EFLM
EUROPEAN FEDERATION OF CLINICAL CHEMISTRY AND LABORATORY MEDICINE

Eurachem
A FOCUS FOR ANALYTICAL CHEMISTRY IN EUROPE

Verification of precision and bias

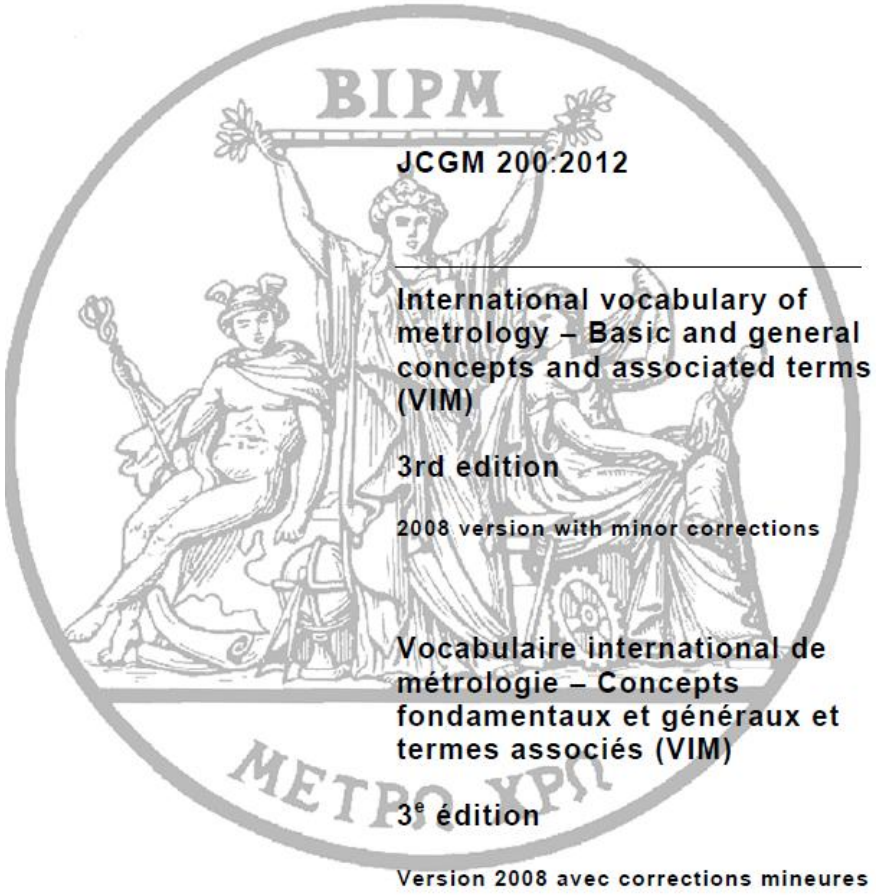
Elvar Theodorsson

Validation and verification of measurement methods

- Procedures aiming at establishing realistic expectations with the analyst and confidence with the end-user that the methods are fit for the intended purposes

Validation

- According to VIM 3, **verification** is “provision of objective evidence that a given item fulfills specified requirements” and
- **Validation** is “verification, where the specified requirements are adequate for the intended use”



Method validation

- Method validation is *a specific kind of validation* “the process of defining an analytical requirement, and confirming that the method under consideration has performance capabilities consistent with what the application requires”
- Method validation includes procedures that both 1) establish the performance characteristics and limitations of a measurement method (e.g., trueness, precision, recovery, linearity, robustness) and 2) establish whether the performance characteristics of the measurement method being investigated are fit for the intended purpose

Method verification

- Procedures to test to what extent the performance data obtained by manufacturers during method validation can be **reproduced** in the environments of end-users
- Possible if the method (reagents, procedure and the measurement instrument) is **manufactured** by a company or other reliable source which has performed proper method validation and who is providing you with the detailed results, a single laboratory method validation is not needed.

Method validation is performed to a varying extent depending on its intended use

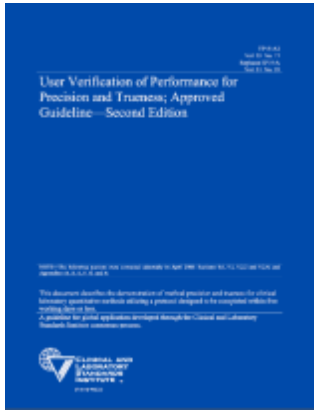
- **Single laboratory method validation** is appropriate where the method is used for a specific purpose in a specific laboratory by personnel with the appropriate training.
- **Full method validation** includes, in addition to the procedures employed in single laboratory validation an interlaboratory study (collaborative study/ collaborative trial) with many measurement instruments several operators etc. The performance characteristics of the measurement method over extended periods of time are also studied in full method validation, including the effects of lot-to-lot variation etc.
- **Full diagnostic method validation** is establishing the diagnostic properties of the method e.g. in health and disease

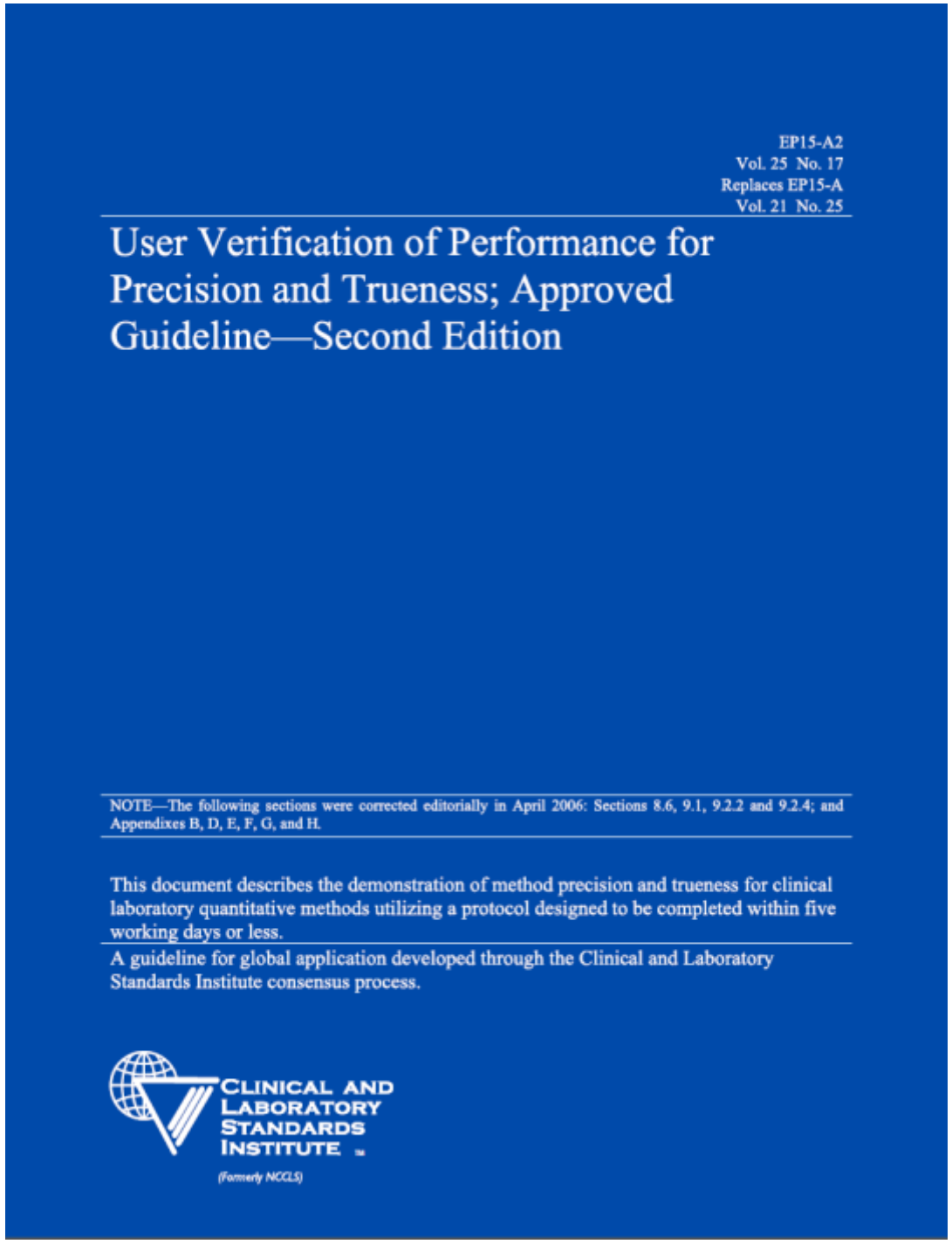
Verification of measurement methods

- In Vitro Diagnostic (IVD) medical devices are in Europe regulated by a third EC Medical Device Directive, the IVD medical device Directive 98/79/EC which has been mandatory in Europa since December 2003
- Local verification practices have commonly been established over time and are frequently influenced by accreditation and certification authorities. Published practices for end-user verification officially endorsed both by the end-users and by the companies have appeared only recently

