



ELSEVIER

Contents lists available at ScienceDirect

## Clinical Biochemistry

journal homepage: [www.elsevier.com/locate/clinbiochem](http://www.elsevier.com/locate/clinbiochem)

## Measurement uncertainty in laboratory reports: A tool for improving the interpretation of test results

Andrea Padoan<sup>a,b,\*</sup>, Laura Sciacovelli<sup>b</sup>, Ada Aita<sup>a,b</sup>, Giorgia Antonelli<sup>a,b</sup>, Mario Plebani<sup>a,b</sup><sup>a</sup> Department of Medicine - DIMED, University of Padova, via Giustiniani 2, 35128 Padova, Italy<sup>b</sup> Department of Laboratory Medicine, University-Hospital of Padova, via Giustiniani 2, 35128 Padova, Italy

## ARTICLE INFO

## Keywords:

Measurement uncertainty  
ISO15189:2012  
Laboratory reports  
Reference intervals  
Reference change value  
Clinical decision point

## ABSTRACT

**Background:** Measurement uncertainty (MU) estimation has been introduced by ISO 15189 for the accreditation of clinical laboratories. Although MU reporting is not required, its inclusion in medical reports is of potential assistance to physicians in results interpretation.

**Methods:** MU reporting was evaluated with respect to different test purposes, namely comparison with reference intervals (RI), patient monitoring or comparison with clinical decision limits. Clinical Biochemistry, Hematology, Coagulation and Clinical Immunology measurands were used as examples. Assuming Gaussian RI distribution, the probability of retesting due to MU was determined by simulations. Significant MU variations were compared against the reference change value (RCV) and clinical decision limits.

**Results:** Three potential scenarios emerged for RI. For 12 measurands, depending on the MU interval, a potential change in results interpretation was found only for Sodium and S-Protein. On considering only the results within RI, simulations confirmed that up to 8.6% of MU intervals encompassed the RI limits, thus potentially leading to retesting. For tests used in patient monitoring, significant MU variations were comparable to those calculated by RCV, with the exception of CEA. For tests results evaluated with respect to clinical decision limits, on including MU, the clinical interpretation may be improved (e.g. for tPSA).

**Conclusion:** The findings made in the present study, which considers real MU data and hypothetical results obtained for a series of measurands, support the concept that MU may aid the physician's interpretation thus ensuring reliable clinical decision making.

### 1. Introduction

Medical laboratories should guarantee that laboratory reports contain all the information required for the correct interpretation of tests results. In laboratory medicine the communication of results to physicians, a process included in the post-analytical phase, is of crucial importance, as the interpretation of results is always performed as a comparison. For this purpose, a series of different types of information, usually included in laboratory reports in order to facilitate the interpretation of test results, includes the measurand reference interval (RI), diagnostic cut-offs and decision limits, as well as the reference change value (RCV), where appropriate. However, as stated by the Guide to the expression of Uncertainty in Measurement (JCGM 100:2008) “the result of a measurement is only an approximation or estimate of the value of the measurand and thus is completed only when accompanied by a statement of the uncertainty of that estimate” [1]. In recent years, with the adoption of ISO 15189:2012 for the accreditation of medical laboratories, a requirement made is the estimation of measurement

uncertainty (MU) [2]. Nevertheless, the ISO 15189:2012 does not specify methods for estimating MU, and the inclusion or exclusion of MU in laboratory reports is left to the laboratory's discretion, the ISO 15189:2012 only stating that “Upon request, the laboratory shall make its estimates of measurement uncertainty available to laboratory users” [2]. However, the reporting of MU in medical reports does call for considerations to be made and discussed in advance by laboratory medicine experts, in order to guarantee that physicians are provided with the support needed for the correct interpretation of test results. In fact, the reporting of MU may effect a modification of test result interpretation and of clinical reasoning. Moreover, as recently pointed out, when using MU to assist results interpretation (according to ISO 15189:2012), it is also important to choose the most appropriate available model according to the fit-for-purpose of tests [3–5].

In order to ascertain the appropriateness of including the MU in laboratory medical reports, the modality to use, and to focus on how this information could influence the interpretation of test result, the present paper reports on and evaluates a series of different scenarios

\* Corresponding author at: Department of Medicine – DIMED, University of Padova, via Giustiniani 2, 35128 Padova, Italy.  
E-mail address: [andrea.padoan@unipd.it](mailto:andrea.padoan@unipd.it) (A. Padoan).

<https://doi.org/10.1016/j.clinbiochem.2018.03.009>

Received 24 November 2017; Received in revised form 12 March 2018; Accepted 12 March 2018  
0009-9120/ © 2018 The Canadian Society of Clinical Chemists. Published by Elsevier Inc. All rights reserved.

**Table 1**

Test purposes and uncertainty. Clinically significant components for estimating measurement uncertainty for the different tests purposes.

