

# Full method validation in clinical chemistry

Elvar Theodorsson<sup>1</sup> · Bertil Magnusson<sup>2</sup>

Received: 18 January 2017 / Accepted: 6 June 2017 / Published online: 29 June 2017  
© The Author(s) 2017. This article is an open access publication

**Abstract** Clinical chemistry is subject to the same principles and standards used in all branches of metrology in chemistry for validation of measurement methods. The use of measuring systems in clinical chemistry is, however, of exceptionally high volume, diverse and involves many laboratories and systems. Samples for measuring the same measurand from a certain patient are likely to encounter several measuring systems over time in the process of diagnosis and treatment of his/her diseases. Several challenges regarding method validation across several laboratories are therefore evident, but rarely addressed in current standards and accreditation practices. The purpose of this paper is to address some of these challenges, making a case that appropriate conventional method validation performed by the manufacturers fulfils only a part of the investigation needed to show that they are fit for purpose in different healthcare circumstances. Method validation across several laboratories using verified commercially available measuring systems can only be performed by the laboratories—users themselves in their own circumstances,

and need to be emphasised more by the laboratories themselves and accreditation authorities alike.

**Keywords** Validation · Bias · Verification · Commutability · Diagnostic uncertainty

## Introduction

The core of method validation in general, including that of “closed” measuring systems intended for healthcare, is the investigation of whether their properties are adequate for the intended use [1–5]. A single laboratory validation/verification is sufficient if the same measuring system is always used when analysing all samples from a population of patients. However, this is seldom the case in clinical chemistry. Patients are commonly diagnosed and their treatment combined with monitoring initiated at large University hospitals to be continued at a smaller hospital and one or two primary healthcare physicians (Fig. 1).

Even if the measuring systems, for example, for measuring the concentrations of glycated haemoglobin in whole blood from diabetics are validated and found fit for the intended use when investigating one or a handful of measuring systems in ideal situations under the control of manufacturers, they may not necessarily be fit for the intended use when the patient utilises the services of several laboratories using different measuring systems, in different real-life situations and even perhaps performs point-of-care measurements himself/herself. The manufacturers cannot be expected to shoulder the responsibility for their measuring systems in any constellation of laboratories and users. That responsibility rests with the users—the healthcare organisations.

---

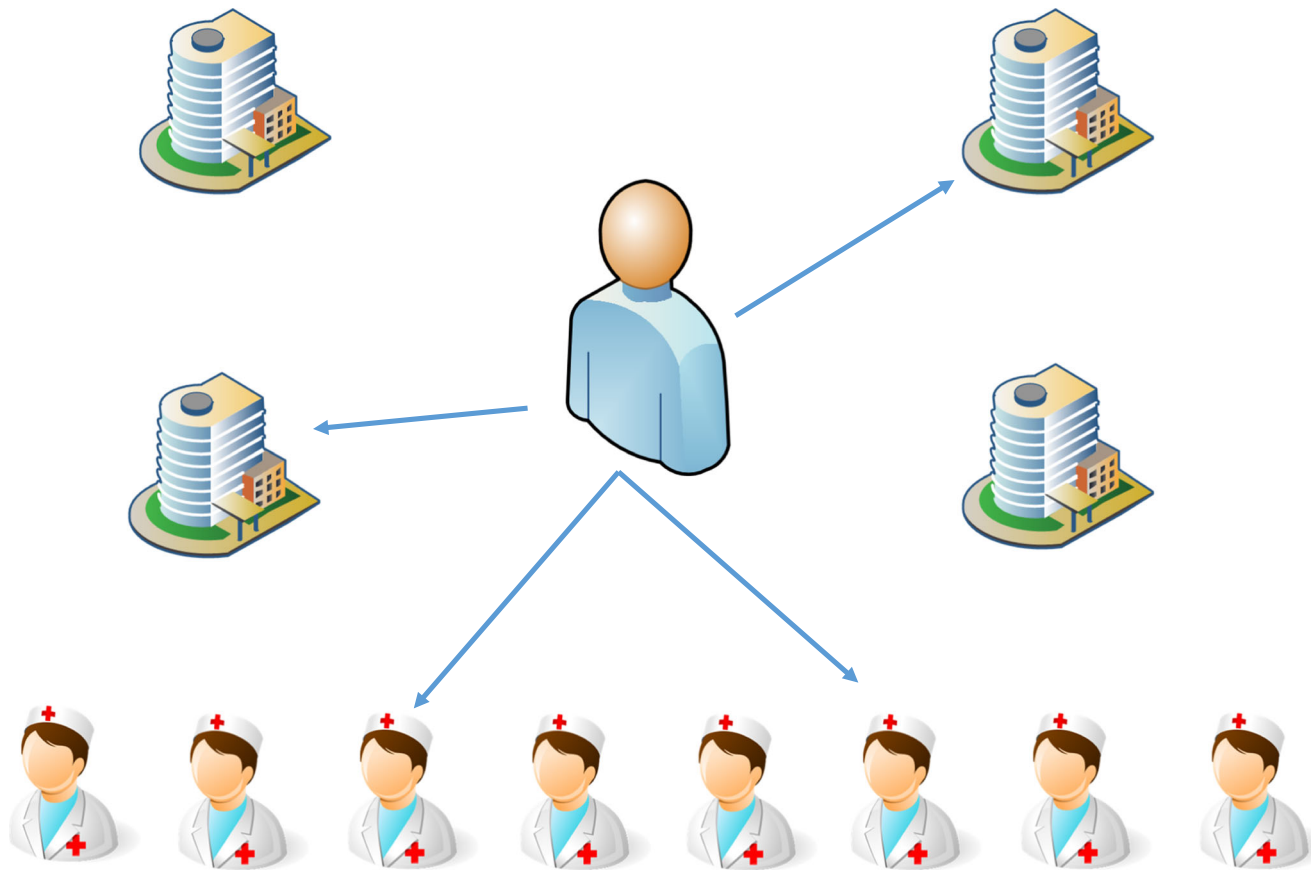
Presented at the Eurachem Workshop on Method Validation, May 2016, Gent, Belgium.

---

✉ Elvar Theodorsson  
elvar.theodorsson@liu.se  
Bertil Magnusson  
bertil.magnusson@sp.se

<sup>1</sup> Department of Clinical Chemistry and Department of Clinical and Experimental Medicine, Linköping University, Linköping, Sweden

<sup>2</sup> RISE, Research Institutes of Sweden, Borås, Sweden



**Fig. 1** Illustration of the common situation where a patient (*centre of illustration*) is being treated by two primary healthcare physicians (*bottom of illustration*) and by specialists at two different hospitals where both primary healthcare centres and the hospitals measure the blood concentration of, e.g. glycated haemoglobin by

different measuring systems. Each patient in the population may, furthermore, be cared for by different combinations of hospitals and primary healthcare physicians

This paper intends to provide a brief overview of validation practices in clinical chemistry and laboratory medicine and makes a case for extensions to these validation practices that should be and need to be performed by laboratory and other healthcare personnel during their use of the measuring systems (including pre- and postanalytical factors) in patient care. Practices in this vein have the potential to substantially contribute to minimising diagnostic uncertainty in the interest of the patients and healthcare providers alike.