Verification

- The **EP15-A2** protocol from **CLSI**
  - Uses control material with assigned concentration (e g from external quality control) or certified reference materials
  - Does not test for matrix effects which may occur in patient materials
- Practical and pragmatic method **using patient samples** and common samples for internal quality control
  - Bias is tested by comparison with a well-established methods using at least 20 patient samples
  - Variation within- and between series is measured using the normally used stable materials for internal quality control at least twice daily during two weeks





http://www.clsi.org/source/orders/free/ep15a2f.pdf

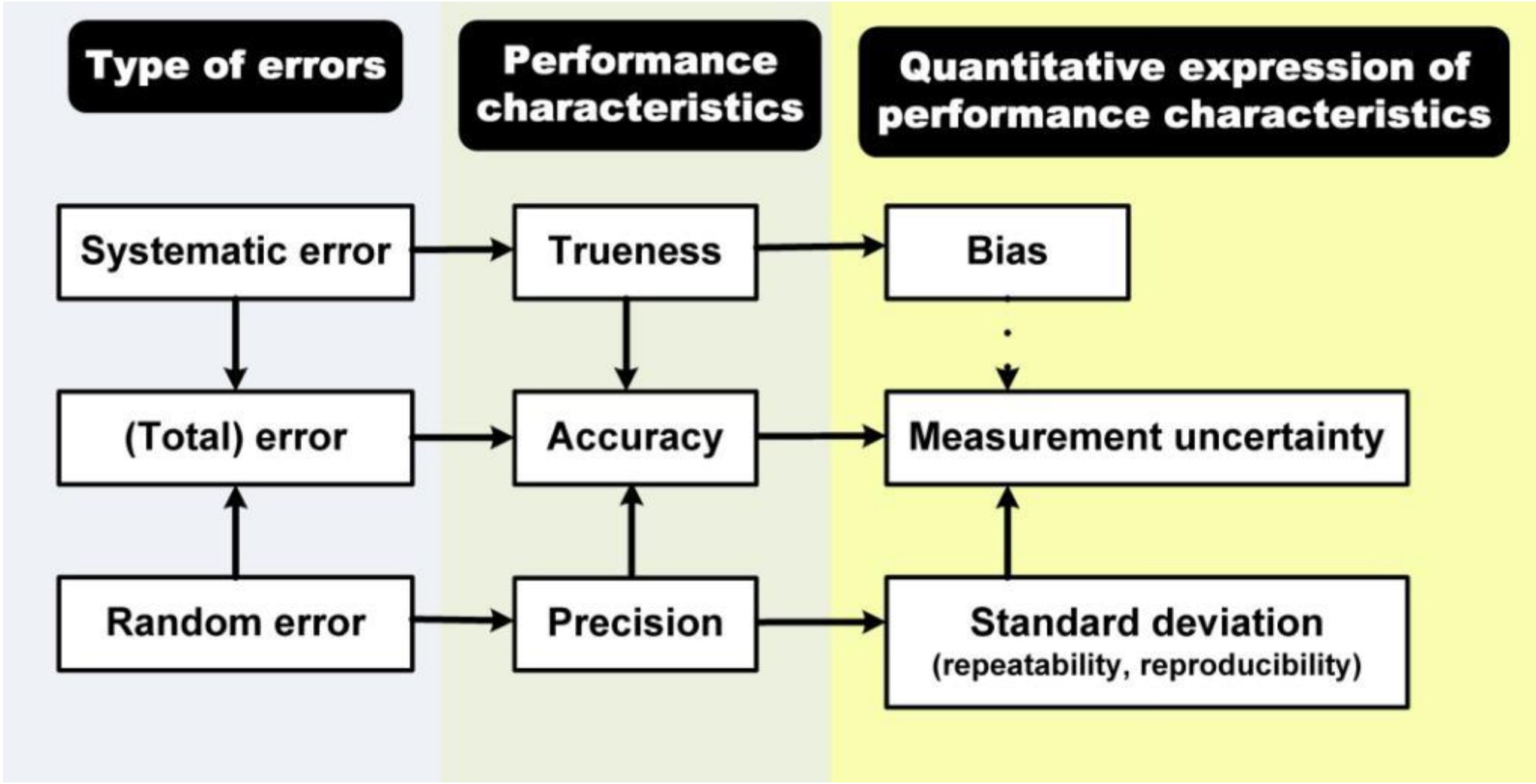
Verification of measurement methods

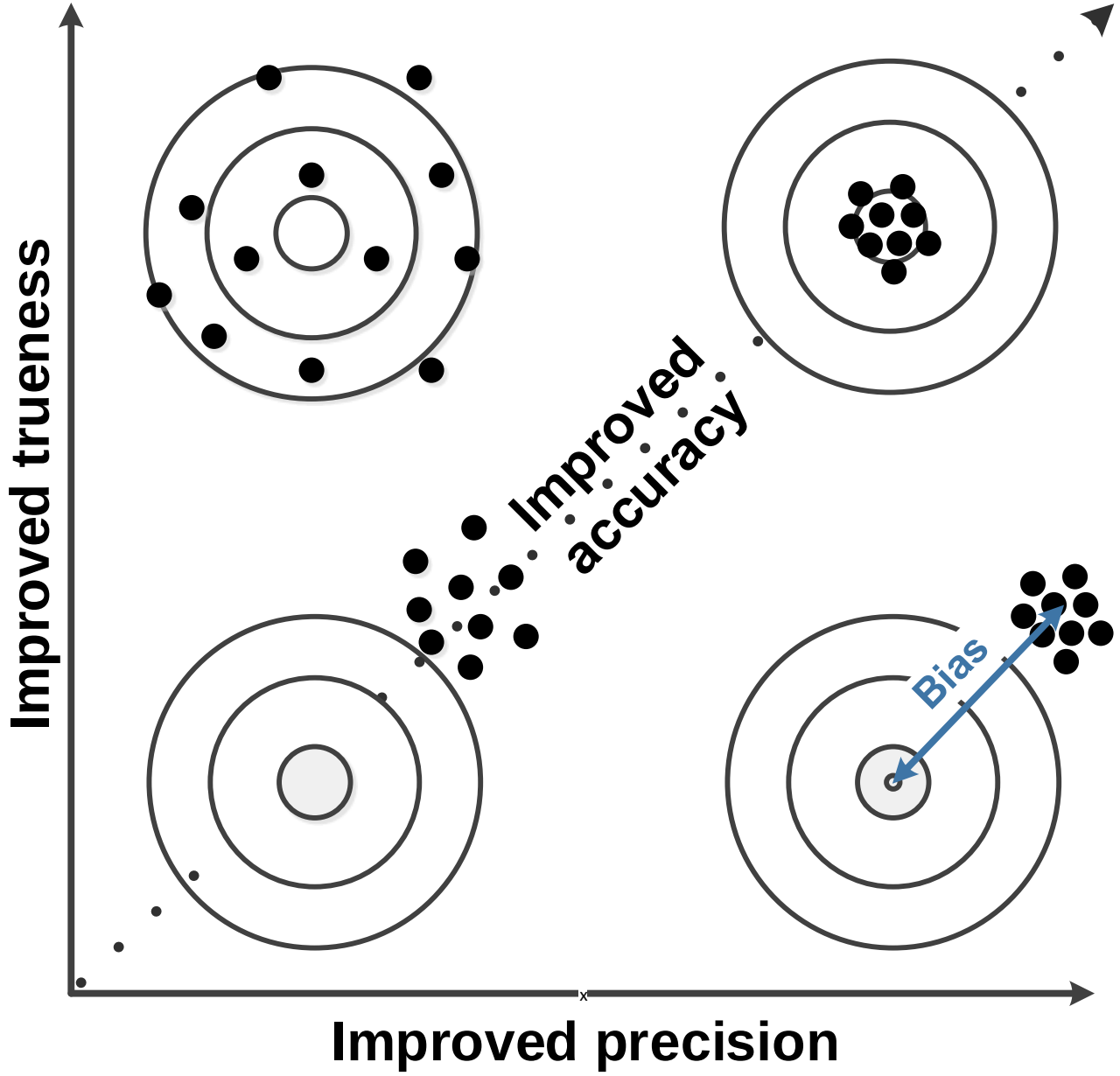
- Clinical laboratories commonly measure in the order of 20 to 200 natural patient samples having as wide concentration range as possible, using both the method being replaced and the new method. At least two pooled patient samples may also be used, and this may actually be an advantage when the medical decision point is close to the detection limit of the method.
- Suitable stable materials for internal quality control are measured at two levels in at least three replicates for at least five consecutive days but preferably at least two weeks for estimating imprecision and for establishing initial control limits for the internal quality control procedures. Linear regression, bias-plot and analysis of variance techniques are used to determine bias, imprecision, matrix effects

