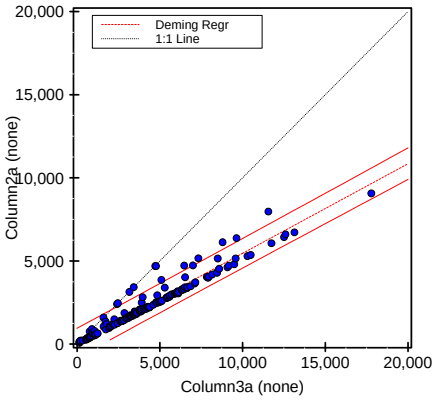
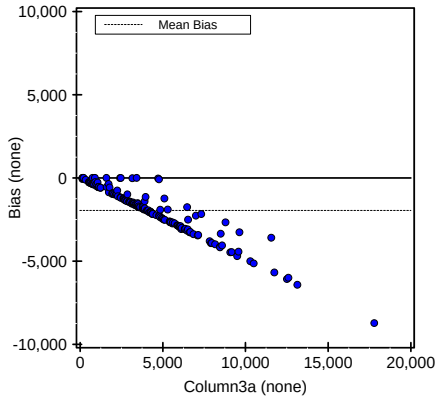
Alternate (Quantitative) Method Comparison

X Method: Column3a **Y Method: Column2a**

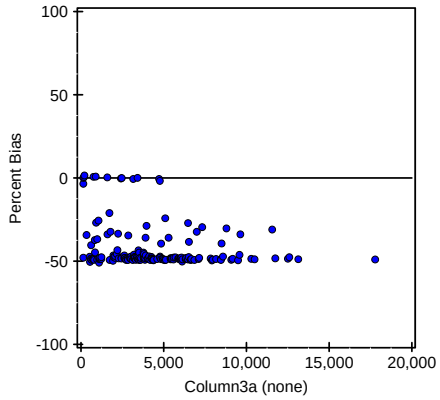
Scatter Plot



Bias



Percent Bias



Regression Analysis

	Deming	Regular
Slope:	0.539 (0.516 to 0.562)	0.530 (0.507 to 0.553)
Intercept:	69.0 (-51.8 to 189.9)	109.0 (-11.6 to 229.6)
Std Err Est:	444.3	443.4

95% Confidence Intervals are shown in parentheses

Supporting Statistics

Corr Coef (R):	0.9634	SubRange Bounds:	None
Bias:	-1949.1	Points (Plotted/Total):	166/166
X Mean ± SD:	4378.9 ± 2995.9	Outliers:	None
Y Mean ± SD:	2429.8 ± 1648.2	Scatter Plot Bounds:	95% CI
Std Dev Diffs:	1475.9		

Experiment Description

	X Method	Y Method
Expt Date:	28 Sep 2016	27 Sep 2016
Rep SD:	1	1
Result Ranges:	142 to 17789	74 to 9066
Units:	none	none
Analyst:	VK	VK
Comment:		

