EFLM Paper

Mauro Panteghini*, Ferruccio Ceriotti, Graham Jones, Wytze Oosterhuis, Mario Plebani and Sverre Sandberg, on behalf of the Task Force on Performance Specifications in Laboratory Medicine of the European Federation of Clinical Chemistry and Laboratory Medicine (EFLM)

Strategies to define performance specifications in laboratory medicine: 3 years on from the Milan Strategic Conference

https://doi.org/10.1515/cclm-2017-0772
Received August 29, 2017; accepted September 5, 2017; previously published online October 14, 2017

Abstract: Measurements in clinical laboratories produce results needed in the diagnosis and monitoring of patients. These results are always characterized by some uncertainty. What quality is needed and what measurement errors can be tolerated without jeopardizing patient safety should therefore be defined and specified for each analyte having clinical use. When these specifications are defined, the total examination process will be "fit for purpose" and the laboratory professionals should then set up rules to control the measuring systems to ensure they perform within specifications. The laboratory community has used different models to set performance specifications (PS). Recently, it was felt that there was a need to revisit different models and, at the same time, to emphasize the presuppositions for using the different models. Therefore, in 2014 the European Federation of Clinical Chemistry and Laboratory Medicine (EFLM) organized a Strategic Conference in Milan. It was felt that there was a need for more detailed discussions on, for instance, PS for EQAS, which measurands should use which models to set PS and how to set PS for the extra-analytical phases. There was also a need to critically evaluate the quality of data on biological variation studies and further discussing the use of the total error (TE) concept. Consequently, EFLM established five Task Finish Groups (TFGs) to address each of these topics. The TFGs are finishing their activity on 2017 and the content of this paper includes deliverables from these groups.

Keywords: biological variation; outcome; performance specifications.

***Corresponding author: Mauro Panteghini,** Research Centre for Metrological Traceability in Laboratory Medicine (CIRME), University of Milan, Via G.B. Grassi 74, 20157 Milan, Italy, Phone: +39 02 3904 2806, Fax: +39 02 50319835, E-mail: mauro.panteghini@unimi.it
Ferruccio Ceriotti: Central Laboratory, Fondazione IRCCS Ca' Granda, Ospedale Maggiore Policlinico, Milan, Italy
Graham Jones: SydPath, St Vincent's Hospital, Sydney, NSW, Australia
Wytze Oosterhuis: Department of Clinical Chemistry and Haematology, Zuyderland Medical Center, Heerlen, The Netherlands
Mario Plebani: Department of Laboratory Medicine, University-Hospital of Padova, Padova, Italy
Sverre Sandberg: Norwegian Quality Improvement of Laboratory Examinations (Noklus), Haraldsplass Deaconess Hospital, Bergen, Norway; Department of Global Public Health and Primary Health Care, University of Bergen, Bergen, Norway; and Norwegian Porphyria Centre, Laboratory of Clinical Biochemistry, Haukeland University Hospital, Bergen, Norway

Introduction

Performance specifications (PS) in laboratory medicine should ideally identify criteria that specify (in numerical terms) the quality required for laboratory test information that can satisfy clinical needs for improving patients' outcomes. Measurement uncertainty should in principle be within limits based on medical relevance making the results clinically acceptable and reliable for clinical decision-making and patient management [1]. If PS are not objectively defined and fulfilled, there is a risk of letting the variation in laboratory result overwhelm the clinical information supplied, even causing negative effects on patients' outcomes [2, 3]. What degree of quality is needed to guarantee patient safety should therefore be precisely defined and specified for each analyte.

The debate focusing on how these limits should be defined, in order to answer the essential question about what amount of medical harm due to measurement error is acceptable to let go undetected, started many years ago [4]. In 1999, this was addressed in the organization of the

IFCC-International Union of Pure and Applied Chemistry (IUPAC) conference in Stockholm, where a hierarchy of sources for deriving the PS of laboratory measurements was established for the first time [5]. However, in a review of progress made in the 10 years following the conference, two of the organizers recognized challenges in the practical implementation of concepts and the need to reappraise the topic by simplifying the approaches and encompassing further aspects, such as extra-analytical phases [6]. In consequence, the Strategic Conference "Defining Analytical Performance Goals – 15 years after the Stockholm Conference" was organized in Milan on November 2014 by the European Federation for Clinical Chemistry and Laboratory Medicine (EFLM) to investigate to what extent the advocated hierarchy was still valid or if it has to be changed or expanded [7].

