Clinical Biochemistry 57 (2018) 37–40

Contents lists available at ScienceDirect

Clinical Biochemistry

journal homepage: www.elsevier.com/locate/clinbiochem

Review

# Deriving proper measurement uncertainty from Internal Quality Control data: An impossible mission?

Ferruccio Ceriotti*

*Clinical Laboratory, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Milano, Italy*

ARTICLE INFO

*Keywords:*
Measurement uncertainty
Internal Quality Control
Bias
Imprecision

ABSTRACT

Measurement uncertainty (MU) is a "*non-negative parameter characterizing the dispersion of the quantity values being attributed to a measurand, based on the information used*". In the clinical laboratory the most convenient way to calculate MU is the "top down" approach based on the use of Internal Quality Control data. As indicated in the definition, MU depends on the information used for its calculation and so different estimates of MU can be obtained. The most problematic aspect is how to deal with bias. In fact bias is difficult to detect and quantify and it should be corrected including only the uncertainty derived from this correction. Several approaches to calculate MU starting from Internal Quality Control data are presented. The minimum requirement is to use only the intermediate precision data, provided to include 6 months of results obtained with a commutable quality control material at a concentration close to the clinical decision limit. This approach is the minimal requirement and it is convenient for all those measurands that are especially used for monitoring or where a reference measurement system does not exist and so a reference for calculating the bias is lacking. Other formulas including the uncertainty of the value of the calibrator, including the bias from a commutable certified reference material or from a material specifically prepared for trueness verification, including the bias derived from External Quality Assessment schemes or from historical mean of the laboratory are presented and commented. MU is an important parameter, but a single, agreed upon way to calculate it in a clinical laboratory is not yet available.

## 1. Introduction

In general, the result of a measurement is only an approximation or estimate of the value of the measurand and thus it is complete only when accompanied by a statement of the uncertainty of that estimate (measurement uncertainty (MU)). MU is a "*non-negative parameter characterizing the dispersion of the quantity values being attributed to a measurand*," [1], it defines an interval within which the true value of the measured quantity is expected to lie with a stated level of confidence; it assumes that all significant systematic errors can be identified and corrected within some defined uncertainty so that all uncertainty components can be treated in the same manner (standard deviation). As indicated in its definition, the value of MU is "*based on the information used*".

Analytical variation is not the only source of variation that affects laboratory results, and in their interpretation also biological variation has to be taken into account, but the present paper focalizes only on the analytical variation aspects. ISO 15189:2012 [2] specifically requires that laboratory determine MU for each measurement procedure, so the topic of MU and the debate on how to calculate it has become of interest in the scientific literature [3–9].

An important question is which type of information on method variation should we include into the calculation of MU? A second question, related to the first, is how should we obtain this information?

The scope of the present note is to analyze the problem and to indicate some options on how to reply to the two questions.

## 2. Identification of sources of variation

A (not exhaustive) list of the sources of variation of a measured quantity value that contribute to MU is presented in Table 1.

The classification of the types of errors distinguish between random and systematic errors (bias), both contributing to the MU. Actually, as told before, the theory requires that systematic errors, when detected and significant, should be corrected, including into the calculation only the additional MU introduced with this correction. But if the theory is clear, its application to real life is not so straightforward. In fact, looking at the sources of variation listed in Table 1, only two sources

*Abbreviations:* CRM, certified reference material; EQAS, external quality assessment scheme; IQC, Internal Quality Control; MU, measurement uncertainty; RMS, root mean square

* Corresponding author at: Clinical Laboratory, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Via Francesco Sforza 28, 20122 Milano, Italy.
*E-mail address:* ferruccio.ceriotti@policlinico.mi.it.

https://doi.org/10.1016/j.clinbiochem.2018.03.019
Received 20 January 2018; Received in revised form 26 March 2018; Accepted 28 March 2018
Available online 30 March 2018
0009-9120/ 

**Table 1**
Sources of variation of a measured quantity value.

- Repeatability of the analytical system
- Calibration
  - Value assigned to the calibrator(s) and its uncertainty
  - Frequency of calibration
  - How calibration is performed
- Stability of the reagents on board
- Lot to lot variation
- Frequency of maintenance
- Skill of the operators
- Environmental conditions

can be clearly identified as causing specifically random error (repeatability of the analytical system) or systematic error (value assigned to the calibrator); all the others introduce systematic error, if analyzed under a short timeframe (days), or random variation around the long term mean, if seen over a longer period (weeks or months). If we consider these sources of variation (e.g. reagent lot to lot variation, or calibration related variations) as significant systematic errors we should act for eliminating them as soon as possible, but here we find two types of difficulties: the time needed to identify the presence of a bias (by definition we need replicate measures) and how we can correct them.

