Clinical Biochemistry 57 (2018) 23–28

Contents lists available at ScienceDirect

Clinical Biochemistry

journal homepage: www.elsevier.com/locate/clinbiochem

The role of external quality assessment in the verification of in vitro medical diagnostics in the traceability era

Federica Braga*, Sara Pasqualetti, Mauro Panteghini

Research Centre for Metrological Traceability in Laboratory Medicine (CIRME), University of Milan, Milan, Italy

ARTICLE INFO

Keywords:
External quality assessment
Measurement uncertainty
Standardization

ABSTRACT

Once an in-vitro diagnostic (IVD) measuring system has been marketed and introduced into daily practice, the possible sources of degradation of its performance are numerous. It is therefore essential to put in place a continuous post-market surveillance of the quality of performance of the IVD system and of the laboratories that perform measurements in clinical setting. The participation to external quality assessment (EQA) schemes that meet specific metrological criteria is central to the evaluation of performance of clinical laboratories in terms of standardization and clinical suitability of their measurements. In addition to the use of commutable materials, in this type of EQA it is necessary to assign values (and uncertainty) to them with reference procedures and to define and apply clinically permissible analytical performance specifications to substantiate the suitability of laboratory measurements in the clinical setting. Unfortunately, there are still few permanent EQA programs fully covering these requirements because some practical constraints, including technical and economic aspects, which limit their introduction. It is, however, clear that these issues should be quickly overcome, since EQA schemes are in a unique position to add substantial value to the practice of laboratory medicine, by identifying analytes that need improved harmonization and by stimulating and sustaining standardization initiatives that are needed to support clinical practice. Importantly, this will definitively help those manufacturers that produce superior products to demonstrate the superiority of those products and oblige end users (and consequently industry) to abandon assays with demonstrated insufficient quality.

1. Introduction

Clinicians and patients expect laboratory results to be equivalent, independent of the laboratory producing them and of employed measuring systems. To reach this goal, it is essential to implement a measurement standardization process based on metrological traceability of patient results to higher order references [1]. To this regard, the main requirements are: a) the identification of an unbroken metrological traceability chain to implement the appropriate transfer of the measurement trueness to commercial calibrators and patient results, b) the estimate of measurement uncertainty budget, with definition of permissible limits for clinical application of measurements, and c) a post-market surveillance surveying the suitability of in vitro diagnostics (IVD) for clinical use and of laboratory performances in using them [2]. Professionals, IVD manufacturers and clinical laboratories are the heading actors of this process and their roles and responsibilities have been previously discussed [3,4]. Using the standardization of enzyme measurements as an example, Infusino et al. [5] have further described the main components and expected consequences of this process, summarized in Table 1.

2. Traceability implementation: A long and bumpy road

By selecting suitable reference materials and/or identifying laboratories performing reference measurement procedures (RMP), IVD manufacturers have to define a calibration hierarchy to assign traceable values to their commercial calibrators [2,4]. During this process, the measurement bias has to be appropriately eliminated in order to guarantee the trueness transfer. Furthermore, IVD manufacturers should estimate the measurement uncertainty associated to commercial calibrators and compare it to uncertainty budget limits, which represent a proportion of the uncertainty budget allowed for clinical results [6]. IVD manufacturers assume, therefore, total responsibility for supplying products of acceptable quality in terms of traceability and measurement uncertainty of the system. For this reason, the commercial measuring system, composed by platform, reagents, calibrators and control materials (to be used for the verification of system alignment) should be considered as a whole [4].

* Corresponding author at: Research Centre for Metrological Traceability in Laboratory Medicine (CIRME), Via G.B. Grassi 74, Milano, Italy.
E-mail address: federica.braga@unimi.it (F. Braga).

https://doi.org/10.1016/j.clinbiochem.2018.02.004
Received 2 January 2018; Received in revised form 31 January 2018; Accepted 7 February 2018
Available online 09 February 2018

Table 1
Main steps and expected consequences of an effective measurement standardization process.

- In vitro diagnostic manufacturers implement traceability to higher order references as defined by professional bodies (e.g., Joint Committee on Traceability in Laboratory Medicine)
- Users (and industry) abandon non-specific methods impossible to standardize
- External quality assessment schemes provide commutable materials and trueness-based grading to check and survey measurement traceability
- Professionals establish clinically permissible measurement errors
- Individual laboratories monitor their performance by participating to external quality assessment that meets metrological criteria and by applying permissible limits

The CE ("European Communities") mark is a certification that should be applied by manufacturers on the whole measuring system to indicate its compliance with the European legislative requirements [the European Union (EU) Directive and, in the foreseeable future, the new EU Regulation] requesting, among other things, to ensure traceability of commercial measuring systems to recognized higher order references. The CE marking on IVD products suffers, however, from important limitations. First, CE mark by itself does not guarantee that the manufacturer has transferred trueness successfully. Secondly, CE mark does not mean that uncertainty of the calibrators meets clinical needs.

