Med TechNet

Online Continuing Education for the Clinical Laboratory Professional

Med TechNet Online Services✦ East Amherst, NY November 1997

DECISION MAKING IN THE CLINICAL LABORATORY: A QUANTITATIVE AND STATISTICAL APPROACH FOR METHODS EVALUATION

Robert L. Klick

Associate Professor
Department of Clinical Laboratory Sciences
State University of New York at Buffalo

Double click on the 's for a Med TechNet tip!

At a glance ...

Introduction

Clinical laboratories are constantly changing to meet customers' needs, operate on a financially sound basis, successfully pass proficiency programs and comply with the mandates of regulatory agencies. Several aspects of meeting customers' needs from a laboratory's perspective are producing information which is accurate, precise and available in a timely manner. Laboratory

information which is both accurate and precise allows members of health care delivery team to make the best possible decisions concerning the care and treatment of the patient.

Federal and state governmental bodies as well as accreditation organizations now require clinical laboratories to demonstrate and document the performance of the assays in use and to evaluate all non-exempt new assays being introduced in the laboratory. The Clinical Laboratory Improvement Act of 1988 (CLIA 88) published testing performance standards for individual analytes.[1]

The CLIA 88 analyte performance standards are expressed as total fixed error limits or as standard deviation total error limits. Implicit in the published total fixed error limits is that the method's combined inaccuracy and imprecision should be less than the allowable error limits. The intent of the analytical fixed error limits is to insure that analytical errors will not invalidate the medical usefulness of the test results and that the method's performance compare favorably the results form other laboratories.

The Joint Commission 1996 Comprehensive Accreditation Manual for Pathology and Clinical Laboratory Services has an objective of Improving Organizational Performance (section 1).[2] In this section, improvements which benefit patients are listed. Essential activities to improve the quality of patient care are: performance measurement, performance assessment and performance improvement. This process of performance measurement, assessment and improvement can be applied to laboratory methods evaluation and will be discussed in this article.

CLIA 88 regulations regarding test performance are covered in subparts J (Patient Test Management) and K (Quality Control).

Elements of the law cover:

- **Test Methodologies**
- **Normal Ranges (Reference Intervals)**
- **Test Interferences**
- **Performance Claims**
  - **Accuracy**
  - **Precision**
  - **Sensitivity (Analytical Detection Limit)**
  - **Linearity**
  - **Specificity (Interferences And Cross-reactivity)**

If a new test is being considered for introduction by the laboratory, a needs assessment should be done to determine the expected number of analyses which will be performed daily by the laboratory and when, during the day, specimens will be received and what the expected turnaround time is for the test. The information gathered from the need assessment is essential in equipment and method selection as well as cost analysis.

After a needs assessment has defined the expected workload and turn a round time, the laboratory administration should establish a working group to carry out the method selection, evaluation, comparison, user training and implementation. The group should make an objective search of vendors who supply equipment and/or reagents and quality control materials for the new test. An excellent comprehensive source of product information can be found in the ***Clinical Laboratory Reference*** published yearly as a supplement to ***(MLO) Medical Laboratory Observer***, Medical Economics Publishing Inc., Five Paragon Drive, Montvale, NJ 07645-1742. Members of the working group should also contact and tap the collective wisdom of colleagues who are performing the test. Remember that these users have had to answer many of the same questions with which *you* are now grappling.

The process of implementing a new test requires a considerable amount of time,

energy and cost, hence the administration must be prepared to make an investment in the project. Depending on the regulations which govern clinical testing in a particular state and the test complexity, the laboratory will be required to establish the performance characteristics of the assay and may be required to successfully pass one or more rounds of proficiency testing prior to patient testing.

After acquiring the equipment, associated reagents and supplies, the working group should become familiar with the use, calibration of the assay and any special procedures (ex. dilutions) or data reduction required.

Precision Studies

After the familiarization phase, precision of the method should be established. The least robust assessment of reproducibility is ***Within Run Precision***. Within run precision can be quickly evaluated by running pooled sera/blood or quality control material multiple times to establish reproducibility characteristics in terms of standard deviation (sd) and percent coefficient of variation (%CV). A minimum of twenty replications at medically important concentrations or activities should be assayed and used to calculate the above parameters of precision. The author highly recommends that a personal computer with spreadsheet software and printer be available in each clinical laboratory for data reduction, statistical analysis and graphic representation. Three excellent software companies offering spreadsheet programs are: Excel version 5 or later (Microsoft Corp.), Quattro Pro version 5 or later (Corel) and Lotus 123 (Lotus Development Corporation). All of these products are exceptionally good for use in the laboratory but the buyer should investigate these and others to insure that they contain the features the user desires. I will be using Excel version 5 or Inplot 4 (GraphPad Software) for calculations and graphs in this article.

Within run precision is computed, by entering the data in a column, selecting descriptive statistics of the replicate determinations, the mean ($\bar{x}$) and sd along with other statistic information is calculated. The %CV is not returned, but can be easily calculated by selecting a cell to place the result and typing =(sd/$\bar{x}$) x 100 and enter. (See **Appendix 1**, page 24, details.) Note, the numerical values for sd and $\bar{x}$ can be entered into the equation or the cell locations from the descriptive statistics output can be used.

Always inspect the results to determine if any of the data points appears to be spurious (not drawn from the same population as the other observations). If a value is markedly different from the others, the possibility exists that the value is an outlier and can be eliminated from the replicate determinations. If the suspect observation exceeds the mean ± 3sd, the probability of this occurrence is 0.26% or one in 385 times by random change alone. Hence with a small sample size, the likelihood of exceeding ± 3sd from the mean is small and the questionable data point can be removed from the replicates. After the outlier is removed the corrected mean, sd and %CV must be calculated.

What are some common causes of outliers in clinical analyses? Several possible causes include insufficient sample to allow the proper volume to be pipetted, a blockage of the pipette probe by specimen debris or a calculation or transposition error. Note, if in a small number of replicates, two values exceed the mean ± 3sd, the user should investigate and identify the cause of the disparate results, because it is highly unlikely that two such observations in a small population would occur by chance alone!

Manufacturers often provide information about the performance of a new instrument or technology in the form of within run, between day or total precision. As users of the new equipment, you should be assured that your precision is statistically no different than that of the manufacturer. If the precision of your method is different (worse) than that of the manufacturer, a problem exists which needs to be identified and corrected. The problem may originate with the equipment, reagents, supplies or the operators.

The statistical test which can be applied to test the user's precision versus that of the manufacturer is the **Sample Variance Compared to Some Value, Chi-Squared Test.** The Chi-squared calculated value (χ^2) is:

$$\chi^2 = \frac{User\ sd^2\ (n-1)}{Manufacturer\ sd^2}$$

Where n-1 degrees or freedom is the number of determinations minus 1. The critical Chi-square value for n-1 at a given confidence level (usually 95%) is obtained from a Chi -square table. If the calculated χ^2 value exceeds the critical χ^2 value, then the variance of the method as compared to the manufacturer's claim is different at that confidence level. The reason for the difference should be explored and corrected. If the calculated χ^2 is smaller that the critical χ^2, than the variance of the method being evaluated is not statistical different than the manufacturer's claims. The user can conclude that precision of the method has been validated. Note the Chi-square test can be applied to within run, between day and total precision estimates if comparable estimates are provided by the manufacturer.

EXAMPLE: Precision Data Verification

Example of verification of users precision based of manufacturer's precision data:

A clinical instrument manufacturer states that for total cholesterol the within run precision (sd) was 2.6 mg/dL at 200 mg/dL. The user determined the method's within run precision to be 3.0 mg/dL in a replication study on a serum analyzer 24 times with a mean of 212 mg/dL. Is the user's precision the same as that of the manufacturer?

$$Calculated\ \chi^2 = \frac{(3.0\ mg/dL)^2\ x\ 23}{(2.6\ mg/dL)^2} = 30.6$$

The critical Chi-square value at the 95% confidence level for 23 degrees of freedom is 35.17, hence the calculated χ^2 value (30.6) is less the critical χ^2 value (35.17). The conclusion is that the precision of the method is not different than the manufacturer's precision at the 95% confidence level.

Before evaluating the acceptability of the within run precision, based on the CLIA total fixed error limits, you may wish to ask the question, is the precision of the new method the same as or different than the old method's precision. If the precision of the two methods are different, it is advantageous for the precision of the new method to be better than that of the old method and certainly not worse! The statistical test which answers this question is the **F-test** for analysis of variance (sd^2). The null hypothesis (question being evaluated) states that the variability (variance) between the two methods are the same. The F-test can accept the null hypothesis that the variances of the two method are the same at some confidence level or reject the null hypothesis. Normally, the null hypothesis is tested at the 95% confidence level, which implies that a difference in variances of the two methods would occur by chance along 1 in 20 times or less (ex. 1 in 35 times).

EXAMPLE: F-TEST

Use the F-test to determine if a method's variance is statistically equal to that of a second method:

In the cholesterol evaluation study, the method under evaluation had a standard deviation of 2.95 mg/dL and the comparison method (old method) had a standard deviation of 4.02 mg/dL (both estimates are based on replication experiments using the same control material whose mean cholesterol concentration was 199 mg/dL). There were 10 measurements by the method under evaluation and 17 measurements by the old method.

$$F = \frac{Larger\ (s.d.)^2}{Smaller\ (s.d.)^2} = \frac{4.02^2}{2.95^2} = 1.86$$

Degrees of freedom = n-1

Look up the F critical value 16 degrees of freedom for the numerator and 9 degrees of freedom for the denominator in the F table (**Table 1**).

F critical = 2.98 (p=0.05)

F calc. > F critical; Reject Null Hypothesis

F calc. < F critical; Accept Null Hypothesis

TABLE 1
Critical Value of F (p<0.05)
Degrees of Freedom in the Numerator

Degrees of Freedom in the Denominator	1	2	3	4	5	6	7	8	9	10	11	12	14	16	20
1	161.00	200.00	216.00	225.00	230.00	234.00	2[illegible]0	239.00	241.00	242.00	243.00	244.00	245.00	246.00	248.00
2	18.51	19.00	19.16	19.25	19.30	19.33	19.36	19.37	19.38	19.39	19.40	19.41	19.42	19.43	19.44
3	10.13	9.55	9.28	9.12	9.01	8.94	8.88	8.84	8.81	8.78	8.76	8.74	8.71	8.69	8.66
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.93	5.91	5.87	5.84	5.80
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.78	4.74	4.70	4.68	4.64	4.60	4.56
6	5.99	5.14	4.67	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.03	4.00	3.96	3.92	3.87
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.63	3.60	3.57	3.52	3.49	3.44
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.34	3.31	3.28	3.23	3.20	3.15
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.13	3.10	3.07	3.02	2.98	2.93
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.97	2.94	2.91	2.86	2.82	2.77

Is the difference in precision between the two methods statistically significant?