### Causes of variation/uncertainty in clinical chemistry

Before discussing this topic, it is worthwhile recalling the following concepts:

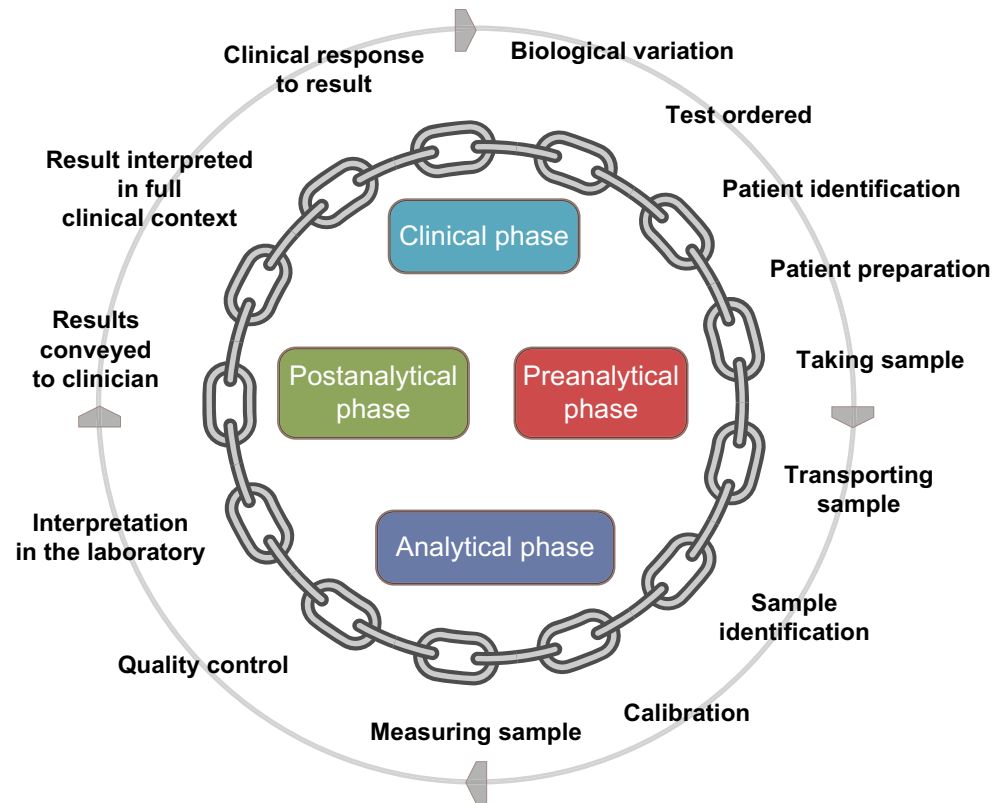
The *measurement procedure* is commonly called measurement method (as in the term method validation and in ISO/IEC 17025) or examination procedure (ISO 15189).

*Diagnostic uncertainty* is the uncertainty physicians and other healthcare personnel optimally need to count in when faced with challenges in diagnosis or when monitoring treatment effects. It is the combined uncertainty of all diagnostic measures taken, including anamnesis, physical examination, imaging and laboratory, and furthermore the uncertainty in the full diagnostic validation of the diagnostic measures, including diagnostic sensitivity, diagnostic specificity and diagnostic decision limits.

*Analytical uncertainty* is the combined uncertainty for a certain measurement result of a certain measurand for all measuring systems in a conglomerate of laboratories catering for a population of patients.

*The total testing chain* in clinical chemistry involves several possible sources of uncertainty from the clinical decision to order a test through the biological variation inherent in all mammals, the preanalytical, analytical and postanalytical phases to the use of the test results days, weeks and months on end for monitoring the effects of treatment (Figs. 2, 3).

**Fig. 2** Sources of uncertainty in the total testing chain in clinical chemistry

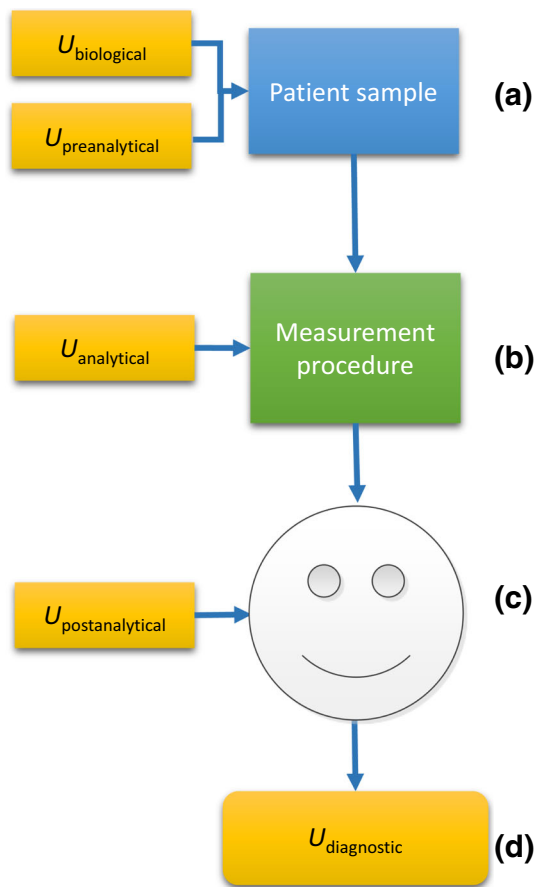


The *clinical phase* involves the knowledge and skills of the healthcare personnel in the use of biomarkers for diagnosing and monitoring treatment effects, including the understanding, e.g. of the effects of biological variation, drugs, interferences on the results. The preanalytical phase involves preparing the patient for sampling, e.g. making sure that samples to be compared are taken in a standardised manner. Biological variation is sometimes included in the preanalytical phase. The analytical phase including the uncertainty in this phase (analytical uncertainty) involves all measuring systems and laboratories that a patient potentially encounters. The postanalytical phase deals with the interpretation of the measurement results in the context of the patient(s). Successful handling of the postanalytical phase is highly dependent on the knowledge and skills of the laboratory- and other healthcare personnel. The clinical phase involves understanding of the pathophysiology of diseases and the strengths and weaknesses of individual biomarkers in diagnosis and in monitoring of treatment effects. Healthcare personnel acquires knowledge in this area during their basic training, but recurrent opportunities for continuous educational activities which include aspects of laboratory medicine are needed to optimise the clinical phase in the total testing chain. Engagement of laboratory personnel is crucial to make this happen in any healthcare organisation.

Understanding of the uncertainty caused by *biological variation* [6–10] (which is frequently in the order of twice the measurement uncertainty of a single measuring system) and its influence on diagnosis and monitoring is crucial. Biological variation is a homeostatic biological mechanism whereby the body keeps the concentration of the measurand varying around an individual set point which commonly differs amongst individuals. Knowledge of biological variation and skills in handling this uncertainty component must be an integral part of medical decision-making since biological variation cannot be regulated neither in humans nor in other living organisms.

*Preanalytical variation* is the variation caused by differences in patient preparation, in the techniques and equipment used taking the sample and when transporting the sample to the laboratory. For example, the effect of gravitation on body fluids and molecules dissolved in them decreases the concentration of cells and large molecules by 8 % to 10 % about 30 min after a patient changes body position from vertical (standing up) to horizontal (laying down).