Recently, we retraced the introduction of correctly standardized assays in clinical enzymology [7]. Years after the availability of reference measurement systems for the most popular enzymes, some manufacturers continue to market assays giving results that are not traceable to them. Aminotransferases represent an important example, in which assay performance in terms of traceability vary considerably also within users of instruments from the same manufacturer [8]. This is mainly due to the use on the same platforms of various reagents with different analytical selectivity for these enzymes. For measuring aminotransferases, almost all manufacturers still market assays with or without the addition of pyridoxal-5-phosphate (P-5′-P). Assays without P-5′-P activation give, however, significantly different results compared to assays with P-5′-P activation and are often unable to fulfil quality specifications when aminotransferase results are compared to the RMP [9]. More importantly, the still significant differences in measuring aminotransferases carry the risk that many individuals might be misclassified when clinicians start using universally recommended cut-offs [10].

The commercial availability of measuring systems with different analytical selectivity points to the need for clinical laboratories to take responsibility to move to assays displaying similar selectivity when compared to higher order references. In doing this, it is not enough to base the choice on the manufacturer's declaration, but it is mandatory to rely on experimental studies and on a careful inspection of characteristics of commercial products. Our recent experience evaluating the level of standardization of serum alkaline phosphatase (ALP) measurement is a nice exemplification of the current state of things [11]. In this study, only about one fourth of participant laboratories were able to fulfil the desirable goal for bias when their results were compared with target values set by a reference laboratory. Interestingly, the ability to meet the goal was clearly dependent on the measuring system used [11]. Using traceability information reported by the manufacturers, one could largely explain this unsatisfactory situation. For instance, Roche Diagnostics was anchored to an outdated ALP method. By the way, based on the results of the mentioned study, this manufacturer has recently recalibrated their ALP assays to the 2011 IFCC RMP to fulfil current requirements of the traceability to higher order references. On the other hand, alternatives that do not comply with the EU Directive, e.g., the Beckman AU system declaring traceability to an internal "master" calibrator that is not a high metrological order material, are still remaining undisturbed on the market.

Keeping the uncertainty of commercial calibrators to a level fulfilling clinical needs is also a disregarded issue. In our laboratory, serum creatinine measurement is performed with the enzymatic assay on the Abbott Diagnostics Architect c16000 platform. The calibrator of the measuring system is traceable to the NIST SRM 967 reference material. Fig. 1 depicts the metrological traceability chain and measurement

