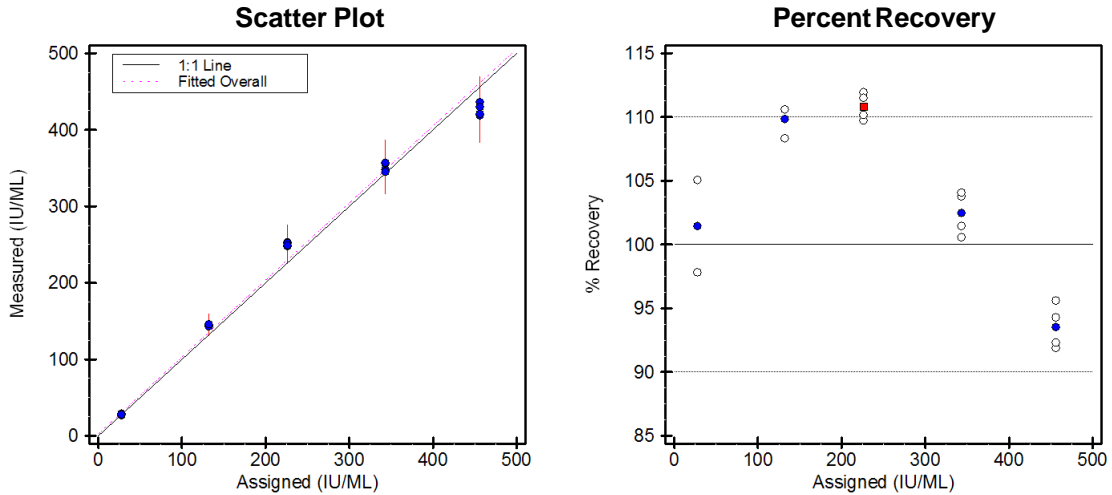


Accuracy and Linearity



Linearity Summary

	N	Slope	Intercept	Error
Overall	5	1.006	2.5	8.2%

LINEAR within Allowable Systematic Error of 10.0%

Statistical Analysis

	Assigned Pct	Mean	Percent Recovery	Accuracy	Reportable Range	Linearity
S01	27.6	28.0	101.4	Pass	-	Pass
S02	132	145.0	109.8	Pass	-	Pass
S03	226	250.5	110.8	Fail	--	Pass
S04	343	351.5	102.5	Pass	-	Pass
S05	456	426.5	93.5	Pass	-	Pass

See User's Specifications on the next page for Pass/Fail criteria

Evaluation of Results

The Accuracy and Linearity of RF were analyzed on Excimer 250 over a measured range of 28.0 to 426.5 IU/ML. Reportable Range was not verified. This analysis assumes accurate assigned values. Allowable systematic error (SEa) was 10.0%. The accuracy test FAILED. The maximum deviation for a mean recovery from 100% was 10.8%. 4 of 5 mean recoveries were accurate within the SEa. 20 of 20 results were accurate within the allowable total error (TEa) of 20.0%. The results are LINEAR.

Accepted by: _____
Signature

Date

Accuracy and Linearity

Experimental Results

S01	28	29	27	28
S02	145	143	146	146
S03	253	252	248	249
S04	348	356	345	357
S05	436	419	421	430

x: Excluded from calculations

User's Specifications

Allowable Total Error: 20.0%
Systematic Error Budget: 50%
Allowable Systematic Error: 10.0%

Supporting Data

Analyst: J Jones
Date: 29 Nov 2000
Units: IU/ML
Reportable Range: -
Value Mode: Preassigned
Controls: -
Reagent: -
Calibrators: -
Comment: Linearity Set

Accepted by: _____

Signature

Date

Linearity Report Interpretation Guide

For EP Evaluator® purposes, Linearity experiments are experiments that use specimens with defined concentrations. This includes Calibration Verification, Accuracy, and Reportable Range, as well as Linearity. The Linearity module can also verify Precision. This means you can verify three of the four CLIA '88 requirements with a single experiment.

User-selectable options determine which of these parameters the report verifies. Also, the user may request Pass/Fail flags against a specific allowable error criterion, or he/she may simply report selected statistical measures.

Experiment Procedure: Replicate measurements are made on 3-11 specimens, with (known) concentrations spread across the reportable range. Ideally, the lowest and highest specimens should challenge the limits of the range.

Accuracy (or Recovery)

Definition: The ability to recover the correct amount of analyte present in the specimen.

Verification process: Accuracy can be verified only when the "correct" amount of analyte (the **Assigned Value**) is known. While it is possible to determine recovery using a single replicate, one gets a more reliable estimate when 2 to 4 replicates are assayed.

Key statistic: $\text{Recovery} = 100 \times \text{Measured Mean} / \text{Assigned Value}$

Reportable Range

Definition: As used in CLIA, this term refers to the Analytical Range or Assay Range -- the maximum range of values that can be assayed accurately without dilution. The CAP term "Analytical Measurement Range" (AMR) is a synonym for Reportable Range.

Verification Procedure: Reportable Range is verified if two conditions are met: 1) the assigned values of the lowest and highest specimens are within proximity limits of the Reportable Range limits, and 2) these two specimens are acceptably accurate.

Proximity Limits define how close the lowest and highest specimens must be to the Reportable Range limits.

Calibration Verification

verifies whether a method is properly calibrated. Calibration Verification is identical to verifying both Accuracy and Reportable Range. The only difference is that the report is titled to match the regulatory requirement.

CLIA requires a minimum of three specimens, each assayed in duplicate. Two specimens challenge the lower and upper limits of the reportable range. The third specimen is somewhere in between.

Linearity

Several definitions are in common use. Among them:

- **Traditional Linearity** (CAP Visual Inspection): Draw a scatter plot with assigned values on the X-axis and measured mean on the Y-axis. If it looks like a straight line, the method is linear.
- **Statistical Linearity** (CLSI EP6P and EP6-A): These procedures determine acceptability based on statistical significance (i.e., p-values) rather than medical significance. EP Evaluator® does not compute Statistical Linearity.
- **Clinical Linearity:** The method is linear if it is possible to draw a straight line that passes within a user-defined allowable error of each specimen point.

Related concepts:

- **Best Fit Line:** If the user opts to verify Linearity, this line is obtained using the Clinical Linearity algorithm. Otherwise it is a regular linear regression line.
- **Outliers:** When verifying Linearity, the program first tries to determine an acceptable line using all specimens. If it fails, it then tries to find some subset of at least three specimens that are linear within allowable error. Specimens not in this acceptable subset are classified as outliers.
- **Slope and Intercept:** Coefficients of the Best Fit Line. The ideal slope is 1.00; the ideal intercept is zero.
- **Observed Error:** For Clinical Linearity, the minimum allowable error that could be defined for a data set and still have it be linear.
- **Standard Error of Estimate:** For regular regression, measures dispersion of the data points around the Best Fit Line.
- **Residual:** The difference between the best fit line and either an individual result or a mean measured value, depending on context.

Precision

Definition: Ability to obtain the same result upon repeated measurement of a specimen.

Verification Process: Measure the specimen many times. Compute the SD and CV, and verify that they are acceptably small. While 2-4 replicates are adequate for assessing accuracy, a minimum of 10 (and preferably 20 or more) is required to verify Precision.

