

Opinion Paper

Ian Farrance*, Tony Badrick and Robert Frenkel^a

Uncertainty in measurement and total error: different roads to the same quality destination?

<https://doi.org/10.1515/cclm-2018-0421>

Received April 22, 2018; accepted June 7, 2018

Abstract: The debate comparing the benefits of measurement uncertainty (uncertainty in measurement, MU) with total error (TE) for the assessment of laboratory performance continues. The summary recently provided in this journal by members of the Task and Finish Group on Total Error (TFG-TE) of the EFLM put the arguments into clear perspective. Even though there is generally strong support for TE in many laboratories, some of the arguments proposed for its on-going support require further comment. In a recent opinion which focused directly on the TFG-TE summary, several potentially confusing statements regarding ISO15189 and the *Evaluation of measurement data – Guide to the expression of uncertainty in measurement* (GUM) were again promulgated to promote TE methods for assessing uncertainty in laboratory measurement. In this opinion, we present an alternative view of the key issues and outline our views with regard to the relationship between ISO15189, uncertainty in measurement and the GUM.

Keywords: bias; estimated glomerular filtration rate (eGFR); ISO15189; measurement uncertainty; total error; uncertainty in measurement.

The debate comparing the benefits of measurement uncertainty (uncertainty in measurement, MU) with total analytical error (TAE or TE) for the assessment of laboratory performance continues. The summary provided in this journal by members of the Task and Finish Group on Total Error (TFG-TE) of the EFLM put the arguments into

clear perspective [1]. The arguments, both for and against, were presented in an even-handed and scientific manner. Even though there is generally strong support for TE in many laboratories, some of the arguments proposed for its on-going support require further comment. In a recent opinion which focused directly on the TFG-TE summary [2], several potentially confusing statements regarding ISO15189 and the *Evaluation of measurement data – Guide to the expression of uncertainty in measurement* (usually referred to as the GUM) [3, 4] were again promulgated in order to promote TE methods for assessing uncertainty in laboratory measurement. Even though we have tried to address some of these issues previously, the misunderstandings appear to continue [5]. In this opinion, we present an alternative view of the key issues in order to guide those who may accept all of the Westgard statements at face value.

This opinion is not in itself a criticism of the TE concept, but an alternative view regarding some of the reasons advanced for its preference over MU. We can certainly agree with many of Westgard's comments, but those relating to ISO15189 and the requirements for a GUM "bottom-up" approach appear misleading. We also believe that whatever can be achieved by TE can also be achieved by MU and GUM; but the reverse is not necessarily true (that is, not everything that can be achieved by MU and GUM can also be achieved by TE).

As there are several important concepts interwoven within the Westgard article [2], it is difficult to know where to actually start. However, we assume readers will be familiar with the total error concept and the definition of uncertainty in measurement (which is a very general and widely applicable definition) [6]. Our main comments with regard to the Westgard article are:

1. In the Westgard article, there is the implication that ISO15189 required MU to be evaluated according to the GUM [2]. In particular, "TAE vs. MU became an issue when ISO15189 was published in 2007 and required that 'uncertainty of results' be considered in the accreditation of a medical laboratory ... Another ISO document, GUM became the obvious source for this new requirement ...". More definitive statements to

^aFormer affiliation: National Measurement Institute, West Lindfield, NSW, Australia.

*Corresponding author: Ian Farrance, Discipline of Laboratory Medicine, School of Health and Biomedical Sciences, RMIT University, Bundoora, Victoria 3083, Australia, E-mail: ian.farrance@rmit.edu.au

Tony Badrick: RCPA Quality Assurance Programs, St Leonards, NSW, Australia

Robert Frenkel: Roseville, NSW, Australia

this effect are provided on the Westgard internet site [7, 8]. To start, ISO15189 was first formally published in 2003 and not 2007 as indicated in the Westgard article. However, all editions of ISO15189 certainly did require the assessment of MU, but the implication that MU must be determined by the GUM “bottom-up” approach is incorrect. ISO15189 has never mandated the “bottom-up” approach, has never included GUM as a normative reference, and only mentioned GUM as a subsidiary reference in a bibliography [3, 9–11]. This lack of specification has enabled laboratories to choose the procedure (or combination of procedures) which may be used to demonstrate their performance quality. Since its inception in 2003, ISO15189 has never suggested any particular approach for determining MU. Westgard is certainly correct in stating that the 2012 edition of ISO15189 does give some suggestions as to how internal quality control (IQC) data may be used for determining MU, but this still does not specify a particular procedure [3].