Calculate the F-value:

F calc. = 1.86 < F critical = 2.98; therefore there is greater than a 5% probability of observing such a large difference in variances by chance alone and although different, they are not statistically different.

Statistical estimates of imprecision (within run precision) can be used as a first criteria of acceptable or unacceptable performance. ***Within run sd should be less than the CLIA total error limit*** for the particular analyte being evaluated. Ideally the random error (imprecision) should be 0.25 or 0.33 of the total error limit.

Total Error Limits

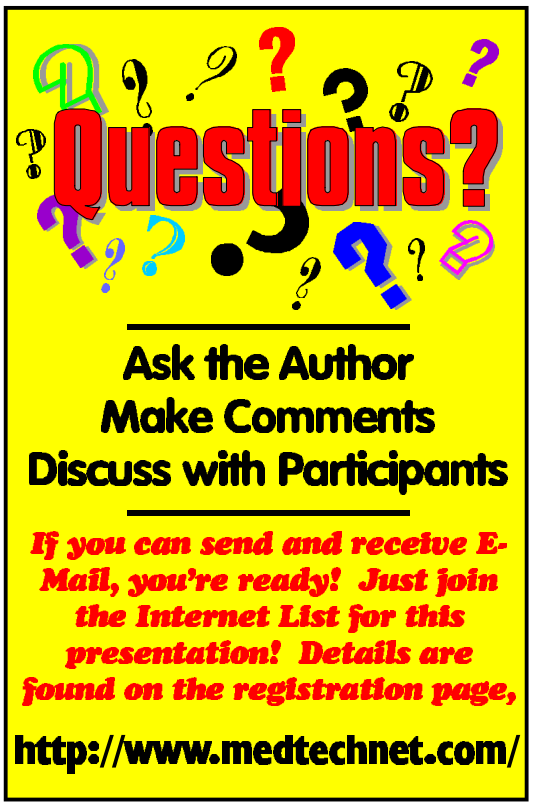
Performance standards (total error limits) have been established using a number of criteria. One approach is based on the intra-individual variation of an analyte. Fraser, has proposed that analytical imprecision should be equal to or less than one half the normal intra-individual variation.[3] Additionally, CLIA 88 regulations define fixed limit goals in absolute terms or multiples of standard deviations for a particular analyte. Furthermore, as published by Koch, D.D. and Peters, P., a method's total error should be less than one quarter the total error limit as defined under CLIA 88. [4] Reasons for this criteria are, first, the chance of proficiency testing failure approaches zero (assuming no inaccuracy in the method) and second, effective quality assurance can be maintained with less frequent use of quality control testing materials. This latter consideration allows users to lower quality control costs in the laboratory while still identifying significant changes in the imprecision of the method. If within run precision is not acceptable, correct the problem or abandon the method.

It should be readily apparent that within run precision estimates do not reflect an assay's imprecision in a realistic manner. A truer estimate of precision must take into account assay imprecision as a function of time as well as within run precision. Imprecision estimates which evaluate the effect of time can be achieved by determining ***Between Day Precision***. Furthermore ***Total Assay Precision*** estimates includes the contributions of both ***Within Run and Between Day Precision.***

$$Total\ Precision = \sqrt{sd^2_{Within\ Run} + sd^2_{Between\ Run}}$$

Total assay precision should be used to decide if the random error of the method meets the CLIA total fixed error limits. Note that the initial within run precision should first be used to decide if the method's reproducibility is acceptable given the CLIA allowable error limits or other error limits which users may establish.

The next section illustrates a concise statistical approach to the calculation of within and between as well as total precision estimates for a method. In this example a serum sample is split and cholesterol was determined twice within a run (beginning and end of the run) on each of a minimum of ten days. The simulated data and associated calculations are shown below.



Estimate of Assay Total Precision

Estimate of Assay Within Run Precision
Estimate of Assay Between Day Precision
Estimate of Assay Total Precision

Replicate Cholesterol Control Values (mg/dL)= x_1 and x_2

	x_1	x_2	Daily $\bar{x}$	$\bar{x}^2$	$\sigma^2 = (\lvert x_1 - x_2 \rvert / \sqrt{2})^2$
Day 1	201	200	200.5	40200.25	0.500
Day 2	204	204	204	41616	0
Day 3	210	205	207.5	43056.25	12.504
Day 4	201	206	203.5	41412.25	12.504
Day 5	200	206	203	41209	18.005
Day 6	198	201	199.5	39800.25	4.501
Day 7	204	208	206	42436	8.002
Day 8	206	199	202.5	41006.25	24.507
Day 9	201	198	199.5	39800	4.501
Day 10	208	206	207	42849	2.001
			$\sum \bar{x} = 2033$	$\sum \bar{x}^2 = 413385.25$	$\sum \sigma^2 = 87.025$

WITHIN RUN PRECISION (Standard deviation of Differences) = σ_w

$$\sigma_W = \sqrt{\frac{\Sigma\sigma^2}{n}} = \sqrt{\frac{87.025}{10}} = 2.95 \ mg/dL$$

BETWEEN DAY PRECISION (Corrected Standard deviation) = σ_B

$$\sigma_{\bar{x}}^2 = \frac{\left[\Sigma\bar{x}^2 - \frac{(\Sigma\bar{x})^2}{n}\right]}{n-1} = \frac{\left[413385.25 - \left(\frac{2033^2}{10}\right)\right]}{10-1} = 8.48\ mg/dL$$

$$\sigma_B = \sqrt{\sigma_{\bar{x}}^2 - \frac{\sigma_W^2}{2}} = \sqrt{8.48\ mg/dL - \frac{(2.95\ mg/dL)^2}{2}} = 2.03\ mg/dL$$

TOTAL PRECISION (ACCOUNTS FOR BOTH σ_W and σ_B) = σ_T

$$\sigma_T = \sqrt{\sigma_W^2 + \sigma_B^2} = \sqrt{(2.95\ mg/dL)^2 + (2.03\ mg/dL)^2} = 3.58\ mg/dL$$

Within run precision is determined as the standard deviation (σ_W) of the difference of the duplicates. To calculate σ_W, first the variance (σ^2) of the differences must be calculated.

$$Variance_W = \sigma^2 = \left(\frac{|x_1 - x_2|}{\sqrt{2}}\right)^2$$

$$Within\ Run\ sd_W = \sigma_W = \sqrt{\frac{\Sigma\sigma^2}{n}} = \sqrt{\frac{87.025}{10}} = 2.95\ mg/dL$$

Between day precision is determined as the standard deviation (σ_B) of the differences between the mean of all and the daily means and correcting for (subtracting) within run precision. To calculate σ_B, first the variance is determined.

$$Variance_B = \sigma_{\bar{x}}^2 = \frac{\Sigma\bar{x}^2 - \frac{(\Sigma\bar{x})^2}{n}}{n-1} = \frac{413385.25 - \frac{2033^2}{10}}{10-1} = 8.48\ mg/dL$$

$$Between\ Run\ sd_W = \sigma_B = \sqrt{\sigma_{\bar{x}}^2 - \frac{\sigma_W^2}{2}} = \sqrt{8.48\ mg/dL - \frac{(2.95\ mg/dL)^2}{2}} = 2.03\ mg/dL$$

Total precision of the method is estimated by combining both within run and between day precision. Total standard deviation (σ_T) is calculated as the square root of within run variance plus between run variance.[5]

$$Total\ Precision,\ sd_T = \sigma_T = \sqrt{\sigma_W^2 + \sigma_B^2} = \sqrt{(2.95\ mg/dL)^2 + (2.03\ mg/dL)^2} = 3.58\ mg/dL$$

The following Table contains examples of the criteria for acceptable performance (total fixed error limits) as published in the CLIA 88 Federal Register, Vol. 57, No. 40, February 28:7149-68,1992. The complete list of performance criteria can be found in the reference cited above. Note the performance criteria are expressed in several different ways; first, as an absolute value in the units of the analyte, second, as the target value ± multiples of the standard deviation for that analyte or the target value ± a percent (%) of the actual value. Each manner of expressing acceptable criteria is illustrated in the table below.

Is the cholesterol method **acceptable** in terms of **Random Error**? Listed above is the total error limit for serum cholesterol. The allowable error is the target value ± 10%. Using the example above, the serum cholesterol target value was 200 mg/dL. The total error is 200 mg/dL x 0.1 = 20 mg/dL and the acceptable total error limits is 180 to 220 mg/dL.

First, always determine if within run is acceptable. If the precision is not acceptable, the random error must be reduced or the method should be abandoned. If the within run precision is acceptable, determine if total precision is acceptable. The goal is to have

CLIA 88 CRITERIA FOR ACCEPTABLE PERFORMANCE (TOTAL FIXED ERROR LIMITS)

ANALYTE	ACCEPTABLE PERFORMANCE
Erythrocyte Count	Target Value ± 8%
Prothrombin Time	Target Value ± 15%
Cholesterol	Target Value ± 10%
Blood Alcohol	Target Value ± 25%
α-Fetoprotein (Tumor Marker)	Target Value ± 3 SD
Sodium	Target Value ± 4 mmol/L
Calcium, Total	Target Value ± 1 mg/dL
IgE	Target Value ± 3 SD
IgG	Target Value ± 25%

imprecision less than the total error limit or a much better criteria is ⅓ or ¼ total error limit greater than the method's sd:

Method sd < Total Fixed Error Goal
3 x Method sd < Total Fixed Error Goal
4 x Method sd < Total Fixed Error Goal

In the cholesterol data above the within run precision was 2.95 mg/dL at 200 mg/dL. Three times within run precision = 8.85 mg/dL (3 x 2.95 mg/dL), which is much less than the CLIA total error limit of 20 mg/dL at 200 mg/dL. Within run precision is acceptable. The total precision estimate was 3.58 mg/dL. Three times the total precision = 10.74 mg/dL. This value is again much less than the total fixed error limit of 20 mg/dL. The precision of this method is satisfactory to meet the CLIA goals and to assure the user that the probability of failing proficiency testing is very small, assuming no inaccuracy. [6]

Meeting CLIA Criteria

With assurance that the method's precision is acceptable, next the method must be evaluated to determine if accuracy meets CLIA criteria. Accuracy verification can be determined using calibration materials traceable to the National Institute of Standards and Technologies (NIST), assayed control material and proficiency testing materials (http://www.nist.gov/). All must be compatible with the method under investigation. [7]

A simple protocol for estimation of inaccuracy involves analyzing one of the materials referenced above, a minimum of three times to minimize random error and calculating the mean. It is essential that the concentration or activity of the material analyzed be **within the linear range** of the assay! A point estimate of the method's inaccuracy is the target value minus the mean analyzed value. This difference divided by the target value expressed as a percentage is the % deviation.

$$Inaccuracy\ as\ \%\ Deviation = \frac{|\bar{x}\ measured - Target\ Value|}{Target\ Value} X\ 100$$

Ehrmeyer and Laessig have published Accuracy Verification Tolerance Limits for many commonly performed analytes. [7] The total cholesterol accuracy verification tolerance limit is ± 4%. In the example above, the target concentration of the cholesterol control was 200 mg/dL. The method under evaluation had a mean for all days of 203.3 mg/dL. The deviation from the expected value of 200 mg/dL was 1.65%. The inaccuracy (1.65%) at 200 mg/dL is less than the ±4% accuracy tolerance limit, hence the conclusion is that the assay meets accuracy goals. Note, accuracy of an assay should be assessed at all important medical decision concentrations. If the inaccuracy of the assay exceeds the accuracy verification tolerance limit, the problem must be corrected. One of the first corrective actions to be taken is to recalibrate the instrument. Then, new standards should be employed in calibration.

