Opinion Paper

Marc H.M. Thelen* and Marith van Schrojenstein Lantman

When bias becomes part of imprecision: how to use analytical performance specifications to determine acceptability of lot-lot variation and other sources of possibly unacceptable bias

https://doi.org/10.1515/cclm-2023-1303
Received November 15, 2023; accepted January 28, 2024;
published online February 8, 2024

Abstract: ISO 15189 requires laboratories to estimate the uncertainty of their quantitative measurements and to maintain them within relevant performance specifications. Furthermore, it refers to ISO TS 20914 for instructions on how to estimate the uncertainty and what to take into consideration when communicating uncertainty of measurement with requesting clinicians. These instructions include the responsibility of laboratories to verify that bias is not larger than medically significant. If estimated to be larger than acceptable, such bias first needs to be eliminated or (temporarily) corrected for. In the latter case, the uncertainty of such correction becomes part of the estimation of the total measurement uncertainty. If small enough to be acceptable, bias becomes part of the long term within laboratory random variation. Sources of possible bias are (not limited to) changes in reagent or calibrator lot variation or calibration itself. In this paper we clarify how the rationale and mathematics from an EFLM WG ISO/A position paper on allowable between reagent lot variation can be applied to calculate whether bias can be accepted to become part of long-term imprecision. The central point of this rationale is to prevent the risk that requesting clinicians confuse changes in bias with changes in the steady state of their patients.

Keywords: measurement uncertainty; bias; imprecision; lot variation; analytical performance specifications

***Corresponding author: Marc H.M. Thelen**, Scientific Director, SKML, Foundation for Quality Assurance in Laboratory Medicine, Nijmegen, The Netherlands; and Department of Laboratory Medicine, Radboud University Medical Centre, Toernooiveld 300NL-6525 EC, Nijmegen, The Netherlands, Phone: +31 63113 2997, E-mail: mthelen@skml.nl. https://orcid.org/0000-0003-1771-669X

Marith van Schrojenstein Lantman, SKML, Foundation for Quality Assurance in Laboratory Medicine, Nijmegen, The Netherlands; Department of Laboratory Medicine, Radboud University Medical Centre, Nijmegen, The Netherlands; and Result Laboratory for Clinical Chemistry, Amphia Hospital, Breda, The Netherlands. https://orcid.org/0000-0002-5454-990X

The concept of analytical performance specifications

To make sure that measurement methods that claim uniformity of results indeed do so within quantified limits of uncertainty ISO 17511:2020 defines how methods need to be traceable to higher references via an unbroken chain of methods and materials for the transfer of values [1]. To make sure IVD providers implement the concepts in this standard the EU has harmonized this and other ISO standards to the IVDR legislation [2]. What neither ISO17511:2020 nor the IVDR define is which uncertainty is “good enough”, indicating the methods are fit for the intended use. Meanwhile, both clinicians and patients assume that the results provided by medical laboratories reflect the true value of the particular measurands in their body. Even if they are aware of the possibility that repetition of the analysis on the same or a different instrument in the same laboratory or in a different laboratory, possibly using a different measurement procedure, might yield a different absolute value for the same measurement, they assume that these differences are within clinically meaningful limits. It is the responsibility of the laboratory to make this assumption valid, or at least safe, for which they need the scientific societies for laboratory medicine to define clinically acceptable measurement uncertainty limits. The EFLM strategic meeting Milan 2014 has set the stage to determine a rationale for the determination of analytical performance specifications (APS) [3]. Two important and highly cited papers from that meeting gave guidance for choosing models to determine the APS [4] and on dividing the budget for uncertainty of an APS irrespective from what model was assigned [5].

