Andrea Padoan*, Giorgia Antonelli, Ada Aita, Laura Sciacovelli and Mario Plebani

# An approach for estimating measurement uncertainty in medical laboratories using data from long-term quality control and external quality assessment schemes

## Abstract

**Background:** The present study was prompted by the ISO 15189 requirements that medical laboratories should estimate measurement uncertainty (MU).
**Methods:** The method used to estimate MU included the: a) identification of quantitative tests, b) classification of tests in relation to their clinical purpose, and c) identification of criteria to estimate the different MU components. Imprecision was estimated using long-term internal quality control (IQC) results of the year 2016, while external quality assessment schemes (EQAs) results obtained in the period 2015–2016 were used to estimate bias and bias uncertainty.
**Results:** A total of 263 measurement procedures (MPs) were analyzed. On the basis of test purpose, in 51 MPs imprecision only was used to estimate MU; in the remaining MPs, the bias component was not estimable for 22 MPs because EQAs results did not provide reliable statistics. For a total of 28 MPs, two or more MU values were calculated on the basis of analyte concentration levels. Overall, results showed that uncertainty of bias is a minor factor contributing to MU, the bias component being the most relevant contributor to all the studied sample matrices.
**Conclusions:** The model chosen for MU estimation allowed us to derive a standardized approach for bias calculation, with respect to the fitness-for-purpose of test results. Measurement uncertainty estimation could readily be implemented in medical laboratories as a useful tool in monitoring the analytical quality of test results since they are calculated using a combination of both the long-term imprecision IQC results and bias, on the basis of EQAs results.

**Keywords:** external quality assessment schemes (EQAs); internal quality controls (IQC); ISO 15189; measurement procedures (MPs); measurement uncertainty (MU); medical laboratory accreditation.

**\*Corresponding author: Andrea Padoan,** Department of Medicine – DIMED, University of Padova, Padova, Italy; and Department of Laboratory Medicine, University-Hospital of Padova, via Giustiniani 2, 35128, Padova, Italy, Phone: +390498212801, Fax: +390498211981, E-mail: andrea.padoan@unipd.it
**Giorgia Antonelli, Ada Aita and Mario Plebani:** Department of Medicine – DIMED, University of Padova, Padova, Italy; and Department of Laboratory Medicine, University-Hospital of Padova, Padova, Italy. http://orcid.org/0000-0002-0270-1711 (M. Plebani)
**Laura Sciacovelli:** Department of Laboratory Medicine, University-Hospital of Padova, Padova, Italy

# Introduction

Medical laboratory results are widely recognized as a tool of fundamental importance in making a reliable diagnosis, monitoring diseases and providing information conducive to clinical decision-making. However, medical laboratories should guarantee that their measurement procedures (MPs) results are fit for clinical purposes, and this requirement calls for the long-term monitoring of the quality and reliability of results. Since its inception, the International Standard ISO 15189 for medical laboratory accreditation has called for the calculation of measurement uncertainty (MU) to be included in each MP [1]. Interestingly, because MPs are used to describe the whole measurement process, including the specific analytical procedure, all processes which contribute to uncertainty in the test results should be considered when calculating MU. The international vocabulary of metrology (VIM) has defined MU as a "non-negative quantity that characterizes the dispersion of the values that could reasonably be attributed to the measurand" [2]. For a given test result, MU thus represents the interval associated with a defined probability in which the true result should lie. In addition, this interval should fall within limits which guarantee fitness for the clinical purpose of the tests in question [3, 4]. Measurement uncertainty goals for defining fitness-for-purpose limits may be based on clinical outcome studies, biological variation, state of the art, recommendations from an expert group or professional opinions [5, 6].

The components which require consideration in calculating MU are systematic error (bias) and random errors. Bias is inversely related to the degree of trueness of a measurement, while random error represents imprecision and is defined as the standard deviation of a series of measurements [6]. In contrast to the Guide to the Expression of Uncertainty in Measurement model (GUM) [7], which recommends the identification and elimination of bias at an early stage of the measurement process, the Westgard "total error approach" includes both an imprecision and bias component [8]. As discussed by Farrance et al. [9], either procedure may be used to evaluate MU as ISO 15189 does not specify any particular approach. In either situation, however, the true value of the measurand cannot be known, as uncertainty associated with determining bias, in addition to assay imprecision, provide uncertainty as to the true value. In a similar situation to assay imprecision, true bias cannot be determined: whether determined using a reference method or derived from an external quality assessment scheme (EQAs), there will always be an uncertainty associated with this determination. For example, in certified reference materials (CRM) the measurand concentration is associated with an MU, the value being given with the CRM as the best estimate of the "true" level. In contrast, if bias is referred to the target value derived from EQA results (or a consensus value), it is calculated using the statistics which provide the distribution of results (with or without the use of reference methods or reference materials).

