Disregarded Measurement Uncertainty Contributions and Their Magnitude in Measuring Plasma Glucose

Astrid Petersmann, MD[1,2] (iD)**, Rainer Macdonald, PhD[3]** (iD)**, and Matthias Nauck, MD[1,4]**

Abstract

Background: Each measurement is subject to measurement uncertainty (MU). Consequently, each measurement of plasma glucose concentration used for diagnosis and monitoring of diabetes mellitus (DM) is affected. Although concepts and methods of MU are well established in many fields of science and technology, they are presently only incompletely implemented by medical laboratories, neglecting MU of target values of internal quality control (IQC) materials.

Methods: An empirical and practical approach for the estimation of MU based on the analysis of routine IQC using control samples with assigned target values is presented. Its feasibility is demonstrated exemplarily by analyzing IQC data from one year obtained for glucose employing the hexokinase method with IQC of two different concentrations.

Results: Combined relative extended ($k = 2$) MU comprising bias, coefficient of variation (CV), and MU of the target values assigned to control materials were about 9% with a lower (~ 56 mg/dL; ~3.1 mmol/L) and 8% with a higher (~ 346 mg/dL; ~19.2 mmol/L) concentration sample, analyzing IQC of one year from three different devices.

Conclusions: Estimation of MU in this study is quite reliable due to the large number of IQC data from one year. The MU of the target values of the commercial control material in this study was considerably larger than other MU contributions, ie, standard deviation and bias. In the future, the contribution of MU of commercial IQC should be addressed more carefully and technologies to measure glucose should be geared toward smaller MU possible, as needed, especially for glucose concentration measurements in diagnosis and management of DM.

Keywords

analytical measurements, bias, imprecision, metrological controls, quality control, uncertainty of measurement

Introduction

Plasma glucose concentration measurements are essential in diagnosis and management of diabetes mellitus (DM). Consequently, requirements for analytical quality need to be adequate for this purpose. For diagnosis of DM, the measured plasma glucose is compared to fixed cutoffs that apply all over the world.[1,2] Under ideal circumstances, a patient sample would receive the same result eg, for plasma glucose, regardless of the medical laboratory used. This is a prerequisite to use commonly defined guidelines and recommendations, containing well-defined and globally applied cutoff values. In reality, however, the results will differ and the true value[3] of a measurand like the plasma glucose concentration cannot be determined, regardless of the method used or the effort spent. Depending on the country, a deviation of about $\pm 10\%$ to $\pm 15\%$ from the target value is allowed.[4,5] Applied to a patient sample sent to a randomly

chosen medical laboratory, results may vary between 113.5 mg/dL (6.3 mmol/L) and 138.7 mg/dL (7.7 mmol/L) if the "true" value of the sample was 126.1 mg/dL (7.0 mmol/L). Some countries allow even larger deviation in their external quality control schemes of up to 15%.

[1]Institute of Clinical Chemistry and Laboratory Medicine, University Medicine Greifswald, Greifswald, Germany
[2]Institute of Clinical Chemistry and Laboratory Medicine, University Oldenburg, Oldenburg, Germany
[3]Physikalisch-Technische Bundesanstalt (PTB), Berlin, Germany
[4]German Center for Cardiovascular Research (DZHK), partner site Greifswald, Germany

Corresponding Author:
Rainer Macdonald, PhD, Physikalisch-Technische Bundesanstalt, Abbestrasse 2-12, Berlin 10587, Germany.
Email: Rainer.Macdonald@ptb.de

As a quantitative indication of closeness of a measurement result to the true value, the concept of measurement uncertainty (MU)[6,7] was developed and has been accepted widely during recent years. To demonstrate the current variability and consequences of MU of plasma glucose measurements, it is helpful to look at data from external quality assurance programs.[8,9] Despite all efforts to harmonize results, eg, by using reference methods for target value assignments in external quality control schemes, target values for internal quality controls (IQCs) are generally not assigned using reference methods. Reference methods are highly accurate measurement procedures with a clearly smaller MU compared to routine methods for a given measurand.[3] Poor target value assignment for IQCs is, at present, rarely considered a contribution to MU.

