

Opinion Paper

Federica Braga* and Mauro Panteghini

Performance specifications for measurement uncertainty of common biochemical measurands according to Milan models

<https://doi.org/10.1515/cclm-2021-0170>

Received February 5, 2021; accepted March 5, 2021;

published online March 17, 2021

Abstract

Objectives: Definition and fulfillment of analytical performance specifications (APS) for measurement uncertainty (MU) allow to make laboratory determinations clinically usable. The 2014 Milan Strategic Conference have proposed models to objectively derive APS based on: (a) the effect of analytical performance on clinical outcome; (b) biological variation components; and (3) the state of the art of the measurement, defined as the highest level of analytical performance technically achievable. Using these models appropriately, we present here a proposal for defining APS for standard MU for some common biochemical measurands.

Methods: We allocated a group of 13 measurands selected among the most commonly laboratory requested tests to each of the three Milan models on the basis of their biological and clinical characteristics. Both minimum and desirable levels of quality of APS for standard MU of clinical samples were defined by using information obtained from available studies.

Results: Blood total hemoglobin, plasma glucose, blood glycosylated hemoglobin, and serum 25-hydroxyvitamin D3 were allocated to the model 1 and the corresponding desirable APS were 2.80, 2.00, 3.00, and 10.0%, respectively. Plasma potassium, sodium, chloride, total calcium, alanine aminotransferase, creatinine, urea, and total bilirubin were allocated to the model 2 and the corresponding

desirable APS were 1.96, 0.27, 0.49, 0.91, 4.65, 2.20, 7.05, and 10.5%, respectively. For C-reactive protein, allocated to the model 3, a desirable MU of 3.76% was defined.

Conclusions: APS for MU of clinical samples derived in this study are essential to objectively evaluate the reliability of results provided by medical laboratories.

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

Background

The main scope of laboratory medicine is to provide useful information for correct clinical decision-making in order to significantly contribute to the quality of health care [1]. Particularly, to make laboratory determinations clinically usable and to ensure that the measurement variability does not gain the upper hand obscuring the clinical information supplied by the test results, it is essential to estimate the measurement uncertainty (MU) of results produced by procedures intended to measure biological measurands, and to verify that the estimated MU is within objectively defined analytical performance specifications (APS) [2]. Recently published documents have recommended how medical laboratories should correctly estimate MU, basically endorsing the so-called “top-down” approach using appropriate internal quality control (IQC) data and commercial calibrator information [2–4]. MU at the level of clinical samples should be the combination of all uncertainty contributions accumulated across the entire traceability chain, starting with MU of reference materials, extending through the *in vitro* diagnostic (IVD) manufacturers and their process for assignment of calibrator values and MU, and ending with the random variability of measuring systems. Very important to the scope of the present paper, the ISO Technical Specification 20914 dealing with the MU estimation also emphasizes that the magnitude of MU should be suitable for a result to be used in a medical decision [3]. Therefore, the definition of an

*Corresponding author: Federica Braga, UOC Patologia Clinica, ASST Fatebenefratelli-Sacco, Via GB Grassi 74, 20157 Milan, Italy; and Research Centre for Metrological Traceability in Laboratory Medicine (CIRME), University of Milan, Milan, Italy, Phone: +390239042743, Fax: +390250319835, E-mail: federica.braga@unimi.it. <https://orcid.org/0000-0003-3562-7180>

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

allowable MU is essential to ascertain if estimated MU for a given laboratory result may significantly affect its interpretation [5–7].

The Strategic Conference organized in Milan in 2014 by the European Federation of Clinical Chemistry and Laboratory Medicine (EFLM) defined three models for establishing APS, i.e., model 1, based on the effect of analytical performance on the clinical outcome; model 2, based on components of biological variation (BV) of the measurands; and model 3, based on the state of the art of the measurement (defined as the highest level of analytical performance technically achievable) [8, 9]. 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 be primarily directed toward the measurand and its biological and clinical characteristics. Accordingly, criteria for allocating laboratory measurands to different models for APS were elaborated and practical principles were proposed [10]. Briefly, model 1 applies for measurands that have a central role in diagnosis and monitoring of a specific disease; model 2 should be used for measurands under strict metabolic control; and model 3 should be used for measurands that cannot be assigned to the first two models.