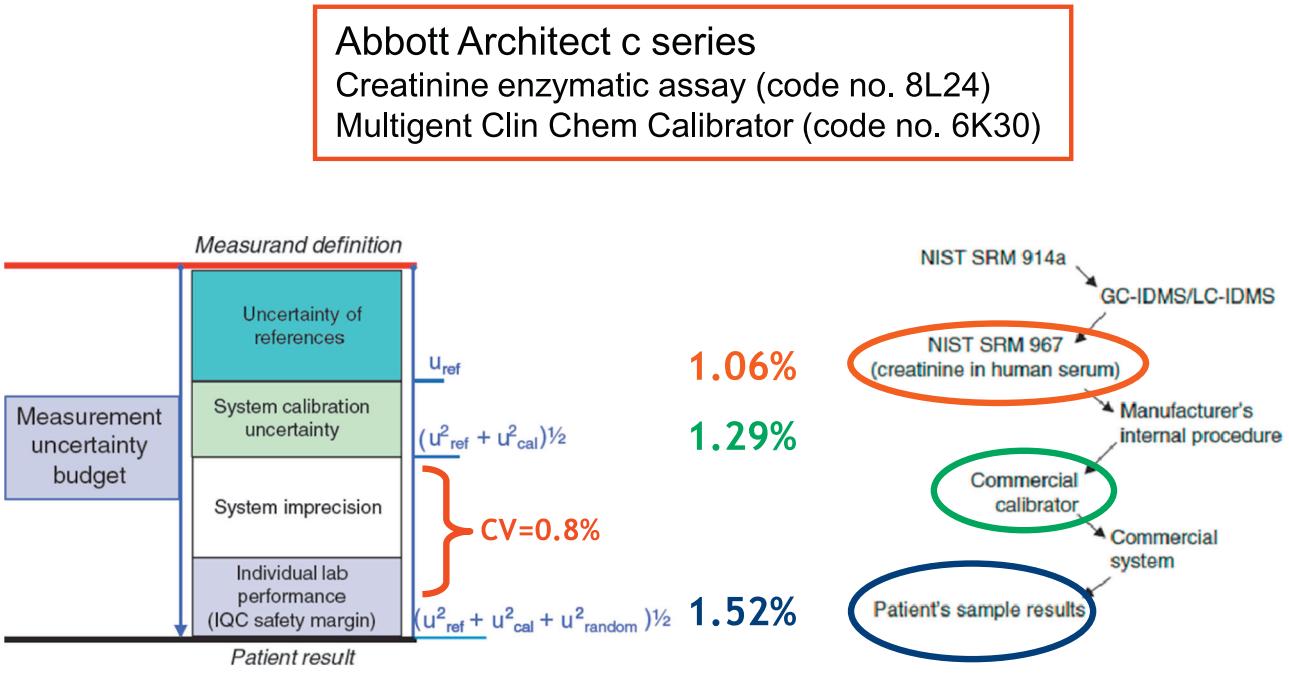

Fig. 1. Metrological traceability chain (right) and measurement combined standard uncertainties (left) of Abbott Architect enzymatic creatinine assay, as described by the manufacturer. Note that performance specifications for uncertainty of serum creatinine measurement on clinical samples are 3.0% (desirable quality) and 1.5% (optimum quality) [12]. Data of long-term CV (including measuring system imprecision and individual laboratory performance) are from the authors' laboratory.
U, standard uncertainty; IQC, internal quality control; GC-IDMS, gas chromatography-isotope dilution mass spectrometry; LC-IDMS, liquid chromatography-isotope dilution mass spectrometry.

uncertainty of this measuring system, as declared by the manufacturer. Note that the measuring system as described perfectly fulfils analytical performance specifications (APS) for combined uncertainty of serum creatinine measurement on clinical samples. After the introduction of a new lot of calibrator, we observed, however, a constant overestimation (in average, +8%) of creatinine results during the participation in the regional external quality assessment (EQA) scheme. As the analytical imprecision of the method was optimal, we verified the trueness of the measuring system by using the NIST SRM 967a reference material [12]. This confirmed the presence of significant positive bias when the measuring system was calibrated with the new lot of calibrator. Consequently, the combined uncertainty experimentally obtained using this calibrator lot on clinical samples was much higher than the desirable performance goal [12]. After inquiring, we discovered that Abbott Diagnostics considered as validation criterion for traceability of different calibrator lots an internal specification of ± 5% from the target value of SRM 967a level 1. Our experience showed that this criterion is too large to comply with the combined uncertainty goal for creatinine measurements in clinical samples and raised important concerns about the manufacturer's calibrator value-assignment protocol probably inadequate for ensuring suitable quality of serum measurements.

3. The fallibility of the “peer group” approach in EQA

One could argue that, in evaluating EQA results, the approach using the “peer group”-based value as reference may mitigate the scenario of heterogeneity of marketed assays. However, this approach has major limitations that configure an insurmountable fallibility. First, the definition of “peer group” is heterogeneous per se. Is a ‘peer group’ consisting of a) the same model instrument/reagents/calibrator from one manufacturer, b) the same instrument family from one manufacturer, c) instruments from different manufacturers that use the same reagent and calibrator, or d) methods with the same measurement principle with different reagents and calibrators? Considering traceability of commercial measuring systems, each of these definitions exposes to flaws. Recently, we reported two clarifying experiences about this issue [13,14].

To comply with EU Directive demand, Abbott Diagnostics correctly validates the traceability of the ALP assay for the Architect c series systems by comparison to results from IFCC RMP, providing the users with an “experimental” (i.e., IFCC standardized) calibration factor [15]. Using this factor, however, we observed a constant positive error during the participation to the regional EQA scheme when our ALP results were compared with the median of the Architect users' group. This performance was completely unexpected having previously verified the perfect alignment of our Architect ALP assay to the IFCC RMP during the study mentioned above [11]. Nevertheless, we decided to investigate once again the trueness of our Architect ALP assay by performing a dedicated correlation experiment with the IFCC RMP. The results of this further evaluation confirmed that, despite our EQA results, the Architect ALP assay in our hands, calibrated with the ‘experimental’ factor, is optimally standardized, the estimated combined measurement uncertainty being 2%, well within the previously advocated desirable performance specification of ± 3% [7,13]. However, this is probably not true when the assay is calibrated with the “theoretical” factor, derived by the manufacturer from the p-nitrophenol molar extinction coefficient of an optimized reaction for ALP, that Abbott provides as ‘first choice’ to its users, declaring however a non-negligible bias (−7%) between results obtained with two calibration factors. As the great majority (87%) of laboratories participating to regional EQA scheme employing the Architect analytical systems declared to using the latter factor, the ‘peer-group’ (i.e., Architect system-specific) consensus value used in the EQA was therefore expected to be strongly influenced by this type of calibration adopted by the majority of laboratories. Consequently, this significantly lowers the EQA value used as a reference for evaluating the performance of participating laboratories and could explain the (apparent) positive error of our laboratory. Based on these results, we expect that Abbott Diagnostics withdraw the so-called ‘theoretical’ calibration factor from the Architect ALP package insert, recommending users only employing the factor obtained by correlation results using clinical samples with RMP-assigned values [15].