Despite the unavoidable presence of MU, technologies to measure glucose should be geared toward smaller MU, as needed. In addition, the complete MU needs to be considered in order to provide sufficient care for DM patients. In this article, we propose an empirical approach for the estimation of MU and its different components based on frequent (daily) IQC data. In addition to commonly considered contributions of bias and imprecision, we include the MU of commercial control samples, called target value uncertainty (TVU) in the following, as an important input to the overall MU.

Methods

The basic considerations and concepts of MU have been laid down in the Guide to the Expression of Uncertainty in Measurement (GUM)[6,7] of the International Organisation for Standadisation and the International Electrotechnical Commission (ISO/IEC). For details of GUM and its application to laboratory medicine, we refer the interested reader to other sources.[8-13]

The GUM allows a variety of methods to estimate MU, which usually comprises several components. In brief, some of them may be determined by a top-down or so-called Type-A method, ie, investigation of the statistical distribution and variability of results with the help of repeated measurements under repeating conditions. The MU is given by the standard deviation in this case and is often called the standard measurement uncertainty. This may be used as MU of other measurements with the same method under the same conditions.

Since direct observation of the variability of results is not always feasible for estimation of standard MU, there is alternately a Type-B method, which is a bottom-up approach to harness all other type of information about the probability distribution of the results around the unknown true value. This information could be uncertainties stated in calibration certificates of measuring devices used (eg, balances, pipettes, thermometers etc.), could be due to digitization of quantity values, or could be a known systematic measurement deviation—to just mention a few examples. A Type-B MU is estimated by analyzing all potential sources and components contributing to the uncertainty and combining them in a mathematical model of the measurement process to yield the combined MU for the whole process. Although the Type-B approach is applicable universally, the mathematical model of the measuring procedure may become complex easily and analyzing the uncertainty budget could be unhandy.

On the other hand, it is very common in laboratory medicine to employ control samples with known target values for IQC. Hence, these IQC data can be analyzed easily to estimate the MU of a measuring procedure based on (mainly) the Type-A method, if the control material behaves very similar to patient samples. In fact, a combined MU, complementing the Type-A method with Type-B information about the MU of the target value of the control material, is proposed as follows.

Consider a medical laboratory frequently performing IQC for a certain measurand X, eg, plasma glucose, on one or more devices employing two different control samples *A* and *B* with two different target values. It is assumed that the measurement procedure is stable, ie, the IQC results under consideration were all accepted. The medical laboratory can calculate for a certain number of IQC measurements the means as best estimates of the expectation values of the measurands, the standard deviations as standard MU, and the systematic measurement deviations. The overall MU for the measurand consists mainly of three components (see Supplement A for details), ie,

$$u(x) = \sqrt{\left(1 + \frac{1}{n}\right) \cdot s^2(x) + \delta^2 + u^2(x_0)} \qquad (1)$$

The three components contributing to the MU are given by

(I)   the standard deviation $s(x)$,
(II)   bias δ and
(III)   the target value uncertainty (TVU) $u(x_0)$ of the target value x_0.

The number of IQC results analyzed per control sample is denoted as n. The TVU of the target value $u(x_0)$ is often neglected, which will be discussed later. Consideration of systematic measurement deviation or bias as a component of MU needs some explanation.

Although strict interpretation of GUM requires that all results are corrected for known bias,[14] this correction is usually not common in laboratory medicine as well as in other analytical measurements for several good reasons.[15-17] Some of them are evident as well with the practical examples discussed below. Hence, we follow the procedures to estimate MU consistent with GUM, but *without* correcting bias, as described in references.[15,16] Accordingly, the bias of a method could be estimated as mean bias δ from repeated replicate measurement runs, and its uncertainty $u(\bar{\delta})$ from its variance and standard deviation, as discussed above. If, however,

bias is not corrected, the *assumed* mean bias is zero. Therefore, the corresponding experimental estimate of its variance is $V(\delta) \approx \sum \delta_i^2 / n$, ie, the uncertainty is given by the root mean square of all observed bias of repeated replicate measurements which becomes the bias itself if it is estimated from a single replicate measurement run ($i = n = 1$).