Test purpose	Examples	Components to include in measurement uncertainty
Primarily for monitoring patients over time	E.g. tumour markers, immunosuppressive drugs	Imprecision only [3,4]
Comparison with reference intervals, either established in the same laboratory or verified in the laboratory using appropriate procedures	E.g. hormones	Imprecision only [3]
Usually compared with a clinical decision point	E.g. glucose, ions	Imprecision, bias and bias uncertainty [3]

regarding the role of MU in test results interpretation, when the attention is focused on RIs, clinical decision points and RCV, respectively, as shown in Table 1. This study requires skills both on essentials and advanced statistical concepts, such as probability distributions, mean and standard deviation, percentiles and the Monte Carlo simulations, and methods for understanding and implementing the proposed approach, but in another paper, a proposal for “what information on measurement uncertainty should be communicated to clinicians, and how” is available as guidance as to how the outcome of this type of study are applicable to clinical laboratories [6].

## 2. Materials and methods

Two assumptions regarding RI were made in this study: a) the RI distributions are assumed to be Gaussian, b) RI, established by either the laboratory or the manufacturer, are carefully verified by the laboratory. In this context, the total variance of the distribution of reference values consists of: biological (intra-individual,  $SD_I^2$ , and inter-individual,  $SD_G^2$ , variance) and analytical ( $SD_A^2$ ) variation and for each individual, both the distribution values around the personal homeostatic set-points, with variance  $SD_I^2$  and the distribution homeostatic set-points of the reference population, with variance  $SD_G^2$ , can be considered normal [7–9]. Based on these considerations, and given a measurand, the variance attributable to the distribution of the reference values ( $SD_T^2$ ), can be calculated by  $SD_T^2 = SD_A^2 + SD_I^2 + SD_G^2$ , and  $SD_T$  can be derived by using RI with the following formula  $SD_T = (URL - LRL)/3.92$ , where URL and LRL are the upper and the lower RI limits, respectively, and 3.92 derived from the z factor 1.96 multiplied by 2, for a RI covering the 95% interval of the reference values [10,11]. The reference population mean can be derived by mean =  $(URL + LRL)/2$ . Therefore, by assuming a) and b), a test result between mean  $\pm 2 * SD_T$  is within its RI [10]. Imprecision was estimated by using the long-term IQC data (of the latest 6 months), while the expanded MU was estimated by applying the Nordtest approach, with further detailed information available in a previous publication [5,12]. Briefly, MU were estimated by including imprecision or a combination of imprecision, bias and bias uncertainty, depending on the fit-for-purpose of test results (Table 1). The weighted average of IQCs variance were used as imprecision, while bias and bias uncertainty were estimated by External Quality Assessment schemes (EQAs) [5]. The approach used was flexible and feasible, based on the data available for each measurement procedure. Measurement bias was estimated using commutable EQA materials where target values were assigned using high order reference materials and/or reference measurement procedures. If such materials were not available, the consensus value related to the specific diagnostic system was used [5].

The Index of Individuality (II) was calculated using the formula  $(CV_A^2 + CV_I^2)^{1/2}/CV_G$  from Petersen et al. [13].

Monte Carlo analyses were performed by simulating random generated normal distribution of  $n = 200$  values. For each measurand considered in the simulation, normally distributed data were generated by using the mean and the total standard deviation, calculated using the above formulas. Finally, for any test result included in the interval between URL and LRL, a total of 1000 iterations were performed to calculate the probability that the corresponding MU interval would

include the upper or lower RI ( $p_{inc}$ ). The Monte Carlo standard error estimate of  $p_{inc}$  was also calculated using the standard deviation of  $p_{inc}$ . For the same measurands used in the Monte Carlo simulations, data from the Laboratory Information System (LIS) of the Department of Laboratory Medicine of the University-Hospital of Padova where obtained, in a time period of 2 months (from November to December 2017). For the following measurands ALT, Sodium, Potassium, Urea, Cholesterol, Iron, Hemoglobin and MCV results from outpatients (non-hospitalized subjects referring directly to the laboratory for blood testing) were included; for Lactate, S-Protein, C-Protein and D-Dimer, also the inpatients (hospitalized patients) were included, due to the limited number of determinations for these measurands in the group of the outpatients. For test results included in the interval between URL and LRL, the real percentage of MU intervals that would include the upper or lower RI were estimated.

Bidirectional, 95% probability reference change value (RCV) was calculated using the formula  $RCV = 2^{1/2} * Z * (CV_A^2 + CV_I^2)^{1/2}$  [14], while the MU critical difference was derived through multiplying the corresponding MU by a factor equal to 2.83, which allows two serial measurements and the two corresponding MU to be taken into account when a test result is used mainly for monitoring patients over time [15].

