Review

Mauro Panteghini*

Redesigning the surveillance of *in vitro* diagnostic medical devices and of medical laboratory performance by quality control in the traceability era

Abstract: IVD manufacturers have total responsibility in terms of the traceability of marketed *in vitro* diagnostic medical devices (IVD-MD). This includes the provision of a quality control (QC) material as a part of the measuring system, suitable for traceability verification and alignment surveillance by end-users in daily practice. This material [to be used for the internal QC (IQC) component I as described in this paper] should have unbiased target values and an acceptability range corresponding to analytical performance specifications (APS) for suitable (expanded) measurement uncertainty (MU) on clinical samples. On the other hand, medical laboratories (by the IQC component II as described in this paper) should improve the IQC process and its judging criteria to establish a direct link between their performance, estimated as MU of provided results, and APS defined according to recommended models to apply corrective actions if the performance is worsening with the risk to jeopardize the clinical validity of test results. The participation to external quality assessment (EQA) programs that meet specific metrological criteria is also central to the evaluation of performance of IVD-MDs and of medical laboratories in terms of harmonization 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 to them with selected reference procedures and to define and apply maximum allowable APS to substantiate the suitability of laboratory measurements in the clinical setting.

Introduction

The application of traceability concepts to the analytical quality control (QC) was the object of an editorial I wrote for this journal in 2010 [1]. In some presentations about the topic, I compared the content of that paper discussing the QC subject to the "The origin of the world" painting by Gustave Courbet. ("L'origine du monde" is a picture painted in oil on canvas by the French artist Gustave Courbet in 1866). Indeed, reading the paper one may find in embryo many of the concepts that have been developed and debated in the following decade, most of them discussed in contributions published in *Clinical Chemistry and Laboratory Medicine* (*CCLM*). The time of writing was at the end of the first decade following the release by the European Union of the Directive on *in vitro* diagnostic medical devices (IVD-MD), which asked IVD manufacturers to ensure traceability of their measuring systems to recognized higher-order references by implementing a reliable transfer of the measurement trueness from the highest level of the metrological hierarchy to commercial calibrators used in medical laboratories [2]. This represented the milestone starting the "traceability era", aiming to improve equivalence of laboratory measurement results through more structured approaches for standardization and introducing a legal background for the use of metrologically sound measuring systems in Laboratory Medicine, highlighting that IVD-MD poor performance may compromise the patient safety. For the first time in the history of our profession, it was clear that, for these aspects, medical laboratories should rely on the IVD manufacturers, which entirely assume the responsibility of implementing traceability of their products to the highest available level. The manufacturers were also asked to estimate the measurement uncertainty (MU) of assay calibrators, when used in conjunction with other components (platform

*Corresponding author: Prof Mauro Panteghini**, Centre for Metrological Traceability in Laboratory Medicine (CIRME), University of Milan, Via GB Grassi 74, 20157 Milano, Italy, E-mail: mauro.panteghini@unimi.it

and reagents) of a given IVD-MD. In turn, medical laboratories were called to verify the consistency and the suitability of declared performance about IVD-MD metrological traceability and MU during routine operations performed in accordance with the manufacturer's instructions. The need to monitor the efficacy of traceability implementation on a continuous basis requires however that both the internal quality control (IQC) and the EQA programs should be redesigned in terms that may appear revolutionary to provide proper information, becoming an additional indispensable pillar sustaining the "temple of laboratory standardization" [3, 4].

Redesigning IQC

As the IQC, as traditionally carried out in medical laboratories, does not provide enough information about metrological traceability of IVD-MD in terms of assay standardization [5], in the last years, a number of proposals on how to rethink IQC in the metrological traceability era have been made [6–13]. Basically, the concept that the two sources of measurement error (the bias against higher-order references and the random MU) have different causes requiring different control approaches is now widely accepted and medical laboratories should therefore establish and maintain separate approaches for estimating and minimizing them [14]. In a 'Perspectives' paper on this journal, we have elaborated in detail a practical proposal suggesting that, to obtain information about traceability and its correct implementation (including the suitability of MU on clinical samples), the IQC used to monitor the analytical performance of IVD-MD in individual laboratories should be properly reorganized into two independent components, one devoted to checking the alignment of the IVD-MD and, indirectly, to verifying the consistency of manufacturer's declared traceability during routine operations, therefore highlighting the possible sources of systematic error of measurements [IQC component I (IQC-I)], and the latter structured for estimating MU due to random effects [IQC component II (IQC-II)] (Table 1) [12]. By recognizing the need of rethinking IQC, our proposal was largely supported by the *CCLM* Editors in a following editorial [15].

Checking the alignment of the IVD-MD by IQC-I

Although the bias of an IVD-MD should be appropriately corrected by IVD manufacturer before placing it on the

Table 1: Main characteristics of the two internal quality control (IQC) components. Adapted from ref. [12].