Measurement precision/imprecision

- Closeness between indications or measured quantity values obtained by replicate measurements on the same or similar objects under the specified conditions of measurement
- The quantitative expression of precision is the standard deviation (SD) or relative standard deviation (CV/CV %)
- The standard deviation of the estimate of the standard deviation is inversely proportional to the square root of the number of replicates

Harmonised terminology – VIM 3




Improved trueness
Improved precision
Improved accuracy
Bias

Repeatability imprecision

- When the same measurement procedure, same operators, same measuring system, same operating conditions and same location, and replicate measurements on the same or similar objects over a short period of time
- **A short period** of time is usually less than a working day of 8 hours
- Example of repeatability condition is when a stable control material or the same unknown sample is measured repeatedly on the same day
- A prudent and cost effective number of replicate measurements for estimating repeatability imprecision are in the order of 15

Intermediate measurement precision

- When a set of conditions that includes the same measurement procedure, same location, and replicate measurements on the same or similar objects over an extended period of time, but may include other conditions involving changes
- Intermediate measurement imprecision includes variation due to new calibrations, new reagent lots, new operators etc.
- The concept of between-days, between series, inter-series imprecision has earlier been used to describe this type of imprecision

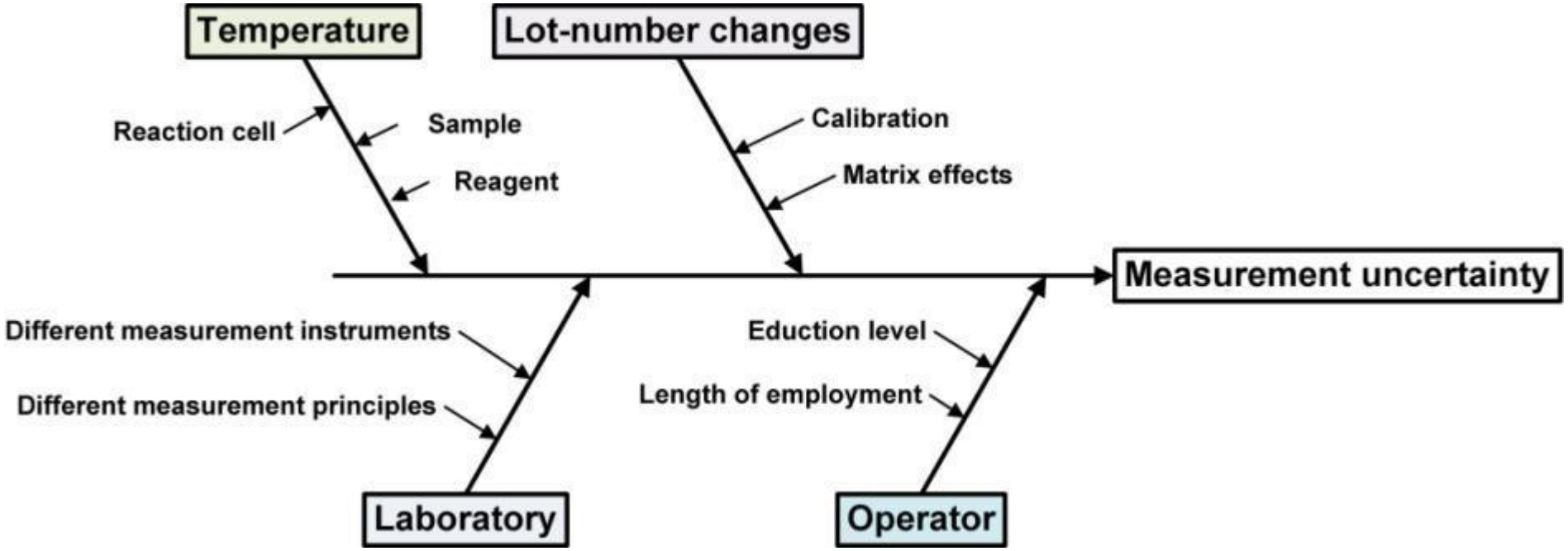
Intermediate measurement precision

- Intermediate imprecision is usually measured using stable control materials in two different concentrations which are measured routinely/daily over extended periods of time for at least 1 year, but preferably during 2-3 years
- It is crucial that all sources of variation included in intermediate imprecision including e.g. lot-number changes are included in sufficient/appropriate number of occurrences

Intermediate and repeatability precision

- If the numbers of results obtained in each series/day are the same, common two-way analysis of variance (ANOVA) can be used to calculate the total SD and its components of SD within and between series. However, as is commonly the case in clinical laboratories, the number of replicate observations in the series is unequal, more advanced ANOVA and **variance component analysis** models catering for unequal number of observations each day/series should be used

Fishbone/cause and effect/Ischikawa diagram



Measurement bias

- Bias in the preparation of the calibrator, including erroneous volume measurements or weighing of calibrators
- Using sample matrix for the calibrators which differs from the matrix in the samples
- Interferences/matrix effects in the samples, e.g. the colour of bilirubin and haemoglobin in icteric and haemolytic samples in laboratory medicine or the presence of high concentrations of lipids or proteins in the sample (hyperlipidaemia or myeloma). Manufacturers commonly use samples from healthy subjects for their validation studies, and the real influence of matrix effects on the methods may be fully evident only when the methods are fully introduced in diagnosing and monitoring seriously ill patients.

Measurement bias

- The presence of molecules in the sample specifically interfering with the reagents used in the measurement process, e.g. heterophilic antibodies (e.g. human antibodies against mouse IgG frequently used in immunoassays).
- Uncorrected loss of measurand at extraction
- Instability of the sample during transport or storage

Determining measurement bias

- Purchasing **certified reference materials** from companies or organizations of high metrological competence and comparing the stated concentration with the concentration your own methods shows
- Comparing the concentrations your method measured in natural samples with the concentrations **a reference method** measured in the same sample
- Participating in programs for **external quality control**. Most of these programs are based on consensus concentrations in modified control samples, but some few are based on comparison to reference methods. The latter are frequently preferable.