Reportable Range

The **Reportable Range** of an assay is determined by **linearity studies**. Implicit in the reportable range is determining the **Least Detectable Dose (LDD)** and **Maximum Dose Limit (MDL).** Determination of the reportable range can be done using commercially available linearity materials with analyte values traceable to NIST standards, assayed quality control materials or clinical samples with both low and very high analyte concentrations

which have been accurately diluted and analyzed multiple times to obtain mean values.

The following example using cholesterol linearity materials illustrates the calculations of MDL. The assigned target cholesterol concentration of the linearity solution was 800 mg/dL. Summarized in the table is the dilution method, calculated target value, measured concentrations, mean concentration and percent deviation from the target value.

mL Soln/mL Diluent	Target Value (mg/dL)	Meas. Value #1	Meas. Value #2	Meas. Value #3	Meas. $\bar{x}$	% Dev.
0.1 mL/0.9 mL	80	81	80	80	80.34	+0.4
0.2 mL/0.8 mL	160	161	160	161	160.7	+0.4
0.4 mL/0.6 mL	320	318	319	318	318.3	-0.5
0.5 mL/0.5 mL	400	400	397	396	397.7	-0.6
0.6 mL/0.4 mL	480	476	477	473	475.3	-1.0
0.8 mL/0.2 mL	640	631	625	627	627.7	-1.9
1.0 mL/0 mL	800	727	732	741	733.3	-8.3

Using the maximum allowable inaccuracy of ±4% of the target value, it is apparent that the deviation is unacceptable (-8.3%) at a total cholesterol of 800 mg/dL. Hence the upper limit of linearity (MDL) is 640 mg/dL. The CLIA tolerances can be used to determine at what analyte concentration the deviation exceeds the total error limit, but please realize that this approach does not allow for random error! Investigators have suggested that in general the allowable deviation should not exceed 5% of the target value. In the example given above, the MDL would be the same using the ±4% or ±5%.

A graph of the relationship between the target value and the measured values should always be plotted for visual inspection to confirm the calculated deviations. **Figure 1** is a plot of the linearity data. The regression line of this plot was determined by regressing the five lowest target concentrations versus the mean measures values. Using the lower concentrations to determine the regression line allows for a visual inspection to determine the distance a value is from the line.

Least detectable dose (LDD) is in general defined as the measured response at zero dose of the analyte ±2 sd. [7] If the calibration curve has a positive slope, then LDD is the response at zero +2 sd and if the calibration curve has a negative slope, then LDD is the response at zero -2 sd. The concept of the minimum ±2 sd is acceptable for general clinical testing, but if the information is or could potentially be used in a medicolegal context (ex. urine drug screening) then the laboratory must insure that the probability of the false positive in extremely low. If the laboratory must guard against a false positive result, a more appropriate estimate of LDD might be the response at zero +3.5 or 4 sd. The following table summarizes the probability that a truly negative (zero concentration) specimen would exceed a

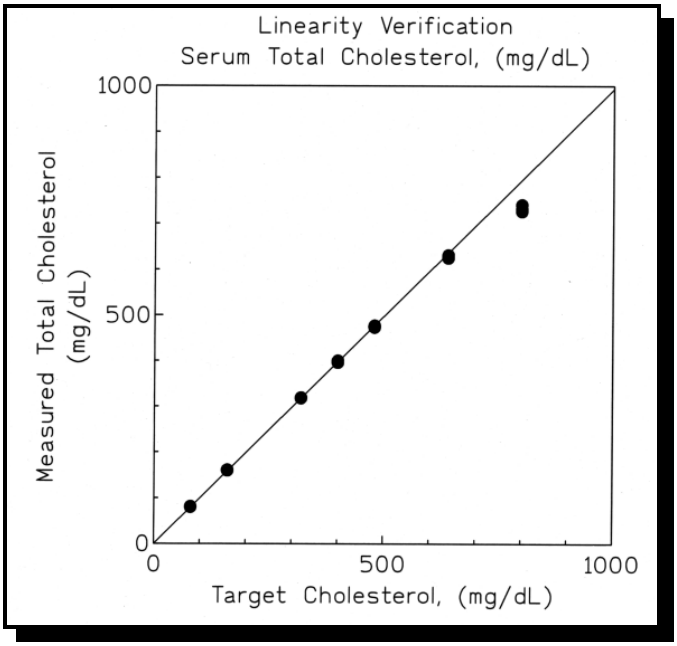


Figure 1

stated sd above the mean by random chance alone (false positive).

Probability of False Positive as Multiples of SD above Zero		
> Zero + 2 sd	2.28%	1 in 44 times
> Zero + 3 sd	0.13%	1 in 769 times
> Zero + 3.5 sd	0.02%	1 in 5000 times
> Zero + 4 sd	0.003%	1 in 33,330 times

The probability of a truly negative specimen being a false positive due to chance alone is 2.28% using the detection limit of zero +2 sd. Stating the concept another way, if a negative specimen is analyzed repeatedly, one in 44 times a result would equal 2 or more sd above the response at zero dose. For routine clinical analytes, the response at zero dose +2 sd is an acceptable criteria for LDD.

For elicit drug screening in the workplace or other medicolegal applications a 1 in 44 false positive results would be disastrous. A more appropriate criteria of LDD might be zero +3.5 sd. Using the zero plus 3.5 sd criteria of LDD, only 1 in 5,000 truly negative specimen would equal or exceed the limit and be a false positive result. Laboratories can set LDD by choosing the most appropriate multiple of the sd for their situation.

LDD is determined by analyzing a control or specimen with a very low analyte concentration or activity multiple times (20 times is ideal but no less than 10 times) and calculating the sd. The specimen selected for the replicate testing should have a matrix as similar to the actual samples as possible. For example, it would be inappropriate to analyze a serum based specimen, if the actual clinical samples were urine. Unfortunately it is many times difficult or impossible to find specimens with very low analyte concentrations, hence compromises are made. For example, total cholesterol aqueous calibrators which are viscosity adjusted to simulate sera may be the best available possibility. A delipidated serum pool would be better but more difficult to obtain.

As an example, a 25 mg/dL cholesterol calibrator was analyzed 20 times. The results are: 23, 22, 25, 28, 21, 24, 24, 25, 27, 28, 24, 23, 25, 25, 20, 28, 22, 27, 25 and 29. The sd of the replicates is 2.5 mg/dL and is an estimate of imprecision as a very low concentration. If the assayed concentration is zero at a dose of zero, then LDD = Response at Zero Dose + (2 x sd).

LDD = 0 mg/dL + (2 x 2.5 mg/dL) = 5.0 mg/dL

The **analytical range** of the cholesterol method extends from **5 to 680 mg/dL**. The linearity has been assessed at low, normal and high physiological concentrations to establish if accuracy is acceptable at these levels. In most cases the analytical range for a method may be used as the **Reportable Range**. One cautionary note is that if the matrix (eg, viscosity and/or ionic strength) of the linearity materials or the material used for LDD are different than the clinical specimen, the reportable range and analytical range may be somewhat different. Many instrument manufacturers publish or incorporate in the equipment algorithms for reportable ranges. The manufacturers' reportable ranges should still be confirmed using a evaluation protocol similar to the one outlined in this presentation. It is always good laboratory practice to establish the performance of the method in your laboratory and never accept without conformation the claims of manufacturers or other laboratories.

Recovery Studies

Recovery Studies are used to establish what portion (percent) of the analyte in the specimen is being measured. Ideally, the exact

amount of analyte in the specimen should be detected by the method and recovery would is 100%.

If the recovery is different from 100%, **proportional error** exists. Proportional error at important medical decision concentrations may compromise the usefulness of the method.

To conduct a recovery study, an appropriate specimen is divided in two. To a known volume of the aliquot, a known volume of a high analyte concentration or activity is added (spiked specimen). Next, to a equal volume of the aliquot, a known volume of an appropriate diluent is added (baseline specimen). It is important to keep the ratio of diluent to specimen volumes as small as possible. Ideally, the volume of analyte or diluent should be no greater than 10% of total volume to minimize dilutional (matrix) effects relative to neat clinical specimen.

EXAMPLE: Recovery Experiment

An easy way to prepare the spiked aliquot is to use the highest linearity material available or a patient specimen with a high concentration. For this example, the high linearity material had a cholesterol concentration of 800 mg/dL. The specimen aliquot was a quality control (QC) material with a normal cholesterol level.

Preparation of Baseline and Spiked Aliquot

Spiked Aliquot = 1.9 *mL QC sera* + 0.1 *mL cholesterol*, 800 *mg/dL*

Baseline Aliquot = 1.9 *mL QC sera* + 0.1 *mL of diluent* (*saline*)

Next the two aliquots are analyzed in duplicate or, better still, in triplicate and the mean values calculated.