The fifth symposium Cutting Edge of Laboratory Medicine in Europe (CELME), held in October 2023 in Prague, Czech Republic [6] had the aim to translate the theoretical concepts and models from the Milan strategic meeting to practical recommendations. This paper documents the content of one of the lectures of that conference. One important conclusion from another presentation at that conference was that, in order to be practical, we may need to reserve the term analytical performance *specifications* for those performance requirements which can currently be met by the current analytical performance *characteristics* (APC). By definition *state of the art* specifications are compatible with this definition, whereas APS determined by either clinical outcome (simulations) or based on biological variation data may be unreachable with currently available technology. The conference concluded that in such cases the term APS is preferably replaced with analytical performance *goals* (APG). Such goals define the specifications for the future and should be used by the IVD industry to define their ambitions in product development. Meanwhile laboratories need specifications which can and must be met today. Therefore, different models may be needed for the same clinical application of the same measurand at the same time in order to separate needed APGs from reachable APSs. Another conclusion from the 2023 CELME meeting was that laboratories need more guidance on how to use 'their' part of the uncertainty budget. Although the paper of Braga et al. [5] defines that at least 50 % of the measurement uncertainty allowable (MAU) should be available for laboratories, this paper does not provide guidance on how laboratories should spend those 50 % on the different sources of variation. This paper identifies those sources of variation and provides a division of the laboratory uncertainty budget between manageable sources which vary with low frequency, such as reagent lot changes and those which vary with high frequency.

The concept of measurement uncertainty

Errors in measurement resulting in inaccuracy can be distinguished between systematic and random sources. Random errors are quantified as imprecision whereas systematic errors are referred to as bias (Figure 1). From the combination of bias and imprecision the total error (TE) can be calculated. However, since bias has a systematic source, it can be assumed to be constant for the period the source of the bias is present. Because bias is constant for the time it is present, it becomes a manageable part of error, where users of a method can identify and quantify bias as the difference of the mean value from an assigned true value, resulting in their ability to either accept, reject or correct for bias. Especially when the source of the bias is not certain, laboratories need to monitor its constantness and the appropriateness of its correction with regular intervals.

Figure 1 depicts the relationship between the systematic and random sources of error combined in the TE related to the quantities trueness and precision that together define accuracy [7]. Trueness is estimated and expressed as bias and precision is estimated and expressed as imprecision. Bias and imprecision combine to inaccuracy of the single estimate of a measurand in a particular sample. Inaccuracy as a measure for TE is useful in external quality assessment (EQA) that aims to quantify bias from a

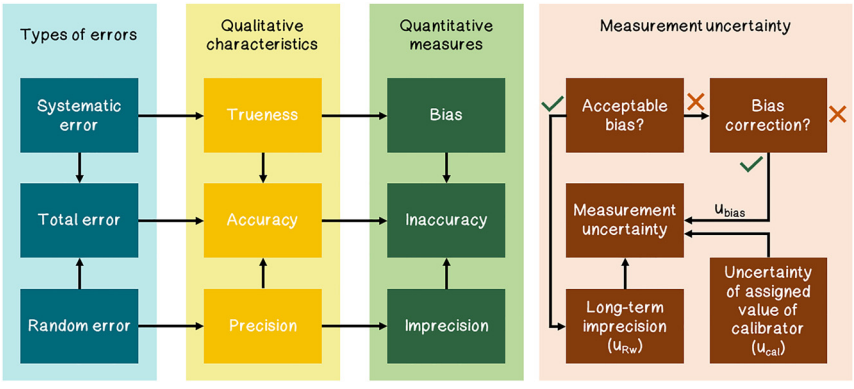

Figure 1: The relationship between types of error (blue), the quality characteristic they are related to (orange), and the quantitative measure related to it (green), adapted from Theodorsson et al. [7]. Bias and imprecision combine to inaccuracy or if bias is maintained within medically acceptable limits, bias will become part of long term imprecision (u_{rw}), which together with the uncertainty of the value assignment of the calibrator (u_{cal}) and – in cases where uncorrected bias is judged as medically significant – the uncertainty of the correction of the bias (u_{bias}), are combined to calculate to express measurement uncertainty (brown). Note that there is no arrow between acceptable bias and measurement uncertainty.

