Validation in Clinical chemistry

Elvar Theodorsson

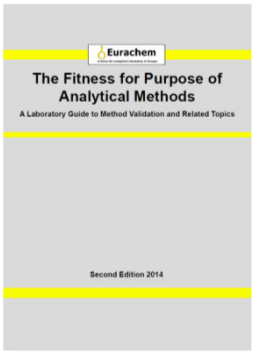
LINKÖPING UNIVERSITY

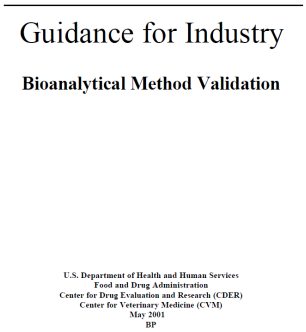
Overview

- Validation in Clinical chemistry
- Validation vs verification
- Single laboratory validation
- Full validation
- Full diagnostic validation
- Handling diagnostic uncertainties

LINKÖPING UNIVERSITY

Method validation in clinical chemistry follows the established standards and procedures accepted by all disciplines of chemical metrology.




Guidance for Industry

Bioanalytical Method Validation

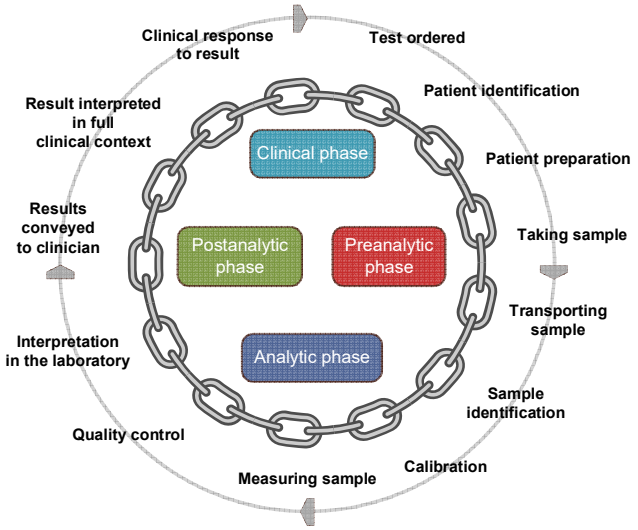
U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Veterinary Medicine (CVM)
May 2001
BP

LINKÖPING UNIVERSITY

Clinical chemistry is a high-volume, highly automated activity

LINKÖPING UNIVERSITY

The total testing chain in clinical chemistry



Biological variation 1956

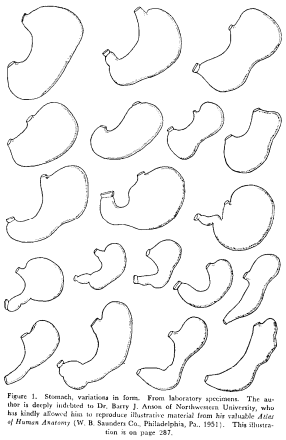


Figure 1. Stomach, variations in form. From laboratory specimens. The author is deeply indebted to Dr. Barry J. Anson of Northwestern University, who has kindly allowed him to reproduce illustrative material from his valuable *Atlas of Human Anatomy* (W. B. Saunders Co., Philadelphia, Pa., 1951). This illustration is on page 287.

ROGER J. WILLIAMS

Professor of Chemistry
Director, Biochemical Institute
University of Texas

Biochemical Individuality

THE BASIS FOR THE GENETOTROPHIC CONCEPT

New York · John Wiley & Sons, Inc.
London · Chapman & Hall Limited

Biological variation

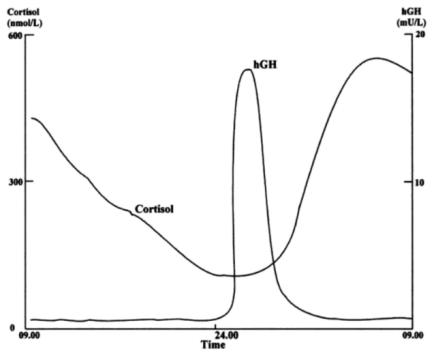


Figure 1.7 Concentrations of Serum Cortisol and Growth Hormone over a Typical Day with a Usual Sleep/Wake pattern