Besides staff at the medical wards, also laboratory personnel are responsible for assessing preanalytical issues such as haemolysis in the sample and errors in sample transport. It is crucial to register and regularly monitor such events for possible lack of conformance using computerised systems in



**Fig. 3** Components of diagnostic uncertainty when using chemical measurements in diagnostic medicine. Diagnostic uncertainty (D) is the combination of all the other uncertainty components (including A-C)

order to monitor their incidence and prevalence preferably as internationally agreed quality indicators [11], aiming to reduce preanalytical errors as much as possible. The Working Group on Laboratory Errors and Patient Safety of the International Federation of Clinical Chemistry and Laboratory Medicine has agreed on such quality indicators [12–15] which include misidentification errors, transcription errors, incorrect sample type, incorrect fill level, transportation and storage problems, contamination, haemolysed and clotted samples, data transcription errors and inappropriate turnaround times. Most importantly this register is crucial in deciding where and when to efficiently use the resources of the laboratory organisation for self-improvement and as an aid to their clinical colleagues in improving their knowledge and skills in preanalytics by educational activities, preferably delivered in person to individuals and groups. Since the influence of both biological and preanalytical variation on the patient’s diagnosis is highly dependent on the knowledge and skills of all involved in the clinic and in the laboratory alike [11], these factors should be included in the evaluation of the overall uncertainty estimates.

The *analytical phase* is usually conceived as fully in the hands of commercial producers of measuring systems and reagents, even though individual laboratories are crucial in monitoring the entire conglomerate of measuring systems.

The postanalytical uncertainty is caused by suboptimal technical facilities or routines in conveying the results to the healthcare personnel and/or lack of knowledge and skills in interpreting the results by the laboratory personnel and end-users [12, 16–18].

### Standardisation and harmonisation in clinical chemistry

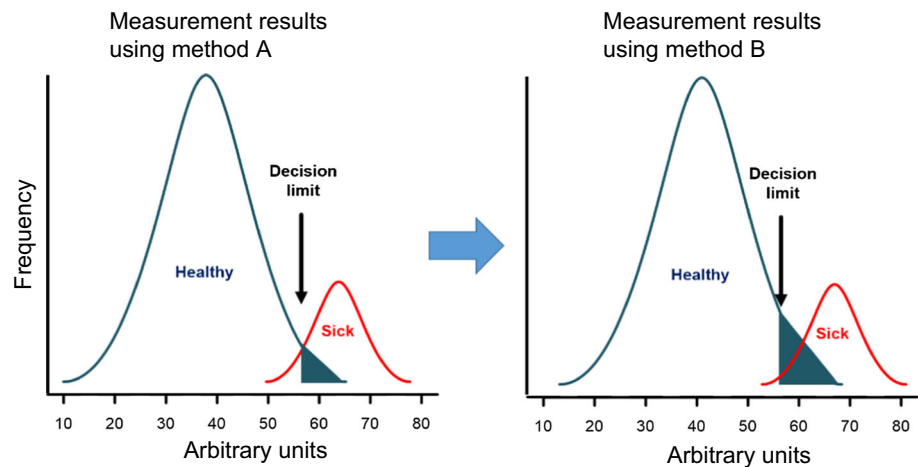
If measurements systems give different (biased) results for the same patient sample, it risks confusion amongst patients and their doctors. Furthermore, monitoring and treatment practices risk being implemented erroneously due to the bias, since clinical practice guidelines [19–21] that inform about proper actions for diagnosis and treatment are optimally based on unbiased test results (Fig. 4).

Absence of bias can only be assumed in very rare cases. In many cases, guidelines are based on measurement results obtained with a single, non-standardised device. Even worse, for guidelines based on studies performed in the past it is often not known in what manner the measurement scale used in the study relates to measurement scales, calibrators and selectivity of current devices. This can be a problem even if the same measurement principle and method is used, due to uncontrolled method drift. It is also common that “old” cut-off points are used for measurement results obtained with “new” methods. Therefore, the uncertainty of reference intervals and clinical decision limits is essential when counting in the postanalytical uncertainty.

A general comment concerns the definition of standardisation. In the field of clinical chemistry, some authors have developed the tendency to use definitions for standardisation and harmonisation that deviate from those generally used in measurement science or metrology. In fact, standardisation is defined in ISO/IEC Guide 2:2004 (Standardisation and Related Activities—General Vocabulary) as “activity of establishing, with regard to actual or potential problems, provisions for common and repeated use, aimed at the achievement of the optimum degree of order in a given context”. Standardisation can be achieved in different ways, for example, by developing standards with consensus scales (e.g. the SI units or International Units of WHO standards).

Clinical practice guidelines [19–21] that inform about proper actions for diagnosis and treatment are based on unbiased test results. *Standardisation* aims at achieving equivalent results by applying calibrators traceable to SI

**Fig. 4** A bias of +5 arbitrary units in this case means that an increased number of healthy persons are falsely diagnosed as sick as shown by the increase in the *dark triangular area* in the figure to the right compared to the figure to the left



and the use of reference measurement procedures [22–26]. Standardisation is accomplished when equivalent results are obtained by different clinical laboratory tests conducted by different laboratories using valid traceability chains established between the measurement results and a stable endpoint, be it the SI, the value of internationally agreed reference material (RM) or a value obtained with a reference method.

Standardisation is not possible when internationally agreed RM, and corresponding reference measurement procedures are not available. *Harmonisation* is then the second best and in fact the only option. It aims at achieving equivalent results amongst different measurement procedures commonly using fresh patient samples [27–31]. Unfortunately, less than 10 % of measurands (60 of more than 600) in a typical university hospital laboratory of clinical chemistry and laboratory medicine are as yet traceable to SI.

Standardised and harmonised clinical laboratory test results [24–26] improve the quality of healthcare by ensuring reliable screening, diagnosis and supporting appropriate treatments. They also reduce the risk of diagnostic and treatment errors that may be caused by unnecessary variation in test results. They lower healthcare cost by avoiding false-positive or false-negative results from non-standardised/harmonised tests. Such results risk unnecessary follow-up diagnostic procedures and treatments.

Standardisation is the method of choice for obtaining equivalence of measurement results. It has the unique advantage that when measurement results provided by reference methods or values assigned to RM are traceable to the SI units, this allows maintenance of proper calibration over time and across locations. Standardisation has proven particularly successful for well-defined measurands existing in only a single molecular form (e.g. small molecules like creatinine and cholesterol) in clinical samples.

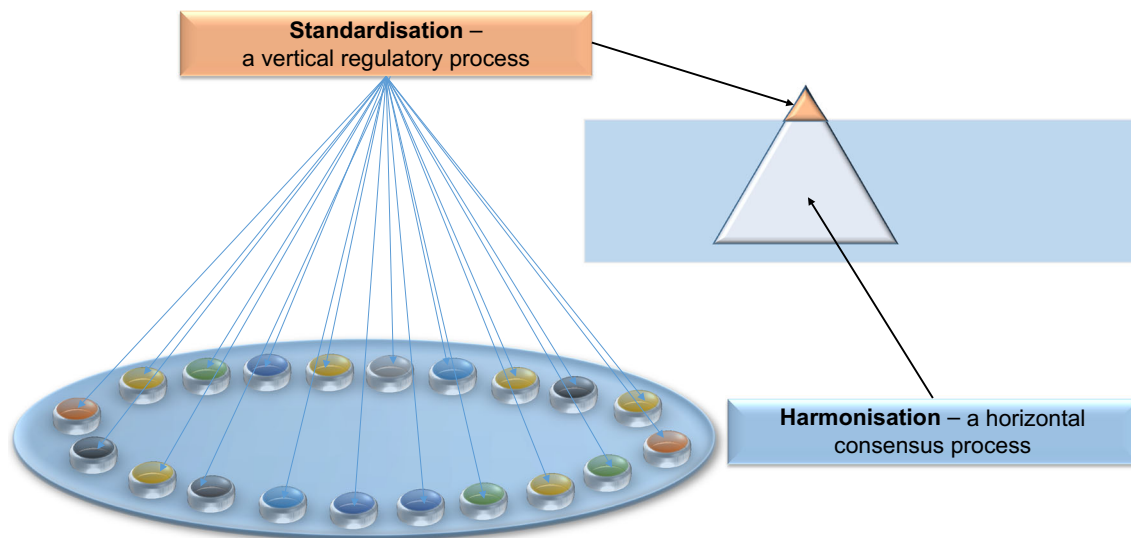
Harmonised methods work through consensus and are valid during a particular period in time. They do not share the ability of standardised methods to maintain trueness over extended periods of time. Harmonisation is usually based on the use of natural patient samples for comparing methods [28]. The advantage of harmonisation is that it is able to address the tests that as yet cannot be standardised (Fig. 5).

