What are Sigma-metrics?
Benchmarking Quality, Optimizing QC

October 25th, 2015
EFLM Continuing Postgraduate Course in Clinical Chemistry
and Laboratory Medicine

15th EFLM Continuing Postgraduate Course in Clinical Chemistry and Laboratory Medicine:
"How to assess the quality of your method?"
October 24-25, 2015,Zagreb, Croatia

Sten Westgard, MS
Westgard QC, Inc.

Westgard QC

Outline of the Talk

- **Why do We need to worry about quality?**
- A brief introduction to Six Sigma
  - Counting defects: How does healthcare perform?
- Calculating Sigma-metrics
  - Setting Goals for Quality
  - Measuring Performance
  - Examples of Current Performance
- Tools for Sigma-metrics
  - Sigma-metric Equation
  - Method Decision Chart

Westgard QC

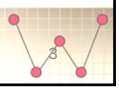
Why is Determining the Quality of the method OUR job?

- (Isn't every method on the market a quality method?)

  "Conclusion 7-1. The 510(k) clearance process is not intended to evaluate the safety and effectiveness of medical devices with some exceptions. The 510(k) process cannot be transformed into a premarket evaluation of safety and effectiveness as long as the standard for clearance is substantial equivalence to any previously cleared device."

  Institute of Medicine 2011: Medical Devices and the Public's health: the FDA 510(k) Clearance Process at 35 years, prepublication copy

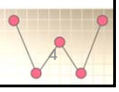
Westgard QC



Three Glucose methods: Are they acceptable?

Method A	Method B	Method C
CV = 2.3	CV = 1.9	CV = 1.9
Bias = 2.1	Bias = 4.2	Bias = 0

Westgard QC



Outline of the Talk

- Why do We need to worry about quality?
- **A brief introduction to Six Sigma**
- Counting defects: How does healthcare perform?
- Calculating Sigma-metrics
  - Setting Goals for Quality
  - Measuring Performance
  - Examples of Current Performance
- Tools for Sigma-metrics
  - Sigma-metric Equation
  - Method Decision Chart

Westgard QC

Six Sigma – A Way to Think About Errors

- Defects Per Million (DPM)
- Scale of 0 to 6
- 6 is world class (3.4 dpm)
- 3 is minimum for any business or manufacturing process (66,807 dpm)

Westgard QC

Two ways to Determine Sigma

- Count Defects, convert to DPM, look up in Sigma table
  - Short Term Sigma typically used
  - Most common method of calculating Sigma
- Measure Variation
  - Sigma-metric Equation

DPM	Sigma Short Term	Sigma Long Term	Yield	Cpk
3.4	6	4.5	99.99966	2
5	5.9	4.4	99.99954	1.97
9	5.8	4.3	99.99915	1.93
13	5.7	4.2	99.9987	1.9
21	5.6	4.1	99.9979	1.87
32	5.5	4	99.9968	1.83
48	5.4	3.9	99.995	1.8
72	5.4	3.9	99.993	1.77
108	5.2	3.7	99.989	1.73
159	5.1	3.6	99.984	1.7
233	5	3.5	99.98	1.67
337	4.9	3.4	99.97	1.63
483	4.8	3.3	99.95	1.6
687	4.7	3.2	99.93	1.57
968	4.6	3.1	99.90	1.53
1,350	4.5	3	99.87	1.5
1,866	4.4	2.9	99.81	1.47
2,555	4.3	2.8	99.74	1.43
3,467	4.2	2.7	99.65	1.4
4,661	4.1	2.6	99.5	1.37
6,210	4	2.5	99.4	1.33
8,198	3.9	2.4	99.2	1.3
10,724	3.8	2.3	98.9	1.27
13,903	3.7	2.2	98.6	1.23
17,864	3.6	2.1	98.2	1.2
22,750	3.5	2	97.7	1.17

Current Laboratory Performance

Sample Sigma-metrics:

Hemolyzed serum sample:
4.1 sigma

Control exceeds limits:
3.4 Sigma

Biggest problems:
Incorrect name/request (2.9)
Report takes too long (2.8)