For practical reasons, the MU is commonly used to state an interval of measurement values in which the true value is believed to lie in with a high (typically 95%) probability. This interval is calculated with the help of the extended measurement uncertainty

$$U(x) = k u(x), \tag{2}$$

where k is the so-called coverage or extension factor leading to a desired coverage probability p or level of confidence, depending on the probability distribution of the quantity X. For normal distributions, $k(p = 95\%) = 1.96 \approx 2$. For higher levels of confidence or other distributions, the coverage factor has other values.[6] The coverage interval is then stated as $x \pm U(x)$.

Results

The MU per eqn. (1) has been exemplarily estimated by analyzing IQC data from one year obtained in one laboratory on three different instruments of the same type (Dimension Vista, Siemens Healthineers, Eschborn, Germany) for glucose employing the hexokinase method with control samples of two different concentrations (A and B). In total, 2291 results of samples A and B obtained with all three devices have been analyzed. First, the data for each device and control sample were approved to be normally distributed. The monthly means observed with each device for both control samples are shown in Figure 1. The average number of results per month for each mean was approximately $n = 32$ (range 25-37). The error bars shown in Figure 1 display the standard deviations of the mean and are of the same size as the symbols in this representation. Results are all quite close to the assigned target values of the control samples.

Figure 2 shows monthly coefficients of variation $CV(C) = s(C)/\overline{C}$ observed with the three devices for control samples A and B, where $\overline{C}$ denotes the mean concentration. Accordingly, the variability of the results expressed as CV are very similar for the devices, ranging from about 1% to 3%.

Figure 3 displays the monthly relative systematic deviation $\delta_{rel} = \delta / \overline{C}$ between measured means $\overline{C}$ and target values of the control samples. It varies between −0.6% and +3.6% with the low concentration sample A, and between −1% and +4.8% for sample B. Obviously, the relative bias shows a variability, which is of the same order of magnitude as the CV and even changes sign in some cases. These observations are clear indications that, indeed, correction of results by subtraction of bias is not reasonable. Instead, it is advisable to consider bias as
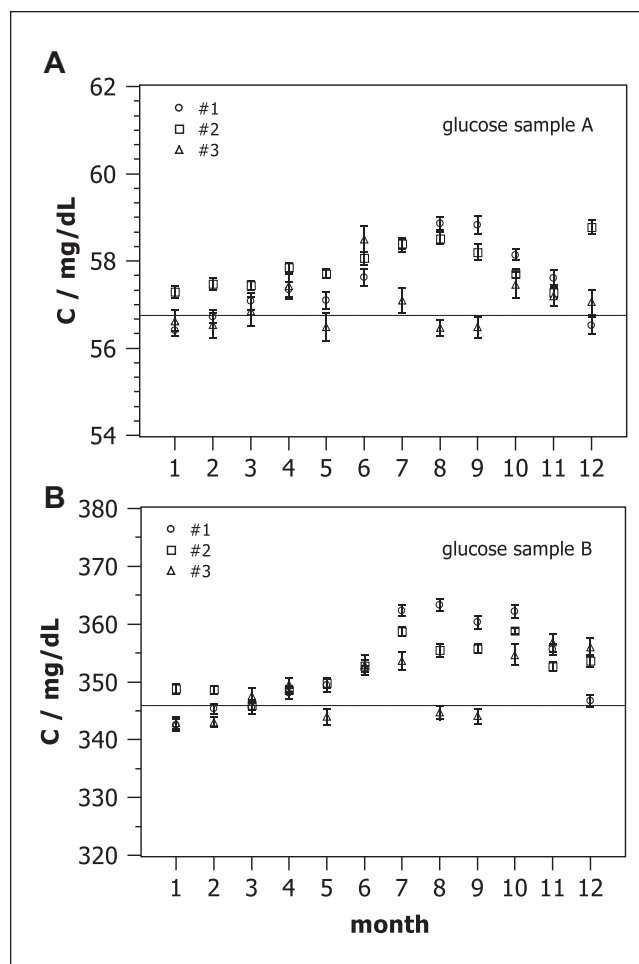

Figure 1. Monthly means of glucose concentration measured with three devices #1, #2, and #3 on two control samples A ($C_{0A} = 56.75 mg / dL$) and B ($C_{0B} = 345.9 mg / dL$) during internal quality control. (Figure in mmol/L can be found in supplements.)

another component of the measurement uncertainty, as stated above. It is worth noting that the three devices are almost of the same age and calibration is performed every two weeks with the same calibrator.

