Perspectives

Federica Braga, Sara Pasqualetti, Francesca Borrillo*, Alessia Capoferri, Mariia Chibireva, Leila Rovegno and Mauro Panteghini

# Definition and application of performance specifications for measurement uncertainty of 23 common laboratory tests: linking theory to daily practice

https://doi.org/10.1515/cclm-2022-0806
Received August 17, 2022; accepted September 30, 2022; published online October 26, 2022

**Abstract:** Laboratories should estimate and validate [using analytical performance specifications (APS)] the measurement uncertainty (MU) of performed tests. It is therefore essential to appropriately define APS for MU, but also to provide a perspective on suitability of the practical application of these APS. In this study, 23 commonly ordered measurands were allocated to the models defined during the 2014 EFLM Strategic Conference to derive APS for MU. Then, we checked if the performance of commercial measuring systems used in our laboratory may achieve them. Most measurands (serum alkaline phosphatase, aspartate aminotransferase, creatine kinase, γ-glutamyltransferase, lactate dehydrogenase, pancreatic amylase, total proteins, immunoglobulin G, A, M, magnesium, urate, and prostate-specific antigen, plasma homocysteine, and blood red and white cells) were allocated to the biological variation (BV) model and desirable APS were defined accordingly (2.65%, 4.75%, 7.25%, 4.45%, 2.60%, 3.15%, 1.30%, 2.20%, 2.50%, 2.95%, 1.44%, 4.16%, 3.40%, 3.52%, 1.55%, and 5.65%, respectively). Desirable APS for serum total cholesterol (3.00%) and urine albumin (9.00%) were derived using outcome-based model. Lacking outcome-based information, serum albumin, high-density lipoprotein cholesterol, triglycerides, and blood platelets were temporarily reallocated to BV model, the corresponding desirable APS being 1.25%, 2.84%, 9.90%, and 4.85%, respectively. A mix between the two previous models was employed for serum digoxin, with a 6.00% desirable APS. In daily practice by using our laboratory systems, 16 tests fulfilled desirable and five minimum APS, while two (serum albumin and plasma homocysteine) exceeded goals, needing improvements.

**Keywords:** analytical performance specifications; measurement uncertainty; metrological traceability.

## Introduction

The estimation of measurement uncertainty (MU) of laboratory results is requested to obtain accreditation of medical laboratories according to ISO 15189:2012 standard [1]. The ISO Technical Specification 20914:2019 provides a guidance on how to estimate MU using the so-called "top-down" approach by combining all sources of MU present in the selected traceability chain [2]. In particular, the MU at the level of clinical samples ($u_{result}$) should be the combination of all uncertainty contributions represented by: 1) MU of higher-order references, 2) MU of *in vitro* diagnostics medical device (IVD-MD) calibrators, and 3) MU associated with the random variability of commercial measuring systems [3]. ISO 20914:2019 also emphasizes that the magnitude of estimated MU should be suitable for a result to be used in medical decisions. The definition of an allowable MU is therefore essential to ascertain if estimated MU for a given laboratory result may significantly affect its interpretation [4]. The Strategic Conference of the European Federation of Clinical Chemistry and Laboratory Medicine (EFLM), held in 2014 in Milan, established the criteria for defining analytical performance specifications (APS), based on three models: a) the effect of analytical performance on clinical outcome; b) the biological variation (BV) of the measurand; and c) the state of the art of the

***Corresponding author: Francesca Borrillo**, UOC Patologia Clinica, ASST Fatebenefratelli-Sacco, Via GB Grassi 74, 20157 Milano, Italy, E-mail: francesca.borrillo@unimi.it
**Federica Braga, Sara Pasqualetti, Alessia Capoferri, Mariia Chibireva, Leila Rovegno and Mauro Panteghini,** Research Centre for Metrological Traceability in Laboratory Medicine (CIRME), University of Milan, Milan, Italy. https://orcid.org/0000-0003-3562-7180 (F. Braga)

measurement (defined as the highest level of analytical performance technically achievable) [5, 6]. In a following paper, the criteria for assigning different measurands to each of the three models were made explicit, considering the clinical use of the measurand and its biological characteristics [7].

In a previous study, we defined APS for MU for a first group of 13 measurands according to the above-mentioned models, also briefly providing a preliminary information about their applicability in daily practice [8]. In the present paper, we expanded this approach by analyzing an additional group of 23 measurands, which measurements are frequently requested in Laboratory Medicine. The aim of our study was: a) to categorize the selected measurands according to the Milan models, b) to define APS for MU (at desirable and minimum quality level) using the available information preliminary checked in terms of robustness, and c) to estimate for each test the $u_{result}$ and compare it with set forth APS to see if today's measuring systems are able to meet them.

## Evaluated measurands and search for information associated to selected Milan models

We identified a list of 23 measurands among the most requested tests in our healthcare system. We considered (in alphabetical order): albumin (serum and urine), serum alkaline phosphatase (ALP), serum aspartate aminotransferase (AST), serum creatine kinase (CK), serum digoxin, serum γ-glutamyltransferase (GGT), serum high-density lipoprotein cholesterol (HDL), plasma homocysteine, serum immunglobulin A (IgA), G (IgG), and M (IgM), serum lactate dehydrogenase (LDH), serum magnesium, serum pancreatic amylase (P-AMY), blood platelets, red blood cells (RBC), serum total cholesterol, serum prostate-specific antigen (PSA), serum total proteins, serum triglycerides, serum urate, and white blood cells (WBC).