Complex large-molecular measurands that exist in several molecular forms (e.g. lutropin, follitropin, human chorionic gonadotropin) are difficult to standardise. Consensus is required on the unique definition of the measurands based on solid research findings and understanding of the clinically and metrologically relevant molecular forms that are needed both in RM and the patient samples. We are currently only in the very beginning of a long process of accomplishing this for all relevant measurands in laboratory medicine.

The use of a single central laboratory has been the rule when establishing laboratory result-based clinical guidelines [28]. Knowledge of their performance in the complex uncertainties conglomerates of laboratories using different measuring systems is in its infancy.

## Method validation in clinical chemistry

*Single laboratory method validation* is appropriate when a method is used for a specific purpose in one laboratory. *Full method validation* in a conglomerate of laboratories includes, in addition to the procedures of single laboratory validation, a study of the fitness for the intended use of measuring systems in a number of locations, several operators, etc. including a study of the performance characteristics of the measuring systems over extended periods of time including the effects of lot-to-lot variations.



**Fig. 5** Standardisation using traceable and internationally agreed RM and appropriate reference measurement procedures is optimal. Unfortunately, only about 10 % of measurands in laboratory medicine today are traceable to SI (illustrated by the tip of the iceberg analogy

*Full diagnostic method validation* is an investigation of the diagnostic properties of the method (diagnostic sensitivity, diagnostic specificity and diagnostic decision limits, etc.) and the added value the method brings to the clinical diagnosis and monitoring of treatment effects. It is used for establishing the diagnostic properties of the method in health and disease [32–35], a major undertaking demanding that the diagnosis in question is independently established by methods other than the one being tested.

*Diagnostic validation* investigates to what extent a conglomerate of measuring systems that samples from a patient are likely to encounter can reproduce the conditions that existed during the original full diagnostic method validation.

The conglomerate of laboratories should minimise the analytical uncertainty since results can be produced and reported by any laboratory within the conglomerate. The contribution of pre- and postanalytical uncertainty also needs to be minimised by systematic monitoring of errors and other sources of uncertainty and collaboration with the clinically active personnel. The analytical uncertainty is preferably estimated by stabilised samples for internal quality control for measuring precision and using commutable samples, e.g. using split-sample techniques for estimating bias as described below.

### Precision

Precision is the quantitative expression of random error usually by the coefficients of variation monitored under specific conditions. *Repeatability conditions* exist when the same examination procedure, same operators, same

on the right). The consensus process of harmonisation using natural patient samples can, however, always be used

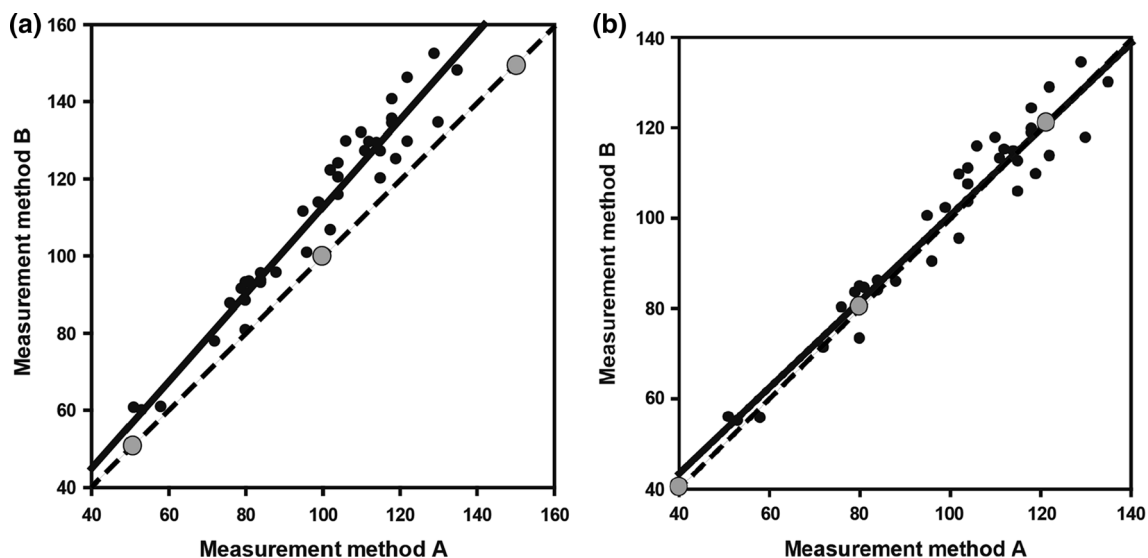
measuring system, same operating conditions and same location are used for replicate measurements on the same or similar objects over a short period of time, usually less than a working day of 8 h. *Reproducibility conditions* includes the same or different measurement procedure, different location, and replicate measurements on the same or similar objects over an extended period, but may include other conditions involving changes. *Intermediate precision* includes conditions in between the extremes of repeatability and reproducibility. It is usually estimated by daily examinations over extended periods of time for at least 1 year. All sources of variation included in intermediate precision including, e.g. lot number changes are included in appropriate number of occurrences. The intermediate precision can refer to one measuring system or to all measuring systems in the conglomerate of laboratories.

### Bias

Bias is an estimate of a systematic measurement error. The qualitative concept trueness—in this case lack of trueness—is quantitatively expressed as bias. It is optimally estimated using commutable certified RM or by comparing the average concentration measured in a natural patient sample with the method to be tested with the average concentration measured in the same sample using a reference method.

### Commutability

Commutability is a property of a material/sample demonstrated by “the closeness of agreement between the relation



**Fig. 6** **a** Lack of commutability of a RM (grey dots and broken line) compared to natural patient samples (black dots and black solid line). Commutability in clinical chemistry describes a RM ability to react in the same way as patient specimens in laboratory measurements. **b** A

commutable RM (grey dots and broken line) overlaps with natural patient samples (black dots and solid line)

among the measurement results for a stated quantity in this material, obtained according to two given measurement procedures, and the relation obtained among the measurement results for other specified materials” [1] (Fig. 6). Commutability is thus “the equivalence of the mathematical relationship between the results of different measurement procedures for a RM and for representative samples from healthy and diseased individuals” [36]. Natural patient samples are by definition commutable.

When a traceability chain is established, it is crucial to include commutable materials in the procedures for determining the concentrations in secondary RM, working calibrators and product calibrators (Fig. 7) in order that the results ultimately measured in the patient samples are comparable. Omission or disregard of this fundamental necessity contributes to the bias frequently found between measuring systems and methods from different manufacturers even for traceable measurement methods.