Clin Chem Lab Med 2011;49(3):463–470 © 2011 by Walter de Gruyter · Berlin · New York. DOI 10.1515/CCLM.2011.067

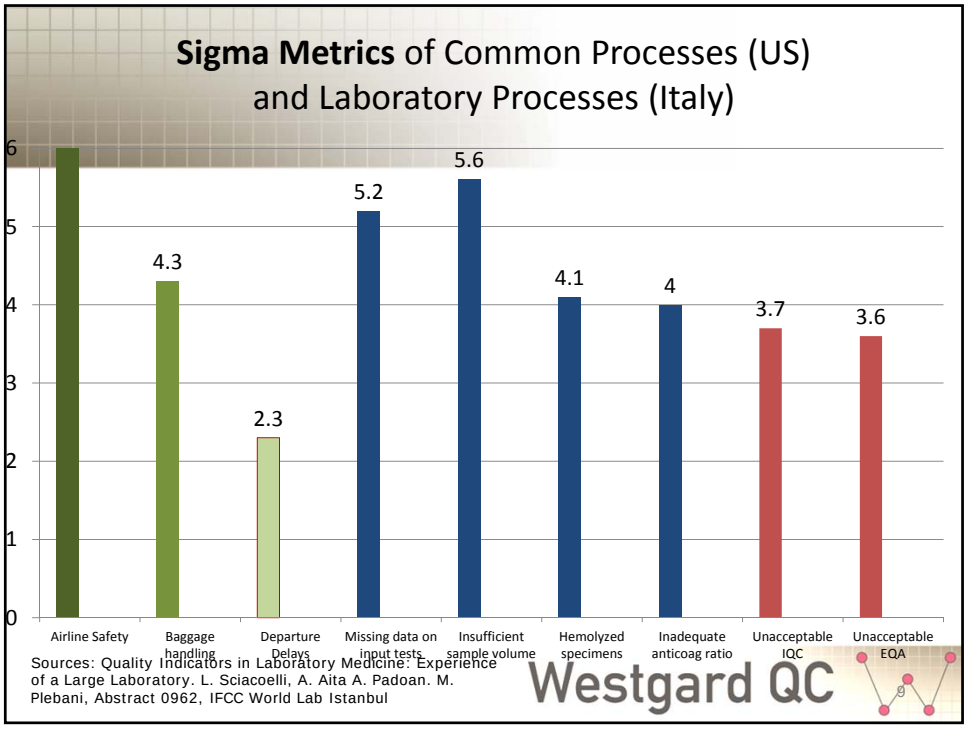
Quality indicators and specifications for key analytical-extranalytical processes in the clinical laboratory. Five years' experience using the Six Sigma concept