For derivation of corresponding APS for MU, each measurand was allocated to the best suited Milan model on the basis of its clinical and biological characteristics, and corresponding information was searched and retrieved as follows:

(a) Outcome-based model (model 1). According to previous discussion [6–8], six measurands were in principle allocated to this model: albumin (serum and urine), total cholesterol and HDL, triglycerides, and platelets. We searched peer-reviewed literature for outcome studies dealing with their clinical use and evaluating the impact of random analytical variability on clinical outcomes (systematic search strategy shown in Supplementary File 1 in the online Data Supplement). The APS for MU were derived by identifying the random variability of patient results corresponding to the misclassification rate which was considered clinically acceptable. However, for four measurands (serum albumin, HDL, triglycerides, and platelets), no outcome-based data in literature were retrieved. Therefore, to derive APS we temporarily allocated those measurands to the BV model [7].

(b) BV-based model (model 2). Except for digoxin (see below), all other measurands were allocated to this model. For these measurands, we first retrieved from the EFLM BV database [9] the publications with the highest rate when evaluated for their compliance to the 14 BV data critical appraisal checklist quality items (BIVAC-QI) [10]. We also searched in literature other studies deriving BV of measurands and evaluated them according to BIVAC-QI and other practical guidances (systematic search strategy shown in Supplementary File 2) [11]. For each measurand, the study with the highest score was identified and used to derive APS for MU by the adaption of the classical formulas for deriving analytical goals for random variability from intra-individual BV ($CV_I$), i.e., ≤0.50 $CV_I$ for desirable and ≤0.75 $CV_I$ for minimum quality level, respectively [4, 8, 12]. We are aware that selecting the study with the highest score does not exactly correspond to the strategy employed by the EFLM working group, which is conversely based on the meta-analysis of available data (although we are aware that in their approach, the group applied different weights to reflect the quality of the articles included in the meta-analysis). When studies of elevated grade are available, we consider better using estimates from such studies alone more than the meta-analysis results. Indeed, the use of meta-analysis can be criticized if included studies show significant heterogeneity and different qualities, especially if the majority are of C grade according to BIVAC-QI, as is not rarely the case in the EFLM database.

(c) Model 1&2. According to the concepts elegantly discussed by Fraser [13], drugs which serum concentrations are monitored in laboratory need a specific approach when deriving APS, based on fundamental pharmacokinetic theory and average elimination half-life of the drug. Although the concentration of drugs does not fluctuate randomly around a homeostatic set

point, this approach has a relationship with biological knowledge. On the other hand, therapeutic drug monitoring (TDM) is linked to the patient outcome in defining the levels of drug which are potentially toxic or when the treatment can be ineffective. Accordingly, we considered that directly allocating digoxin (and more in general therapeutically monitored drugs) to Milan models 1 or 2 can be incorrect and a sort of hybrid model between the two models, that we called model 1&2, was tentatively proposed.

Note that in this paper all MU data are reported as standard relative MU (u). They can be expanded (U) at 95.45% level of confidence by multiplying by a coverage factor of 2.

# Measurands belonging to model 1

## Urine albumin

Together with the estimate of glomerular filtration rate, urine albumin is the first-level marker of kidney damage. According to the recommendations of the Kidney Disease Improving Global Outcomes initiative, an urine albumin >30 mg/day, equivalent to urine albumin-to-creatinine ratio (ACR) >30 mg/g, detected for ≥3 months, represents a criterion for the diagnosis of chronic kidney disease (CKD) [14]. Furthermore, on the basis of urine albumin values, three different severity categories (A1, 10–30 mg/g; A2, 30–300 mg/g; and A3, >300 mg/g) can be defined. Studies have suggested that A2 category patients may undergo regression of renal disease if early treated [15]. Accurate detection of the A2 group is therefore central for disease progression monitoring in clinical practice. Ko et al. aimed to estimate the impact of MU on ACR results in classifying patients in A2 category [16]. ACR quality standards were proposed based on the number of ambiguous cases defined as 'cases possibly reclassified in a different severity category because of MU of ACR test'. As there is no guideline regarding the number of ambiguous cases acceptable for clinical purposes, they referred to the classical consensus for setting generally applicable quality goals solely based on biology, accepting 25% (minimum) and 12% (desirable) increases in total result variability. MU levels generating approximately the same number of ambiguous results were proposed as the minimum (17.0%) and desirable (9.0%) APS for standard MU of ACR, respectively. Considering that the standard MU of urine creatinine measurements (<2.3% in our laboratory) does not significantly influence the established goals for ACR, as shown in Table 4 of the article of Ko et al. [16], we adopted for urine albumin the same APS proposed by these Authors for ACR test.

## Serum total cholesterol

Even if more recent guidelines directly focused on low-density lipoprotein cholesterol (LDL), total cholesterol measurements are often requested for the definition of risk for cardiovascular disease (CVD) and some risk scores still ask for knowledge of these values. Petersen and Klee evaluated the influence of analytical variability of total cholesterol measurements on the number of low-risk individuals misclassified as false positive, i.e., individuals at high risk [17]. A misclassification in terms of approximately 5.0% and 7.5% false positive results was obtained with MU of 3.0% (desirable) and 7.0% (minimum), respectively (data from Supplementary Table 2A of ref. [17]).

# Measurands temporarily belonging to model 2

## Serum albumin

Ceriotti et al. summarized the central role of serum albumin measurements in many clinical conditions, recommending to allocate this measurand in the model 1 for deriving APS [7]. However, we were unable to retrieve from literature outcome-based studies about the impact of analytical variability. Therefore, considering that albumin has a major role in assuring stability of colloid osmotic pressure, we temporarily allocated this measurand to the model 2. To this end, we retrieved the publication of Carobene et al. scored as 'A' according to BIVAC-QI [18]. With an average $CV_I$ for serum albumin of 2.50%, APS for MU of 1.25% (desirable) and 1.88% (minimum) were derived.