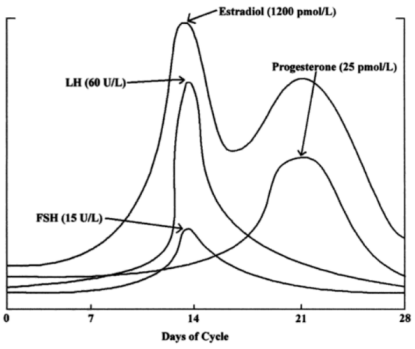


Figure 1.8 Concentrations of Four Hormones During a Classical 28-Day Menstrual Cycle

Biological variation

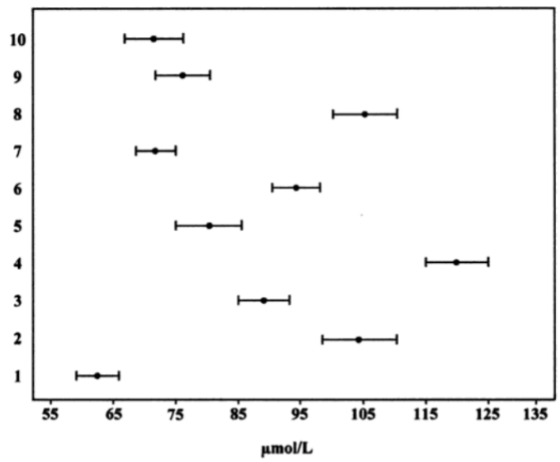


Figure 1.10 Mean Values and Absolute Ranges of Serum Creatinine in Four Samples Taken from Each of 10 Apparently Healthy Men.

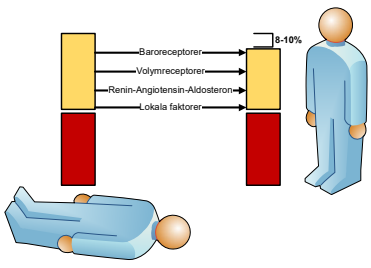
The age and sex matched reference interval for men aged 18–55 years is 64–120 µmol/L.

Biological variation

- The same in health and disease for the same individual

Sampling variation

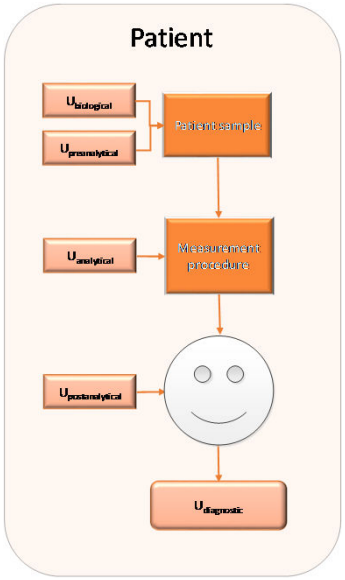
- Rest
- Venous stasis
- Mixing of the sample with stabilizers/anticoagulants
- Transport of the sample to the laboratory



Postanalytical variation

- Conveying of results to the user
- Interpretation of the results by the user

Diagnostic uncertainty consist of several uncertainties



Verification is a much more common activity in Clinical chemistry compared to 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"

Verification

- 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 since December 2003
- Verification practices have commonly been established over time and are frequently influenced by accreditation and certification authorities.