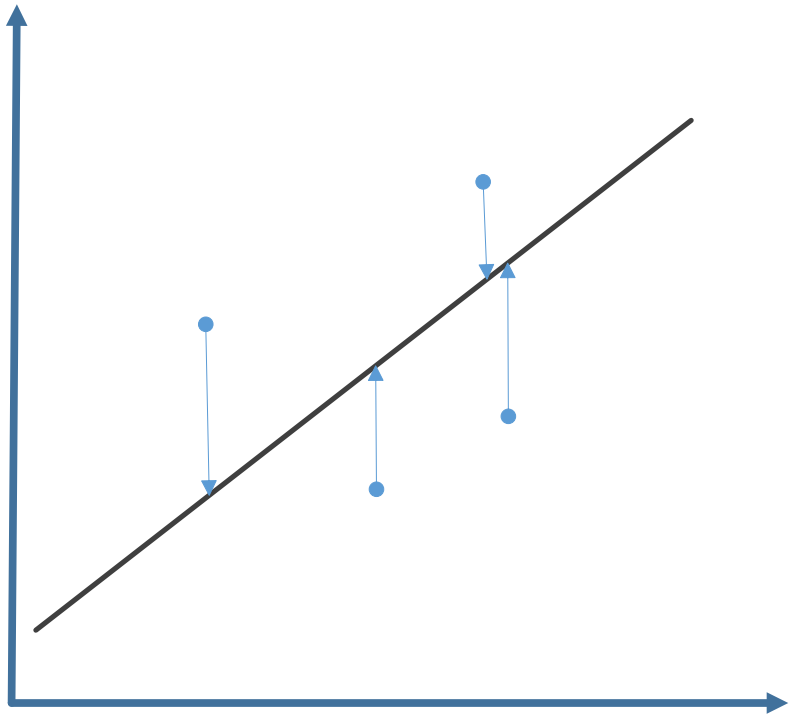
Determining measurement bias

- Measuring the **recovery of the measurand in spiked natural samples**
- Comparing the **serial dilution of a natural sample** or that of a **spiked natural sample with the serial dilution of the calibrator** in the calibration curve

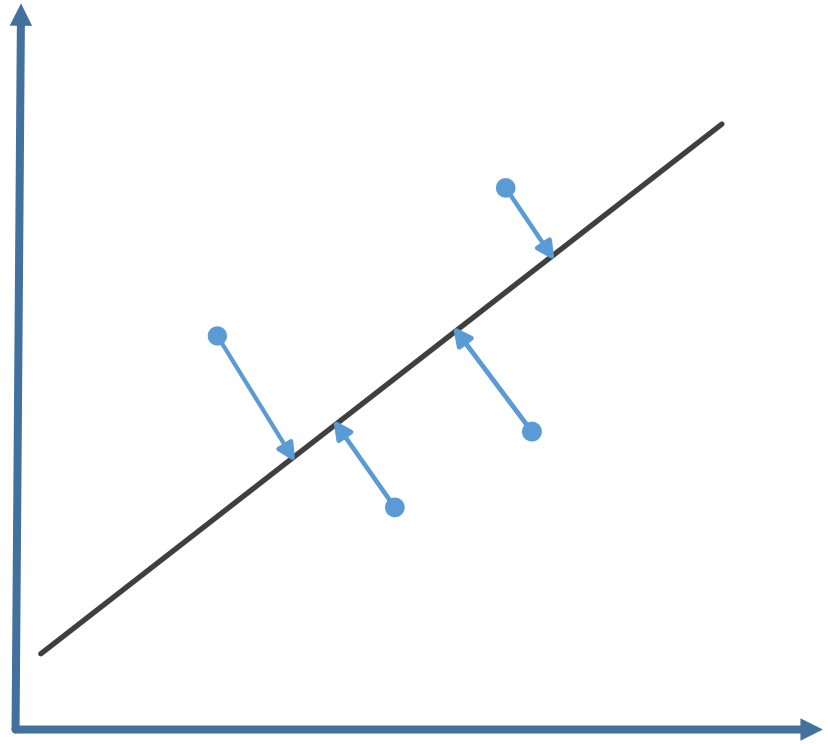
Determining measurement bias

- Making **studies of possible interferences/selectivity**.
- This is evidently very different amongst different measurement methods and fields of study. In laboratory medicine the studies of interferences by bilirubin, haemoglobin, lipids, proteins and drugs are amongst the most important. VIM 3 defines selectivity as “property of a measuring system, used with a specified measurement procedure, whereby it provides measured quantity values for one or more measurands such that the values of each measurand are **independent of other measurands or other quantities** in the phenomenon, body, or substance being investigated”

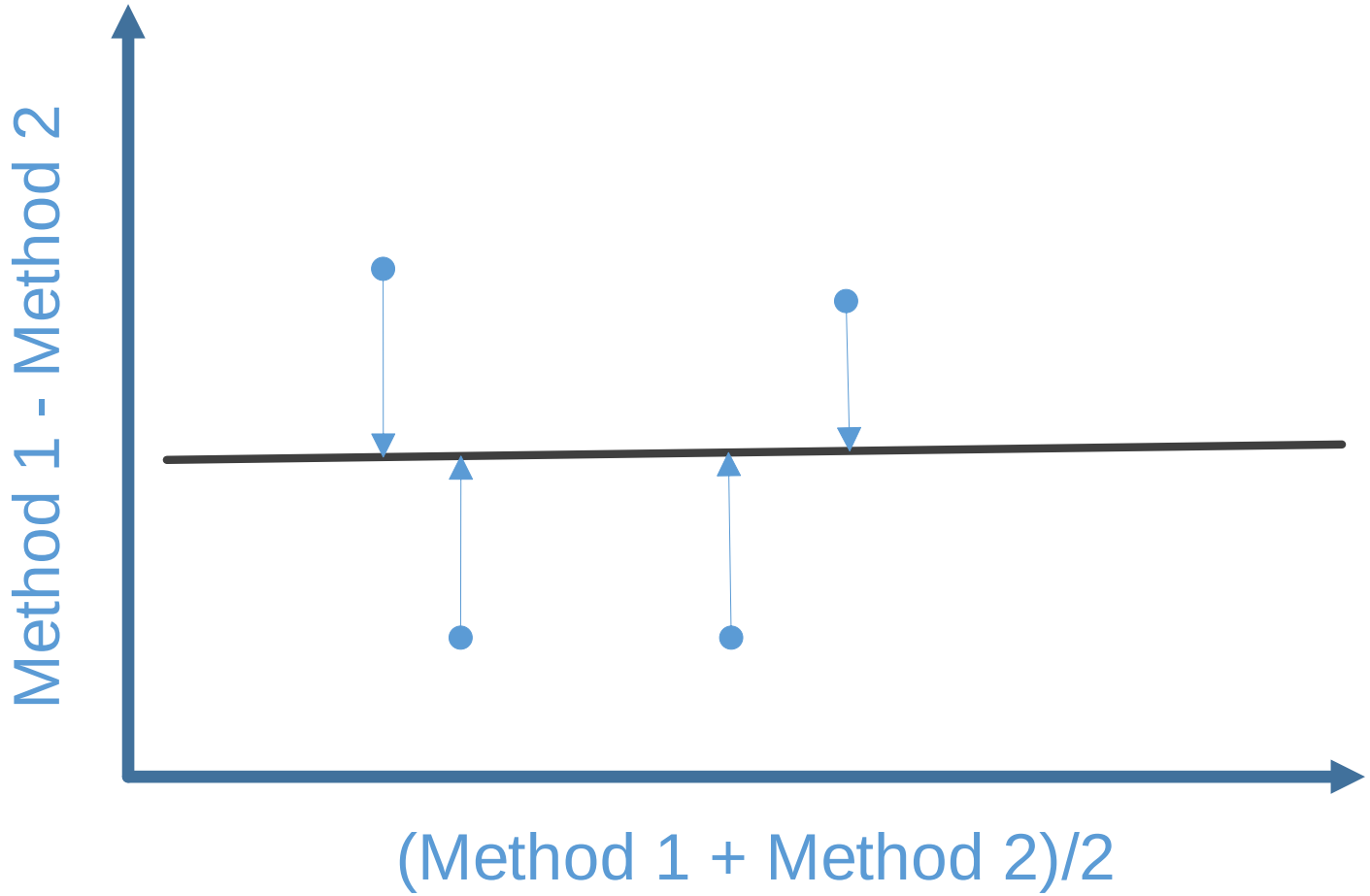
Common linear regression



Orthogonal linear regression



Adcock (1878) = Deming (1943)
Passing Babloc (1988)
Bartlett (1949)