	IQC component I	IQC component II
Aim	Testing IVD-MD alignment according to manufacturer's specifications	Checking IVD-MD variability (lot-to-lot variations, analytical drifts, etc.)
Materials	Control materials supplied by the IVD-MD's manufacturer with system-specific assigned values and acceptability range	Third-party control materials, commutable, with concentrations at clinical decision limits
Scope	Acceptance/rejection of analytical runs	Provide data for measurement uncertainty calculation from its random sources
Rules	Results within a stated acceptability range	Fulfil allowable performance specifications

market, during the daily activity the system alignment may undergo some changes. IQC-I is properly devoted to monitoring the magnitude and the acceptability of these changes. To this aim, IVD manufacturers are asked to provide end-users with a QC material (hereafter referred to as IQC-I material) suitable for daily surveillance of the IVD-MD performance, when working according to the manufacturer's indications. End-users must strictly observe these indications, as only operating in conformity with them the intended purpose of the marketed IVD-MD can be warranted, including the performance declared in terms of metrological traceability. IVD manufacturers are requested to provide IQC-I materials as a qualified and integral part of the IVD-MD; these materials should be designed for daily monitoring of the IVD-MD alignment, with appropriate target values and acceptability range [16]. For effective surveillance of traceability, IQC-I target values must be indeed assigned (and hopefully certified about their traceability) to permit to end-users to confirm that the IVD-MD performance is properly unbiased (or negligibly biased). Accordingly, the common manufacturers' practice to derive mean values of their QC materials (when offered as part of the IVD-MD) from replicates performed using the same IVD-MD with no trueness check regarding assigned values, should be abandoned.

Estimating the random sources of MU by IQC-II

According to the ISO 15189:2012 standard, medical laboratories should know the MU of their results for assessing whether employed IVD-MD are suitable for clinical use [11,

17]. This requirement originated a great debate on how MU should be estimated in practice (sometimes with some scepticism about the utility of do it), with many contributions published on this journal [11, 14, 17–23]. The Guide to the Expression of Uncertainty of Measurement (GUM) approach [24], establishing a model for evaluating and combining all relevant MU sources of a measurement procedure, was endorsed by reference material suppliers and laboratories performing reference measurement procedures [25, 26]. However, within the medical laboratories the application of GUM was too complicated and encountered many practical objections and a substantial rejection for use in daily practice. As a practical alternative, the so-called "top-down" approach was proposed, based on the estimate of laboratory MU results by using IQC data to derive the random components of MU (u_{Rw}) and commercial calibrator information (u_{cal}), the latter combining all uncertainties introduced by the manufacturer's selected calibration hierarchy for the measurand, beginning with the highest available reference (u_{ref}) down to the assigned value of the calibrator for the commercial IVD-MD ($u_{value\ assignment}$) [22, 27]. In other words, IVD-MD that operates in virtually unbiased conditions (as obtained by the correct implementation of system traceability) produce results on clinical samples that have an associated MU combining two major sources due to: (a) the accumulated MU of the corresponding traceability chain, and (b) the MU due to effects associated with the variability of the IVD-MD over time when employed by the individual laboratory. The ISO 20914:2019 Technical Specification describes the optimal conditions for obtaining u_{Rw} as "intermediate within-laboratory precision" [27]. It should be estimated from consecutive 6-month IQC daily data to also capture systematic sources of MU, such as those caused by different lots of reagents, different calibrations, different environmental conditions, etc. Once u_{Rw} has been correctly estimated, it must be combined with u_{cal}, provided by the IVD-MD manufacturer to obtain MU on clinical samples as follows: $\sqrt{(u_{cal}^2 + u_{Rw}^2)}$ [12, 22, 28].

The characteristics of the IQC-II material to be used for u_{Rw} estimate should be carefully considered (Table 1). First, the material should be different from that used for IQC-I. Second, it should be commutable as results obtained on non-commutable materials may not reflect performances achieved by the same IVD-MD on clinical samples in terms of u_{Rw} [29]. Regarding this point, as the use of native biological samples (commutable by definition) for the evaluation of MU over an extended period is not feasible, the use of adequate commercial QC materials or frozen sample pools is unavoidable [30, 31]. Finally, IQC-II materials should have analyte concentrations close to clinical decision thresholds

or, at least, to reference limits employed in the medical application of the test. This is important because, for most, if not all laboratory tests, MU may vary with analyte concentrations [32, 33]. Figure 1 summarizes the main steps to correctly estimate MU in medical laboratories.

Redesigning EQA

In the last decade, many efforts have been dedicated on clarifying and discussing the specific requirements for the applicability of information provided by EQA in the evaluation of the performance of participating laboratories in terms of traceability of their measurements [1, 4, 25, 26, 34–44]. All experts agreed that EQA programs may have a central role but only if they meet specific requirements (Table 2). EQA programs that meet metrological criteria have indeed unique benefits that add substantial value to the practice of Laboratory Medicine (Table 3) [36, 37, 39, 45–48]. All the involved stakeholders should be oriented on configuring and implementing EQA that is effective in the post-market verification of IVD-MD suitability. In extreme cases, they can provide strong supportive evidence to allow advice about IVD-MDs with demonstrated insufficient quality for their abandonment by users (and consequently by IVD manufacturers). As an example, several of the larger IVD companies that have historically provided only alkaline picrate-based creatinine reagents later produced and introduced enzymatic assays, which are more suitable for clinical use of creatinine measurements [49, 50]. EQAS organizers (and the laboratory professionals too) have to ask themselves what is the ultimate scope of what they are offering. Accordingly, they should retune the design to assess the quality of measurement results defined as suitable for clinical use, independent of the IVD-MD type and the possibility to just fulfil the manufacturer's specifications, and act accordingly.