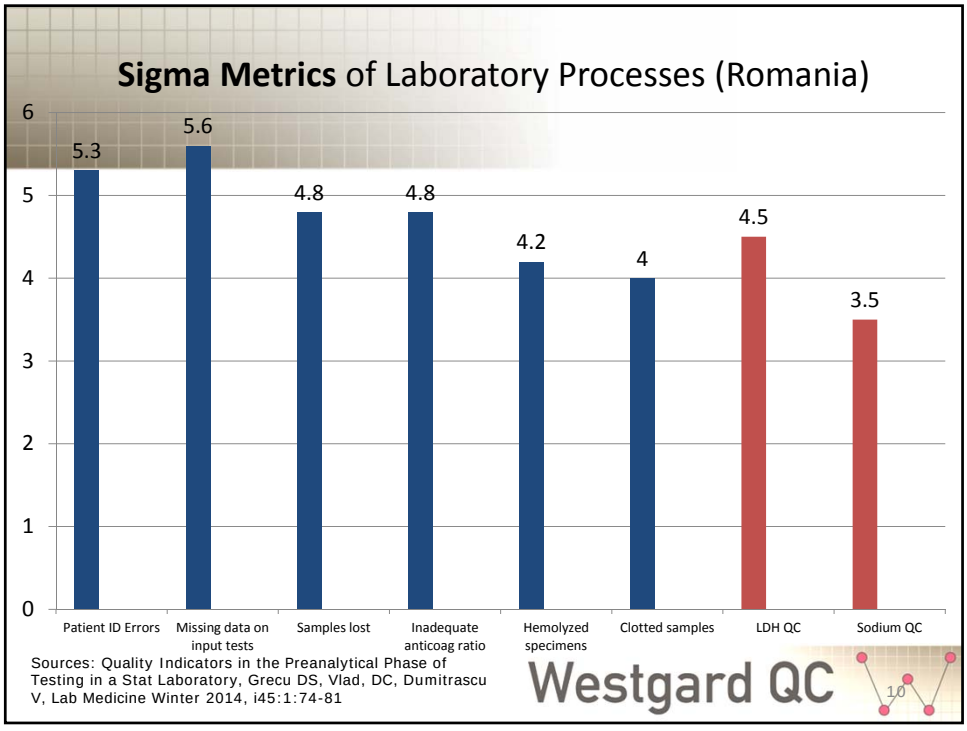
Mª Antònia Llopis[1,*], Gloria Trujillo[2], Mª Isabel Llovet[3], Ester Tarrés[4], Mercè Ibarz[5], Carme Biosca[6], Rosa Ruiz[7], Mª Jesus Alsina Kirchner[3], Virtudes Alvarez[7], Glòria Busquets[8], Mª Vicenta Doménech[9], Carme Figueres[10], Joana Minchinela[1], Rosa Mª Pastor[11], Carmen Perich[10], Carmen Ricós[12], Mireia Sansalvador[2] and Margarita Simón Palmada[13]

[1] Laboratori Clínic Barcelonès Nord i Vallès Oriental (ICS-Metropolitana Nord), Badalona, Spain
[2] Laboratori Clínic Bages (ICS-Catalunya Central), Manresa, Spain
[3] Laboratori Clínic Hospital Verge de la Cinta (ICS-Terres de l'Ebre), Tortosa, Spain
[4] Laboratori Clínic Berguedà (ICS-Catalunya Central), Berga, Spain
[5] Laboratori Clínic Hospital Arnau de Vilanova (ICS-Lleida), LLeida, Spain
[6] Servei de Bioquímica Clínica Hospital Germans Trias i Pujol (ICS-Metropolitana Nord), Badalona, Spain
[7] Laboratori Clínic L'Hospitalet (ICS-Metropolitana-Sud), L'Hospitalet, Spain

tions that are more robust than the preliminary ones proposed in a previous study by the same group.
Methods: The yearly average was recorded for each indicator in each laboratory, the yearly interlaboratory median was calculated, and the changes occurring were studied to determine their continuity in the 5-year period. For each indicator, the average of the yearly medians was calculated and the results transformed to the Six Sigma scale to estimate the degree of control over the related process. It was suggested to establish the yearly interlaboratory median as the desirable specification for each indicator.
Results: The medians for most indicators were stable during the period studied. Thus, the specifications proposed in the first study were considered robust in these cases. The Six Sigma statistic provided added value in this study because it enabled detection of processes that should be improved, in which case the specifications proposed were considered provisional despite their stability. After identifying processes that have the greatest impact on patient safety, the group set a specification of 0%, regardless of the actual specification obtained, although the members are conscious of the difficulty in attaining this level of quality. Certain processes that

Westgard QC


Sigma Metrics of Common Processes (US) and Laboratory Processes (Italy)
6
5
4
3
2
1
0
4.3
2.3
5.2
5.6
4.1
4
3.7
3.6
Airline Safety
Baggage handling
Departure Delays
Missing data on input tests
Insufficient sample volume
Hemolyzed specimens
Inadequate anticoag ratio
Unacceptable IQC
Unacceptable EQA
Sources: Quality Indicators in Laboratory Medicine: Experience of a Large Laboratory. L. Sciacoelli, A. Aita A. Padoan. M. Plebani, Abstract 0962, IFCC World Lab Istanbul
Westgard QC
9