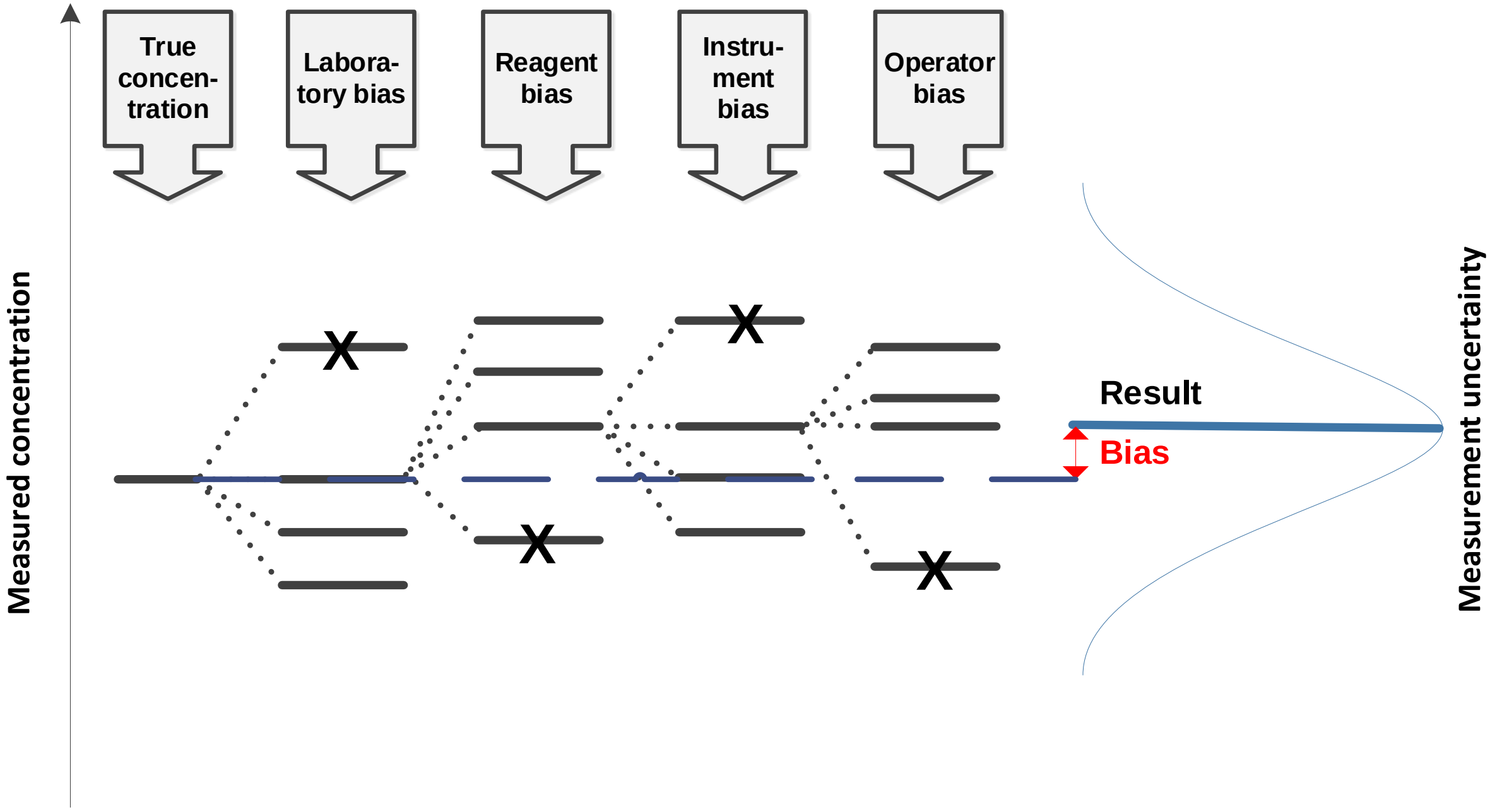
Bias plot
Mean difference plot (Tukey, 1977)
Eksborg (1981)
Bland-Altman (1983)

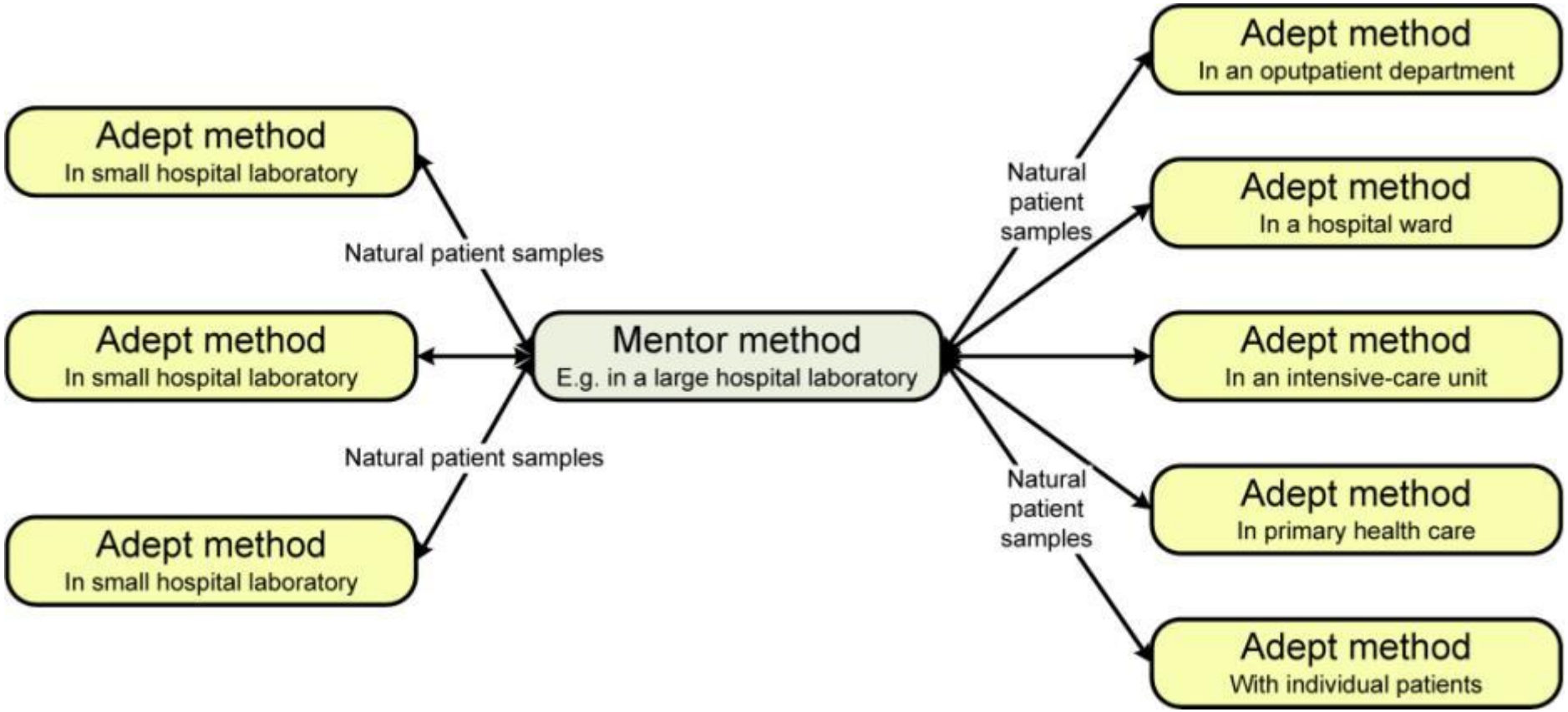
Beware of the use of the correlation coefficient

