Defining Analytical Performance Specifications

Matthias Orth

IFCC Committee on Clinical Laboratory Management - http://www.ifcc.org/ifcc-education-division/emd-committees/c-clm/

Satellite Educational Workshop on Intelligent Clinical Laboratory Management: Impacts on Quality System Improvement

Hilton Durban - October 22, 2017

How do others rate our performance in laboratory medicine services?

Table 1. Percentage and Aggregate Number of Ratings

Laboratory Service Category*	Excellent, % (No.)	Good, % (No.)	Average, % (No.)	Below Average, % (No.)	Poor, % (No.)
Quality/reliability of test results	45.6 (1939)	42.9 (1823)	9.8 (416)	1.3 (56)	0.4 (15)
Staff courtesy	50.4 (2069)	37.1 (1523)	9.7 (398)	2.2 (89)	0.6 (25)
Accessibility of pathologist	51.7 (1823)	34.1 (1201)	11.5 (406)	2.1 (74)	0.7 (23)
Accessibility of laboratory manager	46.5 (1524)	36.0 (1178)	13.7 (449)	2.7 (87)	1.2 (38)
Phlebotomy services	37.8 (1313)	43.7 (1515)	14.4 (501)	3.1 (106)	1.0 (35)
Test menu adequacy	36.7 (1427)	46.9 (1826)	14.0 (543)	1.7 (68)	0.7 (26)
Accessibility of laboratory staff	47.3 (1913)	36.5 (1475)	12.4 (500)	2.7 (111)	1.1 (44)
Courier services	38.0 (1039)	41.1 (1124)	15.6 (428)	3.2 (87)	2.1 (57)
Routine test TAT	33.5 (1389)	44.7 (1855)	17.0 (704)	3.4 (142)	1.4 (56)
Laboratory management responsiveness	40.4 (1380)	40.1 (1372)	14.4 (492)	3.6 (123)	1.5 (51)
Inpatient stat test TAT	36.7 (1177)	41.7 (1338)	15.0 (480)	4.4 (142)	2.2 (71)
Critical value notification	44.3 (1833)	39.3 (1624)	11.4 (470)	3.1 (128)	1.9 (79)
Clinical report format	33.7 (1396)	46.0 (1905)	15.5 (644)	3.1 (127)	1.7 (71)
Outpatient stat test TAT	33.6 (1170)	40.3 (1407)	17.4 (605)	6.2 (216)	2.6 (89)
Esoteric test TAT	17.1 (629)	38.0 (1398)	32.9 (1212)	8.9 (328)	3.1 (116)

* TAT indicates turnaround time.

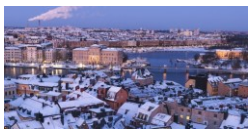
4329 respondents

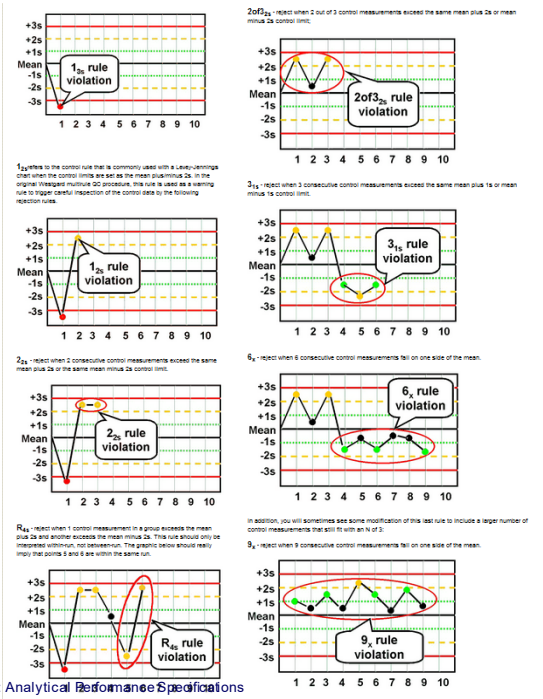
responsibility for processes out of the laboratory

