Opinion Paper

Francesca Borrillo* and Mauro Panteghini

State-of-the-art model for derivation of analytical performance specifications: how to define the highest level of analytical performance technically achievable

Abstract: To be accurate and equivalent among assays, laboratory results should be traceable to higher-order references and their quality should fulfill maximum allowable measurement uncertainty (MU) as defined to fit the intended clinical use. Accordingly, laboratory professionals should estimate and validate MU of performed tests using appropriate analytical performance specifications (APS). Current consensus supports the derivation of APS by using one of the three models established by the European Federation of Clinical Chemistry and Laboratory Medicine (EFLM) Strategic Conference held in Milan in 2014. It is recognized that some models are better suited for certain measurands than for others and the attention should be primarily directed towards their biological and clinical characteristics. Among others, model 3 should reflect the state of the art of the measurements that can be defined as the best analytical performance that is technically achievable. Taking serum C-reactive protein and ferritin as examples, here we describe the theoretical premises and the experimental protocol to be used to derive APS for MU when a measurand is allocated to this model. Although the model lacks a direct relationship with clinical outcomes, useful information about the in vitro diagnostic medical device performance and the average quality of provided results may be obtained.

Keywords: measurement uncertainty; metrological traceability; analytical performance specifications

Conceptual background

Laboratory customers, i.e., clinicians and patients, essentially ask equivalent results from all laboratories that fulfill quality requirements for clinical use. The agreed approach to ensure a satisfactory response to this demand is to implement metrological traceability to a common standard through an unbroken sequence of calibrations, each contributing to the measurement uncertainty (MU) of results [1]. MU was recently the subject of a specific contribution to this journal and readers should refer to this article for further information [2].

Here we would like to remember the role of MU as key performance indicator and pivotal management tool for both medical laboratories and in vitro diagnostic (IVD) manufacturers, because the magnitude of obtained MU should be suitable for a patient result to be used in a medical decision [3]. The ISO/TS 20914:2019 has provided a practical guidance for the MU estimate using a ‘top-down’ approach, also clearly pointing out that “estimating the uncertainty of the results produced is of very limited value unless it can be compared with the allowable MU [meaning, analytical performance specifications (APS)] based on the quality of results required for medical use” [4]. The inspiring concept behind this approach relies on the definition of MU across the entire traceability chain, starting with the MU of reference materials ($u_{ref}$), extending through the IVD manufacturers and their processes for assignment of calibrator values and MU ($u_{cal}$), and ending with the random variability of IVD medical device (IVD-MD) ($u_{Rw}$) (Figure 1) [5]. We note the assumption behind the MU concept that the bias should be appropriately corrected by IVD manufacturers during the trueness transfer along the metrological traceability chain and the MU of the correction ($u_{bias}$) combined with other sources of MU. Theoretically, correct alignment of IVD-MD is therefore expected before it goes to market, although in practice this is sometimes not achieved. During daily use the system alignment may however undergo changes due to systematic sources of MU, such as...
those caused by different lots of reagents [6–8]. In agreement with the recommended ISO approach, this bias is incorporated in the MU of clinical samples \( [u(y)] \) through the \( u_{ref} \) estimate,\(^1\) and can be tolerated provided the \( u(y) \) fulfills the predefined APS. Finally, it is important to note that, although reference providers, IVD manufacturers, and medical laboratories have different roles and independent tasks in the implementation and application of metrological concepts, their performances contribute together to the \( u(y) \) value [9]. Therefore, MU associated with each step of the calibration hierarchy should be governed to obtain a final combined \( u(y) \) on clinical samples that fulfills the established APS [1, 10–12].

### Defining APS for MU

After the 1st European Federation of Clinical Chemistry and Laboratory Medicine (EFLM) Strategic Conference, held in Milan in 2014, objective criteria for defining APS became available. These criteria are founded on three models: model 1, based on the effect of analytical performance on clinical outcomes; model 2, based on components of biological variation of the measurand; and model 3, based on state of the art (SA) of the measurement [13]. The models use different principles and do not constitute a hierarchy; therefore, some models are better suited for certain measurands than for others, and the attention should primarily direct toward the measurand and its biological and clinical characteristics [14]. Criteria have been proposed for selecting an appropriate model to determine APS for different laboratory measurands [15].