The Milan Strategic Conference

Although many of the ideas established in 1999 were supported, in the Milan conference the models for PS were based on three completely different theoretical frameworks (Table 1) [8]. Accordingly, the recommended approaches for defining PS should preferentially be based on the effect of measurement performance on clinical outcome or on the biological variation of the measurand. If PS based on these models could not be made, state-of-the-art could be used. Importantly, it should be noted that these three models use different principles and do not necessarily constitute a hierarchy. An important innovative aspect of the new consensus is the recognition that some models are better suited for certain measurands than for others; the attention is therefore primarily directed towards the measurand and its biological and clinical characteristics [9]. Another novel aspect is the focus on the quality of the information. Independent of the model, high quality studies or data must be available. For instance, for many decades biological variability has been promoted as a source of analytical PS. This can be a good and usable model for many measurands, but it must be underlined that studies and papers with information about biological variation have to be critically appraised before they can be used to set PS [10–12]. The possibility to elaborate specifications at different levels of quality (i.e. minimum, desirable and optimum) to move, in case, from desirable to minimum PS is also an option (Figure 1). This will also make it possible for *in vitro* diagnostics (IVD) manufacturers to work for improving the quality of assay performance. In general, proposed PS should always be accompanied by a statement of the rationale, the source and the quality of the evidence behind the recommendation, together with the suggested application [8].

Table 1: Models to be used for defining performance specifications according to the 2014 Milan Strategic Conference.

Model 1: Based on the effect of test performance on clinical outcomes
- a. Based on direct outcome studies – investigating the impact of performance of the test on clinical outcomes;
- b. Based on indirect outcome studies – investigating the impact of 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 of the measurement, defined as the highest level of performance technically achievable

Adapted from Ref. [8].

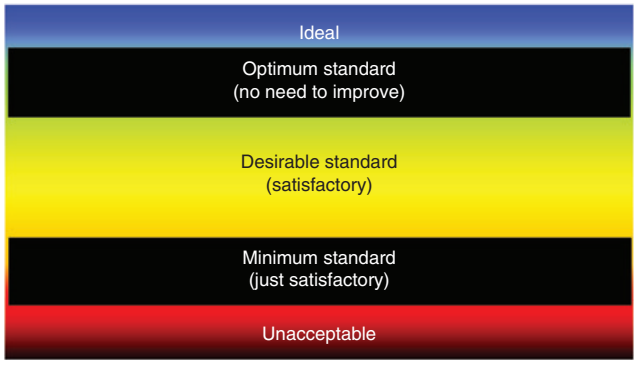

Figure 1: Grading different quality levels of performance specifications.

Another feature of the Milan conference is the concept that the application of the analytical PS can be modulated depending on its use. During the conference, Braga et al. [13] described for the first time the rationale behind the definition of recommended limits for combined measurement uncertainty across the entire metrological traceability chain. In comparison with the traditional approach, they turned the problem upside down, focusing first on the PS for combined uncertainty associated with patient results and recommending that the higher order references should display uncertainty at most equal to 1/3 of these PS. Similarly, criteria for IVD manufacturers that can be achieved for their calibrators were defined in order to leave enough uncertainty budget for the individual laboratories to produce clinically acceptable results on clinical samples [14]. Individual laboratories who provide patient results have of course a central role in correctly defining PS to be applied in daily practice. Recent examples in the

literature show that the current quality of laboratory performance for many common analytes is probably not as satisfactory as expected when PS are selected according to the Milan models [15–17].

The Strategic Conference heritage

EFLM initiated Task and Finish Groups

The main outcome of the Strategic Conference was the creation of an EFLM Task Force (TF) on Performance Specifications in Laboratory Medicine (TF-PS). Under the TF-PS, five Task and Finish Groups (TFG) were established dealing with the main topics of the conference [7]. According to the EFLM policy, TFG are expected to complete their deliverables within 2 years. In this report, we highlight their major accomplishments.

The TFG on Allocation of Laboratory Tests to Different Models for Performance Specifications (TFG-DM) has worked on the criteria for assigning different measurands to each of the three Milan models [18]. The work done by the TFG-DM represents a fundamental step towards proposing practical principles for how to allocate measurands to different models. These principles will allow evaluations and clear indications for improvement; indeed, not all the PS will be immediately reachable, but they will highlight which limitations of the current technology should be prioritized and solved.