The concentration recovered is the spiked aliquot concentration minus the baseline aliquot concentration.

$$\text{Mean Analyte Recovered} = \bar{x}\ \text{Spiked}_{conc} - \bar{x}\ \text{Baseline}_{conc}$$

$$\text{Mean Cholesterol Recovered} = 228\ mg/dL - 189\ mg/dL = 39\ mg/dL$$

The concentration of the analyte added is calculated as the concentration of that added, times the volume added divided by the total volume of the aliquot.

$$\text{Concentration Analyte Added} = \text{High Analyte Conc.} \times \frac{\text{vol. Analyte}}{\text{total volume}}$$

$$\text{Cholesterol Added} = 800\ mg/dL \times \frac{0.1\ mL}{2.0\ mL} = 40\ mg/dL$$

The mean recovery for the method is calculated as the concentration recovered times 100 divided by the concentration added.

$$\%\ \text{Recovery} = \frac{\text{Concentration Recovered}}{\text{Concentration Added}} \times 100$$

$$\%\ \text{Cholesterol Recovery} = \frac{39\ mg/dL}{40\ mg/dL} \times 100 = 97.5\%$$

In the example above the recovery was 97.5%. Is that magnitude of proportional error acceptable? To determine acceptability of the proportional error (PE), a medical decision concentration must be stated. Let us evaluate proportional error at 200 mg/dL, the cutoff between normal risk and moderately elevated risk of development of cardiovascular complications. Using the CLIA total error limit of ± 10% of the target value, at 200 mg/dL, ± 20 mg/dL in acceptable. The point estimate of proportional error is mean recovery (as a fraction) times the medical decision concentration subtracted from the medical decision concentration. [8]

$$\text{Proportional Error} = \left|(\text{Recovery} \times \text{Decision}_{Conc}) - \text{Decision}_{Conc}\right|$$

$$\text{Cholesterol Method PE} = \left|(0.975 \times 200\ mg/dL) - 200\ mg/dL\right| = 5\ mg/dL$$

A proportional error of 5 mg/dL is clearly much less than the CLIA total error limit of 20 mg/dL at 200 mg/dL, hence the conclusion is that recovery is acceptable.

Interference Studies

Interference Studies are performed to assess the effect of compounds which potentially alter test results in either a negative or positive manner. Commonly evaluated interfering compounds include: hemoglobin (hemolysis), lipids (lipemia), bilirubin (icterus), anticoagulants and preservatives, vitamins (dietary supplementation), analgesics (aspirin and acetaminophen), drugs (ex. lithium) and drug metabolites. Interference studies are carried out in much the same manner as are recovery studies. In interference studies a known amount of the interfering compound is added to a specimen to produce a sample which contains a high concentration of the interferant. In general, drugs should be tested at 5 to 10 times the upper limit of the therapeutic range. Preservatives and anticoagulants should be evaluated at twice the normal additive specimen concentration, thus simulating a "short draw" in phlebotomy. An excellent source of information about interfering compounds and recommended evaluation procedures can be found the NCCLS EP7-P guidelines. [9]

In immunoassays, in addition to the potential interferences listed above, users need to provide information about the specificity (cross-reactivity) of the antibody with closely related antigens or haptens. Most manufacturers of immunoassays provide fairly extensive assay specificity information, but because of a unique patient population, laboratories may need to evaluate potential cross-reactivity of a new drug.

Provided below are protocols for the preparation of the interferants: lipids, hemoglobin and bilirubin.

Lipid Solution, 1,000 mg/dL
A simulated lipemic serum specimen can be easily produced by obtaining an I.V. fat emulsion solution. One such preparation is Intralipid, 20% (w/v), KabiVitrum Inc, Clayton, NC 27520. This solution contains 20 g of soybean oil per 100 mL. The addition of 0.5 mL of Intralipid and 9.5 mL of water (reagent type I) to a 10 mL lyophilized normal control will produce a serum based material with an added triglyceride concentration of 1,000 mg/dL. This preparation will simulate a fairly lipemic serum specimen, but a greater amount of Intralipid can be added to produce a higher added triglyceride concentration.

Hemoglobin Solution, 5000 mg/L
Hemolysis in a serum or plasma specimen can be prepared by the addition of free hemoglobin. To prepare a solution of free hemoglobin collect a tube of heparinized blood. Centrifuge, 1000 xg for 10 minutes, the blood as you normally would to obtain plasma. Decant the plasma and add isotonic (sodium chloride, 0.156 mol/L) saline at a volume approximately equal to twice the packed RBC volume. Gently re-suspend the RBCs, centrifuge, 1000 xg for 10 minutes, again and decant the supernatant. Repeat the saline wash three more times as indicated above. After the final centrifugation and decanting of the supernatant, add reagent type I water of a volume equal to the volume of the RBCs and mix. The RBCs can be lysed by mechanical disruption (ex. tissue grinder) or refrigeration overnight (hyposmotic manipulation). The solution is next centrifuged at 2000 xg for 30 minutes to remove the stroma.

Analysis of the free hemoglobin can be done using the Drabkin's cyanomethemoglobin method or a direct spectral method. Several

direct spectral methods for the determination of free hemoglobin concentration are discussed by Fairbanks, Virgil F. et al. [10] One method described in the publication requires that the plasma be diluted 11-fold with a sodium carbonate solution, 0.942 mol/L, prior to determining the absorbance at 380 nm, 415 nm and 450 nm against the sodium carbonate solution. Since the hemolysate is too concentrated a further 100-fold dilution is required.

$$Free\ Hemoglobin,\ g/L = \{0.836[2Abs_{415} - (Abs_{380} + Abs_{450})]\}\ x\ 100$$

To make the appropriate dilution to yield a final hemoglobin concentration of 5000 mg/L, apply the following equation.

$$Vol\ Hb\ Soln = \left[\frac{Desired\ Hb.(mg/L)}{Hemolysate(mg/L)}\right] X\ Tot.Vol.$$

EXAMPLE: Interferant Preparation

If 10 mL of a 5,000 mg/L hemoglobin solution is required and the hemolysate hemoglobin is 70,000 mg/L, what volume of the hemolysate must be diluted to a final volume of 10 mL with reagent grade water?

$$Vol\ Hb\ Soln\ (mL) = \left[\frac{5{,}000\ mg/L}{70{,}000\ mg/L}\right] X\ 10\ mL$$
$$= 0.714\ mL$$

Ten mL of a 5,000 mg/L hemoglobin solution is prepared by diluting 0.714 mL of hemolysate to the final volume. Ten mL of the 5,000 mg/L hemoglobin can be added to reconstitute a lyophilized 10 mL normal control.

Bilirubin Solution, 20 mg/dL

Commercially available bilirubin solutions of approximately 20 mg/dL can be purchased and used to reconstitute normal control materials. This approach is simple and easy, but expensive since relatively large volumes are required to reconstitute a control vial. A second approach is to prepare a 20 mg/dL bilirubin solution from crystalline bilirubin. Highly purified crystalline bilirubin can be purchased inexpensively from many chemical vendors.

Weight out 20 mg of crystalline bilirubin on a plastic weighing boat and transfer to a 100 mL volumetric flash. Dissolve the bilirubin by adding 1.0 mL of dimethylsulfoxide (DMSO). Note, DMSO should also be used to wash the weighing boat of any residual bilirubin. Completely dissolve the bilirubin by swirling and next add 2.0 mL of sodium carbonate, 0.1 mol/L and add approximately 70 mL of reagent grade water. Adjust the solution to a final pH of 7.4 by drop-wise addition of 0.1 mol/L hydrochloric acid while monitoring with a pH electrode and meter. Finally dilute to 100 mL with reagent water. This solution has a bilirubin concentration of 20 mg/dL. [11]

Using a lyophilized 10 mL normal control, reconstitute the control vial by the addition of 10 mL of the bilirubin solution, 20 mg/dL. This control material should be stored in the dark to prevent light mediated degradation.

Finally, a normal control of the same lot as used to prepare the interference solutions should be reconstituted with 10 mL of reagent water. The mean values of this control analyzed in triplicate will serve as the analyte target concentrations or activities. The normal control materials containing the added lipid, hemoglobin and bilirubin should also be analyzed in triplicate and the mean values determined. A point estimate of the effect of

each interferant on the assay can be determined.

$$\%\ Interference = \frac{Conc_{(With\ Interferent)} - Conc_{(Target)}}{Conc_{(Target)}} X\ 100$$

Using this equation to calculate percent interference, both positive and negative types of interference in the assay can be quantified at known levels of the interfering substance.

EXAMPLE: Calculation of Interference

The mean total cholesterol in the target control was 199 mg/dL and the mean total cholesterol in the control with 5,000 mg/L hemoglobin added was 206 mg/dL.

$$\%\ Interference\ (at\ 5{,}000\ mg/L\ Hb) = \frac{206\ mg/dL - 199\ mg/dL}{199\ mg/dL} X\ 100 = 3.5\%$$

In this case, hemoglobin at 5,000 mg/L falsely increased the target total cholesterol concentration by 7 mg/dL or 3.5%. This level of interference is acceptable given the CLIA guidelines of ± 10% but just acceptable at a total fix error of 4%. Calculation of the percent interference due to the other interfering substances are determined in the same manner.

Evaluation of immunoassay cross-reactivity can be evaluated in several different ways. One method, employed in competitive immunoassays, replaces calibrators with the potential interfering compound as assay calibrators. The calibration curves are constructed and analyzed to determine the concentrations at which a 50% displacement of the labelled antigen or hapten from the antibody occurs. This method accurately determines cross-reactivity but is more complicated to carry-out. The following protocol is simple and in most cases, provides adequate estimates of antibody cross-reactivity.

Evaluation of Immunoassay Cross-reactivity

Obtain:

A. A sample negative for the analyte (ie. drug or hormone)
B. Potential interferant of known concentration

Prepare Solutions:

A. 0.1 mL Diluent + 0.9 mL Sample (zero conc.), mix
B. 0.1 mL interferant + 0.9 mL of Sample (zero conc.), mix

Analyze both solutions in triplicate, calculate $\bar{x}$

$$\%\ Cross\text{-}reactivity = \frac{\bar{x}\ Conc.\ Soln.\ B - \bar{x}\ Conc.\ Soln.\ A}{Concentration\ of\ Interferent\ Added} X\ 100$$

Example:

$$Interferant\ Concentration = 1000 \mu g/dL$$

$$Added\ interferant\ to\ soln\ B = 1000 \mu g/dL\ X\ \frac{0.1 mL}{1.0 mL} = 100 \mu g/dL$$

Results:

$$Solution\ A = 0.5 \mu g/dL$$
$$Solution\ B = 6.0 \mu g/dL$$

$$\%\ Cross\text{-}reactivity = \frac{6.0 \mu g/dL - 0.5 \mu g/dL}{100 \mu g/dL} = 5.5\%$$

Comparison Studies: Final Phase

The final phase of a methods comparison study is to analyze actual clinical samples

which have the broadest possible analyte concentration range. The team should be collecting these specimens in advance of the final phase of the evaluation. Ideally 120 specimens should be analyzed, but many evaluations are done with fewer samples. A random sampling of 128 abstracts published for the 1997 annual meeting of Clinical Chemistry revealed that the mean number of clinical specimen analyzed in methods comparison studies was 101.8 and the median was 63.[12] Larger numbers of samples in the comparison study provide more confidence in the parameters which describe the relationship between the methods.