In a second study, we reported another example demonstrating the risks of using measuring-system consensus values in EQA [14]. The CE-marked Architect measuring system for serum transferrin determination offers two calibrator material options, the Abbott Specific Proteins Multiconstituent Calibrator (SPMC, ref. 1E78) or the Plasmaproteins Cal (PC, ref. 11200D), both declared to be traceable to the reference material ERM-DA470/IFCC, although with a markedly different expanded uncertainty (1.83% for SPMC and 4.17% for PC, respectively). Considering that the expanded uncertainty of transferrin value in the ERM-DA470/IFCC is 3.39%, it is clear that the declared uncertainty of SPMC is not combined with that of the reference material, as actually recommended [4,6]. This highlights a first issue that makes the uncertainties of the two calibrators not comparable. More importantly, given the availability of two options warranting the CE mark provided by Abbott to calibrate the Architect transferrin assay, one is expecting equivalence in transferrin results obtained by interchanging SPMC with PC. This was specifically investigated in a correlation study showing, however, a systematic error between the two calibrator options (in average, 7.8% biased) with a high probability to markedly influence the clinical use of the test [14]. As outcome of this study, we expect that the manufacturer carefully review the internal protocol for transferring trueness (and estimating uncertainty) from ERM-DA470/IFCC to calibrators of Architect transferrin measuring system.

4. How to improve the post-market surveillance of IVD medical devices: The role of EQA

Given the examples reported above, it is clear that we must improve the post-market surveillance in terms of traceability and uncertainty of IVD products. As previously discussed [2–7], this surveillance relies substantially on the analytical quality control, which should be, however, redesigned in agreement with metrological concepts. In particular, the participation to EQA programs that meet specific metrological criteria is mandatory for the evaluation of performance of laboratories in terms of traceability and clinical suitability of their measurements. Firstly, the value (and uncertainty) of control materials used for EQA should be assigned with RMP [or, if a RMP is lacking, by reference institutions (possibly including the manufacturer releasing that specific measuring system)] and their commutability should be demonstrated. Secondly, the adoption of APS based on the clinically acceptable measurement error is essential to verify the suitability of laboratory measurements in the clinical setting.

4.1. Quality of EQA target

The value assignment by RMP to EQA materials allows objective evaluation of the performance of laboratory measurements through a trueness-based (instead of inferior consensus-based) grading of the competency of participating clinical laboratories. To cover this important requirement, it is essential that value of EQA materials be assigned by a competent reference laboratory, able to provide results within narrow limits of uncertainty [16].

4.2. Commutability of EQA materials

Commutability is defined as the equivalence of the mathematical relationships between the results of different measurement procedures for a control material and for representative clinical samples [17,18]. Commutability of control materials is one of the most important aspects affecting the design and interpretation of EQA programs, as only the use

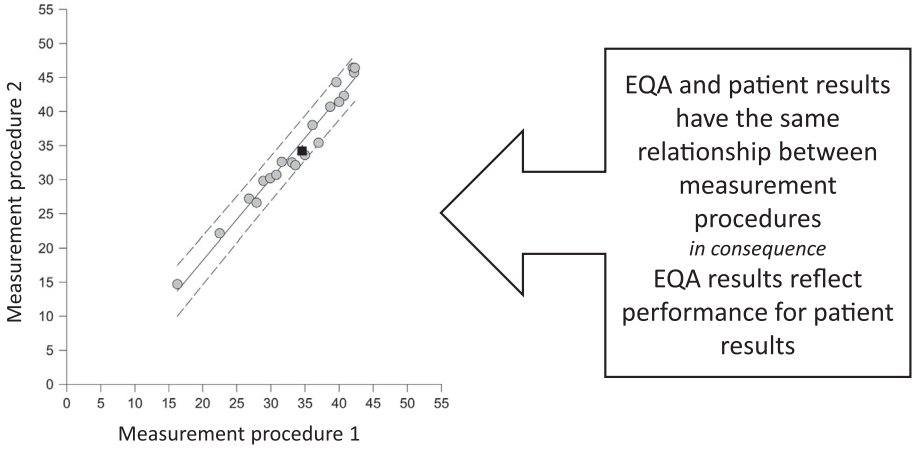

Fig. 2. Schematic diagram showing the behaviour and the characteristics of a commutable external quality assessment (EQA) material (black square) when compared with native samples (grey circles). Continuous line, regression line; dashed lines, 95% prediction interval.
Note: Measurement procedure 1 in x-axis should be reference measurement procedure when available.

Table 2
Constraints limiting the introduction of External Quality Assessment (EQA) schemes that meet metrological criteria. Adapted from ref. 20.

