

Perspectives

Federica Braga*, Sara Pasqualetti, Elena Aloisio and Mauro Panteghini

The internal quality control in the traceability era

<https://doi.org/10.1515/cclm-2020-0371>

Received March 23, 2020; accepted March 24, 2020

Abstract: To be accurate and equivalent, laboratory results should be traceable to higher-order references. Furthermore, their quality should fulfill acceptable measurement uncertainty (MU) as defined to fit the intended clinical use. With this aim, *in vitro* diagnostics (IVD) manufacturers should define a calibration hierarchy to assign traceable values to their system calibrators. Medical laboratories should know and verify how manufacturers have implemented the traceability of their calibrators and estimate the corresponding MU on clinical samples. Accordingly, the internal quality control (IQC) program should be redesigned to permit IVD traceability surveillance through the verification by medical laboratories that control materials, provided by the manufacturer as a part of measuring systems, are in the clinically suitable validation range (IQC component I). Separately, laboratories should also monitor the reliability of employed IVD measuring systems through the IQC component II, devoted to estimation of MU due to random effects and to obtaining MU of provided results, in order to apply prompt corrective actions if the performance is worsening when compared to appropriate analytical specifications, thus jeopardizing the clinical validity of test results.

Keywords: internal quality control; measurement uncertainty; standardization.

“Everything should be made as simple as possible, but not simpler.”

Albert Einstein

***Corresponding author: Federica Braga**, Research Centre for Metrological Traceability in Laboratory Medicine (CIRME), University of Milan, Milan, Italy, E-mail: federica.braga@unimi.it

Sara Pasqualetti, Elena Aloisio and Mauro Panteghini: Research Centre for Metrological Traceability in Laboratory Medicine (CIRME), University of Milan, Milan, Italy

Introduction

An appropriately organized analytical (internal and external) quality control (QC), redesigned to meet metrological concepts, has been recognized as one of the pillars sustaining the “temple of laboratory standardization”, together with the classical elements of the reference measurement system (higher-order reference materials, reference measurement procedures and reference laboratories performing them), the availability of traceable reference intervals/decision limits and the setting of targets for measurement uncertainty (MU) that fit for purpose [1]. The need to apply metrological traceability concepts to the analytical QC for surveying assay standardization was pointed out, for the first time, in an editorial published in this journal 10 years ago [2]. Since then, a number of proposals about how to rethink QC in the metrological traceability era have been provided, and the concept that the two sources of measurement error (the bias against higher-order references and the random error) have different causes requiring different control approaches is now widely accepted [1, 3–13]. According to Oosterhuis et al. [14], “medical laboratories should therefore establish and maintain routines for estimating and minimizing them *separately*”.

In implementing and verifying traceability of patient results, the main responsibilities of medical laboratory professionals in their daily activity are: (1) the verification of the availability and quality of information about metrological traceability of *in vitro* diagnostics (IVD) and their MU, (2) the surveillance of IVD device traceability and (3) the estimate of MU due to the random effects, useful for the calculation of MU of patient results. We can argue if the internal QC (IQC), as traditionally carried out in medical laboratories, provides enough information to fulfill these goals and if the information is reliable in terms of assay standardization. What appears clear from the published experiences is that sometimes we probably have an optimistic perception of analytical quality in medical laboratories, due to the conventional QC approaches for evaluating their performances [10]. Jassam et al. [15] pointed out that the analytical performance of results provided by medical laboratories frequently does not meet the quality for supporting the application of

evidence-based clinical guidelines. They identified two possible reasons for this poor situation: an insufficient robustness of currently used technologies and/or (and in our opinion, more likely) a sub-optimal control over the IQC process. On the other hand, to establish whether the quality of clinical measurements fits the purpose, it is necessary to define what degree of quality is needed and what measurement error can be tolerated without jeopardizing patient safety. This should be done according to the consensus established at the 2014 European Federation of Clinical Chemistry and Laboratory Medicine (EFLM) Strategic Conference, by applying the recommended models and grading different quality levels for analytical specifications according to the current performance of measuring systems [16, 17].

The Clinical and Laboratory Standards Institute (CLSI) describes the IQC principles in a general guideline discussing how to plan and design related procedures as an everyday practice composed by a set of activities based on QC material testing for the confirmation of process stability and reliability of analytical measurements [18]. However, there is no mention on how IQC should be designed to contribute in evaluating and monitoring metrological traceability of IVD measuring systems. We previously proposed that, to fulfill metrological expectations, the IQC used to monitor the analytical performance of IVD measuring systems should be more properly reorganized into two independent components, one devoted to checking the alignment of the measuring system and, indirectly, to verifying the consistency of manufacturer's declared traceability during routine operations performed in accordance with the manufacturer's instructions (IQC component I), and the latter structured for estimating MU due to random effects (IQC component II) [2, 5, 6, 8]. In our opinion, this permits to balance the complication of metrology theory and the practical simplicity needed for adoption by medical laboratories, to establish a direct link between the performance characteristics of measuring systems and the QC rules derived from the EFLM-proposed

models, and, last but not least, to improve control of the bias component. Table 1 summarizes the main characteristics of the two IQC components that we will further discuss in this paper as a novel approach to the IQC program.