Unique specimens which contain potential interferences should also be included in the study. It is imperative that the range of concentrations or activities in the sample be broad. Several investigators suggest that adequacy of range can be evaluated with the correlation coefficient (r). Cary, R.N. et at, have suggested that $r \geq 0.99$.[8] **Note, that the correlation coefficient (r) is used to establish the adequacy of analyte range and not to reach conclusions about the performance of the methods.** Some investigators have used r as a primary descriptive parameter in evaluating relationships between methods. A better and more quantitative parameter to evaluate random error between methods is the standard deviation of the residuals about the regression line which is also called the standard error of the estimate (S_{yx}). The calculation and use of S_{yx} are discussed in a following section.

Three types of error can be observed between methods. They are **random error, constant error and proportional error**. One or more of these errors may be observed between methods. By evaluation of the types and magnitude of the errors, decisions can be made about the acceptability of clinical methods.

Random Error

Random error is due to imprecision of measurement of the methods and is randomly distributed about the regression line. The magnitude of random error can be quantitatively estimated by the standard deviation of the residuals, (S_{yx}). S_{yx} is calculated as the square root of the sum of squares of y at the line minus the actual y value divided by n-2.

$$S_{yx} = \sqrt{\frac{\sum (y_i - Y_i)^2}{n - 2}}$$

where: y_i = actual y value
Y_i = y at the line

Most all statistical software packages will calculate S_{yx}, but some spreadsheet programs will not directly calculate S_{yx} under the regression option. Hence an estimate of S_{yx} can be obtained by requesting residuals, standard residuals (residuals in sd multiples) and a residual plot. From the residuals, the standard deviation of the residuals can be obtained by using the descriptive statistic function. A good estimate of S_{yx} can be obtained in this manner when the number of samples exceeds 30. In the random error data presented in **Figure 2** with 38 samples in the comparison, the correct calculation of S_{yx} was 14.4 and the spreadsheet calculation of S_{yx} was 14.5, an error of less than 1%. The reason for the error is that the spreadsheet uses n-1 in the sd calculation rather than n-2 as required in the correct S_{yx} calculation. For comparisons using smaller numbers of sample, the error in estimation of S_{yx} using n-1 in the calculation becomes larger and the equation above should be used.

Excel 5 will calculate S_{yx} by using the **STEYX** command in function wizard. The format would

be **STEYX({y-values},{x-values})** and will return S_{yx}. An example,is:

STEYX({5,7,9,11,13,15,17,12},{6,8,11,12,13,18,15,14}) = s_{yx} = 1.629.

Also Excel 5 calculates the **standard error of the regression line** (S_{yx}) when regression analysis is performed.

When graphing methods comparison data, it is customary to plot the reference or existing method on the x-axis and to plot the method under evaluation on the y-axis. Also it is convenient to graph both axes using the same scale, in this manner the plots are square and a line from the lower left to top right corner represents perfect agreement between methods. These conventions make is easy to evaluate the types and magnitude of error which may occur.

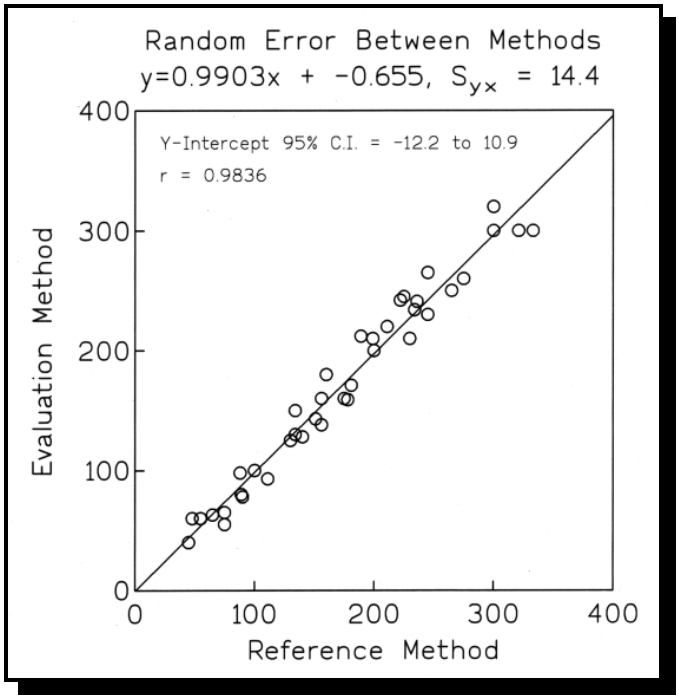


Figure 2

Figure 2, illustrates appreciable random error between the two methods. Note the scatter of the points about the line, S_{yx} equals 14.4. Units of S_{yx} are the units of analyte concentration or activity. **Figure 3**, illustrates a reduction in random error, S_{yx} equals 7.79. The S_{yx} of Figure 3 is approximately half that observed in **Figure 2**. Visual comparison of the two figures reveals that **Figure 3** has data points clustered more closely about the line. As random error decreases S_{yx} decreases. The slope in both **Figures 2 and 3** is close to 1.000 and the y-intercept is close to zero.

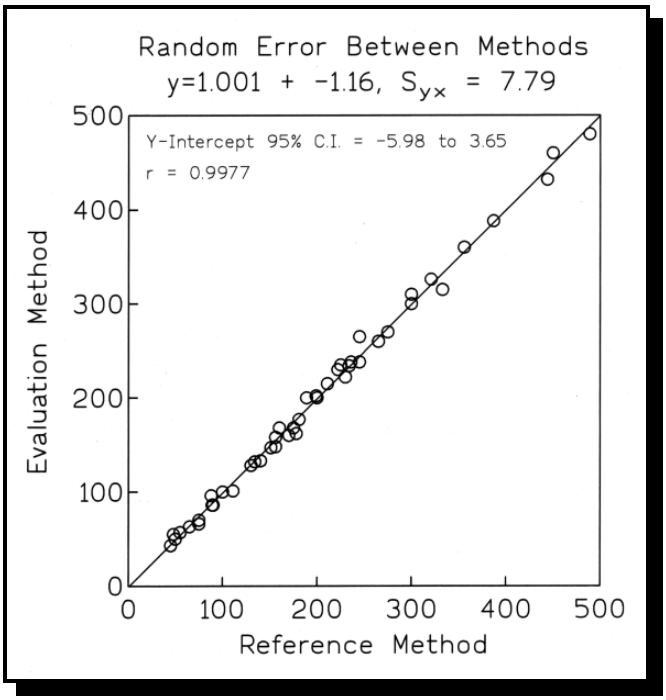


Figure 3

Figure 4, illustrates the effect of concentration range on the correlation coefficient. In **Figure 2**, r = 0.9836, the range of x-values was 48 to 333 with an S_{yx} of 14.4. An r of less than 0.99 indicates the range is not broad enough to insure adequate estimates of relationships between methods. In **Figure 4**, r is now greater than 0.99 (r= 0.993) with a similar S_{yx} of 15.0. The main difference between **Figure 2 and 4** is that the range of sample analyte concentrations has been extended to 489 from the previous maximum of 333. Extending the range increased the correlation coefficient (r) when the random error (S_{yx}) remained essentially the same in the two comparisons.

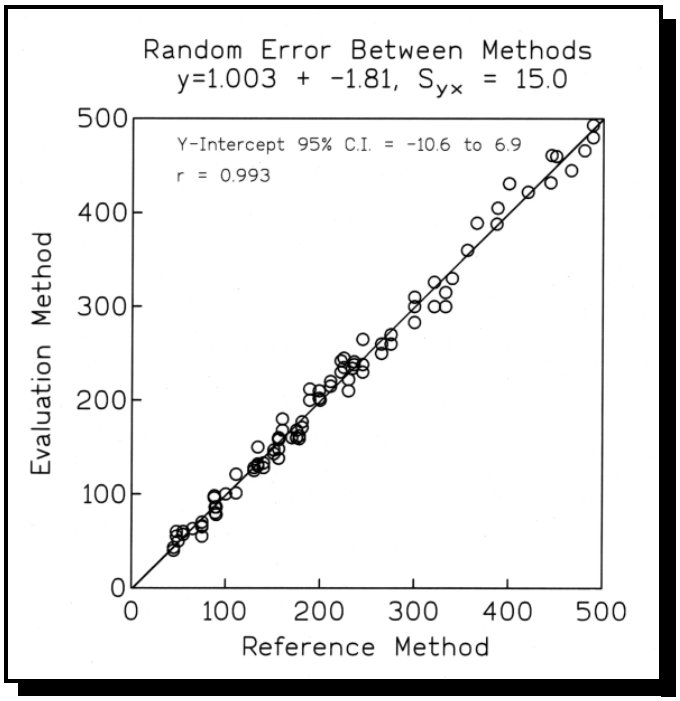


Figure 4

An important point to remember about least squares regression analysis is that the model assumes that the x-axis values (independent variables) are the true concentration and assigns all error of estimation to the y-axis values (dependent variables). In reality both x and y methods have some degree of imprecision and a better regression model would partition the imprecision of measurement between both the x and y variables. The Deming's regression model recognizes this problem of imprecision in analytical methods and gives a better regressional relationship when both methods have appreciable random error. [13]

Recently observed comparison data from two serum total iron binding capacity methods revealed a slope of approximately 0.91 and a sizable bias by least squares linear regression. Deming's regression yielded a slope of approximately 0.99 and insignificant bias. Total iron binding capacity is one of the least precise tests done in clinical chemistry and random error should be attributed to both methods. Deming's regression was the more appropriate model is this case. If the imprecision of the x method is small then only minor differences will occur between the two regression models.

Most commercial spreadsheets only perform least squares regression, so if Deming's regression is required, special software will need to be purchased or the data sent to a company for data analysis. I would recommend that both least squares and Deming's regression software be available for use in the laboratory.