- Technical aspects: lack of certified control materials or difficulties to prepare commutable samples
- Practical considerations: complicated logistics of distribution of frozen samples
- Educational limitations: lack of awareness of which quality factors make an EQA important
- Economic concerns: higher costs

of commutable samples allows transferability of participating laboratory performance to clinical samples (Fig. 2) [19]. Unfortunately, EQA samples are frequently not validated for commutability. In principle, commutability is assumed based on how the control materials were prepared. It may be reasonable for single donation, while potential limitations exist for spiked or supplemented pools or for materials that are more artificial. The use of single-donor samples, which is preferable to overcome commutability problems, may limit however the achievement of desired concentrations. Another important practical limitation of single donation is the difficulty to obtain adequate volumes of samples needed for preparing sufficient amount of control materials. On the other hand, pooled samples have the potential limitation that interactions of components such as proteins may cause modification of the matrix [20]. The Clinical and Laboratory Standards Institute (CLSI) has described a rigorous protocol to collect blood, obtain serum, prepare a pool and freeze aliquots under conditions that do not alter the commutability characteristics [21]. For example, blood donated for EQA schemes may be collected in plastic 'transfusion' bags. Failure to separate the serum promptly (as recommended in the guidelines) can result in leaching of materials from the plastic (especially plasticizers), which can cause commutability problems.

4.3. Analytical performance specifications for EQA

In addition to the use of commutable materials with assigned value by RMP, EQA programs should evaluate results of participating laboratories using adequate APS. To this regard, the 1st Strategic Conference of the European Federation of Clinical Chemistry and Laboratory Medicine (EFLM) held in 2014 revisited the hierarchy of models for deriving APS [22,23]. The recommended approaches should rely on the effect of measurement performance on clinical outcome (model 1) or on the biological variation of the measurand (model 2). If APS based on these models could not be made, state-of-the-art, defined as the highest level of performance technically achievable, could be used (model 3). Ceriotti et al. [24] has worked on the criteria for assigning different measurands to each of the three EFLM models, preparing a preliminary list related to common measurands. We previously recommended the addition of APS derived from these models to the EQA scheme categorization published by Miller et al. [20] as criteria to evaluate the performance of laboratories participating to EQA [7,25]. Miller's categories 1 and 2, which fulfil the metrological requirements highlighted under items 4.1 and 4.2 above, should be each split in two sub-categories: 1/2A, in which EFLM models 1 and 2 for APS are applied, and 1/2B, in which other low-order models are employed [7,23,25]. The possibility of grading different levels of quality (i.e., minimum, desirable and optimum) to move, in case, from desirable to minimum goals and, in the meantime, ask IVD manufacturers to work for improving the quality of assay performance is also an important aspect to consider [23].

4.4. Towards a perfect EQA

Currently, there are few permanent EQA programs covering all metrological requirements described above. Table 2 lists major constraints supposed to limit their introduction [20].

In describing factors influencing choice of EQA schemes by users, James et al. [26] were unable to include the value-assignment by RMP and the type of employed performance limits, while demonstration of material commutability only appeared at the last point of the list. The value (and uncertainty) assignment to control materials by an accredited reference laboratory involves some economic efforts by EQA organizers, frequently raising expediency concerns. Furthermore, the majority of EQA schemes use lyophilized materials, which usually behave differently from authentic clinical specimens. Finally, there is a wide variation in the APS being used by EQA providers, which adds further confusion [27]. In particular, there seems to be a sort of tug-of-war between the adoption of wider goals, surely more practical and less exclusive, and the use of tighter goals that would instead help to improve test quality and laboratory performances [23]. Anyway, because of different APS used in EQA schemes, different information about measuring systems and/or performance of participants is potentially being conveyed.

Jones et al. [28] has recently identified six basic elements that need to be considered and described to significantly improve EQA: a) the nature of the EQA material, including its commutability; b) the procedure used to assign the target value; c) the data set to which APS are applied; d) the analytical property being assessed (i.e., total error, bias, imprecision); e) the rationale for the APS selection; and f) type(s) of model used to set them.

4.5. Some laudable examples

Although the implementation of EQA programs fulfilling abovementioned requirements is difficult, their unique benefits have been incontrovertibly proved [11,29–34]. We previously summarized these advantages, highlighting that EQA schemes that meet metrological

criteria are in a unique position, permitting to identify analytes that need improved harmonization and to stimulate standardization initiatives by creating evidence about intrinsic equivalence (or not) of results from different measuring systems [35]. What appears clear from the published experiences is that sometimes we probably have an optimistic perception of analytical quality in clinical laboratories, due to the traditional approaches for evaluating their performances. The INPUtS project, aiming to evaluate the performance of the measurement of 17 common chemistry analytes across five European countries, reported that six analytes (plasma sodium, chloride, calcium, magnesium, total proteins and ALP) are currently unable to meet the minimum APS [36]. Similarly, Kristensen et al. [37], in evaluating performances of five common analytes measured by immunoassays, showed that the analytical bias of ferritin and free thyroxine measurements exceeds desirable APS, independently of the evaluated measuring system. In these cases, questionable results should force manufacturers to investigate and fix the problems and then improve the performance. Some examples show that this approach works.