If a RM is not commutable, the results from routine methods cannot be properly compared with the assigned value of the RM when determining a possible bias [37, 38]. Observed bias may in this case be either due to the non-commutability of the RM or due to the differing specificities of the methods. Non-commutable RM used in validation results in wrong estimation of bias [38, 39].

### Proficiency testing

In proficiency testing, individual laboratory results are compared with a consensus value or assigned value. Since the stabilised control materials—that may or may not be

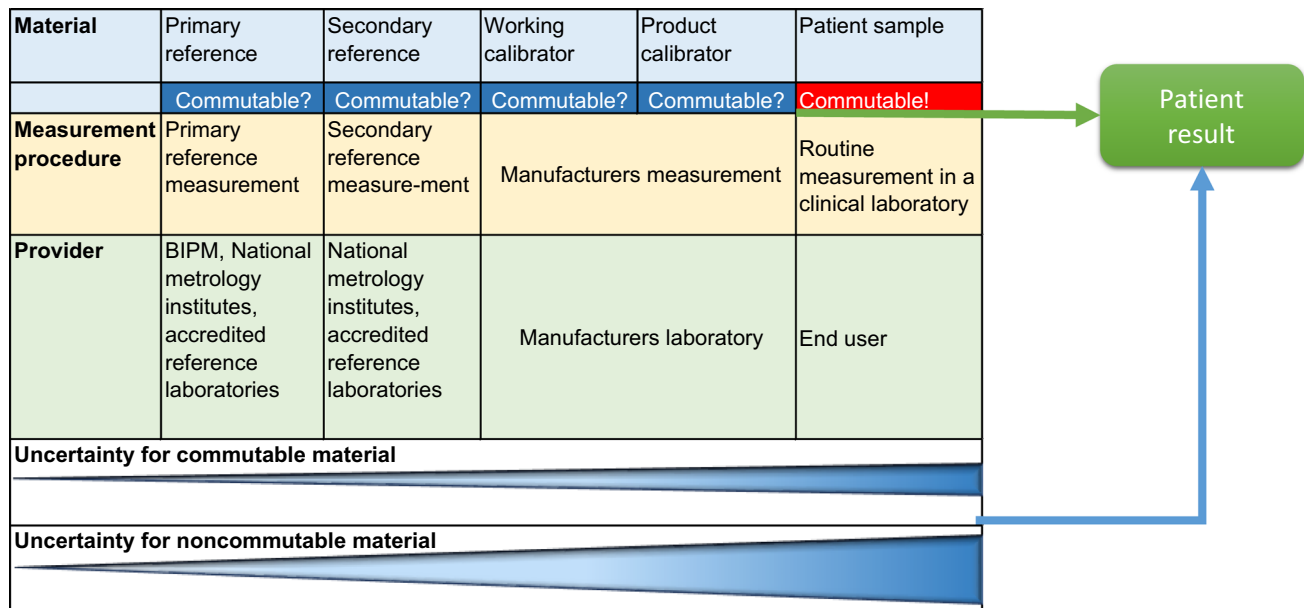
commutable—are commonly used, the averages of participants’ results grouped by measuring system or method commonly differ. Therefore, participants’ performances are commonly evaluated against an assigned value, which in clinical chemistry is most often determined as the participants’ consensus value. This bias information is, however, valuable for monitoring the performance of individual measuring systems and methods. Furthermore, accreditation and certification organisations keep data from proficiency testing in high regard and find them essential for obtaining and maintaining accreditation and certification.

Participating in a proficiency testing programme applying singleton measurements of the samples will provide a check on the estimated uncertainty (the combination of precision and bias) instead of trueness. Optimal estimation of trueness requires replicate measurements and calculation of the average and the difference (bias) between the average and the assigned value.

Some organisations/companies running proficiency testing schemes occasionally use fresh patient samples in their surveys. This practice substantially decreased the bias between different measuring systems and methods because the manufacturers commonly use natural patient samples which are commutable in their efforts to establish and maintain traceability to certified RM and reference methods.

### Split samples for estimation of bias within a conglomerate of laboratories

Running a proficiency testing scheme requires sophisticated logistics and computerisation outside the scope of



**Fig. 7** Traceability chain of RM involves reference measurement procedures and measurement procedures of lower metrological order including routine measurement procedures. If non-commutable RM is used for calibration in one or more of the measurement steps

conglomerates of laboratories. However, the laboratory conglomerate always maintains logistics for sending patient samples between the laboratories, e.g. from a small laboratory analysing a limited number of measurands to a larger laboratory analysing a comprehensive selection of measurands. Let's imagine using this already established and well-maintained logistic function for estimating bias. In this case, a laboratory (adept) sends a patient sample that it has already analysed to a central laboratory (mentor) which measures the sample using its normal automation and measuring systems and methods. However, in this case the sample result is not reported to healthcare as a patient result but as a result for internal use in the laboratory conglomerate for estimation of the bias between the methods used by the mentor and adept laboratories.

Such a split-sample mentor-adept scheme does evidently not establish or maintain traceability of the measuring systems and methods in the conglomerate of laboratories. However, it provides valuable information about the calibration and other technical parameters of the different measuring systems that influence the trueness and thereby the uncertainty when measuring patient samples that are analysed at different locations/laboratories with the laboratory conglomerate. This bias information is then most commonly used to identify measuring systems that need recalibration, maintenance or full blown overhaul rather than for secondary adjustment of the calibration functions to reduce bias.

performed, there is a risk of bias and increased uncertainty in the traceability chain as shown at the *bottom* of the figure

The advantages of natural patient samples are: (1) the material is commutable and has similar matrix properties, (2) they are available without cost for all laboratories accepting routine patient samples, (3) there is a general agreement that theoretically all measuring systems and reagents should result in identical results when analysing the same patient samples. This is not always the case.

### Fitness for purpose/fitness for intended use evaluation

Fitness for purpose is “the property of data produced by a measurement process that enables a user of the data to make technically correct decisions for a stated purpose” [40]. When defining the concept Thompson and Ramsey [40] referred to Tonks study from 1963 [41] that the allowable limits of error for a measurand should be one quarter of the reference interval and expressed as percentage of the mean of the reference interval. Thereby, the concept of “fitness for purpose” was from the outset coupled to the concept of “analytical quality specifications”/ “analytical performance specifications” widely used in clinical chemistry [42–50].

In decision theory, fitness for purpose is “the property of a result when it provides the maximum utility” [5]. Decisions on fitness for purpose may therefore be based on informed professional judgement and an agreement between the laboratory and the users of the laboratory [5].



Estimating fitness for purpose has also been defined as reaching externally stated requirements of “target measurement uncertainty” [51] or “property of a result of a measurement when the uncertainty provides minimal total average costs” [5]. Such fitness for purpose evaluations may, for example, be performed in proficiency testing schemes, e.g. using *z*-scores.

Whereas evaluation of fitness for purpose/fitness for intended use has been narrowed to reaching an agreed “target measurement uncertainty” in some parts of the sciences of metrology [51] including VIM 2.34, it has maintained its original “maximum utility” [5] scope in clinical chemistry and is known as analytical quality or analytical performance specifications [48]. Fitness for purpose remains the property of results produced by measuring systems that enables a user of the data to make clinically correct decisions for a stated purpose.