Depending on the test purpose, it might also be important to consider different models for calculating MU [4, 9, 10]. If the interpretation of a specific test result is largely provided by making a comparison with, e.g. 1) previous results from the same patient or 2) a reference interval established in the same laboratory or with the same MP, bias becomes largely irrelevant and MU in this setting concerns only the imprecision of the assay [9]. On the other hand, when the comparison is made with a clinical decision point, the bias should be considered in the MU estimation when the clinical decision limits have been derived from multiple MPs [9, 10].

ISO 15189 allows flexibility for laboratories, stating that the "laboratory shall define the performance requirements for the MU of each MP and regularly review estimates of MU". Furthermore, several guidelines and manuscripts available in the literature propose different theoretical approaches for MU estimation [7, 11–14].

The aim of the present study was to describe an approach for estimating the MU for a large number of tests in medical laboratories and report the results obtained. In particular, the contributions of imprecision, bias and bias uncertainty, related to the overall MU were inspected and compared.

## Materials and methods

The study was performed at the Department of Laboratory Medicine, University-Hospital of Padova in the year 2016.

Initially, the method used for estimating MU included the identification of:

–  quantitative tests, performed using both already approved/validated methods (i.e. CE-IVD) and "in house" validated methods;
–  classification for tests in relation to their main clinical purpose, diagnosis and/or monitoring;
–  criteria for estimating the different components of MU (bias, imprecision, bias uncertainty).

Measurement uncertainty components were calculated by using the formula suggested by NordTest [12] (Equation 3), which includes the: 1) imprecision component, calculated by the long term standard deviation of the internal quality controls (IQC) materials; 2) bias component, expressed as root mean square error of bias, calculated using the last eight EQAs exercises; 3) bias uncertainty, calculated using the standard error statistics of the EQAs. Each of these components has similar statistical properties (i.e. they represent the square roots of the variances of their respective estimators, and have the same MU unit).

$$u_{\text{imp}} = \sqrt{\frac{\sum_{i=1}^{k}(n_i-1)s_{QC_i}^2}{\sum_{i=1}^{k}(n_i-1)}} \tag{1}$$

$$\text{RMS}_{\text{bias}} = \sqrt{\frac{\sum_{i=1}^{m}(\text{bias}_i)^2}{m}} \tag{2}$$

$$U = k \cdot \sqrt{u_{\text{imp}}^2 + \text{RMS}_{\text{bias}}^2 + u_{\text{bias}}^2} \tag{3}$$

For each MP, a careful consideration was made of the degree of association between the measurand concentration levels, imprecision and bias. For those MPs where bias, imprecision or both of them were dependent on concentration levels, two or more MU values were estimated; otherwise, a single MU was calculated for each MP. The long-term IQC results (of the latest 6 months) were used to calculate the imprecision component of uncertainty for each IQC concentration level. In cases in which the imprecision values of each concentration level were similar, the pooled imprecision was calculated by taking the square root of the combination of the variances of each long-term IQC series of results (eq. 1). Equation 1 represents the weighted average of IQCs variance ($s_{QC}^2$), considering the series from $i=1$ to $k$ IQC levels, each one calculated with the corresponding number of determinations ($n_i$); in particular, when the number of determinations $n_i$ (used to calculate each $s_{QC}^2$) are the same for each IQC level, the equation become equal to the arithmetic mean of $s_{QC}^2$. Subsequently, for each MP, the EQA results obtained during the survey period 2015–2016 were used to estimate the bias. Only satisfactory EQAs results were kept, whereas results not complying with

established EQA criteria (i.e. results with z-scores >2 or <−2) were discarded. The approach used was based on feasibility possibilities. In order to calculate the bias component the value obtained with the reference procedure (reference MPs of higher metrological order) was used when available in EQAs, when the results concerning the specific diagnostic system failed to show standardization problems, or when EQAs control material did not present commutability problems. Otherwise, the consensus value related to the used diagnostic system was employed.

The latest eight EQAs were chosen to calculate both $RMS_{bias}$ (eq. 2), and the uncertainty of bias $u_{bias}$. Equation 2 represents the square root of the mean of the squares of bias deviations, where $m$ represents the number of EQAs results used ($m=8$). Regarding the calculation of $u_{bias}$, initially each standard deviation given for the specific EQA result was divided by the square root of the number of participating laboratories; following the average of these quantities was then defined as $u_{bias}$. However, when EQAs used the median or other robust statistical methods to calculate the consensus value, $u_{bias}$ was multiplied by 1.25, according to the method for MU calculation specified in ISO 13528 : 2015 [15].