2. VIM provides the definitive definition of MU [6]: Clause 2.26 states *measurement uncertainty* is a “non-negative parameter characterizing the dispersion of the quantity values being attributed to a measurand, based on the information used”. There is nothing about how the uncertainty inherent in all laboratory results is to be calculated, quantified or described. Total error is not mentioned but nor is any alternative method, so perhaps ISO15189 is not as restrictive as has been suggested. In some regard, MU can be considered as just another jargon, which replaces imprecision and bias (systematic error). The narrow definition of MU implied throughout the Westgard article overlooks the wider view allowed by GUM itself and the “top-down” procedures described by several professional organizations at an early stage in the development of MU applications. Such early examples of a suggested “top-down” approach were available in the *Handbook for the calculation of measurement uncertainty in environmental laboratories* (2003) [12]; *EA guidelines on the expression of uncertainty in quantitative testing* (2003) [13]; and somewhat later in *Eurolab, Guide to the evaluation of measurement uncertainty for quantitative test results* (2006) [14].
3. The first edition of ISO15189 was released in 2003 [10], with government-funded pathology services in Australia requiring full ISO15189 accreditation in 2005. All pathology laboratories (and many POCT facilities) which receive government funding (including fee for service payments) are accredited to this ISO standard. Even though a mandatory accreditation scheme had

been in operation for many years before the introduction of ISO15189, its introduction was achieved in a progressive manner with time for laboratories to seek assistance if this was required. Any suggestion that a full “bottom-up” GUM approach should be used for routine pathology testing was never seriously contemplated. All laboratories had existing quality procedures based on IQC and external proficiency testing (external quality assessment, EQA) already mandated by the national accreditation scheme. Imprecision (standard deviation, standard uncertainty) was determined from replicate measurements of control material over an extended period (intermediate precision), and bias from EQA when direct traceability studies were impractical or unavailable. Some laboratories used total error methods while some preferred the separate representation of imprecision and bias. The widely utilized (within Australia) guide, *The uncertainty of measurement in quantitative medical testing – A laboratory implementation guide* [15], published by the AACB in 2004 to assist with the introduction of ISO15189, provided a mixture of total error and MU procedures. In most circumstances, a change of terminology was all that was required, with a possible extension to include a fitness for purpose component [15]. Irrespective of the suggestions for a “top-down” approach as described in various early editions of well publicized guidelines [12–14], laboratories were free to choose whatever procedure they considered appropriate. The chosen approach was then more generally applied to all quantitative test procedures within that laboratory. In the same manner that ISO15189 does not mandate any particular procedure for assessing laboratory performance, we understand that the requirements of CLIA are similar.

4. We can certainly agree that the main difference between total error and GUM is the treatment of bias (systematic error). This topic has indeed provided much debate in the medical sciences literature and is still on-going, with various alternative procedures having been devised which take the effect(s) of bias into account [16, 17]. Bias in analytical methods (and method instrument combinations) is a significant issue for many medical laboratory tests. However, having separate procedures for the representation of imprecision and bias is simple and inherently logical. The ISO15189 requirement is to document (and understand) the particular process chosen.
5. Even though the Westgard article discusses bias with respect to its inclusion in the TE procedure, other useful approaches for the treatment of bias are not

considered. In particular, the VIM definition of MU allows for the inclusion of “components arising from systematic effects” as may occur in estimates of intermediate precision (for example, where changes in bias may arise as a result of a change in calibrator or reagent). In addition, with instrument and method group differences for many measurands clearly evident in EQA reports, alternative strategies such as test “harmonization” are being developed to minimize bias [18], or where result interpretation is by comparison to a population reference interval or by comparison to a previous result, the effect of bias is largely irrelevant provided it remains analytically consistent and consistent with the comparator. Meanwhile, we certainly agree with Westgard that in situations where bias can be reduced or eliminated, both TE and MU essentially become an assessment of imprecision. Whatever method is chosen to provide an uncertainty estimate, GUM outlines the manner in which this can be described in the form of a standard deviation.

6. Since its introduction in 1993, the GUM has provided the basic framework for evaluating MU. In contrast to total error, the preferred GUM approach is based on the assumption that all systematic errors are identified and corrected at an early stage in the measurement process. The quality of a measurement is then expressed by taking account of the uncertainty associated with random errors plus the uncertainty associated with any correction for systematic error (bias). However, section F.2.4.5 of the GUM specifically describes the procedure where a significant systematic effect may be taken into account by enlarging the uncertainty assigned to the result: “An example is ... replacement of an expanded uncertainty U with $U+b$, where U is an expanded uncertainty obtained under the assumption $b=0$ ”. This situation is directly comparable to the calculation of total error where the uncertainty includes both an imprecision and bias component. However, having given some credence to the total error approach, the preferred GUM procedure is to evaluate random and systematic uncertainties separately. Supporters of GUM actually prefer the separate reporting or representation of imprecision and bias. This separate representation of imprecision and bias in many respects is essentially equivalent to the bias versus imprecision plots or *method decision charts* described by Westgard [19].
7. “Ignoring the uncertainty in the estimate of the bias is seldom a problem because it is small compared