#### The state-of-the-art model (Milan model 3)

Following the Milan conference, Ceriotti et al. depicted the workflow for assignment of a measurand to APS models [15]. In particular, the model 3 (SA) should be in principle used when a measurand has neither central diagnostic role nor strict homeostatic control. In addition, the model can be temporarily used also for those measurands still waiting for the definition of outcome-based APS and for which the biological variation-based model should not be used because a strict homeostatic control is lacking. Therefore, we have two categories of measurands that may belong to the SA model: (a) measurands that cannot be included in the first two models, because they have not the appropriate biological and clinical characteristics; and (b) measurands that theoretically should be allocated to other models but do not have available data (in this case the model is used as temporary allocation).

A highly debated issue relates to the definition of SA in relation to the APS establishment. In literature, SA has been defined as: (a) the highest level of analytical performance technically achievable by field methods (considered as the best option in the Milan consensus); (b) the analytical performance achieved by a certain percentage of laboratories, e.g., in an External Quality Assessment (EQA) (considered as an alternate option by Milan consensus); or (c) the average analytical performance declared for that test by the most relevant IVD manufacturers. Other important perspectives, such as cost-effectiveness, convenience, risk management, etc., can be also considered. Although no official agreement on how to set APS based on SA model is available, we believe that the SA definition for deriving APS should be related to an aspirational approach that links this definition to the best quality available. This does not mean that

---

\(^1\) ISO/TS 20914:2019 defines \( u_{ref} \) as “uncertainty component under conditions of within-laboratory precision” (i.e., the uncertainty for a given IVD-MD in the same laboratory over an extended period (e.g., six consecutive months) that includes routine changes to measuring conditions, for example, lot changes of reagents, calibrators, instrument maintenance, etc.).

**Figure 1:** Sources of measurement uncertainty (\( u \)) across the entire metrological traceability chain contributing to the estimate of \( u \) of clinical sample \( [u(y)] \). Note that, once correctly estimated, this value should be compared with the maximum allowable \( u \), i.e., the established analytical performance specification, to show if the magnitude of \( u \) of the in vitro diagnostic medical device (IVD-MD) employed by the medical laboratory is suitable for use in clinical decision making. Adapted from Ref. [2]. \( u_{ref} \), uncertainty of selected higher-order reference; \( u_{bias} \), uncertainty of the bias correction; \( u_{cal} \), uncertainty of IVD-MD calibrator; \( u_{ref} \), assay precision under intermediate reproducibility conditions obtained by an individual laboratory using the IVD-MD [4].

ISO/TS 20914:2019 defines \( u_{ref} \) as “uncertainty component under conditions of within-laboratory precision” (i.e., the uncertainty for a given IVD-MD in the same laboratory over an extended period (e.g., six consecutive months) that includes routine changes to measuring conditions, for example, lot changes of reagents, calibrators, instrument maintenance, etc.).
the top-performing IVD-MD at a certain stage should replace less well-performing devices, but that the former should be used as ‘reference’ in terms of quality provided to place alongside the traditional strategy of commercial competition, the production of IVD-MDs that may improve the average analytical performance [16–18]. Comparative information should become a stimulus to improve the analytical quality of available IVD-MDs by identifying, if any, weaknesses that can be ameliorated. The APS criteria can be based on the model presented in the next paragraph, where the performance of the top-performing IVD-MD is identified as ‘desirable’ and the other methods should at least have the ‘minimum’ quality. Users should pay special attention to underperforming IVD-MDs, which are in use and, only in this case, the involved stakeholders can be requested to improve the IVD-MD performance by correcting anomalies identified during the described comparative approach.

Proposed approach to define SA of MU as the highest level of performance technically achievable using the ISO/TS 20914 guidance for the MU estimate

Taking for reference the SA definition as the highest level of analytical performance technically achievable in medical laboratories using currently available IVD-MDs in common use, which we consider, as explained above, the preferable statement in terms of quality improvement, an approach to define SA for MU has been proposed [19]. Figure 2 displays in detail the process steps. Specifically, it requires to collect the information about the entire calibration hierarchy of the selected IVD-MDs for the evaluated measurand to understand if \( u_{cal} \) values provided by the respective manufacturers have been combined in advance with \( u_{ref} \) if any. In case this has not been done (which corresponds to the great majority of cases), it is necessary to estimate \( u_{cal} \) by combining \( u_{ref} \) of the stated higher-order reference, which should be retrieved from publicly available information, e.g., from the reference material’s certificate of analysis, with the MU of commercial calibrator as provided by the manufacturer. How to derive \( u_{Rw} \) using the Internal Quality Control (IQC) program has been previously described in detail (Figure 3) [6]. Briefly, it should be estimated from six consecutive months IQC daily data to also include systematic sources of MU, using a commutable control material that should be different from that usually employed to check the alignment of IVD-MD and with concentrations corresponding to the decision level used for the clinical application of the test [2, 4, 6–8]. Establishing APS at the ‘minimum quality level’ as being 50 % greater than the desirable one follows the classical paradigm for deriving APS for random variability [10].