## Serum HDL

HDL plays a pivotal role in the definition of CVD risk and, similarly to total cholesterol, model 1 should be applied. However, we did not find outcome-based studies and, consequently, model 2 was temporarily applied as the measurand has a steady state in blood of normolipemic individuals. We used the paper by Aarsand et al. [19], estimating the $CV_I$ of this measurand at 5.67%, to derive APS for MU at desirable (2.84%) and minimum (4.26%) quality level.

### Serum triglycerides

Triglyceride measurements are part of the complete lipid profile and their value is also needed to estimate LDL using appropriate formulas. The Adult Treatment Panel III guideline of the U.S. National Cholesterol Education Program adopted the following categorization of serum triglyceride concentrations: physiologic (<150 mg/dL); borderline high (150–199 mg/dL); high (200–499 mg/dL); and very high (≥500 mg/dL) [20]. Given these clearly defined decision thresholds, serum triglycerides should be preferentially allocated to the model 1, even because serum triglyceride can be a biologically challenging analyte, showing relatively high variability if pre-analytical aspects, such as fasting state, are not controlled or a particular lifestyle (e.g., vegetarian diet, alcohol consumption) is adopted. In spite of this caveats, lacking outcome-based studies, the BV model was temporarily employed. Aarsand et al. defined an average triglyceride $CV_I$ of 19.8% [19]. Consequently, APS of 9.90% (desirable) and 14.85% (minimum) were derived.

### Blood platelets

Blood platelet measurements are essential in conditions of excess of bleeding or clotting, and both thrombocytosis and thrombocytopenia can be associated with serious and life-threatening conditions. Specific concentration thresholds have been defined for platelet transfusion [21]. Therefore, for this measurand, the model 1 would apply better. However, even for platelets the lack of outcome-based data to establish APS required the temporary use of model 2. In applying this model, a critical issue that belongs to employed methods in studies evaluating BV of blood cells should be however considered. These measurands have poor stability when blood is stored frozen, so that the usual approach of storing at −80 °C all collected samples until analyses for performing all measurements in the same analytical run and to eliminate the influence of between-run analytical variation is not appropriate. We will discuss this issue in detail when APS for RBC and WBC are described. What we would like to underline here is that even some papers scored 'A' in the EFLM database are not correctly considering and managing this aspect. Conversely, Pineda-Tenor et al. [22] correctly managed the issue by using a previously published experimental design for unstable analytes [23]. In their study, the estimated $CV_I$ for platelets was 9.70%, with derived APS for MU of 4.85% (desirable) and 7.28% (minimum).

## Measurands belonging to model 2

### Serum enzymes

Among Milan models, the one based on the BV has been proposed to derive APS for serum enzymes [24]. In recommending this, Carobene et al. focused on their relatively stable concentrations in healthy individuals and on very different clinical applications that make difficult to define outcome-based APS [25]. For instance, as serum ALP increases may originate from both hepatobiliary and bone disease, and the measurement of ALP alone is unable to diagnose a specific disease, this situation fits better with model 2. Similarly for GGT, AST, and LDH, an increase of which can occur as a result of several hepatic and extra-hepatic pathologies [26]. In a paper scored 'A' according to BIVAC-QI, $CV_I$ for ALP, GGT, AST, and LDH were estimated to be 5.3%, 8.9%, 9.5%, and 5.2%, respectively [25].

For CK and P-AMY, the allocation to models for deriving APS can be more difficult. As previously discussed for serum ALT [8], the clinical role for these enzymes is more defined. Serum CK is the first-level laboratory test in cases of suspected skeletal muscle damage, while P-AMY, although less clinically performing than pancreatic lipase, is still employed for detecting acute pancreatitis [26]. This could require the use of model 1. However, as at the present time outcome-based studies are not available to enable APS setting for these two enzymes, it is rational to use the BV-based model. In the same paper mentioned above [25], $CV_I$ for CK and P-AMY were 14.5% and 6.3%, respectively.

Table 1 shows the derived APS for all mentioned enzymes.

### Serum total proteins

Serum protein quantification does not have a role in a specific disease or clinical condition, and a deviation of their concentrations from the reference interval can be found in a variety of states. Furthermore, the relatively long half-life of the most representative proteins and the strict hormonal control of the body water content make the total protein concentration in serum stable enough [7]. For this reason, the measurand should be allocated to the model 2. From the EFLM database, we retriewed one paper graded

**Table 1:** Milan model allocation and analytical performance specifications (APS) for standard measurement uncertainty on clinical samples ($u_{result}$) for the evaluated measurands.