Accepted by: _____ _____

Signature Date

Alternate (Quantitative) Method Comparison

X Method: Column3a **Y Method: Column2a**

Experimental Results

	Spec ID	Results X	Results Y	Bias	Calc'd Y	SEE Factor
1	43	142	137	-5	145.6	0.0
2	10	142	74	-68	145.6	-0.2
3	67	144	144	0	146.7	0.0
4	74	187	188	1	169.8	0.0
5	18	208	211	3	181.2	0.1
6	17	338	222	-116	251.3	-0.1
7	51	525	276	-249	352.1	-0.2
8	71	549	272	-277	365.0	-0.2
9	69	600	309	-291	392.5	-0.2
10	109	616	367	-249	401.1	-0.1
11	68	637	324	-313	412.5	-0.2
12	90	716	367	-349	455.0	-0.2
13	11	757	762	5	477.2	0.6
14	110	761	390	-371	479.3	-0.2
15	66	796	404	-392	498.2	-0.2
16	13	833	522	-311	518.1	0.0
17	14	851	470	-381	527.8	-0.1
18	139	894	901	7	551.0	0.8
19	3	917	670	-247	563.4	0.2
20	35	985	501	-484	600.1	-0.2
21	130	986	623	-363	600.6	0.1
22	6	1059	787	-272	640.0	0.3
23	42	1098	539	-559	661.0	-0.3
24	72	1105	566	-539	664.8	-0.2
25	154	1127	579	-548	676.6	-0.2
26	73	1131	588	-543	678.8	-0.2
27	150	1230	631	-599	732.2	-0.2
28	70	1233	646	-587	733.8	-0.2
29	34	1592	1596	4	927.3	1.5
30	88	1604	1060	-544	933.8	0.3
31	101	1720	1357	-363	996.3	0.8
32	7	1740	882	-858	1007.1	-0.3
33	82	1785	1208	-577	1031.4	0.4
34	9	1937	972	-965	1113.3	-0.3
35	91	1960	1047	-913	1125.7	-0.2
36	148	1982	1039	-943	1137.6	-0.2
37	79	2065	1072	-993	1182.3	-0.2
38	114	2069	1097	-972	1184.5	-0.2
39	55	2164	1175	-989	1235.7	-0.1
40	2	2207	1249	-958	1258.9	0.0
41	5	2247	1493	-754	1280.5	0.5
42	95	2280	1177	-1103	1298.3	-0.3
43	46	2417	2408	-9	1372.1	2.3
44	93	2437	1267	-1170	1382.9	-0.3
45	166	2449	1262	-1187	1389.4	-0.3
46	86	2465	2462	-3	1398.0	2.4
47	158	2621	1400	-1221	1482.1	-0.2
48	75	2632	1364	-1268	1488.0	-0.3
49	25	2649	1376	-1273	1497.2	-0.3
50	160	2700	1384	-1316	1524.7	-0.3
51	142	2742	1389	-1353	1547.3	-0.4
52	78	2836	1465	-1371	1598.0	-0.3
53	135	2843	1442	-1401	1601.8	-0.4
54	39	2860	1870	-990	1610.9	0.6
55	45	2976	1537	-1439	1673.5	-0.3
56	50	2978	1562	-1416	1674.6	-0.3
57	145	2985	1551	-1434	1678.3	-0.3
58	87	3028	1598	-1430	1701.5	-0.2
59	1	3035	1556	-1479	1705.3	-0.3
60	143	3130	1586	-1544	1756.5	-0.4
61	40	3158	3138	-20	1771.6	3.1
62	41	3187	1658	-1529	1787.2	-0.3
63	20	3205	1727	-1478	1796.9	-0.2
64	89	3232	1699	-1533	1811.5	-0.3
65	163	3248	1688	-1560	1820.1	-0.3
66	61	3258	1690	-1568	1825.5	-0.3
67	8	3319	1687	-1632	1858.4	-0.4
68	57	3376	1761	-1615	1889.1	-0.3
69	161	3422	3422	0	1913.9	3.4
70	144	3440	1765	-1675	1923.6	-0.4
71	60	3451	1766	-1685	1929.6	-0.4
72	16	3477	1963	-1514	1943.6	0.0
73	132	3480	1839	-1641	1945.2	-0.2
74	116	3514	1937	-1577	1963.5	-0.1
75	138	3523	1785	-1738	1968.4	-0.4
76	97	3548	1841	-1707	1981.9	-0.3
77	58	3609	1833	-1776	2014.8	-0.4
78	76	3635	1885	-1750	2028.8	-0.3
79	22	3778	2081	-1697	2105.9	-0.1
80	24	3784	1979	-1805	2109.1	-0.3
81	136	3801	1993	-1808	2118.3	-0.3
82	156	3806	1959	-1847	2121.0	-0.4
83	141	3807	2065	-1742	2121.5	-0.1
84	157	3831	1973	-1858	2134.4	-0.4
85	107	3866	1966	-1900	2153.3	-0.4
86	94	3882	2008	-1874	2161.9	-0.3
87	149	3903	2499	-1404	2173.3	0.7
88	65	3912	2088	-1824	2178.1	-0.2
89	162	3964	2822	-1142	2206.1	1.4
90	106	4035	2076	-1959	2244.4	-0.4
91	115	4080	2101	-1979	2268.7	-0.4
92	112	4126	2181	-1945	2293.5	-0.3
93	59	4193	2126	-2067	2329.6	-0.5
94	29	4258	2238	-2020	2364.7	-0.3
95	103	4266	2206	-2060	2369.0	-0.4
96	117	4284	2199	-2085	2378.7	-0.4
97	127	4359	2213	-2146	2419.1	-0.5
98	83	4361	2228	-2133	2420.2	-0.4
99	54	4381	2213	-2168	2431.0	-0.5
100	49	4389	2227	-2162	2435.3	-0.5