Arch Pathol Lab Med. 2009;133:38–43

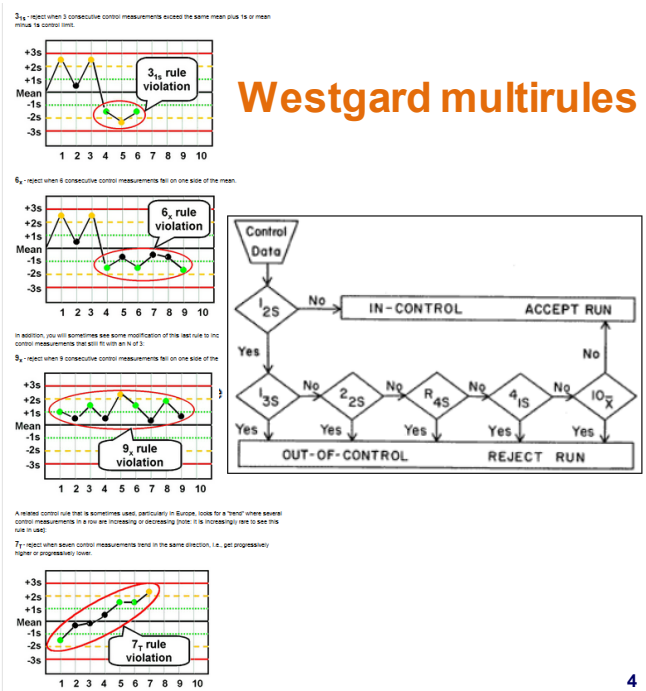
Analytical quality criteria to be covered

- performance criteria for daily routine quality controls
- performance criteria for EQAS
- performance criteria for tests with numeric as well as for alpha-numeric results
- use of reference method values and/or method specific values for EQAS
- optional: quality specifications for calculated tests





Westgard multirules



Variation of test results

1. preanalytic variation
2. analytical variation (imprecision and bias)
3. biological variation within a single subject

A Study of the Accuracy and Precision of Clinical Chemistry Determinations in 170 Canadian Laboratories

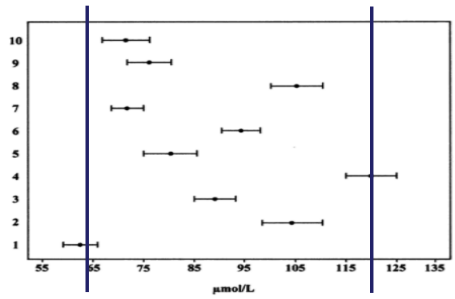
David B. Tonks

Clinical chemistry. 1963;9(2):217-33

¼ reference interval

Table 1.7 Some Studies on the Within-Subject Biological Variation of Serum Sodium and Serum Urea (CV, %)

No. of subjects	Time span (weeks)	Sex	Sodium	Urea	Country
11	2	M	0.7	12.3	Denmark
10	4	M	0.9	14.3	USA
10	8	M	0.6	9.5	Germany
14	8	F	0.5	11.3	Germany
9	12	M	1.4	13.6	USA
11	15	M	0.6	15.7	Denmark
37	22	M	0.5	11.1	England
15	40	M&F	0.7	13.9	Scotland



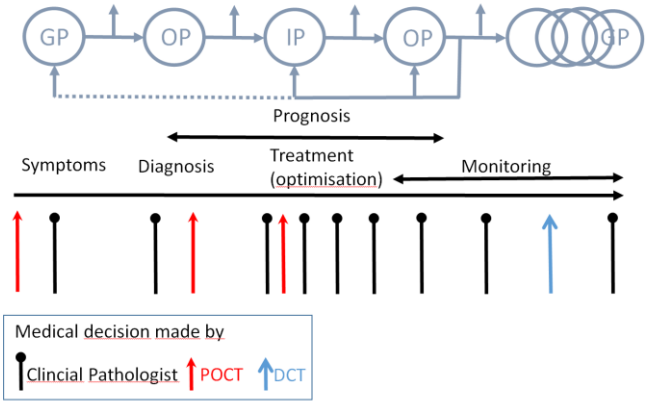
CG Fraser: Biological Variation. From Principles to Practice 2001

Figure 1.10 Mean Values and Absolute Ranges of Serum Creatinine in Four Samples Taken from Each of 10 Apparently Healthy Men.
The age and sex matched reference interval for men aged 18–55 years is 64–120 µmol/L.[5]

1999 Stockholm consensus conference statement hierarchy of models to set analytical quality specifications

1. Evaluation of the effect of analytical performance on clinical outcomes in specific clinical settings
2. Evaluation of the effect of analytical performance on clinical decisions in general:
   a. Data based on components of biological variation
   b. Data based on analysis of clinicians' opinions
3. Published professional recommendations:
   a. From national and international expert bodies
   b. From expert local groups or individuals
4. Performance goals set by:
   a. Regulatory bodies
   b. Organisers of EQA schemes
5. Goals based on the current state of the art:
   a. As demonstrated by data from EQA or Proficiency Testing schemes
   b. As found in current publications on methodology

Identical performance criteria in real labs, POCT and DCT ?