Serum C-reactive protein and ferritin as illustrative measurands of APS for MU derivation according to the Milan model 3

For illustrating the derivation of APS for MU according to the SA model, we will use two measurands belonging to the two

![Figure 2](image-url): Proposed steps to estimate the state of the art of measurement uncertainty performance for a given measurand as the highest level of performance technically achievable by field methods. IVD-MDs, in vitro diagnostic medical devices; \( u_{cal} \), uncertainty of IVD-MD calibrator; \( u_{Rw} \), IVD-MD precision under intermediate reproducibility conditions; APS, analytical performance specification.
Different categories pertaining to this model, i.e., serum C-reactive protein (CRP) as a measurand that, due to its characteristics, cannot be allocated to the other two models, and serum ferritin that theoretically should be allocated to model 1 but does not have available data.

CRP is one of the most requested tests in medical laboratories [20]. It is a sensitive acute phase protein and its concentrations in serum rapidly increase during inflammatory processes. However, an elevation of CRP is not diagnostic of any one specific disease as it occurs in many diseases involving tissue damage or inflammation. Furthermore, CRP is a biologically challenging analyte [21–23]. The meta-analyzed biological variation data made available on the EFLM database (https://biologicalvariation.eu/) give 34.1 % for intra-individual and 83.6 % for interindividual variability, respectively. This very high biological variation is probably deriving from the inclusion in the studies of asymptomatic subjects with transitory illness lacking the mandatory steady state condition [24]. Considering that neither of the first two Milan models appear suitable for CRP, the SA model represents the best option to derive APS for this measurand (Figure 4).

By selecting four IVD-MDs, used in approximately 70 % of our Region laboratories, we retrieved the information about the calibration hierarchy and $u_{\text{cal}}$ for each one [19]. The $u_{\text{Rw}}$ was experimentally estimated as described in Figure 3, using a fresh-frozen serum pool with a CRP concentration close to 10 mg/L, corresponding to the 99th percentile limit of reference value distribution, employed as decision level for detecting sub-clinical infection. Finally, $u(y)$ for each IVD-MD was calculated as $\sqrt{u_{\text{ref}}^2 + u_{\text{cal}}^2 + u_{\text{Rw}}^2}$ and the system with the best analytical performance identified. Data (the full information is available in the relevant paper [19]) showed that the Abbott Architect performance in terms of $u(y)$ (3.76 %) may represent the desirable SA of the CRP measurements to be employed for the definition of APS for MU as “the highest level of analytical performance technically achievable”. This desirable APS can also be modulated to 5.64 % as minimum quality level in accordance with the equation: $3.76 + \frac{1}{2}3.76$ % [19]. Accordingly, all other CRP IVD-MDs evaluated in the study fulfilled the minimum APS for MU. Together with data demonstrating the good harmonization of CRP measurements, this information confirms that the analytical status of this measurand provides fit-for-purpose quality for its clinical use when measured by a representative number of end-user measurement procedures [25, 26]. Potential drawbacks for CRP measurements are confined to employed protocols for assigning traceable values to the IVD-MD calibrators to fulfill during this process MU limits, which represent a fraction of the APS allowed for medical laboratory results, and to changes in reagent and calibrator lots that may influence the IVD-MD intermediate reproducibility.

Circulating ferritin concentrations reflect the amount of intracellular ferritin, the main iron storage protein [27].
As they are used in the decision-making process of iron deficiency anemia (IDA) and the test results can directly influence patient’s management, the outcome-based model for deriving APS should be used [28]. Even recently, we searched available literature for papers evaluating the impact of ferritin analytical variability on clinical outcomes related to IDA, but no such studies could however be retrieved [29]. On the other hand, serum ferritin is a sensitive acute phase reactant with its variation in acute and chronic inflammation not related to iron storage, and this may influence the protein homeostasis in terms of biological variation. In this situation, the SA model for deriving APS can be temporarily used (Figure 5).

For defining SA for ferritin MU, we applied the protocol described in Figure 2 on four IVD-MDs, covering more than 70 % of ferritin measurements in our Region [29]. For each system, we collected from the corresponding manufacturer the information about the metrological traceability and \( u_{\text{cal}} \).