With the help of the results shown in Figures 2 and 3, we estimated the monthly combined MU considering the three components, ie, the analytical MU of the laboratory, the bias, and the MU of the target value, as given with eqn. (1), provided the uncertainties of the target values of the controls are known. For the control samples under discussion, the manufacturer provided specific $\pm 3s$ ranges for the target values, determined from multiple measurements, from which TVU have been calculated for sample A as $u(C_{0A}) = 2.10 mg / dL$ ($u(C_{0A}) / C_{0A} = 3.7\%$) (0.12 mmol/L) and for sample B as $u(C_{0B}) = 10.03 mg / dL$ ($u(C_{0B}) / C_{0B} = 2.9\%$) (0.56 mmol/L), respectively. Using these values, the relative combined MUs shown in Figure 4 have been obtained. Estimation of MU becomes more reliable allowing for general statements if larger number of repeats are analyzed, eg, from IQC measurements of one year. Table 1 summarizes the
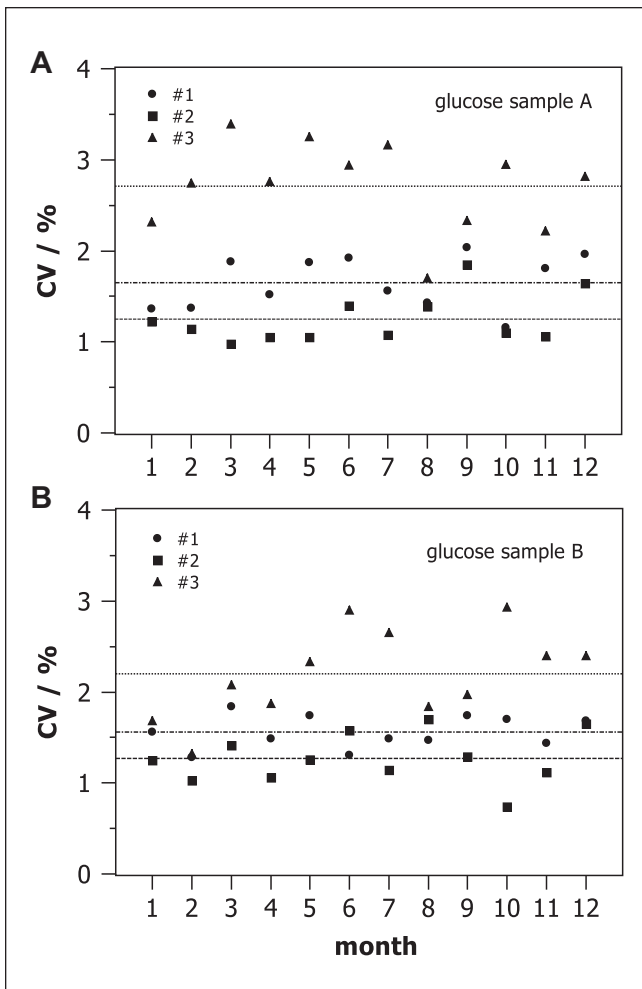
Figure 2. Monthly coefficient of variation of glucose concentrations measured with three devices #1, #2, and #3 on two control samples A and B of different concentration during internal quality control.

corresponding analysis of the above discussed internal quality controls for one year. The extended ($k = 2$) combined relative MU corresponding approximately to 95% confidence intervals is close to 9% (range 8%-11%) with sample A and 8% (range 6%-12%) with sample B. In addition, absolute values for $U(k = 2)$ are given for completeness.