The need for urgently improving alignment to higher order references (i.e., the ERM-DA470/IFCC reference material) of commercial assays for serum albumin measurement was shown in 2011 from two independent sources [38,39]. EQA survey data showed an extremely large bias addressing to some inconsistency in the assignment of values to the manufacturer's calibrators. Quite recently, Bachmann et al. [40] have revaluated the state of standardization of 24 commercial albumin measuring systems, showing a general improvement in their metrological alignment. In particular, the Tina-quant Albumin Gen. 2 immunoturbidimetric assay carried out on the Cobas platforms (Roche Diagnostics), judged quite inaccurate in our original experience [38], has shown a perfect recovery of the ERM-DA470k/IFCC reference material, supporting an optimal traceability of its calibrator.

In 2013, in evaluating the traceability chain for glycated hemoglobin (HbA_{1c}) as measured in our laboratory, we noted that the combined standard uncertainty associated with the measurement of clinical samples (~2.0%) was higher than the minimum acceptable target derived from biological variability of the measurand [41]. Then, we advocated advances in the quality of HbA_{1c} measurement, by reducing uncertainty associated with higher order metrological references and by increasing the precision of commercial HbA_{1c} assays [41]. Working on both issues, we were recently able to demonstrate that, for this analyte, it is possible to reduce the combined uncertainty associated with the measurement of clinical samples to a level (~1.45%) close to the desirable goal [42].

4.6. Do not forget the post-analytical phase: The case of serum folate

To fulfil the requirements of EU Directive, in 2016 Roche Diagnostics has replaced the 'home-made' calibrated folate assay with an assay traceable to WHO NIBSC 03/178 International Standard. After the recalibration, we experienced a significant shift in the average folate measured values in our laboratory. Particularly, at serum folate concentrations around the lower reference limit (LRL) of the old Roche assay, we observed a positive bias of ~50% vs. the new Roche assay recalibrated assay [43]. Taking into account this difference, the shift from 4.6 μg/L (Roche recommended LRL for old calibration) to 3.9 μg/L (Roche recommended LRL for recalibrated assay) appeared to be inconsistent. Consequently, a misleading overestimate of the prevalence of folate deficiency was expected if the recalibrated Roche assay is used together the manufacturer's newly recommended LRL, with potential consequences on patient's safety. The experimental estimate of LRL using Roche WHO traceable assay on a population free of folate supplementation actually revealed that this value (1.3 μg/L) is far lower than that reported by the manufacturer in the assay package insert [44].

5. Conclusion

The organization and participation to EQA schemes that meet metrological criteria is a central requirement allowing an appropriate post-market surveillance of IVD medical devices in terms of standardization and clinical suitability of laboratory measurements. Conventional EQA using non-commutable materials, consensus 'peer' group assessment and not clinically oriented APS must be discontinued. We should stop comparing "our site's apple to another site's oranges". Only by well-structured EQA is possible objectively evaluating the reliability of commercial measuring systems and the quality of measurements provided by clinical laboratories. Despite the majority of EQA programs still does not cover these requirements, some excellent experiences have shown the feasibility to "create a perfect external control system" [45].