The model 1 ("outcome-based PS") should be applied when the measurand has a central and well-defined role in the decision making of a specific disease or a given clinical situation and test results should be interpreted through established decision limits. In other words, the test results should directly influence the outcome for the patient or society. For the optimal application of this model, the measurement should be standardized, so it is possible to define a common, method-independent threshold and consequently the impact of the measurement error in terms of clinical misclassification [19]. HbA_{1c} represents a good example. As in the Diabetes Control and Complications Trial (DCCT), patients in poor glycemic control had HbA_{1c} concentrations >64 mmol/mol, while those in good glycemic control had values <53 mmol/mol, it can be estimated that, to properly classify an individual with an HbA_{1c} value of 58.5 mmol/mol, the measurement error should not exceed ±5.5 mmol/mol, amounting to a relative total error (TE) of ±9.4% (5.5/58.5) [20–22]. Indeed, if the measurement error is greater, a patient with an HbA_{1c} of 58.5 mmol/mol could be randomly misclassified into both glycemic control categories (good or poor) and this obviously would not be acceptable.

The model 2 ("biological variability-based PS") should be applied when the measurand is in a "steady state" status when a subject is in good health. The TFG-DM recognizes these conditions as: (a) a situation where a measurand has to be kept at a certain concentration level in the blood otherwise the body will suffer showing symptoms (the measurands is under strict homeostatic control, e.g. plasma ions); and (b) a situation where a measurand has de facto a stable concentration, but deviations from this concentration will not in itself cause symptoms (e.g. plasma metabolites, such as urate and creatinine). For measurands not in "steady state" (e.g. urinary analytes for which the concentrations vary to maintain the corresponding plasmatic concentrations stable, compensating for dietary provision, water supplementation or loss, etc.), this model is more difficult to apply.

Finally, when a measurand cannot be placed in either model 1 or in model 2, it can be placed in model 3 ("state of the art-based PS"). This model can be temporarily used also for those measurands still waiting for the definition of outcome-based PS or while waiting for robust biological variability data.

There is a very wide variation in the analytical PS being used by EQAS providers, which adds further confusion to an already difficult situation [23]. Why are the PS in EQAS so different? Because they mean different things in different programs. In general, if PS have regulatory impact, they are looser; if PS have aspirational intention of quality improvement, they may be tighter. Different EQAS may also have different PS dependent on the analytical quality already present in the laboratories of their participants. Therefore, a harmonization through a collaborative effort, starting from clear definitions of elements to describe PS in EQAS, is needed. The TFG on Performance Specifications for EQAS (TFG-PEQAS) has identified six basic elements that need to be considered: (a) the nature of the EQAS material, including its commutability, which may affect the result interpretation; (b) the procedure used to assign the target value; (c) the data set to which PS are applied; (d) the analytical property being assessed (i.e. TE, bias, imprecision); (e) the rationale for the selection of the PS; and (f) type(s) of model used to set PS [24].

For result interpretation in EQAS, there can be different types of analytical PS, depending on the number of replicates of each EQAS sample performed by the participants. If participants measure in singlicate, a TE specification is needed (when only a single measurement is performed, the result includes effects of both bias and imprecision

errors and these cannot therefore be separated). If EQAS ask for multiple measurements of the same sample, a bias as well as an imprecision specification can be used for judgement of an EQAS result, providing that the scheme is also able to estimate the random component of the measurement uncertainty of individual participants.

The TFG-PEQAS has recommended the use of one of the PS models from the Milan conference [24]. In agreement with Infusino et al. [25], the addition of PS derived from Milan models to the EQAS categorization previously published by Miller et al. [26] as criteria to evaluate the performance of laboratories participating to EQAS should be promoted. Miller's categories 1 and 2, which fulfil the metrological requirements (commutable samples, value-assigned with reference measurement procedure), should be each split in two sub-categories: 1/2A, in which Milan models 1 and 2 for PS are applied, and 1/2B, in which other low-order models to establish PS are employed [25]. As initial steps, the TFG-PEQAS encourages EQAS organizers to provide structured descriptions of PS and to review PS providing their rationale; in a more consolidated manner, they recommend developing common EQAS PS through a collaborative process and using relevant Milan model(s), in order to support uniform performance and high quality in the total examination process [24].