Lenters-Westra E, Slingerland RJ. Six of Eight Hemoglobin $_{A1c}$ Point-of-Care Instruments Do Not Meet the General Accepted Analytical Performance Criteria. Clinical Chemistry 2010;56(1):44

Diagnosis and monitoring: $CV_{analytical} < 0.5\ CV_{within\text{-}subject}$

Screening: $CV_{analytical} < 0.5 \sqrt{CV_{within\text{-}}^2 + CV_{between\text{-}subject}^2}$

Elevitch FR, et al. Am J Clin Pathol 1979:71:624

Challenges of HTA/outcome studies for diagnostic procedures

Qualifying performance testing in the medical laboratory by HTA is a yet unresolved challenge Reid, M. C., M. S. Lachs, et al. (1995). JAMA **274**: 645-51

General concept of laboratory medicine which only delivers data to the attending physicians such as the presence or absence of a certain disease. Most meta-analyses for diagnostic test studies still pool diagnostic sensitivity and sensitivity values only Willis, B. H. and M. Quigley (2011). BMC Med Res Methodol **11**: 27

diagnostical and analytical performance goals of a certain laboratory test might even have to be defined for different clinical situations and have to be revised in specified intervals thereafter Sandberg, S., and Thue, G. Scand J Clin Lab Invest. 1999;59:531

„Evidence on current practice indicates that clinical practice has changed to such a degree that the original research question is no longer relevant to UK practice" Czoski-Murray, C., M. Lloyd Jones, et al. (2012). Health Technol Assess **16**(50): i-xvi, 1-159.

Challenges of a general acceptance of the Stockholm criteria

Recommendations not widely introduced because data not available for many tests or concept not applicable (e.g. graphical presentation of titers, numerical + alphanumerical results, extreme analytical ranges)

In particular in immunoassays and mass-spectrometry, data highly dependent on method / matrix

Most data on biological validation on "simple Clinical Chemistry tests“

Skipping too many (complex) tests by giving no recommendations at all

2014 Milan consensus conference statement hierarchy of models to set analytical quality specifications

1. Evaluation of the effect of analytical performance on clinical outcomes in specific clinical settings (very few analytes)
2. Evaluation of the effect of analytical performance on clinical decisions in general:
   a. Data based on components of biological variation (scrutinizing data)
3. Other goals
   a. From national and international expert bodies
   b. From expert local groups or individuals
   c. Regulatory bodies
   d. Organisers of EQA schemes
   e. As demonstrated by data from EQA or Proficiency Testing schemes
   f. As found in current publications on methodology

Pre-Analytical and Post-analytical Performance Goals - TBD

http://www.efcclm.eu/files/efcc/2%20CCLM-Consensus%20Statement.pdf

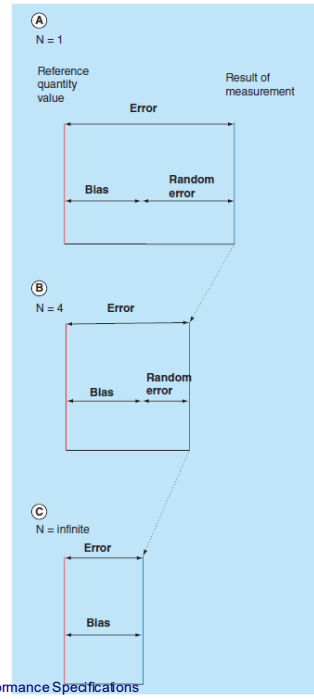
Measurement of „true“ value and correct medical interpretation of test result „*(selecting the correct language)*“

Test result has deviation from „true“ value (total analytical error TAE or permissible uncertainty (pU))

pU consists of dispersion of results **(„random error“)** and systematic deviation from „true“ value, called **„bias“**

Preanalytic effects lead to

- Gross errors (e.g. sample mixup)
- Unsuitable results (e.g. wrong timing of TDM or in provocation test)
- Systematic in- or decrease of result caused by instability of analyte or by interference (hemolysis), unpredictable instability by recentrifugation of gel tubes or barricor tubes



components of error (random and systematic (bias) error) of
(A) a single result of measurement,
(B) the mean of four replicate measurements and
(C) the mean of infinite number of measurements, which eliminates the random error component