Defining performance specifications for measurement uncertainty

For MU, the relevant goal that should be fulfilled is that related to the allowable random variability of patient results, as the correct trueness transfer along the traceability chain should permit to achieve unbiased (or negligibly biased) results [2, 11, 12].

The Milan model 1 (“outcome-based APS”) evaluates, directly or indirectly, the impact of a certain analytical variation on laboratory data interpretation and related clinical consequences. As studies investigating the direct impact of performance of laboratory measurements on clinical outcome are seldom available, indirect outcome studies are usually performed [13]. To be included in this model, the measurand should have a central and well-defined role in the decision making of a specific disease or clinical situation and test results should be interpreted through established decision limits [9, 10]. Therefore, the model applies better to measurands which measurements are harmonized, when it is possible to define a common,

method-independent threshold and consequently the impact of the MU in terms of patient misclassification. The APS for MU are defined by identifying the MU level corresponding to the misclassification rate which is considered clinically acceptable [14, 15].

The Milan model 2 (“BV-based APS”) should be adopted when the measurand concentrations in a certain biological fluid are under strict homeostatic control or when the measurand has stable concentrations, i.e. is in a “steady state” status when a subject is not sick. The APS for MU are calculated from intra-individual BV (CV_I) of measurand through an adaptation of the classical formula for deriving analytical goals for random variability [11, 16]. This approach is based on the concept that the total variation of laboratory results is dependent from standard MU and CV_I according to the following formula adding both sources of variation linearly as variances: $(MU^2 + CV_I^2)^{1/2}$. Being the contribution of CV_I to the total variation incompressible, it is therefore necessary to minimize MU that should be modulated according to the specific CV_I of a given measurand. It is agreed that desirable standard MU should be at maximum equal to $0.50 CV_I$.

When a measurand cannot be included in either model 1 or model 2, it should be placed under model 3 (“state of the art-based APS”). To derive APS for MU by using this model it is necessary to identify the highest quality of analytical performance that is at present technically achievable [8]. The major components of the process are as follows: (1) assess combined MU for the selected measurand as recommended by ISO Technical Specification 20914 by using widely employed measuring systems; (2) identify the MU from the best performing system as the ‘desirable’ APS; and (3) establish the ‘minimum’ APS as being 50% greater than the desirable one. We previously reported the example of the measurement of intact human chorionic gonadotropin (hCG), intended as a test to detect pregnancy [4]. Intact hCG is indeed not primarily used for clinical diagnosis (situation to be fulfilled for using model 1) and it is not normally produced in healthy status (situation to be fulfilled for using model 2).

Finally, grading different levels of quality for APS (e.g., minimum and desirable) is also important because it may stimulate IVD manufacturers to work for improving the quality of assays in case of unacceptable or minimum performance [9, 17]. In particular, if the desirable APS is defined, the quality level can be further modulated to minimum as follows: minimum APS = desirable APS + $0.50 \times$ desirable APS. If minimum APS is only established, the quality level can be further modulated to desirable as $0.67 \times$ minimum APS.

Selection of measurands and their categorization according to the models

Now that the theory has been consolidated, it is necessary to move to practice. Particularly, it is required to assign each measurand assayed in the medical laboratory to one of the three models described above and to define APS for its MU for suitable clinical application of the test. In this study, in line with the premises discussed above, we defined APS for MU of the most commonly requested biochemical measurands after their categorization according to the proper model. We identified 13 measurands after examining the yearly number of test requests received by the laboratory of ‘Luigi Sacco’ academic hospital in Milan and selecting the most requested measurands in different biochemistry categories (e.g., metabolites, ions, enzymes, etc.), excluding coagulation factors and cell components. We considered blood total hemoglobin, plasma glucose, blood glycated hemoglobin (HbA_{1c}), plasma creatinine, plasma urea, plasma total bilirubin, plasma sodium, plasma potassium, plasma chloride, plasma total calcium, plasma alanine aminotransferase (ALT), plasma C-reactive protein (CRP), and serum 25-hydroxyvitamin D₃ [25(OH)D₃].

