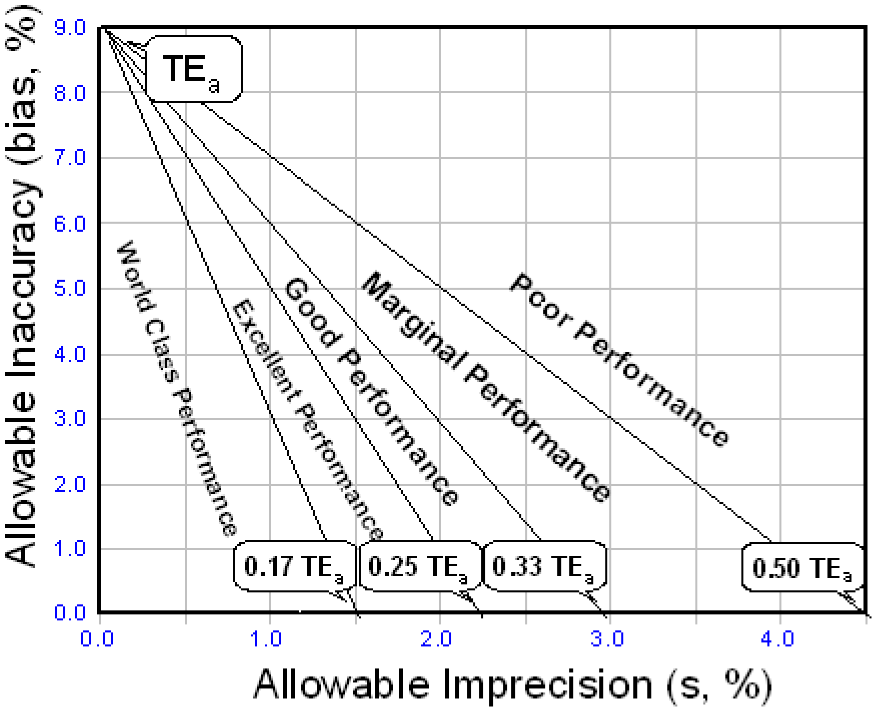
Add bias + 2 times the observed SD, bias + 2SD < TE_a;
Add bias + 3 times the observed SD, bias + 3SD < TE_a;
Add bias + 4 times the observed SD, bias + 4SD < TE_a;
Add bias + 6 times the observed SD, bias + 5SD < TE_a.

Rather than choose between these recommendations, all four can be incorporated into a graphical decision tool – a Method Decision Chart. The chart is simple to construct, minimizes the need for additional calculations, and provides a graphical picture that simplifies the interpretation and judgment of method performance.

How to construct a Method Decision Chart?
First, express the allowable total error as a percentage of the medical decision concentration. Most CLIA allowable errors are already given in percent. For those given in concentration units, express the allowable error as a percent of the medical decision concentration of interest, i.e., divide the allowable error by the medical decision concentration and multiply by 100 to express as a percentage.

Express observed SD (s,%) and bias (bias,%) in percent.

Use a sheet of graph paper and complete the following steps:



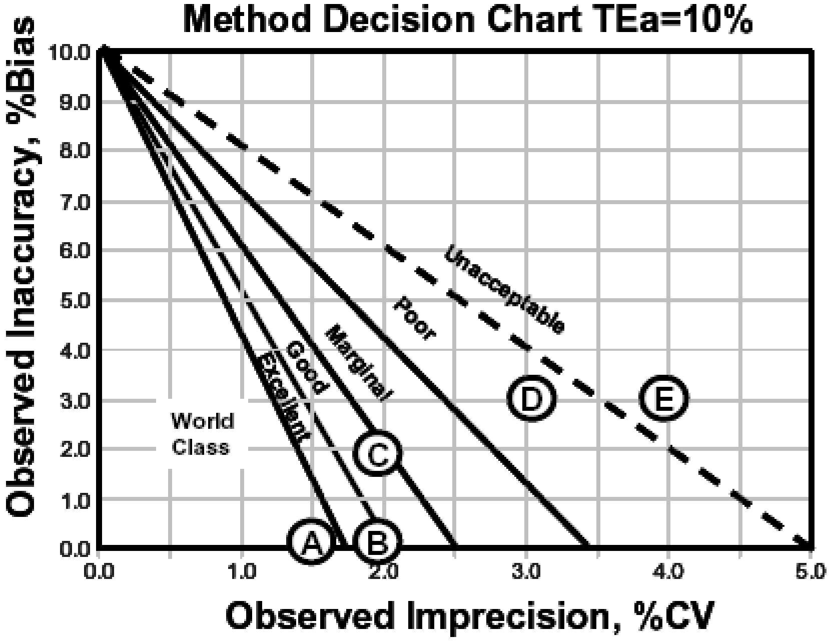
1. Label the y-axis "Allowable inaccuracy, (bias,%)" and scale from 0 to TE_a, e.g., if TE_a is 10%, scale the y-axis from 0 to 10% in increments of 1%.

2. Label the x-axis "Allowable imprecision, (s,%) and scale from 0 to 0.5 TE_a, e.g., if TE_a is 10%, scale the x-axis from 0 to 5% in increments of 0.5%.
3. Draw a line for bias + 2 SD from TE_a on the y-axis to 0.5 TE_a on the x-axis, e.g., if TE_a is 10%, draw the line from 10% on the y-axis to 5% on the x-axis.
4. Draw a line for bias + 3 SD from TE_a on the y-axis to 0.33 TE_a on the x-axis, e.g., if TE_a is 10%, draw the line from 10% on the y-axis to 3.33% on the x-axis.
5. Draw a line for bias + 4 SD from TE_a on the y-axis to 0.25 TE_a on the x-axis, e.g., if TE_a is 10%, draw the line from 10% on the y-axis to 2.5% on the x-axis.
6. Draw a line for bias + 5 SD from TE_a on the y-axis to 0.20 TE_a on the x-axis, e.g., for TE_a = 10%, draw the line from 10% (y-axis) to 2.0% (x-axis).
7. Draw a line for bias + 6 SD from TE_a on the y-axis to 0.17 TE_a on the x-axis, e.g., if TE_a is 10%, draw the line from 10% on the y-axis to 1.7% on the x-axis.
8. Label the regions **"unacceptable**," **"poor**," **"marginal**," **"good**," **"excellent**," and "**world class**" as shown in the figure.