The **Precision Index** is the ratio of SD to Allowable Random Error (defined below). The ideal -- and probably unattainable -- Precision Index is zero. A value of 1.00 indicates

Linearity Report Interpretation Guide

borderline acceptability. Any further increase in SD would exceed allowable error.

The **95% Confidence Interval** (CI) for the Precision Index indicates how much sampling variation might be expected. The CI narrows as the number of replicates increases.

Allowable Total Error (TEa), and the Error Budget

TEa states the laboratory's policy for how much error is medically (or administratively) acceptable. Regulatory requirements represent an upper limit. Example: the CLIA limit for Sodium is 4 mmol/L.

Total Error has two major components: **Systematic Error** (synonym Bias) and **Random Error** (synonym Imprecision). The **Error Budget** allocates a fraction of the Allowable Total Error for Systematic Error, and another fraction for Random Error. Establishing an appropriate Error Budget allows the lab to control accuracy and precision separately, with reasonable confidence that Total Error will also remain in control. Recommended ranges of values are 25-50% for the Systematic Error Budget and 16-25% for the Random Error Budget.

Pass or Fail?

The program reports Pass/Fail for Accuracy and Linearity based on Allowable Systematic Error (SEa). Pass/Fail for Precision is based on Allowable Random Error (REa).

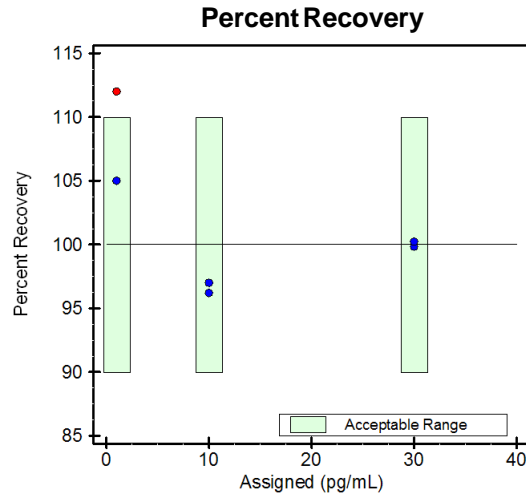
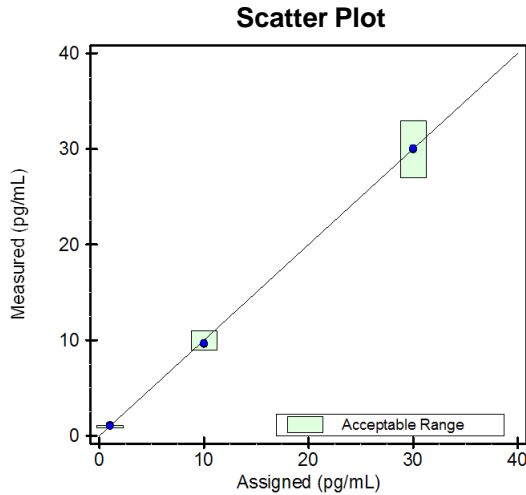
- A specimen passes Accuracy if its mean measured value is within SEa of the Assigned Value.
- The experiment passes Linearity if it is possible to draw a straight line (on the scatter plot of mean measured value vs. assigned value) that passes within +/- SEa of each specimen point.
- A specimen passes Precision if SD does not exceed REa.
- The experiment passes Reportable Range if 1) the assigned values of the lowest and highest specimens are within proximity limits of the Reportable Range Limits, and 2) these two specimens also pass accuracy.

Preliminary Report

The word PRELIMINARY printed diagonally across the report indicates that the data is incomplete, and the report is not acceptable as a final report. Some or all of the statistics may be missing.

The Linearity report is preliminary if there are less than three specimens.

SIMPLE ACCURACY



Statistical Analysis and Experimental Results

	Target		Measured Values		(1) Accuracy
	Range	Midpoint			
FT3-CalS1	0.9 to 1.1	1.00	1.12	1.05	Fail
FT3-CalS4	9 to 11	10.00	9.62	9.7	Pass
FT3-CalS5	27 to 33	30.00	29.95	30.07	Pass

(1) Accuracy passes if all measured values lie within the Target Range. 'x' indicates an excluded result.

Supporting Data

Analyst: mkf
 Date: 21 Oct 2009
 Units: pg/mL
 Reportable Range: -
 Controls: rew 333 exp 01 Dec 2010
 Reagent: -
 Calibrators: -
 Comment:

Evaluation of Results

The Accuracy of Free T3 was analyzed on 601540 over a range of 1.00 to 30.00 pg/mL. Reportable Range was not verified. The accuracy test FAILED. All replicate measurements were within the target range for 2 of 3 specimens. Overall, 5 of 6 replicates were within their target ranges.

Accepted by: _____
 Signature

 Date

Simple Accuracy Report Interpretation Guide

The Simple Accuracy experiment tests whether measurements fall within a defined Target Range. The experiment requires at least two specimens, assayed in duplicate. Ideally, the lowest and highest specimens should challenge the limits of the Reportable Range. Verifying Reportable Range in conjunction with the Simple Accuracy experiment is optional, but recommended.

Accuracy (or Recovery)

Definition: The ability to recover the correct amount of analyte present in the specimen.

Verification process: Accuracy is verified when all replicate measurements at each level lie within the Target Range at that level.

Key statistic: Recovery = $100 \times \text{Measured Value} / \text{Midpoint of Target Range}$

Reportable Range

Definition: As used in CLIA, this term refers to the Analytical Range or Assay Range -- the maximum range of values that can be assayed accurately without dilution. The CAP term "Analytical Measurement Range" (AMR) is a synonym for Reportable Range.

Verification Procedure: Reportable Range is verified if two conditions are met: 1) the Target Range midpoints for the lowest and highest specimens are within proximity limits of the Reportable Range limits, and 2) these two specimens are acceptably accurate.

Proximity Limits define how close the lowest and highest specimens must be to the Reportable Range limits.

Data Requirements

The experiment accepts up to 11 specimens, and up to 5 replicates per specimen. The minimum is 2 specimens, assayed in duplicate. A Target Range is required for each specimen. On the Scatter and Percent Recovery plots, the plotted X value is the midpoint of the target range.

Pass or Fail?

- A specimen passes Accuracy if all measured results fall within the target range.
- The experiment passes Reportable Range if 1) the midpoints of the target ranges for the lowest and highest specimens are within proximity limits of the Reportable Range Limits, and 2) these two specimens also pass accuracy.