The Milan conference revitalized the TE vs. measurement uncertainty discussion [9, 27]. If TE is still useful in evaluating single EQAS results, there is also a need to look carefully to the measurement uncertainty as this helps very much in the identification of important sources of bias [1, 28]. These concepts were taken forward by the TFG on Total Error (TFG-TE) established after the conference, commissioned with exploring, developing and coming up with a proposal for how to correctly use the TE concept and how to possibly combine PS for bias and imprecision in a more scientifically sound way. The TFG-TE has recalled criticisms to the conventional biological variability model for deriving allowable TE to be used in assessing quality of laboratory measurements [29]. Using this approach that includes summing of mutual exclusive terms, an overestimation of the permissible TE was demonstrated [30]. An alternative model, in which the maximum permissible bias and imprecision are interrelated and described in a curve and the permissible TE calculated from each point of the graph, has therefore to be considered to give a more realistic description of the permissible TE [31].

We already mentioned the need, recognized during the Strategic Conference, to improve information on the biological variability of laboratory analytes by applying more stringent criteria in the selection and review of available studies [11]. A dedicated TFG on Biological Variation Database (TFG-BVD) was established with the final aim to generate an EFLM biological variation database only including high quality studies and updated data after a careful appraisal by a very detailed checklist including 14 items. Structured searches for biological variability studies have been performed by four different subgroups for lipids, enzymes, metabolites and kidney-related analytes. Papers are categorized as A, B, C and D, depending on their methodological quality, with category A papers indicating high quality and D poor quality [32]. The database, hosted on the EFLM website, is currently under development and will also include derived PS for bias and imprecision of different measurands.

After the 1999 Stockholm Conference, much evidence was collected on the frequency and partition of errors in laboratory medicine and the vulnerability of both pre- and post-analytical phases has been highlighted as well as the risk for quality and patient safety [33]. Consequently, criteria for setting and harmonization of extra-analytical quality indicators have been developed and data collected [34]. This, in turn, should provide the premise to define reliable PS for the extra-analytical phases (Figure 2) [35, 36]. This was indeed the term of reference assigned to the TFG on Performance Specifications for the Extra-Analytical Phases (TFG-PSEP) [37].

According to the Milan consensus statement, in principle PS for extra-analytical phases should follow the same models as for analytical PS [8]. It is however difficult to apply model 2 to extra-analytical phases, so models 1 and 3 are therefore more usable. The TFG-PSEP has recognized that PS based on a reliable state-of-the-art, defined on surveyed quality indicator data, is the most feasible and attainable criterion to be quickly applicable [38]. The TFG-PSEP has agreed in fixing the limits for evaluation of the quality of laboratory extra-analytical performance at the 25th and 75th percentile of the distribution of data collected on extra-analytical quality indicators by the IFCC Working Group on Laboratory Error and Patient Safety. The performance of individual laboratories for each indicator is then classified according to following three quality levels in order to

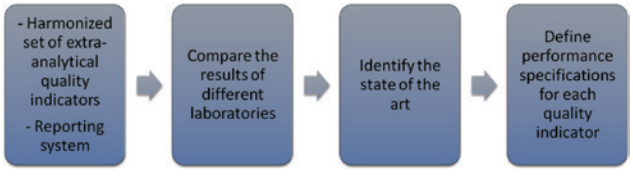

Figure 2: Proposed roadmap to define performance specifications in extra-analytical phases.

Table 2: Clinical impact of biased HDL cholesterol-risk multipliers, simulated in men with initial SCORE of 4%.

Direct method	Laboratories, n	HDL cholesterol median (range), mmol/L	Mean bias in the hypertriglyceridemic sample[a]	SCORE >5%, n (%)
Reference	1	1.08 [HDL multiplier, 1; SCORE = 4%]	0%	–
Abbott	18	1.06 (0.98–1.10)	−2.8%	0
Beckman	39	1.00 (0.79–1.17)	−6.5%	2 (5%)
Olympus	8	0.98 (0.92–1.03)	−7.4%	0
Roche	113	0.87 (0.68–1.25)	−19.4%	71 (63%)
Siemens	14	0.79 (0.62–1.20)	−22.2%	10 (71%)

Adapted from Ref. [41]. SCORE, Systematic Coronary Risk Evaluation. [a]Triglyceride concentration, 6.78 mmol/L.

allow laboratories to evaluate how they are placed in comparison with other laboratories and if improvement actions are needed [38]:

- individual results <25th percentile of value distribution: high quality performance,
- individual results between 25th and 75th percentile of value distribution: medium quality performance,
- individual results >75th percentile of value distribution: low quality performance.