References

[1] M. Panteghini, Traceability as a unique tool to improve standardization in laboratory medicine, Clin. Biochem. 42 (2009) 236–240.
[2] M. Panteghini, Implementation of standardization in clinical practice: not always an easy task, Clin. Chem. Lab. Med. 50 (2012) 1237–1241.
[3] M. Panteghini, Application of traceability concepts to analytical quality control may reconcile total error with uncertainty of measurement, Clin. Chem. Lab. Med. 48 (2010) 7–10.
[4] F. Braga, M. Panteghini, Verification of in vitro medical diagnostics (IVD) metrological traceability: responsibilities and strategies, Clin. Chim. Acta 432 (2014) 55–61.
[5] I. Infusino, G. Schumann, F. Ceriotti, M. Panteghini, Standardization in clinical enzymology: a challenge for the theory of metrological traceability, Clin. Chem. Lab. Med. 48 (2010) 301–307.
[6] F. Braga, I. Infusino, M. Panteghini, Performance criteria for combined uncertainty budget in the implementation of metrological traceability, Clin. Chem. Lab. Med. 53 (2015) 905–912.
[7] I. Infusino, E. Frusciante, F. Braga, M. Panteghini, Progress and impact of enzyme measurement standardization, Clin. Chem. Lab. Med. 55 (2017) 334–340.
[8] R. Jansen, N. Jassam, A. Thomas, C. Perich, P. Fernandez-Calle, H. Correia AP Faria, C. Weykamp JH Barth, C. Cobbaert, M. Thelen, C. Ricós, A category 1 EQA scheme for comparison of laboratory performance and method performance: an international pilot study in the framework of the calibration 2000 project, Clin. Chim. Acta 432 (2014) 90–98.
[9] K. Goossens, K. Van Uytfanghe, L.M. Thienpont, Trueness and comparability assessment of widely used assays for 5 common enzymes and 3 electrolytes, Clin. Chim. Acta 442 (2015) 44–45.
[10] M. Panteghini, K. Adeli, F. Ceriotti, S. Sandberg, A.R. Horvath, American liver guidelines and cutoffs for "normal" ALT: a potential for overdiagnosis, Clin. Chem. 63 (2017) 1196–1198.
[11] F. Braga, E. Frusciante, I. Infusino, E. Aloisio, E. Guerra, F. Ceriotti, M. Panteghini, Evaluation of the trueness of serum alkaline phosphatase measurement in a group of Italian laboratories, Clin. Chem. Lab. Med. 55 (2017) e47–e50.
[12] S. Pasqualetti, I. Infusino, A. Carnevale, D. Szőke, M. Panteghini, The calibrator value assignment protocol of the Abbott enzymatic creatinine assay is inadequate for ensuring suitable quality of serum measurements, Clin. Chim. Acta 450 (2015) 125–126.
[13] E. Aloisio, E. Frusciante, S. Pasqualetti, M. Quercioli, M. Panteghini, Traceability of alkaline phosphatase measurement may also vary considerably using the same analytical system: the case of Abbott Architect, Clin. Chem. Lab. Med. (2017) Dec 19. pii: /j/cclm.ahead-of-print/cclm-2017-1007/cclm-2017-1007.xml https://doi.org/10.1515/cclm-2017-1007 (Epub ahead of print, in press).
[14] S. Pasqualetti, A. Carnevale, E. Aloisio, A. Dolci, M. Panteghini, Different calibrator options may strongly influence the trueness of serum transferrin measured by Abbott Architect systems, Clin. Chim. Acta 478 (2018) 119–120.
[15] D. Armbruster, D. Yahalom, L. Lennartz, M. Orth, Metrological traceability of Architect amylase and alkaline phosphatase assays to IFCC reference methods, Clin. Chem. 61 (2015) S208–S209.
[16] D. Stöckl, H. Reinauer, Development of criteria for the evaluation of reference method values, Scand. J. Clin. Lab. Invest. 212 (Suppl) (1993) 16–18.
[17] W.G. Miller, H.G. Schimmel, R. Rej, N. Greenberg, F. Ceriotti, C.J. Burns, J.R. Budd, C. Weykamp, V. Delatour, G. Nilsson, F. MacKenzie, M. Panteghini, T. Keller, J.E. Camara, I. Zegers, H.W. Vesper, IFCC working group recommendations for assessing commutability part 1: general experimental design, Clin. Chem. (2018) (in press).
[18] G. Nilsson, J.R. Budd, N. Greenberg, V. Delatour, R. Rej, M. Panteghini, F. Ceriotti, H.G. Schimmel, C. Weykamp, T. Keller, J.E. Camara, C.J. Burns, H. Vesper, F. MacKenzie, W.G. Miller, IFCC working group recommendations for assessing commutability part 2: use of the difference in bias between a reference material and clinical samples, Clin. Chem. (2018) (in press).
[19] L.M. Thienpont, D. Stöckl, B. Friedecký, J. Kratochvíla, M. Budina, Trueness verification in European external quality assessment schemes: time to care about the quality of the samples, Scand. J. Clin. Lab. Ivest. 63 (2003) 195–202.
[20] W.G. Miller, G.R. Jones, G.L. Horowitz, C. Weykamp, Proficiency testing/external