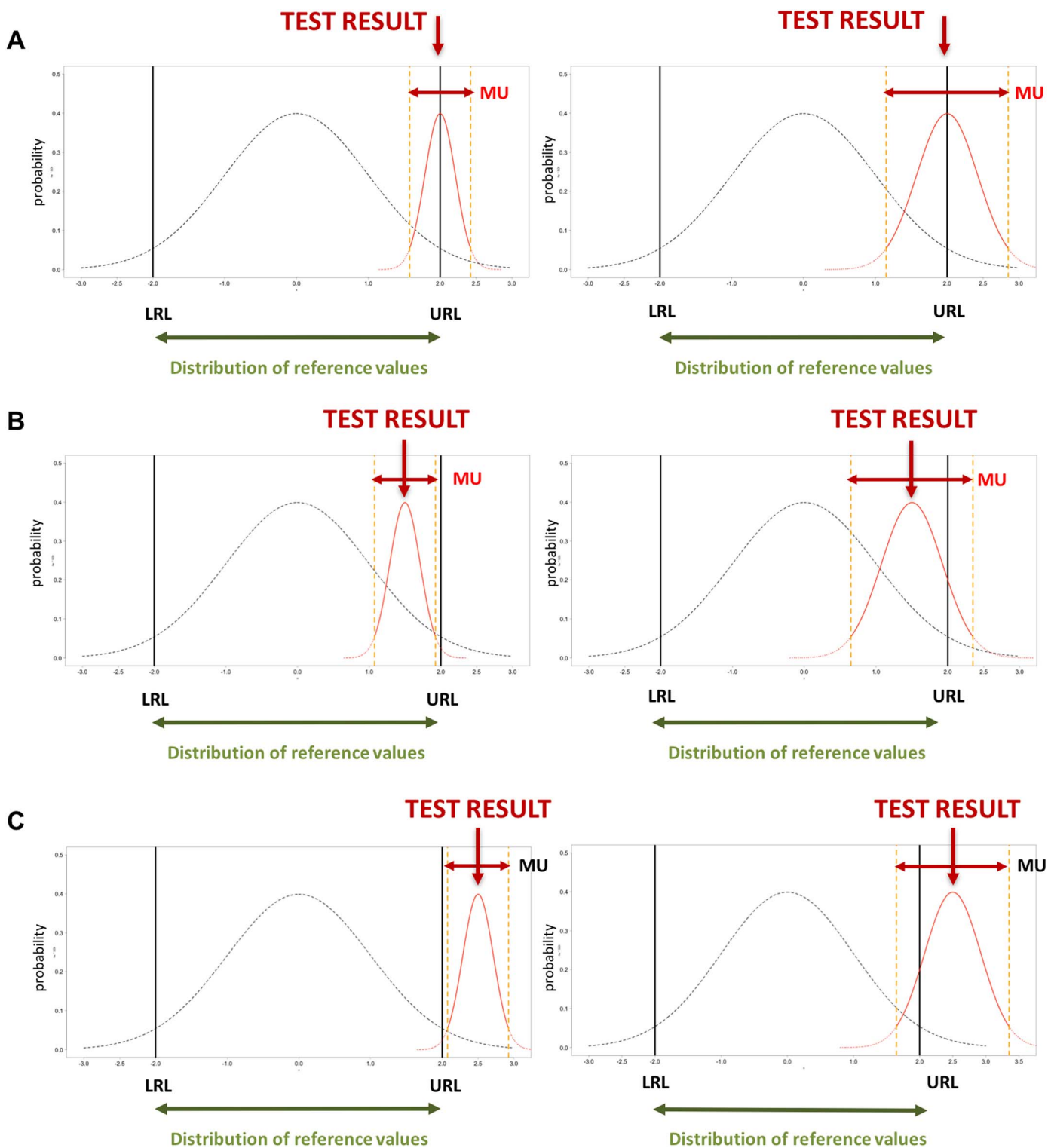
Data for biological within- and between-subject variations were obtained with the Westgard database on biological variation (<https://www.westgard.com/biodatabase1.htm>, accessed on May 2017). R for statistical computing v 3.3.1 software, (Statistical Computing, Vienna, Austria) and Microsoft Excel (Microsoft® Corporation, Redmond, Washington, USA) were used for the statistical analyses.

## 3. Results

### 3.1. Test results compared with established (or verified) RI

On considering hypothetical test results lying around the upper reference limit (URL) of the distribution of reference values (e.g. at values equal to mean plus 1.5, 2 or 2.5 times  $SD_T$ ), different scenarios may emerge when MU was included in the laboratory report (Fig. 1) and considered for test interpretation. The first, more straightforward Scenario 1 regards the situation in which the test result (x) overlaps to URL ( $x = \text{mean} + 2 * SD_T$ ) (Fig. 1, panel A). The MU interval spreads 50% to the left and 50% to the right of test results, irrespective of the MU interval extent and the extent of biological and analytical variations. In this case, if a repeat test is requested on the same sample, the probability of the new result being inside or outside RI should be equal. Other hypothetical scenarios are represented by test results within RI (Scenario 2) or above the URL (Scenario 3). An example of the former case is when a hypothetical test result is at  $x = \text{mean} + 1.5 SD_T$ ; once observed with its MU, the MU interval may include the upper reference interval limit (Fig. 1, panel B). Likewise, also in Scenario 3 the MU interval may include the URL (Fig. 1, panel C). Scenarios 2 and 3 represent the two situations in which the inclusion of MU in the test result might lead to different clinical conclusions. In Scenarios 2 and 3, the contribution of MU to results interpretation became relevant when the ratio of  $SD_A/SD_T$  was elevated.