It has been proposed that at the end of each year of collection, quality indicator data from participating laboratories will be processed and analyzed, so allowing the update of the 25th and 75th percentiles to be used as PS for the following year. However, this will happen only if the average performance of all participants is improving.

Stimulating studies using Milan models to obtain PS

After the Milan conference, much work has been performed to produce high quality data that can be used to set PS using the different models.

Difficulties in directly connecting laboratory testing to clinical outcomes are well known [39]. For this reason, defining PS using "indirect" outcome data (i.e. Milan model 1b) can be easier. Horvath et al. [40] have thoroughly described the prerequisites to obtain this information. Using model 1b, some recent studies have shown the clinical effects of varying analytical performance of some important tests, present in the model 1 list compiled by the TFG-DM [18].

Langlois et al. [41] have performed a simulation study showing the clinical impact of some analytical bias caused by hypertriglyceridemia on commercial assays for direct measurements of HDL and LDL cholesterol. For instance, considering HDL cholesterol and its use as risk multiplier in the systematic coronary risk evaluation (SCORE) prediction model estimating 10-years risk of cardiovascular mortality [42], when the two assays showing a negative bias of ~20% at serum triglyceride concentration ~6.8 mmol/L were used, SCORE of 4% in male individuals were falsely brought to a value >5% in approximately 2/3 of subjects (Table 2) [41]. In these cases, clinicians may wrongly modify treatment from lifestyle changes alone to drug therapy.

Given their clinical importance, cardiac troponins are one of the first biomarkers for which PS has been defined in terms of allowable misclassification rates [43]. Performing duplicate measurements, Sheehan et al. [44] calculated how many times the result of the second replicate fell in a different diagnostic group, so that defining the percentage of misclassified patients with suspected acute myocardial infarction (AMI) based on assay imprecision. Very recently, Lyon et al. [45] have applied a simulation model for estimating the misclassification rate of patients with suspected AMI when a highly sensitive troponin I assay in conjunction with its 99th percentile limit is employed. For example, a false positive rate of ~1% was obtained when both bias and imprecision of measurements were kept around 10%.

As plasma glucose plays a relevant role in diagnosis of diabetes and its values are used to define glycemic-related conditions, the outcome-based model should be preferred to derive PS for its measurement. By the way, Horvath et al. [40] stated that studies investigating the direct impact of performance of laboratory tests on clinical outcome are not necessary when: a) the clinical decisions associated with the test results are well defined, b) evidence about the diagnostic accuracy of the test to classify patients for these clinical decisions is available and is generalizable to the patient population, and c) the consequences of correct/incorrect classification are established. As all these conditions are clearly fulfilled by plasma glucose testing, model 1b can be safely adopted. Hyltoft Petersen pioneered the topic simulating

the influence of analytical bias and imprecision of fasting plasma glucose measurement on the misclassification of healthy subjects as diabetics (false positives) and of diseased subjects as healthy (false negatives) [46]. A simulation based on the current definition of impaired fasting glucose (IFG) status (i.e. fasting glucose between 6.1 and 6.9 mmol/L) has been recently done by Pasqualetti et al. [47]. They estimated that to properly classify an individual with a fasting plasma glucose of 6.5 mmol/L as IFG, the measurement error should not exceed ±6.15% (0.4/6.5). Indeed, if the measurement error is greater, a subject with a fasting glucose of 6.5 mmol/L could be randomly classified as either healthy or diabetic and this obviously would not be acceptable [47]. In a simulation analysis, a +6.15% error in plasma glucose measurement resulted in 7.7% of IFG subjects misdiagnosed as diabetes and 18.1% of healthy individuals classified as IFG. Conversely, a −6.15% error implied the shift of 6.2% subjects from diabetes to IFG category and of 12.6% IFG subjects to healthy group [48]. Interestingly, the authors showed the similarity of PS derived from the proposed model 1b approach and PS obtained by model 2 (i.e. biological variability-based), confirming the equivalence of the two models advocated by the Strategic Conference when measurands, such as fasting plasma glucose, with well-defined biological and clinical characteristics are considered [47].

The need to improve the biological variability studies has been already mentioned [12]. To this aim, the EFLM Working Group on Biological Variation (WG-BV) is working to a European project using a biobank of samples from 91 healthy subjects to be used to produce high quality data [49]. The WG-BV has recently published biological variability data for nine enzymes and creatinine in serum [50, 51]. A preliminary report is also available for serum ions [52]. Importantly, the obtained intra-individual CVs in these very strictly controlled studies are basically lower than those reported in the database available online [53]. This may translate in more stringent analytical PS, when derived using Milan model 2, than previously accounted for.