Unfortunately, the vast majority of current EQA programs are often not adequate to assess traceability of laboratory results: most schemes do not pay sufficient attention to the quality of the samples and use peer-group means (or other indicators of central tendency) for grading performance of participants, so the real benefit of participation in EQA as post-marketing surveillance of laboratory performance in terms of standardization/harmonization remains modest [43]. In evaluating EQA results, the approach using the peer group-based value as reference is believed to mitigate the scenario of heterogeneity of marketed IVD-MDs. However, the definition of "peer group" is heterogeneous by itself, alternatively consisting of: (a) the
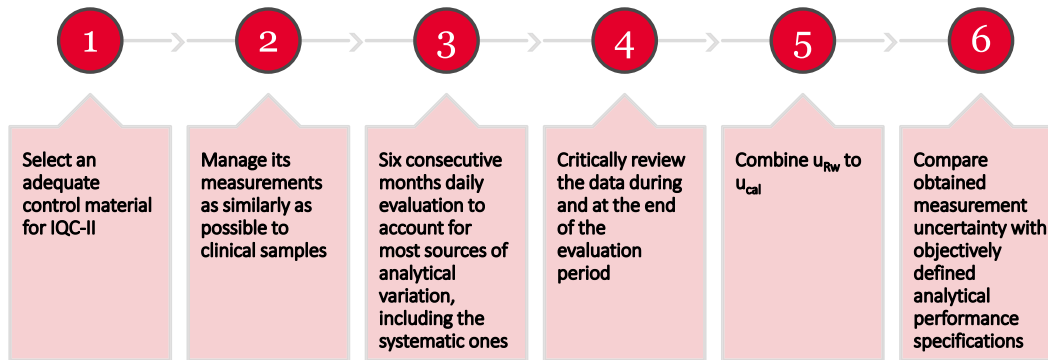
Figure 1: Main steps to be performed to correctly estimate measurement uncertainty (MU) in medical laboratories. IQC-II, internal quality control component II; u_{Rw}, random MU component under conditions of within-laboratory precision (ISO/TS 20914:2019); u_{cal}, calibrator MU, which combines the MU of the higher-order references selected by the IVD manufacturer for implementing traceability with the MU deriving from the process for assignment of calibrator values.

Table 2: Requirements for the applicability of external quality assessment (EQA) results in the evaluation of the performance of participating laboratories in terms of traceability of their measurements.

Feature	Aim
EQA material value-assigned with reference measurement procedures or strictly controlled procedures, if the reference procedure is lacking	To check the traceability of employed IVD-MD to reference measurement systems and the performance of participating laboratories against higher-order references
Proved commutability of EQA materials	To allow transferability of participating laboratory performance to the measurements of patient samples
Use of objectively defined analytical performance specifications	To verify the suitability of laboratory measurements in clinical setting

Table 3: Unique benefits of external quality assessment programs meeting requirements listed in Table 2. Adapted from ref. [36].

– Giving objective information about quality of individual laboratory performance
– Creating evidence about intrinsic standardization status/equivalence of the examined IVD-MDs
– Serving as management tool for the medical laboratory and IVD manufacturers, forcing them to investigate and eventually fix the identified problem
– Helping those manufacturers that produce superior products to demonstrate the superiority of those products in terms of metrological traceability
– Identifying measurands that need improved harmonization and stimulating and sustaining standardization initiatives that are needed to support clinical practice guidelines
– Abandonment by users (and consequently by IVD manufacturers) of non-selective methods and/or of IVD-MDs with demonstrated insufficient quality

same IVD-MD 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 IVD-MDs, each of these definitions exposes to flaws [51, 52]. Importantly, adopting traditional EQA approaches, which disregard commutability of employed control materials and adopt consensus-based values as comparators, we may have an optimistic perception of analytical quality in medical laboratories, creating a situation where they can meet governmental regulations despite consistently reporting biased results [39, 44].

Quality of EQA target

The first requirement is the value assignment of EQA materials with reference measurement procedures when available or strictly controlled procedures if a reference procedure is lacking. This allows objective evaluation of the performance of measurements by participating laboratories against the selected higher-order reference, instead of inferior group-based grading. Note that for measurands for which a reference procedure is not available, IVD-MD-dependent target values should be used to evaluate the performance of participating laboratories, but also in this case the assigned target values to the EQA materials should be determined by reference institutions (possibly including the manufacturer releasing that specific IVD-MD), acting under strictly controlled conditions to avoid biased results and keep their MU as lowest as possible, and not as a peer group mean [1, 39].

Commutability of EQA materials

Commutability has been recently the subject of a specific contribution on this journal and readers should refer to that article for more information [29]. Regarding EQA, commutability of control materials represents another important aspect affecting the applicability of program results in the evaluation of the performance of participating laboratories, as only the use of commutable control samples allows transferability of performance to clinical samples [34, 39, 43, 44, 53, 54]. Many aspects of preparation of EQA materials may potentially affect commutability [44]. Danilenko et al. have described a rigorous protocol to collect blood, obtain serum, prepare a pool, and freeze aliquots under conditions that do not alter the commutability characteristics [55], while Jones et al. have stressed the need that the EQA material matrix and its commutability should be specified by providers, because the interpretation of differences between results in an EQA program is strongly dependent on the nature of the employed material [38]. Sometimes the usefulness of non-commutable EQA material is justified with a scope identified in the demonstration that the employed IVD-MD is performing the way its manufacturer intends it to perform. As laboratory professionals, we are, however, interested in expanding our horizon to know whether the quality of our measurement results is suitable for clinical use, independent of the IVD-MD type and the possibility to fulfil the manufacturer's specifications.