## 3. Estimation of MU

To estimate MU it is necessary to quantify the effect of all the possible sources of variation. To this aim the classical bottom-up approach proposed by the GUM [10] requiring the quantification of the variation introduced by each source, is not practical for a clinical laboratory: too many types of tests, difficulty in quantifying specifically the contribution of any single variable e.g. of imprecision of volume dispensing or temperature control. Moreover, a specific effort to quantify the effect of single variables is superfluous because most of the relevant information needed can be gathered through the analysis of the results of the Internal Quality Control (IQC) (the so called "top-down approach").

Abbreviations used:

| | |
|---|---|
| $u$ | "standard uncertainty" and equated to standard deviation. In the current article "$u$" is usually equated to CV. |
| CV | "coefficient of variation" or "relative standard deviation"; calculated as standard deviation/mean expressed as percentage |
| $U$ | "expanded uncertainty" product of standard uncertainty and a coverage factor k = 2 for a 95.4% coverage probability |
| $R_w$ | "intermediate precision" as defined in VIM paragraph 2.22 [1] |
| $u(R_w)$ | random component of standard uncertainty |
| $u(cRef)$ | uncertainty of the target value of the calibrator |
| $u(bias)$ | systematic component of standard uncertainty |

### 3.1. How to estimate MU from the IQC data

The IQC provides enough information to estimate MU provided that these three conditions are fulfilled: 1) the quality control material used is commutable, thus reproducing reliably the behavior of the patient samples; 2) the measurand concentration in the control is at a clinically relevant level and 3) the control material is measured exactly as any patient sample for a sufficiently long period of time. In fact, only if the time period on which the QC results are collected is sufficiently long, all short time variations of the mean (bias) due to lot to lot changes or calibration effects or changings of the analyzer's performances related to the deterioration of some components (photometer lamp, liquid delivery apparatus, washing system etc.) etc., can be included into the calculation, giving their contribution to the overall variation. This information is available at different concentration levels and thus it is possible to evaluate if different uncertainty values are needed in relation to the concentration. It depends upon two considerations:

- The precision and bias profile of the method
- The clinically relevant concentration value (decision limits)

If the CV is relatively constant at different concentrations or if the clinically relevant concentration value is close the value of one of the control materials, just one MU value may be sufficient. If the CV vary considerably with the concentration or if the clinically relevant concentration values are more than one (e.g. lower and upper limits of the reference intervals), more than one uncertainty value should be used.

Minimal requirement (personal opinion, supported also by [3], scenario 1 and [11]):

Just consider intermediate precision ($R_w$) (6 months of IQC data with at least 20 results) of one IQC material (concentration at decision level), multiplied by the coverage factor (k = 2 for a 95.4% coverage probability):

$$u = R_w = CV \quad Rw \times 2 = U$$

This approach can be sufficient for most tests, particularly those without traceability to a defined reference material or reference method, or when mainly used for monitoring. However, this approach has several limitations:

- It includes only the random components of MU of the intermediate precision
- It does not include the sources of uncertainty due to the previous steps of the traceability chain and the contribution of the bias. It may include uncertainty of the value assignment of the calibrator if more than one lot of reagent or calibrator was used during the period of data sampling for IQC.

If this approach is used a note as the following one may be added to the report:

> "*The reported measurement uncertainty was calculated by considering only the component due to the variation observed in the control materials over the last 6 months. Additional components, such as possible bias from reference values, may be provided on request.*"

As told before, the IQC data allow to calculate only the random component of MU, including the one related to short time variation of bias (intermediate precision). The possible bias caused e.g. by an incorrect value assignment to the calibrator is not included in this calculation (or only partly if more than 1 lot of calibrator is used). The problems on how to consider the fraction of MU caused by the presence of bias are of two types: 1. how to detect it; 2. correct for it (adding the uncertainty generated by the correction) or just add it into the calculation.