| Measurand | Milan model | APS for $u_{result}$, % | |
|---|---|---|---|
| | | Desirable | Minimum |
| Urine albumin | Outcome | 9.00 | 17.0 |
| Serum total cholesterol | Outcome | 3.00 | 7.00 |
| Serum albumin | 2Temp[a] | 1.25 | 1.88 |
| Serum HDL cholesterol | 2Temp | 2.84 | 4.26 |
| Serum triglycerides | 2Temp | 9.90 | 14.9 |
| Blood platelets | 2Temp | 4.85 | 7.28 |
| Serum alkaline phosphatase | Biological variation | 2.65 | 3.98 |
| Serum aspartate aminotransferase | Biological variation | 4.75 | 7.13 |
| Serum creatine kinase | Biological variation | 7.25 | 10.9 |
| Serum γ-glutamyltransferase | Biological variation | 4.45 | 6.68 |
| Serum lactate dehydrogenase | Biological variation | 2.60 | 3.90 |
| Serum pancreatic amylase | Biological variation | 3.15 | 4.73 |
| Serum total proteins | Biological variation | 1.30 | 1.95 |
| Serum immunoglobulin G | Biological variation | 2.20 | 3.30 |
| Serum immunoglobulin A | Biological variation | 2.50 | 3.75 |
| Serum immunoglobulin M | Biological variation | 2.95 | 4.43 |
| Serum prostate-specific antigen | Biological variation | 3.40 | 5.10 |
| Serum magnesium | Biological variation | 1.44 | 2.16 |
| Serum urate | Biological variation | 4.16 | 6.24 |
| Plasma homocysteine | Biological variation | 3.52 | 5.27 |
| Red blood cells | Biological variation | 1.55 | 2.33 |
| White blood cells | Biological variation | 5.65 | 8.48 |
| Serum digoxin | 1&2[b] | 6.00 | 9.00 |

[a]2Temp indicates measurands temporarily allocated to the biological variation model because outcome-based data are lacking. [b]A hybrid model specifically developed for drugs (see text for more details).

'A' estimating a $CV_I$ of 2.60% [19]. Derived APS for MU were 1.30% (desirable) and 1.95% (minimum).

## Serum IgG, IgA, and IgM

Immunoglobulins represent a heterogeneous group of glycoproteins, synthesized by plasmacells, with antibody function. Increased concentrations are associated with infectious, inflammatory, or autoimmune conditions as well as with malignant diseases. Therefore, these measurands do not have a role in a specific disease; furthermore, since their serum concentrations are tightly controlled by homeostatic mechanisms, they should be allocated to the model 2. Regarding the BV of immunoglobulins, Ford et al. gave the more accurate information about $CV_I$, with 4.40% for IgG, 5.00% for IgA, and 5.90% for IgM [27]. Corresponding APS for MU are reported in Table 1.

## PSA

PSA is tissue-specific, but not cancer-specific because increased serum PSA concentrations occur in benign prostatic hyperplasia, prostatitis, and following interventions involving the gland [28]. Although blood is not the biological compartment where PSA is physiologically secreted, serum PSA concentrations are stable when a subject is in good health. The measurand can be therefore allocated to model 2. Carobene et al. estimated an average $CV_I$ of 6.80%, defining APS for MU of 3.40% (desirable) and 5.10% (minimum) [29].

## Serum magnesium

Magnesium is the second most abundant intracellular cation, playing a key role in cellular energy metabolism. Its blood concentrations are maintained by a dynamic balance between intestinal absorption, renal excretion, and bone and soft tissue deposition. As other serum ions, magnesium should be allocated to the model 2 [8]. We identified a paper scored 'A' from which a $CV_I$ of 2.88% was obtained [19], and APS for MU of 1.44% (desirable) and 2.16% (minimum) were derived.

## Serum urate

Urate is the final product of purine nucleoside catabolism. Its concentrations in blood are under strict homeostatic control so that the measurand has to be allocated to the BV-based model [7]. In the same paper already quoted for other measurands [19], a $CV_I$ of 8.32% was estimated and APS for MU of 4.16% (desirable) and 6.24% (minimum) derived.

### Plasma homocysteine

Homocysteine is a sulfur-containing amino acid involved in the methionine metabolism. Its measurement is appropriate in case of suspected homocystinuria (an inherited disorder of the methionine metabolism), in patients with previous venous or arterial thromboembolism, and in those with folate and cobalamin deficiency. For its stable biological behaviour, homocysteine should be assigned to model 2. We identified the paper by Garg et al. as the best one, when judged according to the BIVAC-QI [30]. The $CV_I$ derived from this study was 7.03%, and APS for MU were 3.52% (desirable) and 5.27% (minimum).

### Blood cells

RBC and WBC counts are stable in healthy people, so it is reasonable to derive APS using the BV-model. As already mentioned for platelets, the instability of blood cells in samples stored frozen requires however a different experimental design when BV of these measurands is estimated. In particular, blood samples must be immediately assayed after collection, and the generated between-day CV by this protocol should be estimated by assaying in each run a control material having a concentration near the mean of the subjects studied and then subtracted from the total variation of data to obtain the BV estimates [11, 23]. As this essential aspect is not explicitly considered in the BIVAC-QI [10], we revaluated all retrieved papers dealing with blood cell BV by specifically focusing on this issue in addition to other items present in the checklist. The study by Pineda-Tenor et al. correctly applied an experimental design for instable analytes [22]. Therefore, we considered the estimated $CV_I$ as accurate: 3.10% for RBC and 11.3% for WBC, respectively. Table 1 is showing the derived APS.

## Measurand belonging to model 1&2

### Serum digoxin

Digoxin is a cardiac glycoside obtained from digitalis plants. Although it is less frequently used than in the past because of the availability of newer drugs, digoxin is still needed for treatment of supraventricular arrhythmias because of its activity on atrioventricular nodal conduction [31]. As it has a narrow therapeutic range, close TDM is necessary that should be carried out together with clinical monitoring [32]. Using a theoretical model based on digoxin pharmacokinetic and biological knowledge, Fraser determined the desirable analytical variation goal for this drug as follows: $CV \leq ¼ [(2^{T/t} - 1)/(2^{T/t} + 1)] \times 100\%$, where T is the time interval between doses and t is the average elimination half-life of drug [13]. Considering that digoxin is usually given as a single daily dose and the average half-life for individuals without impaired renal function is 38.4 h, the desirable APS can be fixed at 6.0% and the quality level modulate to minimum goal of 9.0% (6.0% + ½ 6.0%), as previously described [8].