Redesigning analytical performance specifications (APS)

For establishing the suitability of a test measurement, acceptability limits must be used to properly identify laboratories (and, in case, IVD-MDs) that require corrective actions. As we are faced with medical laboratory measurements, simple statistical criteria are not enough [56]. Rather, measurement variability should fall within limits based on medical relevance so that results are reliable for clinical decision-making and patient management [12]. If APS 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 patient outcomes [57]. What degree of quality is needed to guarantee patient safety should therefore be precisely defined and specified for each measurand. The European Federation of Clinical Chemistry and Laboratory Medicine (EFLM) made a landmark contribution by organizing in 2014 a Strategic Conference in which a consensus was reached in defining models for establishing APS [58, 59]. (An entire issue of *CCLM* [2015; vol. 53, issue 6] was dedicated to the publications of all contributions given during the Conference).

Three models were proposed: model 1, based on the effect of analytical performance on clinical outcomes [60, 61]; model 2, based on components of biological variation of the measurand [62]; and model 3, based on state of the art of the measurement (defined as the highest level of analytical performance technically achievable) [63, 64]. An aspect of novelty of this approach was to emphasize that certain models are better suited for some measurands than for others, and the attention should therefore primarily direct toward the measurand and its biological and clinical characteristics [65]. So that, the three models do not necessarily constitute a hierarchy.

The EFLM Conference was a milestone originating several important outcomes [66]. Ceriotti et al. [67] discussed practical principles for allocating measurands to different models, as identifying the most appropriate model to derive APS is an essential step towards the definition of suitable measurement quality. The model 1 ("outcome-based APS") 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 thresholds. The model 2 ("biological variation-based APS") should be applied when the measurand is in a "steady state" status when a subject is in good health. Finally, when a measurand lacks the characteristics to be placed either in model 1 or in model 2, it can be placed in model 3 ("state of the art-based APS"). This model can be temporarily used also for those measurands still waiting for the definition of outcome-based APS or while waiting for robust biological variation data (Figure 2). The myth of state of the art as a 'rescue' model when APS correctly obtained with other models appear too stringent for a certain measurand should be however dismantled. On the other hand, the model 2 should not be used for measurands not having strict homeostatic control. It is inadvisable to use the biological variation-based model to derive APS for any measurand just because the biological variation information is more easily obtainable than outcome-based one. Measurands with defined role in diagnosis of a specific disease should be tested in outcome-based studies and appropriate APS defined. It is a fact, however, that outcome-based information is still frequently not available and performing these types of studies is therefore a fundamental requirement for making recommendations about APS for measurands that should be allocated to this model [33]. To this aim, investigating the impact of analytical performance of the test on clinical (mis)classifications and thereby on the
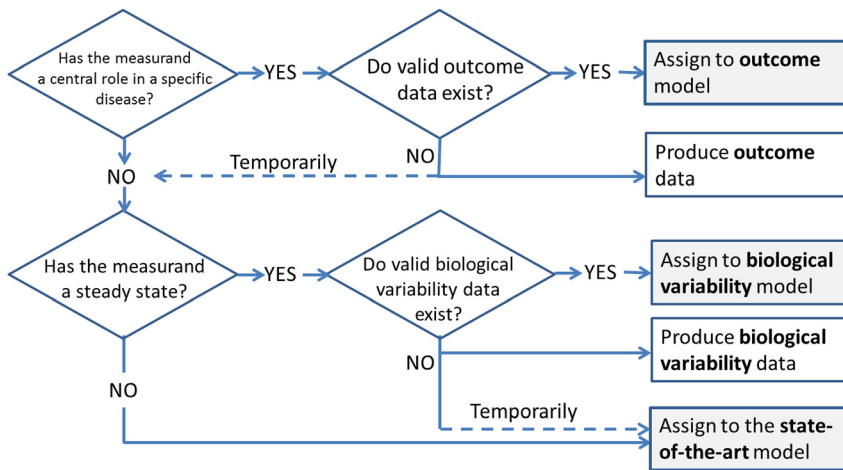
Figure 2: Workflow for assignment of a measurand to an analytical quality specification model as defined by the 2014 EFLM strategic conference. Adapted from Ceriotti et al. [67].

probability of patient outcomes by simulation studies is a practically passable option [68].

Recently, we published the results of the APERTURE, a project for establishing Analytical Performance Specifications for Measurement Uncertainty of 39 common laboratory measurands using these models [32, 33, 69]. Both minimum and desirable quality levels of APS for standard MU of clinical samples were defined by using information obtained from available literature preliminarily checked in terms of robustness. Table 4 summarizes the model allocation together with APS for standard MU on clinical samples for the selected measurands to be used in laboratory practice to validate MU of employed IVD-MDs and to ascertain if estimated MU for a given laboratory result may significantly affect its interpretation [70]. The recently reported case of plasma electrolyte measurements showed the efficacy of MU evaluation by using objectively derived APS in driving laboratories to improve the quality of provided results [71].