Conclusions

This paper presents an update of the activities initiated or stimulated after the EFLM Strategic Conference, held in Milan less than 3 years ago. Concepts about PS for EQAS, which measurands should use which models to set PS and how to set PS for the extra-analytical phases, together with the need for more quality data on biological variation and for further discussion about the TE concept were debated. In principle, the concepts here reported should be employed for all measurands used in the clinical setting. It is expected that this work will now be taken forward, possibly by consolidating some of these activities in a permanent structure within EFLM. The EFLM Executive Board is discussing how to carry on the issue.

Author contributions: All the authors have accepted responsibility for the entire content of this submitted manuscript and approved submission.
Research funding: None declared.
Employment or leadership: None declared.
Honorarium: None declared.
Competing interests: The funding organization(s) played no role in the study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the report for publication.

References

1. Panteghini M. Application of traceability concepts to analytical quality control may reconcile total error with uncertainty of measurement. Clin Chem Lab Med 2010;48:7–10.
2. Panteghini M. Implementation of standardization in clinical practice: not always an easy task. Clin Chem Lab Med 2012;50:1237–41.
3. Braga F, Panteghini M. Verification of in vitro medical diagnostics (IVD) metrological traceability: responsibilities and strategies. Clin Chim Acta 2014;432:55–61.
4. Fraser CG. The 1999 Stockholm Consensus Conference on Quality Specifications in Laboratory Medicine. Clin Chem Lab Med 2015;53:837–40.
5. Kallner A, McQueen M, Heuck C. The Stockholm Consensus Conference on Quality Specifications in Laboratory Medicine, 25–26 April 1999. Scand J Clin Lab Invest 1999;59:475–585.
6. Hyltoft Petersen P, Fraser CG. Strategies to set global analytical quality specifications in laboratory medicine: 10 years on from the Stockholm Consensus Conference. Accred Qual Assur 2010;15:323–30.
7. Panteghini M, Sandberg S. Defining analytical performance specifications 15 years after the Stockholm conference. Clin Chem Lab Med 2015;53:829–32.
8. Sandberg S, Fraser CG, Horvath AR, Jansen R, Jones G, Oosterhuis W, et al. 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 2015;53:833–5.
9. Panteghini M, Sandberg S. Total error vs. measurement uncertainty: the match continues. Clin Chem Lab Med 2016;54:195–6.
10. Carobene A. Reliability of biological variation data available in an online database: need for improvement. Clin Chem Lab Med 2015;53:871–7.
11. Bartlett WA, Braga F, Carobene A, Coşkun A, Prusa R, Fernandez-Calle P, et al. A checklist for critical appraisal of studies of biological variation. Clin Chem Lab Med 2015;53:879–85.