Each selected measurand was associated to one of the three Milan models on the basis of its biological and clinical characteristics. APS for MU of clinical samples were then defined as follows:

- (a) for measurands belonging to the model 1, we searched peer-reviewed literature for outcome studies dealing with the main clinical use of the measurand and evaluating the impact of random analytical variability on clinical outcomes;
- (b) for measurands belonging to the model 2, we retrieved from the EFLM BV database [18] the publications rated as ‘A’, indicating full compliance to the 14 BV data critical appraisal checklist quality items (BIVAC-QI) [19]. From these publications, we derived CV_I estimates needed to calculate APS for MU;
- (c) for measurands belonging to the model 3, we referred to recent studies defining the highest level of analytical performance technically achievable in terms of MU.

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

Measurands belonging to model 1

Plasma glucose

Since plasma glucose determination plays a relevant role in diagnosis of diabetes and its values are used to define glycemic-related conditions, model 1 should be preferred to derive APS [20]. Hyltoft Petersen et al. [21] simulated the influence of analytical variability of plasma glucose measurement on the misclassification of healthy subjects as diabetics [false positives (FP)] and of diseased subject as healthy [false negatives (FN)]. In particular, for unbiased results a random variability (as CV) <2.0% gave a rate of FP and FN of 5.0%. On the other hand, it is important that the application of different levels of suitable analytical quality is considered in the context of the intended use of the plasma glucose measurement. For instance, for diagnosing gestational diabetes stricter glucose APS have been advocated [22]. Tighter APS may be also required to allow for the variation seen in the pre-analytical phase due to variability of glucose stability *in vitro* [20].

Blood HbA_{1c}

HbA_{1c} test also plays a crucial role in the monitoring and diagnosis of diabetes, with clearly defined decision limits [23]. A simulation study by Nielsen et al. [24] evaluated the influence of analytical variability of HbA_{1c} measurement on the number of undiagnosed patients with diabetes. With a maximal variability of 3.0% the percentage of undiagnosed patients with diabetes in the study population was 2%. This misclassification rate increased to 3.7% if the analytical variation grew up to 3.7% (data derived from Fig. 2 of ref. [24]).

Total hemoglobin in blood

The measurement of total hemoglobin in blood is central to diagnose anemia and decision levels for its clinical application are available (i.e., 130 g/L in men, 120 g/L in non-pregnant women, and 110 g/L in children) [25]. Therefore, model 1 should be adopted. Among all the selected measurands belonging to this model, total hemoglobin was however that for which we and other authors [13] were unable to identify solid studies evaluating the impact of analytical variability on diagnostic accuracy. For total hemoglobin, a number of old studies interrogating clinicians

as to what changes in laboratory results are judged as being important are available, but consensus at the EFLM Strategic Conference definitely considered this approach too subjective [8]. One paper in some ways represents a [partial] exception by providing a model accepting a false positive rate of 5% in classifying patients [26]. In this study, the authors identified a 2.8% goal for analytical variability of total hemoglobin measurements.