Bioanalysis (2014) 6(21), 2855–2875

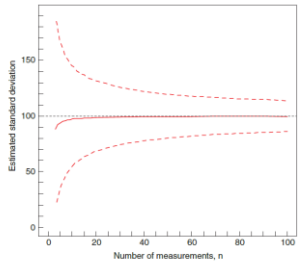
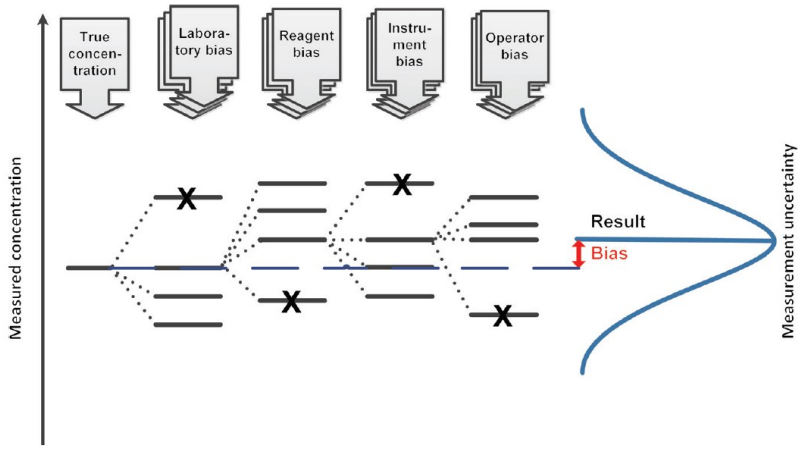


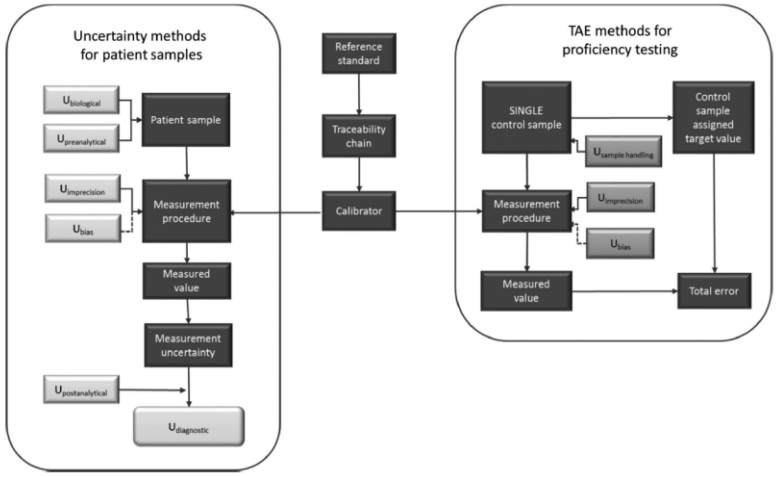
Figure 1: Confidence limits (95%) for an estimated standard deviation (broken lines), obtained by Monte Carlo simulation (20,000 replicates per n value).
Limits were calculated as described in ref. [7], equation 6.102, p. 344). Horizontal axis: Number of values from which the standard deviation is calculated. Drawn line: mean estimated standard deviation.

bias may be indistinguishable from imprecision



CCLM https://doi.org/10.1515/cclm-2017-0341

pU methods for patient samples and TAE methods for proficiency testing



target value is defined for the proficiency testing sample, which is used for calculating **error**

In patient samples, **uncertainty** methods estimate the confidence we can have in the measurement result for the purpose of diagnosis

Proficiency testing and measurement uncertainty are related through the **traceability chain** to the reference standard

CCLM https://doi.org/10.1515/cclm-2017-0341

Legal framework for performance criteria

FDA
clinical validity (accuracy with which test identifies, measures, or predicts presence or absence of a clinical condition or predisposition in a patient)

CLIA
safety and effectiveness of the test system. does not address the clinical validity of any test

J Lab Med 2015; 39(1): 26–69 DE GRUYTER

On 19 September 2014, the current version of the "Guideline of the German Medical Association on Quality Assurance in Medical Laboratory Examinations" was published. It featured an introduction by the German Medical Association.