Table 2 shows data that include biological (expressed as CV) and analytical variation, calculated by internal quality control (IQC), and



**Fig. 1.** Reference intervals limits, expanded measurement uncertainty (MU) and test results interpretation. The same Gaussian distributed reference interval (RI) is shown with its upper and lower limits (URL and LRL, respectively), estimated by the formulas reported in Materials and Methods. Figure reports the three possible different scenarios obtained by the inclusion of MU to laboratory report (panels A, B and C). Panel A represents the situation when a test result is equal (overlapped) to the upper reference limit (URL), Scenario 1. Panel B illustrates a hypothetical test result within RI, Scenario 2. Panel C illustrates a hypothetical test result above the URL, Scenario 3. For each panel, two different MU were superimposed to the distribution of reference values; on the left a smaller MU (width equal to  $0.1 \times SD_T$ ), on the right a larger MU (width equal to  $0.185 \times SD_T$ ) are shown both as means of probability distribution but also as confidence limits (vertical dashed lines).

the MU obtained on evaluated measurands. As shown in Table 1, MU was calculated by using analytical variation alone. An example of this is Sodium, for which the RI was 136 mmol/L to 145 mmol/L: on considering a hypothetical result of 146.2 mmol/L (Scenario 3), the

corresponding interval of results (obtained by using MU) was between 144 mmol/L and 148.5 mmol/L, thus including the URL (145 mmol/L). Likewise, for Sodium values of 143.9 mmol/L (Scenario 2, representing mean +  $1.5 \times SD_T$ ), the results interval is 141.7–146.2 mmol/L, URL

**Table 2**  
Results obtained for a series of 12 illustrative measurands, for the Clinical Biochemistry and Hematology and Coagulation areas. Two hypothetical series of results were considered, the first at mean + 2.5 × SD<sub>r</sub> and the other at mean + 1.5 × SD<sub>r</sub> (B), with their corresponding MU intervals, which are compared to corresponding RI. Data for biological variation and the individuality index (II) are also included.

Measurands	Westgard BV <sup>a</sup>		IOC data		Index of Individuality		Reference Interval (RI) limits		Total standard deviation <sup>b</sup>	SD <sub>r</sub>	SD <sub>A</sub> /SD <sub>T</sub>		MU <sup>c</sup>	Hypotheticals result		MU interval		Hypotheticals result		MU interval		Upper limit	
	CV <sub>i</sub>	CV <sub>G</sub>	CV <sub>A</sub>	SD <sub>A</sub>	CV <sub>A</sub>	SD <sub>A</sub>	Lower	Upper			Lower limit	Upper limit		2.5 SD <sub>r</sub>	1.5 SD <sub>r</sub>	Lower limit	Upper limit	Lower limit	Upper limit	Lower limit	Upper limit		
ALT (U/L)	19.4	41.6	4.4	1.7	0.48	10	50	10.2	10.2	0.17	3.4	55.5	52.1	58.9	45.3	41.9	48.7	41.9	48.7				
Lactate (mmol/L)	27.2	16.7	2.4	0.04	1.64	0.5	2.2	0.4	0.4	0.09	0.10	2.4	2.4	2.5	2.4	1.9	2.1	1.9	2.1				
Sodium (mmol/L)	0.6	0.7	0.9	1.13	1.55	136	145	2.3	2.3	0.49	2.26	146.2	144.0	148.5	143.9	141.7	146.2	141.7	146.2				
Potassium (mmol/L)	4.6	5.6	0.7	0.03	0.83	3.4	4.5	0.3	0.3	0.11	0.06	4.7	4.6	4.7	4.4	4.3	4.4	4.3	4.4				
Urea (mmol/L)	12.1	18.7	2.3	0.12	0.66	2.5	7.5	1.3	1.3	0.09	0.24	8.2	7.9	8.4	6.9	6.7	7.2	6.7	7.2				
Cholesterol (mmol/L)	5.95	15.3	1.3	0.06	0.40	2	6.19	1.1	1.1	0.06	0.12	6.8	6.6	6.9	5.7	5.6	5.8	5.6	5.8				
Iron (μmol/L)	26.5	23.2	1.22	0.2	1.14	11.6	31.3	5.0	5.0	0.04	0.4	34.0	33.6	34.4	29.0	28.6	29.4	28.6	29.4				
Hemoglobin (g/L)	2.85	6.8	0.82	0.14	0.44	140	175	8.9	8.9	0.02	0.28	179.8	179.5	180.1	170.9	170.6	171.2	170.6	171.2				
MCV (fl)	1.4	4.85	0.7	0.55	0.32	80	96	4.1	4.1	0.13	1.1	98.2	97.1	99.3	94.1	93.0	95.2	93.0	95.2				
S-Protein (%)	5.8	63.4	2.6	5.5	0.10	74	146	18.4	18.4	0.30	11	155.9	144.9	166.9	137.6	126.6	148.6	126.6	148.6				
C-Protein (%)	5.6	55.2	2.9	0.77	0.11	70	140	17.9	17.9	0.04	1.54	149.6	148.1	151.2	131.8	130.2	133.3	130.2	133.3				
D-Dimer (μg/L)	23.3	26.5	6.25	16.8	0.91	0	400	102.0	102.0	0.16	33.6	455.1	421.5	488.7	353.1	319.5	386.7	319.5	386.7				