The new background

Implementing and maintaining metrological traceability

Laboratory Medicine effectively contributes to the medical decision-making process when results provided on clinical samples are accurate, i.e. associated to a clinically tolerable measurement error, and equivalent, i.e. no matter where and how they are obtained [19]. Basic is the implementation of the traceability of IVD measuring systems to recognized high-order references in order to express the results obtained by the commercial procedure in terms of the values obtained at the highest available level of the metrological hierarchy [20]. The European legislation (the European Union [EU] Directive and, in the near future, the new EU Regulation 2017/746 on IVD medical devices) requires IVD manufacturers to accomplish this [21, 22].

In implementing traceability, IVD manufacturers have a key role and unique responsibilities (Table 2). Information regarding reference materials and reference measurement procedures, useful for assigning traceable unbiased values to commercial calibrators, can be obtained from the Joint Committee for Traceability in Laboratory Medicine database [23]. The value assignment to calibrator materials should be managed by correcting any bias against references and, once the bias has been corrected, the uncertainty of this correction should be included in the MU of calibrator value (u_{cal}) [13]. Manufacturers have the duty of estimating such MU that must include all uncertainties pertaining to the selected calibration hierarchy for the specific measurand, beginning with the highest

Table 1: Main characteristics of the two internal quality control (IQC) components discussed in this paper.

	IQC component I	IQC component II
Aim	Testing IVD measuring system alignment according to manufacturer's specifications	Checking measuring system variability (lot-to-lot variations, analytical drifts, etc.)
Materials	Control materials supplied by the system'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
Rules	Results within a stated acceptability range	Fulfill allowable performance specifications

Table 2: Key responsibilities of *in vitro* diagnostics (IVD) manufacturers in order to fulfill the European legislative requirements about traceability of their measuring systems to recognized higher-order references.

-
- Identification of appropriate higher-order metrological references
 - Definition of a calibration hierarchy to assign traceable values to measuring system calibrators and bias correction during the trueness transfer process
 - Estimation of combined measurement uncertainty of calibrators
 - Fulfillment of measurement uncertainty specifications for calibrators, which represent a proportion of the uncertainty budget allowed for medical laboratory results
-

available reference down to the assigned value of the calibrator for the commercial IVD medical device. The u_{cal} should represent a portion of the total MU budget allowed for medical laboratories in order to leave enough uncertainty budget usable by individual laboratories to produce clinical results of acceptable quality [6, 24].

IVD manufacturers are also asked to provide end-users with technical documentation, instructions for use and a QC material suitable for post-market surveillance of the measuring system 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 measurement system can be warranted, including the performance declared in terms of traceability [1]. IVD manufacturers are therefore requested to provide QC materials as a qualified part of the measuring system; these materials, representing the IQC component I, should be designed for daily monitoring of the measuring system alignment, with appropriate target values and acceptability range. This relies on the concept that, if the traceability of the measuring system to higher-order references is granted by the manufacturer, QC materials, which are part of the whole system, have to be a suitable surrogate of the employed (and declared) reference to permit routine checking of the correctness of system alignment to such reference.

Estimating measurement uncertainty

In lay terms, MU represents the interval of possible values for a measurand for which an x result was obtained. Recently published documents have clarified how MU should be calculated and why this information matters in medical laboratories [13, 25, 26]. MU associated with laboratory results should combine two uncertainty

components: the u_{cal} and the MU accounting for random sources (u_{Rw}) [13, 26]. As mentioned earlier, the former combines the MU of the higher-order reference selected by the IVD manufacturer for implementing traceability with the MU deriving from the process for assignment of calibrator values, including the bias correction, if any.

u_{Rw} gives information about the stability of the measuring system over time and its variability when employed by an individual laboratory. Major variables that may affect the imprecision of the measuring system and individual laboratory performance in using it are summarized in Table 3. As the measuring system imprecision is an intrinsic characteristic of the system itself, it is under the manufacturer's responsibility and it is not controllable by end-users. IVD manufacturers should be aware, however, that if the system imprecision is intrinsically too high, it does not leave margin for MU to the end-users causing u_{Rw} to exceed the desirable performance specifications [27]. When the individual laboratory estimates u_{Rw} through the IQC component II (see below), the two components (i.e. intrinsic system imprecision and individual laboratory performance) are evaluated together.