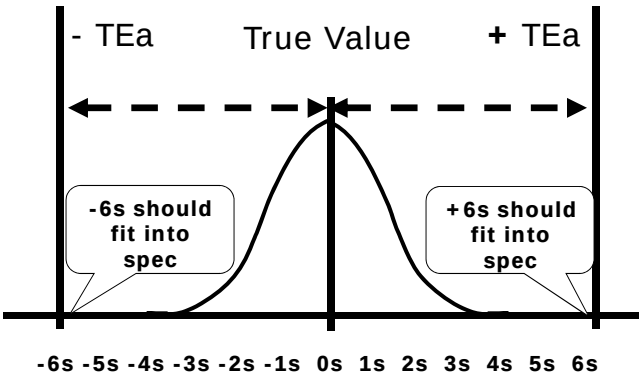

Sigma Metrics of Laboratory Processes (Romania)
6
5
4
3
2
1
0
5.3
5.6
4.8
4.8
4.2
4
4.5
3.5
Patient ID Errors
Missing data on input tests
Samples lost
Inadequate anticoag ratio
Hemolyzed specimens
Clotted samples
LDH QC
Sodium QC
Sources: Quality Indicators in the Preanalytical Phase of Testing in a Stat Laboratory, Grecu DS, Vlad, DC, Dumitrascu V, Lab Medicine Winter 2014, i45:1:74-81
Westgard QC
10

Outline of the Talk

- Do we need to worry about quality?
- A brief introduction to Six Sigma
  - Counting defects: How does healthcare perform?
- **Calculating Sigma-metrics**
  - Setting Goals for Quality
  - Measuring Performance
  - Examples of Current Performance
- Tools for Sigma-metrics
  - Sigma-metric Equation
  - Method Decision Chart

Westgard QC

Six Sigma and Total Allowable Error:



Westgard QC

Quality requirements: many options (and Milan 2014)

- First choice: clinical outcome studies (evidence-based, but only applicable and available for a few analytes)
- Second best: Biologic-derived goals ("Ricos goals") [available for many analytes but now seen as flawed – see next presentation]
- Last choice: Everything else ("Best" state of the art)
  - RCPA (Royal College of Pathologists of Australasia)
  - PT and EQA goals
  - CLIA PT criteria
  - RiliBÄK (Germany)

Westgard QC

Three Glucose methods: Are they acceptable?

Glucose Total Allowable Error = 10% (CLIA)

Method A	Method B	Method C
CV = 2.3	CV = 1.9	CV = 1.9
Bias = 2.1	Bias = 4.2	Bias = 0

Westgard QC

14

From Tolerance Limits to Total Allowable Error (TEa)

- **TEa in the literature**

  Bias + 1.65 SD or Bias + 2SD
  (1974) Westgard, Carey, Wold

  Bias + 3SD (1991) Laessig and Ehrmeyer

  Bias + 4SD (1991) Westgard and Burnett

  Bias + 6SD (2001) "Six Sigma"

Westgard QC

How do we measure Sigma performance for analytical tests?

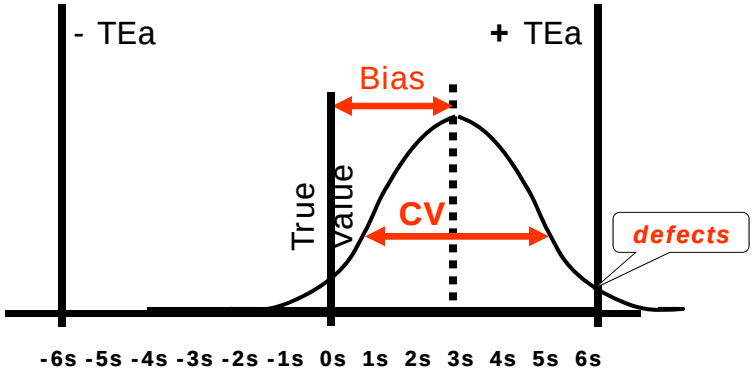
Measure Variation

- Do we measure imprecision (CV)?
- Do we measure inaccuracy (bias)?