Identification of outlier data points about the regression line should begin with the a visual inspection of plotted data. **Figure 5**, reveals by visual inspection one observation (indicated by the arrow) is suspiciously different from all other points about the line.

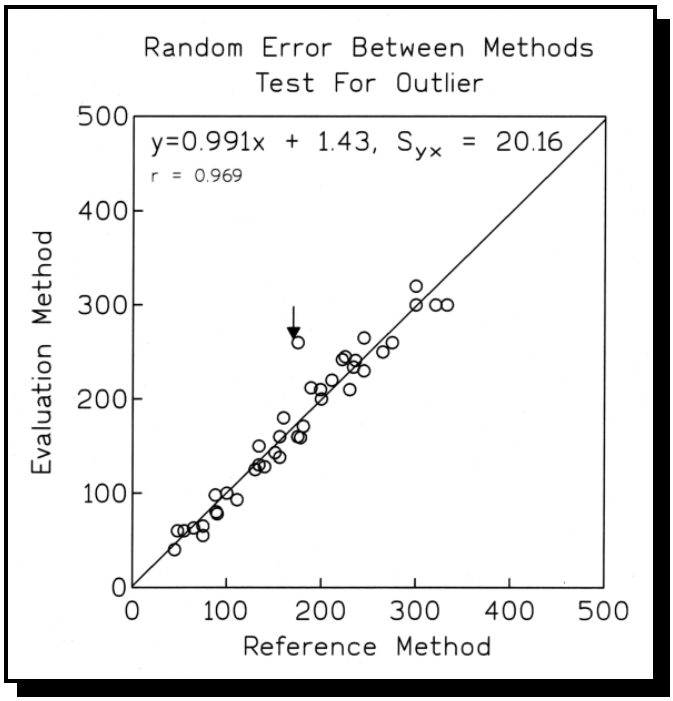


Figure 5

Is this observation spurious for some reason and what criteria should be applied to accept of reject the data point as part of the population? A good criteria for rejecting the

data is when the point is greater than 3.5 time S_{yx} form the regression line. The data point in question is 175 by the reference method equals and 260 by the evaluation method (175, 260). Solving for what the evaluation method should have given (value at the regression line) when the reference method was 175, yielded 174.86.

$$y = mX_c + b = (0.991)(175) + 1.43 = 174.86$$

where: m = slope
X_c = x value
b = y-intercept

The actual value (260) minus the value at the line (174.86) equals 85.14. This y-axis distance from the line (85.14) divided by S_{yx} (20.16) equals 4.22 times S_{yx} from the line. The data point in question is 4.22 times S_{yx} from the regression line and clearly exceeds 3.5 S_{yx}. The point can be rejected as an outlier.

Figure 5 contained the same data as in **Figure 2**, except for the inclusion of the outlier. This one spurious point increases S_{yx} by 5.76 units (40%) from 14.4 to 20.16. After an outlier has been rejected it is **essential** that the regression analysis be repeated to establish the corrected equation of the line and associated parameters.

Proportional Error

Proportional error produces results by one method which are some multiple (percentage) of results by the second method. Proportional error is evaluated by assessing the slope of the regression line. **If the slope at some confidence interval is different from 1.00, then significant proportional error exists between methods**. In proportional error, S_{yx} increases with the magnitude of the error but the y-intercept will not be significantly different from zero.

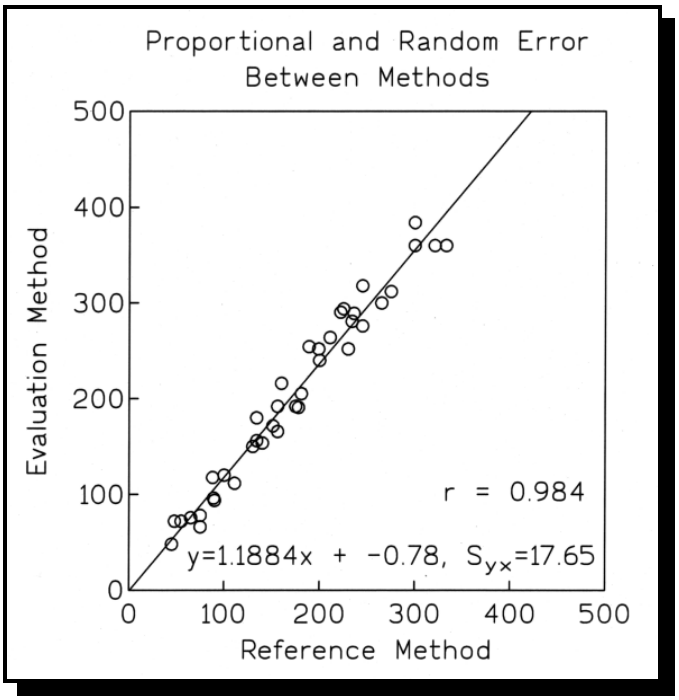


Figure 6

Figure 6, illustrates proportional error. The slope is 1.188, which implies that the average value for the evaluation method (y-axis) was 1.188 times that of the reference method (x-axis). The 95% confidence interval of the slope is 1.11 to 1.26 and does not include 1.00, hence we conclude that proportional error does exist. S_{yx} is very large at 17.65 also indicating proportional error. The y-intercept is not significantly different from zero (y = -0.78). The reasons for proportional error are first that one method is not completely measuring all the specimen analyte. Recovery studies discussed earlier in this paper will help confirm this possibility. Recovery in one method will be significantly less than 100%. A second possible cause of proportional error is calibration error of one or both methods. Recalibration of one or both methods may rectify the problem of proportional error.

The bias ($\bar{x}$ - $\bar{y}$) is the difference between methods at the means. The bias should be less than the CLIA total error limit. In this example, bias equals 203 - 172 = 31. The

CLIA total error limit is 10% of the target value. The mean x was 172 and 10% of this value is 17.2. Clearly proportional error at the mean is unacceptable (31 is greater than the allowable error of 17.2).

Systematic error may be described as the inaccuracy of a method when compared to a method with established accuracy. Bias is one measure of systematic error. In addition to the inaccuracy at the mean, it is important to calculate the systematic error at medical decision concentrations. For total cholesterol, laboratorians would like to be able to assure users that systematic error is acceptable at 200 mg/dL, the borderline between normal and moderately increased cardiovascular risk. **Systematic error (SE) is the difference at the medical decision concentration between the reference method and the evaluation method obtained by regression analysis.**[8]

$$\textbf{Systematic Error (SE)} = (mX_c + b) - X_c$$

Where:

m = slope
X_c = Medical decision concentration
b = y-intercept

Using **Figure 6** at an example, at a total cholesterol of 200 mg/dL, the systematic error is 36.9 mg/dL.

$$SE = [(1.1884)(200\ mg/dL) + (-0.78)] - 200\ mg/dL = 36.9\ mg/dL$$

The CLIA total error limit at 200 mg/dL is 20 mg/dL. Clearly the systematic error of 36.9 mg/dL is greater than the recommended total error limit (20 mg/dL) and the performance of the method is unacceptable.

Constant Error

A third type of analytical error which can exist between methods is **Constant Error**.

Constant error is caused by interferences in the analytical samples. These interferences will produce a constant difference regardless of analyte concentration between the evaluation and reference methods. The nature of the interference may be either positive of negative with respect to the reference value. Using regression analysis, **Constant error is manifest by a y-intercept which is significantly different from zero.**

Figure 7, illustrates constant error between the reference and evaluation method. The y-intercept is 19.3 and the 95% confidence interval of the intercept is 7.8 to 30.9. The confidence interval does not include zero. The 95% confidence intervals of the slope and y-intercept are shown in Figure 7 as the dashed lines. Note the 95% confidence intervals do

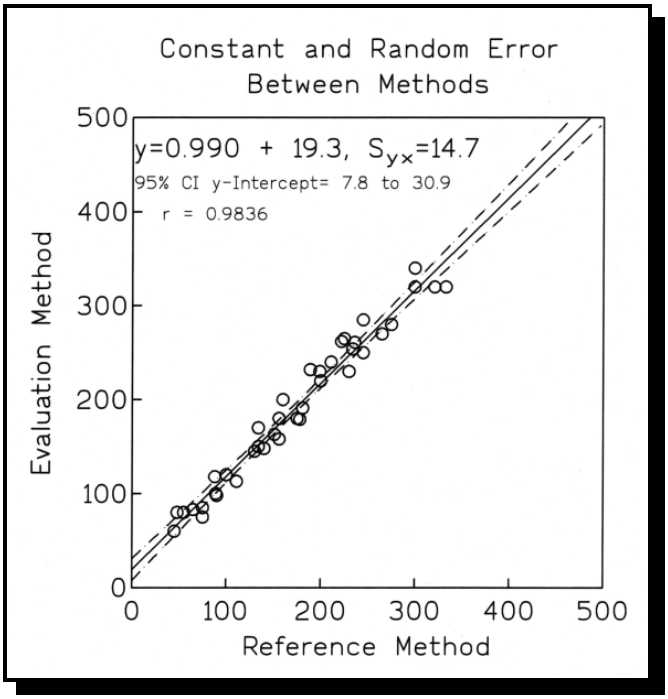


Figure 7

not include zero at the y-intercept. Evaluators are fairly assured that this magnitude of off-set in the y-intercept would not happen by chance alone. The cause of the constant error may not readily be apparent by regression analysis, but interference studies discussed earlier in the presentation will be helpful in the elucidation. Systematic error is attributable constant error in **Figure 7**, at a medical decision concentration of 200 mg/dL total cholesterol is 17.3 mg/dL. Total allowable error is 20 mg/dL at 200 mg/dL, so the systematic error is acceptable, but inaccuracy consumes nearly all of allowable error!

The final criteria of acceptable method performance is total error. **Total error is a measure of systematic plus random error and should be less than the CLIA total error limit**.

$$\text{Total Analytic Error} = \text{Systematic Error} + 3 \times sd_{total}$$
$$\text{Total Analytic Error} = \text{Systematic Error} + 4 \times sd_{total}$$

The total analytic error criteria of SE plus 4 x sd_{total} is the ideal measure of a method's performance. If total analytic error, calculated by this equation is less than CLIA total error limit, then the chance of proficiency failure is remote. Total analytic error calculated as SE plus 3 x sd_{total} is still a robust measure of performance.[8]

Final Performance goal!!