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

Alternate (Quantitative) Method Comparison

X Method: Column3a **Y Method: Column2a**

Experimental Results

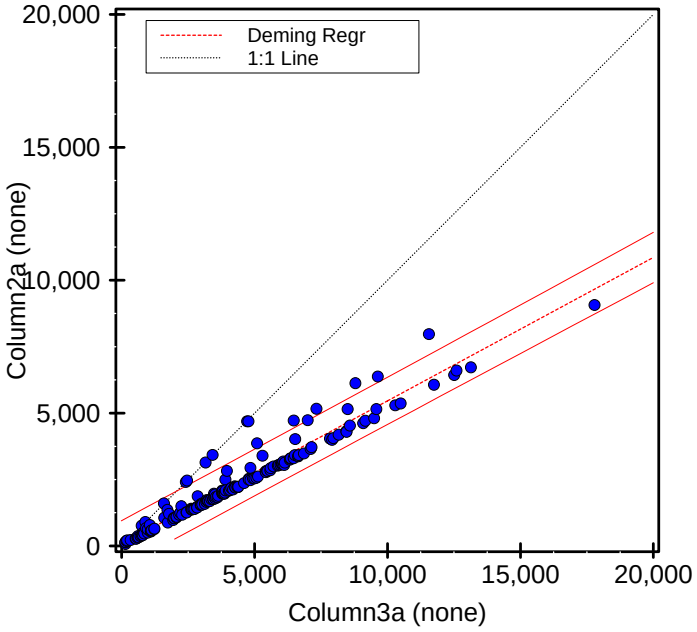
	Spec ID	Results X	Results Y	Bias	Calc'd Y	SEE Factor
101	4	4607	2363	-2244	2552.8	-0.4
102	63	4721	4691	-30	2614.3	4.7
103	153	4772	4688	-84	2641.8	4.6
104	15	4780	2521	-2259	2646.1	-0.3
105	64	4787	2493	-2294	2649.9	-0.4
106	151	4844	2479	-2365	2680.6	-0.5
107	53	4848	2936	-1912	2682.7	0.6
108	159	4947	2539	-2408	2736.1	-0.4
109	102	4994	2541	-2453	2761.5	-0.5
110	33	5070	2566	-2504	2802.4	-0.5
111	126	5095	3859	-1236	2815.9	2.3
112	96	5130	2609	-2521	2834.8	-0.5
113	81	5302	3395	-1907	2927.5	1.1
114	113	5418	2805	-2613	2990.0	-0.4
115	129	5430	2772	-2658	2996.5	-0.5
116	28	5477	2830	-2647	3021.9	-0.4
117	122	5486	2819	-2667	3026.7	-0.5
118	108	5592	2849	-2743	3083.9	-0.5
119	147	5610	2925	-2685	3093.6	-0.4
120	30	5717	2988	-2729	3151.2	-0.4
121	155	5876	3024	-2852	3237.0	-0.5
122	77	5914	3049	-2865	3257.5	-0.5
123	92	5977	3056	-2921	3291.4	-0.5
124	105	6013	3102	-2911	3310.8	-0.5
125	36	6034	3105	-2929	3322.2	-0.5
126	31	6058	3079	-2979	3335.1	-0.6
127	123	6078	3172	-2906	3345.9	-0.4
128	52	6115	3043	-3072	3365.8	-0.7
129	152	6137	3136	-3001	3377.7	-0.5
130	120	6330	3243	-3087	3481.7	-0.5
131	125	6342	3284	-3058	3488.2	-0.5
132	100	6363	3298	-3065	3499.5	-0.5
133	47	6471	3301	-3170	3557.8	-0.6
134	164	6475	4718	-1757	3559.9	2.6
135	99	6517	3420	-3097	3582.6	-0.4
136	32	6527	4019	-2508	3587.9	1.0
137	133	6635	3370	-3265	3646.2	-0.6
138	98	6647	3377	-3270	3652.6	-0.6
139	124	6665	3420	-3245	3662.3	-0.5
140	128	6856	3477	-3379	3765.3	-0.6
141	27	6999	4732	-2267	3842.4	2.0
142	121	7111	3649	-3462	3902.8	-0.6
143	119	7145	3714	-3431	3921.1	-0.5
144	12	7330	5159	-2171	4020.9	2.6
145	62	7836	4038	-3798	4293.7	-0.6
146	118	7915	3990	-3925	4336.3	-0.8
147	38	7968	4056	-3912	4364.8	-0.7
148	104	8167	4185	-3982	4472.1	-0.6
149	80	8459	4317	-4142	4629.5	-0.7
150	56	8462	4293	-4169	4631.2	-0.8
151	84	8503	5150	-3353	4653.3	1.1
152	21	8587	4529	-4058	4698.6	-0.4
153	85	8794	6125	-2669	4810.2	3.0
154	37	9087	4617	-4470	4968.1	-0.8
155	146	9168	4712	-4456	5011.8	-0.7
156	134	9500	4803	-4697	5190.8	-0.9
157	19	9582	5150	-4432	5235.0	-0.2
158	131	9637	6373	-3264	5264.6	2.5
159	140	10293	5289	-5004	5618.3	-0.7
160	26	10497	5359	-5138	5728.3	-0.8
161	23	11561	7971	-3590	6301.9	3.8
162	137	11748	6065	-5683	6402.8	-0.8
163	44	12510	6431	-6079	6813.6	-0.9
164	111	12600	6600	-6000	6862.1	-0.6
165	165	13141	6720	-6421	7153.8	-1.0
166	48	17789	9066	-8723	9659.7	-1.3