Revision of the "Guideline of the German Medical Association on Quality Assurance in Medical Laboratory Examinations – Rili-BAEK"

J Lab Med 2015; 39(1): 26–69

Legal background behind RiliBÄK

EU IVD directive

German Medical Devices Act ("Medizinproduktegesetz")

German Medical Devices Operator Ordinance ("Medizinproduktebetreiberverordnung")

German Medical Association ("Bundesärztekammer")

RiliBÄK
every professional employing laboratory tests in human healthcare is obliged to comply to all regulations specified in RiliBÄK

part A (the description of a quality management system closely resembling DIN EN ISO norm 15189 as a framework for structural quality) **(GROSS ERROR)**

part B with extensive appendices covering analytical performance goals in internal as well as in external quality programs in tabulated form for 84 selected quantitative and 50 semiquantitative tests in hematology, hemostaseology, clinical chemistry, TDM, endocrinology, serology in different matrices (such as serum, plasma, whole blood, urine, cerebrospinal fluid) as well as for genetical and microbiological tests and sperm analysis **(RANDOM and SYSTEMATIC ERROR)**

J Lab Med 2015; 39(1): 26–69

Special Part B1: Quantitative tests in medical laboratories

1. Principles of quality assurance
2. 1. Minimum requirements are listed that need to be met to asses the quality of quantitative results of examinations in medical laboratories.
3. 2. All quantitative tests performed by medical laboratories are subject to IQC.
4. 3. All measurands listed in table B1 a to c are subject to EQA

J Lab Med 2015; 39(1): 26–69

Special Part B1: 2. Carrying out quality assurance

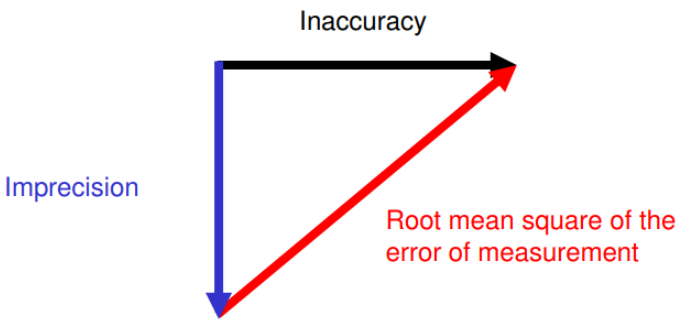
1. Internal quality assurance

1. Carrying out individual measurements of control samples
2. Evaluating the results of the individual measurements of control samples
3. Calculating and evaluating the root mean square of the error of measurement after completing a control cycle.
4. Establishing internal laboratory limits of permissible error for measurands that are not listed in Table B1
5. Point-of-care testing with unit-use reagents
6. Measurands with small test frequencies
7. Documentation

2. External quality assurance (round robin test)

J Lab Med 2015; 39(1): 26–69

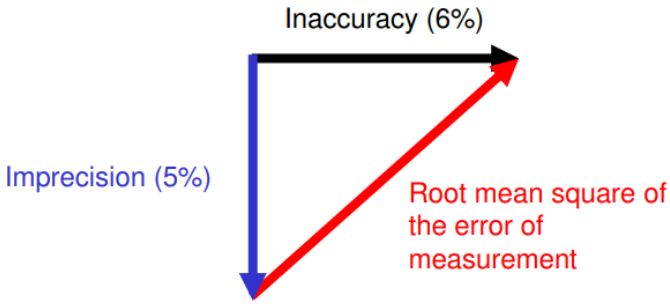
principle: root mean square of the error of measurement

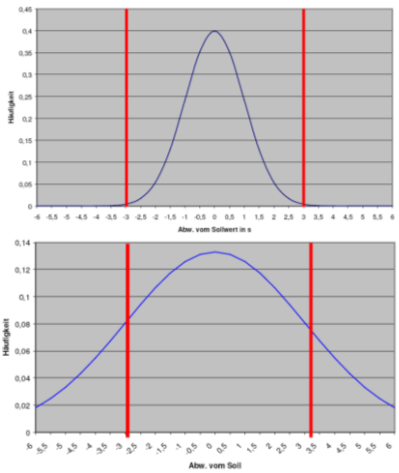


$a^2 + b^2 = c^2$

Pythagoras of Samos (570 BC – 510 BC)

error limits





old RiliBÄK
Total error: 2 x 5% + 6% = 16 %

New RiliBÄK Root mean square of the error of measurment: $\sqrt{5^2 + 6^2}$ = 7.8 %

Calculation of root mean square of measurement deviation (RSMD)

$$\Delta_{max} = \sqrt{k^2 * s_{ep}^2 + \delta_{ep}^2}$$

k=3, coverage factor for calculating the internal laboratory deviation limits
s_{ep}, empirical standard deviation of the control sample measurements used in the calculations during the pre-evaluation period
δ_{ep}, systematic deviation of measurement of the control sample measurements used in the calculations during the evaluation period (ep)

Procedure for non-tabulated tests with new control samples (new control cycle)