Deriving analytical performance specifications according to Milan 2014 consensus

As mentioned before, objective criteria for defining analytical performance specifications (APS) for different measurands were defined during the EFLM Strategic Conference held in Milan in 2014 [28]. Three models based on (a) the effect of analytical performance on clinical outcome, (b) the components of biological variation of the measurand and (c) the state of the art of the measurement (defined as the highest level of analytical performance technically achievable), respectively, were proposed [16]. One of the innovative aspects of this recommendation was to emphasize that certain models are better suited for some measurands than for others, with

Table 3: Variables which can influence the random uncertainty, related to the measuring system itself or to the individual laboratory using it.

Measuring system	Individual laboratory
Reagent lot variability	Environmental conditions
Calibrator lot variability	Different operators
Reagent/calibration stability	Instrument maintenance
Measuring equipment	Material preparation

attention primarily directed toward the biological and clinical characteristics of each measurand [29]. Grading different levels of quality (i.e. minimum, desirable and optimum) for APS was also highlighted as very important because it may stimulate IVD manufacturers to work for improving the quality of assays in case of unacceptable or minimum performance [17].

IQC component I: checking the alignment of the IVD measuring system

Characteristics of IQC component I material

Table 4 lists the characteristics of QC materials that should be provided by the IVD manufacturer as an integral component of the CE (“Communautés Européennes”)-marked measuring system. Firstly, material concentrations should be in line with clinically relevant decision limits of the measurand. Recently, using the highly sensitive assay for measuring cardiac troponin T (hs-TnT) as an example, we showed the relevance of monitoring the performance of the measuring system at all clinically relevant thresholds [30]. In patients admitted to the emergency department with symptoms suspicious for acute coronary syndrome (ACS), the diagnosis of non-ST-segment elevation acute myocardial infarction can be ruled out, with a negative predictive value of >99.5%, on a single hs-TnT measurement if the marker concentration is undetectable [31]. Even relatively small variations in assay performance at concentration near the limit of detection (LOD) may hence affect the classification of patients with suspected ACS. Accurate calibration of the

hs-TnT assay in the low range of concentrations is therefore of the utmost importance and, among the tools that laboratories need to check the performance of hs-TnT, is a QC material with concentration near the assay LOD to monitor baseline drifts following assay calibration [32]. Unfortunately, QC materials offered by the manufacturers together with highly sensitive troponin assays do not usually cover very low marker concentrations, leaving assays vulnerable to potential drifts, associated with reagent and calibrator lot changes, platform-to-platform variability and different schemes of instrument maintenance, as well as instrument decline and malfunctioning, that may pass unnoticed.

Secondly, for effective surveillance of traceability, QC target values must be assigned in order to confirm that the measuring system performance is properly unbiased. Note that for this type of materials commutability is not required because they are designed to monitor the specific measuring system for which they are provided. In practice, however, manufacturers of measuring systems commonly derive mean values of their QC materials just from replicates performed by independent laboratories using the same measuring system, with no explicitly certified quality, and frequently provide them as approximate targets, only for convenience, with no trueness claims regarding assigned values. Our recent experience with the use of the Alinity c measuring system for alanine aminotransferase (ALT) illustrates the importance of the issue and possible practical mishaps. The method (Activated ALT, code no. 08P1820) uses a three-level IQC material for system alignment verification provided by Abbott Diagnostics as part of the CE-marked measuring system (Multichem S Plus Technopath, ref. 08P88). In our daily practice, we noticed a systematic relevant underestimation for our ALT measuring system, declared traceable to the IFCC reference procedure, at level-1 QC, corresponding to average physiological ALT concentrations. Figure 1 displays sequential actions we took to correct the problem, including repetition on new aliquots of the QC material, recalibration of the system and then QC repetition, replacement of reagent tank, change of the reagent lot and, finally, manufacturer technical intervention. As all these actions were ineffective, trueness investigation became mandatory to show if the repeated out-of-control results were due to a true system misalignment from the declared IFCC reference. The investigation showed a very good correlation between the ALT patient results by our procedure and those by the IFCC reference procedure, confirming the perfect standardization of the Alinity ALT measuring system, despite the marked underestimation of level-1 QC value, which was clearly wrongly assigned [33].

Table 4: Characteristics of the quality control (QC) material for checking the alignment of the IVD measuring system (internal QC component I).