Table 1 summarizes the Milan model allocation and APS for standard MU on clinical samples ($u_{result}$) for the discussed measurands.

## MU of laboratory tests evaluated and compared with established APS

As mentioned in the 'Introduction' section, the 'top-down' approach for MU estimate, as recommended by the ISO Technical Specification 20914:2019 to medical laboratories, relies on the definition of MU across the entire calibration hierarchy assuming that all the significant systematic error (bias) is estimated and corrected by the IVD-MD manufacturer, and the uncertainty of correction ($u_{bias}$) defined [3]. $u_{result}$ is calculated using the equation $\sqrt{(u_{cal}^2 + u_{Rw}^2)}$, where $u_{cal} = \sqrt{(u_{ref}^2 + u_{value\ assignment}^2 + u_{bias}^2)}$, if not negligible, and $u_{Rw}$ is defined by ISO/TS 20914:2019 as "uncertainty component under conditions of within-laboratory precision" [3]. Because more than one metrological traceability option may be available for the transfer trueness process and MU of IVD-MD may be influenced by the selected traceability chain [33], $u_{cal}$ should be always provided as combined with uncertainties introduced by higher levels of the selected calibration hierarchy. In practice, however, few, if none, manufacturers provide $u_{cal}$ estimated as described above. What manufacturers are usually providing on request is $u_{value\ assignment}$ of commercial calibrator, so that the laboratory should independently retrieve, when available, the corresponding $u_{ref}$ on the basis of higher-order reference declared by the IVD-MD manufacturer and combine it to the former to obtain the correct $u_{cal}$ estimate.

For the IVD-MDs we tested in this study to compare obtained $u_{result}$ to the established APS for the 23 selected tests, none of the manufacturers of IVD-MDs provided $u_{cal}$ combined with the corresponding $u_{ref}$, even when, as in the majority of cases, this was available. Manufacturers were asked for metrological traceability information in order to identify the higher-order references (materials and/or procedures) used to assign traceable values to

their calibrators and obtain a description of the applied calibration hierarchy. Abbott and Roche issued traceability and MU documents where only the names of higher-order references of each analyte are enlisted, without any indication regarding the internal procedure followed for the implementation of the selected metrological traceability. Beckman Coulter and Sysmex, in providing certification of the calibrator uncertainty (as $u_{value\ assignment}$) also attached documents reporting a general description of applied calibration hierarchy. So that, in the majority of cases, we retrieved by ourselves $u_{ref}$, when available, from the certificate of analysis of the stated reference material found on the website of material supplier. A different approach was used for enzymes that (except for P-AMY) in the Abbott Alinity c system are calibrated using a calibration factor instead of a calibrator material. In this case, $u_{result}$ was obtained as previously described in a validation study [34].

Basics to the 'top-down' approach for estimating MU is the correct estimate by the medical laboratory of the random variability ($u_{Rw}$). $u_{Rw}$ should be derived from internal quality control (IQC) data as described in detail in a previous paper [35]. In this study, results of daily measurements of control materials obtained over a 6-month period were employed as recommended in the ISO 20914:2019 Technical Specification [2]. As previously described, the IQC material employed for $u_{Rw}$ estimate should be different from that used for the verification of measuring system alignment, be commutable and, if possible, with measurand concentrations near to the employed clinical cut-points [3, 35, 36]. When commercially unavailable, in-house IQC materials should be prepared (e.g., by arranging pools of selected samples). In addition, IQC material measurements should be performed randomly inside the analytical run, mimicking measuring conditions of clinical samples, under properly verified system alignment [35].

Table 2 reports MU contributions for evaluated measurands when measured with the measuring systems available in our laboratory, highlighting the fulfillment (or not) of the established APS. Our data can be used for directly answering to the question: "how many measurands can achieve the recommended level of MU and can the tested measuring systems hit these targets?". As shown, 16 tests fulfilled desirable and 5 minimum APS for MU, while only 2 assays (serum albumin and plasma homocysteine) exceeded them. MU associated with a serum albumin result of approximately 40 g/L (3.54%) was about two times higher than the minimum APS derived from BV (≤1.88%). Depending from the biology and strict homeostatic control of serum albumin, the analytical quality required for its measurement is extremely high and MU should be as much as possible small. In previous papers, we discussed in detail the issue of MU of serum albumin and how MU of the currently available reference material (i.e., ERM-DA470k/IFCC) is probably not small enough to guarantee the performance needed in terms of MU for the clinical usefulness of the test [4, 37, 38]. We previously recommended that no more than one third of APS for MU should be consumed by the $u_{ref}$, letting the remaining allowable uncertainty available for other MU sources in the lower parts of calibration hierachy, i.e., $u_{value\ assignment}$ and $u_{Rw}$ [33, 39]. Serum albumin is therefore representative of a measurand for which it should be a priority to significantly reduce the uncertainty associated to the upper levels of metrological chain. On the other hand, parallel strategies focused on reducing the contribution of the lowest parts of the traceability chain to the total uncertainty budget should also be envisaged.

