

# Validation of New Methods

Laboratory Medicine Residency Didactic Conference

June 10, 2003

# Sources

- Chapter 13, *Tietz Textbook of Clinical Chemistry*, 3<sup>rd</sup> ed.
- “Toxicology Assay Validation Procedure”, Clinical Chemistry Laboratory, YNHH, written by Petrie Rainey, November, 1995.
- The Westgard Web Lessons: [www.westgard.com/lesson.htm](http://www.westgard.com/lesson.htm)
- “Method Performance Specifications” section of the CAP “Laboratory General Checklist” ([www.cap.org](http://www.cap.org))
- CLIA-88, Subpart K, 493.1213 Standard; “Establishment and verification of method performance specifications”; refer to [www.vh.org/adult/provider/pathology/CLIA/CLIAHP.html](http://www.vh.org/adult/provider/pathology/CLIA/CLIAHP.html)

# Basic Definitions: Accuracy

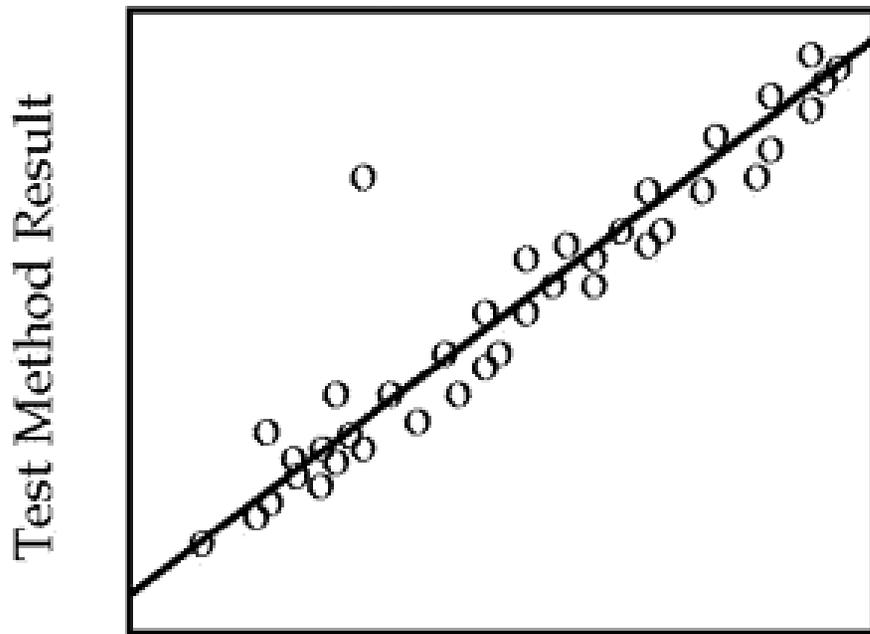
- Defined as “the closeness of the agreement between the measured value of an analyte and its ‘true’ value.”
- Of course, the catch is defining a “true” value.
- A number of organizations exist to define and provided standardized reference materials:
  - National Reference System for the Clinical Laboratory (NRSCL)
  - The Standards Committee of the AACC
  - National Institute of Standards and Technology (NIST)
- Accuracy is often expressed using the concept of *error*.

# Basic Definitions: Systematic Error

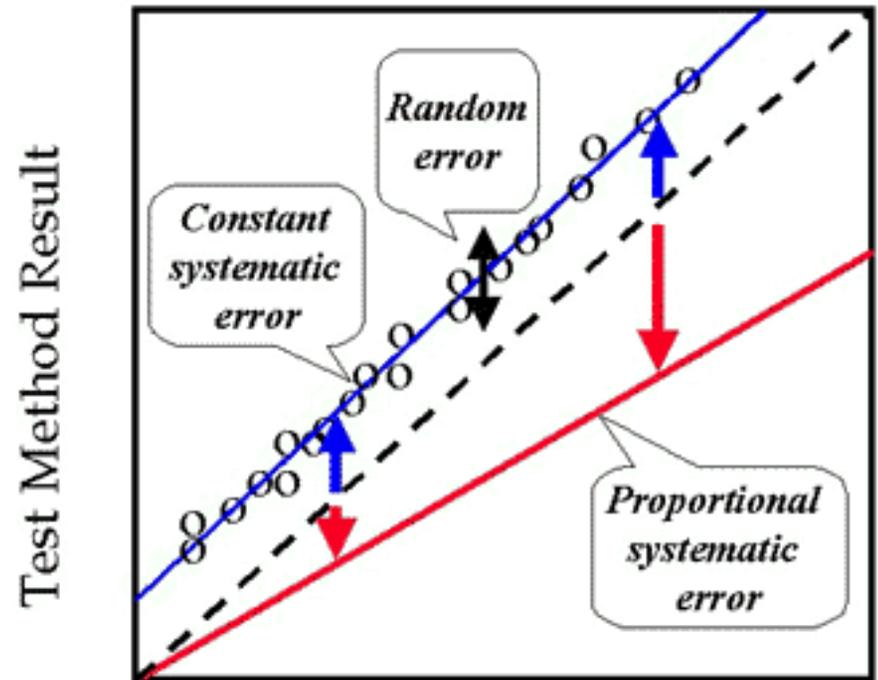
- A measure of the agreement between the average measured value and the “true” value.
- Best described by a correlation plot between measured and reference values.
- Can be divided into two categories
  - *Constant Systematic Error*: Manifests as a constant difference between the measured and true values and corresponds to the “vertical offset” of a correlation plot. Often caused by an interfering substance. Can be practically eliminated by a proper “blank” as long as the interference is constant.
  - *Proportional Systematic Error*: Varies with the magnitude of the value, often linearly, and corresponds to a non-unity slope of the correlation plot. Generally reflects a failing of the methodology. Sometimes, may be simply fixed by recalibration and, other times, may be more insidious.
- “Bias” is defined simply as the error expressed as a percent.