7.12.98 EN Official Journal of the European Communities L 331/1

I

(Acts whose publication is obligatory)

DIRECTIVE 98/79/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL
of 27 October 1998
on *in vitro* diagnostic medical devices

THE EUROPEAN PARLIAMENT AND THE COUNCIL OF THE EUROPEAN UNION,

trade, and whereas the need to establish harmonised rules has been confirmed by a comparative survey of national legislations carried out on behalf of the Commission;

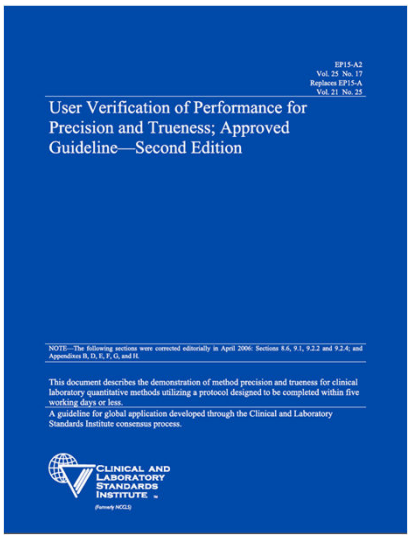
Verification

- 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.

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

Verification



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

LINKÖPING UNIVERSITY

Commutability

- "The equivalence of the mathematical relationships between the results of different measurement procedures for a reference material and for representative samples from healthy and diseased individual".
- This material characteristic is of special importance for measurement procedures that are optimized for measuring analytes directly in patient samples.
- The commutability of a reference material is measurement procedure specific and its assessment requires special experimental designs.

LINKÖPING UNIVERSITY

Types of validation in clinical chemistry

- **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 variations etc.
- **Full diagnostic method validation** is establishing the diagnostic properties of the method e.g. in health and disease

Basic requirements for method validation 1(2)

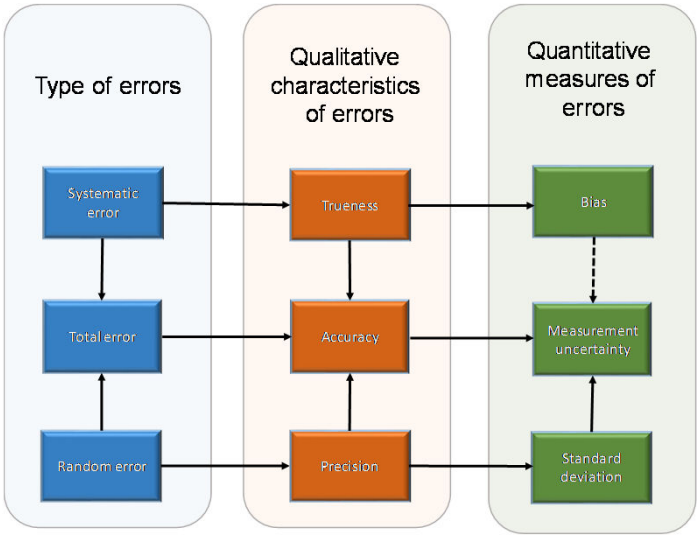
- The method should be fully developed and optimized
- A written standard operating procedure (SOP) for the method should be available
- The measurement instruments to be used should be regularly technically controlled and well maintained
- The persons performing the measurements should have sufficient training and experience for the task

Basic requirements for method validation 2(2)

- Appropriate calibrators should be available and a supply (for at least 1 year) of suitable stable materials (for at least 2 concentration levels) for internal quality control purposes
- The needs of the end user regarding fit-for-purpose of the method should be known

Fit for purpose = ”Analytical quality specifications”

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



Menditto A. M. Patriarca, M. Magnusson B (2007). "Understanding the meaning of accuracy, trueness and precision." Accred Qual Assur **12: 45-47**.

Precision

- 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
- Precision is measured as its opposite - imprecision

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 precision are in the order of 15

Reproducibility imprecision 1(4)

- 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

Reproducibility imprecision 2(4)

- 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

Reproducibility imprecision 3(4)

- 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

Reproducibility imprecision 4(4)

- Condition of measurement, out of a set of conditions that includes different locations, operators, measuring systems, and replicate measurements on the same or similar objects
- Validating measurement reproducibility is done as part of full method validation

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).

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

Estimating 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.