known target concentration. To be able to address bias as a separate source of error, imprecision needs to be separated from total inaccuracy. To do this properly a class 1 EQA scheme is needed with multi sample statistics calculated from the results in commutable samples with value assignment and given uncertainty to higher order reference methods [8, 9]. In contrast to the use of TE in EQA, the TE concept is not appropriate for the expression of the uncertainty of routine patient samples to requesting clinicians, as no laboratory would bother clinicians with results indicated as 'inaccurate'. This calls for the use of measurement uncertainty (MU). If bias is maintained within medically acceptable limits it will behave like a component of long term imprecision with low frequency changes and combines with other sources of imprecision to MU (Figure 1, brown box). ISO TS 20914:2019 [10] has been developed to give practical guidance for laboratories to estimate the uncertainty of their measurements. The graphical workflow in that standard (Figure 2) requires the laboratory to identify whether a medically significant bias is present, and if so, this bias has to be resolved by the manufacturer or (as long as unresolved) corrected for by the laboratory.

The check whether bias is acceptable also implies that bias can be smaller than medically significant, but larger than zero. This means that once accepted all sizes and directions of bias become a source of long-term variation. In ISO TS 20914:2019 this long term variation (u_{Rw}) is estimated as the coefficient of variation (CV) in the results of internal control materials of a period long enough to contain all sources of long-term variation such as maintenance, calibration and changes in the production lot of reagents and calibrators [10, 11]. Together with the u_{Rw}, the uncertainty of the assigned value of the calibrator (u_{cal}) as provided by the manufacturer and the uncertainty of the correction of bias (u_{bias}) (if applicable) form part of the calculation to estimate the MU, or $u_{(y)}$ in Figure 2. Bias as such is not a separate ingredient of these calculations. The reason for that is, although obvious, but poorly understood; as long as the bias is medically acceptable and therefore accepted it will become part of the long term imprecision u_{Rw}. Addition of a separate bias component into the calculation to estimate MU would either result in double counting the bias source or would require cleaning u_{Rw} from the components caused by accepted bias, which – if not impossible – is unpractical. In case the uncertainty of the correction of the bias (u_{bias}) introduced by a certain reagent/calibrator lot is larger than the bias it is correcting for, such correction can better be replaced by rejection of