# Correlation/Comparison Plot

“Comparison Plot”



Comparative Method Result



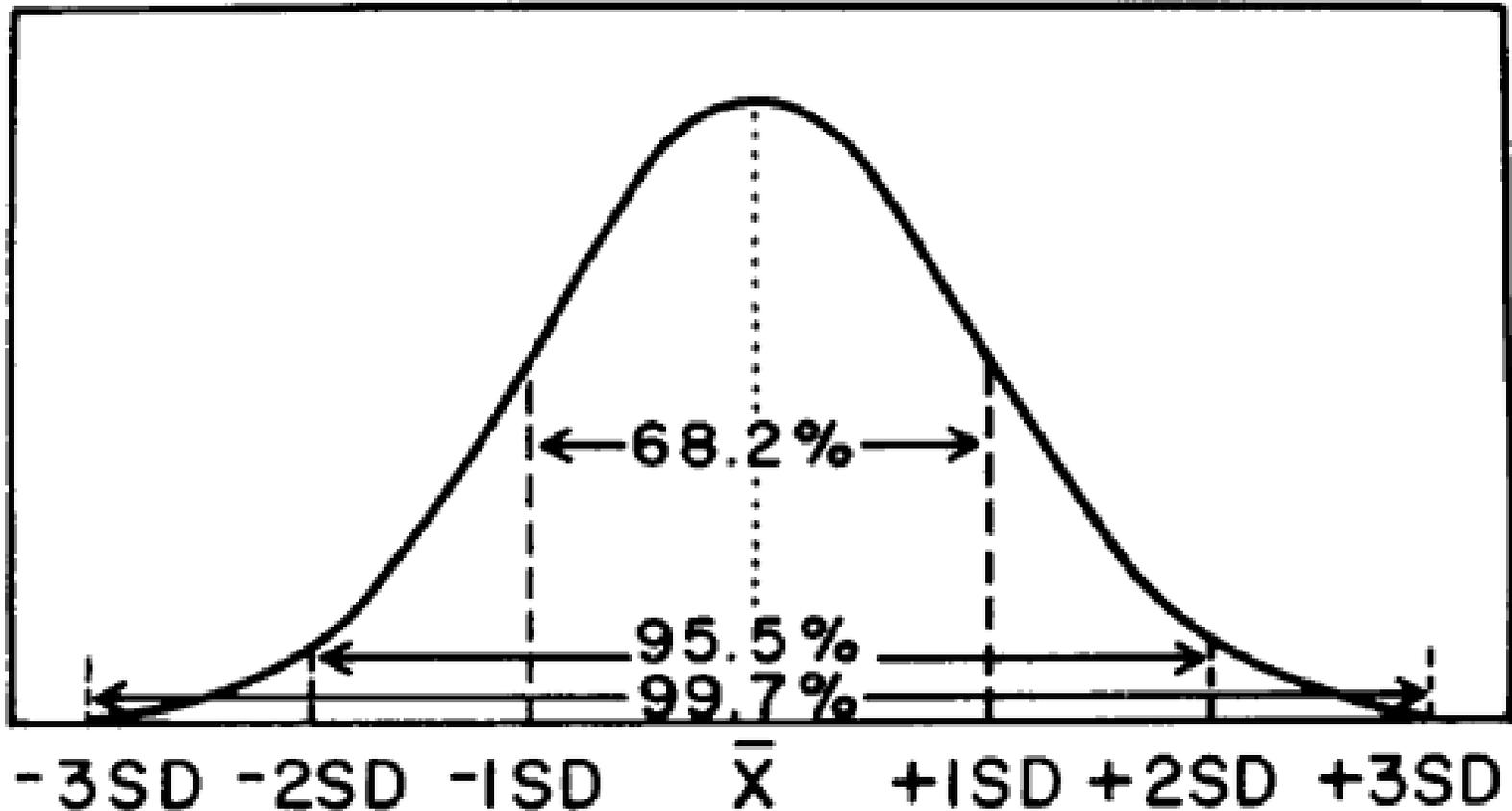
Test Method Result

Comparative Method Result

# Basic Definitions: Random Error

- Fluctuations of the measured values about their mean due to random factors.
- An indicator of *analytical precision*.
- In a correlation plot, reflected by the deviations of the measured values from a straight line.
- When conforming to a Gaussian or Normal distributions, these fluctuations are mathematically encapsulated by the *standard deviation*, which can be used to predict statistical probabilities.
- Although not strictly a component of analytical accuracy, random error can be considered to contribute to the “correctness” of a reported result and is included in the *total error*.

# Gaussian Distribution



# Basic Definitions: Total Error

- Sum of the systematic error and the random error (2 – 4 times the standard deviation of the imprecision).
- Meant to be a “worst case” incorrect answer.
- As both the systematic and the random error can be value-dependent, it is necessary to define the total error at each of the important parameter values (i.e. at the clinical decision points) and compare to the pre-defined medical requirements for the assay.