Preliminary Report

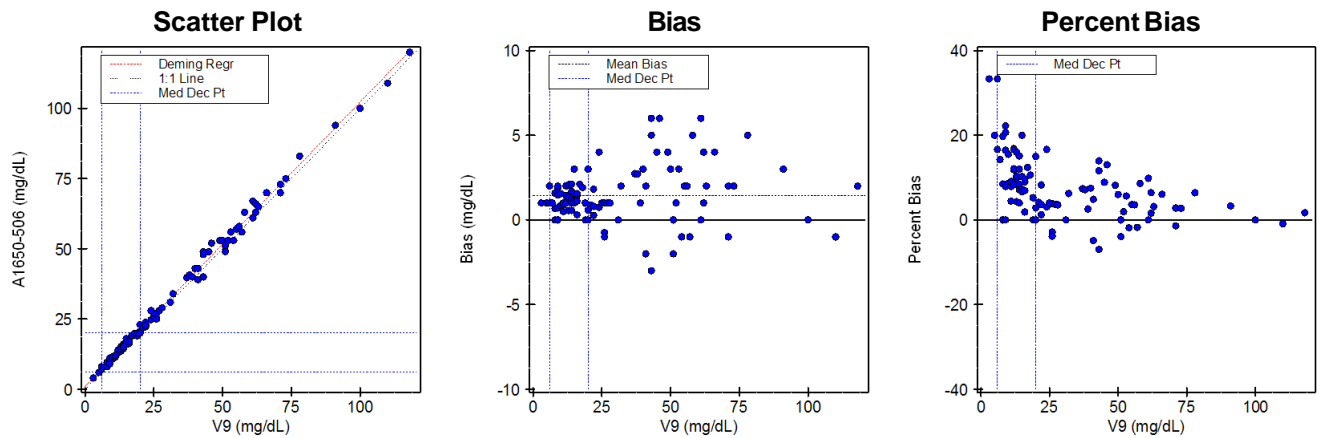
The word PRELIMINARY printed diagonally across the report indicates that the data is incomplete, and the report is not acceptable as a final report. Some or all of the statistics may be missing.

This report is preliminary if there are fewer than two specimens, or fewer than two replicates per specimen.

Alternate (Quantitative) Method Comparison

X Method: V9

Y Method: A1650-506



Regression Analysis

	Deming	Passing-Bablok	Regular
Slope:	1.013 (1.000 to 1.025)	1.011 (1.000 to 1.028)	1.010 (0.998 to 1.023)
Intercept:	1.051 (0.577 to 1.525)	0.938 (0.666 to 1.045)	1.113 (0.639 to 1.587)
Std Err Est:	1.575	--	1.574
SMAD:	1.037	0.991	1.037

95% Confidence Intervals are shown in parentheses

Medical Decision Point Analysis

Calculated by Deming Regression (R>=0.9)

X Method MDP	Y Method Pred. MDP	95% Conf. Limits	
		Low	High
6	7.1	6.7	7.5
20	21.3	21.0	21.6

Supporting Statistics

Corr Coef (R):	0.9980	Y Mean ± SD:	31.223 ± 24.642	Points (Plotted/Total):	109/110
Bias:	1.425	Std Dev Diffs:	1.588	Outliers:	1
X Mean ± SD:	29.798 ± 24.338	SubRange Bounds:	None	Scatter Plot Bounds:	None

Experiment Description

	X Method	Y Method
Expt Date:	01 Apr 2003	01 Apr 2003
Rep SD:	1	1
Result Ranges:	3.00 to 118.00	4.00 to 120.00
Units:	mg/dL	mg/dL
Reagent	ABC R1234 exp 31 Dec 2009	ABC R1234 exp 31 Dec 2009
Calibrators	-	-
Analyst:	mkf	mkf
Comment:		

Accepted by: _____

Signature

Date

Alternate (Quantitative) Method Comparison

X Method: V9

Y Method: A1650-506

Experimental Results

Spec ID		Results			Calc'd	SEE	Spec ID		Results			Calc'd	SEE	
		X	Y	Bias	Y	Factor			X	Y	Bias	Y	Factor	
1	1	22	23.82	1.82	23.327	0.3	51	45	50	53	3	51.678	0.8	
2	10	21	21.88	0.88	22.315	-0.3	52	46	39	40	1	40.540	-0.3	
3	100	49	53	4	50.666	1.5	53	47	61	67	6	62.816	2.7	
4	1009	78	83	5	80.029	1.9	54	48	51	51	0	52.691	-1.1	
5	1022	55	57	2	56.741	0.2	55	49	19	19	0	20.290	-0.8	
6	1025	O	8.5	90	81.5	9.668	51.0	56	5	37	39.73	2.73	38.515	0.8
7	1057	58	63	5	59.779	2.0	57	50	53	56	3	54.716	0.8	
8	11	15	16.09	1.09	16.239	-0.1	58	51	15	16	1	16.239	-0.2	
9	1102	63	65	2	64.841	0.1	59	52	7	8	1	8.139	-0.1	
10	1104	61	61	0	62.816	-1.2	60	53	31	31	0	32.440	-0.9	
11	1132	62	63	1	63.829	-0.5	61	54	51	49	-2	52.691	-2.3	
12	1133	110	109	-1	112.431	-2.2	62	55	57	56	-1	58.766	-1.8	
13	1164	71	73	2	72.942	0.0	63	56	12	13	1	13.202	-0.1	
14	1165	56	58	2	57.754	0.2	64	57	54	53	-1	55.728	-1.7	
15	12	9	9.71	0.71	10.164	-0.3	65	58	100	100	0	102.305	-1.5	
16	13	18	19.91	1.91	19.277	0.4	66	59	71	70	-1	72.942	-1.9	
17	14	13	13.56	0.56	14.214	-0.4	67	6	14	14.56	0.56	15.227	-0.4	
18	15	8	8.68	0.68	9.152	-0.3	68	60	43	40	-3	44.590	-2.9	
19	16	11	11.96	0.96	12.189	-0.1	69	61	26	27	1	27.377	-0.2	
20	17	10	10.86	0.86	11.177	-0.2	70	62	3	4	1	4.089	-0.1	
21	18	13	15.08	2.08	14.214	0.5	71	63	14	15	1	15.227	-0.1	
22	19	10	11.55	1.55	11.177	0.2	72	64	8	8	0	9.152	-0.7	
23	2	20	20.57	0.57	21.302	-0.5	73	65	5	6	1	6.114	-0.1	
24	20	24	24.74	0.74	25.352	-0.4	74	66	11	12	1	12.189	-0.1	
25	21	17	19.11	2.11	18.264	0.5	75	67	46	52	6	47.628	2.8	
26	22	16	16.3	0.3	17.252	-0.6	76	68	118	120	2	120.531	-0.3	
27	23	14	15.18	1.18	15.227	0.0	77	69	39	40	1	40.540	-0.3	
28	24	14	16.11	2.11	15.227	0.6	78	7	16	17.55	1.55	17.252	0.2	
29	25	9	10.48	1.48	10.164	0.2	79	70	6	8	2	7.127	0.6	
30	26	12	13.42	1.42	13.202	0.1	80	71	6	7	1	7.127	-0.1	
31	27	11	11.87	0.87	12.189	-0.2	81	72	24	28	4	25.352	1.7	
32	28	15	16.54	1.54	16.239	0.2	82	73	20	20	0	21.302	-0.8	
33	29	12	14.02	2.02	13.202	0.5	83	74	20	23	3	21.302	1.1	
34	3	14	15.68	1.68	15.227	0.3	84	75	8	8	0	9.152	-0.7	
35	30	8	9.58	1.58	9.152	0.3	85	76	19	20	1	20.290	-0.2	
36	31	16	17.43	1.43	17.252	0.1	86	77	5	6	1	6.114	-0.1	
37	32	13	14.28	1.28	14.214	0.0	87	78	9	9	0	10.164	-0.7	
38	33	19	19.99	0.99	20.290	-0.2	88	79	9	11	2	10.164	0.5	
39	34	22	22.8	0.8	23.327	-0.3	89	8	9	10.86	1.86	10.164	0.4	
40	35	11	11.49	0.49	12.189	-0.4	90	80	31	31	0	32.440	-0.9	
41	36	13	14.34	1.34	14.214	0.1	91	81	9	11	2	10.164	0.5	
42	37	12	13.45	1.45	13.202	0.2	92	82	9	9	0	10.164	-0.7	
43	38	22	22.27	0.27	23.327	-0.7	93	83	91	94	3	93.192	0.5	
44	39	26	25.25	-0.75	27.377	-1.4	94	84	6	8	2	7.127	0.6	
45	4	38	40.71	2.71	39.528	0.8	95	85	25	26	1	26.365	-0.2	
46	40	22	22.8	0.8	23.327	-0.3	96	86	32	34	2	33.453	0.3	
47	41	66	70	4	67.879	1.3	97	87	12	14	2	13.202	0.5	
48	42	41	39	-2	42.565	-2.3	98	88	45	49	4	46.616	1.5	
49	43	43	49	6	44.590	2.8	99	89	7	8	1	8.139	-0.1	
50	44	73	75	2	74.967	0.0	100	9	16	17.11	1.11	17.252	-0.1	