Finally, the formula in Equation 3 (eq. 3) was used to estimate MU, using a coverage factor equal to 2 (k in eq. 3) that was employed to adjust MU to a specific level of confidence; a factor of 2 gives a confidence level of 95.4% (the true value probability for Z=2).

## Statistical analyses

The non-parametric Kruskal-Wallis test was used to evaluate differences across several groups and p-value were adjusted by Bonferroni's method for multiple comparisons. Stata v 13.1 (Statacorp, College Station, TX, USA) and Microsoft® Excel 2011 for Mac (Microsoft® Corporation, Redmond, WA, USA) software were used for all the statistical analyses.

# Results

The Nordtest, Eurachem, Eurolab and GUM approaches were evaluated for their applicability, considering the data already available in our laboratory [7, 12–14]. The Nordtest approach was chosen for estimating MU, as it has proved to be adequate for most MPs. A total of 263 MPs were studied for MU estimation; Figure 1 summarizes the relative percentages for types of sample matrix included in the study. Of the MPs evaluated, 39.5% were for serum specimens, 35.4% plasma, 9.5% urine, and 11.4% blood; as shown in the Figure 1, overall these cases accounted for >95% of the MP types included in the study. MPs were subdivided on the basis of four different laboratory areas (clinical biochemistry, clinical and molecular biology, coagulation and hematology and diagnostic and clinical immunology).

Table 1 summarizes the results for MU estimation of all the MPs included in the study. The bias component
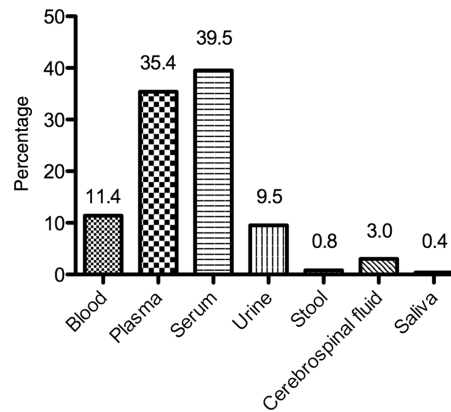


**Figure 1:** Sample types, expressed as percentages of total measurement procedures.

of MU was quantifiable for 190 MPs, but was not calculated in the remaining 22 MPs due to: a) low number of laboratories participating (less than six laboratories, eight MPs) precluded a reliable estimate of the consensus value related to diagnostic system group; b) lack of standard deviations in EQAs reports (two MPs); c) low number of EQAs results obtained during the survey period (12 MPs). Nor was the bias component of MU considered in the MU calculation of 51 MPs because the latter were used mainly for monitoring patients; therefore, only the imprecision component was chosen for MU estimation.

For a total of 28 MPs, two or more MU values were calculated, imprecision and/or bias being dependent on concentration levels. Figure 2 shows the contribution of different components (imprecision, bias, and bias uncertainty) on MU estimation, overall and subdivided according to matrix types. In addition, a more detailed description of some key measurands is also provided in Table 2, to indicate the MU information used in the comparison. Overall, and more generally for all the matrix types evaluated, the results showed that bias uncertainty was a minor contributory factor in MU, although in MPs using stool samples, the magnitude of this component is comparable with that of imprecision and bias. In contrast, the bias component of MU represents the most relevant contributory factor in blood and plasma samples. For cerebrospinal fluid, serum and urine the contribution of bias is not significantly different to the uncertainty contribution related to the imprecision component.

# Discussion

The purpose of medical laboratories is to provide reliable information for improving patient health. This goal not

**Table 1:** Measurement uncertainty (MU) results obtained for all the measurement procedures (MPs) included in the study, subdivided by the different laboratory areas.

| Laboratory areas | MU estimation | | | | Total tests | Type of matrices | MU calculated for different levels |
|---|---|---|---|---|---|---|---|
| | Imprecision | Bias | | | | | |
| | MP used for monitoring patients | MPs with calculable trueness | EQAs with insufficient statistics | | | | |
| | | | Number | Specifications | | | |
| Clinical biochemistry | 51 | 136 | 8 | Trueness not included as the number of participating labs for bias estimation was limited | 195 | Urine (25), blood (5), serum (83), plasma (72), cerebrospinal fluid (8), saliva (1), stool (1) | 16 |
| Clinical and molecular biology | 0 | 1 | 1 | MP with EQAs without AV standard deviation | 2 | Blood (1), stool (1) | 1 |
| Coagulation and hematology | 0 | 33 | 12 | MPs with insufficient number of EQA results | 45 | Plasma (21), blood (24) | 11 |
| Diagnostic and clinical immunology | 0 | 20 | 1 | MPs with EQAs without AV standard deviation | 21 | Serum (21) | 0 |
| Total | 51 | 190 | 22 | | 263 | | 28 |