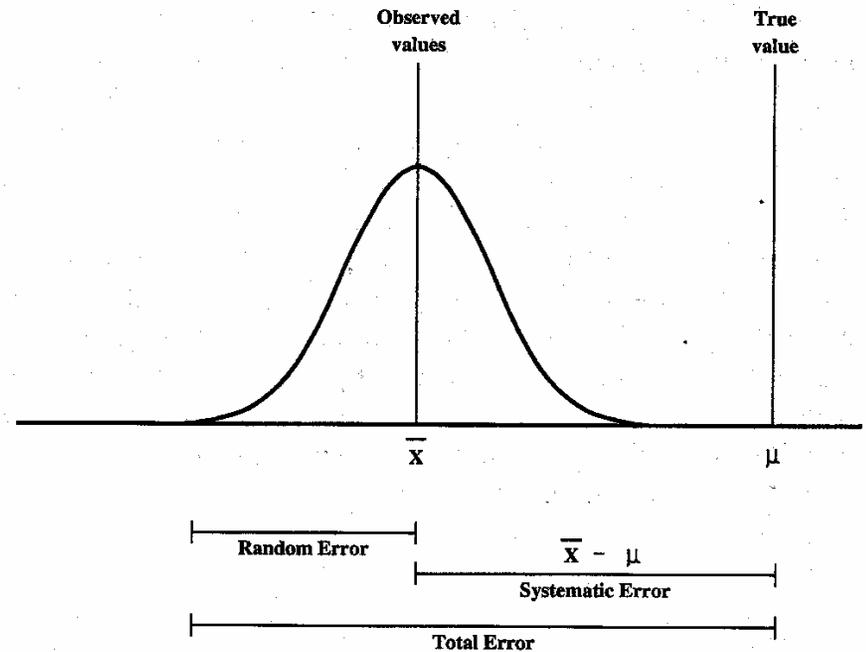


FIGURE 13-3. The total error concept of accuracy. (Modified from Westgard, J.O., de Vos, D.J., Hunt, M.R., et al.: Method Evaluation. Houston, American Society for Medical Technology, 1978.)

# Basic Definitions: Precision

- Same as random error.
- Always basically determined by repeated analysis of samples with expected “correct” answers.
- Three categories:
  - Within-run precision
  - Between-run or within-day precision
  - Day-to-day precision (best estimate of total precision and for calculation of the total error)

# Basic Definitions:

## Coefficient of Variation

- Coefficient of Variation (C.V.) is a way to express imprecision.
- $C.V. = \text{standard deviation} \div \text{mean}$
- Although, strictly, the C.V. can vary with the magnitude of the value, it is often used as if it doesn't. Hence, in this way it expresses a proportional variability. Across a limited range this assumption is probably warranted.
- In general, coefficients of variation should be below 5%, and should rarely exceed 10%, across the medically-relevant range of analyte concentrations.