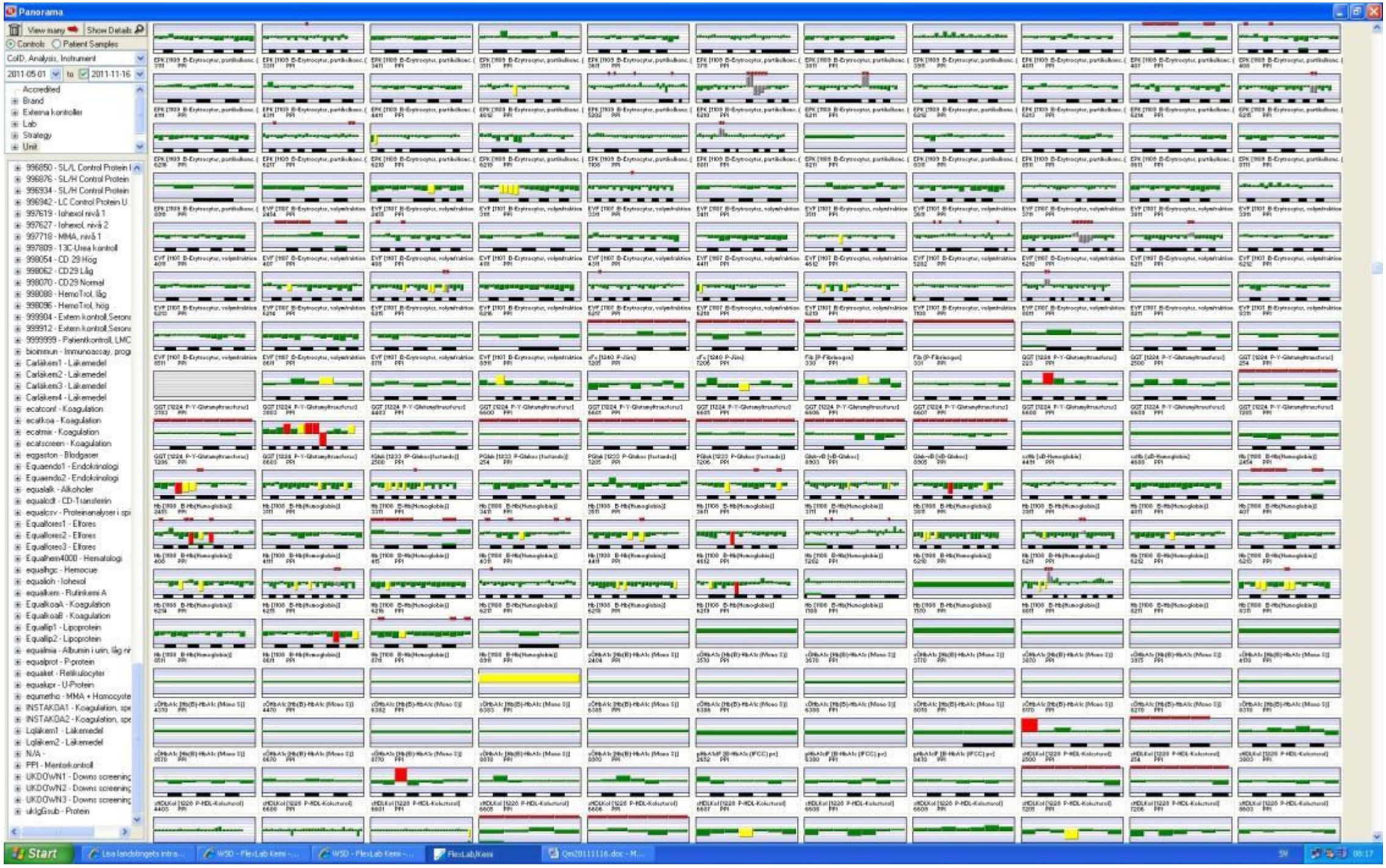
- Any correlation coefficient between two methods can be improved (made closer to 1) by increasing the range of concentrations measured.

A conglomerate of laboratories

- In healthcare, the samples from a patient are over time likely to be measured in different laboratories using different methods as the patient visits primary care and different levels of hospital care.

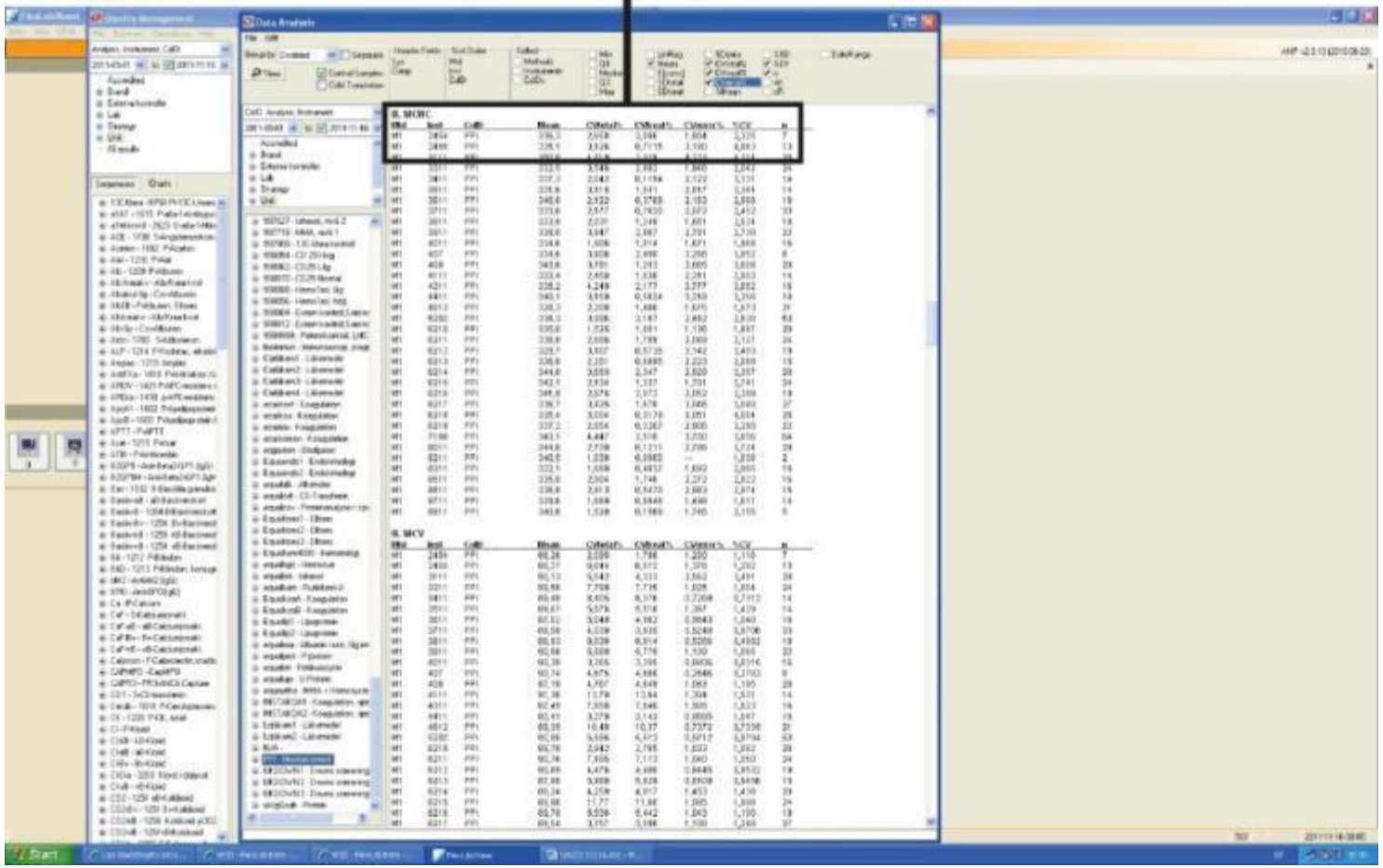

True concen-tration
Labora-tory bias
Reagent bias
Instru-ment bias
Operator bias
Measured concentration
Result
Bias
Measurement uncertainty