Total Analytic Error < CLIA Total Error Limit

A final question needs to be asked: are the patient sample means by the two methods the same? If the means are the same statistically, then the existing reference intervals can be safely used for the new method. If the means are different, then new reference intervals may need to be established for a defined population. Judgement of the medical significance of mean differences would need to be considered. The statistical test to be employed is the **Paired-t test** of sample means. Using the data from **Figure 2**, where the analytical error was primarily random, the two-tail probability (p=0.322) is 32 times out of 100 the differences in means would occur by chance alone. The Excel printout of the paired-t is shown in **Table 2**.

TABLE 2
Paired-t Test for Data in Figure 2

	Reference Method	Evaluation Method
Mean	173.4871795	171.1282051
Variance	6383.940621	6473.167341
Observations	39	39
Pearson Correlation	0.9835696	
Hypothesized Mean Difference	0	
df	38	
t Stat	0.987825363	
P(T≤t) one-tail	0.16474331	
t Critical one-tail	1.685953066	
P(T≤t) two-tail	0.322	
t Critical two-tail	2.024394234	

A generally accepted significant difference in means occurs when $p \leq 0.05$. Hence the conclusion is that there is no significant difference in the means of the two methods and that the same reference intervals could be applied for the new method.

Table 3 contains the results of the paired-t test for the data used in Figure 6. The methods

comparison revealed proportional error with a slope of 1.1884. The two-tailed probability ($p=1.62^{-10}$) is much less than p=0.05, hence the null hypothesis is rejected. The two means are not the same and a new reference interval would most probably need to be established.

TABLE 3
Paired-t Test for Data in Figure 6

	Reference Method	Evaluation Method
Mean	173.4872	203.0526316
Variance	6482.874111	9353.240398
Observations	38	38
Pearson Correlation	0.984350976	
Hypothesized Mean Difference	0	
df	37	
t Stat	-8.166339186	
P(T≤t) one-tail	1.62E-10	
t Critical one-tail	1.687094482	
P(T≤t) two-tail	1.61E-10	
t Critical two-tail	2.026190487	

Summary

Underlying this paper, has been the goal, as stated in the Joint Commission Comprehensive Manual of Performance Measurement, performance assessment and performance improvement.[2] Covered in this paper are many objective techniques to measure performance in terms of precision, accuracy, interference, reportable range and others. The CLIA 88 total fixed error limits now provide defined performance standards against which analytical methods can be assessed. Meeting or exceeding acceptable performance standards for a method in large measure assures laboratorians that results will satisfy medical care needs and pass proficiency testing in a consistent manner.

Finally, performance improvement should be an important part of method evaluation and selection by the laboratory. Precision and accuracy can be objectively compared with existing procedures or other methods under consideration. The ultimate goal being that total analytical error should be reduced with each succeeding methodology. Methods evaluation should be integrated in to routine laboratory operations as a mechanism to promote continual quality improvement.

Appendix 1: Descriptive Statistics using Excel

Data is entered on an Excel spreadsheet. Raw Data is in the first column. Row 1 (cell A1) contains the label for the column below (see **Figure 8**).

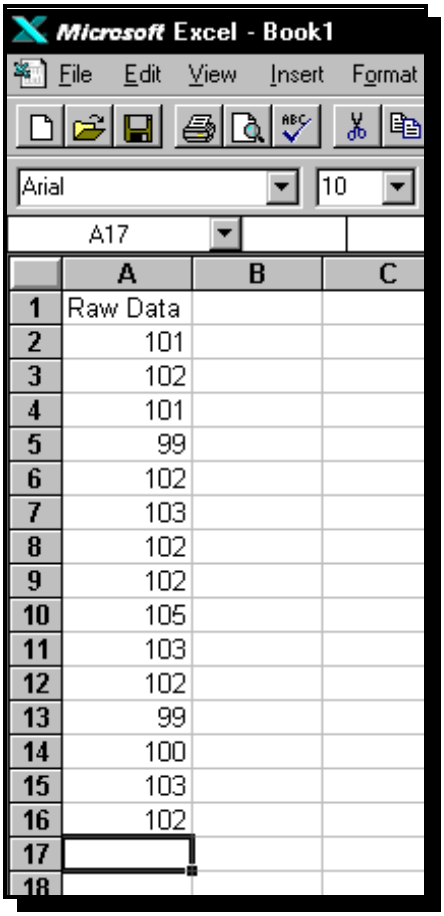


Figure 8: Entering data

(Excel 7 was used for this example; however, the steps are similar in Excel 5).

Next, drag the mouse across the data (include label cell), to select the block of data; click on **Tools** from the menu bar, then select **Data Analysis** from the drop down menu (see **Figure 9**). (If Data Analysis does not show up on this menu, check your Add-Ins. The module must be installed before using.)

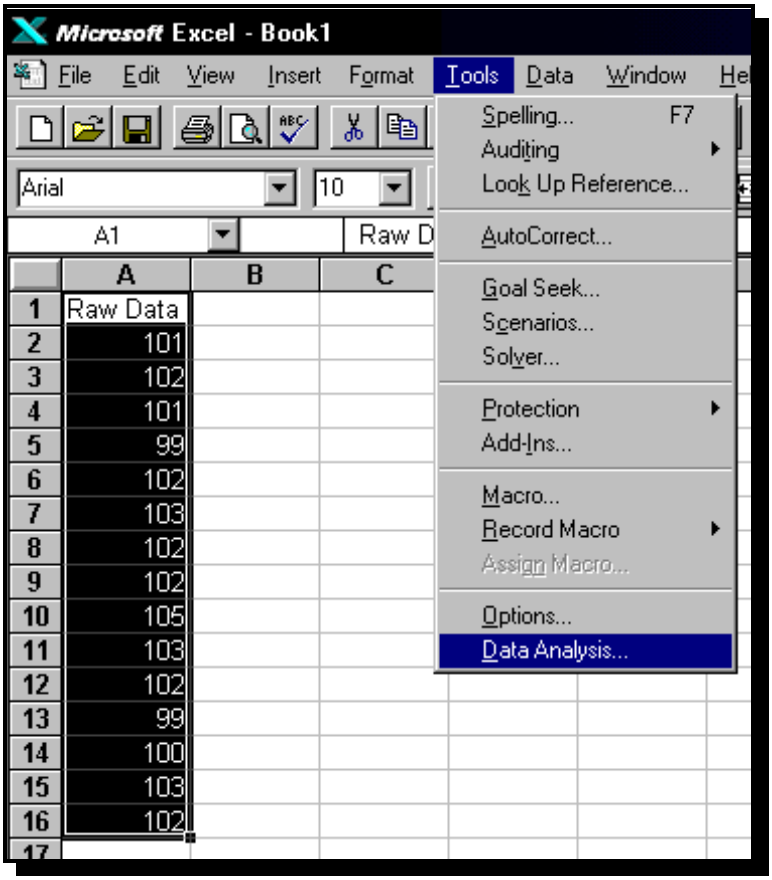


Figure 9: Select data and run Data Analysis

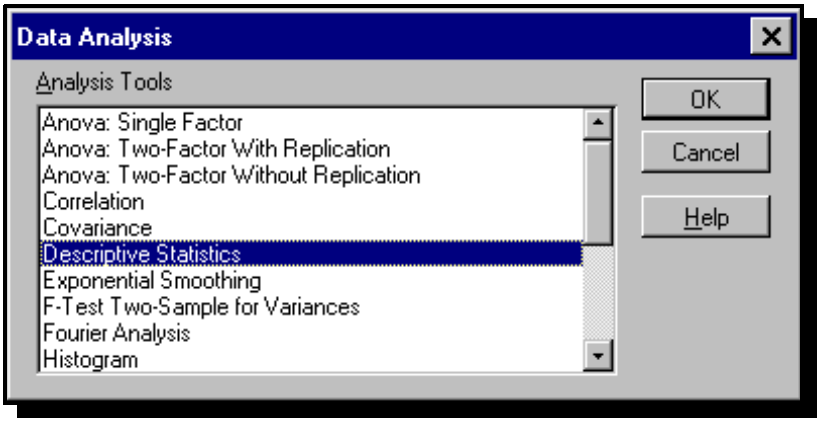


Figure 10: Selecting a data analysis package

The next dialog box (**Figure 10**) lists all the data analysis packages available. To obtain basic statistical parameters, click on **Descriptive Statistics**, then the OK button.

The next dialog box (**Figure 11**) allows the user to specify the Input Range (in this case, A1:A16); the "Labels in First Row" is checked, since the first row of our selection contains the column labels. The output (calculations) will be placed in the cells beginning with B1 (ie, B1 is the upper, left-hand cell of the output block). Also note that "Summary Statistics" is checked to provide additional statistical information.

Clicking the OK button performs the calculations and places the results in the designated area of the spreadsheet. To view all labels and numbers completely, click **Format >> Column >> AutoFit Selecton** (**Figure 12**).