Validation criteria for IQC-I

For IQC-I, the acceptability range, which defines the tolerance of value deviation from the unbiased target (obtained by the manufacturer as the mean value of replicate measurements of control material on the same IVD-MD optimally calibrated to the selected reference measurement system), should permit the suitable application of test results in clinical conditions. As previously discussed [12], evaluating how much the possible IVD-MD misalignment influences u_{Rw}, used for the calculation of MU of patient results, should be the ultimate criterium for result validation. Indeed, sudden changes in the alignment of the IVD-MD, due to poor calibration or to change in reagent or calibrator lots, causing shifts in QC results may be responsible of unacceptable u_{Rw} estimated by the IQC-II. Therefore, the acceptability range

for IQC-I should correspond to APS for (expanded) MU discussed in the previous paragraphs [7]. The APS should be directly indicated in the IQC chart, then the relationship between the performance of the IVD-MD in terms of measurement alignment and the desired quality is immediately perceivable.

Acceptability criteria for IQC-II

Once obtained, IQC-II data should be first critically reviewed and proper decisions related to their management taken before moving on to the u_{Rw} estimate [12]. Then, MU on clinical samples should be calculated and, to ascertain if estimated MU for a given laboratory test may affect its interpretation, APS for MU derived as described above should be employed to apply prompt corrective actions if the IVD-MD performance is worsening (Figure 1).

Acceptability of EQA results

In an article published in the *CCLM* special issue dedicated to the contributions given during the EFLM Conference mentioned above, Jones highlighted the wide variation in the definition of APS used by EQA programs, calling for a harmonization through collaborative efforts [72]. It was recommended that APS models from the Conference be used but, just as important, the EQA program should state which aspect of the analytical quality is being evaluated. If EQA programs meet requirements described in Table 2, permitting the evaluation of the performance of participating laboratories in terms of traceability of their measurements, the deviation of a laboratory measurement from the value assigned to the EQA material by the higher-order procedure should stay within the allowable MU limits for that

Table 4: Model allocation according to the EFLM strategic conference consensus and analytical performance specifications (APS) for standard measurement uncertainty (MU) on clinical samples for the measurands evaluated in the APERTURE project. Adapted from refs. [32, 33, 69].

Measurand	APS for standard MU, %	
	Desirable	Minimum
Outcome-based model		
Plasma glucose	2.00	3.00
Blood HbA$_{1c}$	3.00	3.70
Blood total hemoglobin	2.80	4.20
Serum total cholesterol	3.00	7.00
Urine albumin	9.00	17.0
Serum 25-hydroxyvitamin D$_3$	10.0	15.0
Temporarily belonging to biological variation model[a]		
Serum albumin	1.25	1.88
Serum HDL cholesterol	2.84	4.26
Serum triglycerides	9.90	14.9
Blood platelets	4.85	7.28
Biological variation model		
Serum sodium	0.27	0.40
Serum potassium	1.96	2.94
Serum chloride	0.49	0.74
Serum total calcium	0.91	1.36
Serum creatinine	2.20	3.30
Serum urea	7.05	10.6
Serum total bilirubin	10.5	15.7
Serum alanine aminotransferase	4.65	6.98
Serum alkaline phosphatase	2.65	3.98
Serum aspartate aminotransferase	4.75	7.13
Serum creatine kinase	7.25	10.9
Serum γ-glutamyltransferase	4.45	6.68
Serum lactate dehydrogenase	2.60	3.90
Serum pancreatic amylase	3.15	4.73
Serum total proteins	1.30	1.95
Serum immunoglobulin G	2.20	3.30
Serum immunoglobulin A	2.50	3.75
Serum immunoglobulin M	2.95	4.43
Serum prostate-specific antigen	3.40	5.10
Serum magnesium	1.44	2.16
Serum urate	4.16	6.24
Plasma homocysteine	3.52	5.27
Red blood cells	1.55	2.33
White blood cells	5.65	8.48
Serum free triiodothyronine	2.35	3.53
Serum free thyroxine	2.80	4.20
State-of-the-art model		
Serum C-reactive protein	3.76	5.64
Serum thyroid stimulating hormone	2.89	4.34
Model 1 & 2[b]		
Serum digoxin	6.00	9.00

[a]Indicates measurands temporarily allocated to the biological variation model because outcome-based data are lacking. [b]A hybrid model specifically developed for drugs was proposed (see ref. [33] for more details).

measurand, which specify (in numerical terms) the analytical quality required to deliver laboratory test information that would satisfy clinical needs [1]. This will permit EQA participants to understand the effect that the quality of laboratory data has on the way they are used in patient care, including the traceability of the calibration and the test result equivalence among laboratories (i.e., result standardization) [73].

If MU APS are not fulfilled, two main causes should be considered and possibly investigated depending from the behaviour of EQA results provided by the participating laboratory. If EQA results by an individual laboratory in following exercises are randomly distributed both above and below outside the limits of allowable APS for expanded MU, it will be necessary to identify which of the three MU contributions (u_{ref}, u_{cal}, u_{Rw}) is too high for that measurement and working to improve it. During the mentioned EFLM Conference, we proposed a rationale for the definition of recommended limits for combined MU across the entire metrological traceability chain [6]. Focusing first on the APS for combined MU associated with patient results, we recommended that specific MU limits at different levels of the traceability chain should be defined as APS fractions [70, 74, 75]. Criteria for IVD manufacturers that can be achieved for their calibrators should be defined to leave enough MU budget for the individual laboratories to produce clinically acceptable results on clinical samples [74, 76]. If allowable MU limits are exceeded in EQA, the participating laboratory should first verify all analytical conditions that may affect its u_{Rw} (e.g., change of reagent lots, calibrators, instrument maintenance, etc.), checking IQC data (components I and II) in the period in which EQA exercises were carried out. If all these aspects are well under control, it would be necessary to review the manufacturers' protocol to assign value and corresponding MU to the calibrators or the MU associated with the reference measurement system selected by the IVD manufacturer for implementing traceability. Indeed, the selection of different types of calibration hierarchies for the same measurand may lead to different u_{cal}, sometimes making it more difficult to achieve the APS for MU [4, 6].