BV = biological variation; CV<sub>i</sub> = intra-subject biological variation; CV<sub>G</sub> = inter-subject biological variation; CV<sub>A</sub> = analytical variation calculated using IOC available data. TV was calculated as suggested by Haeckel et al. [9].  
<sup>a</sup> Derived from Westgard online database (<https://www.westgard.com/biodatabase1.htm>, accessed on May 2017).  
<sup>b</sup> Derived from RIs as described in Materials and methods.  
<sup>c</sup> Calculated using only SD<sub>A</sub>.

again included. On considering the 12 measurands shown in Table 2, a similar situation was found only for S-Protein. The highest SD<sub>A</sub>/SD<sub>T</sub> ratios were those for Sodium and for S-Protein (Table 2). In order to test the hypothesis that the SD<sub>A</sub>/SD<sub>T</sub> ratio was associated with the probability of having one RI included in the MU interval, a series of Monte Carlo simulations were performed. In this iterated analysis, the probability that the MU intervals included the URL (p<sub>inc</sub>) was calculated only for tests results between URL and the lower reference limit (LRL); the results, reported in Table 3, show that the ratios SD<sub>A</sub>/SD<sub>T</sub> were closely correlated with p<sub>inc</sub> (Spearman r = 0.995, p < 0.01), these probabilities ranging from 8.61% of Sodium to 0.18% of Hemoglobin. However, the Individuality Index (II) was not associated with p<sub>inc</sub>. These simulation results undertaken the assumptions that the within- and the between-subjects variance are normally distributed; in order to evaluate the transferability of these findings into real life, a large series of data were obtained from the Laboratory Information System (LIS) for the same measurands evaluated by simulations. For each measurand, the real percentage of MU intervals that would include the upper or lower RI were calculated and reported in Table 3. By carefully evaluating Q-Q plots, especially for ALT, Lactate, Urea, Iron, MCV, S-Protein, C-Protein and D-dimer, a skewed data distribution was found with respect to the normal distributed data of Na, K, Cholesterol, Hemoglobin.

A final consideration to make is that all the results obtained for the URL will be identical to those to be obtained on considering the lower limit of RIs, due to assumption of RI distribution symmetry.

### 3.2. Tests results compared with a previous result in patient monitoring

Table 4 shows examples of measurands with their significant variations, calculated by using both RCV and MU, with respect to their decision limits. With the exception of CEA, significant MU differences were lower (albeit for some measurands the difference was slight) than those obtained for significant RCV differences, for all the studied measurands.

### 3.3. Tests results compared with a diagnostic decision level

Table 5 reports examples of some measurands with their significant variation calculated with respect to the diagnostic decision level. For tPSA, the calculated MU was 1 μg/mL; negative results are therefore those ≤ 3 μg/mL. On the contrary, a tPSA value of > 3 μg/mL possibly becomes positive if interpreted by considering MU. Analogously, a value of 6.8 mmol/L of glucose will possibly be positive when MU is included with the test result.

## 4. Discussion

Medical laboratory reports contain laboratory test results and the information necessary to enable requesting physicians to make the correct interpretation; this, in turn, is of crucial importance in reliable clinical decision making and providing effective patient care. Therefore, it is of paramount importance that the information provided is clear, concise and fully interpretative, as the format, the content and the communication strategies of laboratory reports significantly affect the interpretation and utilization of laboratory data by clinicians [16].

ISO 15189:2012 accreditation introduces the estimation of MU for the measurement procedure, which provides a quantitative overview of the quality of a test result [17]. ISO 15189:2012 states that MU should be readily available, while laboratories are not obliged to include MU in their reports. Yet, at least in some specific clinical settings, MU might improve clinical decision making, contributing to ameliorating patient care [17]. Including MU in a test result should prompt the clinician to shift focus from the actual numerical value (the result) to a surrounding interval of probable results as, for simplicity, MU could be viewed as the confidence that can be placed in a test result. From this viewpoint, it is important to bear in mind that the importance MU calculation may



**Table 3**

Monte Carlo simulation results and, in comparison, real data results calculated by using the Laboratory Information System, for the measurands of the Clinical Biochemistry and Hematology and Coagulation areas included in the study.

	SD <sub>A</sub>	SD <sub>T</sub>	SD <sub>A</sub> /SD <sub>T</sub>	RIs	Monte Carlo Simulation Results		Real data results	
					Probability that MU includes RIs (P <sub>inc</sub> ) (%)	SE <sup>a</sup> of P <sub>inc</sub> (%)	Total number of observations	Percentage of MU that includes RIs (%)
ALT (U/L)	1.7	10.2	0.17	10–50	2.51	0.05	526	2.88
Lactate (mmol/L)	0.04	0.4	0.09	0.5–2.2	1.32	0.04	669	3.51
Sodium (mmol/L)	1.13	2.3	0.49	136–145	8.61	0.09	2540	8.80
Potassium (mmol/L)	0.03	0.3	0.11	3.4–4.5	1.32	0.03	2808	1.6
Urea (mmol/L)	0.12	1.3	0.09	2.5–7.5	1.20	0.03	1919	0.94
Cholesterol (mmol/L)	0.06	1.1	0.06	2–6.19	0.65	0.02	3455	0.75
Iron (μmol/L)	0.2	5.0	0.04	11.6–31.3	0.47	0.02	1617	0.42
Hemoglobin (g/L)	0.14	8.9	0.02	140–175	0.18	0.01	7896	0.22
MCV (fL)	0.55	4.1	0.13	80–96	1.89	0.04	7896	1.63
S-Protein (%)	5.5	18.4	0.30	76–146	5.76	0.07	356	12.9
C-Protein (%)	0.77	17.9	0.04	70–140	0.51	0.02	338	1.2
D-Dimer (μg/L)	6.25	102	0.16	0–400	2.47	0.05	3905	1.5