Westgard QC

Tool #1: Sigma metric equation for analytical process performance

Sigma-metric = $(TE_a - Bias)/CV$



Westgard QC

17

Example Sigma-metric Calculation

- 3 levels of a cholesterol study, Clin Chem July 2014
  - CLIA PT criterion for acceptability = 10%
  - Total Precision (CV): 1.0% 0.9% 1.0%
  - Bias : 3.0% 2.5% 2.3%

- Sigma = (10 – 3) / 1.0
  = 7.0 / 1.0
  = 7.0
- Sigma = (10-2.5) / 0.9
  = 7.5 / 0.9
  = 8.3
- Sigma = (10 – 2.3) / 1.0
  = 7.7 / 1.0
  = 7.7

Average Sigma = (7.0 + 8.3 + 7.7) / 3 = 7.67

Westgard QC

Is quality consistent across all labs and manufacturers? What does the Data say?

- **Big Picture: recent data comparing instrument performance**
- Case studies: what individual lab studies can tell us
- Tools for Assessment and Assurance
  - Sigma-metric Equation
  - Method Decision Chart
  - OPSpecs Chart

Westgard QC

Comparison of 6 Competitors on 8 chemistry analytes

- 20 patient serum samples
- Comparison against reference methods or all-method-trimmed-mean

"Additionally, large laboratory effects were observed that caused interlaboratory differences >30%."

"There is a need for improvement even for simple clinical chemistry analytes. In particular, the interchangeability of results remains jeopardized by assay standardization issues and individual laboratory effects."

Clinical Chemistry 60:6
855–863 (2014)

General Clinical Chemistry

Measurements for 8 Common Analytes in Native Sera Identify Inadequate Standardization among 6 Routine Laboratory Assays

Hedwig C.M. Stepman,[1] Ulla Tikkainen,[2] Dietmar Stöckl,[3] Hubert W. Vesper,[4] Selvin H. Edwards,[4] Harri Laitinen,[2] Jonna Pelanti,[2] and Linda M. Thienpont,[1]* on behalf of the participating laboratories[5]

BACKGROUND: External quality assessment (EQA) with commutable samples is essential for assessing the quality of assays performed by laboratories, particularly when the emphasis is on their standardization status and interchangeability of results.

METHODS: We used a panel of 20 fresh-frozen single-donation serum samples to assess assays for the measurement of creatinine, glucose, phosphate, uric acid, total cholesterol, HDL cholesterol, LDL cholesterol, and triglycerides. The commercial random access platforms included: Abbott Architect, Beckman Coulter AU, Ortho Vitros, Roche Cobas, Siemens Advia, and Thermo Scientific Konelab. The assessment was done at the peer group level and by comparison against the all-method trimmed mean or reference method values, where available. The considered quality indicators were intraassay imprecision, combined imprecision (including sample-matrix interference), bias, and total error. Fail/pass decisions were based on limits reflecting state-of-the-art performance, but also limits related to biological variation.

RESULTS: Most assays showed excellent peer performance attributes, except for HDL- and LDL cholesterol. Cases in which individual assays had biases exceeding the used limits were the Siemens Advia creatinine (−4.2%), Ortho Vitros phosphate (8.9%), Beckman Coulter AU triglycerides (5.4%), and Thermo Scientific Konelab uric acid (6.4%), which lead to considerable interassay discrepancies. Additionally, large laboratory effects were observed that caused interlaboratory differences of >30%.

CONCLUSIONS: The design of the EQA study was well suited for monitoring different quality attributes of assays performed in daily laboratory practice. There is a need for improvement, even for simple clinical chemistry analytes. In particular, the interchangeability of results remains jeopardized both by assay standardization issues and individual laboratory effects.