quality assessment: current challenges and future directions, Clin. Chem. 57 (2011) 1670–1680.
[21] CLSI, Preparation and Validation of Commutable Frozen Human Serum Pools as Secondary Reference Materials for Cholesterol Measurement Procedures; Approved Guideline. CLSI Document C37-A, Wayne, PA, Clinical and Laboratory Standards Institute, 1999.
[22] S. Sandberg, C.G. Fraser, A.R. Horvath, R. Jansen, G. Jones, W. Oosterhuis, P.H. Petersen, H. Schimmel, K. Sikaris, M. Panteghini, 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. 53 (2015) 833–835.
[23] M. Panteghini, F. Ceriotti, G. Jones, W. Oosterhuis, M. Plebani, S. Sandberg, 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, Clin. Chem. Lab. Med. vol. 55, 2017, pp. 1849–1856.
[24] F. Ceriotti, P. Fernandez-Calle, G.G. Klee, G. Nordin, S. Sandberg, T. Streichert, J.L. Vives-Corrons, M. Panteghini, Criteria for assigning laboratory measurands to models for analytical performance specifications defined in the 1st EFLM Strategic Conference, Clin. Chem. Lab. Med. 55 (2017) 189–194.
[25] S. Pasqualetti, F. Braga, M. Panteghini, Pre-analytical and analytical aspects affecting clinical reliability of plasma glucose results, Clin. Biochem. 50 (2017) 587–594.
[26] D. James, D. Ames, B. Lopez, R. Still, W. Simpson, P. Twomey, External quality assessment: best practice, J. Clin. Pathol. 67 (2014) 651–655.
[27] G.R. Jones, Analytical performance specifications for EQA schemes – need for harmonisation, Clin. Chem. Lab. Med. 53 (2015) 919–924.
[28] G.R.D. Jones, S. Albarede, D. Kesseler, F. MacKenzie, J. Mammen, M. Pedersen, A. Stavelin, M. Thelen, A. Thomas, P.J. Twomey, E. Ventura, M. Panteghini, EFLM Task Finish Group, Analytical Performance Specifications for EQAS (TFG-APSEQA). Analytical performance specifications for external quality assessment - definitions and descriptions, Clin. Chem. Lab. 55 (2017) 949–955.
[29] L.M. Thienpont, D. Stöckl, J. Kratochvíla, B. Friedecký, M. Budina, Pilot external quality assessment survey for post-market vigilance of in vitro diagnostic medical devices and investigation of trueness of participants' results, Clin. Chem. Lab. Med. 41 (2003) 183–186.
[30] W.G. Miller, G.L. Myers, E.R. Ashwood, A.A. Killeen, E. Wang, G.W. Ehlers, D. Hassemer, S.F. Lo, D. Seccombe, L. Siekmann, L.M. Thienpont, A. Toth, State of the art in trueness and interlaboratory harmonization for 10 analytes in general clinical chemistry, Arch. Pathol. Lab. Med. 132 (2008) 838–846.
[31] H.C. Stepman, U. Tiikkainen, D. Stöckl, H.W. Vesper, S.H. Edwards, H. Laitinen, J. Pelanti, L.M. Thienpont, Measurements for 8 common analytes in native sera identify inadequate standardization among 6 routine laboratory assays, Clin. Chem. 60 (2014) 855–863.
[32] A. Carobene, F. Ceriotti, I. Infusino, E. Frusciante, M. Panteghini, Evaluation of the impact of standardization process on the quality of serum creatinine determination in Italian laboratories, Clin. Clin. Acta. 427 (2014) 100–106.
[33] A. Mosca, R. Paleari, A. Carobene, C. Weykamp, F. Ceriotti, Performance of glycated hemoglobin (HbA1c) methods evaluated with EQAS studies using fresh blood samples: still space for improvements, Clin. Chim. Acta 451 ((2015) 305–309.
[34] A.S. Bargnoux, L. Piéroni, J.P. Cristol, N. Kuster, P. Delanaye, M.C. Carlier, S. Fellahi, A. Boutten, C. Lombard, A. González-Antuña, V. Delatour, E. Cavalier, Société Française de Biologie Clinique (SFBC), Multicenter evaluation of cystatin C measurement after assay standardization, Clin. Chem. 63 (2017) 833–841.
[35] S. Ferraro, F. Braga, M. Panteghini, Laboratory medicine in the new healthcare environment, Clin. Chem. Lab. Med. 54 (2016) 523–533.
[36] C. Weykamp, S. Secchiero, M. Plebani, M. Thelen, C. Cobbaert, A. Thomas, N. Jassam, J.H. Barth, C. Perich, C. Ricós, A.P. Faria, 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. 55 (2017) 203–211.
[37] G.B. Kristensen, P. Rustad, J.P. Berg, K.M. Aakre, 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. 62 (2016) 1255–1263.
[38] I. Infusino, F. Braga, R. Mozzi, C. Valente, M. Panteghini, Is the accuracy of serum albumin measurements suitable for clinical application of the test? Clin. Chim. Acta 412 (2011) 791–792.
[39] S.K. Van Houcke, P. Rustad, H.C. Stepman, G.B. Kristensen, D. Stöckl, T.H. Røraas, S. Sandberg, L.M. Thienpont, Calcium, magnesium, albumin, and total protein measurement in serum as assessed with 20 fresh-frozen single-donation sera, Clin. Chem. 58 (2012) 1597–1599.
[40] L.M. Bachmann, M. Yu, J.C. Boyd, D.E. Bruns, W.G. Miller, State of harmonization of 24 serum albumin measurement procedures and implications for medical decisions, Clin. Chem. 63 (2017) 770–779.
[41] F. Braga, M. Panteghini, Standardization and analytical goals for glycated hemoglobin measurement, Clin. Chem. Lab. Med. 51 (2013) 1719–1726.
[42] D. Szőke, A. Carnevale, S. Pasqualetti, F. Braga, R. Paleari, M. Panteghini, More on the accuracy of the Architect enzymatic assay for hemoglobin A1c and its traceability to the IFCC reference system, Clin. Chem. Lab. Med. 54 (2016) e71–e73.
[43] S. Ferraro, A. Panzeri, M. Panteghini, Tackling serum folate test in European countries within the health technology assessment paradigm: request appropriateness, assays and health outcomes, Clin. Chem. Lab. Med. 55 (2017) 1262–1275.
[44] S. Ferraro, A. Panzeri, S. Borille, D. Szőke, M. Panteghini, Estimation of the reference interval for serum folate measured with assays traceable to the WHO International Standard, Clin. Chem. Lab. Med. 55 (2017) e195–e196.
[45] P.H. Petersen, Is it possible to create a perfect external control system? Scand. J. Clin. Lab. Invest. 58 (1998) 265–268.