Discussion

The combined MU as shown in Table 1 may be used as MU of other measurements with the same method under the same conditions if data are normally distributed. It was checked that at least 95% of the results were within the range $\bar{C} \pm U(C)$. Since the MU consists of three contributions, ie, standard deviation (I), bias (II), and TVU (III) (cf. eqn. 1), it is worth analyzing the results obtained with respect to these contributions. From Table 1, it is obvious that contribution (III) ie, the MU of the control material is largest, followed by
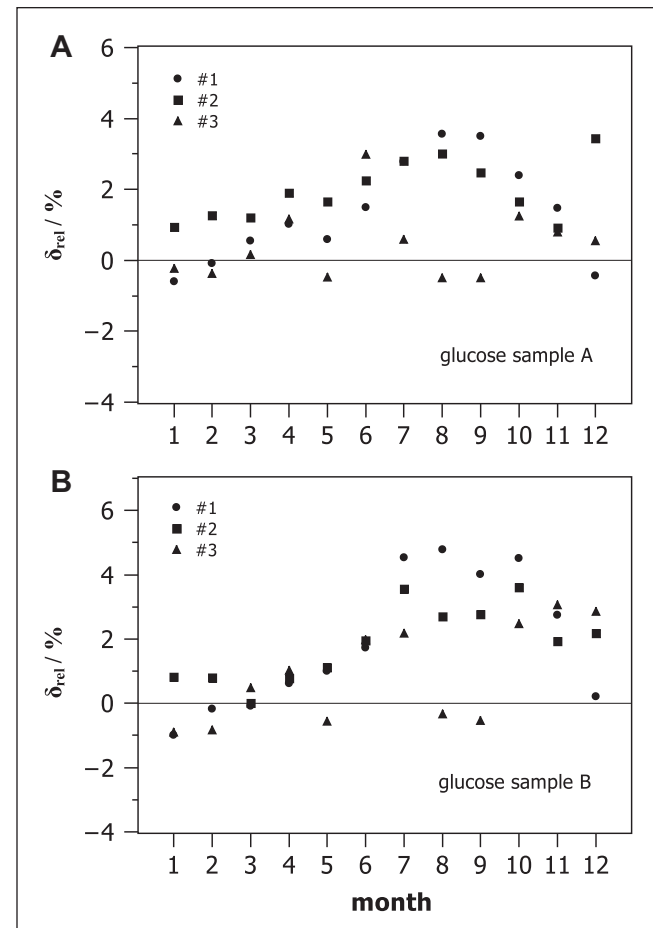

Figure 3. Monthly relative measurement deviation between measured mean glucose concentrations and target values of two control samples A and B during internal quality control observed with three devices #1, #2, and #3.

contribution (I) given by the standard deviation with devices #1 and #3. Part (II) given by the bias has the smallest contribution for devices #1 and #3, whereas the order between standard deviation and bias is reversed with #2. A graphical quantitative representation of the different contributions to the combined MU is exemplarily shown in Figure 5 for the yearly analysis obtained with data from device #1. Results with the other two devices look very similar and therefore are not shown here.