### Performance specifications—target measurement uncertainty

The Stockholm Conference held in 1999 on “Strategies to set global analytical quality specifications in laboratory medicine” advocated the following hierarchical structure for performance 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 using (a) data based on components of biological variation, or (b) analysis of clinicians’ opinions; (3) published professional recommendations from (a) national and international expert bodies, or (b) expert local groups or individuals; (4) performance goals set by (a) regulatory bodies, or (b) organisers of external quality assessment (EQA) schemes; and (5) goals based on the current state of the art as (a) demonstrated by data from EQA or proficiency testing scheme, or (b) found in current publications on methodology [52].

The conference “Defining analytical performance specifications” the 1st Strategic Conference of the European Federation of Clinical Chemistry and Laboratory Medicine in Milan 2014 maintained and simplified the criteria in an attempt to improve its applications for various stakeholders [48]. *Model 1*. Based on the effect of analytical performance on clinical outcomes (1) Direct outcome studies—investigating the impact of analytical performance of the test on clinical outcomes; (2) Indirect outcome studies—investigating the impact of analytical performance of the test on clinical classifications or decisions and thereby on the probability of patient outcomes, e.g. by simulation or decision analysis. *Model 2*. Based on components of biological variation of the measurand. *Model 3*. Based on state of the art.

Optimal clinical/patient outcomes remain the “reasons for being” in clinical chemistry and should, whenever proper data are available, remain at the top of the list of performance specifications for laboratories; however, tempting it may seem to regress to purely technical/metrological specifications including “target measurement uncertainty” and state of the art determined, e.g. by performance in proficiency testing schemes.

Optimal performance specifications should evidently cover the entire total testing process (Figs. 2, 3) including the pre- and postanalytical phases [11, 13, 53, 54]. Since clinical decision limits are based on studies where all phases of the total testing process have been involved, they are usually counted in when model 1 (see above) is used. A primary task of laboratories and conglomerates of laboratories is to establish and maintain systems to minimise pre-and postanalytical errors and to monitor their occurrences. If and when pre- and postanalytical errors can be expressed as uncertainty components, they should evidently be included in performance specifications in the same manner as measurement uncertainties [48].

### The European in vitro diagnostics IVD directive

In vitro diagnostic (IVD) medical devices are in Europe regulated by the IVD Directive 98/79/EC [55] which has been mandatory since December 2003.

ISO 17511:2013 (In vitro diagnostic medical devices—Measurement of quantities in biological samples—Metrological traceability of values assigned to calibrators and control materials) [56] is the standard showing how to achieve traceability in accordance with EU legislation. The fact that it is a harmonised standard means that it is recognised at EU level as describing how the legislation (IVD directive) should be implemented.

ISO 17511 [56] describes several different possible traceability chains, which can all be used to achieve standardisation (albeit it only within a particular measurement system in the last case):

- Cases with primary reference measurement procedure and primary calibrator(s) giving metrological traceability to SI.
- Cases with international conventional reference measurement procedure (which is not primary) and international conventional calibrator(s) without metrological traceability to SI.
- Cases with international conventional reference measurement procedure (which is not primary) but no international conventional calibrator and without metrological traceability to SI.

- Cases with international conventional calibrator (which is not primary) but no international conventional reference measurement procedure and without metrological traceability to SI.
- Cases with manufacturer's selected measurement procedure but neither international conventional reference measurement procedure nor international conventional calibrator and without metrological traceability to SI.

### Validation versus verification

The IVD directive [55] states “The traceability of values assigned to calibrators and/or control materials must be assured through available reference measurement procedures and/or available RM of a higher order”. (98/79/EC, Annex1 (A) (3) 2nd paragraph). “Higher order” is not defined in the directive and neither was implementing legislation beyond assigning responsibility for assuring traceability to national notified bodies. Furthermore, harmonisation for the methods that are not traceable is not either mentioned in the directive.

One of the crucial advantages of the IVD directive is that it emphasises standardisation/traceability of measurement methods and puts the responsibility for validation on the shoulders of the manufacturers. The responsibility of the users/laboratories then becomes to verify the measurement methods—to investigate to what extent the performance data obtained by manufacturers during method validation can be reproduced in the environments of the end-users.

Verification practices have commonly been established over time and are naturally influenced by accreditation and certification authorities. The EP15-A2 protocol from CLSI [57] is commonly used for this purpose and uses stabilised control material with assigned concentrations or certified RM. Another pragmatic method involving commutable materials is to measure a range of concentrations in at least 20 natural patient samples both by the established method and by the new method to estimate bias and to measure at least two concentrations of stabilised control materials at least twice daily for at least 10 days to estimate repeatability and intermediate reproducibility.

### Limitation of the IVD directive and current verification practices

The IVD directive [55] has done Clinical chemistry in Europe service in emphasising traceability and clarifying the responsibilities of metrology institutes and manufacturers of measuring systems. However, the IVD directive

risks complacency amongst the users of the measuring systems and methods since it puts the overwhelming responsibility for the overall quality of measurements in clinical chemistry on the shoulders of the manufacturers of measuring systems. Furthermore, it only demands the verification of each measuring system independently, and not as a part of a conglomerate of measuring systems all potentially reporting to the same client.

The manufacturers of measuring systems are usually in no position to do full method validations (as defined earlier in this paper) and are therefore unable to supply the end-users with information about bias and reproducibility precision to be expected and possibly verified in typical conglomerates of laboratories for a certain population. The users of measuring systems in conglomerates of laboratories in clinical chemistry therefore need to look for analytical performance specifications/goals [46, 48–50, 58, 59] appropriate for the patient population their laboratories serve preferably in close collaboration with their clinical colleagues. The priorities within the conglomerate of laboratories should then be to fulfil these analytical performance goals not only in the analytical phase of the total testing process, but also in the pre- and postanalytical phases. Using commutable control materials including split natural patient samples will serve well in this effort. The main purpose of bias control within a conglomerate of laboratories using commutable materials is to identify measuring systems in need of technical overhaul and primary calibration. Secondary adjustment of calibrations [60] is rarely required when calibrations are properly performed and the measuring systems are in optimal technical condition.

### Conclusions

Samples for measuring the same measurand from a certain patient are likely to encounter several measuring systems over time in the process of diagnosis and treatment of his/her diseases. The conglomerate of laboratories serving a population of patients will serve the interest of their patients even better if they minimise even further the part of diagnostic uncertainty caused by analytical uncertainty and improve the traceability/harmonisation of the measuring systems. A full method validation is a study of fitness for purpose including all the measuring systems in a number of laboratories. Clinical decision limits and clinical guidelines will thereby be appropriately used.

**Acknowledgements** The authors acknowledge with gratitude the substantial contribution of the reviewers and editor to improving the quality of this manuscript.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