Process for repeated failures of column 3 at the end of control cycles („event“ according to § 2 Medical Products Safety Plan Ordinance)

Open discussion whether different analytical performance standards might be acceptable between real laboratory tests and point of care tests

Mueller, C., A. Scholer, et al. (2004). N Engl J Med **350**: 647-54
Straseski, J. A., M. E. Lyon, et al. (2011). Clin Chem **57**: 1566-73
J Lab Med 2015; **39**: 26–69

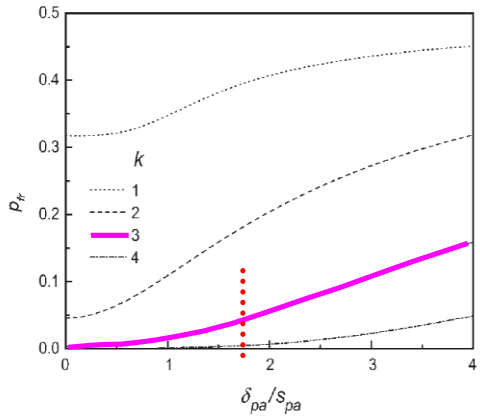


Figure 2 False rejection probability as a function of the systematic error, expressed in multiples of the standard deviation, both observed during a pre-analytical period. Parameter is the coverage factor k.

D

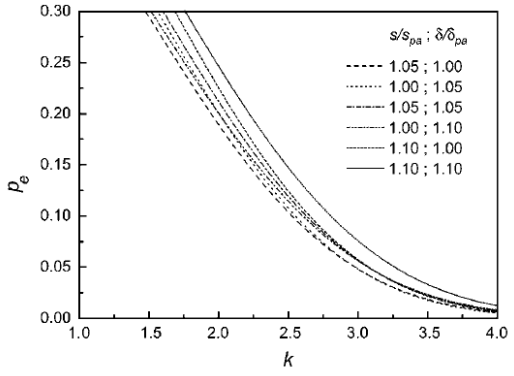


Figure 3 Error detection probability vs. coverage factor *k* for different size and type of error, $\delta_{pa}/s_{pa} = 1.7$.

empirical: $d_{pa}/s_{pa} = 1.7$

Macdonald, R. (2006). J Lab Med **30**: 111-7

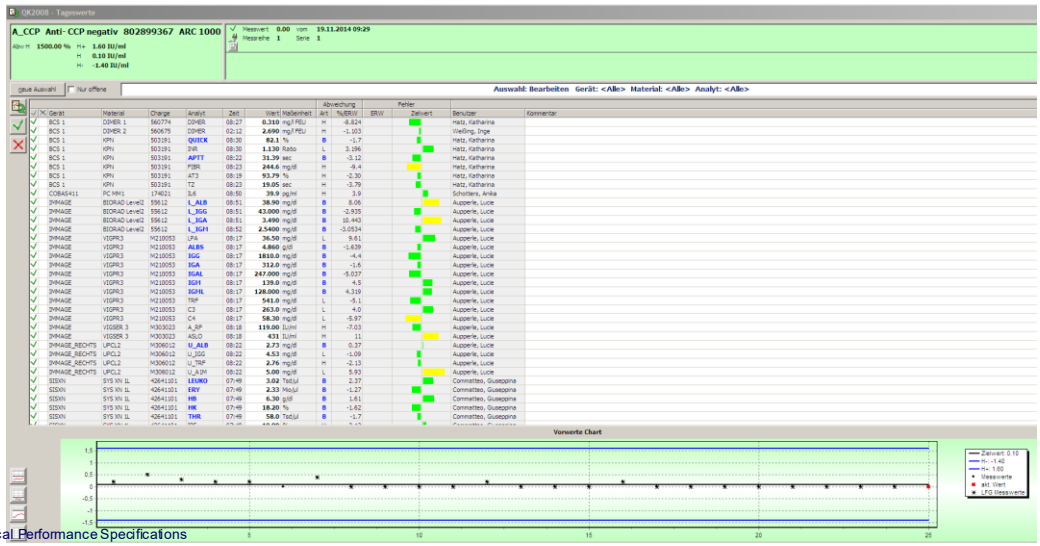
reference method (RMV) or method-specific consensus value (NV)