Values marked with an "X" were excluded from the calculations. Outliers "O" were also excluded.

Alternate (Quantitative) Method Comparison

X Method: V9

Y Method: A1650-506

Experimental Results

Spec ID	Results			Calc'd Y	SEE Factor	Spec ID	Results			Calc'd Y	SEE Factor
	X	Y	Bias				X	Y	Bias		
101 90	27	28	1	28.390	-0.2	106 95	40	43	3	41.553	0.9
102 91	28	29	1	29.402	-0.3	107 96	41	43	2	42.565	0.3
103 92	62	66	4	63.829	1.4	108 97	26	25	-1	27.377	-1.5
104 93	52	53	1	53.703	-0.4	109 98	43	48	5	44.590	2.2
105 94	15	18	3	16.239	1.1	110 99	9	11	2	10.164	0.5

Values marked with an "X" were excluded from the calculations. Outliers "O" were also excluded.

Alternate (Quantitative) Method Comparison Report Interpretation Guide

There are many reasons for doing method comparison studies. Perhaps the most common:

- To determine the relationship of the Medical Decision Points (MDPs) of an old method with those of a new method. In other words, "Can I continue to use the same MDPs with the new method?"
- To validate a new method being brought into the lab, by demonstrating that it is statistically identical to the method currently in use.

The statistical tool used is linear regression. Bottom line -- the methods can be considered statistically identical if:

- The slope is 1.00 (within 95% confidence)
- The intercept is 0.00 (within 95% confidence)
- The predicted Y MDPs are equal to the X MDPs (within 95% confidence)

Not all regressions are method comparisons. This Report Interpretation assumes that X and Y are alternative methods for measuring the same quantity, and that the purpose of the experiment is to determine whether X is statistically identical to Y. If the purpose is to predict weight as a function of height, or to predict APTT levels from Heparin levels, some of the interpretive comments may not apply.

Regression Approaches

The report shows at least two, and (optionally) three sets of regression coefficients.

Regular Regression: This is the ordinary least squares regression line commonly provided in spreadsheets and general statistical software. It is shown only to provide a familiar frame of reference; it is not used to estimate Medical Decision Points. The problem with using regular regression to compare methods is that it assumes the X method is measured with no random error -- not very likely for clinical laboratory results. Regular regression almost always underestimates the true slope, sometimes by a very significant amount.

Deming Regression: This approach assumes that both the X and Y methods are subject to measurement error. In theory, a **Representative SD** (precision estimate) is input for each method. In practice, only the ratio of the two precisions affects the calculation. If exact precisions are unknown, entering 1.0 for both Representative SDs says "these methods have about the same precision", and gives reasonable results in most cases.

Several studies have shown that Deming Regression is the best approach to use when the two methods are expected to be identical, and the data is well-distributed and free of outliers. It can, however, be seriously affected by outliers. EP Evaluator® provides the option to automatically exclude

extreme outliers, or the user can exclude them manually.

All Regression Lines on the EP Evaluator® graphs are Deming Regression Lines. When MDPs are estimated by linear regression, Deming linear regression is used.

Passing-Bablok Regression: Passing-Bablok regression is a non-parametric regression technique developed specifically to be resistant to outliers.

Main strengths: There is no need to exclude perceived outliers, either manually or automatically. Like Deming, it does not assume that X is free from error. Comparative studies show that it performs about as well as Deming Regression in most cases, and better than Deming when outliers are present.

Main weaknesses: While Passing-Bablok provides confidence intervals for the slope and intercept, it does not give confidence intervals for predicted Medical Decision Points. This is a serious deficiency if a primary objective of the study is to evaluate equivalence of the MDPs. Passing-Bablok is also computationally intensive, particularly for large N, and it may be unreliable for very small N. EP Evaluator® does not show Passing-Bablok statistics when $N < 10$ or $N > 250$.

Removing Outliers

An outlier is a point so far from the others as to arouse suspicion that it was generated by a different mechanism. Some common causes: typing a number with the decimal point in the wrong place, analyzing the wrong sample, or entering incorrect specimen identification. The best way to deal with an outlier is to (manually) determine its cause and correct it. Another option is to use a statistical procedure to remove outliers automatically.

EP Evaluator® uses a somewhat complex iterative algorithm to identify outliers. The goal is eliminate points whose distance from the regression line exceeds 10 times the Standard Error of Estimate (SEE), where SEE is computed not from the full data set, but from the data set with outliers excluded. (When outliers are included, the SEE is over-stated. Also, the regression coefficients are suspect.)

An outlier is, by definition, a rare occurrence. If the mathematical algorithm excludes more than 5% of the data points, the report is stamped PRELIMINARY. This indicates that the automatic procedure has failed. The user should disable automatic outlier detection, and exclude outliers manually if necessary.

Interpreting your Results

When interpreting a method comparison report, there are two areas which must be addressed:

Alternate (Quantitative) Method Comparison Report Interpretation Guide

- First, is the QUALITY OF THE DATA adequate to accurately draw conclusions?
- Second, what conclusions can be drawn from those data?

These issues MUST be addressed in this order. If the data quality is not adequate, then any additional conclusions drawn from those data may well be wrong.

Data Quality Statistics

The most important elements of a good method comparison study are a reasonable N (number of x-y pairs) and a good distribution of results. Generally a good experiment will include 30 to 50 specimens with their results distributed more or less evenly across the method's reportable range.

Results Range: The minimum and maximum values of X and Y. It is inappropriate to draw conclusions outside the range of data studied. When evaluating MDPs, it is important to include data points that cover the full range of MDPs.

Result Range Analysis: This (optional) table shows how the X values are distributed within the range. A relatively even distribution is desirable. If 99% of the values are at the low end and 1% are at the high end, with none in the middle, the regression slope is almost totally determined by the handful of high points.

Points (Plotted/Total): More commonly called N, the number of x-y pairs in the regression. "Plotted" is the number on which calculations are based. The difference between Plotted and Total is points that were excluded, either manually or by the automatic outlier removal procedure. CLSI considers N=40 to be the minimum for a good method comparison study. Increasing N improves the quality up to a point, but a good distribution of data is much more important than a large N.