- JCGM (2012) International vocabulary of metrology—Basic and general concepts and associated terms (VIM 3). Bureau International des Poids et Mesures. [http://www.bipm.org/utis/common/documents/jcgm/JCGM\\_200\\_2012.pdf](http://www.bipm.org/utis/common/documents/jcgm/JCGM_200_2012.pdf). Accessed 1 Feb 2017
- Fearn T, Fisher SA, Thompson M, Ellison SLR (2002) A decision theory approach to fitness for purpose in analytical measurement. *Analyst* 127(6):818–824. doi:10.1039/B111465d
- Thompson M, Fearn T (1996) What exactly is fitness for purpose in analytical measurement? *Analyst* 121:275–278
- Magnusson B, Örnemark U (2014) Eurachem guide: the fitness for purpose of analytical methods—a laboratory guide to method validation and related topics, 2nd edn. Eurachem. [www.eurachem.org](http://www.eurachem.org)
- Thompson M, Ellison SLR (2006) Fitness for purpose—the integrating theme of the revised Harmonised protocol for proficiency testing in analytical chemistry laboratories. *Accred Qual Assur* 11(8–9):373–378. doi:10.1007/s00769-006-0137-5
- Fraser CG, Cummings ST, Wilkinson SP, Neville RG, Knox JD, Ho O, MacWalter RS (1989) Biological variability of 26 clinical chemistry analytes in elderly people. *Clin Chem* 35(5):783–786
- Simundic AM, Bartlett WA, Fraser CG (2015) Biological variation: a still evolving facet of laboratory medicine. *Ann Clin Biochem* 52(Pt 2):189–190. doi:10.1177/0004563214567478
- Fraser CG, Hyltoft Peterson P (1993) Desirable standards for laboratory tests if they are to fulfill medical needs. *Clin Chem* 39:1447–1455
- Ricos C, Alvarez V, Perich C, Fernandez-Calle P, Minchinela J, Cava F, Biosca C, Boned B, Domenech M, Garcia-Lario JV, Simon M, Fernandez PF, Diaz-Garzon J, Gonzalez-Lao E (2015) Rationale for using data on biological variation. *Clin Chem Lab Med CCLM FESCC* 53(6):863–870. doi:10.1515/cclm-2014-1142
- Ricós C, Alvarez V, Cava F, García-Lario JV, Hernández A, Jiménez CV, Minchinela J, Perich C, Simón M (1999) Current databases on biological variation: pros, cons and progress. *Scand J Clin Lab Invest* 59(7):491–500
- Plebani M, Sciacovelli L, Aita A, Pelloso M, Chiozza ML (2015) Performance criteria and quality indicators for the pre-analytical phase. *Clin Chem Lab Med CCLM FESCC* 53(6):943–948. doi:10.1515/cclm-2014-1124
- Sciacovelli L, Plebani M, Garcia del Oino Castro I, Lippi G, Sumarac Z, Furtado Veira K, West JB, Ivanov A (2016) Laboratory Errors and Patient Safety (WG-LEPS). International federation of clinical chemistry and laboratory medicine (IFCC). [http://217.148.121.44/MqiWeb/resources/doc/Quality\\_Indicators\\_Key\\_Processes.pdf](http://217.148.121.44/MqiWeb/resources/doc/Quality_Indicators_Key_Processes.pdf). Accessed 16 Sept 2016
- Plebani M, Sciacovelli L, Aita A, Padoan A, Chiozza ML (2014) Quality indicators to detect pre-analytical errors in laboratory testing. *Clin Chim Acta* 432:44–48. doi:10.1016/j.cca.2013.07.033
- Plebani M, Sciacovelli L, Marinova M, Marcuccitti J, Chiozza ML (2013) Quality indicators in laboratory medicine: a fundamental tool for quality and patient safety. *Clin Biochem* 46(13–14):1170–1174. doi:10.1016/j.clinbiochem.2012.11.028
- Plebani M, Chiozza ML, Sciacovelli L (2013) Towards harmonization of quality indicators in laboratory medicine. *Clin Chem Lab Med CCLM FESCC* 51(1):187–195. doi:10.1515/cclm-2012-0582
- Skeie S, Perich C, Ricos C, Araczkowski A, Horvath AR, Oosterhuis WP, Bubner T, Nordin G, Delport R, Thue G, Sandberg S (2005) Postanalytical external quality assessment of blood glucose and hemoglobin A1c: an international survey. *Clin Chem* 51(7):1145–1153. doi:10.1373/clinchem.2005.048488
- Kristoffersen AH, Thue G, Sandberg S (2006) Postanalytical external quality assessment of warfarin monitoring in primary healthcare. *Clin Chem* 52(10):1871–1878. doi:10.1373/clinchem.2006.071027
- Favaloro EJ, Lippi G, Adcock DM (2008) Preanalytical and postanalytical variables: the leading causes of diagnostic error in hemostasis? *Semin Thromb Hemost* 34(7):612–634. doi:10.1055/s-0028-1104540
- Wils J, Fonfrede M, Augereau C, Watine J (2014) Further comments on “Critical review of laboratory investigations in clinical practice guidelines: proposals for the description of investigation”. *Clin Chem Lab Med CCLM FESCC* 52(8):e155–e157. doi:10.1515/cclm-2013-1045
- Aakre KM, Langlois MR, Watine J, Barth JH, Baum H, Collinson P, Laitinen P, Oosterhuis WP (2013) Critical review of laboratory investigations in clinical practice guidelines: proposals for the description of investigation. *Clin Chem Lab Med CCLM FESCC* 51(6):1217–1226. doi:10.1515/cclm-2012-0574
- Trenti T, Schunemann HJ, Plebani M (2016) Developing GRADE outcome-based recommendations about diagnostic tests: a key role in laboratory medicine policies. *Clin Chem Lab Med CCLM FESCC* 54(4):535–543. doi:10.1515/cclm-2015-0867
- Armbruster D (2013) Accuracy controls: assessing trueness (bias). *Clin Lab Med* 33(1):125–137. doi:10.1016/j.cll.2012.10.002
- Bais R, Armbruster D, Jansen RT, Klee G, Panteghini M, Passarelli J, Sikaris KA, Results IWGoAeFT (2013) Defining acceptable limits for the metrological traceability of specific measurands. *Clin Chem Lab Med CCLM FESCC* 51(5):973–979. doi:10.1515/cclm-2013-0122
- Armbruster DA (2009) Measurement traceability and US IVD manufacturers: the impact of metrology. *Accred Qual Assur* 14(7):393–398. doi:10.1007/s00769-009-0535-6
- Armbruster DA (2013) Implementation of traceability: is the IVD industry’s approach really fulfilling obligations? In: 7th CIRME international scientific meeting: metrological traceability & assay standardization, May 24th, 2013, Stresa
- Armbruster D, Miller RR (2007) The Joint Committee for Traceability in Laboratory Medicine (JCTLM): a global approach to promote the standardisation of clinical laboratory test results. *Clin Biochem Rev* 28(3):105–113
- Miller WG, Tate JR, Barth JH, Jones GR (2014) Harmonization: the sample, the measurement, and the report. *Ann Lab Med* 34(3):187–197. doi:10.3343/alm.2014.34.3.187
- Miller WG, Eckfeldt JH, Passarelli J, Rosner W, Young IS (2014) Harmonization of test results: what are the challenges; how can we make it better? *Clin Chem* 60(7):923–927. doi:10.1373/clinchem.2012.201186
- Miller WG, Myers GL (2013) Commutability still matters. *Clin Chem* 59(9):1291–1293. doi:10.1373/clinchem.2013.208785
- Gantzer ML, Miller WG (2012) Harmonisation of measurement procedures: how do we get it done? *Clin Biochem Rev* 33(3):95–100
- Miller WG, Myers GL, Lou Gantzer M, Kahn SE, Schonbrunner ER, Thienpont LM, Bunk DM, Christenson RH, Eckfeldt JH, Lo