For large *n*, the quantity $\Delta \approx \sqrt{s^2 + \delta^2}$ introduced with Figure 5 describes the root mean square (rms) of the measurement deviation from the target value as introduced in reference.[18] Obviously, the combined MU $u(C)$ is identical to Δ if the uncertainty of the target values of the control samples is neglected, ie, assuming $u(C_0) = 0$, which is common practice as already mentioned above. In fact, Δ comprises that part of the MU that reflects the quality of the laboratory results in terms of variability and bias, but without the uncertainty of the target value of the control material assigned by its manufacturer. It is presently the basis for quality assessment of
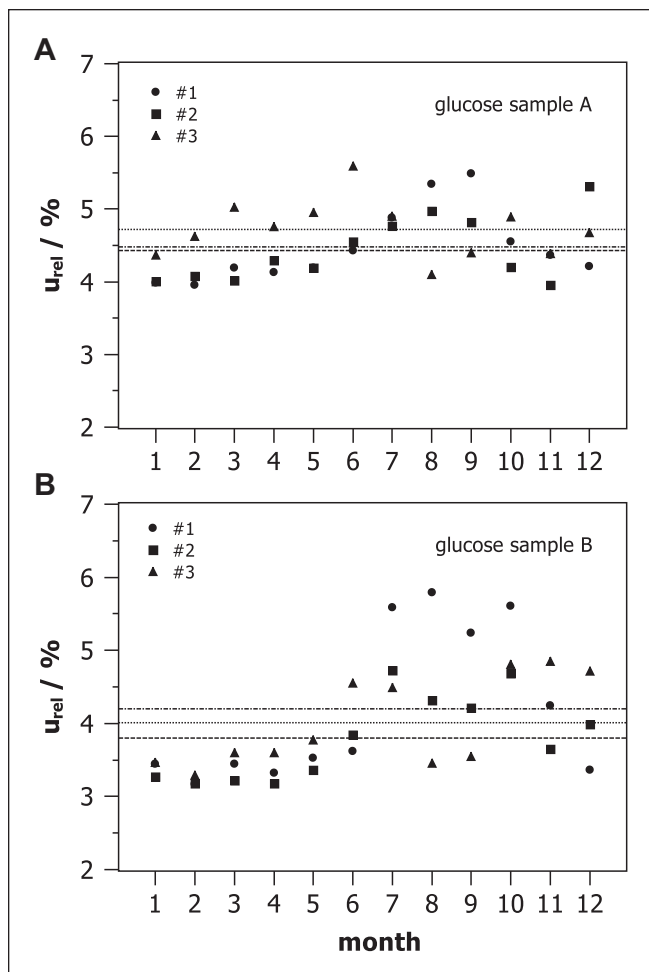
Figure 4. Monthly relative combined measurement uncertainty estimated for three devices #1, #2, and #3 with two control samples A and B from internal quality assessment.

Table 1. Analysis of IQC Data for Glucose (Hexokinase Method) with Two Control Samples on Three Devices for One Year.

Device	Parameter		Sample A	Sample B
#1	$\bar{C}$ / mg/dL		57.56	352.90
	(I)	s / mg/dL	1.26	9.17
	(II)	δ / mg/dL	0.81	6.99
	(III)	TVU / mg/dL	2.10	10.03
	$U(C)$ / mg/dL ($k = 2$)		**5.20**	**30.84**
	$U(C)/C$ / % ($k = 2$)		**9.04**	**8.74**
#2	$\bar{C}$ / mg/dL		57.89	352.54
	(I)	s / mg/dL	0.88	6.04
	(II)	δ / mg/dL	1.30	6.63
	(III)	TVU / mg/dL	2.10	10.03
	$U(C)$ / mg/dL ($k = 2$)		**5.30**	**25.38**
	$U(C)/C$ / % ($k = 2$)		**9.16**	**7.72**
#3	$\bar{C}$ / mg/dL		57.04	349.40
	(I)	s / mg/dL	1.66	9.42
	(II)	δ / mg/dL	0.41	3.50
	(III)	TVU / mg/dL	2.10	10.03
	$U(C)$ / mg/dL ($k = 2$)		**5.43**	**28.58**
	$U(C)/C$ / % ($k = 2$)		**9.52**	**8.02**

Abbreviations: $\bar{C}$: mean concentration, IQC: internal quality control, s : standard deviation, δ : systematic measurement deviation, TVU : target value uncertainty, $U(C) = k \cdot u(C)$ extended MU (see text for details) and relative extended MU, both given in bold. The number of IQC results for each device and control was for #1: nA = 380, nB = 374; for #2: nA = 381, nB = 380; for #3: nA = 395, nB = 387. Target values (concentrations) provided by the manufacturer of the control samples are given with Fig. 1. (Table in mmol/L can be found in supplements.)

quantitative analytical results in laboratory medicine as required by the German guidelines (Rili-BAEK),[19] whereas the TVU of the control material is neither part of that assessment nor well known to physicians as an important contribution to MU of measurement results such as plasma glucose. For sample A of this study, the contribution of TVU almost doubled the overall MU, as can be seen in Figure 5. Ignoring this considerable contribution in practice may blind medical laboratories in realizing the analytical quality they could provide. Therefore, it is important to include the assignment of TVU for control materials (as well as calibration materials) into legal quality requirements for analytical quality.