Correlation Coefficient (R): R generally corresponds to the width of an ellipse drawn around the data. The narrower the ellipse relative to its length, the higher R will be. If there lots of error, the width will be greater and will result in a lower R.

R ranges from -1 to 1. Zero means there is absolutely no relationship. +1 or -1 means there is a perfect relationship, and a very high-quality regression. An R of 1.000 could be achieved just as easily with a slope and intercept of 1.000 and 0.0 as with a slope and intercept of 0.5 and 400 respectively. In other words, *it specifies the degree of correlation, not the degree to which the two methods match.*

In a method comparison setting, R has special significance:

- A small R may be a sign that the Results Range is inadequate. Adding samples to increase the range of X will improve both the R value, and the quality of the study.

- If R is less than a user-selectable cutoff value (0.90, 0.95, or 0.975), regression is not used to evaluate Medical Decision Points. Instead, they are evaluated by the method of Partitioned Biases.

Interpreting the Regression Statistics

Assuming that the quality of the data is adequate, you may proceed to interpreting the results.

Slope, Intercept, and their Confidence Intervals: When two methods are statistically identical, the 95% confidence interval for the slope includes 1.00, and the 95% confidence interval for the intercept includes 0.0.

Example: If the 95% CI for the slope is 0.92 to 1.02, 1.00 is included in the interval. However, if the 95% CI is 0.82 to 0.92, 1.00 is not included in the interval.

If the experiment were repeated with different data, the slope and intercept would be a bit different. But 95% of such estimates are expected to fall within the confidence interval.

Medical Decision Point Analysis: A Medical Decision Point is an analyte concentration at which medical decisions change. If the concentration is to one side of the MDP, one decision is made; if on the other side of the MDP, a different decision is made. For example, Fasting Plasma Glucose above 126 mg/dL (7 mmol/L) indicates hyperglycemia which, if confirmed, establishes a diagnosis of diabetes. For obvious reasons, it is particularly important that the two methods agree at the MDPs.

When the two methods are statistically identical, the 95% Confidence Interval for each Y MDP includes the corresponding X MDP.

Standard Error of Estimate (SEE): measures the spread of the x-y data around the linear regression line. If both methods have the same constant precision SD across the full analytical range, SEE should be about 1.4 times the precision SD.

Bias, and its Relationship with Regression

Bias is the difference Y-X. The **Bias Plot** is a scatter plot with X on the x-axis, and Y-X on the y-axis. The ideal bias plot would have all points falling exactly on the zero line. That is unlikely to occur in practice, because both X and Y are measured with some random error. A good bias plot is centered on the zero-line, and forms an envelope of approximately constant width about it.

Constant Bias is present when Y is consistently greater than (or less than) X by a constant amount. The bias plot forms a constant-width envelope around the average bias line instead of the zero line. The regression intercept

Alternate (Quantitative) Method Comparison Report Interpretation Guide

measures constant bias. In fact, if the slope is exactly 1.000, the regression intercept is equal to the average bias.

Proportional Bias is present when Y differs from X in a way that is proportional to X. For example, Y may be consistently 5% higher than X instead of 5 units higher. On the Bias plot, the points center around an upward or downward-sloping line instead of a horizontal line. The regression slope is a measure of proportional bias.

The **Method of Partitioned Biases** comes into play when R is "small", as defined by the cutoff value (0.90, 0.95, or 0.975). In this situation, the Bias Plot is divided into three segments, with the same number of points in each segment. It is assumed that bias is approximately constant within each segment. This segmented structure provides an estimate of bias and its 95% confidence interval at the Medical Decision Points.

Preliminary Report

The word PRELIMINARY printed diagonally across the report indicates that the data is incomplete, and the report is not acceptable as a final report. Some or all of the statistics may be missing. Causes:

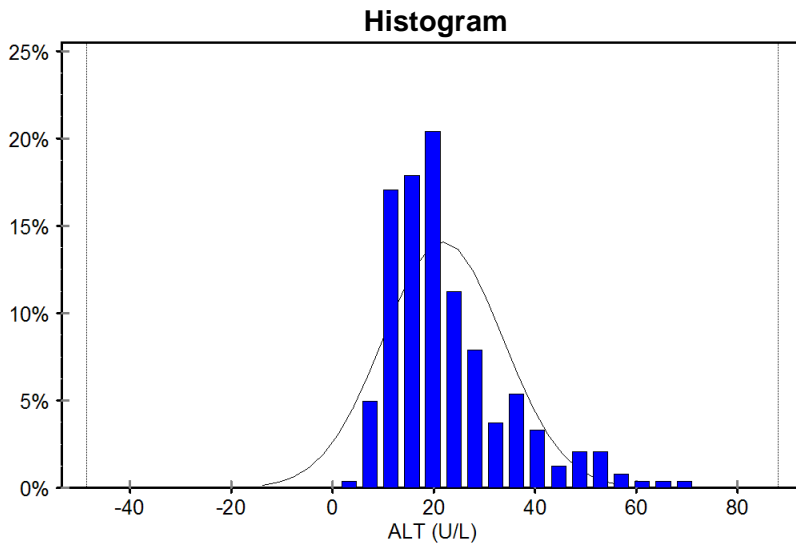
- Less than 3 unexcluded x-y pairs.
- More than 5% of points are outliers.
- Excluding outliers reduced the range of X by more than 50%. The range of X is a significant aspect of data quality, and it should be confirmed by the analyst rather than by a mathematical algorithm.

Reference Interval Estimation: Combined

Central 95% Interval (N = 240)

	Lower		Upper		Confidence Ratio
	Value	90% CI	Value	90% CI	
Nonparametric (CLSI C28-A)	8	6 to 9	54	49 to 65	0.21
Alternatives:					
Transformed Parametric	8	7 to 8	52	48 to 57	0.12
Parametric	-1	-3 to 1	46	44 to 48	0.09

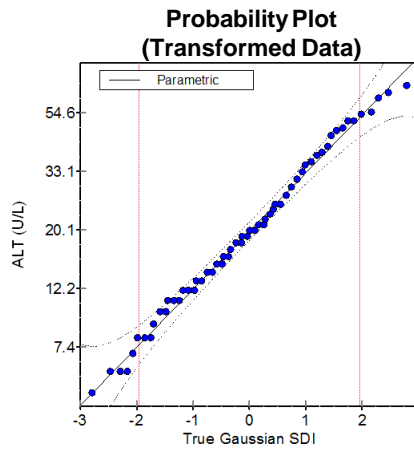
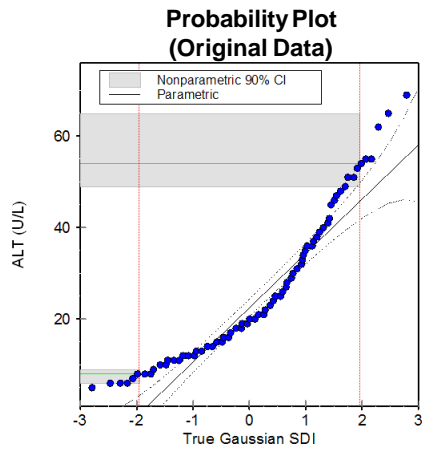
Confidence Limits for Nonparametric CLSI C-28A method computed from C28-A Table 8.