-
- Concentration levels in line with clinically relevant threshold(s)
 - Unbiased target value. Note that commutable materials are not needed because these are measuring system dedicated, with the possibility of correction of bias due to matrix-related effect resulting from the interaction of reagents and “matrix-modified” material
 - Acceptability range according to the suitable application of test results in a clinical setting
 - Enough stability to monitor the performance of the measuring system under the influence of components potentially deteriorating it
-

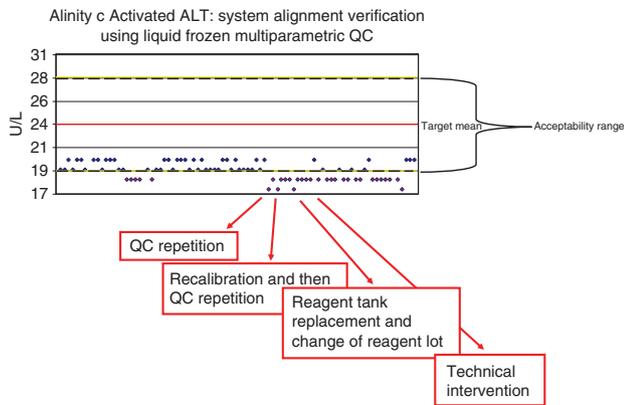


Figure 1: Level-1 quality control (QC) results by alanine aminotransferase (ALT) measuring system in the authors' laboratory, and unsuccessful actions implemented to correct the problem. See text for more information.

In verifying what the QC manufacturer declared in its insert, we found no accuracy claims regarding the assigned values and a sentence stating that ALT mean values were derived from independent laboratories without any trueness check, confirming the high vulnerability of this approach adopted to assign QC values in terms of metrological traceability. Overall, taking care of the metrological quality of the assigned values to offered QC materials is not an issue for manufacturers, with very few exceptions [34].

Thirdly, the acceptability range, which defines the tolerance of value deviation from the target, should permit the suitable application of test results in clinical conditions. Information reported on QC data sheets shows that the acceptability range provided by manufacturers is usually based on the statistical dispersion of data obtained by n laboratories (e.g. ± 2 SD or $\pm 20\%$ of the mean value), with no relationship with clinically suitable APS. Sometimes data from interlaboratory programs are used in the determination of validation range, with the risk to include results from laboratories working under biased analytical conditions, and quite often both mean QC values and ranges are provided only “as guide”, with the recommendation that each laboratory should establish its own acceptability limits. Finally, the same IQC statistical limits are commonly used for all tests carried out on the same measuring system, even if their clinical use and biological behavior are different. This explains why on many analytical platforms it is only possible to set statistical dispersion parameters for IQC program and not predefined measurand-dependent APS.

Lastly, enough stability of the QC material is required to monitor the performance of the measuring system

under all variables influencing or deteriorating it. Sometimes it could be difficult to obtain the optimal stability in all different QC manufactured batches. This may further create difficulties in the interpretation of possible decreasing trends in QC results, by confounding a real drift of the measuring system with the analyte instability in the QC material.

Criteria for interpreting IQC component I and establishing validation limits

In general, IQC programs should detect situations when IVD measuring systems are out of control and results of poor quality are at risk of being produced [18]. This is fully dependent on the criteria applied to the interpretation of IQC results and on the definition of desirable quality. In the last 40 years, a huge amount of literature has been published about recommended interpretative approaches for IQC to be implemented by end-users of IVD medical devices [35–40]. The most popular approach, originally proposed by Westgard et al. [35], is based on a combination of multiple interpretative statistical rules based on different multiples of SD from the QC mean. Although not very recent and limited in the number of surveyed laboratories, some audit has shown that $\sim 60\%$ of medical laboratories used multirules for data analysis, whereas the remainder reported only a single 2-SD rule to define an out-of-control situation [41]. Doubtless, the latter, i.e. checking if the single QC value is (or is not) in the acceptable range (a sort of “macro-evaluation”), represents the first interpretative criterion adopted by medical laboratories to accept/reject analytical runs. QC results outside the acceptability range promote immediate corrective actions to bring the situation under control again and before patient results related to the samples analyzed in the affected run are released. What is clear, however, is that this approach, if used in isolation, is not enough sensitive and precocious enough to prevent poor quality results on clinical samples. The interpretation of QC results by checking their temporal trend without waiting for out-of-control signals is surely more effective. As the effectiveness, in terms of early warning, of this “longitudinal evaluation” may be influenced by the frequency of QC testing, the establishment of this frequency is central. Parvin recommended to adapt the frequency to events, both expected and unexpected, potentially influencing the quality of results [42]. Although the determination of the optimal QC frequency may require the application of sophisticated statistical models, the organization of laboratories

for measuring clinical samples can drive decisions [18]. For clinical samples assayed in batch, QC should be at minimum tested at the beginning and end of each analytical run. In continuous mode, which is now typical for laboratories using total automation, QC materials should be measured periodically along with clinical samples. CLSI C24 called this type of QC schedule “bracketed QC” and suggests considering several potential influencing factors in defining QC frequency [18]. In our core-lab facility, working by a continuous first-come/first-serve processing of samples [43], we perform QC component I three times per day, i.e. every 8 h.