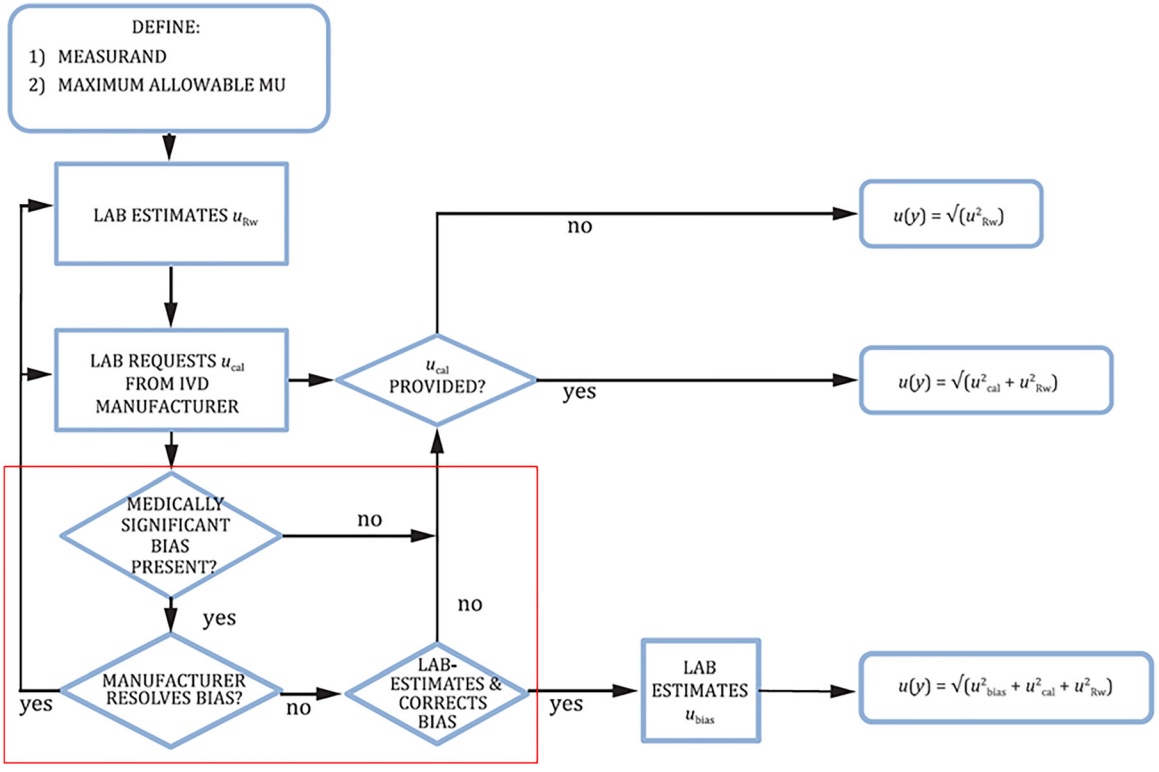

Figure 2: Graphical representation of the workflow to estimate the uncertainty of measurement ($u_{(y)}$) in laboratory medicine, adopted from [10]. The red box indicates the part in which the bias must either be accepted as a source that contributes to the end-user uncertainty or be corrected.

the lot, or in case suboptimal performance is judged in the interest of patient care, the lot should be accepted without correction.

Sources of variation

In any measurement procedure different sources of variations can be identified which can differ in both frequency and amplitude (Figure 3). The within-run repeatability of a measurement procedure for which all sources with possible impact on variation are kept constant is seldomly the relevant magnitude of variation as experienced by the requesting physicians. For them measurement uncertainty is a source to be taken into account along with other sources of uncertainty including, but not restricted to the biological variation of their patients. This is explicitly acknowledged in the 2023 revision of ISO 15189:2022 [12] for laboratories when communicating measurement uncertainty with requesting physicians.

An important source that contributes to measurement uncertainty is the variation introduced by changes in the lots of reagent, including that of the calibrator. Guidance documents like CLSI EP-26A [13] give practical instruction for between reagent lot variation studies with important recommendations on the materials with between lot commutability. However, as many laboratories do in practice, this protocol only compares a new lot to the current lot in use and therefore allows for long term drifts which could get magnitudes larger than the allowable bias and also larger than can be anticipated when the reported between lot CV communicated by the IVD provider is considered. In their paper on acceptable between-lot variation, the EFLM quality committee working group ISO/A [14, 15] has proposed that whether the bias between lots of reagents, including calibrators should be judged as acceptable depends on its opportunity to get noticed by requesting physicians. Figure 4 illustrates how between-lot variation is hard to distinguish as long as the within-lot variation has a similar or larger magnitude. However, the same between-lot variation is instantly recognizable when the within-lot variation is decreased to become smaller than the between-lot variation. Since IVD manufacturers have done a great job in improving mechanical and optical instrument stability and reproducibility as well as reagent stability and homogeneity, this is not a just a theoretical concept. Together with longer lasting periods of lot stability and consequent less frequent lot changes, this has resulted in an increased awareness of between-lot differences, necessitating the need to manage this variation between acceptable limits [16, 17].

The proposal of the EFLM WG is to relate the acceptable between-lot variation (u_{brlot}) to the total u_{Rw}. Since the u_{Rw} without the u_{brlot} solely exists of the within-lot variation component (u_{wrlot}); a relative distribution between u_{wrlot} and u_{brlot} is relevant. The proportion of the budget available for u_{brlot} decreases with increasing frequency of results generated for individual patients within a typical lot of reagent. The rationale for this being that requesting physicians can get familiarized with the combination of within person biological variation and within reagent lot variation for patients with multiple assessments within a reagent lot, and therefore might confuse changes in results introduced by lot changes with changes of the steady state of their patients (Figure 5).