There are few options to detect the presence of an "absolute" bias (bias from an internationally recognized reference): the participation in category 1 or 2 External Quality Assessment Schemes (EQAS) or the measurement of commutable certified reference materials. EQA programs have been classified by Miller et al. into 6 categories according to their ability to evaluate laboratory performance [12]. Programs in category 1 and 2 are the most desirable because they use commutable materials with target values assigned by reference measurement procedures. Unfortunately, the category 1 or 2 EQAS are available for a limited number of measurands, and commutable certified reference materials, besides being extremely costly, are few in number. A possible alternative is the identification of a "relative" bias by the use of a trueness material provided by the manufacturer of the analytical system. The bias identified in this way is relative to the analytical system in use; however, if the value assignment has been done correctly, respecting the traceability chain, this type of material could be

considered similar to a certified reference material (CRM). Unfortunately for very few, if any, of these materials, the value of MU is available.

Other references that may be used for the calculation of the relative bias are the peer group means of inter-laboratory IQC. The limitation of these data is caused by the possible heterogeneity of results used for the calculation of the mean (different analytical systems, calibrations, lot of reagents but also errors in laboratory classification). For these reasons the uncertainty around the target value is usually quite large. Other relative references are the results of EQAS category 3–5. In this case results are few (usually maximum of 12/year) so statistically weak and require long periods to obtain the information.

As discussed before there are several possible approaches for including the bias component into the MU calculation and different ways of calculation, some examples are the following:

1. add the uncertainty of the value assigned to the calibrator
2. add the bias from a reference material (CRM or trueness material from the manufacturer)
3. add the bias calculated from EQAS or from interlaboratory IQC data
4. add the bias from the previous period (historical mean of the laboratory)

No matter the way of detection, when laboratory staff identifies a significant bias, it is responsible for correcting it. So, if bias has been corrected, it must not be included into MU calculation even though the uncertainty of the correction itself should be included. Therefore, the only portion of bias to be included in MU calculation is the one that cannot be corrected, or it is small enough to be accepted (the uncertainty related to the correction might be greater than the bias itself).

### 3.1.1. Add the uncertainty of the value assigned to the calibrator [COFRAC, fourth approach [13]]

Random component = $u(R_w)$ = six months CV; Bias component = uncertainty of the target value of the calibrator = u(cRef)

Being the uncertainty expressed as expanded uncertainty it has to be divided by 2.

$$u(cRef) = \frac{U_{CAL}}{2}$$

$$\text{expressed as\%}\ \ u(cRef)\% = \frac{u(cRef)}{CAL} \times 100$$

$$\boldsymbol{U = 2 \times \sqrt{u(R_w)^2 + u(cRef)\%^2}}$$

where

$U_{CAL}$ expanded uncertainty of the value assigned to the calibrator
CAL concentration value assigned to the calibrator

Note: $U_{CAL}$ is not always easy to obtain from the manufacturers, and its value is not always reliable. When multiple calibrators are required for assay standardization, the procedure for assessing calibrator uncertainty is even more difficult. In this situation, regression analysis of the calibrator standard curve and the calculation of standard error (standard uncertainty) of the regression will be required.

### 3.1.2. Add the bias from a reference material (CRM or trueness material from the manufacturer) [COFRAC, second approach, modified [13]]

Random component = $u(R_w)$ = six months CV

Bias component = $u(bias)$ = it has two components: the uncertainty of value assigned to the reference ($u(cRef)$) and the amount of bias.

$$u(cRef) = \frac{U_{ref}}{2}\ \text{(if available)}, \qquad bias = \overline{x} - Xref$$

$$u(bias) = \sqrt{\left(\frac{bias}{\sqrt{3}}\right)^2 + u(cRef)^2}$$

$$\boldsymbol{U = 2 \times \sqrt{u(R_w)^2 + u(bias)^2}}$$

where bias/ $\sqrt{3}$ is the standard uncertainty of the absolute bias when taken as a uniform distribution.

Note: u(cRef) usually not available for manufacturers' trueness materials.