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

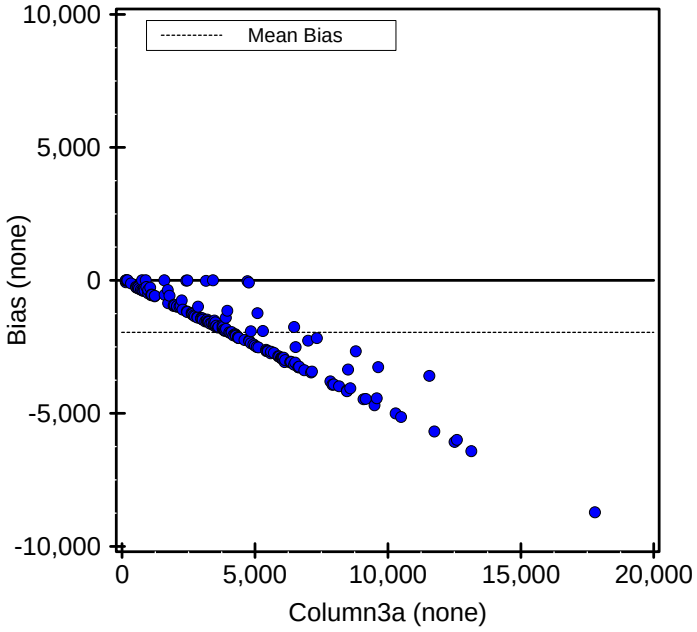
Alternate (Quantitative) Method Comparison

X Method: Column3a **Y Method: Column2a**

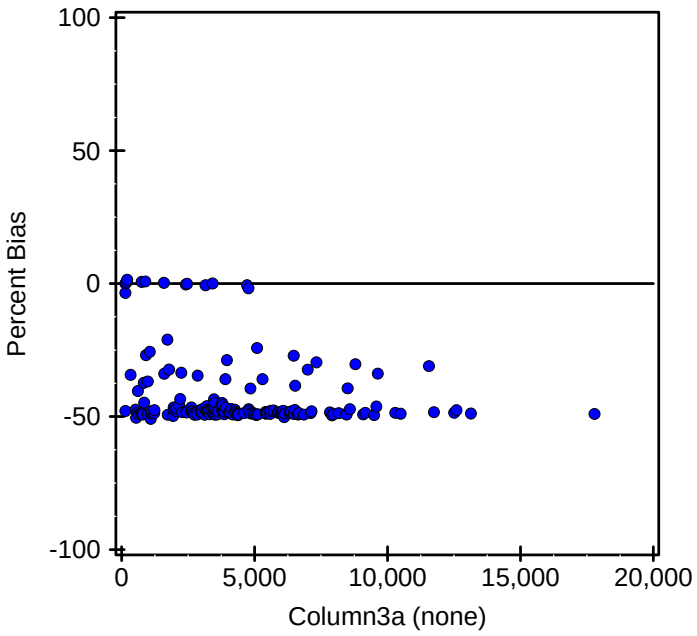
Scatter Plot



Bias



Percent Bias



This page contains a larger, working copy of the same graphs that appear on page 1.

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 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 an analyte concentration at which medical decisions change. If the concentration is to one side of the MDP, one decision is make; 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.