For a series of illustrative measurands, reference intervals (RIs), analytical (SD<sub>A</sub>), total (SD<sub>T</sub>) variation and the ratio between SD<sub>A</sub> and SD<sub>T</sub> are reported. For each measurand, P<sub>inc</sub> is the probability that any “normal test result” (falling inside RIs) will have MU intervals that include the upper (or lower) RIs limit (P<sub>inc</sub>). This is the probability of obtaining a particular test result in which its MU includes one (not both) of the RIs limits.

<sup>a</sup> P<sub>inc</sub> standard errors (SE) are derived from the distribution of Monte Carlo results. Monte Carlo runs = 10,000. Real data were obtained by the Laboratory Information system of the Department of Laboratory Medicine of the University-Hospital of Padova in a time period from November to December 2017.

**Table 4**

Significant variations calculated for six illustrative measurands, considering both reference change value (RCV) and measurement uncertainty (MU) with respect to a hypothetical result at clinical decision cut-off.

Measurands	Westgard BV <sup>a</sup>		IQC data	Index of Individuality (II)	MU <sup>b</sup>	Hypothetical result at decision limit		Significant variation based on:	
	CV <sub>I</sub>	CV <sub>G</sub>				CV <sub>A</sub>	RCV <sup>b</sup>	MU <sup>b</sup>	
CEA (μg/L)	12.7	55.6	3.7	0.24	1.4	5.0	6.8	7.0	
CA 15–3 (kU/L)	6.1	62.9	3.8	0.11	3.2	37.5	45.0	42.0	
CA 125 (kU/L)	27.4	54.6	4.5	0.51	9.2	48.0	84.9	61.0	
CA 19–9 (kU/L)	16.0	130.5	8.7	0.14	5.2	37.0	55.7	44.3	
AST (U/L)	12.3	23.1	2.5	0.54	11.0	45	60.7	60.5	
Creatinine (umol/l)	5.95	14.7	1.2	0.41	4.0	104	121.3	109.7	

BV = biological variation; CV<sub>I</sub> = intra-subject biological variation; CV<sub>G</sub> = inter-subject biological variation; CV<sub>A</sub> = analytical variation calculated using available IQC data, expressed as percentage.

<sup>a</sup> Derived from the Westgard online database.

<sup>b</sup> Calculated as described in [Materials and methods](#).

**Table 5**

Significant limits for MU intervals of four illustrative measurands, with respect to their corresponding decision limits.

Measurand	IQC data	MU <sup>a</sup>	Decision limit	Significant limit for MU	
				Cut-off	Lower limit
Glucose (mmol/L)	1.5	0.4	7	6.6	7.4
tPSA (μg/L)	3.7	1.0	4	3.0	5.0
HbA1c (mmol/mol)	1.7	3.6	48	44.4	51.6
Troponin I (ng/L)	7.0	2.7	16	13.3	18.7

CV<sub>A</sub> = analytical variation calculated using IQC data; HbA1c = glycated hemoglobin; tPSA = total prostate specific antigen level.

<sup>a</sup> Calculated as described in [Materials and methods](#).

differ depending upon the primary clinical application of the test result, as recently stated in the literature, and summarized in [Table 1](#) [3,4,17]. This approach consists on the inclusion of imprecision, or a combination of imprecision, bias and bias uncertainty in MU. However, it should be noted that reagents lot-to-lot variability can deliver bias (that will possibly appear in IQC and external quality assessment data, shifting the reference points during time), possibly delivering a further source of error, especially considering the aim of improving the overall decision making based on sound analytical and metrological principles. In this study, for estimating MU we used the approach we have recently used

to estimate measurement uncertainty in our laboratory to achieve the accreditation according to ISO 15189:2012 [5]. This approach attracted many criticisms from ISO 15189 accreditors, especially because JCGM 100:2008 stated that for MU steps are to be taken to minimize bias, e.g. by recalibration [1]. However, the discussion on bias remains still open and Theodorson E. in a recent work stated that “Among myths about measurement uncertainty methods is that they demand that bias be eliminated before uncertainty calculations and estimated can be made. Uncertainty approaches claim that bias should be eliminated when identified. However, if bias cannot be eliminated, or when bias elimination risks increasing the overall measurement uncertainty, this can be handled as any other type B uncertainty” [19]. Further, despite bias should be ideally estimated relative to the reference material or reference measurement procedure that was used through a calibration hierarchy to assign the value of the end-user calibrator of the routine measurement procedure being considered, this approach can be currently applied only to few measurands. Therefore, in this study, in the absence of reference materials or reference measurement procedures, the diagnostic system was used, thus allowing to reduce the variability due to a mix of different measurement methods and procedures [5].