Serum 25(OH)D3

25(OH)D3 concentrations in serum is currently thought to be the best estimate of vitamin D status of an individual. Although different sets of guidelines for defining clinically relevant states of vitamin D status, especially vitamin deficiency and insufficiency, have been proposed [27], in the light of 25(OH)D3 clinical role, model 1 for APS derivation should be preferred. Stöckl et al. [28] analyzed the impact of analytical variability on interpretation of clinical results by using decision limits commonly used in assessing 25(OH)D3 results for the diagnosis of mild or moderate vitamin insufficiency (30 µg/L or 20 µg/L, respectively) or deficiency (12 µg/L). Assuming a tolerance limit of 20% misclassifications, the quality goal for analytical variability was defined as ≤15%. However, the authors themselves mentioned that the misclassification rate of 20% was chosen on an arbitrarily basis, being probably a too permissible outcome in the field of vitamin status evaluation. In a following clarification they stated indeed that a maximum analytical variability of 10% would be desirable [29]. Quite recently, Cavalier et al. [30] has proposed APS for MU based on the physiological variation of 25(OH)D3 concentrations over time as follows: MU <13.6% to detect a difference at $p < 0.05$ ('minimum' MU) and 9.6% to detect a difference at $p < 0.01$ ('desirable' MU). Interestingly, these APS are not so different from the above-mentioned proposals formulated by the Thienpont's group using the simulation approach [28, 29].

Measurands belonging to model 2

Plasma ions

Plasma ions (i.e., sodium, potassium chloride, and total calcium) are the typical measurands that should be allocated to model 2 [10]. Their concentrations are tightly controlled by

homeostatic mechanisms, including hormonal control and renal function. For the BV data of these measurands, only one paper graded 'A' according to BIVAC-QI was retrieved [31]. The CV_I estimates of sodium (0.53%), potassium (3.92%), chloride (0.98%), and total calcium (1.81%) were used to derive APS for MU.

Plasma creatinine and urea

As the kidney function finely controls their plasma concentrations and assures their stability when a subject is in good health, these measurands should be assigned to the BV-based model. The Aarsand's paper previously mentioned for plasma ions [31] also represents the reference for deriving urea CV_I (14.1%), while another paper from the EuBIVAS project provided the CV_I of creatinine (4.40%) [32].

Total bilirubin in plasma

The bilirubin metabolism is a physiologic process devoted to the elimination of catabolic products that arise from the destruction of red blood cells. Bilirubin concentrations in plasma are therefore well controlled and the measurand is allocable to model 2. Aarsand et al. [31] estimated the CV_I of this measurand at 20.9%.

Plasma ALT

Allocation of plasma ALT to the correct model to derive APS is a tricky issue. Although the definition of its cut-offs has been largely debated [33], this enzyme has surely a clinical role in various liver diseases. This would require the use of model 1. However, at this point in time outcome-based data are not available to enable APS setting for ALT. The EFLM Working Group on Biological Variation has suggested that, since the enzyme demonstrated rather stable catalytic concentrations in healthy individuals, it is rational to use the BV-model to derive APS. The group, in a paper graded 'A' by themselves, has reported a mean CV_I estimate of 9.3%, which in our opinion is not properly expressing a strong enzyme stability, even because a number of subjects and/or drawn blood samples in the study were *a priori* excluded from the analysis because outliers [34].

Measurands belonging to model 3

Plasma CRP

CRP is the most sensitive of the acute phase proteins and its concentrations in plasma increase rapidly in many diseases involving tissue damage or inflammation. Besides not having a role in the decision making of a specific disease, this measurand is a biologically challenging analyte and this makes impossible to derive reliable CRP BV data [35, 36]. Considering that neither of the first two Milan models are suitable for CRP, the model 3 should be used to derive APS for this measurand. A recent paper, by estimating MU of four widely used measuring systems, has defined as desirable a MU for CRP of 3.76% [37].

Table 1 summarizes the Milan model allocation and APS for standard MU on clinical samples for the selected biochemical measurands previously discussed. To provide a preliminary perspective on suitability of the application of these APS, we also added for comparison information on current state-of-the-art performance of our medical laboratory. The MU associated with the laboratory measurements was estimated by using IQC data to derive the random components of MU and commercial calibrator information, as previously described [2]. In particular, random variability of measuring systems was estimated as CV from 6-month consecutive measurement data of a

serum-based fresh-frozen control material, randomly analyzed daily during ordinary laboratory activity [4]. As can be seen, the majority of the desirable MU in the table seem achievable and laboratories can expect to meet the corresponding APS, some are barely achievable (i.e., total calcium and CRP), and some (i.e., plasma sodium and chloride) not achievable, then laboratory professionals may know that industry improvements are required [17].