Express the observed SD and bias in percent, then plot the point whose x-coordinate is the observed imprecision and y-coordinate is the observed inaccuracy. This point is called the "operating point" because it describes how the method operates. Judge the performance of the method on the basis of the location of the operating point, as follows:

- A method with **unacceptable performance** does not meet the requirement for quality, even when the method is working properly. It is not acceptable for routine operation.
- A method with **poor performance** might have been considered acceptable prior to the recent introduction of the principles of Six Sigma Quality Management, but industrial benchmarks now set a minimum standard of 3-Sigma performance for a routine production process, thus performance in the region between 2- Sigma and 3-Sigma is not satisfactory.
- A method with **marginal performance** provides the necessary quality when everything is working correctly. However, it may be difficult to manage in routine operation, may require 4 to 8 controls per run, and a Total QC strategy that emphasizes well-trained operators, reduced rotation of personnel, more aggressive preventive maintenance, careful monitoring of patient test results, and continual efforts to improve the method performance.
- A method with **good performance** meets the requirement for quality and can be well-managed in routine operation with 2 to 4 control measurements per run using multirule QC procedures or a single control rule having 2.5s control limits.
- A method with **excellent performance** is acceptable and should be well-managed in routine operation with only 2 control measurements per run using a single control rule with 2.5s or 3.0s control limits.
- A method with **world class performance** is usually the easiest to manage and control, generally requiring 1 or 2 control measurements per run and a single control rule with wide limits, such as 3.0s or 3.5s.

Example:
The following examples illustrate the evaluation of cholesterol methods, where the CLIA requirement for acceptable performance is an allowable total error of 10%



Note: the chart above contains several errors: The y axis should be labeled as "Allowable Inaccuracy, %Bias" and the x axis should say "Allowable Imprecision, %SD"

A. A cholesterol method with a CV of 1.5% and a bias of 0.0% provides world class quality, as shown by the operating point labeled A, whose x-coordinate is 1.5 and y-coordinate is 0.0. This method is clearly acceptable and will be easy to manage and control in routine operation using 2 control measurements per run and a single control rule having 3.5s control limits.

B. A cholesterol method with a CV of 2.0% and bias of 0.0% provides excellent performance, as shown by operating point B. This method is clearly acceptable and will be controllable in routine service using 2 control measurements and a single control rule having 3.0s or 2.5s control limits.

C. A cholesterol method with a CV of 2.0% and a bias of 2.0% has an operating point that falls on the line between excellent performance and good performance, as shown by point C. A careful assessment of QC is required and will show that a multirule procedure with a total of 4 control measurements per run may be necessary to guarantee that desired quality is achieved by this method.

D. A cholesterol method having a CV of 3.0% and a bias of 3.0% satisfies the specifications of the National Cholesterol Education Program (NCEP). To assess whether these performance specifications are adequate, an operating point can be plotted with a y-coordinate of 3.0% and an x-coordinate of 3.0%, as shown by the point labeled D in the accompanying figure. Such a method would have "marginal" performance, which means that the quality of the test results will be okay if everything is working perfectly, but it will be very difficult to detect problems and maintain the desired quality during routine service operation.

E. A cholesterol method with a CV of 4.0% and a bias of 3.0% may be representative of the type of screening methods encountered in shopping malls and pharmacies. As shown by operating point E, such a method does not provide the quality necessary to meet the CLIA requirement for acceptable performance

IX. Other Statistics

A. Deming regression

This refers to an alternate way of calculating regression statistics when the range of data isn't as wide as desired for ordinary linear regression (i.e., the correlation coefficient doesn't satisfy the criterion of being 0.99 or greater). An assumption in ordinary linear regression is that the x-values are well known and any difference between x and y-values is assignable to error in the y-value. In Deming regression, the errors between methods are assigned to both methods in proportion to the variances of the methods. This requires additional information about the performance of the methods, particularly the ratio of the variances of the two methods**.**

B. Passing-Bablock regression

Another alternate regression procedure is called Passing-Bablock regression, after the authors who described the technique. The slopes are calculated for every combination of two points in the data set, then the slopes are ordered and ranked, and the median value is selected as the best estimate. There is no need for additional information about the relative SDs of the test and comparative methods.

X. References

A. Westgard J. O.: Basic Method Validation, Westgard Quality Corporation
B. CLIA, 42CFR 42 § 493.1253 Standard: Establishment and verification of performance specifications.
C. Lumsden, J.H.: Laboratory test method validation
D. CAP Master All Common Checklist 07.9.2013, Page 26 to 32, Method Performance Specifications
E. Sarewitz S.J.: CAP Accreditation Requirements for Validating Laboratory Tests, 7/9/13
F. Jennings L., Van Deerlin V.M., Gulley M.L.: Recommended Principles and Practices for Validating Clinical Molecular Pathology Tests
G. Loeffelholz M.: Test Method Verification in the Microbiology Laboratory

H. Clark R.B., Lewinski M.A., Loeffelholz M.J., Tibbetts R.J. Cumitech 31A; Verification and Validation of Procedures in the Clinical Microbiology Laboratory