Estimating 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

Estimating 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”

Full full diagnostic 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]**		

LINKÖPING UNIVERSITY

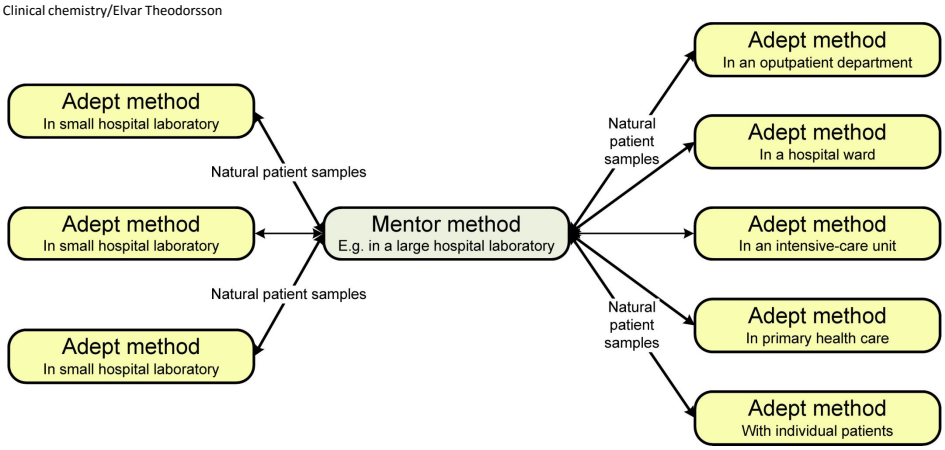
Table 1. Definition and calculation of parameters/concepts describing diagnostic properties of measurement methods.

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.

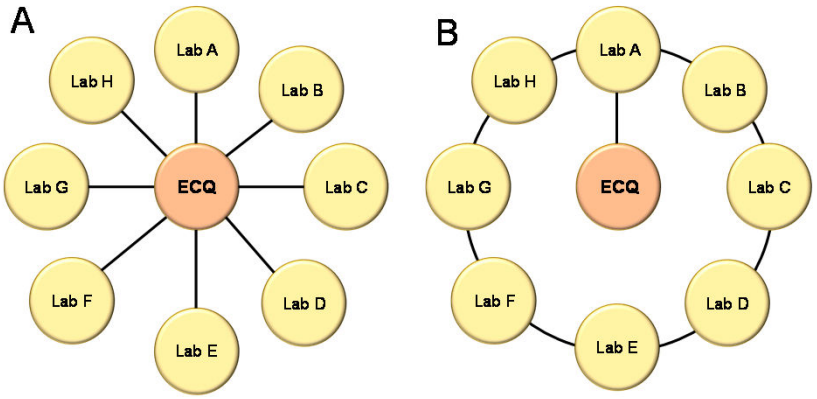
Full diagnostic method validation

LINKÖPING UNIVERSITY



LINKÖPINGS UNIVERSITET

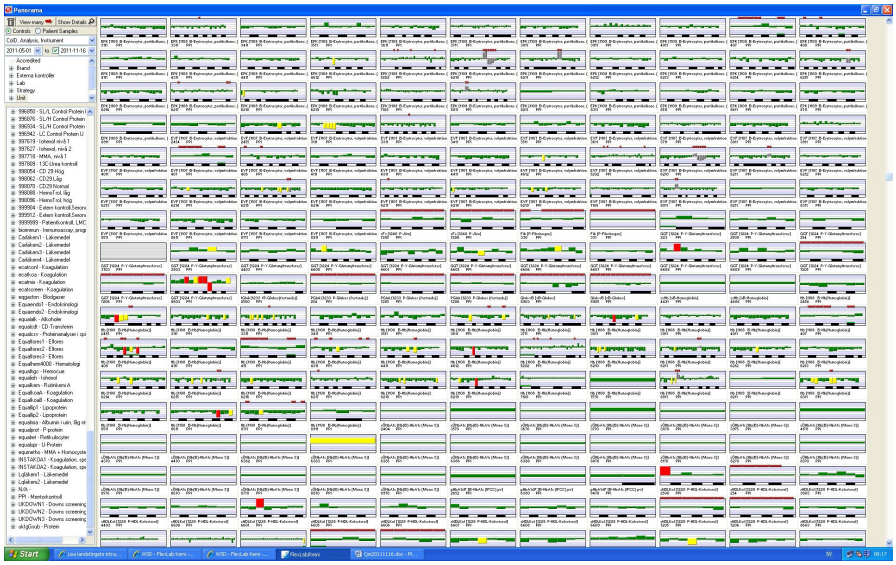
Two different principles for proficiency assessment