EQAs, external quality assessment schemes; AV, assigned value.

only entails MPs being fit for their clinical purpose, but also laboratories caring about the quality of MPs results, using routine evaluation of IQC, participating in proficiency testing or EQAs and comparing their own results with those of other laboratories [16]. As MU is an important consideration when interpreting test results, it is a critical aspect of all MPs [9].

In some recently issued guidelines and studies, MU estimation for medical laboratories has been discussed,



**Figure 2:** Contribution to measurement uncertainty (MU) of imprecision ($u_{imp}$), bias root mean square error (RMS$_{bias}$) and bias uncertainty ($u_{bias}$), expressed as percentages.
Contributions to MU are calculated Overall (considering all the MPs with a calculable trueness) or subdividing by matrix types. Bars represent the median percentage values. *p < 0.05.

and the applicability of suggested approaches has been evaluated against tools already available in medical laboratories for monitoring the quality of test results, such as IQC and EQAs results, and focusing on MU components. To this end, advantage was taken of the long-term data collected for laboratories on IQC and EQAs, which provided reliable statistics for MU estimation. In fact, ideally the bias estimation should be performed by comparing results from field methods with reference methods and using patients samples to avoid commutability problems. However, as suggested by Tran et al. [17] this approach in studying bias is not frequently practicable and realistic for medical laboratories; in contrast, the determination of bias based on an EQAs peer group mean (consensus value) could represent a possible and standardized alternative. Based on this premise, the Nordtest approach was chosen for MU estimation as it was suitable for estimating MU for most of the MPs evaluated, although its applicability was limited by the presence of sufficient statistics for both IQC and EQA results. Furthermore, the Nordtest approach represented the actual "dispersion of the values that could reasonably be attributed to the measurand" as it included all the components of MU, namely imprecision, bias and uncertainty of bias.

Imprecision data were available for all the 263 MPs evaluated, whereas the bias component and the bias uncertainty obtained from EQAs could not be estimated for some MPs (8.3%). In particular, for two EQAs, statistics included only the assigned value, while the
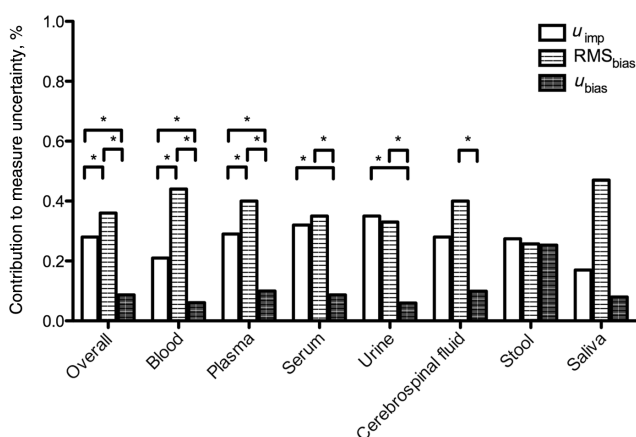
**Table 2:** Illustrative measurement uncertainty (MU) results, obtained from 7 key measurands comparing plasma and urine matrices.

| | Plasma | | | | Urine | | | |
|---|---|---|---|---|---|---|---|---|
| | $u_{imp}$ | $RMS_{bias}$ | $u_{bias}$ | MU | $u_{imp}$ | $RMS_{bias}$ | $u_{bias}$ | MU |
| Sodium, mmol/L | 1.13 | 0.65 | 0.41 | 2.73 | 1.81 | 1.85 | 0.22 | 5.18 |
| Potassium, mmol/L | 0.03 | 0.04 | 0.01 | 0.10 | 1.68 | 1.70 | 0.22 | 4.79 |
| Chlorine, mmol/L | 1.15 | 1.44 | 0.36 | 3.76 | 2.89 | 3.98 | 0.46 | 9.88 |
| Calcium, mmol/L | 0.02 | 0.06 | 0.01 | 0.12 | 0.04 | 0.10 | 0.01 | 0.22 |
| Glucose, mmol/L | 0.09 | 0.11 | 0.03 | 0.29 | 0.16 | 0.29 | 0.02 | 0.66 |
| Urea, mmol/L for plasma and mol/L for urine | 0.20 | 0.50 | 0.06 | 1.08 | 0.01 | 0.01 | 0.01 | 0.03 |
| Magnesium, mmol/L | 0.01 | 0.02 | 0.01 | 0.05 | 0.49 | 0.50 | 0.05 | 1.42 |