The possible advantages or disadvantages of including MU in medical laboratory reports have not yet been adequately discussed, although the potential utility of MU in clinical decision-making is advocated in ISO 15189:2012 as well as other manuscripts available in literature [2,17–19]. The present study takes into account some real

MU data, in addition to a series of hypothetical results for some common measurands, and evaluates the differences in clinical interpretations given to the combination of test results and their MU values.

Firstly, we considered the interpretation of a test result with respect to a reference interval that had been either established or verified by the laboratory; in this case, MU estimation should include only imprecision (Table 1). Three scenarios can be identified (Fig. 1). In theory, Scenario 2 may include situations in which, although the test result falls within the RI range, the MU interval could include the URL (or LRL). In this situation, based on the inclusion of the MU interval, the clinical interpretation of test results should take the patient's clinical data into account. As shown in Table 2, sodium (but not potassium), could be involved in situations like Scenario 2, and a sodium value of 144 mmol/L, with its MU, could encompass (and exceed) the upper reference limit (145 mmol/L). Although the MU interval does not include critical values for sodium, patients might be considered alternatively in a "normal" or in borderline hypernatraemia conditions, depending on the interpretation given to MU. However, to assure patient safety, particularly in critical circumstances, clinicians may decide to correct sodium levels on the basis on an initial hypernatremia. Interestingly, if a test result and its MU value are not interpreted in the right clinical setting, Scenario 2 might induce clinicians to request the re-testing of a specimen. Table 3 reports, for the series of measurands evaluated, the probability that a "normal" subject, with a test result falling within RI, will have MU that encompass at least one of the RIs limits. This probability ( $p_{inc}$ ) specifically indicates that a test result may be retested due to misinterpretation arising from the inclusion of the MU interval in the laboratory report; this is extremely unlikely for several measurand, especially those with a low  $SD_A/SD_T$  ratio.  $p_{inc}$  values were then compared with the real percentage calculated on a series of laboratory results retrieved from the LIS and, irrespective on data distribution, results for the majority of measurands were overlapping.

In addition, another aspect that may be consider regards the little utility of using RI for that measurands with  $II$  is  $< 0.6$  (or  $> 1.4$ ). This fact should not limit the clinical utility of MU as we found that  $p_{inc}$  is not associated to  $II$ . However, as stated by ISO 15189:2012, the MU value, calculated for each quantitative measurement procedure, is considered of practical utility to clinicians when the interpretation of test results is used as "meaningful comparison of a patient value with a previous value of the same type or with a clinical decision value" [2]. Therefore, the limitations given by measurands individuality on RI does not influence MU interpretation.

The case is different when test results are used for monitoring disease, the interpretation being made mainly by comparison with a previous result. It is also important to bear in mind that in this situation imprecision alone should be included in the MU estimation because, their bias being similar for both measurements will be cancelled. Table 4, which reports data for some measurand commonly used for monitoring the patient's disease status, shows that the significant RCV variation is generally wider than that of MU. Therefore, in these situations (e.g. for tumour markers), MU is not of clinical utility, the RCV being the most appropriate value to be considered. Importantly, therapeutic drug monitoring (TDM) is a completely different case, and MU (not RCV) should be considered for determining differences in serial measurements.

Where a test result is compared with a clinical decision point, the MU calculation should include bias (Table 1). Table 5 lists some measurands in this category, such as glucose, cardiac troponin, tPSA and HbA1c. For these measurands, the interpretation of patients' results is mainly performed by comparison with respect to a defined clinical decision point. At least for some of these measurands, such as Glucose and tPSA, the clinical interpretation of test results could differ if MU is included in laboratory reports. For example, the diagnostic cut-off of fasting plasma glucose (FPG) for diabetes mellitus diagnosis is 7.0 mmol/L; the lower limits obtained by MU (0.4 mmol/L) for a value of 7.0 mmol/L become 6.6 mmol/L. Therefore, the clinician may

consider an FPG result ranging between 6.6 mmol/L and 7.0 mmol/L as an initial alert suggesting diabetes. Yet there are other ways of enabling a diagnosis of diabetes, such as 1) repeating FPG after a few weeks, 2) confirming diabetes by either evaluating the HbA1c result or requesting a 75 g oral glucose tolerance test [20]. Differently, patients with elevated tPSA (i.e.  $> 4$  ng/mL) and/or suspicious signs at digital rectal examination are eligible for prostate biopsy, especially if elderly [21]. Table 4 shows that the tPSA MU is 1 ng/mL. Therefore, with a value of tPSA  $> 3$  ng/mL, the probability of having a real value above 4 is not null and, according to the European Association of Urology (EAU) guidelines, the clinician can consider the opportunity of requesting a verification of PSA levels after a few weeks to exclude persistently elevated tPSA, which calls for prostate biopsy, thus ensuring that the diagnostic workup is more reliable [21].

In conclusion, the present paper bears out that the reporting of MU in medical laboratory reports can improve the clinical decision making. The results obtained, which suggest how to calculate MU in different situations, and for a series of different measurands, confirm the concept that physicians can adjust their decisional level by giving a more accurate interpretation of MU of a test, in order to facilitate clinical decision-making and maximise patient safety. The reduction in analytical variability should be emphasised, because the ratio  $SD_A/SD_T$  seems to represent a measure of how MU can provide "weight" in the frequency of interpretation of test results related to MU inclusion in laboratory reports.