Selection Criteria:
 Bounds None
 Filter None

Statistics:
 Mean 22.5 U/L
 SD 11.9
 Median 19.5
 Range 5 to 69
 N 240 of 240
 Distinct values 50
 Zeroes 0
 Central 95% Index 6.0 to 235.0

Analyst mkf
 Expt. Date 13 Apr 2000



Normalizing Transformation

Exponent	0.00 (log)
Constant	0.00

Accepted by: _____

Signature

Date

Results Listing

Result	Count	Result	Count	Result	Count	Result	Count
5	1	18	10	31	5	46	1
6	3	19	13	32	1	47	2
7	1	20	15	33	1	48	2
8	5	21	11	34	2	49	1
9	3	22	8	35	2	51	3
10	4	23	5	36	6	53	1
11	11	24	3	37	3	54	1
12	13	25	11	38	2	55	2
13	13	26	5	39	3	62	1
14	13	27	1	40	3	65	1
15	10	28	6	41	1	69	1
16	11	29	2	42	1		
17	9	30	5	45	2		

Values marked with an "X" were excluded from the calculations.

Establish Reference Interval Report Interpretation Guide

Establishing a Reference Interval

A Reference Interval (RI) in this case refers to the central 95% range -- or "normal" range -- for endogenous analytes of a healthy patient. RIs are established by assaying a large number of specimens from healthy people and using the resulting sample to estimate the the central 95% range of the population. "Normal Range" is a pun. The statistical meaning is that it represents (often inaccurately) a Gaussian distribution of results. The clinical meaning is that it is derived from a healthy population.

While estimating the central 95% may seem like a simple task, there are some complicating factors:

- It is inappropriate to establish a reference interval for an exogenous analyte (i.e., a drug) using this program. This warning also applies to cutoff values for endogenous analytes such as critical levels of CKMB. In these cases, use the ROC Curve approach.
- The population used to establish the RI must be the same population to which the RI applies. Determine a PSA on a male population, not female lab employees. Similarly, determine first trimester bHGC's on women in that stage, not on non-pregnant women.
- RIs can differ dramatically with age, gender, and lifestyle. Hemoglobin concentrations are quite different for a pediatric population than for an adult population. Cholesterol concentrations are much lower for a Japanese population than for an American population.
- The number of specimens is important. CLSI recommends using at least 120 specimens. This is the smallest sample size to obtain both an RI estimate and a 90% confidence interval for it without making an assumption about the population distribution. All other things being equal, a larger sample size gives a more accurate RI and a tighter confidence interval.
- The quality of the RI cannot be judged solely by the width of the confidence interval. A parametric RI estimate will almost always have a narrower confidence interval than a nonparametric RI estimate. However, if the population distribution is materially non-Gaussian, the narrow parametric confidence interval is computed from a false assumption.

Definition of Statistical Terms

Reference Interval. The range of values that includes the central 95% of the population. The two endpoints of the Reference Interval are called the **reference limits**. Occasionally a reference interval is also calculated for some

other percent, like 90% or 80%. This may be appropriate if it is not possible to obtain enough reference subjects to calculate a 95% interval.

Confidence Interval (CI). A measure of the precision of the Reference Interval estimate. For example, suppose you estimate upper reference limit for Haptoglobin at 193 mg/dL, with a 90% confidence interval of 175-210. If you repeat the calculation on many different samples, you would expect 90% of the estimates to be between 175 and 210.

Confidence Ratio The ratio of the average confidence interval width to the reference interval width:

$$0.5 * (\text{URLU} - \text{URLL} + \text{LRLU} - \text{LRL}) / (\text{URL} - \text{LRL})$$

A value of 0.10 or less is desirable. Values over 0.30 are flagged in the report. The primary determinant of the confidence ratio is sample size -- confidence ratio improves as sample size increases.

Distribution. A statistical concept describing what percent of the population values fall into various ranges. A familiar example is the bell-shaped **Gaussian Distribution** curve, where 95% of the values lie within +/- 2 SD of the mean.

Nonparametric Method. A method of calculating a Reference Interval that makes no assumption about the shape of the population distribution. The simplest nonparametric method is that recommended by CLSI C28A -- obtain 120 samples and list them in ascending order. The lower reference limit is the 3rd sample, and the upper reference limit is the 118th sample. With as few as 40 results it is still possible to estimate a Reference Interval using selected sample ranks, but we refer to it as the Nonparametric Index method instead of the CLSI method. (CLSI C28-A recommends no fewer than 120 results.)

Parametric Method. A method of calculating a Reference Interval which assumes the population either follows a Gaussian distribution or can be converted to a Gaussian distribution -- perhaps by taking a log or square root. The simplest parametric method is to calculate the reference limits as Mean +/- 2 SD. The **Transformed Parametric Method** first attempts to change the scale of the data so it has a Gaussian distribution. Then it computes mean +/- 2 SD, and converts the answer back to the original units.

A parametric method based on a false assumption may be unreliable. For example, when the distribution is not Gaussian the parametric method may give a negative value for the lower reference limit. Not only is the estimate unreliable, the 90% confidence interval is also overly optimistic. It is based on the same false assumption as the

Establish Reference Interval Report Interpretation Guide

reference interval itself. Don't use the Reference Interval from mean ± 2 SD unless it is very clear that the curve really is Gaussian. If Mean - 2SD is negative, then the curve is not Gaussian.

Histogram. A bar chart that illustrates the distribution of results. Each rectangle represents a equal-width range of values, with height proportional to the number of results in the range. A Gaussian distribution curve is plotted in the background.

Probability Plot. A chart that helps you judge whether your data is Gaussian, as well as assessing how reasonable the Reference Interval estimate is. Plotted points represent specimens from your sample. The Y axis is result value, and the X axis is SD's of a hypothetical Gaussian distribution.

- If your data is Gaussian, the points lie in a straight line.
- Where the scatter of dots crosses ± 2 SD is the nonparametric Reference Interval estimate.
- If you draw a straight line that passes through the sample mean, and has the slope of the sample SD, it intersects ± 2 SD to give the parametric Reference Interval estimate.
- If the points fall very close to this *true Gaussian line*, it is likely that the parametric and nonparametric Reference Interval calculations will agree.

Normalizing Transformation (Box-Cox Transformation). With the aid of a moderately complex algorithm, the software attempts to transform the data into a Gaussian model so simple Gaussian statistics can be calculated. This normalizing transformation produces an exponent and a constant. It really does not matter to you what these two numbers are. The only thing that really matters is the degree to which the data is linearly distributed in the transformed probability plot.