In addition, we experimentally estimated \( u_{\text{Ref}} \) on fresh-frozen serum pools with a ferritin concentration around 70 μg/L, a value in the range indicating an unlikely IDA [28]. Contrary to the CRP situation, \( u_{\text{cal}} \) for ferritin did not include \( u_{\text{ref}} \) of the employed WHO reference materials because it is not available from the provider [29, 30]. Nevertheless, end-user calibrator assigned values had an MU that strongly influenced the overall \( u \) (y) of measurement results, with Roche \( u_{\text{cal}} \) 3.5 to 6 times higher than those of other IVD-MD calibrators [29]. As last step, \( u \) (y) was calculated as \( \sqrt{(u_{\text{cal}}^2 + u_{\text{Ref}}^2)} \). Data (the full information is available in the relevant paper [29]) showed that the Abbott Architect i2000 has the best performance in terms of \( u \) (y), and the corresponding value, i.e., 4.31 %, can be used to define desirable APS for ferritin MU. As previously described, it can be modulated to 6.47 % (4.31 + ½4.31 %) as APS of minimum quality level. Using this minimum goal, three out of the four evaluated IVD-MDs appeared able to fulfil the requirement. The Roche Cobas e601 \( u \) (y) (7.0 %) exceeded the minimum allowable MU (6.5 %), because of the elevated value of \( u_{\text{cal}} \) when compared to the other manufacturer calibrators’ provided information [29].

Concerning serum ferritin, we previously highlighted as information about the traceability implementation and on how the IVD manufacturers validate the calibrator traceability to the selected higher order WHO reference materials is not available [30]. Manufacturers provide \( u_{\text{cal}} \) without any information about the robustness of the value transfer protocol that may even markedly influence it (and therefore \( u \) (y)). Furthermore, although the MU concept assumes that the bias should be appropriately corrected by IVD manufacturers, in the same study we observed a sizeable bias of ferritin results among different IVD-MDs [30]. For instance, the Abbott system, i.e., the IVD-MD that showed the best \( u \) (y) according to the SA protocol described above, showed a 30 % positive bias when compared with Beckman Access, the only IVD-MD that correctly recovered the WHO 94/572 assigned value, and for this reason selected as surrogate reference
procedure [30]. In principle, this bias was not included in the approach of derivation of APS for MU, but, if present, it can markedly reduce the quality of the top-performing IVD-MD. Other experiences showed that the practical application of the traceability concept is still often poorly managed by IVD manufacturers [31–33].

**Concluding remarks**

Traditionally, the SA concept has been merely associated with the average performance obtainable, e.g., from results of EQA programs. In this case, the advantage of the ease of data availability is evident. However, if SA should become a model for APS derivation at least for certain measurands (those that in principles do not fulfil criteria to be allocated to other two Milan models), a different definition focusing on the best quality available should be endorsed and another approach to obtain information conceived. Although the model in principle lacks a direct relationship with clinical needs and appears to be in some way ‘static’ by making a picture of the current IVD-MD performance, our proposed approach to define SA of MU as the highest level of performance technically achievable using the ISO/TS 20914:2019 guidance for the MU estimate may provide useful information about the IVD-MD performance and the quality of provided results. From what described in this article, the evaluation of the SA of IVD-MD performance in terms of MU is however not easy and probably requires coordinated efforts. Furthermore, bias arising from an insufficient correction during the implementation of traceability to higher-order references, if any, may complicate the MU estimate and the definition of its SA. Even if from the proposed categorization it is expected that only IVD manufacturers that do not fulfil SA minimum APS should provide improvement of their products to align their quality to others IVD-MDs, it should be noted that the field is dynamic and new IVD-MDs could be developed, even changing the product categorization and SA information. In this scenario, it is possible that certain manufacturer introduces an improved IVD-MD with a higher performance and it theoretically might result in an update of the SA APS. At the moment, we do not have practical evidence about the effects of the approach proposed in this paper on the market competition in terms of quality improvement and the possible manufacturers’ reaction to the published information. We hope however that our proposal may provide a useful approach in the practical definition of SA to be employed as a model for APS, contributing to improve the knowledge of the quality provided by available IVD-MDs.

**Research ethics:** Not applicable.

**Informed consent:** Not applicable.

**Author contributions:** The authors have accepted responsibility for the entire content of this manuscript and approved its submission.
Competing interests: The authors state no conflict of interest.

Research funding: None declared.

Data availability: Not applicable.

References