Concluding remarks

In this paper, we elaborated a synopsis of APS for MU of the most commonly requested biochemical measurands, according to models developed and recommended by the EFLM Strategic Conference held in Milan in 2014, to be used in laboratory practice to validate MU of employed measuring systems. We developed this process according to two main aspects: (1) allocation of each measurand to one of the three models on the basis of its biological and clinical characteristics, and (2) definition of APS for MU by reviewing available literature and selecting adequate information.

Identifying, for a specific measurand, the most appropriate model to derive APS is an essential step towards the definition of suitable MU. We believe that it is not acceptable to use the BV-based model to derive APS for all measurands just because the BV information is more easily

Table 1: Milan model allocation, analytical performance specifications (APS) for standard measurement uncertainty (MU) on clinical samples for the selected biochemical measurands, and corresponding current standard MU performance in the authors' laboratory. Sources of proposed APS are described in the text.

Measurand	Milan model	APS for standard MU, %		U_{result}^a
		Desirable	Minimum	
Plasma glucose	Outcome-based	2.00	3.00 ^b	1.20% [276 mg/dL] ^c
Blood HbA _{1c}	Outcome-based	3.00	3.70	1.50% [48.5 mmol/mol] ^d
Blood total hemoglobin	Outcome-based	2.80	4.20 ^b	0.90% [131.2 g/L] ^e
Serum 25-Hydroxyvitamin D ₃	Outcome-based	10.0	15.0	9.20% [30 µg/L] ^f
Plasma sodium	Biological variation ^g	0.27	0.40	0.83% [153.2 mmol/L] ^c
Plasma potassium	Biological variation ^g	1.96	2.94	1.10% [6.4 mmol/L] ^c
Plasma chloride	Biological variation ^g	0.49	0.74	1.26% [100.0 mmol/L] ^c
Plasma total calcium	Biological variation ^g	0.91	1.36	1.04% [12.4 mg/dL] ^c
Plasma creatinine	Biological variation ^g	2.20	3.30	1.39% [5.75 mg/dL] ^c
Plasma urea	Biological variation ^g	7.05	10.6	2.26% [99.4 mg/dL] ^c
Plasma total bilirubin	Biological variation ^g	10.5	15.7	1.89% [4.95 mg/dL] ^c
Plasma alanine aminotransferase	Biological variation ^g	4.65	6.98	2.20% [88 U/L] ^c
Plasma C-reactive protein	State of the art	3.76	5.64 ^b	4.25% [11.4 mg/L] ^c

^aRepresentative MU associated with laboratory results, combining the MU of calibrator value and the MU accounting for random sources. In turn, 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. In brackets, the concentration level of the measurand to which MU was estimated. ^bAs desirable APS + 0.50 desirable APS. ^cAbbott Alinity c measuring system. ^dAbbott Architect c4000 measuring system. ^eSysmex XN-9000 measuring system. ^fDiaSorin Liaison measuring system. ^gMU APS calculated as 0.50 CV_i (desirable) and 0.75 CV_i (minimum).

obtainable. Measurands with defined role in diagnosis and monitoring of a specific disease should be tested in outcome-based studies and appropriate APS defined. It is a fact, however, that in the last years there has been much progress in improving the quality of BV estimation. By contrast, in our study the assignment of an APS for model 1 appears to be often based on a single study. Performing high quality outcome-based studies is therefore a fundamental requirement for making stronger recommendations about APS for measurands that should be allocated to this model.