The longitudinal QC evaluation can be done using two main judging criteria: (a) checking for significant trend of QC values and (b) evaluating how much the measuring system misalignment influences u_{Rw} , used for the calculation of MU of patient results. In applying the first criterion, decreasing or increasing trends of IQC results toward the upper or lower limit of the acceptability range are investigated, even if the macroevaluation of individual QC results shows that they are still within the acceptability range. This approach allows laboratories to implement corrective actions earlier, before the measuring system definitively gets worse. The second criterion relates to the assessment of the influence of variability of measuring system alignment on u_{Rw} . Shifts in QC results due to poor calibration or to change in reagent or calibrator lots, which occur during daily activity, may once again go undetected if QC values are judged against the acceptability range. However, the sudden change in the alignment of the measuring system may be responsible of unacceptable u_{Rw} . One may easily note that what discussed earlier for longitudinal QC evaluation is a type of visual interpretation that was objectified by the Westgard’s rules [35]. What is lacking is the link with the new scientific background described in the previous chapter “The new background”. To obtain this, the acceptability range for QC component I should correspond to APS for MU derived according to the appropriate Milan model and it should be set based on unbiased target value of the material obtained by the manufacturer as the mean of replicate measurements on the same measuring system optimally calibrated to the selected reference. Figure 2 displays a case study where, by using the acceptability range provided by the manufacturer, the QC shift, causing an unacceptable u_{Rw} , was not alarming. On the contrary, when the acceptability range was defined according to the appropriate APS for MU, the shift of the measuring system alignment, significantly affecting MU of patient results, was noticed and corrective actions immediately put in place.

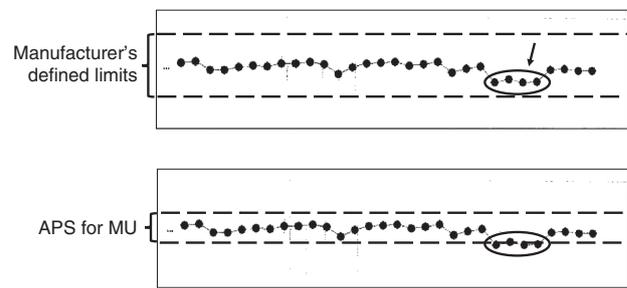


Figure 2: Example of the impact of the different definition of acceptability limits for quality control (QC) component I. Measuring system: Architect c4000 enzymatic assay for hemoglobin A_{1c} (Abbott Diagnostics), as described in ref. [44]. QC material: BioRad Lyphochek Diabetes control material (ref. 740), mean value: 76.3 mmol/mol. The upper panel displays the situation using the defined acceptability range provided by the manufacturer. The QC shift (arrow), causing an excessively elevated measurement uncertainty (MU) on clinical samples, was not identified. However, if the acceptability range was defined according to the appropriate analytical performance specification (APS) for MU ($\pm 3.8\%$, expanded MU for a coverage factor = 2), the shift of the measuring system would have been identified and corrective actions immediately put in place. Dashed lines in the graphs show the upper and lower QC component I acceptability limits.

IQC component II: estimating the random sources of measurement uncertainty

Estimate u_{Rw} using IQC component II

From a statistical point of view, the u_{Rw} represents the dispersion of replicated measurements performed on a QC material, expressed as SD or CV. The ISO 20914:2019 technical specification defines the optimal condition for obtaining u_{Rw} as that estimating “intermediate within-laboratory precision” [26]. Accordingly, QC measurements should be performed over a sufficiently extended period able to include all changes in measuring conditions influencing u_{Rw} reported in Table 3. The recommended time span is 6 months, as it allows to cover those significant sources of u_{Rw} [13, 26]. Table 5 summarizes the steps to be followed to obtain a robust estimate of u_{Rw} . Once u_{Rw} has been estimated, it must be combined with u_{cal} , provided by the manufacturer of measuring system, to obtain MU on clinical samples as follows: $\sqrt{(u_{cal}^2 + u_{Rw}^2)}$ [13, 26].

Characteristics of IQC component II material

The characteristics of the QC material to be used for u_{Rw} estimate have been previously described in detail [5, 6, 13].

Table 5: Steps to obtain the estimate of random uncertainty (u_{Rw}) using the internal quality control (IQC) component II.