Plasma homocysteine appeared not far (5.67% vs. 5.27%) from fulfilling the minimum APS for MU. It should however be noted that Abbott does not provide for their assay $u_{ref}$ making the $u_{result}$ underestimated. More importantly, given the availability of mass spectrometry-based reference measurement procedures, listed in the database of Joint Committee on Traceability in Laboratory Medicine (JCTLM), manufacturers should use them (and not an internal standard) in order to provide traceability to higher-order references and assure that the bias is appropriately eliminated [40]. $u_{Rw}$ of the employed measuring system was also quite wide, so that an increase in its precision should be considered by the manufacturer.

Except for these two measurands for which further improvements in the quality of their measurements (at least in the particular tested systems) are needed, in our working conditions the remaining 21 tests fulfilled at least the minimum APS for MU. Further research is certainly needed to see if other commercial measuring systems can achieve or not these APS. However, we should preliminarily consider the proposed APS for MU realistic and not impossible to fulfil. One potential limitation of our study was that only one concentration level for IQC material was employed to obtain $u_{Rw}$ in order to simplify the protocol of this preliminary evaluation study. The importance of more than one level for the accurate estimate of MU of the test has been previously underlined because MU may vary with the analyte concentration [36].

**Table 2:** Characteristics of measuring systems and measurement uncertainty contributions for measurands evaluated in this study.

| Platform/Measurand | Reagent code | Method principle | Calibrator type | Stated traceability | $u_{ref}$, % | $u_{value\ assignment}$, % | $u_{Rw}$, % | $u_{result}$[a,b], % | Tested concentration[c] |
|---|---|---|---|---|---|---|---|---|---|
| Abbott Alinity c | | | | | | | | | |
| Serum albumin | DiAgam ALTUR-L00/ALI | Immunoturbidimetry | MPREK-000 DiAgam | ERM-DA470k/IFCC | 1.61 | 2.50 | 1.93 | **3.54** | 40.5 g/L[d] (119) |
| Urine albumin | 08P04 | Immunoturbidimetry | 08P04 Microalbumin calibrator | BCR CRM 470 | 1.01 | 1.67 | 3.81 | 4.28 | 194.3 mg/L[e] (92) |
| Serum digoxin | 08P37 | Particle-enhanced turbidimetric inhibition immunoassay (PETINIA) | 08P74 Multiconstituent calibrator | USP Grade Digoxin | NA | 1.88 | 4.79 | 5.15 | 3.4 µg/L[d] (55) |
| Serum HDL cholesterol | 07P75 | Homogeneous assay | 09P14/5P56 Lipid Multiconstituent Calibrator | CRMLN–HDL verification set | 3.06 | 0.47 | 2.33 | *3.87* | 65.1 mg/dL[d] (97) |
| Serum magnesium | 08P19 | Enzymatic U.V. | 08P60 Multiconstituent Calibrator | NIST SRM 956d | 0.36 | 0.21 | 1.62 | *1.67* | 3.83 mg/dL[d] (159) |
| Serum pancreatic amylase | 01R04 | Immunoinhibition – enzymatic EPS | 08P65 Clin Chem Calibrator | p-Nitrophenol molar absorptivity | NA | 2.95 | 0.87 | 3.08 | 347 U/L[d] (168) |
| Serum total cholesterol | 07P76 | Enzymatic | 08P60 Multiconstituent Calibrator | CRMLN – Total cholesterol verification set | 0.20 | 0.27 | 1.16 | 1.21 | 255.1 mg/dL[d] (164) |
| Serum triglycerides | 07P77 | Enzymatic colorimetric | 08P60 Multiconstituent Calibrator | ACS Grade Glycerol | NA | 0.27 | 1.08 | 1.11 | 189.5 mg/dL[d] (164) |
| Serum urate | 08P56 | Enzymatic colorimetric | 08P60 Multiconstituent Calibrator | NIST 913b | 0.10 | 0.13 | 1.70 | 1.71 | 9.9 mg/dL[d] (137) |
| | | | | | $u_{bias}$, %[f] | | $u_{Rw}$, % | $u_{result}$[a,b], % | |
| Serum alkaline phosphatase | 08P20 | Enzymatic | Calibration factor (2290) | IFCC RMP | 1.30 | | 1.18 | 1.70 | 428 U/L[d] (99) |
| Serum aspartate aminotransferase | 08P23 | Enzymatic | Calibration factor (6835) | IFCC RMP | 2.70 | | 0.96 | 2.87 | 209 U/L[d] (132) |
| Serum creatine kinase | 08P42 | Enzymatic | Calibration factor (9081) | IFCC RMP | 2.20 | | 1.06 | 2.44 | 531 U/L[d] (161) |
| Serum γ-glutamyltransferase | 07P73 | Enzymatic | Calibration factor (8372) | IFCC RMP | 1.40 | | 1.30 | 1.91 | 174 U/L[d] (167) |
| Serum lactate dehydrogenase | 07P74 | Enzymatic | Calibration factor (11180) | IFCC RMP | 1.60 | | 1.47 | 2.17 | 420 U/L[d] (163) |
| | | | | | $u_{ref}$, % | $u_{value\ assignment}$, % | $u_{Rw}$, % | $u_{result}$[a,b], % | |
| Abbott Alinity i | | | | | | | | | |
| Plasma homocysteine | 09P28 | Chemiluminescence microparticle immunoassay (CMIA) | 09P28-01 Homocysteine Calibrator | Internal Standard (S-adenosyl-L-homocystein in phosphate buffer) | NA | 2.34 | 5.16 | **5.67** | 8.9 µmol/L[d] (130) |