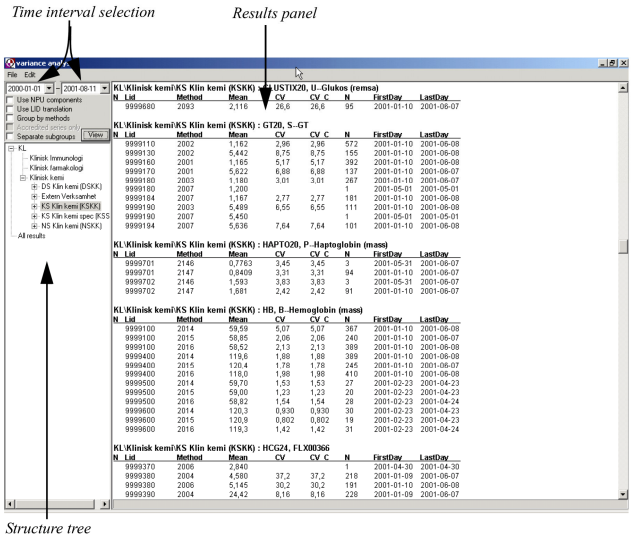


LINKÖPING UNIVERSITY

Norming results

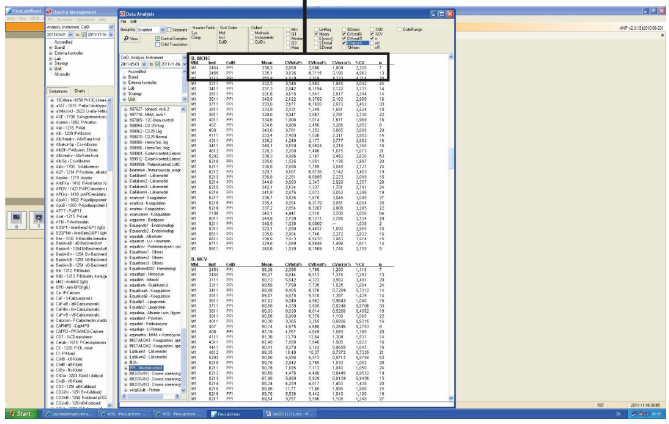
$$\text{Normed result} = \frac{\text{Adept - Mentor}}{\text{Mentor}} * 100$$