Calculations were performed by estimating imprecision ($u_{imp}$ from eq. 1), bias, expressed as the root mean square error of bias ($RMS_{bias}$ from eq. 2) and bias uncertainty. Equation 3 was therefore used for calculating MU (expressed with the same measurement unit as the measurand's concentration).

distribution statistics (standard deviation or interquartile range) were missing. In these and other situations (e.g. where statistical processing of EQAs results does not provide reliable data for specific use or the number of surveys per year is low, or the number of results related to diagnostic system group is low), no calculation can be made for estimating bias components; other approaches are available for this calculation, although they involve more complex procedures. The model proposed here for bias estimation appears to be more a practical solution than a theoretical approach, and further studies are advisable in order to define new strategies for bias calculation, specifically related to theoretical bias references (e.g. previous results, reference intervals and clinical decision points). Furthermore, a possible limitation of this study is that it describes an approach for MU estimation, rather than evaluating the outcome, that will be further investigated.

In conclusion, in this study we provide some usable practical procedures regarding the MU estimation for a series of MPs, routinely used in medical laboratories. In particular, for imprecision component its estimation appears to be a reliable estimation of MU if the correct interpretation of the lab test result is guaranteed on the basis of its clinical purpose. For the bias component, the development of a practical solution for including bias in MU estimation allowed us to derive a standardized approach that considers the source of the bias reference and whether and how bias can be calculated, representing the novelty of this study.

# References

1. ISO 15189: Medical laboratories – particular requirements for quality and competence. Geneva: International Organization for Standardization (ISO), 2012.
2. JCGM 200:2012. International Vocabulary of Metrology – Basic and General Concepts and Associated Terms (VIM 3rd edition). Available at: www.bipm.org. Accessed: Nov 2016.
3. Oosterhuis WP, Theodorsson E. Total error vs. measurement uncertainty: revolution or evolution? Clin Chem Lab Med 2016;54:235–9.
4. Jones GRD. Measurement uncertainty for clinical laboratories – a revision of the concept. Clin Chem Lab Med 2016;54:1303–7.
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. White GH. Basics of estimating measurement uncertainty. Clin Biochem Rev 2008;29:Suppl 1:S53–60.
7. JCGM 100:2008. Evaluation of measurement data – Guide to the expression of uncertainty in measurement, with minor corrections (GUM). Available at: www.bipm.org. Accessed: Sept 2016.
8. Westgard JO. Useful measures and models for analytical quality management in medical laboratories. Clin Chem Lab Med. 2016;54:223–33.
9. Farrance I, Badrick T, Sikaris KA. Uncertainty in measurement and total error – are they so incompatible? Clin Chem Lab Med 2016;54:1309–11.
10. Tate JR, Plebani M. Measurement uncertainty – a revised understanding of its calculation and use. Clin Chem Lab Med 2016;54:1277–9.

11. Matar G, Poggi B, Meley R, Bon C, Chardon L, Chikh K, et al. Uncertainty in measurement for 43 biochemistry, immunoassay, and hemostasis routine analytes evaluated by a method using only external quality assessment data. Clin Chem Lab Med 2015;53:1725–36.

12. Nordic Innovation. Handbook for calculation of measurement uncertainty in environmental laboratories. NT TR 537 – Edition 3.1, 2012. Available at: www.nordtest.info. Accessed: Sept 2016.

13. Guide to the evaluation of measurement uncertainty for quantitative test results. Technical Committee for Quality Assurance in Testing (TCQA). EUROLAB Technical Report No. 1/2006.

14. Quantifying Uncertainty in Analytical measurement. Eurachem/ CITAC Guide CG 4, 3rd ed. Available at: www.eurachem.org. Accessed: Nov 2016.

15. ISO 13528:2015 Statistical methods for use in proficiency testing by interlaboratory comparison. Geneva: International Organization for Standardization (ISO), 2012.

16. James D, Ames D, Lopez B, Still R, Simpson W, Twomey P. External quality assessment: best practice. J Clin Pathol 2014;67:651–5.

17. Tran MT, Hoang K, Greaves RF. Practical application of biological variation and Sigma metrics quality models to evaluate 20 chemistry analytes on the Beckman Coulter AU680. Clin Biochem 2016;49:1259–66.