Finally, this study has some drawbacks, such as the assumption of normality of distribution of RIs, and the limited number of measurands considered. In addition, data on biological variation used this study may indeed be flawed, as recently underlined by findings of the EFLM Working group on Biological Variation, which are delivering an updated estimated of these values. Therefore, other insights should be considered to support the results reported in this study, including new data generated on biological variation which may provide a more evidence-based information [22]. Results obtained in Scenario 1–3 may be replicated to other measurands, namely those with a log-normal distributed data [10].

## Acknowledgements

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit/non-profit sectors.

## References

- [1] JCGM 100:2008, Evaluation of Measurement Data - Guide to the Expression of Uncertainty in Measurement (GUM), Available online [www.bipm.org](http://www.bipm.org), Accessed date: 2 October 2017.
- [2] ISO 15189, Medical Laboratories – Particular Requirements for Quality and Competence, International Organization for Standardization (ISO), 2012.
- [3] G.R.D. Jones, Measurement uncertainty for clinical laboratories - a revision of the concept, *Clin. Chem. Lab. Med.* 54 (2016) 1303–1307.
- [4] J.R. Tate, M. Plebani, Measurement uncertainty - a revised understanding of its calculation and use, *Clin. Chem. Lab. Med.* 54 (2016) 1277–1279.
- [5] 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.
- [6] M. Plebani, L. Sciacovelli, D. Bernardi, A. Aita, G. Antonelli, A. Padoan, What information on measurement uncertainty should be communicated to clinicians, and how? *Clin. Biochem.* (17) (2018) 31180–31183 pii: S0009-9120 <https://doi.org/10.1016/j.clinbiochem.2018.01.017>.
- [7] E.K. Harris, Effects of intra- and interindividual variation on the appropriate use of normal ranges, *Clin. Chem.* 20 (1974) 1535–1542.
- [8] E.K. Harris, Some theory of reference values. II. Comparison of some statistical models of intraindividual variation in blood constituents, *Clin. Chem.* 22 (1976) 1343–1350.
- [9] P.H. Petersen, S. Sandberg, C.G. Fraser, H. Goldschmidt, Influence of index of individuality on false positives in repeated sampling from healthy individuals, *Clin. Chem. Lab. Med.* 39 (2001) 160–165.
- [10] R. Haeckel, W. Wosniok, E. Gurr, B. Peil, Permissible limits for uncertainty of measurement in laboratory medicine, *Clin. Chem. Lab. Med.* 53 (2015) 1161–1171.
- [11] R. Haeckel, W. Wosniok, A new concept to derive permissible limits for analytical

- imprecision and bias considering diagnostic requirements and technical state-of-the-art, *Clin. Chem. Lab. Med.* 49 (2011) 623–635.
- [12] Nordic Innovation, Handbook for Calculation of Measurement Uncertainty in Environmental Laboratories. NT TR 537 – Edition 3.1, Available at [www.nordtest.info](http://www.nordtest.info), (2012), Accessed date: October 2017.
- [13] P.H. Petersen, C.G. Fraser, S. Sandberg, H. Goldschmidt, The index of individuality is often a misinterpreted quantity characteristic, *Clin. Chem. Lab. Med.* 37 (1999) 655–661.
- [14] C.G. Fraser, Reference change values, *Clin. Chem. Lab. Med.* 50 (2012) 807–812.
- [15] Clinical Laboratory Standards Institute (CLSI), Expression of Measurement Uncertainty in Laboratory Medicine, CLSI EP29-A, Wayne, PA, USA, 2012.
- [16] M. Plebani, Errors in clinical laboratories or errors in laboratory medicine? *Clin. Chem. Lab. Med.* 44 (2006) 750–759.
- [17] G.H. White, I. Farrance, AACB uncertainty of measurement working group, uncertainty of measurement in quantitative medical testing: a laboratory implementation guide, *Clin. Biochem. Rev.* 25 (2004) S1–24.
- [18] G.H. White, Basics of estimating measurement uncertainty, *Clin. Biochem. Rev.* 29 (Suppl. 1) (2008) S53–60.
- [19] E. Theodorsson, Uncertainty in measurement and total error: tools for coping with diagnostic uncertainty, *Clin. Lab. Med.* 37 (2017) 15–34, <http://dx.doi.org/10.1016/j.cll.2016.09.002>.
- [20] American Diabetes Association, 2 Classification and diagnosis of diabetes, *Diabetes Care* 40 (2017) S11–S24.
- [21] N. Mottet, J. Bellmunt, M. Bolla, E. Briers, M.G. Cumberbatch, M. De Santis, et al., EAU-ESTRO-SIOG guidelines on prostate cancer. Part 1: screening, diagnosis, and local treatment with curative intent, *Eur. Urol.* 71 (2017) 618–629.
- [22] A.K. Aarsand, T.R. Raas, P. Fernández-Calle, C. Ricós, J. Díaz-Garzón, N. Jonker, et al., The biological variation data critical appraisal checklist: a standard for evaluating studies on biological variation, *Clin. Chem.* 64 (2018) 501–514.