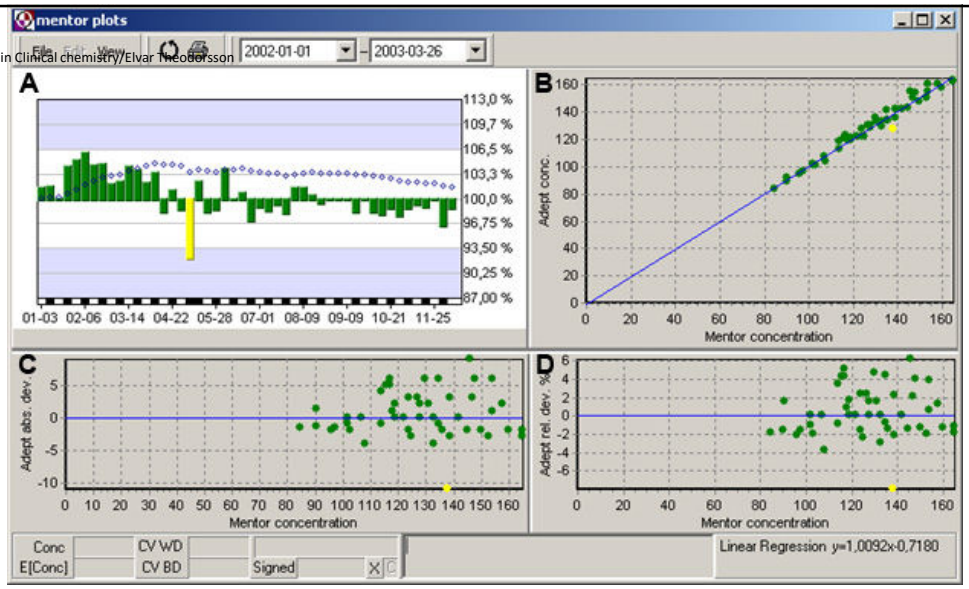


B, MCHC

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



Variance component analysis



LINKÖPINGS UNIVERSITET

Advantages of split samples for quality control

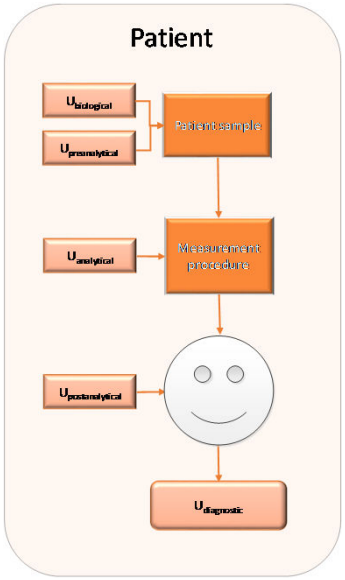
1. The the material has optimal matrix properties (is commutable)
2. The material is available without cost for all laboratories accepting routine patient samples
3. There is general agreement that all measurement systems and reagents should optimally result in identical results when analyzing the same patient samples
4. The methods are optimal for identifying the measurement system(s) in the organization that contribute the largest part of the overall measurement uncertainty due to bias. Split sample methods are laborious in the absence of effective computerized systems, but convenient when properly implemented