1 No.	2 Measurand	3 Permissible relative deviation of a single result or the relative root mean square, respectively	4 Rili-BAEK applicable concentration intervals of columns 3 and 5			5 Permissible relative deviation in EQA	6 Type of target value in EQA
			From	To	Unit		
38	Lactate	11.0%	9	90	mg/dL	18.0%	NV
			1	10	mmol/L		
39	Lactate dehydrogenase (LDH) EC 1.1.1.27	9.0%	100	700	U/L	18.0%	RMV
			1.67	11.7	µkat/L		
40	Leucocytes	6.5%	2	30	10^9/L	18.0%	RMV
41	Lithium	6.0%	0.3	3.5	mmol/L	12.0%	RMV
42	Magnesium	7.5%	0.3	3.5	mmol/l	15.0%	RMV
43	Sodium	3.0%	110	180	mmol/l	5.0%	RMV
44	pCO_2	7.5%	≤35		mmHg	12.0%	NV
		6.5%	>35				

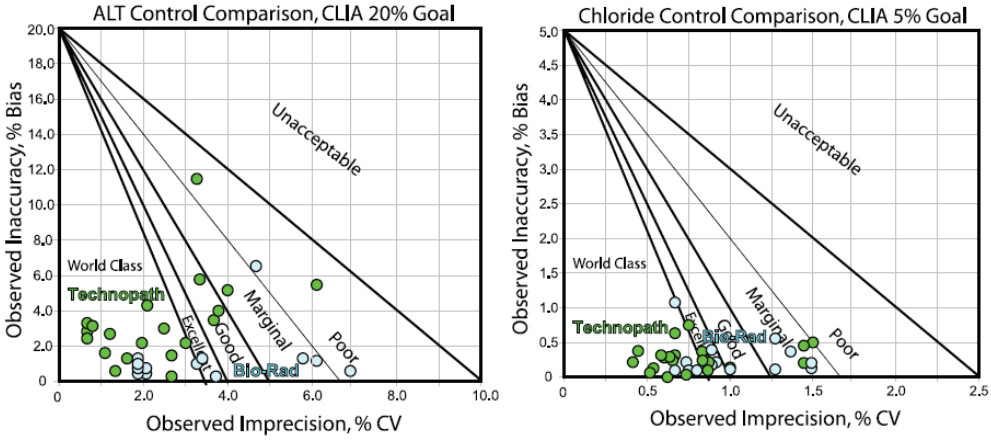
Selection of IQC quality control material based on RiliBÄK specifications (!) (range, target value assignment)

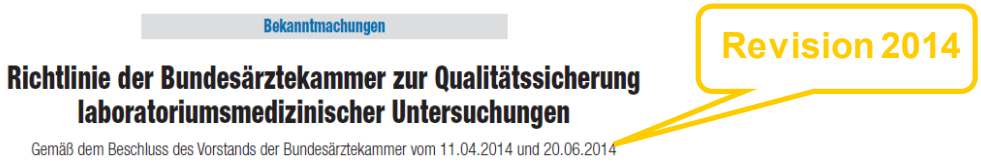
J Lab Med 2015; 39(1): 26–69

- instant assessment of analytical control samples and detection of critical deviations by operator
- automatic calculation of RMSMD is integrated into all major lab information systems



Six Sigma Performance of BioRad and Technopath controls for ALT (left) and Chloride (right)





performance criteria have to be revised in a timely and controlled process

- Alkaline Phosphatase: RMSD reduced from 13% to 11%; EQAS reduced from 21 to 18%
- CA 19-9 replaced by CA 15-3
- FSH added
- Lipase deleted
- pCO2: goals made more complex (2 levels)
- FT4: goals simplified (1 level)
- Transferrin: RMSD reduced from 9.5% to 8.0%; EQAS reduced from 15% to 12%
- FT3: RMSD reduced from 14.5% to 13.0%; EQAS reduced from 24% to 20%
- Vancomycin: EQAS reduced from 21% to 18.0%

J Lab Med 2015; 39(1): 26–69

Quality performance specifications: Challenge of calculated results

(e.g. anion gap, ratios, eGFR)

1. **Error propagation in formulas consisting of test results, constants and estimated factors**
2. **Linearity of uncertainties**
3. **Probability density function of single pU factor? (rectangu triangular, normal, U-form, asymmetrical)**
4. **Reliability of single pU?**
5. **Mathematical model to calculate total pU**

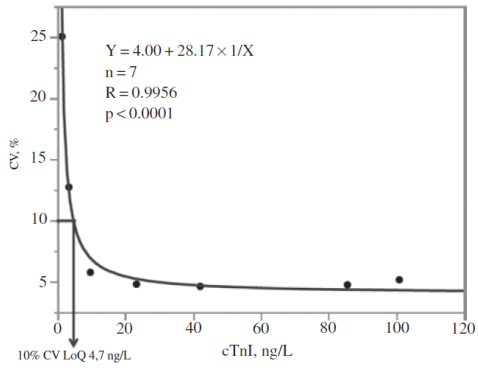


Figure 1: Imprecision profile of the Architect method for cTnI assay. The imprecision profile was obtained in the Authors' laboratory by measuring in 39 different runs seven plasma pools collected from healthy subjects and patients with cardiac disease using three different lots of reagents and calibrators throughout 2 months.