Based on MU as proposed here, accuracy and quality of measurement results in laboratory medicine can be estimated in a comparable and practical way. "Normal" analytical variations can be distinguished from other variations. Proper consideration of MU allows for a comparison and assessment of test results with a reference value or with clinical decision limits based on clear statements about the level of

significance. With the help of MU, serial test results may be assessed whether they are different (or not) with a certain level of significance, as well. A practical conclusion from that is that MU should be the relevant parameter to derive so-called minimal differences[20,21] rather than CV alone.

Official requirements and frequently monitoring of combined MU bear the potential to improve analytical quality of quantitative measurands. Especially for key measurands such as plasma glucose, this can improve patient care for both diagnosis and management of DM. The success of these efforts could be monitored by quality assurance schemes.

Conclusion

All measurements, including those for plasma glucose concentration, are subject to MU. It is important to know and report a reliable and complete MU in order to be able to meet the needs eg, in the diagnosis and management of DM. Employing an empirical and practical approach for the estimation of MU based on analysis of multiple data usually obtained during daily routine IQC using control samples with assigned target values is proposed. TVU of the commercial control materials used with the data set investigated was considerably larger than standard deviation and bias of
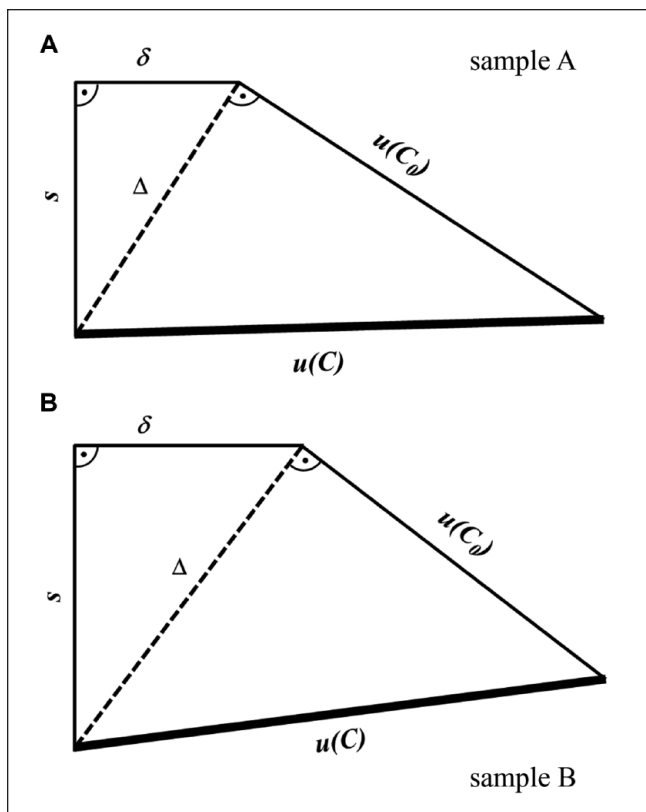
Figure 5. Geometric representation of quantitative contributions of (I) standard deviation s, (II) bias δ, and (III) target value uncertainty TVU ($u(C)_0$) of the control sample to the combined MU $u(C)$, drawn as bold line. The root mean square (rms) of the measurement deviation from target value is denoted as $\Delta \approx \sqrt{s^2 + \delta^2}$ and drawn as dashed lines. Length of all lines are proportional to the values for the corresponding parameters given in Table 1 for device #1.

the measurement procedure. This in turn has a negative impact on the combined MU of medical laboratory results. This problem should be addressed more carefully in the future eg, by legal and normative regulations.

Abbreviations

CV, coefficient of variation; DM, diabetes mellitus; GUM, Guide to the Expression of Uncertainty in Measurement; IQC, internal quality control; Rili-BAEK, Guideline of the German Medical Association on Quality Assurance in Medical Laboratory Examinations; MU; measurement unscertainty; TVU, target values.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

ORCID iDs

Astrid Petersmann [iD] https://orcid.org/0000-0001-9843-2263
Rainer Macdonald [iD] https://orcid.org/0000-0002-4719-5349

Supplemental Material

Supplemental material for this article is available online.