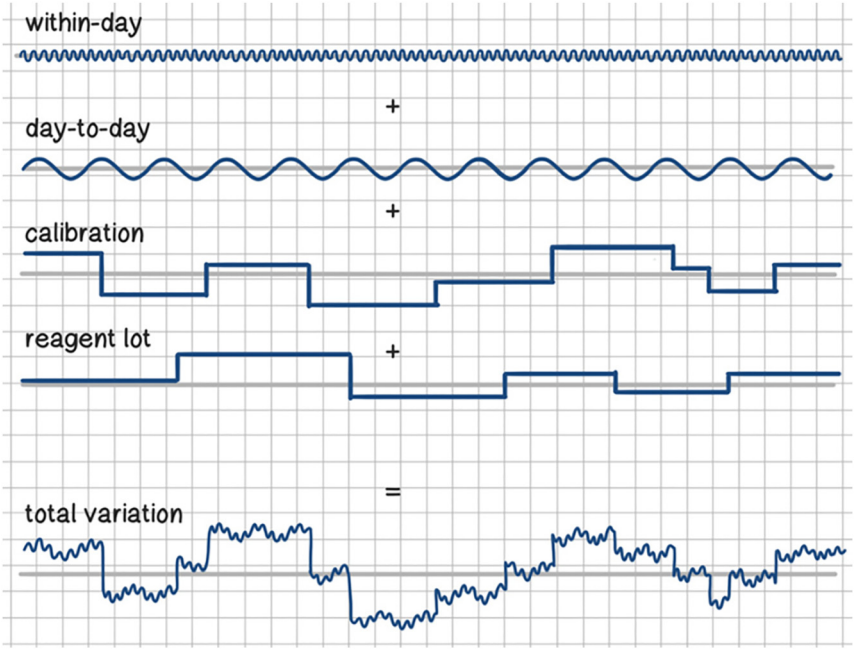

Figure 3: Different sources of within laboratory variation have different amplitudes and frequencies of change [14].

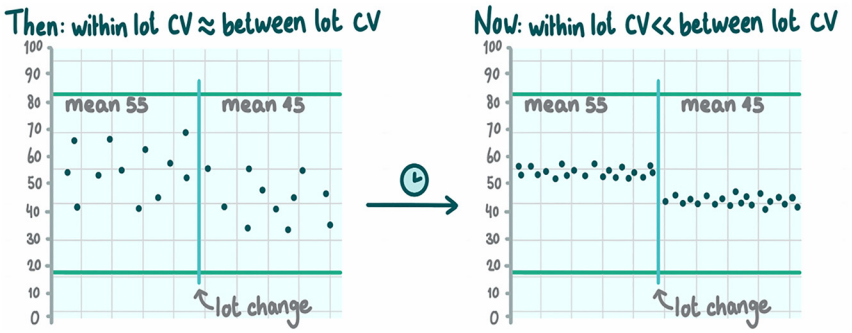

Figure 4: Visualization of the impact of within-lot variation on the recognizability of between-lot variation which is identical in left and right panel.

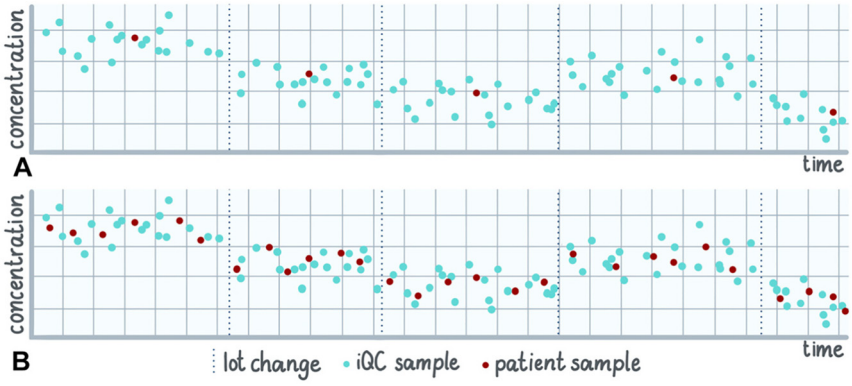

Figure 5: The impact of the number of observations within a period of constant acceptable bias on the possibility to confuse the introduction of variation of such acceptable bias with the change of the steady state of a patient [14].

This makes the rationale and resulting APS for u_{brlot}:u_{wrlot} obvious for measurands used for the monitoring of individual patients. However, it can also be justified to attribute to the use of measurands with incidental application for diagnostic decision purposes. This is because physicians also tend to develop 'feeling' for changes in the fraction of patients flagged as abnormal for a particular measurand. For that reason, also in cases where not the steady state of a patient is monitored, but that of a group, the APS for u_{brlot}:u_{wrlot} can be applicable. For instance, in the monitoring of a national colon cancer screening program it seems obvious that the referral rate for colonoscopy should not suffer from variation in time that can be recognized to depend on reagent lot changes.