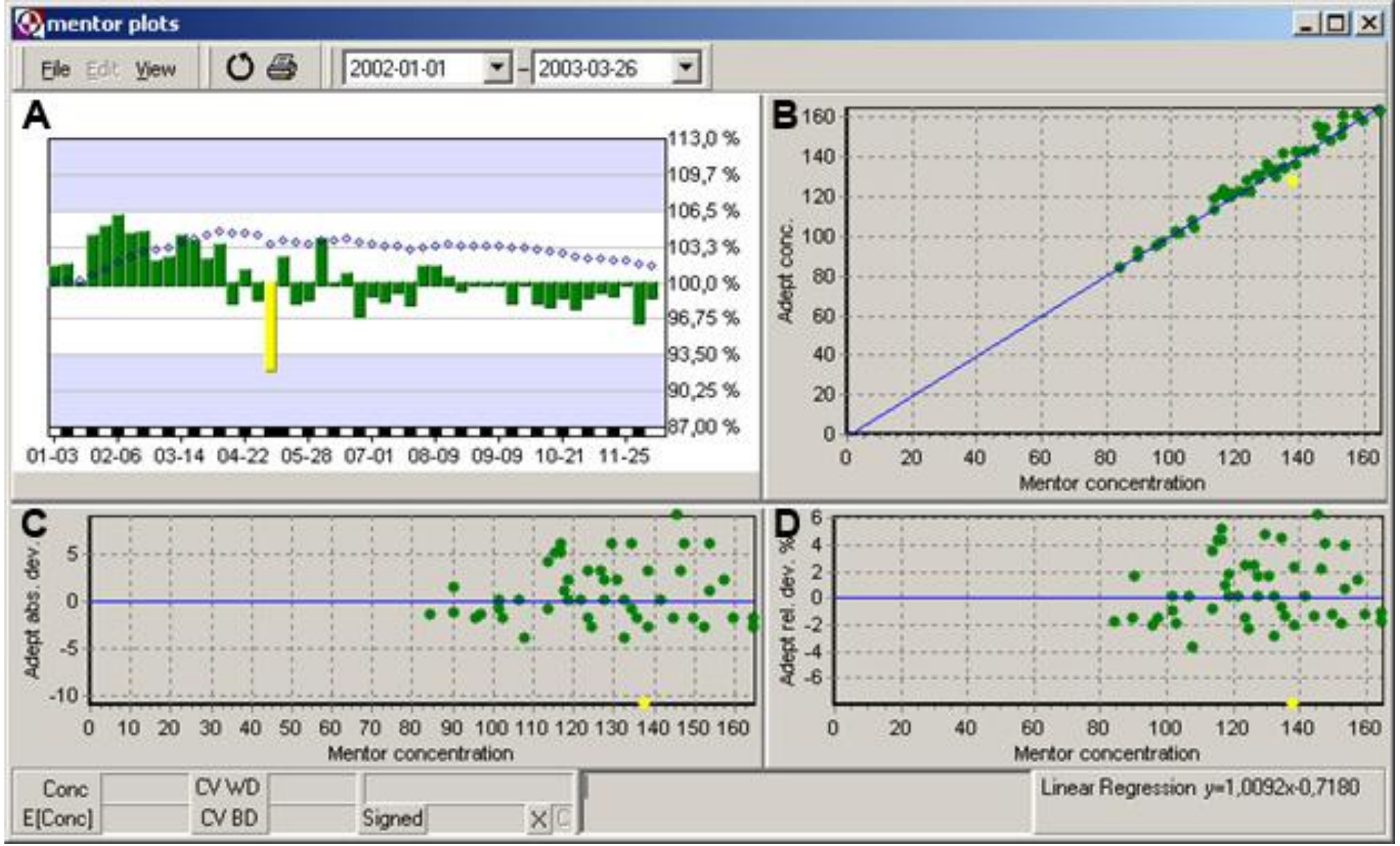

Adept method
In small hospital laboratory
Natural patient samples
Adept method
In small hospital laboratory
Mentor method
E.g. in a large hospital laboratory
Natural patient samples
Adept method
In small hospital laboratory
Natural patient samples
Adept method
In an oputpatient department
Adept method
In a hospital ward
Adept method
In an intensive-care unit
Natural patient samples
Adept method
In primary health care
Adept method
With individual patients


Panorama
View many
Show Details
Controls
Patient Samples
CoID, Analysis, Instrument
2011-05-01
to
2011-11-16
Accredited
Brand
Externa kontroller
Lab
Strategy
Unit
996850 - SL/L Control Protein I
996876 - SL/H Control Protein
996934 - SL/H Control Protein
996942 - LC Control Protein U
997619 - Iohexol nivå 1
997627 - Iohexol, nivå 2
997718 - MMA, nivå 1
997809 - 13C-Urea kontroll
998054 - CD 29 Hög
998062 - CD29 Låg
998070 - CD29 Normal
998088 - HemoTrol, låg
998096 - HemoTrol, hög
999904 - Extern kontroll,Seroni
999912 - Extern kontroll,Seroni
9999999 - Patientkontroll, LMC
bioimmun - Immunoassay, prog
Carlskem1 - Läkemedel
Carlskem2 - Läkemedel
Carlskem3 - Läkemedel
Carlskem4 - Läkemedel
ecatcont - Koagulation
ecatkoa - Koagulation
ecatmix - Koagulation
ecatscreen - Koagulation
eggaston - Blodgaser
Equaendo1 - Endokrinologi
Equaendo2 - Endokrinologi
equalalk - Alkoholer
equalcdt - CD-Transferin
equalcsv - Proteinanalyser i spi
Equaflores1 - Eltores
Equaflores2 - Eltores
Equaflores3 - Eltores
Equalhem4000 - Hematologi
equalhgc - Hemocue
equaloh - Iohexol
equalkem - Rutinkemi A
EqualkoaA - Koagulation
EqualkoaB - Koagulation
Equallip1 - Lipoprotein
Equallip2 - Lipoprotein
equalmia - Albumin i urin, låg nr
equalprot - P-protein
equalret - Retikulocyter
equalupr - U-Protein
equametho - MMA + Homocyste
INSTAKOA1 - Koagulation, spe
INSTAKOA2 - Koagulation, spe
Lqlakem1 - Läkemedel
Lqlakem2 - Läkemedel
N/A -
PPI - Mentorkontroll
UKDOWN1 - Downs screening
UKDOWN2 - Downs screening
UKDOWN3 - Downs screening
uklgGsub - Protein
Start
Lisa landstingets intra...
WSD - FlexLab Kemi -...
WSD - FlexLab Kemi -...
FlexLab/Kemi
Qm20111116.doc - M...
SV
09:17