- SF, Nubling CM, Sturgeon CM (2011) Roadmap for harmonization of clinical laboratory measurement procedures. *Clin Chem* 57(8):1108–1117. doi:[10.1373/clinchem.2011.164012](https://doi.org/10.1373/clinchem.2011.164012)
32. Bossuyt PM, Reitsma JB, Linnet K, Moons KG (2012) Beyond diagnostic accuracy: the clinical utility of diagnostic tests. *Clin Chem* 58(12):1636–1643. doi:[10.1373/clinchem.2012.182576](https://doi.org/10.1373/clinchem.2012.182576)
33. Bossuyt PM, Cohen JF, Gatsonis CA, Korevaar DA, Group S (2016) STARD 2015: updated reporting guidelines for all diagnostic accuracy studies. *Ann Transl Med* 4(4):85. doi:[10.3978/j.issn.2305-5839.2016.02.06](https://doi.org/10.3978/j.issn.2305-5839.2016.02.06)
34. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig L, Lijmer JG, Moher D, Rennie D, de Vet HC, Kressel HY, Rifai N, Golub RM, Altman DG, Hooft L, Korevaar DA, Cohen JF, Group S (2015) STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *BMJ* 351:h5527. doi:[10.1136/bmj.h5527](https://doi.org/10.1136/bmj.h5527)
35. Moons KG, de Groot JA, Linnet K, Reitsma JB, Bossuyt PM (2012) Quantifying the added value of a diagnostic test or marker. *Clin Chem* 58(10):1408–1417. doi:[10.1373/clinchem.2012.182550](https://doi.org/10.1373/clinchem.2012.182550)
36. (ISO) IofS (2003) In vitro diagnostic medical devices—measurement of quantities in biological samples—metrological traceability of values assigned to calibrators and control materials. International Organization for Standardization (ISO), Geneva, Switzerland
37. Miller WG, Myers GL, Rej R (2006) Why commutability matters. *Clin Chem* 52(4):553–554. doi:[10.1373/clinchem.2005.063511](https://doi.org/10.1373/clinchem.2005.063511)
38. Vesper HW, Miller WG, Myers GL (2007) Reference materials and commutability. *Clin Biochem Rev* 28(4):139–147
39. Franzini C, Ceriotti F (1998) Impact of reference materials on accuracy in clinical chemistry. *Clin Biochem* 31(6):449–457
40. Thompson M, Ramsey MH (1995) Quality concepts and practices applied to sampling—an exploratory-study. *Analyst* 120(2):261–270. doi:[10.1039/an9952000261](https://doi.org/10.1039/an9952000261)
41. Tonks DB (1963) A study of the accuracy and precision of clinical chemistry determinations in 170 Canadian laboratories. *Clin Chem* 9:217–233
42. Laessig RH (1990) Medical need for quality specifications within laboratory medicine. *Ups J Med Sci* 95(3):233–244
43. Fraser CG (2015) The 1999 Stockholm consensus conference on quality specifications in laboratory medicine. *Clin Chem Lab Med* 53(6):837–840. doi:[10.1515/cclm-2014-0914](https://doi.org/10.1515/cclm-2014-0914)
44. Dybkaer R (1993) Medical need for quality specifications in clinical laboratories. Truth, accuracy, error and uncertainty. *Ups J Med Sci* 98(3):215–220
45. Fraser CG (1990) Quality specifications in laboratory medicine. *Ups J Med Sci* 95(3):229–232
46. Ceriotti F, Fernandez-Calle P, Klee GG, Nordin G, Sandberg S, Streichert T, Vives-Corrons JL, Panteghini M, Task E, Finish Group on Allocation of laboratory tests to different models for performances (2016) Criteria for assigning laboratory measurands to models for analytical performance specifications defined in the 1st EFLM strategic conference. *Clin Chem Lab Med*. doi:[10.1515/cclm-2016-0091](https://doi.org/10.1515/cclm-2016-0091)
47. Horvath AR, Bossuyt PM, Sandberg S, John AS, Monaghan PJ, Verhagen-Kamerbeek WD, Lennartz L, Cobbaert CM, Ebert C, Lord SJ, Test Evaluation Working Group of the European Federation of Clinical C, Laboratory M (2015) Setting analytical performance specifications based on outcome studies—is it possible? *Clin Chem Lab Med* 53(6):841–848. doi:[10.1515/cclm-2015-0214](https://doi.org/10.1515/cclm-2015-0214)
48. Sandberg S, Fraser CG, Horvath AR, Jansen R, Jones G, Oosterhuis W, Petersen PH, Schimmel H, Sikaris K, Panteghini M (2015) Defining analytical performance specifications: consensus statement from the 1st strategic conference of the European federation of clinical chemistry and laboratory medicine. *Clin Chem Lab Med CCLM FESCC* 53(6):833–835. doi:[10.1515/cclm-2015-0067](https://doi.org/10.1515/cclm-2015-0067)
49. Oosterhuis WP, Sandberg S (2015) Proposal for the modification of the conventional model for establishing performance specifications. *Clin Chem Lab Med CCLM FESCC* 53(6):925–937. doi:[10.1515/cclm-2014-1146](https://doi.org/10.1515/cclm-2014-1146)
50. Thue G, Sandberg S (2015) Analytical performance specifications based on how clinicians use laboratory tests. Experiences from a post-analytical external quality assessment programme. *Clin Chem Lab Med CCLM FESCC* 53(6):857–862. doi:[10.1515/cclm-2014-1280](https://doi.org/10.1515/cclm-2014-1280)
51. De Bievre P (2007) Fitness for purpose is different from a performance specification. *Accred Qual Assur* 12(10):501. doi:[10.1007/s00769-007-0312-3](https://doi.org/10.1007/s00769-007-0312-3)
52. Kallner A, McQueen M, Heuck C (1999) The Stockholm consensus conference on quality specifications in laboratory medicine, 25–26 April 1999. *Scand J Clin Lab Invest* 59(7):475–476
53. Plebani M, Sciacovelli L, Aita A, Chiozza ML (2014) Harmonization of pre-analytical quality indicators. *Biochem Med (Zagreb)* 24(1):105–113. doi:[10.11613/BM.2014.012](https://doi.org/10.11613/BM.2014.012)
54. Plebani M, Astion ML, Barth JH, Chen W, de Oliveira Galoro CA, Escuer MI, Ivanov A, Miller WG, Petinos P, Sciacovelli L, Shcolnik W, Simundic AM, Sumarac Z (2014) Harmonization of quality indicators in laboratory medicine. A preliminary consensus. *Clin Chem Lab Med CCLM FESCC* 52(7):951–958. doi:[10.1515/cclm-2014-0142](https://doi.org/10.1515/cclm-2014-0142)
55. EU (1998) Directive 98/79/EC of the European Parliament and of the Council of 27 October 1998 on in vitro diagnostic medical devices. Eur-Lex, <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31998L0079:EN:NOT>
56. ISO (2003) 17511:2003 In vitro diagnostic medical devices—measurement of quantities in biological samples—metrological traceability of values assigned to calibrators and control materials
57. CLSI (2006) User verification of performance for precision and trueness; approved guideline EP15-A2. Clinical and Laboratory Standards Institute
58. Stepman HC, Stockl D, Twomey PJ, Thienpont LM (2013) A fresh look at analytical performance specifications from biological variation. *Clin Chim Acta* 421:191–192. doi:[10.1016/j.cca.2013.03.018](https://doi.org/10.1016/j.cca.2013.03.018)
59. Panteghini M, Sandberg S (2015) Defining analytical performance specifications 15 years after the Stockholm conference. *Clin Chem Lab Med CCLM FESCC* 53(6):829–832. doi:[10.1515/cclm-2015-0303](https://doi.org/10.1515/cclm-2015-0303)
60. Theodorsson E, Magnusson B, Leito I (2014) Bias in clinical chemistry. *Bioanalysis* 6(21):2855–2875. doi:[10.4155/bio.14.249](https://doi.org/10.4155/bio.14.249)