12. Braga F, Panteghini M. Generation of data on within-subject biological variation in laboratory medicine: an update. Crit Rev Clin Lab Sci 2016;53:313–25.
13. Braga F, Infusino I, Panteghini M. Performance criteria for combined uncertainty budget in the implementation of metrological traceability. Clin Chem Lab Med 2015;53:905–12.
14. Bais R, Armbruster D, Jansen RT, Klee G, Panteghini M, Passarelli J, et al. Defining acceptable limits for the metrological traceability of specific measurands. Clin Chem Lab Med 2013;51:973–9.
15. Kristensen GB, Rustad P, Berg JP, Aakre KM. Analytical bias exceeding desirable quality goal in 4 out of 5 common immunoassays: results of a native single serum sample external quality assessment program for cobalamin, folate, ferritin, thyroid-stimulating hormone, and free T4 analyses. Clin Chem 2016;62:1255–63.
16. Braga F, Frusciante E, Infusino I, Aloisio E, Guerra E, Ceriotti F, et al. Evaluation of the trueness of serum alkaline phosphatase measurement in a group of Italian laboratories. Clin Chem Lab Med 2017;55:e47–50.
17. Weykamp C, Secchiero S, Plebani M, Thelen M, Cobbaert C, Thomas A, et al. Analytical performance of 17 general chemistry analytes across countries and across manufacturers in the INPUtS project of EQA organizers in Italy, the Netherlands, Portugal, United Kingdom and Spain. Clin Chem Lab Med 2017;55:203–11.
18. Ceriotti F, Fernandez-Calle P, Klee GG, Nordin G, Sandberg S, Streichert T, et al. EFLM Task and Finish Group on Allocation of Laboratory Tests to Different Models for Performance Specifications (TFG-DM). Criteria for assigning laboratory measurands to models for analytical performance specifications defined in the 1st EFLM Strategic Conference. Clin Chem Lab Med 2017;55:189–94.
19. Hyltoft Petersen P, Klee GG. Influence of analytical bias and imprecision on the number of false positive results using guideline-driven medical decision limits. Clin Chim Acta 2014;432:127–34.
20. Mosca A, Branca MT, Carta M, Genna ML, Giorda CB, Ghidelli R, et al. Recommendations for the implementation of international standardization of glycated hemoglobin in Italy. Clin Chem Lab Med 2010;48:623–6.
21. Weykamp CW, Mosca A, Gillery P, Panteghini M. The analytical goals for hemoglobin A1c measurement in IFCC units and National Glycohemoglobin Standardization Program units are different. Clin Chem 2011;57:1204–6.
22. Braga F, Panteghini M. Standardization and analytical goals for glycated hemoglobin measurement. Clin Chem Lab Med 2013;51:1719–26.
23. Jones GR. Analytical performance specifications for EQA schemes – need for harmonisation. Clin Chem Lab Med 2015;53:919–24.
24. Jones GR, Albarede S, Kesseler D, MacKenzie F, Mammen J, Pedersen M, et al. Analytical performance specifications for external quality assessment – definitions and descriptions. Clin Chem Lab Med 2017;55:949–55.
25. Infusino I, Frusciante E, Braga F, Panteghini M. Progress and impact of enzyme measurement standardization. Clin Chem Lab Med 2017;55:334–40.
26. Miller WG, Jones GR, Horowitz GL, Weykamp C. Proficiency testing/external quality assessment: current challenges and future directions. Clin Chem 2011;57:1670–80.
27. Westgard JO. Useful measures and models for analytical quality management in medical laboratories. Clin Chem Lab Med 2016;54:223–33.
28. Pasqualetti S, Infusino I, Carnevale A, Szőke D, Panteghini M. The calibrator value assignment protocol of the Abbott enzymatic creatinine assay is inadequate for ensuring suitable quality of serum measurements. Clin Chim Acta 2015;450:125–6.
29. Oosterhuis WP, Bayat H, Armbruster D, Coskun A, Freeman KP, Kallner A, et al. The use of error and uncertainty methods in the medical laboratory. Clin Chem Lab Med 2017 Aug 10. pii: /j/cclm.ahead-of-print/cclm-2017-0341/cclm-2017-0341.xml. doi: 10.1515/cclm-2017-0341. [Epub ahead of print].
30. Oosterhuis WP. Gross overestimation of total allowable error based on biological variation. Clin Chem 2011;57:1334–6.
31. Larsen ML, Fraser CG, Petersen PH. A comparison of analytical goals for haemoglobin A1c assays derived using different strategies. Ann Clin Biochem 1991;28:272–8.
32. Fernandez-Calle P, Diaz-Garzon J, Minchinela J, Roraas T, Aarsand A, Sandberg S. Estimates of within-subject biological variation for total cholesterol, HDL-cholesterol, LDL-cholesterol, triglycerides and apolipoproteins A1, B based on review of published papers using a critical appraisal check list – Results from the EFLM TFG-BVD. Clin Chem Lab Med 2017;55:S1068.
33. Plebani M. Errors in clinical laboratories or errors in laboratory medicine? Clin Chem Lab Med 2006;44:750–9.
34. Plebani M, Astion ML, Barth JH, Chen W, de Oliveira Galoro CA, Escuer MI, et al. Harmonization of quality indicators in laboratory medicine. A preliminary consensus. Clin Chem Lab Med 2014;52:951–8.
35. Plebani M, Sciacovelli L, Aita A, Pelloso M, Chiozza ML. Performance criteria and quality indicators for the pre-analytical phase. Clin Chem Lab Med 2015;53:943–8.
36. Sciacovelli L, Aita A, Padoan A, Pelloso M, Antonelli G, Piva E, et al. Performance criteria and quality indicators for the post-analytical phase. Clin Chem Lab 2016;54:1169–76.
37. Plebani M, on behalf of the EFLM Task and Finish Group on Performance Specifications for the Extra-analytical Phases. Performance specifications for the extra-analytical phases of laboratory testing: Why and how. Clin Biochem 2017;50:550–4.
38. Sciacovelli L, Panteghini M, Lippi G, Sumarac Z, Cadamuro J, De Olivera Galoro CA, et al. Defining a roadmap for harmonizing quality indicators in laboratory medicine: a consensus statement on behalf of the IFCC Working Group "Laboratory Error and Patient Safety" and EFLM Task and Finish Group "Performance specifications for the extra-analytical phases". Clin Chem Lab Med 2017;55:1478–88.
39. Ferraro S, Braga F, Panteghini M. Laboratory medicine in the new healthcare environment. Clin Chem Lab Med 2016;54:523–33.
40. Horvath AR, Bossuyt PM, Sandberg S, St John A, Monaghan PJ, Verhagen-Kamerbeek WD, et al. Setting analytical performance specifications based on outcome studies – is it possible? Clin Chem Lab Med 2015;53:841–8.
41. Langlois MR, Descamps OS, van der Laarse A, Weykamp C, Baum H, Pulkki K, et al. Clinical impact of direct HDLc and LDLc method bias in hypertriglyceridemia. A simulation study of the EAS-EFLM Collaborative Project Group. Atheroscler 2014;233:83–90.
42. Descamps OS, Cooney MT, De Backer G, Graham I. A simple multiplier to calculate the impact of HDL cholesterol on cardiovascular risk estimation using SCORE. Atheroscler 2012;222:564–6.