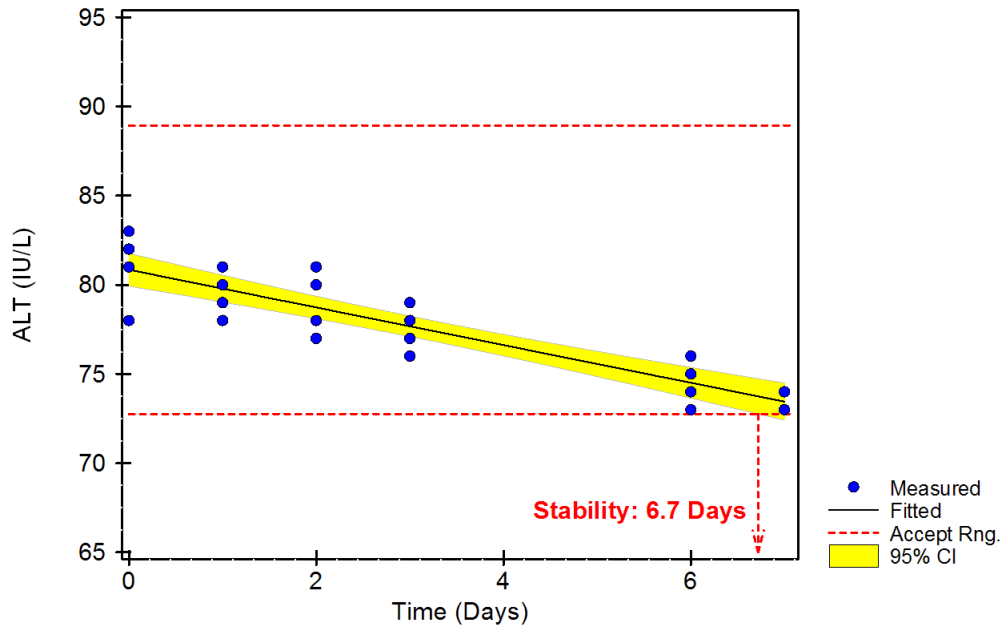
Partitioning Analysis. An analysis to determine whether separate reference intervals are statistically justified for population subclasses. Example: Should separate reference intervals be estimated for men and women, or should there be a single combined reference interval for both? The CLSI guideline recommends the following procedure:

- Collect a preliminary sample consisting of at least 60 reference subjects from each subclass.
- Perform a statistical test to determine whether these samples are significantly different from each other.
- If there is evidence of a difference, collect an additional 60 samples from each subclass, and estimate a separate Reference Interval for each group.

References

1. Harris EK, Boyd JC. *Statistical bases of reference values in laboratory medicine*. New York: Marcel Dekker, 1995:1-61.
2. CLSI. *How to Define and Determine Reference Intervals in the Clinical Laboratory; Approved Guideline*. C28-A 15:4.

Stability



Evaluation: Stability estimate: 6.7 Days

Evaluation Criteria

Allowable Total Error (TEa)	20.0%
% for Instability	50%
Allowable Instability	10.0%
Est. Stable Concentration	81 IU/L
Acceptable Range	73 to 89 IU/L

Fit Statistics

Model	$Y = (80.8) + (-1.056) * \text{Time}$
Slope and 95% CI	-1.056 (-1.283 to -0.829)
Intercept and 95% CI	80.8433 (79.9223 to 81.7644)
R	0.900
SEE	1.4
Experiment Duration	7.0 Days
# Meas. Periods	6
# Data Points	24

Supporting Data

Expt. Date	16 Jul 2009
Analyst:	mkf
Analyte Units	IU/L
Time Units	Days
Comment:	

Accepted by: _____

Signature

Date

Instrument: Eximer
Sample Name: 22346A
Refrigerated

Stability

Results Listing

Time	Mean	SD	CV	Fitted	95%CI		Measured Values			
0	81.0	2.2	2.7	80.8	79.9	81.8	78	81	82	83
1	79.5	1.3	1.6	79.8	79.0	80.5	78	79	80	81
2	79.0	1.8	2.3	78.7	78.1	79.4	77	78	80	81
3	77.5	1.3	1.7	77.7	77.1	78.3	76	77	78	79
6	74.5	1.3	1.7	74.5	73.6	75.4	73	74	75	76
7	73.5	0.6	0.8	73.5	72.4	74.5	73	73	74	74

Values marked with an "X" were excluded from the calculations.

Stability Report Interpretation Guide

The Stability module is designed to evaluate measured drift over time. The report reflects observed change for a single sample under fixed storage conditions. A comprehensive stability study will typically include several reports covering samples near the upper and lower limits of the reportable range, and also multiple storage conditions.

Drift is evaluated by regression analysis of the observed value (Y-axis) vs time (X-axis). Stability duration is the earliest time at which the 95% confidence interval for the fitted curve crosses a predetermined allowable drift.

Experiment Design

The sample is aliquotted into tubes for assay at a later time. One aliquot is assayed at the beginning of the experiment (time=0). At appropriate times, additional aliquots are assayed. The total duration of the experiment should extend past the product's expected stability duration. Within this total duration, at least five time points should be evaluated. Increasing the frequency will increase confidence in the result. To reduce the effects of imprecision, it is recommended that each sample be assayed in duplicate or triplicate.

Example: The claimed or expected shelf life is 5 days. The experiment is run for 7 days. The sample is assayed in triplicate at the beginning of the experiment, and on days 1, 2, 3, 6, and 7.

Allowable Instability

A predetermined allowable instability (allowable drift) must be specified in order to perform the analysis. The results are acceptably stable as long as the 95% confidence interval for the regression line remains within allowable instability bands. In EP Evaluator®, allowable instability is specified as a percent of Total Allowable Error (TEa). This *Instability Budget* is typically 50%.

Example: Based on CLIA limits, TEa for AST is 20%. Assuming a 50% Instability Budget, Allowable Instability is +/- 10% of the concentration at time=0.

Imprecise. If Allowable Instability is too small, no conclusion can be drawn because the data is Imprecise. Imprecise means the confidence interval for the regression line is outside the acceptable stability range even at time=0. Approaches for dealing with this problem include:

- Increase the number of replicate measurements to reduce the width of the regression confidence interval.
- Allowable Instability is unrealistically low. If Allowable Instability is less than 2 X routine CV, a large number of replicates will be required.

Evaluating the Fitted Curve

The curve fit is a regression line determined from the observed data. In most cases an ordinary regression line is used, and the fitting equation is $Y = \text{Intercept} + \text{Slope} \times \text{Time}$. When the duration of the experiment is much longer than the expected stability duration, you may choose to fit an exponential curve: $\text{Log}(Y) = \text{Intercept} + \text{Slope} \times \text{Time}$.

Not Significant. If the 95% confidence interval for the Slope includes Zero, then the slope is not statistically significant. This means that there is no systematic upward or downward trend in the data over the duration of the experiment. You may conclude that there is no drift. Note that this statistic only tests for drift; it does not test for other forms of instability, such as increases in imprecision.

Estimated Stable Concentration and Acceptable Range. The intercept of the fitted curve is an estimate of the stable concentration at time=0. The Acceptable Range is set relative to the Estimated Stable Concentration. For example, suppose Allowable Instability is +/- 10%, and the Estimated Stable Concentration is 50 units. Then the Acceptable Range is 50 +/- 10%, or 45 to 55 units.

Bad Fit. Regression analysis assumes that the fitted curve is a good estimate of overall behavior of the process over time. If this assumption is true, there are three values that might be used to estimate the stable concentration:

- A predetermined Target Value for the sample, established from an independent source.
- The mean measured value of the sample at time=0.
- The intercept of the fitted curve.