1. Provide that the measuring system is running properly and is correctly aligned, through IQC component I data;
2. Run IQC component II material randomly inside the routine analytical run (mimicking analytical conditions of clinical samples);
3. Repeat measurements at least daily for a period (e.g. 6 consecutive months) sufficient to capture most changes in measuring conditions and systematic sources of measurement uncertainty;
4. Do not include gross outliers in the u_{Rw} estimate, but check the measuring system performance and explain the outlier result;
5. At the end of the evaluation period, collect all results and revise the data (exclude explainable outliers, separate data obtained with different lots of control materials, etc.);
6. Calculate mean and SD of replicates;
7. Calculate relative u_{Rw} as $SD/mean \times 100$

The material must be different from that used for the verification of system alignment, i.e. IQC component I material. The latter is used to validate analytical runs and should have different characteristics, described in the previous section of this paper. Importantly, if the component I fails, no other measurements should be performed until the situation is brought back under control. Differently, measurements of component II material for u_{Rw} estimate must be part of an already validated analytical run, as similarly as possible to those of clinical samples. Importantly, the QC component II material should be commutable as results obtained on non-commutable materials may not reflect performances achieved by the same measuring system on clinical samples in terms of random MU [45]. Hage-Sleiman et al. [46] showed the misleading results obtained in the estimate of precision profile when using a non-commutable IQC material, taking as example a highly sensitive troponin assay. As the use of a single biological sample (commutable by definition) for the evaluation of MU over an extended period is not feasible, the use of adequate commercial QC materials or sample pools is unavoidable [47, 48]. Lastly, QC materials for component II should have analyte concentration levels close to clinical decision limits or, at least, to employed reference limits. This is important because, for most, if not all laboratory tests, MU varies with the analyte concentration, usually decreasing with increasing concentrations.

Criteria for interpreting IQC component II and establishing clinical suitability of estimated MU

Once obtained, IQC component II data should be first critically reviewed and proper decisions related to their management taken before moving on to the u_{Rw} estimate. Figure 3 shows three examples of component II data evaluation with respective outcomes.

The ISO 20914 is categorical in stating that “estimating the uncertainty of the results produced is of very limited value unless it can be compared with the allowable MU based on the quality of results required for medical use” [26]. Therefore, in order to ascertain if estimated MU for a given laboratory result may affect its interpretation, clinically allowable MU should be defined. As mentioned before, APS for MU of clinical results should be derived by using one of the three EFLM recommended models [16]. Ceriotti et al. clarified the criteria for assigning different measurands to one of these models [29]. For instance, for cardiac troponin, a marker having a pivotal role in the diagnosis of myocardial infarction, the outcome-based model should be used. Using this model, a standard MU $\leq 10\%$ at the concentration corresponding to the 99th percentile upper reference limit represents the desirable goal [32]. According to the classical Fraser’s paradigm for deriving APS for random variability [49], the quality level can be further modulated to minimum ($10\% + \frac{1}{2} 10\% = 15\%$) and optimal ($\frac{1}{2} 10\% = 5\%$) APS. The biological variation model is used when the measurand has a high homeostatic control. By using published data about the average intra-individual biological variation of measurands with these characteristics and the quality level grading, APS for standard MU can be established [23, 33]. Finally, if the measurand has neither central diagnostic role nor enough homeostatic control, the state-of-the-art model should be used. As an example, we can refer to the human chorionic gonadotropin (hCG) measurement, intended as a test to detect pregnancy. hCG is not used for clinical diagnosis and it is not normally produced in a healthy status. As the definition of “state of the art” requires defining the highest level of achievable analytical performance [16], we compared standard MU of three widely used measuring systems (Roche Modular Evo, Abbott Architect i2000SR and Alinity i), estimated by the approach described in this paper by using a QC component II material, obtaining values of 2.3%, 5.4% and 3.8%, respectively. As a result, the obtained Modular Evo