### 3.1.3. Add the bias calculated from EQAS or from interlaboratory IQC data [14]

Random component = $u(Rw)$ = six months CV

$$\text{Bias component } u(cRef) = \frac{CV_{group}}{\sqrt{n(group)}}$$

$$bias = RMSbias = \sqrt{\frac{\sum (bias_i)^2}{n}}$$

$$u(bias) = \sqrt{(\text{RMSbias})^2 + u(cRef)^2}$$

$$\boldsymbol{U = 2 \times \sqrt{u(R_w)^2 + u(bias)^2}}$$

where:

$CV_{group}$ is the mean CV of the group of methods to which the laboratory belongs, calculated for the period under evaluation (e.g. six months)
n (group) mean number of the laboratories of that specific group of methods in the same period
RMS root mean square

Note: Root mean square bias derives from several EQAS exercises during six months; the calculation of $u(cRef)$ is based on the most common situation in which the reference value derives from a consensus mean and thus is related to the numerosity of the group. Using this approach there is the risk of overestimating the bias component, in fact $u(Rw)$ already includes some bias effects, moreover the bias has a sign, but adding it as square the sign is lost, finally $u(cRef)$ may be significant in case of small groups.

Add the bias component only if significant, i.e. if $RMS_{bias}$ is greater than the uncertainty of the target value ($u(cRef)$).

### 3.1.4. Add the bias from the previous period (historical mean of the laboratory) [15]

The bias component of intermediate precision is minimized by calculating $u(Rw)$ as weighted average of monthly CV. The calculation of weighted average monthly CV's by the following equation is similar to the “error variance” component of an analysis of variance calculation, where only the individual result deviations from their appropriate monthly means are considered. It excludes any variation which may exist between monthly means:

$$\text{Random component} = u(R_w) = average\ monthlyCV = \sqrt{\frac{(n_A - 1) \times CV_A^2 + (n_B - 1) \times CV_B^2 + ..+(n_i - 1) \times CV_i^2}{(n_A + n_B + ... + n_i) - n_{periods}}}$$

To estimate the uncertainty around the value of the historical mean ($u(cRef)$) the average monthly CV is dived by the square root of the mean number of results used to calculate the monthly CV

$$\text{Bias component } 1 = u(cRef) = \frac{Average\ monthly\ CV}{\sqrt{mean\ num\ of\ monthly\ QC\ results}}$$

The bias component of MU is calculated as the root mean square of the biases of the monthly means from the historical mean

Bias component 2 = *bias*=

$$RMSbias = \sqrt{\frac{\sum (bias_i)^2}{n}}$$

$$u(bias) = \sqrt{(\mathrm{RMSbias})^2 + u(cRef)^2}$$

$$\boldsymbol{U = 2 \times \sqrt{u(Rw)^2 + u(bias)^2}}$$

Note. The bias is calculated as RMS bias of the monthly means from the historical mean of the previous period; it implies an unbiased initial situation. It can be applied to a IQC system running from several months. This process depends on the right selection of periods of constant bias. For some analytes monthly may not be the optimum.

## 4. Discussion

It is clear from the previous comments that the "*information used*" will influence the MU estimate. This raises the question of which is the most appropriate MU value? As well discussed by Jones [6] that describes 4 scenarios, the evaluation of which of the possible values of MU is more appropriate depends essentially on the intended use of the result of the measurement. When comparing two consecutive results of the same patient on a short interval of time (scenario 1) the relevant MU for the comparison is the one introduced by the short-term imprecision of the method; on the contrary, when comparing a result with an internationally defined decision limit, also the absolute bias plays an important role, as well as when applying common reference intervals (scenario 4) [6]. ISO 15189 states at paragraph 5.5.1.4 "Upon request, the laboratory shall make its estimates of measurement uncertainty available to laboratory users." [2]. This does not mean that the laboratory should provide data on MU without asking in what perspective those data will be used [16]. It is the responsibility of the laboratory staff to educate and inform the clinicians that interpret and act on laboratory results on other sources of uncertainty than those related to MU. When a clinician requests the information to judge whether a difference from a previous concentration of the same measurand is significant, it may be wise to deliver the reference change value that combines MU with biological variation [17].

So, a standard way of calculating MU might not be correct, on the contrary providing different MU estimates for the same measurand may be confounding both for the patient and for the clinician, moreover usually the laboratory is not aware of the intended use of the measurement result. So, instead of calculating different MU for the same measurand, the way of estimating MU could be different according to the most common use of the measurand: in case of measurands like lipids, that have internationally defined decision limits, an estimate of MU that includes bias could be more relevant than for e.g. tumor markers, that are essentially used for monitoring and where a MU based only on random components might be better. This is an open discussion and it goes beyond the scope of this paper.