# Steps Required to Validate an Assay (in no particular order)

- Establish Accuracy
- Establish Precision
- Establish the Reportable Range (linear range)
- Determine the Analytical Sensitivity and the Lower Limit of Detection
- Investigate the Analytical Specificity (search for interferences)
- Establish the Reference Range (define medically relevant decision points).

# Evaluation of Accuracy

## 1. Comparison with Reference Standards

- Based on repeated analysis of multiple established standard solutions, which can be purchased or prepared locally, across a range of analyte concentrations. Expected values based on a reference or “gold standard” method.
- Standard samples may be aqueous (common) or serum (preferred) based.
- Make a correlation plot and fit to a straight line (if appropriate). Look for both constant and proportional systematic error.

## 2. Recovery Experiment

- Known amounts of the analyte are added to samples containing an unknown amount of the analyte. Measurements are taken with and without the added analyte. The difference between the measurements is compared to the expected value as an indicator of “recovery”.
- This is typically performed across a range of concentrations and analyzed similarly to a correlation plot, above.

## 3. Correlation with a Current or Accepted Method

- Prepared or clinical (preferred) samples are simultaneously analyzed by the two methods (i.e. the new and the old) and compared.
- This step is absolutely required if replacing an existing method.
- Differences may represent inaccuracies with either the old or new method and may need to be considered when establishing the new reference range.

# Evaluation of Precision

- Need to get a sense of both within-run and day-to-day imprecision, at different analyte concentrations.
- For example, using a series of serum standards selected around medically important decision levels, independently assay five aliquots of each on at least five different days.
- Express variations in terms of coefficient of variation (within-run and day-to-day), bias and total error.
- The day-to-day C.V. should be ideally  $< 5\%$  and certainly no worse than  $10\%$ , except at very low levels.
- The bias should not exceed the C.V. (implies a limitation of the assay that is probably correctable).
- The total error should be less than the acceptable performance standard for (CAP) proficiency testing.

# Evaluation of the Reportable Range

- Although not strictly defined, generally implies establishing the *linearity range* for an assay.
- Non-linear assays are allowed, in which case this would translate into the region that fits the non-linear function used to describe the data.
- Evaluate by diluting standards and comparing the measured result to the expected result.
- The reportable range is defined by the highest and lowest points that fall within 1 C.V. of a straight line (or equivalent non-linear curve).
- CLIA-88 requires that the reportable range does not exceed the range of the available calibrators.

# Determination of Analytical Sensitivity and the Lower Limit of Detection

- The analytical sensitivity is defined as the incremental increase in measured signal per incremental increase in analyte concentration.
- Analytical sensitivity is simply described by the slope of the calibration curve.
- An assay with a high analytical sensitivity *should* have a low limit of detection, but not necessarily.
- Hence, one way to estimate the lower limit of detection would be to estimate the minimal detectable increment in signal and calculate the corresponding concentration of the analyte.
- However, this would be inadequate as other factors are probably more important (lower limit of linearity, non-specific signals from blank samples, increased random error at low analyte concentrations, etc.)

# Lower Limit of Detection, cont.

- The lower limit of detection should be investigated with three different approaches; the results then compared and the most sensible limit chosen.
- The first two choices would be (1) the lower limit of linearity and (2) minimal detectable analytical signal or the limit of quantification (where the C.V. exceeds 20%).
- The third method is to prepare a series of clinically-relevant blank samples (if possible) and measure the effective analyte concentration. The lower limit of detection would then be defined as the mean plus three standard deviations (such that there would be less than a 0.3 % chance of confusing analytical noise with a “true” measurement).

# Evaluation of Analytical Specificity

- The effect of common analytical interferences, including plasma hemoglobin, bilirubin, lipemia, etc., must be established.
- Generally, serum is sought with high levels of the above potential interferents and then either
  - (i) the analyte can be added to the samples and a recovery experiment performed or
  - (ii) equal amounts of the samples can be mixed with samples containing pre-measured amounts of the analyte and a mixing study performed.
- Additionally, for every assay, the director should consider any other potential interferent. For example, for TDM, other structurally related drugs should be tested for interference in the assay.

# Establishment of the Reference Range

References may be established in any number of several ways and is a major topic by itself (Chapter 14 in Tietz):

1. Direct determination of the reference range in the laboratory by monitoring analyte values in an appropriate population.
2. Use of a reference range established elsewhere, either based upon the medical literature, test manufacturer or within another clinical laboratory. Must somehow validate the appropriateness of this reference range locally (sample comparison, clinical study, etc.).
3. Establishment of the continued validity of a previous reference range used for an alternative method within the same local laboratory (i.e. “transfer” of the reference range from the old method to the new one).