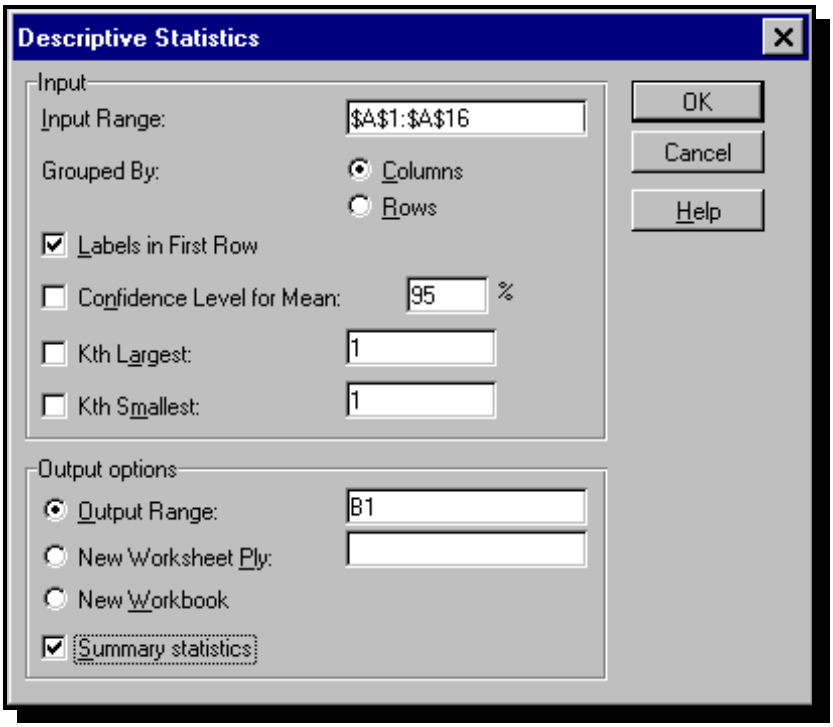


Figure 11

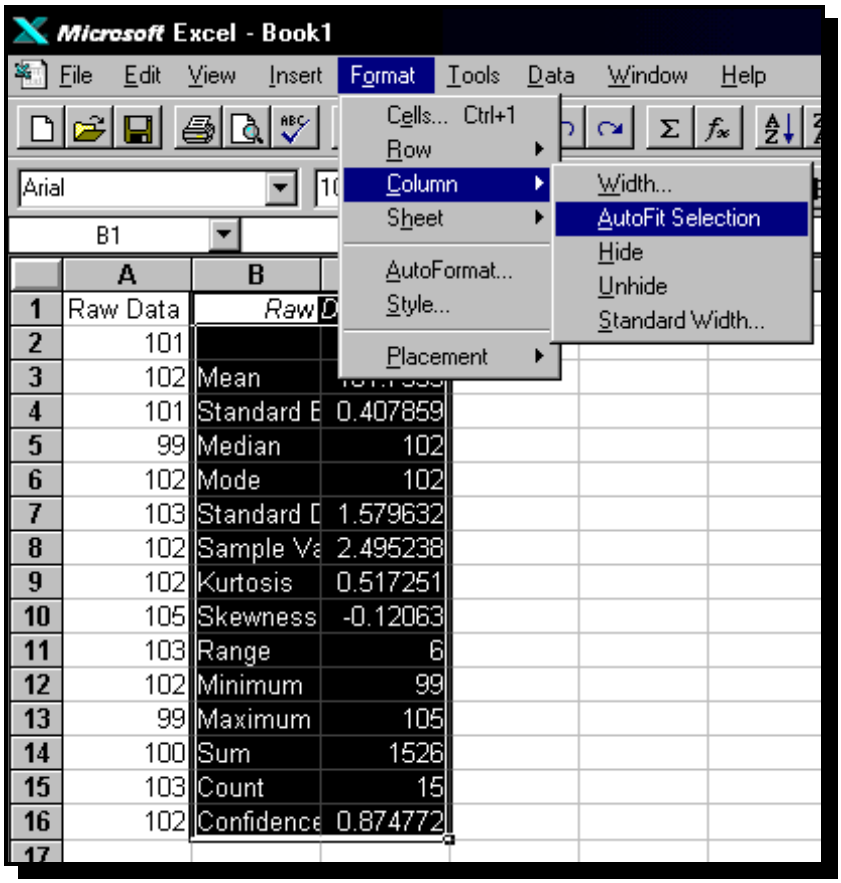
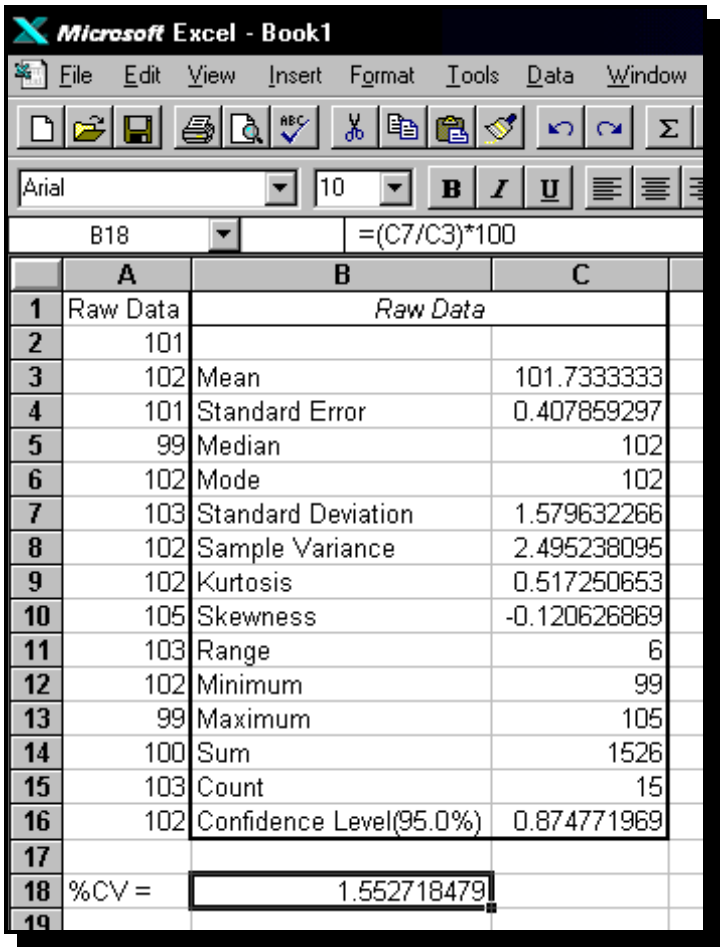


Figure 12

Finally, the %CV value is easily calculated by entering the appropriate formula, **(C7/C3)*100**, in an unused cell; in this case, B18 (see **Figure 13**).

As noted in the presentation, clinical laboratory method comparisons are often better evaluated using Demings regression analysis, which takes into account the measurement error in **both** methods. As a special bonus, this Med TechNet presentation is being distributed with an Excel spreadsheet that will perform the Demings calculations. Be sure to download the file, **MTCon19.xls**, available with this the presentation file (**mtc19pdf.pdf**).

Additional instructions for using the template are found on the **Instructions** worksheet. MTCon19.xls will work with Microsoft's Excel version 5 and higher.



	A	B	C
1	Raw Data	*Raw Data*	
2	101		
3	102	Mean	101.7333333
4	101	Standard Error	0.407859297
5	99	Median	102
6	102	Mode	102
7	103	Standard Deviation	1.579632266
8	102	Sample Variance	2.495238095
9	102	Kurtosis	0.517250653
10	105	Skewness	-0.120626869
11	103	Range	6
12	102	Minimum	99
13	99	Maximum	105
14	100	Sum	1526
15	103	Count	15
16	102	Confidence Level(95.0%)	0.874771969
17			
18	%CV =	1.552718479	
19			

Figure 13

Bibliography

1. Clinical Laboratory Improvement Act; Federal Register, Feb 28, 1992, vol. 57 (40) : 7149-7184 and Federal Register, Jan 19, 1993, vol. 58(11): 5230-5232.

2. Comprehensive Accreditation Manual for Pathology and Clinical Services, 1996. Joint Commission on Accreditation of Health Care Organization, Oakbrook Terrace, Illinois

3. Fraser, CG. Analytical Goals are Applicable to All. J Int Fed Clin Chem. 1990; 2(2): 84-86.

4. Koch, DD and Peters , Jr, T. Selection and Evaluation of Methods: With an Introduction of Statistical Techniques. In Tietz, Fundamentals of Clinical Chemistry, 4th Ed.,1996: 170-181. Editors Burtis, CA and Ashwood, ER, W.B. Saunders Co., Philadelphia

5. Strike, PW Statistical Methods in Laboratory Medicine. 1991, Butterworth-Heinemann Ltd, Oxford, 254-306.

6. Ehrmeyer, SS, Laessig, RH, Leinweber JE and Oryall, JJ. 1990 Medicare/CLIA Final Rules of Proficiency Testing: Minimum Interlaboratory Performance Characteristics (CV and Bias) Needed to Pass. Clin Chem. 1990; 36(10):1736-1740

7. CLIA Watch, Guide to Implementing CLIA Regulations on the Laboratory: Calibration, Calibration Verification and Other Verification Procedures. Sigma Chemical Company, March, 1993.

8. Cary, RN, Garber, CC and Koch, DD. Concepts and Practices in the Evaluation of Laboratory Methods. Workshop #308, Am Assoc Clin Chem, 1993, New York: 1-97.

9. National Committee for Clinical Laboratory Standards, EP7 Interference Testing. 6(13):259-366.

10. Fairbands, VF, Ziesmer, SC, O'Brien, PC Methods for Measuring Plasma Hemoglobin in Micromolar Concentration Compared. Clin Chem. 1992; 38(1):132-140

11. Balistreri, WF and Shaw, LM. Clinical Manifestations of Liver Disease. pg 1378-1404. In Textbook of Clinical Chemistry. Edited by Tietz, Nw., 1996, W.B. Saunders Co., Philadelphia,

12. Abstracts, Scientific Poster 49th Annual Meeting, AACC. Clin Chem. 1997; 43(S6):S1-S306.

13. Cornbleet, PJ and Gochman, N. Incorrect Least-Squares Regression Coefficients in Method-Comparison Analysis. Clin Chem. 1979; 25(3):432-438.

Med TechNet Online Presentations

The online discussion for this presentation can be found in a dedicated Message Area, on the Med TechNet BBS; or, may be accessed by subscribing to the Internet list for this presentation (see below).

Reaching the Med TechNet BBS

The Med TechNet BBS is a menu driven online service with many special and generalized conferences, as well as a complete online software library with thousands of titles, for DOS, Windows 3.1x, Windows 95/NT and OS/2. Users can explore the BBS by telnet'ing to Med TechNet on the Internet at, bbs.medtechnet.com, or dialing in directly, via modem (V.34), at, 1-716-688-1552. (For more information on "telnet", see, http://www.medtechnet.com/telnet.html.)

Online Discussions

Discussions for Med TechNet presentations are available with the author(s) and other presentation participants. Either telnet or dial-in to the BBS and go to the appropriate discussion area in the MESSAGE AREAS. Or, if you prefer, participate via Internet E-Mail. Complete instructions for joining the Internet list for this presentation are available to subscribers who register online at the Med TechNet web site.

Earning CEU's

Earning CEU's for this presentation requires a passing score on the post-quiz. Successful completion of the online, electronically-scored post-quiz earns the participant 4.0 P.A.C.E.® Contact Hours or California State accredited credits.

Registration and quizzes are administered at the Med TechNet web site (http://www.medtechnet.com/). You must be a Med TechNet subscriber to register for CEU's. You must register for a presentation to access the post-quiz.

This is a presentation of Med TechNet Online Services for the Clinical Laboratory. "Talks", such as this one, may be studied at any time within a two year period following the initial posting.

For more information, visit the Med TechNet Web site,
http://www.medtechnet.com/, or
call toll-free, 1-800-836-0720 (M-F, 9a-4p Eastern).

Med TechNet is a service of Western New York Microcomputer, Inc.
PO Box 84
East Amherst, NY 14051, USA