B, MCHC

Mtd	Inst	CoID	Mean	CVtotal%	CVtreat%	CVerror%	%CV	n
M1	2454	PPI	336,3	2,658	2,086	1,804	2,325	7
M1	2455	PPI	335,1	3,126	0,7115	3,180	4,963	13
M1	3111	PPI	350,8	4,719	2,319	4,222	4,214	20
M1	3311	PPI	332,5	3,546	2,992	1,946	2,042	24




mentor plots
File Edit View
2002-01-01
2003-03-26
A
113,0 %
109,7 %
106,5 %
103,3 %
100,0 %
96,75 %
93,50 %
90,25 %
87,00 %
01-03 02-06 03-14 04-22 05-28 07-01 08-09 09-09 10-21 11-25
B
Adept conc.
Mentor concentration
C
Adept abs. dev.
Mentor concentration
D
Adept rel. dev. %
Mentor concentration
Conc
E[Conc]
CV WD
CV BD
Signed
Linear Regression y=1,0092x-0,7180

Full diagnostic method validation

	Participants			
	With disease	**Without disease**		
Positive test	True positives	False positives (type I error)	Total positive	**[PPV]**
Negative test	False negatives (type II error)	True negatives	Total negative	**[NPV]**
	Total with diease	Total without diease		
	[Sensitivity]	**[Specificity]**		

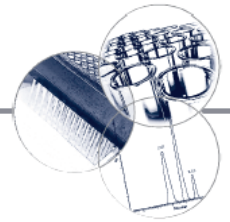
Table 1. Definition and calculation of parameters/concepts describing diagnostic properties of measurement methods.

Parameter/concept	Formula/explanation
Diagnostic sensitivity is the proportion of those with disease who have positive test results	$\text{Sensitivity} = \frac{\text{Number of true positives}}{\text{Total with disease}}$
Diagnostic specificity is the proportion of those without disease who have negative test results	$\text{Specificity} = \frac{\text{Number of true negatives}}{\text{Total without disease}}$
The positive likelihood ratio is the ratio of the true-positive to the false-positive rate	$\text{LR} + = \frac{\text{Sensitivity}}{1 - \text{Specificity}}$
The negative likelihood ratio is the ratio of the false-negative rate to the true-negative rate	$\text{LR} - = \frac{1 - \text{Sensitivity}}{\text{Specificity}}$
DOR combines the concepts of sensitivity, specificity and likelihood ratios into a single number, this is particularly useful for combining study results in systematic reviews	$\text{DOR} = \frac{\text{LR} +}{\text{LR} -}$
ROC curves	ROC curves show diagnostic properties of a measurement method used to classify persons with or without disease as the decision limit between health and disease is changed
PPV is the proportion of those with a positive test result who have the disease; takes into account the prevalence of disease in the target population	$\text{PPV} = \frac{\text{Number of true positives}}{\text{Total number of positives}}$
NPV is the proportion of those with negative test results who do not have the disease; takes into account the prevalence of disease in the target population	$\text{NPV} = \frac{\text{Number of true negatives}}{\text{Total number of negatives}}$

It should be noted that the prevalence of disease in the intended population is crucial for the predictive values, but not for the other parameters.
DOR: Diagnostic odds ratio; NPV: Negative predictive value; PPV: Positive predictive value; ROC: Receiver operating characteristic.

Review

For reprint orders, please contact reprints@future-science.com



Validation and verification of measurement methods in clinical chemistry

The present overview of validation and verification procedures in clinical chemistry focuses on the use of harmonized concepts and nomenclature, fitness-for-purpose evaluations and procedures for minimizing overall measurement and diagnostic uncertainty. The need for mutually accepted validation procedures in all fields of bioanalysis becomes obvious when they implement international accreditation and certification standards or their equivalents. The guide on bioanalytical method validation published by the US FDA in 2001 represents a sensible compromise between thoroughness and cost–effectiveness. Lacking comprehensive international agreements in the field, this document has also been successfully adapted in other fields of bioanalysis. European and international efforts aiming for consensus in the entire field of bioanalysis are currently being made. Manufacturers of highly automated *in vitro* diagnostic methods provide the majority of measurement methods used in unmodified in clinical chemistry. Validated by the manufacturers for their intended use and fitness-for-purpose, they need to be verified in the circumstances of the end-users. As yet, there is unfortunately no general agreement on the extent of the verification procedures needed.

Validation and verification of measurement methods are procedures that aim to establish realistic expectations with the analyst and confidence with the end-user that the methods are fit for their intended purposes. Different fields of bioanalysis have historically lacked a common theoretical and practical ground due not only to differences in the tasks at hand, but also to differences in terminology and in calibration, validation and quality control practices. Recent harmonization efforts in these areas [1,101,102] confirm that all fields of bioanalysis can share the same principles and nomenclature catering for extensive harmonization of guidelines, standards and practices.

In the early 1990s, the US FDA initiated and supported conferences and harmonization work on bioanalytical method validation [2,3] that, in 2001, resulted in the 'FDA Guidance for Industry – Bioanalytical Method Validation' guidelines [4,101]. They have been widely used, being suitable not only for the needs of the pharmaceutical industry but also for bioanalytical methods in general [4]. In fact, lacking similar international guidelines, this FDA document is widely used as standard reference for validation of bioanalytical measurement methods. European efforts in the field of validation (European Medicines Agency's Guidelines on Validation of Bioanalytical Methods) [5] are currently in progress.

The pharmaceutical industry has been and is still a driving force in the development of validation practices given the regulatory environment they have been subject to early on. Clinical laboratories are increasingly being accredited or certified according to ISO 17025, ISO 15189 or other similar quality systems. These laboratories are therefore in need of generally accepted and cost-effective protocols for validation. Theoretically, there are no limits to the extent of validation and verification procedures. However, in practice, there are time and economic constraints. It is therefore crucial that validation and verification efforts are optimized in order to maximize the value gained for the resources spent.

This brief overview of validation and verification methodologies in clinical chemistry attempts to adhere to the currently accepted guidelines in terminology and bioanalytical validation methodologies. The probable over-emphasis on certain aspects, for example, on verification procedures and fitness-for-purpose investigations, may be explained by the authors' background in laboratory medicine and basic research. The current already extensive and increasing use of commercially available measurement instruments and methods underscores the need for agreement on reasonable, but sufficient, methods for end-user verification of the manufacturer's performance claims.

Elvar Theodorsson

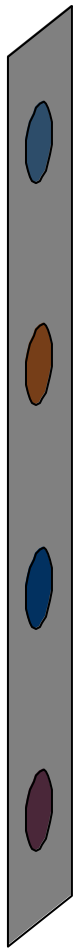
Clinical Chemistry, Department of Clinical & Experimental Medicine, Faculty of Health Sciences, Linköping University, County Council of Östergötland, Linköping, Sweden

Author for correspondence:
Tel.: +46 1328 6720
Fax: +46 1010 33240
E-mail: elvar.theodorsson@liu.se

FUTURE SCIENCE part of fsg

http://www.future-science.com/doi/pdf/10.4155/bio.11.311

What level on the measurement scale is used when quantifying precision?



A. Nominal

0%

B. Ordinal

0%

C. Interval

0%

D. Ratio

0%

Validation should be used instead of verification when…

A. Applying the method on a new measurement system

0%

B. When a new technician is engaged

0%

C. When the method is modified

0%

D. When the control material shows out of control condition

0%

Take a home message

Verification of precision and bias

- In Vitro Diagnostic (IVD) medical devices are in Europe regulated by a third EC Medical Device Directive, the IVD medical device Directive 98/79/EC which has been mandatory in Europa since December 2003
- Local verification practices have commonly been established over time and are frequently influenced by accreditation and certification authorities. Published practices for end-user verification officially endorsed both by the end-users and by the companies have appeared only recently including the EP15-A2 protocol from CLSI.
- Practical and pragmatic method using patient samples and common samples for internal quality control.
  - Bias is tested by comparison with a well-established methods using at least 20 patient samples
  - Variation within- and between series is measured using the normally used stable materials for internal quality control at least twice daily during two weeks