The model 2 should not be used for measurands having not sufficient homeostatic control. In general, one should always be wary of too high biological CV estimates. The formula for combining standard MU and CV_1 is dependent on both distributions being Gaussian and with wide CV_1 , this could not be fulfilled [38]. Furthermore, even if an incorrect estimate due to non-Gaussian distribution of evaluated data can be excluded, high biological CVs for sure represent high individual variability and lack of “steady state” conditions. On the other hand, the myth of state of the art as a ‘rescue’ model when APS correctly obtained with other models appear too stringent for a certain measurand should be dismantled. MU estimate gives objective information about the measuring system quality and becomes helpful in identifying measurands that need analytical improvement for their clinical use only when validated against suitable APS [2, 6]. The case of plasma electrolyte measurement recently reported by us showed the efficacy of MU evaluation by using objectively derived APS in driving laboratories to improve the quality of provided results [17].

APS proposed in this paper represent the total MU budget that should be fulfilled at the level of patient results to make the laboratory test information fit for purpose, when combining MU of the measuring system employed in the individual laboratory to that accumulated along all the steps of metrological traceability chain. To facilitate the achievement of these goal it is therefore essential to define the entity of all MU contributions across the different steps of the metrological traceability chain in use and understand how much of the total MU budget can be used at each level. We recommended that no more than one-third of total MU budget should be consumed by the MU of higher-order references and no more than half of the budget is used when MU is combined at the commercial calibrator level [5, 7]. By applying an upside-down approach, it is therefore possible from APS for MU proposed in this paper to calculate the MU goals that should be fulfilled at each previous step of metrological traceability chain [39]. This may contribute to

achieve MU APS at the bottom of the metrological traceability chain (i.e., on patient samples) making laboratory results able to satisfy clinical needs.

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.

Ethical approval: The local Institutional Review Board deemed the study exempt from review.

References

1. Ferraro S, Braga F, Panteghini M. Laboratory medicine in the new healthcare environment. *Clin Chem Lab Med* 2016;54:523–33.
2. Braga F, Panteghini M. The utility of measurement uncertainty in medical laboratories. *Clin Chem Lab Med* 2020;58:1407–13.
3. ISO/TS 20914:2019. Medical laboratories: practical guidance for the estimation of measurement uncertainty, 1st ed. Geneva: ISO; 2019.
4. Braga F, Pasqualetti S, Aloisio E, Panteghini M. The internal quality control in the traceability era. *Clin Chem Lab Med* 2021;59: 291–300.
5. 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.
6. Infusino I, Panteghini M. Measurement uncertainty: friend or foe? *Clin Biochem* 2018;57:3–6.
7. Braga F, Panteghini M. Defining permissible limits for the combined uncertainty budget in the implementation of metrological traceability. *Clin Biochem* 2018;57:7–11.
8. 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.
9. 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.
10. 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.
11. 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.
12. 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.
13. Smith AF, Shinkins B, Hall PS, Hulme CT, Messenger MP. Toward a framework for outcome-based analytical performance specifications: a methodology review of indirect methods for evaluating the impact of measurement uncertainty on clinical outcomes. *Clin Chem* 2019;65:1363–74.