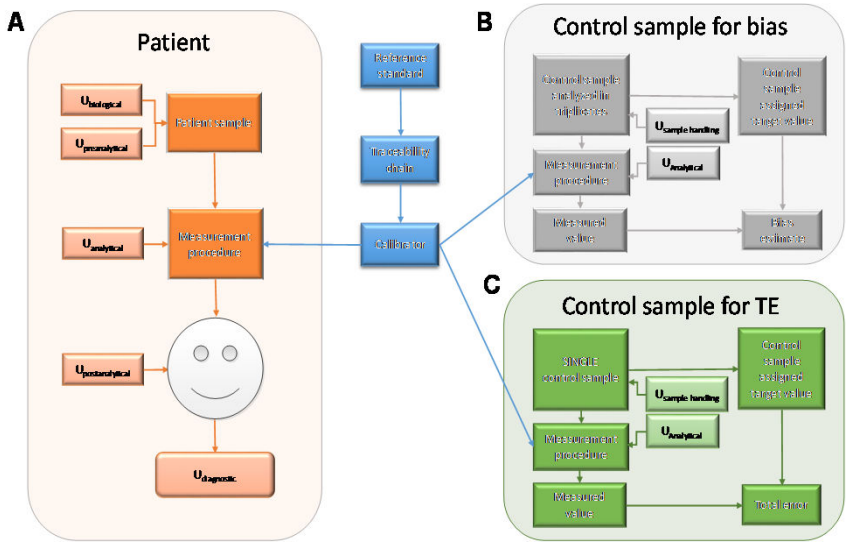
LINKÖPING UNIVERSITY

Analytical quality specifications and diagnostic uncertainty

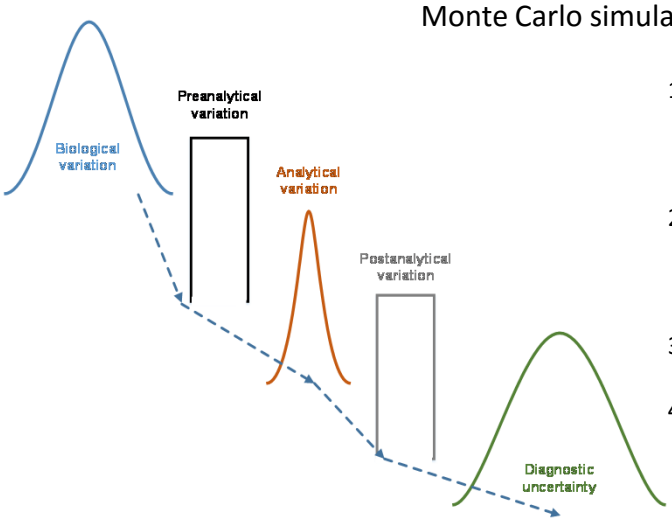
- Controversy in clinical chemistry regarding epistemology of measurement
  - The **realist** view, represented in clinical chemistry by classical **error** methods views measurements as the estimation of "mind-independent properties" of the measurand. Frequentist statistics.
  - **Model-based** methods are represented by measurement **uncertainty** methods which claim that other available information in addition to the measurement result itself should be provided as aid in the proper interpretation of the result. Bayesian statistics.

Diagnostic uncertainty consist of several uncertainties

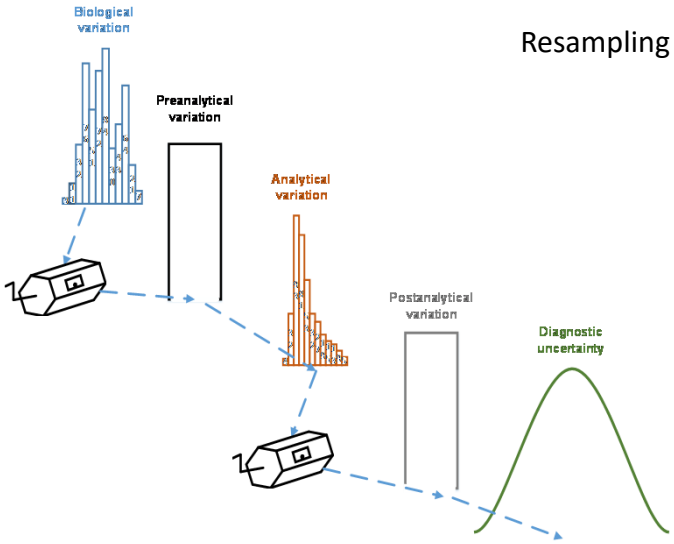




Monte Carlo simulation of **diagnostic uncertainty**



1. No mathematical function (output function) is needed to evaluate the diagnostic uncertainty
2. No assumptions about the input quantities is needed in addition to the assumption that they follow a Gaussian distribution
3. There is no need to calculate partial derivatives
4. It is unaffected by partial derivatives that vanish when estimating input quantities



Resampling estimation of **diagnostic uncertainty**

1. No mathematical function (output function) is needed to evaluate the diagnostic uncertainty
2. No assumptions about the input quantities is needed in addition to the assumption that they follow a Gaussian distribution
3. There is no need to calculate partial derivatives
4. It is unaffected by partial derivatives that vanish when estimating input quantities

LINKÖPINGS UNIVERSITET

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.

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 1038 6720
Fax: +46 1010 33340
E-mail: elvar.theodorsson@liu.se

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 [1,2,10,11] 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 [1,2] that, in 2001, resulted in the 'FDA Guidance for Industry – Bioanalytical Method Validation' guidelines [4,10]. They have been widely used, being suitable not only for the needs of the pharmaceutical industry but also for bioanalytical methods in general [5]. 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) [3] 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 overemphasis 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.

10.4155/BIO.11.311 © Elvar Theodorsson
Bioanalysis (2012) 4(3), 305–320
ISSN 1757-6180

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

LINKÖPINGS UNIVERSITET

Thank you

www.liu.se

LINKÖPING UNIVERSITY