**Table 2:** (continued)

| Platform/Measurand | Reagent code | Method principle | Calibrator type | Stated traceability | $u_{ref}$, % | $u_{value\ assignment}$, % | $u_{Rw}$, % | $u_{result}$[a,b], % | Tested concentration[c] |
|---|---|---|---|---|---|---|---|---|---|
| Sysmex XN-9000 | | | | | | | | | |
| Red blood cells | CU228496 | Impedance technology | XN CAL | ICSH RMP | NA | 0.50 | 0.69 | 0.85 | $4.62 \times 10^{12}$/L[g] (173) |
| White blood cells | BL121531 | Flow citometry | XN CAL | ICSH RMP | NA | 0.70 | 1.49 | 1.65 | $7.52 \times 10^{9}$/L[g] (163) |
| Blood platelets | CU228496 | Impedance technology | XN CAL | ICSH RMP (RBC/platelet ratio method) | NA | 2.15 | 2.41 | 3.23 | $193.2 \times 10^{9}$/L[g] (167) |
| Beckman Coulter AU480 | | | | | | | | | |
| Serum total protein | OSR 6132 | Photometric colorimetric | OSR 66300 System Calibrator | NIST SRM 927c | 0.52 | 0.59 | 1.49 | *1.68* | 68.3 g/L[d] (122) |
| Serum immunoglobulin G | OSR61172 | Immunoturbidimetry | 1 ODR 3021 Serum protein multicalibrator | BCR CRM 470 | 0.52 | 1.23 | 2.00 | *2.40* | 10.1 g/L[d] (123) |
| Serum immunoglobulin A | OSR61171 | Immunoturbidimetry | 1 ODR 3021 Serum protein multicalibrator | BCR CRM 470 | 1.02 | 1.49 | 2.45 | *3.04* | 1.90 g/L[d] (127) |
| Serum immunoglobulin M | OSR61173 | Immunoturbidimetry | 1 ODR 3021 Serum protein multicalibrator | BCR CRM 470 | 1.44 | 1.27 | 2.02 | 2.79 | 0.90 g/L[d] (123) |
| *Roche Cobas 8000* | | | | | | | | | |
| Serum prostate-specific antigen | 08791732190 | Sandwich immunoassay | 08838534 Total PSA CalSet II | WHO International Standard 96/670 | NA | 0.93 | 2.13 | 2.32 | 4.36 µg/L[h] (129) |

$u_{ref}$, standard uncertainty of stated reference; $u_{value\ assignment}$, standard uncertainty of commercial calibrator as declared by the manufacturer; $u_{Rw}$, standard uncertainty associated with the random variability of measuring system; USP, United States Pharmacopeia; NA, not available; CRML, Cholesterol Reference Method Laboratory Network; EPS, ethylene-protected substrate; ACS, American Chemical Society; RMP, reference measurement procedure; ICSH, International Council for Standardization in Haematology. [a]Combining the uncertainties displayed in the previous u columns. [b]Values higher than the desirable APS are in italics; values higher than the minimum APS are in bold. [c]Number of measurements over a 6 month period in parentheses. [d]$u_{Rw}$ estimated on Bio-Rad Liquichek Chemistry Control Level 2 ref. 692. [e]$u_{Rw}$ estimated on Randox Assayed Urine Control level 3, ref. AU2353. [f]Three components contributed to $u_{bias}$: a) the average difference between the obtained mean for the employed reference material and the corresponding IFCC RMP target value, b) the bias variability (expressed as relative SD of individual bias divided by the square root of the number of measurements of reference material), and c) the relative standard uncertainty of the target value assigned to the reference material by the IFCC RMP (for more information, see ref. [34]). [g]$u_{Rw}$ estimated on Bio-Rad Liquichek Hematology Control (X) Level 2 ref. 487. [h]$u_{Rw}$ estimated on in-house ad hoc prepared serum pool. Values higher than the minimum APS are in bold. Values higher than the desirable APS are in italics.

**Research funding:** None declared.
**Author contributions:** All authors have accepted responsibility for the entire content of this manuscript and approved its submission.
**Competing interests:** Authors state no conflict of interest.
**Informed consent:** Not applicable.
**Ethical approval:** Not applicable.