The EFLM lot-lot paper (14) proposed that the u_{brlot} should always be smaller than the u_{wrlot}, even in cases where not more than one observation per patient per reagent lot is sampled (Formula 1).

$$u_{brlot} \leq \frac{u_{wrlot}}{\sqrt{n}}$$

Formula 1. *Proposed division of the u_{Rw} budget between within-reagent-lot (wrlot) and between-reagent (brlot) lot sources of variation. With n representing the number of observations of a typical patient within the usage time of one reagent/calibrator lot [14].*

With increasing numbers of single-patient observations per reagent lot the portion of the budget to be used by between-lot variation further decreases, but is capped at a ratio of 2.4 times (the square root reached when n=6) smaller than the within-lot variation as the paper proposes a maximum value of 6 for n in formula 1.

Practical implementation

Before a laboratory introduces a reagent/calibrator lot into service for patient care, it will calibrate it on an analyser, and run only QC materials (either commercial or – if judged necessary for between lot commutability – pooled patient material), whereas the same commutable QC material is also measured on the previous lot of reagent which is still in-use for patient care. The laboratory can then calculate a short-term mean of the QC-results for the new lot, which allows for comparison to the mean of the current and previous lots as all-lot mean. The laboratory can then decide whether to accept, reject or correct the new lot, before taking it into use for regular patient care. The suitability of the all-lot mean is highly dependent on the number of previous lots used. The uncertainty of the all-lot mean will decrease with the number of previous lots. In a

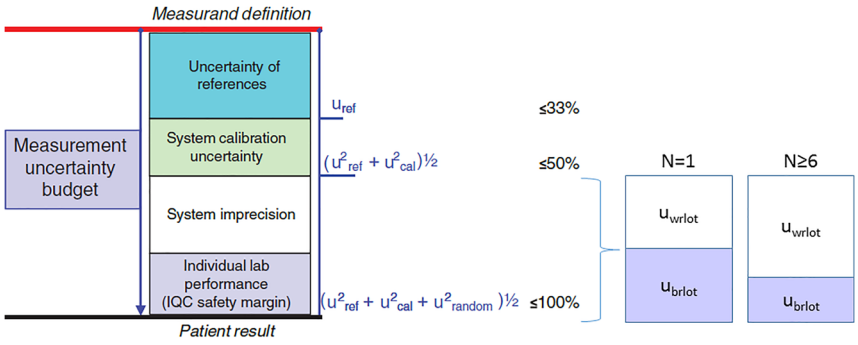

Figure 6: Extension of the graphical representation of the assignment of maximum parts of the uncertainty budget to be used by different parts and parties in the metrological traceability chain [5] with the division of the laboratory budget between manageable sources which can be treated as acceptable bias (purple) and imprecision sources to be accepted as random sources (white).

situation where only one or few previous lots have been used by a laboratory it will be impossible to tell whether the new or the previous lots behave atypical compared to the all-lot mean. In such situations a practical approach is needed where new lots are compared to the mean of all previous lots used so far, with growing certainty of correct decisions with the increase of lot numbers used. At least two scenarios can be helpful in such situations.

(1) Some EQA organisers [18] ask and report lot-based results per method in commutable materials. That can help laboratories to identify whether current and previous lots used by them are typical or atypical in their between lot eccentricity.
(2) Information provided by the IVD manufacturer on both the bandwidth of their reagent/calibrator lot release criteria and the position of every released lot within that bandwidth. In the end all laboratories will find know these facts, so why not inform them on what they will discover sooner or later.

In all cases the commutability of the control material between the lots compared needs to be verified to make sure that between lot differences as assessed are representative for those experienced by patient samples. Although within method, between lot commutability is a quality of a material which is less challenging to fulfil than between method commutability, even such within method between lot commutability is notorious to be not self-evident and therefore needs to be verified.