## 5. Conclusions

Deriving the MU values from the IQC data is not an "impossible mission". But it requires some attention and still presents some unsolved questions. The minimum requirements are long term IQC data, on commutable materials at proper concentrations. With this information it is possible to calculate MU values that cover the random component of MU, that usually represents a significant part of the uncertainty, but does not cover the part related to the possible presence of long term systematic error. Significant bias should be corrected for, including only the uncertainty related to the correction. The decision on how to deal with the bias and how to calculate the bias-related component of uncertainty is still under debate, an internationally agreed upon approach presently is not available. Laboratories should be careful to prevent double counting of bias when adding bias components in MU calculations based on long term imprecision.

## References


[1] JCGM 200, 2008 International Vocabulary of Metrology — Basic and General Concepts and Associated Terms (VIM), International Organization for Standardization Geneva, 2008, p. 104. Available from: http://www.bipm.org/utils/common/documents/jcgm/JCGM_200_2008.pdf , Accessed date: 20 January 2018.

[2] ISO 15189:2012, Medical Laboratories - Requirements for Quality and Competence, 3rd ed., ISO, Geneva, Switzerland, 2012.

[3] G.H. White, Basics of estimating measurement uncertainty, Clin. Biochem. Rev. 29 (2008) S53–60.

[4] J.H. Lee, J.H. Choi, J.S. Youn, Y.J. Cha, W. Song, A.J. Park, Comparison between bottom-up and top-down approaches in the estimation of measurement uncertainty, Clin. Chem. Lab. Med. 53 (2015) 1025–1032.

[5] J.O. Westgard, S.A. Westgard, Measuring analytical quality: total analytical error versus measurement uncertainty, Clin. Lab. Med. 37 (1) (Mar 2017) 1–13.

[6] G.R. Dallas Jones, Measurement uncertainty for clinical laboratories - a revision of the concept, Clin. Chem. Lab. Med. 54 (2016) 1303–1307.

[7] J.R. Tate, M. Plebani, Measurement uncertainty - a revised understanding of its calculation and use, Clin. Chem. Lab. Med. 54 (2016) 1277–1279.

[8] W.P. Oosterhuis, H. Bayat, D. Armbruster, A. Coskun, K.P. Freeman, A. Kallner, et al., The use of error and uncertainty methods in the medical laboratory, Clin. Chem. Lab. Med. 56 (2018) 209.

[9] A. Padoan, G. Antonelli, A. Aita, L. Sciacovelli, M. Plebani, An approach for estimating measurement uncertainty in medical laboratories using data from long-term quality control and external quality assessment schemes, Clin. Chem. Lab. Med. 55 (2017) 1696–1701.

[10] Evaluation of measurement data — guide to the expression of uncertainty in measurement, Available from: https://www.bipm.org/utils/common/documents/jcgm/JCGM_100_2008_E.pdf, (2008) , Accessed date: 20 January 2018.

[11] I. Ćelap, I. Vukasović, G. Juričić, A.M. Šimundić, Minimum requirements for the estimation of measurement uncertainty: recommendations of the joint Working group for uncertainty of measurement of the CSMBLM and CCMB, Biochem. Med. (Zagreb) 27 (2017) 030502, http://dx.doi.org/10.11613/BM.2017.030502.

[12] W.G. Miller, G.R.D. Jones, G.L. Horowitz, C. Weykamp, Proficiency testing/external quality assessment: current challenges and future directions, Clin. Chem. 57 (12) (2011) 1670–1680.

[13] COFRAC, Guide technique d'accreditation pour l'evaluation des incertitudes de mesure en biologie medicale SH GTA14 Rev00, Available from: https://www.cofrac.fr/documentation/SH-GTA-14, (2011) , Accessed date: 20 January 2018.

[14] B. Magnusson, T. Näykki, H. Hovind, M. Krysell, Handbook for Calculation of Measurement Uncertainty in Environmental Laboratories, 1st edition, (2003), p. TR537.

[15] D. Brugnoni, R. Guerranti, P. Iandolo, S. Mattioli, C. Ottomano, F. Ceriotti, Measurement uncertainty calculation in clinical laboratories | La stima dell'incertezza delle misure nel laboratorio clinico, Biochim. Clin. 39 (2015) 108–115.

[16] M.B. Whyte, R.P. Vincent, How the routine reporting of laboratory measurement uncertainty might affect clinical decision making in acute and emergency medicine, Emerg. Med. J. 33 (2016) 278–279.

[17] E.K. Harris, T. Yasaka, On the calculation of a "reference change" for comparing two consecutive measurements, Clin. Chem. 29 (1983) 25–30.