14. Lyon AW, Kavsak PA, Lyon OA, Worster A, Lyon ME. Simulation models of misclassification error for single thresholds of high sensitivity cardiac troponin I due to assay bias and imprecision. *Clin Chem* 2017;63:585–92.
15. Ferraro S, Lyon AW, Braga F, Panteghini M. Definition of analytical quality specifications for serum total folate measurements using a simulation outcome-based model. *Clin Chem Lab Med* 2020;58:e66–8.
16. Adams O, Cooper G, Fraser C, Hubmann M, Jones G, Plebani M, et al. Collective opinion paper on findings of the 2011 convocation of experts on laboratory quality. *Clin Chem Lab Med* 2012;50:1547–58.
17. Pasqualetti S, Chibireva M, Borrillo F, Braga F, Panteghini M. Improving measurement uncertainty of plasma electrolytes: a complex but not impossible task. *Clin Chem Lab Med* 2021;59:e129–32.
18. Available from: <https://biologicalvariation.eu/> [Accessed Jan 2021].
19. 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.
20. Pasqualetti S, Braga F, Panteghini M. Pre-analytical and analytical aspects affecting clinical reliability of plasma glucose results. *Clin Biochem* 2017;50:587–94.
21. Petersen PH, Brandslund I, Jørgensen L, Stahl M, Olivarius Ndf, Johnsen KB. Evaluation of systematic and random factors in measurements of fasting plasma glucose as the basis for analytical quality specifications in the diagnosis of diabetes. 3. Impact of the new WHO and ADA recommendations on diagnosis of diabetes mellitus. *Scand J Clin Lab Invest* 2001;61:191–204.
22. van den Berg SAA, Thelen MHM, Tiel Groenestege WM. Intra-laboratory variation and its effect on gestational diabetes diagnosis. *Clin Chem Lab Med* 2017;55:e216–8.
23. Panteghini M, John WG. Implementation of haemoglobin A1c results traceable to the IFCC reference system: the way forward. *Clin Chem Lab Med* 2007;45:942–4.
24. Nielsen AA, Petersen PH, Green A, Christensen C, Christensen H, Brandslund I. Changing from glucose to HbA1c for diabetes diagnosis: predictive values of one test and importance of analytical bias and imprecision. *Clin Chem Lab Med* 2014;52:1069–77.
25. World Health Organization. Worldwide prevalence of anaemia 1993–2005: WHO global database on anaemia. Geneva, Switzerland: World Health Organization; 2008.
26. Thue G, Sandberg S, Fugelli P. Clinical assessment of haemoglobin values by general practitioners related to analytical and biological variation. *Scand J Clin Lab Invest* 1991;51:453–9.
27. Binkley N, Dawson-Hughes B, Durazo-Arvizu R, Thamm M, Tian L, Merkel JM, et al. Vitamin D measurement standardization: the way out of the chaos. *J Steroid Biochem Mol Biol* 2017;173:117–21.
28. Stöckl D, Sluss PM, Thienpont LM. Specifications for trueness and precision of a reference measurement system for serum/plasma 25-hydroxyvitamin D analysis. *Clin Chim Acta* 2009;408:8–13.
29. Stepman HCM, Thienpont LM. Measurement uncertainty for the analysis of serum 25-hydroxyvitamin D. *Osteoporos Int* 2010;21:1053.
30. Cavalier E, Fraser CG, Bhattoa HP, Heijboer AC, Makris K, Ulmer CZ, et al. Analytical performance specifications for 25-hydroxyvitamin D examinations. *Nutrients* 2021;13:431.
31. Aarsand AK, Diaz-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.
32. Carobene A, Marino I, Coskun A, Serteser M, Unsal I, Guerra E, et al. The EuBIVAS project: within- and between-subject biological variation data for serum creatinine using enzymatic and alkaline picrate methods and implications for monitoring. *Clin Chem* 2017;63:1527–36.
33. Panteghini M, Adeli K, Ceriotti F, Sandberg S, Horvath AR. American liver guidelines and cutoffs for “normal” ALT: a potential for overdiagnosis. *Clin Chem* 2017;63:1196–8.
34. Carobene A, Røraas T, Sølviik 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.
35. Braga F, Panteghini M. Biologic variability of C-reactive protein: is the available information reliable? *Clin Chim Acta* 2012;413:1179–83.
36. Braga F, Ferraro S, Szöke D, Lanzoni M, Panteghini M. Estimate of intraindividual variability of C-reactive protein: a challenging issue. *Clin Chim Acta* 2013;419:85–6.
37. Braga F, Panteghini M. Derivation of performance specifications for uncertainty of serum C-reactive protein measurement according to the Milan model 3 (state of the art). *Clin Chem Lab Med* 2020;58:e263–5.
38. 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.
39. 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.