With this division of the within-laboratory uncertainty budget the graphical representation of the overall division of the uncertainty budget by Braga et al. [5] can be extended to incorporate this division between manageable between-lot sources caused acceptable bias and unmanageable within-lot sources which represent true imprecision (Figure 6).

Responsibility requires capacity

Laboratories are responsible to manage that the uncertainty of measurement is not larger than clinical application and thus allows for useful application of the particular measurand for the intended use. Assignment of different sources of the total MAU between parts of the traceability chain and the parties involved [5] has ensured that laboratories are only responsible for that part of the chain which they can control. Differentiation between APS and APG ensures that laboratories have a realistic challenge to manage their APC within APS, whereas the scientific societies should cooperate with the IVD industry to develop products for which the APS meets the APG. By dividing the part of budget which is available for laboratories between sources that can either be influenced or have to be accepted ‘as is’ can further enable laboratories to assure their APSs are met. This division is fair for laboratories because they cannot be responsible for something they cannot control. Laboratories have the capacity to decide on the acceptance of lot-variation and calibration-variation where in contrast they can only influence their system imprecision by maintenance and/or mitigate its effects by reporting the mean obtained result of multiple estimates rather than one. To use their capacity laboratories need opportunity and competence, with the acquisition of the last being a professional responsibility.

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

1. ISO 17511:2020. In vitro diagnostic medical devices, requirements for establishing metrological traceability of values assigned to calibrators, trueness control materials and human samples. 2020.
2. European Parliament and Council Regulation (EU) 2017/746 of 5 April 2017 on in vitro diagnostic medical devices. Off J Eur Union 2017;117: 176–332.
3. Sandberg S, Fraser FG, 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.
4. 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.
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. 5th symposium CELME 2023, cutting edge of laboratory medicine in Europe, analytical performance specifications, moving from models to practical recommendations. Prague, Czech Republic: Charles University; October 12–13, 2023. http://www.celme2023.cz/programme.php [Accessed 4 Jan 2024].
7. Theodorsson E. Uncertainty in measurement and total error: tools for coping with diagnostic uncertainty. Clin Lab Med 2017;37:15–34.
8. Miller WG, Jones GR, Horowitz GL, Weykamp C. Proficiency testing/external quality assessment: current challenges and future directions. Clin Chem 2011;57:1670–80.
9. Thelen MH, Jansen RT, Weykamp CW, Steigstra H, Meijer R, Cobbaert CM. Expressing analytical performance from multi-sample evaluation in laboratory EQA. Clin Chem Lab Med 2017;55:1509–16.
10. ISO/TS 20914:2019. Medical laboratories — practical guidance for the estimation of measurement uncertainty. 2019.
11. Braga F, Pasqualetti S, Aloisio E, Panteghini M. The internal quality control in the traceability era. Clin Chem Lab Med 2020;59:291–300.
12. ISO 15189:2022. Medical laboratories, Requirements for quality and competence. 2022.
13. CLSI. User evaluation of acceptability of a reagent lot change, 2nd ed. CLSI guideline EP26. Clinical and Laboratory Standards Institute; 2022.
14. van Schrojenstein Lantman M, Çubukçu HC, Boursier G, Panteghini M, BernabeuAndreu FA, Milinkovic N, et al. An approach for determining allowable between reagent lot variation. Clin Chem Lab Med 2022;60: 681–8.
15. Thelen MHM, van Schrojenstein Lantman M, Boursier G, Vanstapel F, Panteghini M. In reply to: limitations in using the EFLM WG-A/ISO approach for assessment of reagent lot variability. Clin Chem Lab Med 2023;61:e218–20.
16. Loh TP, Sandberg S, Horvath AR. Lot-to-lot reagent verification: challenges and possible solutions. Clin Chem Lab Med 2022;60:675–80.
17. Plebani M, Zaninotto M. Lot-to-lot variation: no longer a neglected issue. Clin Chem Lab Med 2022;60:645–6.
18. Stavelin A, Riksheim BO, Christensen NG, Sandberg S. The importance of reagent lot registration in external quality assurance/proficiency testing schemes. Clin Chem 2016;62:708–15.