Performing accurate and precise measurements that are comparable over time and location and across assays is essential for ensuring appropriate clinical and public health practice. One step toward achieving this goal is using assays that are metrologically traceable to a higher-order reference measurement system or harmonized by use of internationally recognized procedures (*1*, *2*). In Europe, the European Union Directive on in vitro diagnostic medical devices requires demonstration of metrological traceability (*3*). Thus, in principle, laboratories using CE-marked assays consisting of calibrator, reagent, and instrument from the same manufacturer (so-called homogeneous systems) can assume accuracy and interchangeability of their measurement results. However, the intrinsic quality of a manufacturer's assay or test system might be confounded by the laboratory using the system. Therefore, an independent assessment of the quality of measurements obtained under routine conditions is essential for ensuring optimal patient care and public health. External quality assessment (EQA),[6] also called proficiency testing, plays a key role in this regard (*4*, *5*). To cover the broad spectrum of sources potentially invalidating the quality of measurements, the assessment schemes need to fulfill certain requirements in terms of

[1] Laboratory for Analytical Chemistry, Faculty of Pharmaceutical Sciences, Ghent University, Ghent, Belgium; [2] Labquality, Helsinki, Finland; [3] ITI-Consulting, Hombeke, Belgium; [4] Centers for Disease Control and Prevention (CDC), Division of Laboratory Sciences, Atlanta, GA.
* Address correspondence to this author at: Laboratory for Analytical Chemistry, Faculty of Pharmaceutical Sciences, Ghent University, Harelbekestraat 72, B-9000 Ghent, Belgium; Fax +32 9 264 81 98; e-mail linda.thienpont@ugent.be.

Received December 12, 2013; accepted March 12, 2014.
Previously published online at DOI: 10.1373/clinchem.2013.220376

[5] See online supplemental data for a list of all participating laboratories.
[6] Nonstandard abbreviations: EQA, external quality assessment; AMTM, all-method trimmed mean; REF, reference methods; SI, Système International d'Unités; TE, total error.

Sigma evaluation of results

Test	A	B	C	D	E	F
Cholesterol	7.67	2.55	3.42	4.25	5.69	3.46
Creatinine	5.7	7.35	5.62	3.58	4.58	5.56
Glucose	4.81	3.96	4.34	5.09	4.71	4.17
HDL	6.56	11.42	11.96	11.29	10.01	10.51
LDL	5.41	n/a	n/a	5.16	3.72	4.06
Phosphate	6.67	6.71	0	3.46	4.82	n/a
Uric Acid	6.98	12.09	15.23	5.68	5.2	6.43
Triglycerides	10.43	5.42	14.18	18.15	8.32	8.02

Average Sigma-metric calculated of 3 levels measured
Approximately 10 labs for each instrument
CLIA goals used

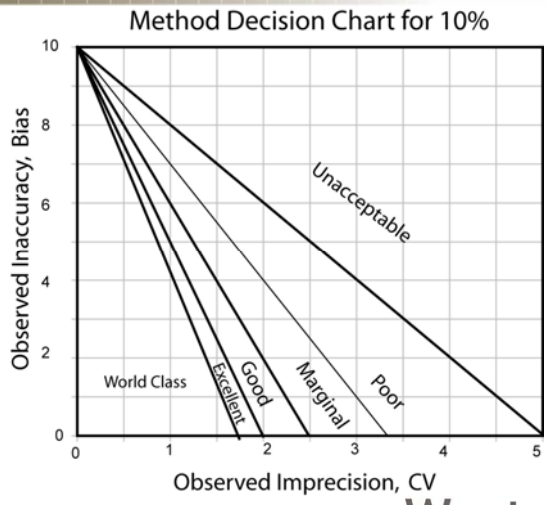
Westgard QC

Standardization Conclusion

- Given conditions: achieving >6-Sigma performance or highest performance among competitors:
  - A: **6 of 8** analytes
  - B: **4 of 7** analytes
  - C: **3 of 7** analytes
  - D: **3 of 8** analytes
  - F: **3 of 7** analytes
  - E: **2 of 8** analytes