If EQA results in following exercises are conversely all above or below the allowable MU limits, appearance of a medically unacceptable measurement bias can be suspected [22]. In this case, the bias against a reference (material or procedure) for that measurand should be estimated by an ad-hoc experiment and the presence of a significant systematic error confirmed [77, 78]. (Note that as reference may act any material or procedure positioned at the top of the corresponding traceability chain, even in the absence of high-order options). Then, the bias value should be included in the estimate of MU of clinical samples [79]. If the recalculated MU is not fulfilling the predefined APS, it is the responsibility of the manufacturer to investigate and

eventually fix the problem with a corrective action (e.g., by improving the calibrator value-assignment protocol). Alternatively, the participating laboratory could introduce a correction factor for the detected bias. However, the use of bias correction factors by individual laboratories may alter the IVD-MD status, depriving the measuring system (and, consequently, the produced results) of the certification originally provided by the manufacturer [22].

In conclusion, through the QC programs, redesigned as summarized in this paper, we can expect to obtain an enhanced post-marketing evaluation of IVD-MDs and of individual laboratory performance in terms of quality and clinical validity of measurements. The hope is that all the involved stakeholders agree that "the times they are a-changing" and come to the new post-marketing surveillance road.

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. Dati F. The new European directive on *in vitro* diagnostics. Clin Chem Lab Med 2003;41:1289–98.
3. Panteghini M. Implementation of standardization in clinical practice: not always an easy task. Clin Chem Lab Med 2012;50:1237–41.
4. Braga F, Panteghini M. Verification of in vitro medical diagnostics (IVD) metrological traceability: responsibilities and strategies. Clin Chim Acta 2014;432:55–61.
5. Jassam N, Yundt-Pacheco J, Jansen R, Thomas A, Barth JH. Can current analytical quality performance of UK clinical laboratories support evidence-based guidelines for diabetes and ischaemic heart disease? – A pilot study and a proposal. Clin Chem Lab Med 2013;51:1579–84.
6. 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.
7. Ceriotti F, Brugnoni D, Mattioli S. How to define a significant deviation from the expected internal quality control result. Clin Chem Lab Med 2015;53:913–8.
8. Topic E, Nikolac N, Panteghini M, Theodorsson E, Salvagno GL, Miler M, et al. How to assess the quality of your analytical method? Clin Chem Lab Med 2015;53:1707–18.
9. Ceriotti F. Deriving proper measurement uncertainty from internal quality control data: an impossible mission? Clin Biochem 2018;57:37–40.
10. Badrick T, Bietenbeck A, Cervinski MA, Katayev A, van Rossum HH, Loh TP, et al. Patient-based real-time quality control: review and recommendations. Clin Chem 2019;65:962–71.
11. Thelen M, Vanstapel F, Brguljan PM, Gouget B, Boursier G, Barrett E, et al. European Federation of Clinical Chemistry and Laboratory Medicine (EFLM) Working Group Accreditation and ISO/CEN standards (WG-A/ISO). Documenting metrological traceability as intended by ISO 15189:2012: a consensus statement about the practice of the implementation and auditing of this norm element. Clin Chem Lab Med 2019;57:459–64.
12. Braga F, Pasqualetti S, Aloisio E, Panteghini M. The internal quality control in the traceability era. Clin Chem Lab Med 2021;59:291–300.
13. Zhou R, Wang W, Padoan A, Wang Z, Feng X, Han Z, et al. Traceable machine learning real-time quality control based on patient data. Clin Chem Lab Med 2022;60:1998–2004.
14. 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 2018;56:209–19.
15. Plebani M, Gillery P, Greaves RF, Lackner KJ, Lippi G, Melichar B, et al. Rethinking internal quality control: the time is now. Clin Chem Lab Med 2022;60:1316–7.
16. Aloisio E, Pasqualetti S, Dolci A, Panteghini M. Daily monitoring of a control material with a concentration near the limit of detection improves the measurement accuracy of highly sensitive troponin assays. Clin Chem Lab Med 2020;58:e29–31.
17. Adams O, Cooper G, Fraser C, Hubmann M, Jones G, Plebani M, et al. Collective opinion paper on findings of the 2011 convocation of experts on laboratory quality. Clin Chem Lab Med 2012;50:1547–58.
18. Westgard JO. Managing quality vs. measuring uncertainty in the medical laboratory. Clin Chem Lab Med 2010;48:31–40.
19. Linko S, Ornemark U, Kessel R, Taylor PD. Evaluation of uncertainty of measurement in routine clinical chemistry-applications to determination of the substance concentration of calcium and glucose in serum. Clin Chem Lab Med 2002;40:391–8.
20. Kallner A. Estimation of uncertainty in measurements in the clinical laboratory. Clin Chem Lab Med 2013;51:2249–51.
21. Zhou R, Qin Y, Yin H, Yang Y, Wang Q. Measurement uncertainty of γ-glutamyltransferase (GGT) in human serum by four approaches using different quality assessment data. Clin Chem Lab Med 2018;56:242–8.
22. Braga F, Panteghini M. The utility of measurement uncertainty in medical laboratories. Clin Chem Lab Med 2020;58:1407–13.
23. Plebani M, Padoan A, Sciacovelli L. Measurement uncertainty: light in the shadows. Clin Chem Lab Med 2020;58:1381–3.
24. JCGM 100:2008. Evaluation of measurement data — guide to the expression of uncertainty in measurement, 1st ed.; 2008. Available from: https://www.bipm.org.
25. Infusino I, Schumann G, Ceriotti F, Panteghini M. Standardization in clinical enzymology: a challenge for the theory of metrological traceability. Clin Chem Lab Med 2010;48:301–7.
26. Infusino I, Frusciante E, Braga F, Panteghini M. Progress and impact of enzyme measurement standardization. Clin Chem Lab Med 2017;55:334–40.
27. ISO/TS 20914:2019. Medical laboratories – practical guidance for the estimation of measurement uncertainty, 1st ed. Geneva, Switzerland: ISO; 2019.
28. Panteghini M. The simple reproducibility of a measurement result does not equal its overall measurement uncertainty. Clin Chem Lab Med 2022;60:e221–2.