43. Panteghini M. Quality requirements for troponin assays – an overview. In: troponin monograph 2012, The Australasian Association of Clinical Biochemists Inc., 2012:53–61.
44. Sheehan P, Blennerhassett J, Vasikaran SD. Decision limit for troponin I and assay performance. Ann Clin Biochem 2002;39:213–6.
45. Lyon AW, Kavsak PA, Lyon OA, Worster A, Lyon ME. Simulation models of misclassification error for single thresholds of high-sensitivity cardiac troponin I due to assay bias and imprecision. Clin Chem 2017;63:585–92.
46. Hyltoft Petersen P, Brandslund I, Jørgensen LG, Stahl M, de Fine Olivarius N, Borch-Johnsen K. Evaluation of systematic and random factors in measurements of FPG as the basis for analytical quality specifications in the diagnosis of diabetes. 3. Impact of the new WHO and ADA recommendations on diagnosis of diabetes mellitus. Scand J Clin Lab Invest 2001;61:191–204.
47. Pasqualetti S, Braga F, Panteghini M. Pre-analytical and analytical aspects affecting clinical reliability of plasma glucose results. Clin Biochem 2017;50:587–94.
48. Pasqualetti S, Braga F, Panteghini M. Impact of measurement error of plasma glucose on clinical classification: a simulation analysis. Clin Chem Lab Med 2017;55:S146.
49. Carobene A, Strollo M, Jonker N, Barla G, Bartlett WA, Sandberg S, et al. Sample collections from healthy volunteers for biological variation estimates' update: a new project undertaken by the Working Group on Biological Variation established by the European Federation of Clinical Chemistry and Laboratory Medicine. Clin Chem Lab Med 2016;54:1599–608.
50. Carobene A, Røraas T, Sølvik U, Sylte MS, Sandberg S, Guerra E, et al. Biological variation estimates obtained from 91 healthy study participants for 9 enzymes in serum. Clin Chem 2017;63:1141–50.
51. Carobene A, Marino I, Coşkun A, Serteser M, Unsal I, Guerra E, et al. The EuBIVAS Project: Within- and between-subject biological variation data for serum creatinine using enzymatic and alkaline picrate methods and implications for monitoring. Clin Chem 2017;63:1527–36.
52. Carobene A, Marino I, Guerra E, Jonker N, Barla G, Bartlett WA, et al. Biological variation estimates obtained from 91 healthy subjects for six electrolytes in serum. EBIOVAR study of the EFLM Working-Group on Biological Variation. Clin Chem Lab Med 2017;55:S170.
53. Minchinela J, Ricos C, Perich C, Fernandez-Calle P, Alvarez V, Domenech MV, et al. Biological variation database and quality specifications for imprecision, bias and total error (desirable and minimum). The 2014 update. Available at: http://www.westgard.com/biodatabase-2014-update.htm. Accessed Aug 2017.