EP Evaluator® uses the intercept of the fitted curve as its estimate of the stable concentration. If the mean measured value at time=0 does not lie within the 95% confidence interval for the intercept, the curve is not a good fit to the observed data, and conclusions about stability duration are questionable. You might consider increasing the frequency of measurement time periods, inspecting observed data for outliers, or increasing the number of replicate measurements.

Pass or Fail

This experiment is not intended as a Pass/Fail test for stability. Instead, its purpose is to determine the rate of drift and, given that rate of drift, the time interval over which the process will produce acceptable results. The expected outcome is that the experiment duration is longer than the shelf life, so the crossover point can be determined. If the duration of the experiment is too short, the report will classify instability as "Not Significant".

Stability Report Interpretation Guide

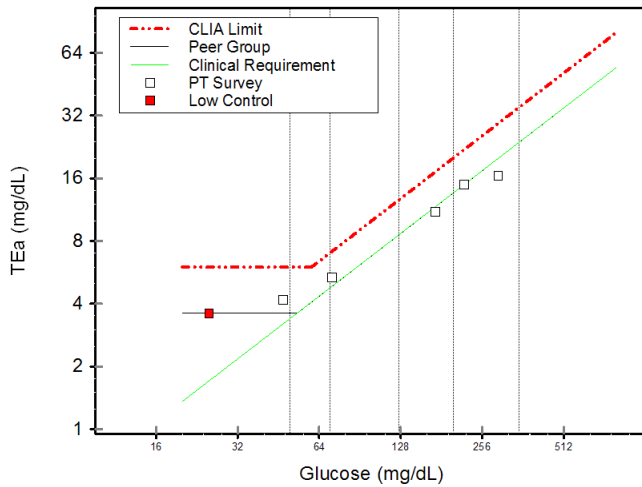
Preliminary Report

The word PRELIMINARY printed diagonally across the report indicates that the data is incomplete, and the report is not acceptable as a final report. Some or all of the statistics may be missing. The Stability report is preliminary if:

- There are no measurements with time=0
- Less than five time periods with valid data
- Data is "imprecise" or a "bad fit", as defined above.

Performance Standards Alternatives

Based on Approach	Allowable Total Error is the greater of:	
	Percent	Concentration
Clinical Requirements	6.8%	--
Regulatory (or quasi-regulatory) Limit	10%	6 mg/dL
Peer Group SD-Based	6.8%	3.6 mg/dL
	3 x PeerCV	3 x Low Ctrl SD



MDP	Allowable Total Error Range (mg/dL)		
	Lowest	Highest	Regulatory
50	3.40	6.00	6.00
70	4.76	7.00	7.00
126	8.57	12.60	12.60
200	13.60	20.00	20.00
350	23.80	35.00	35.00

Reportable Range: 20 to 800 mg/dL

Note: Both axes of the graph are shown on a logarithmic scale.

Peer Group Statistics

	Low Control	Peer Group Statistics from PT Survey				
		1	2	3	4	5
Mean	25	46.9	71.1	171.4	219.1	292.2
SD	1.2	1.4	1.8	3.7	5	5.5
CV (%)	4.8	3.0	2.5	2.2	2.3	1.9
		6	7	8	9	10
Mean	--	--	--	--	--	--
SD	--	--	--	--	--	--
CV (%)	--	--	--	--	--	--

Supporting Data

Reportable Range: 20 to 800 mg/dL
 SD Multiple: 3 SD
 Analyst: Fred Doe
 Date: 07 Dec 2003
 Comment: Real example

Accepted by: _____

Signature

Date

Performance Standards Report Interpretation Guide

A Performance Standard (TEa) defines the quality of results that must be produced. For most analytes there is no single correct TEa. This analysis serves as a starting point for the laboratory to establish its own Performance Standard. The fundamental purpose is to obtain a TEa which is both **achievable** and **defensible**. It is counter-productive to use a TEa which is either too large or too small. For example:

- If you participate in a proficiency (PT) program, and you set your TEa much higher than your provider's grading limit, you may fail PT.
- If, on the other hand, you set your TEa at or below your 3 x your SD from routine QC, you will find the goal difficult or impossible to achieve.

Three Approaches

Medical (Clinical) Requirements. The source of medical requirements used by most labs will be the national medical requirements coming from the working groups at NIH. As of October 2007, values have been established for five analytes: HDL, LDL, and Total Cholesterol, Triglyceride, and Creatinine. Also, some academic institutions have established their own medical requirements. Medical requirements appear on the report exactly as you entered them on the input screen -- either as a percent, a concentration, or greater of concentration/percent.

Regulatory Requirements, such as the CLIA '88 Proficiency Testing limits in the United States. Other jurisdictions and/or regulatory agencies have also established requirements. Like medical requirements, regulatory requirements appear on the report exactly as entered on the input screen. They must be expressed in concentration, percent, or greater of concentration/percent. If your requirement is an SD multiple, your "regulatory" requirements are equivalent to Achievable Requirements.

Achievable Requirements, based on actual performance of clinical laboratory methods on the

field. In general, TEa's can be easily calculated from PT survey results. Furthermore, these values are generally available for almost all the tests routinely performed in most hospital laboratories. On the report, Achievable Requirements are called "Peer Group SD-Based" requirements.

Achievable requirements are computed as the greater of concentration or percent. The percent is 3 x the median CV from from PT survey results. It is a good idea to enter survey results from two different surveys, and at least 6 specimens covering as wide a range of concentrations as possible. Alternatively, you can choose a multiple of 2 or 2.5 instead of 3. One reason to use a lower multiple is that your PT provider's grading limit is 2.5 SD or 2 SD.

The concentration component of TEa is 3 x the SD from a low control run in your lab. You can use any clinically insignificant concentration for the low control.

Report Interpretation

The table at the top of the report shows a TEa estimate for up to three approaches, depending on which data you entered. Data may not be available for some approaches. Also, there may be wide variation among the approaches. This table serves as a guide for setting your TEa.

- If medical requirements are available for an analyte, you should use them if at all possible! The only reason to use anything else is that the medical requirements are not achievable.
- If you have regulatory requirements, you should not set your TEa higher than the regulatory limit. Or, if you do so, you should be able to justify your TEa.
- The SD-Based requirement represents what is commonly achievable in the industry.

Screen Shots

EP Evaluator® Release 9 Simple Precision

File Edit Module Experiment RRE ERI View Utilities Tools Help

Simple Precision

Instrument: DXI-1630

Analyte	Sample	N	Mean	SD	CV
● Beta HCG	Level 2	20 of 20	140.330	7.470	5.3%
● Beta HCG	Level1	20 of 20	7.115	0.436	6.1%
● PSA	Level 2	20 of 20	27.905	2.265	8.1%
● PSA	Level1	20 of 20	1.223	0.053	4.3%
● T4, Total	Level 2	20 of 20	14.215	0.581	4.1%
● T4, Total	Level1	20 of 20	2.940	0.194	6.6%
● Troponin	Level 2	20 of 20	22.727	0.645	2.8%
● Troponin	Level1	20 of 20	0.592	0.024	4.0%
● TSH	Level 2	20 of 20	34.292	1.993	5.8%
● TSH	Level1	20 of 20	0.419	0.026	6.3%

Legend

- ◇ Not calculated
- Insufficient data
- Sufficient data
- FAIL
- PASS
- May need review