29. Braga F, Panteghini M. Commutability of reference and control materials: an essential factor for assuring the quality of measurements in laboratory medicine. Clin Chem Lab Med 2019;57:967–73.

30. Birindelli S, Aloisio E, Carnevale A, Brando B, Dolci A, Panteghini M. Evaluation of long-term imprecision of automated complete blood cell count on the sysmex XN-9000 system. Clin Chem Lab Med 2017;55: e219–22.

31. Krintus M, Panteghini M. Laboratory-related issues in the measurement of cardiac troponins with highly sensitive assays. Clin Chem Lab Med 2020;58:1773–83.

32. Braga F, Panteghini M. Performance specifications for measurement uncertainty of common biochemical measurands according to Milan models. Clin Chem Lab Med 2021;59:1362–8.

33. Braga F, Pasqualetti S, Borrillo F, Capoferri A, Chibireva M, Rovegno L, et al. Definition and application of performance specifications for measurement uncertainty of 23 common laboratory tests: linking theory to daily practice. Clin Chem Lab Med 2023;61:213–23.

34. Miller WG, Jones GR, Horowitz GL, Weykamp C. Proficiency testing/external quality assessment: current challenges and future directions. Clin Chem 2011;57:1670–80.

35. Braga F, Panteghini M. Standardization and analytical goals for glycated hemoglobin measurement. Clin Chem Lab Med 2013;51: 1719–26.

36. Ferraro S, Braga F, Panteghini M. Laboratory medicine in the new healhcare environment. Clin Chem Lab Med 2016;54:523–33.

37. 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.

38. Jones GRD, 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.

39. Braga F, Pasqualetti S, Panteghini M. The role of external quality assessment in the verification of in vitro medical diagnostics in the traceability era. Clin Biochem 2018;57:23–8.

40. Ceriotti F, Cobbaert C. Harmonization of external quality assessment schemes and their role – clinical chemistry and beyond. Clin Chem Lab Med 2018;56:1587–90.

41. Jansen RTP, Cobbaert CM, Weykamp C, Thelen M. The quest for equivalence of test results: the pilgrimage of the Dutch calibration 2.000 program for metrological traceability. Clin Chem Lab Med 2018; 56:1673–84.

42. Badrick T, Stavelin A. Harmonising EQA schemes the next frontier: challenging the status quo. Clin Chem Lab Med 2020;58:1795–7.

43. Badrick T, Miller WG, Panteghini M, Delatour V, Berghall H, MacKenzie F, et al. Interpreting EQA – understanding why commutability of materials matters. Clin Chem 2022;68: 494–500.

44. Jones GRD, Delatour V, Badrick T. Metrological traceability and clinical traceability of laboratory results – the role of commutability in external quality assurance. Clin Chem Lab Med 2022;60:669–74.

45. Thienpont LM, Stöckl D, Kratochvíla J, Friedecký B, Budina M. 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 2003;41:183–6.

46. Braga F, Frusciante E, Infusino I, Aloisio E, Guerra E, Ceriotti F, et al. Evaluation of the trueness of serum alkaline phosphatase

47. Wang J, Wang Y, Zhang T, Zeng J, Zhao H, Guo Q, et al. Evaluation of serum alkaline phosphatase measurement through the 4-year trueness verification program in China. Clin Chem Lab Med 2018;56: 2072–8.

48. Yan Y, Pu Y, Zeng J, Zhang T, Zhou W, Zhang J, et al. Evaluation of serum electrolytes measurement through the 6-year trueness verification program in China. Clin Chem Lab Med 2020;59:107–16.

49. Panteghini M. Enzymatic assays for creatinine: time for action. Clin Chem Lab Med 2008;46:567–72.

50. Klee GG, Schryver PG, Saenger AK, Larson TS. Effects of analytic variations in creatinine measurements on the classification of renal disease using estimated glomerular filtration rate (eGFR). Clin Chem Lab Med 2007;45:737–41.