Westgard QC

A quick non-technical description of Sigma-metric Decision Charts



Free spreadsheet available at westgard.com downloads

Westgard QC

23

2013 Chemistry Analyzer

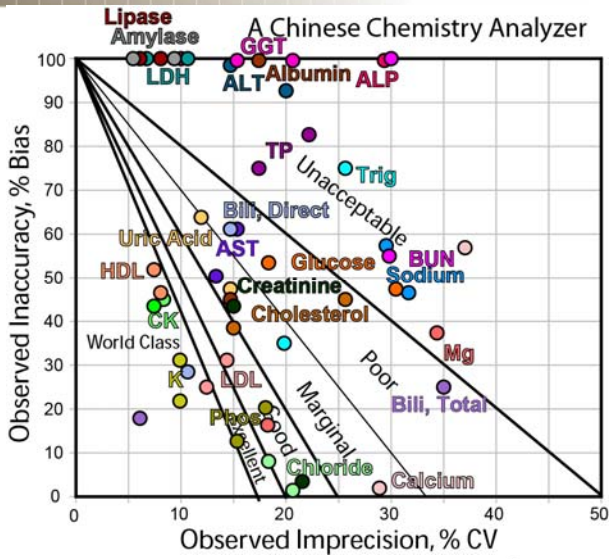
Evaluation of the XXXXXX Automated Chemistry Analyzer. Hyo-Jun Ahn, Hye-Ryun Kim, and Young-Kyu Sun, *J Lab Med Qual Assur* 2013;35:36-46.

Assay	TEa	Level 1	CV%	Bias%	TEa	Level 2	CV%	Bias%
Albumin	**10%**	4.55	1.45%	4.37%	**10%**	3.79	1.75%	10.26%
Alk Phos	**30%**	92.1	8.78%	286.24%	**30%**	366.4	3.19%	281.66%
AST	**20%**	44.2	3.1%	12.12%	**20%**	165.7	2.72%	10.13%
ALT	**20%**	29.4	4.0%	18.65%	**20%**	122.3	2.88%	19.93%
Amylase	**30%**	262.8	2.54%	137.03%	**30%**	836.2	1.59%	124.70%
Bilirubin, Direct	**44.5%**	0.39	4.84%	12.93%	**44.5%**	1.44	6.34%	27.51%
Bilirubin, Total	**63.5%**	0.63	4.2%	12.63%	**20.0%**	2.73	7.19%	5.04%
Calcium	**15.48%**	6.46	4.38%	0.36%	**10.54%**	9.49	3.9%	6.17%
Cholesterol	**10%**	219.8	1.53%	3.88%	**10%**	167.5	2.61%	4.63%
Creatinine Kinase (CK)	**30%**	110	2.17%	13.07%	**30%**	372.2	2.65%	13.71%
Chloride	**5%**	98.9	1.03%	0.05%	**5%**	92.8	0.92%	0.42%
Creatinine	**25.21%**	1.19	3.97%	10.92%	**15%**	4.07	3.25%	0.29%
GGT	**22.11%**	27.9	4.82%	32.92%	**22.11%**	70.4	3.65%	32.76%
Glucose	**10%**	61.4	1.86%	6.53%	**10%**	208.1	3.1%	4.76%
HDL	**30%**	76.1	2.53%	13.94%	**30%**	65.1	2.34%	15.66%
Potassium	**18.05%**	2.77	1.7%	5.56%	**11.42%**	4.38	1.0%	2.55%
Sodium	**2.65%**	151	0.9%	1.24%	**2.84%**	140.6	0.9%	1.64%
Phosphate	**10.7%**	2.73	2.0%	2.29%	**10.7%**	5.42	1.6%	1.31%
Total Protein	**10%**	7.13	1.8%	7.59%	**10%**	5.48	2.3%	8.26%
Triglycerides	**25%**	88.1	5.0%	8.78%	**25%**	57	6.7%	18.98%
Urea Nitrogen	**9%**	13.53	2.8%	19.85%	**9%**	39.89	2.7%	5.0%
Uric Acid	**17%**	3.32	2.4%	8.14%	**17%**	7.03	2.1%	10.85%
LDH	**20%**	246.4	1.2%	136.14%	**20%**	480.7	2.3%	127.49%
Magnesium	**25%**	0.58	8.8%	9.70%	**25%**	1.36	4.6%	4.46%
LDL	**20%**	140.1	2.5%	4.65%	**20%**	106.8	2.9%	6.31%
Lipase	**37.44%**	38.6	2.4%	53.91%	**37.44%**	5.9	2.6%	233.50%