## References


1. ISO 15189: 2012. Medical laboratories: requirements for quality and competence, 3rd ed. Geneva: ISO; 2012.
2. ISO/TS 20914: 2019. Medical laboratories: practical guidance for the estimation of measurement uncertainty, 1st ed. Geneva: ISO; 2019.
3. Braga F, Panteghini M. The utility of measurement uncertainty in medical laboratories. Clin Chem Lab Med 2020;58:1407–13.
4. 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.
5. 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.
6. Panteghini M, Ceriotti F, Jones G, Oosterhuis W, Plebani M, Sandberg S. Strategies to define performance specifications in laboratory medicine: 3 years on from the Milan Strategic Conference. Clin Chem Lab Med 2017;55:1849–56.
7. 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.
8. 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.
9. Aarsand AK, Fernandez-Calle P, Webster C, Coskun A, Gonzales-Lao E, Diaz-Garzon J, et al. The EFLM biological variation database. Available from: https://biologicalvariation.eu/ [Accessed 1 Jun 2022].
10. Aarsand AK, Røraas T, Fernandez-Calle P, Ricos C, Díaz-Garzón J, Jonker N, et al. The biological variation data critical appraisal checklist: a standard for evaluating studies on biological variation. Clin Chem 2018;64:501–14.
11. Braga F, Panteghini M. Generation of data on within-subject biological variation in laboratory medicine: an update. Crit Rev Clin Lab Sci 2016;53:313–25.
12. Bais R, Armbruster D, Jansen RT, Klee G, Panteghini M, Passarelli J, et al. Defining acceptable limits for the metrological traceability of specific measurands. Clin Chem Lab Med 2013;51: 973–9.
13. Fraser CG. Desirable standards of performance for therapeutic drug monitoring. Clin Chem 1987;33:387–9.
14. KDIGO CKD Work Group. KDIGO 2012 clinical practice guideline for the evaluation and management of chronic kidney disease. Kidney Int Suppl 2013;3:1–150.
15. de Boer IH, Afkarian M, Rue TC, Cleary PA, Lachin JM, Molitch ME, et al. Renal outcomes in patients with type 1 diabetes and macroalbuminuria. J Am Soc Nephrol 2014;25:2342–50.
16. Ko DH, Lee SW, Hyun J, Kim HS, Park MJ, Shin DH. Proposed imprecision quality goals for urinary albumin/creatinine ratio. Ann Lab Med 2018;38:420–4.
17. Petersen PH, Klee GG. Influence of analytical bias and imprecision on the number of false positive results using guideline-driven medical decision limits. Clin Chim Acta 2014; 430:1–8.
18. Carobene A, Aarsand AK, Guerra E, Bartlett WA, Coşkun A, Díaz-Garzón J, et al. European Biological Variation Study (EuBIVAS): within- and between-subject biological variation data for 15 frequently measured proteins. Clin Chem 2019;65:1031–41.
19. Aarsand AK, Díaz-Garzón J, Fernandez-Calle P, Guerra E, Locatelli M, Bartlett WA, et al. The EuBIVAS: within- and between-subject biological variation data for electrolytes, lipids, urea, uric acid, total protein, total bilirubin, direct bilirubin, and glucose. Clin Chem 2018;64:1380–93.
20. Expert Panel on detection, evaluation and treatment of high blood cholesterol in adults. Executive summary of the third report of the national cholesterol education Program (NCEP) expert Panel on detection, evaluation and treatment of high blood cholesterol in adults (adult treatment Panel III). J Am Med Assoc 2001;285:2486–97.
21. Kaufman RM, Djulbegovic B, Gernsheimer T, Kleinman S, Tinmouth AT, Capocelli KE, et al. Platelet transfusion: a clinical practice guideline from the AABB. Ann Intern Med 2015;162: 205–13.
22. Pineda-Tenor D, Laserna-Mendieta EJ, Timón-Zapata J, Rodelgo-Jiménez L, Ramos-Corral R, Recio-Montealegre A, et al. Biological variation and reference change values of common clinical chemistry and haematologic laboratory analytes in the elderly population. Clin Chem Lab Med 2013;51:851–62.
23. Pagani F, Panteghini M. Significance of various parameters derived from biological variability for lipid and lipoprotein analyses. Clin Biochem 1993;26:415–20.
24. Infusino I, Frusciante E, Braga F, Panteghini M. Progress and impact of enzyme measurement standardization. Clin Chem Lab Med 2017;55:334–40.
25. Carobene A, Røraas T, Sølvik UØ, Sylte MS, Sandberg S, Guerra E, et al. Biological variation estimates obtained from 91 healthy study participants for 9 enzymes in serum. Clin Chem 2017;63: 1141–50.
26. Panteghini M, Bais R. Serum enzymes. In: Tietz textbook of clinical chemistry and molecular diagnostics, 6th ed; 2018:404–34 pp.
27. Ford RP, Mitchell PE, Fraser CG. Desirable performance characteristics and clinical utility of immunoglobulin and light-chain assays derived from data on biological variation. Clin Chem 1988;34:1733–6.
28. Ferraro S, Bussetti M, Rossi RS, Incarbone GP, Panteghini M. Is pre-biopsy serum prostate specific antigen retesting always justified? A study of the influence of individual and analytical factors on decision making for biopsy referral. Clin Chim Acta 2021;516:77–82.
29. Carobene A, Guerra E, Locatelli M, Cucchiara V, Briganti A, Aarsand AK, et al. Biological variation estimates for prostate specific antigen from the European Biological Variation Study; consequences for diagnosis and monitoring of prostate cancer. Clin Chim Acta 2018;486:185–91.

30. Garg UC, Zheng ZJ, Folsom AR, Moyer YS, Tsai MY, McGovern P, et al. Short-term and long-term variability of plasma homocysteine measurement. Clin Chem 1997;43:141–5.
31. Milone MC, Shaw LM. Therapeutic drugs and their management. In: Tietz textbook of clinical chemistry and molecular diagnostics, 6th ed; 2018:800–31 pp.
32. Lucas C, Donovan P. 'Just a repeat' – when drug monitoring is indicated. Aust Fam Physician 2013;42:18–22.
33. 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.
34. 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.
35. Braga F, Pasqualetti S, Aloisio E, Panteghini M. The internal quality control in the traceability era. Clin Chem Lab Med 2021;59: 291–300.
36. Panteghini M. The simple reproducibility of a measurement result does not equal its overall measurement uncertainty. Clin Chem Lab Med 2022;60:e221–2.
37. Infusino I, Panteghini M. Serum albumin: accuracy and clinical use. Clin Chim Acta 2013;419:15–8.
38. Infusino I, Panteghini M. Measurement uncertainty: friend or foe? Clin Biochem 2018;57:3–6.
39. 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.
40. Satterfield MB, Sniegoski LT, Welch MJ, Nelson BC, Pfeiffer CM. Comparison of isotope dilution mass spectrometry methods for the determination of total homocysteine in plasma and serum. Anal Chem 2003;75: 4631–8.

**Supplementary Material:** The online version of this article offers supplementary material (https://doi.org/10.1515/cclm-2022-0806).