51. Aloisio E, Frusciante E, Pasqualetti S, Quercioli M, Panteghini M. Traceability of alkaline phosphatase measurement may also vary considerably using the same analytical system: the case of Abbott architect. Clin Chem Lab Med 2018;56:e135–7.

52. Pasqualetti S, Carnevale A, Aloisio E, Dolci A, Panteghini M. Different calibrator options may strongly influence the trueness of serum transferrin measured by Abbott architect systems. Clin Chim Acta 2018; 477:119–20.

53. Baadenhuijsen H, Kuypers A, Weykamp K, Cobbaert C, Jansen R. External quality assessment in The Netherlands: time to introduce commutable survey specimens. Lessons from the Dutch "calibration 2000" project. Clin Chem Lab Med 2005;43:304–7.

54. Delatour V, Clouet-Foraison N, Jaisson S, Kaiser P, Gillery P. Trueness assessment of HbA1c routine assays: are processed EQA materials up to the job? Clin Chem Lab Med 2019;57:1623–31.

55. Danilenko U, Vesper HW, Myers GL, Clapshaw PA, Camara JE, Miller WG. An updated protocol based on CLSI document C37 for preparation of off-the-clot serum from individual units for use alone or to prepare commutable pooled serum reference materials. Clin Chem Lab Med 2020;58:368–74.

56. Panteghini M. Reply to Westgard et al.: 'keep your eyes wide … as the present now will later be past'. Clin Chem Lab Med 2022;60:e202–3.

57. Thienpont LM, Van Uytfanghe K, Cabaleiro DR. Metrological traceability of calibration in the estimation and use of common medical decision-making criteria. Clin Chem Lab Med 2004;42:842–50.

58. Panteghini M, Sandberg S. Defining analytical performance specifications 15 years after the Stockholm conference. Clin Chem Lab Med 2015;53:829–32.

59. 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.

60. 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.

61. Petersen PH. Performance criteria based on true and false classification and clinical outcomes. Influence of analytical performance on diagnostic outcome using a single clinical component. Clin Chem Lab Med 2015;53:849–55.

62. Ricós C, Álvarez V, Perich P, Fernández-Calle P, Minchinela J, Cava F, et al. Rationale for using data on biological variation. Clin Chem Lab Med 2015;53:863–70.

measurement in a group of Italian laboratories. Clin Chem Lab Med 2017;55:e47–50.

63. Haeckel R, Wosniok W, Streichert T. Optimizing the use of the 'state-of-the-art' performance criteria. Clin Chem Lab Med 2015;53:887–91.

64. Braga F, Panteghini M. Derivation of performance specifications for uncertainty of serum C-reactive protein measurement according to the Milan model 3 (state of the art). Clin Chem Lab Med 2020;58: e263–5.

65. Panteghini M, Sandberg S. Total error vs. measurement uncertainty: the match continues. Clin Chem Lab Med 2016;54:195–6.

66. Panteghini M, Ceriotti F, Jones G, Oosterhuis W, Plebani M, Sandberg S, 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 2017;55:1849–56.

67. 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.

68. Ferraro S, Lyon AW, Braga F, Panteghini M. Definition of analytical quality specifications for serum total folate measurements using a simulation outcome-based model. Clin Chem Lab Med 2020;58:e66–8.

69. Borrillo F, Pasqualetti S, Panteghini M. Measurement uncertainty of thyroid function tests on a chemiluminescent microparticle immunoassay system needs to be improved. J Appl Lab Med 2022, in press.

70. Panteghini M, Braga F. Implementation of metrological traceability in laboratory medicine: where we are and what is missing. Clin Chem Lab Med 2020;58:1200–4.

71. Pasqualetti S, Chibireva M, Borrillo F, Braga F, Panteghini M. Improving measurement uncertainty of plasma electrolytes: a complex but not impossible task. Clin Chem Lab Med 2021;59:e129–32.

72. Jones GR. Analytical performance specifications for EQA schemes – need for harmonisation. Clin Chem Lab Med 2015;53:919–24.

73. Badrick T, Jones G, Miller WG, Panteghini M, Quintenz A, Sandberg S, et al. Differences between educational and regulatory external quality assurance/proficiency testing schemes. Clin Chem 2022;68:1238–44.

74. Braga F, Panteghini M. Defining permissible limits for the combined uncertainty budget in the implementation of metrological traceability. Clin Biochem 2018;57:7–11.

75. Panteghini M, Braga F, Camara JE, Delatour V, Van Uytfanghe K, Vesper HW, et al. JCTLM Task Force on Reference Measurement System Implementation. Optimizing available tools for achieving result standardization: value added by joint committee on traceability in laboratory medicine (JCTLM). Clin Chem 2021;67:1590–605.

76. Bais R, Armbruster D, Jansen RTP, Klkee 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.

77. Infusino I, Braga F, Mozzi F, Valente C, Panteghini M. Is the accuracy of serum albumin measurements suitable for clinical application of the test? Clin Chim Acta 2011;412:791–2.

78. 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.

79. Bianchi G, Colombo G, Pasqualetti S, Panteghini M. Alignment of the new generation of Abbott alinity γ-glutamyltransferase assay to the IFCC reference measurement system should be improved. Clin Chem Lab Med 2022;60:e228–31.