Westgard QC

24

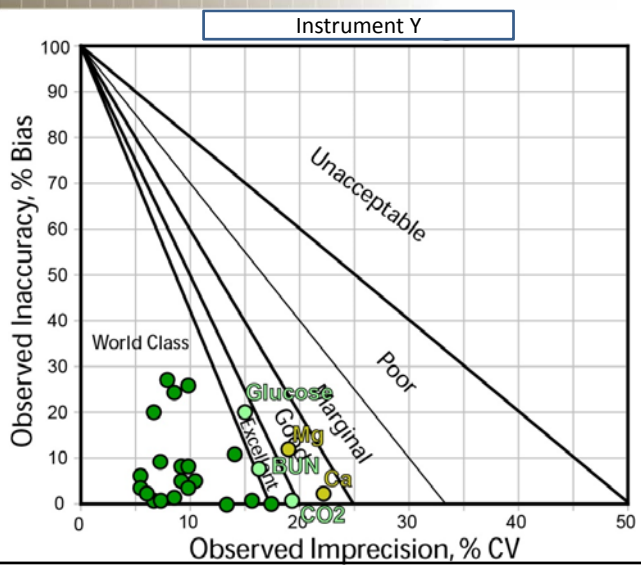
Display of Sigma-metrics:
Normalized Method Decision Chart (26 tests)



Display of Sigma-metrics:
Instrument Y

23 of 25 tests are >5 Sigma

92% of assays



Kern Valley Hospital District
(25 methods)
>6 Sigma:
Sodium
Potassium
Chloride
Creatinine
Total Protein
Albumin
Alk Phos
ALT
AST
Total Bilirubin
Direct Bilirubin
Amylase
Lipase
CK
Phosphorus
Iron
Uric Acid
Cholesterol
Triglycerides
HDL

Three Glucose methods: Are they acceptable?

Glucose Total Allowable Error = 10% (CLIA)

Method A	Method B	Method C
CV = 2.3	CV = 1.9	CV = 1.9
Bias = 2.1	Bias = 4.2	Bias = 0

Sigma A:
= (10 – 2.1) / 2.3
= 7.9 / 2.3
= 3.4

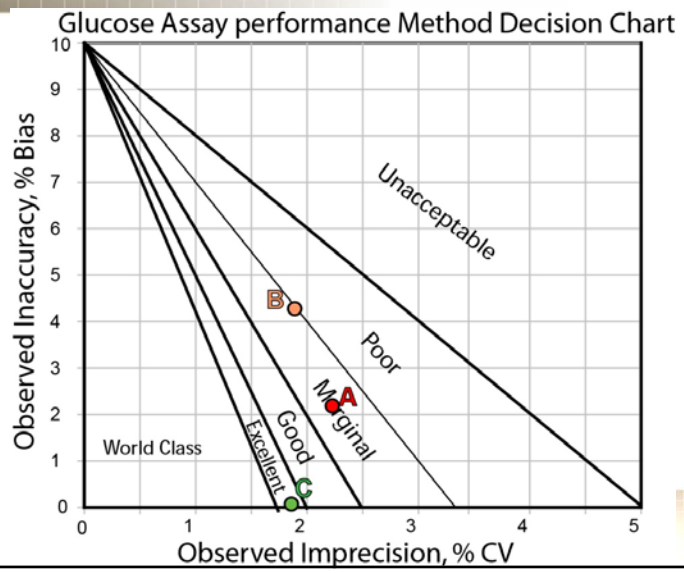
Sigma B:
= (10 – 4.2) / 1.9
= 5.8 / 1.9
= 3.05

Sigma C:
= (10 – 0) / 1.9
= 10 / 1.9
= 5.26

Westgard QC

27

MEDx Chart for Glucose Examples



28

Summary of Sigma-Metrics for Evaluation of Quality

- Sigma-metrics *(concept of hitting the target)*
- Quality Requirements *(size of the target)*
- Method Performance Data *(did we hit it?)*
- *Now what do we do? ACCEPT OR REJECT THE METHOD*
- *Even better: DETERMINE THE RIGHT QC!*

Westgard QC