Clerico, A et al. *CCLM 2017*, 55: 1634-51. doi:10.1515/cclm-2016-0933

*G*uide to the Expression of *U*ncertainty in *M*easurement (**GUM**)

Summary of procedures for evaluating and expressing uncertainty components

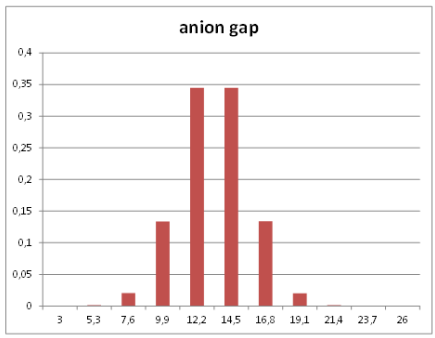
- Specify the *measurand*: (what is being measured and the mathematical functional relationship between the measurand and the input quantities upon which it depends)
- Identify sources of uncertainty
- Identify and correct for systematic error (bias) where possible
- Quantify uncertainty components: determine the standard uncertainty associated with each of the input quantities, including any uncertainty associated with the correction for systematic error. An uncertainty estimate obtained by the statistical analysis of serial observations OR uncertainty estimate obtained by other means (authoritative published report, a calibration certificate, personal experience or a numerical quantity associated with a certified reference material)
- Calculate the value of the measurand: that is, calculate the result of the measurement from the functional relationship which connects the various input quantities to the measurand
- Calculate the combined standard uncertainty of the measurand: that is, calculate the combined standard uncertainty of the measurand from the standard uncertainties (and covariances if present) associated with the various input quantities. These standard uncertainties are combined according to the rules based on the law for the propagation of uncertainties
- Calculate the *expanded uncertainty* of the measurand by applying an appropriate *coverage factor*, *k*. The expanded uncertainty is equal to the combined standard uncertainty of the measurand multiplied by *k*. For medical laboratory applications, *k* is typically given the value of 1.96 (or 2.0). This provides an expanded uncertainty which includes 95.0% (or 95.4%) of the values within the distribution of the measurand. The expanded uncertainty calculated in this manner provides a coverage interval on the assumption that the distribution of the measurand is normal

Monte Carlo Stimulation procedure 'automatically' takes into account any nonlinearities in the functional relationship

- graphical representation of the distribution of the measurand can be obtained directly from the MCS procedure
- significant reduction in the mathematical skills required for most evaluations
- MCS generally provides improved estimates for non-linear models
- MCS provides a coverage interval corresponding to a stipulated coverage probability (normal distribution, 95% for coverage factor of 1.96 or 95.4% for coverage factor of 2.0. For asymmetric distributions the shortest 95% coverage interval is quoted)

Clin Biochem Rev 35 2014 37

reference value	14,5
n	1043088
mean	14,50
STD	2,27
Kurtosis	0,00
2.5%-Percentile	10,05
97.5%-Percentile	18,94



$$u^2(AG) = u^2(Na^+) + u^2(K^+) + u^2(Cl^-) + u^2(HCO_3^-)$$

Thus: $u^2(AG) = (1.2)^2 + (0.10)^2 + (1.5)^2 + (1.2)^2$

$= 1.44 + 0.01 + 2.25 + 1.44$

$= 5.14$

And: $u(AG) = 2.26$, mmol/L (AG standard uncertainty)

Clin Biochem Rev 35 2014 37

Summary

- Lack of outcome-based performance criteria should trigger the use other analytical performance goals lower in hierarchy if widely-accepted both by medical professionals and from the health-economical network
- Performance criteria have to be constituted and revised by medical professionals
- Performance criteria should be established for the complete array of laboratory tests and updated on a regular basis employing different analytical performance goals, in particular goals based on biological variation and the state of the art
- Performance criteria should be mandatory for all tests performed in healthcare (exceptions have to be clearly defined!)
- Results from from EQAS testing schemes can be used in a formalized process to revise performance goals

matthias.orth@vinzenz.de