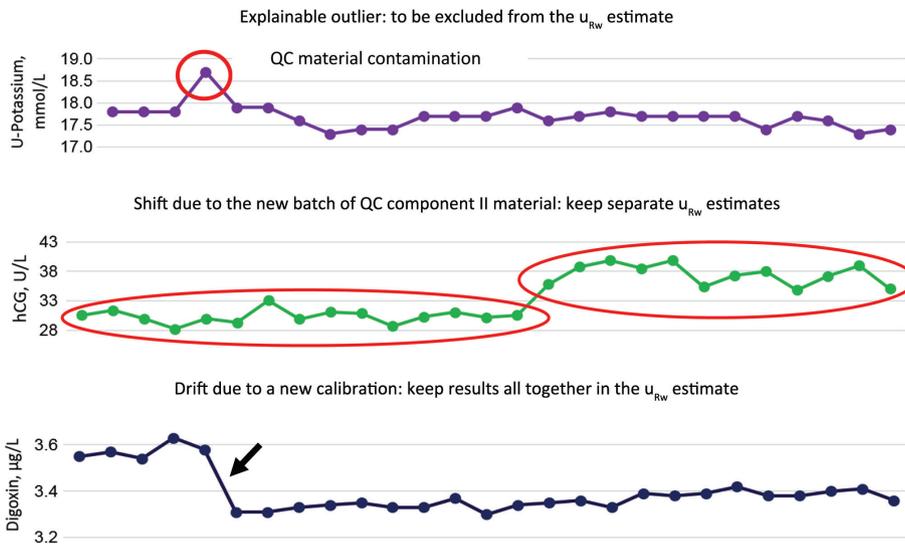


Figure 3: Examples of evaluation of internal quality control (QC) component II data and their related outcomes. u_{Rw} , standard measurement uncertainty due to random error sources.

performance (2.3%) can be considered as representing the state of the art of the measurement to be employed for the definition of APS for hCG MU. If we consider this APS as desirable, we can also modulate the quality level to, e.g. minimum goal ($2.3\% + \frac{1}{2} 2.3\% = 3.5\%$).

Conclusions

According to the European legislation, the CE marking should ensure the availability for medical laboratories of IVD measuring systems with calibrators traceable to higher-order references. IVD manufacturers therefore assume total responsibility in terms of the traceability of marketed measuring systems, including the responsibility to provide a QC material as a part of the system, suitable for traceability verification and alignment surveillance by end-users in daily practice. This material (which corresponds to the IQC component I as described in this paper) should have unbiased target values and an acceptability range corresponding to APS for suitable 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 Milan models, in order to apply prompt corrective actions if the performance is worsening and may jeopardize the clinical validity of test results. Through the IQC program, redefined as proposed in this paper, we expect to obtain an enhanced post-marketing evaluation of laboratory performance in terms of measurement standardization,

complementing a more rigorous premarketing validation by IVD manufacturers.

Acknowledgments: This report summarizes the content of the lectures given during the 13th CIRME international conference, held in November 2019, focused on the topic of IQC and its redefinition in the traceability era.

Author contributions: All authors have accepted responsibility for the entire content of this manuscript and approved its submission.

Research funding: None declared.

Employment or leadership: None declared.

Honorarium: None declared.

Competing interests: Authors state no conflict of interest.

References

1. Braga F, Panteghini M. Verification of in vitro medical diagnostics (IVD) metrological traceability: responsibilities and strategies. *Clin Chim Acta* 2014;432:55–61.
2. 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.
3. Miller WG, Jones GR, Horowitz GL, Weykamp C. Proficiency testing/external quality assessment: current challenges and future directions. *Clin Chem* 2011;57:1670–80.
4. Ceriotti F. The role of External Quality Assessment Schemes in monitoring and improving the standardization process. *Clin Chim Acta* 2014;432:77–81.
5. Braga F, Infusino I, Panteghini M. Role and responsibilities of laboratory medicine specialists in the verification of metrological traceability of in vitro medical diagnostics. *J Med Biochem* 2015;34:282–7.