References

1. World Health Organization. *Classification of Diabetes Mellitus*. Geneva: World Health Organization.
2. American Diabetes Association. Classification and diagnosis of diabetes: standards of medical care in diabetes-2020. *Diabetes Care*. 2020;43(suppl1):s14-s31.
3. Joint Committee for Guides in Metrology (JCGM). *International Vocabulary of Metrology - Basic and General Concepts and Associated Terms (VIM)*. 3rd ed. [2008 version with minor corrections]. JCGM; 2012:200.
4. Bietenbeck A, Geilenkeuser WJ, Klawonn F, et al. External quality assessment schemes for glucose measurements in Germany: factors for successful participation, analytical performance and medical impact. *Clin Chem Lab Med*. 2018;56:1238-1250.
5. Friedecky B, Kratochvila J, Budina M. Why do different EQA schemes have apparently different limits of acceptability? *Clin Chem Lab Med*. 2011;49:743-745.
6. Joint Committee for Guides in Metrology (JCGM). *Evaluation of Measurement Data — Guide to the Expression of Uncertainty in Measurement*. [GUM 1995 with minor corrections]. JCGM; 2008:100.
7. Joint Committee for Guides in Metrology (JCGM). *Evaluation of Measurement Data — An Introduction to the "Guide to the Expression of Uncertainty in Measurement" and Related Document*s. JCGM; 2009:104.
8. White GH. Basics of estimating measurement uncertainty. *Clin Biochem Rev*. 2008;29(suppl1):s53-s60.
9. White GH, Farrance I. Uncertainty of measurement in quantitative medical testing: a laboratory implementation guide. *Clin Biochem Rev*. 2004;25:s1-s24.
10. American Association for Laboratory Accreditation. Policy on estimating uncertainty of measurement for ISO 15189 testing laboratories. https://pdfslide.net/download/link/p903-policy-on-estimating-uncertainty-of-american-association-for-laboratory.html. Accessed June 2020.
11. European Federation of National Associations of Measurement, Testing and Analytical Laboratories. *Measurement uncertainty revisited: alternative approaches to uncertainty evaluation*. Eurolab Technical Report No 1/2007. http://www.eurolab.org. Accessed January 4, 2019.
12. European Federation of National Associations of Measurement, Testing and Analytical Laboratories. *Guide to the evaluation of measurement uncertainty for quantitative test results*. Eurolab Technical Report No 1/2006, August 2006. http://www.eurolab.org. Accessed January 4, 2019.
13. Clinical and Laboratory Standards Institute. *Expression of measurement uncertainty in laboratory medicine; approved guideline*. CLSI document EP29-A. Wayne, PA: CLSI. 2012.

14. Phillips SD, Eberhardt KR, Parry B. Guidelines for expressing the uncertainty of measurement results containing uncorrected bias. *J Res Natl Inst Stand Technol*. 1997;102:577-585.

15. Magnusson B, Ellison SLR. Treatment of uncorrected measurement bias in uncertainty estimation for chemical measurements. *Anal Bioanal Chem*. 2008;390:201-213.

16. Lira IH, Wöger W. Evaluation of the uncertainty associated with a measurement result not corrected for systematic effects. *Meas Sci Technol*. 1998;9:1010-1011.

17. Krouwer JS. Critique of the guide to the expression of uncertainty in measurement method of estimating and reporting uncertainty in diagnostic assays. *Clin Chem*. 2003;49:1818-1821.

18. Macdonald R. Quality assessment of quantitative analytical results in laboratory medicine by root mean square of measurement deviation / Bewertung der Qualität quantitativer analytischer Ergebnisse in der Laboratoriumsmedizin anhand des quadratischen Mittelwertes der Messabweichung. *J Lab Med*. 2006;30:111-117.

19. Revision of the "Guideline of the German Medical Association on Quality Assurance in Medical Laboratory Examinations – Rili-BAEK" (unauthorized english translation). *J Lab Med*. 2015;39:26-69.

20. Keutmann S, Zylla S, Dahl M, et al. Measurement uncertainty impacts diagnosis of diabetes mellitus: reliable minimal difference of plasma glucose results. *Diabetes Ther*. 2020;11:293-303.

21. Petersmann A, Müller-Wieland D, Müller UA, et al. Definition, classification and diagnosis of diabetes mellitus. *Exp Clin Endocrinol Diabetes*. 2019;127:s1-s7.