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, Brugnoli 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. Sciacovelli L, Secchiero S, Padoan A, Plebani M. External quality assessment programs in the context of ISO 15189 accreditation. *Clin Chem Lab Med* 2018;56:1644–54.
10. 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.
11. Ceriotti F. Deriving proper measurement uncertainty from Internal Quality Control data: an impossible mission? *Clin Biochem* 2018;57:37–40.
12. 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.
13. Braga F, Panteghini M. The utility of measurement uncertainty in medical laboratories [published online ahead of print, 2020 Mar 3]. *Clin Chem Lab Med* 2020;58:1407–13.
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. 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.
16. 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.
17. Panteghini M, Ceriotti F, Jones G, Oosterhuis W, Plebani M, Sandberg S, et al. Strategies to define performance specifications in laboratory medicine: 3 years on from the Milan Strategic Conference. *Clin Chem Lab Med* 2017;55:1849–56.
18. CLSI. Statistical quality control for quantitative measurement procedures: principles and definitions. 4th ed. CLSI guideline C24. Wayne, PA: Clinical and Laboratory Standards Institute, 2016.
19. Ferraro S, Braga F, Panteghini M. Laboratory medicine in the new healthcare environment. *Clin Chem Lab Med* 2016;54:523–33.
20. Panteghini M. Traceability as a unique tool to improve standardization in laboratory medicine. *Clin Biochem* 2009;42:236–40.
21. Panteghini M, Forest JC. Standardization in laboratory medicine: new challenges. *Clin Chim Acta* 2005;355:1–12.
22. Regulation (EU) 2017/746 of the European Parliament and of the Council of 5 April 2017 on in vitro diagnostic medical devices and repealing Directive 98/79/EC and Commission Decision 2010/227/EU. *Off J Eur Union* L117, 5th May 2017, 176–332.
23. Panteghini M, Braga F. Implementation of metrological traceability in laboratory medicine: where we are and what is missing [published online ahead of print, 2020 Feb 18]. *Clin Chem Lab Med* 2020;58:1200–4.
24. Braga F, Panteghini M. Defining permissible limits for the combined uncertainty budget in the implementation of metrological traceability. *Clin Biochem* 2018;57:7–11.
25. Infusino I, Panteghini M. Measurement uncertainty: friend or foe? *Clin Biochem* 2018;57:3–6.
26. ISO/TS 20914:2019. Medical laboratories – practical guidance for the estimation of measurement uncertainty, 1st ed. Geneva, Switzerland: ISO, 2019.
27. Aloisio E, Carnevale A, Dolci A, Panteghini M. Novel generations of laboratory instruments should not worsen analytical quality: the case of GEM Premier 5000. *Clin Biochem* 2018;58:128–30.
28. Panteghini M, Sandberg S. Defining analytical performance specifications 15 years after the Stockholm conference. *Clin Chem Lab Med* 2015;53:829–32.
29. Ceriotti F, Fernandez-Calle P, Klee GG, Nordin G, Sandberg S, Streichert T, et al. 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.
30. 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.
31. Body R, Mueller C, Giannitsis E, Christ M, Ordonez-Llanos J, de Filippi CR, et al. The use of very low concentrations of high-sensitivity troponin T to rule out acute myocardial infarction using a single blood test. *Acad Emerg Med* 2016;23:1004–13.
32. 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.
33. Aloisio E, Frusciante E, Pasqualetti S, Infusino I, Krintus M, Sypniewska G, et al. Traceability validation of six enzyme measurements on the Abbott Alinity c analytical system. *Clin Chem Lab Med* 2020;58:1250–6.
34. Multiparametric control. <http://www.diagam.com/upload/documents/FT%20CAL%20CONT/EN%20CONTROL/MULTIPARAMETRICS/MPCONFTEN%20v05-18H09.pdf> (Accessed March 2020).
35. Westgard JO, Barry PL, Hunt MR, Groth T. A multi-rule Shewhart chart for quality control in clinical chemistry. *Clin Chem* 1981;27:493–501.
36. Büttner J, Borth R, Broughton PM, Bowyer RC. IFCC approved recommendation (1983) on quality control in clinical chemistry. Part 4. Internal quality control. *J Clin Chem Clin Biochem* 1983;21:877–84.
37. Petersen PH, Ricós C, Stöckl D, Libeer JC, Baadenhuijsen H, Fraser C, et al. Proposed guidelines for the internal quality control of analytical results in the medical laboratory. *Eur J Clin Chem Clin Biochem* 1996;34:983–99.
38. Westgard JO. Internal quality control: planning and implementation strategies. *Ann Clin Biochem* 2003;40:593–611.
39. Kinns H, Pitkin S, Housley D, Freedman DB. Internal quality control: best practice. *J Clin Pathol* 2013;66:1027–32.
40. Westgard JO, Westgard SA. Quality control review: implementing a scientifically based quality control system. *Ann Clin Biochem* 2016;53:32–50.
41. Housley D, Kearney E, English E, Smith N, Teal T, Mazurkiewicz J, et al. Audit of internal quality control practice and processes in the south-east of England and suggested regional standards. *Ann Clin Biochem* 2008;45:135–9.
42. Parvin CA. Assessing the impact of the frequency of quality control testing on the quality of reported patient results. *Clin Chem* 2008;54:2049–54.

43. Dolci A, Giavarina D, Pasqualetti S, Szőke D, Panteghini M. Total laboratory automation: do stat tests still matter? *Clin Biochem* 2017;50:605–11.
44. Szőke D, Carnevale A, Pasqualetti S, Braga F, Paleari R, Panteghini M. More on the accuracy of the Architect enzymatic assay for hemoglobin A1c and its traceability to the IFCC reference system *Clin Chem Lab Med* 2016;54:e71–3.
45. 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.
46. Hage-Sleiman M, Capdevila L, Bailleul S, Lefevre G. High sensitivity cardiac troponin-I analytical imprecisions evaluated by internal quality control or imprecision profile. *Clin Chem Lab Med* 2019;57:e49–51.
47. 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.
48. Aloisio E, Carnevale A, Pasqualetti S, Birindelli S, Dolci A, Panteghini M. Random uncertainty of photometric determination of hemolysis index on the Abbott Architect c16000 platform. *Clin Biochem* 2018;57:62–4.
49. Fraser CG, Hyltoft Petersen P, Libeer JC, Ricos C. Proposals for setting generally applicable quality goals solely based on biology. *Ann Clin Biochem* 1997;34:8–12.