with the estimate of the SD for imprecision and small compared with the uncertainty in the estimate of that SD, which is ignored in the MU methodology” [2]. This is another statement with which we have some concern. It is presented as a generally applicable concept which suggests that the uncertainty in the bias is always less than the uncertainty in the tests imprecision (“the uncertainty in the estimate of that SD”). It is certainly true that GUM procedures rarely take the uncertainty of the imprecision into account, and total error methods do not take the uncertainty in the bias or the uncertainty in the reference procedure into account. However, to state that one is clearly less than the other is not necessarily true. As bias is generally evaluated as the difference between a test measurement and a reference measurement performed on the same material (specimen), the uncertainty in the bias is essentially the uncertainty associated with the reference measurement procedure. In practice however, the uncertainty in the bias may not be significantly different to the imprecision of the laboratory test procedure. This has been demonstrated recently in a comparison of method imprecision, method bias and uncertainty in the bias for seven key measurands by Padoan et al. [20], which clearly shows that bias and its uncertainty can rival method imprecision as a principal contributor to the overall MU. It can also be observed with regard to the specified uncertainties for reference methods or reference materials as described in the Joint Committee for Traceability in Laboratory Medicine (JCTLM) database [21], with uncertainty values often similar to the corresponding routine laboratory test procedure. The uncertainty in the imprecision (Δu) of the test measurement can be easily calculated as outlined previously [22]. It is largely dependent on the degrees of freedom associated with the estimation of the imprecision of the test procedure (that is; the number of observations associated with the calculation of the imprecision derived from IQC). For many (most?) routine assays in the clinical laboratory, this number can easily exceed 100 or more. For 50 replicate assays used to determine imprecision from IQC, $\Delta u=10\%$ of the estimated method imprecision (standard uncertainty) value. For 100 replicate assays, $\Delta u=7\%$ of the estimated imprecision; and for 200 replicate assay, $\Delta u=5\%$ of the estimated imprecision. An example comparing the Sigma-metric approach to alternative procedures which include the uncertainty associated with the reference method has been described [17].

8. Finally, an important point on which we can all agree. In his article, Westgard suggests that “The major advantage claimed for the MU methodology is that the estimates of uncertainty from different sources can be combined to estimate the uncertainty of a larger process, e.g. the uncertainty of the total testing process could be estimated from the individual uncertainties of the preanalytic, analytic, and postanalytic phases” [2]. This statement in effect acknowledges the position taken by the TE-TFG in their report regarding the value of being able to combine uncertainty components, particularly if this can be expanded to include the diagnostic uncertainty of the whole testing process [1]. In theory this is true, but for this to occur, there is a requirement to define the various uncertainty components in numeric terms (as standard uncertainties). Thus, with the exception of the various components of biological variation which are already aggregated and disaggregated with each other and with analytical MU using the ‘root-sum-square’ procedure, most preanalytical and postanalytical uncertainties have not (as yet) been numerically described. This would also apply to the concept of diagnostic uncertainty as outlined by Theodorsson [23] and by comments in the TE-TEG report [1]. As further suggested by Westgard “This paradigm shift to diagnostic uncertainty is a lofty and revolutionary goal for the future; however, it must be achieved by the evolution of existing quality management practices that deal with the reality of laboratory medicine and the theory of metrology.” [2]. ISO15189:2012 also acknowledges this fact by specifically suggesting that “The relevant uncertainty components are those associated with the actual measurement process, commencing with the presentation of the sample to the measurement procedure and ending with the output of the measured value” (note 1 to item 5.5.1.4) [3]. The well-known ‘root-sum-square’ procedure, which is used to determine the uncertainty of a calculated quantity when given the uncertainties of measured input values (for relatively ‘simple’ linearly related terms such as serum anion gap) [15], is essentially a GUM procedure. The aggregation of uncertainties from the inputs to a measurement equation in order to provide the uncertainty of the output (or calculated quantity value) is indeed a real (and underutilized) benefit of the GUM approach. Simple examples have been described on numerous occasions [15, 22, 24], but the real benefit of the GUM can be shown by its application to more complex examples and to the inclusion of subjective (expert) opinion through a numerical estimation of a type B

uncertainty. The GUM describes how MU may appear as random variations in repeated measurements, as uncertainty documented from previous experience or as uncertainty estimated by professional judgement. The introduction of numerical terms that describe the various components of diagnostic uncertainty or the preanalytical and postanalytical uncertainty components not currently so described, may well be included by this GUM specific procedure. However, the Westgard statement that “The TAE model can be expanded to include a ‘clinical decision interval model’...” will also require the additional components to be described in numerical terms (by the GUM approach?). In situations where uncertainty can already be described in numerical terms the GUM propagation procedures provide a clear benefit. For example, the uncertainty of the calculated serum free testosterone given the uncertainties of the measured serum sex hormone binding globulin and total testosterone, can only be derived using the GUM approach [25, 26], or the uncertainty in the estimated glomerular filtration rate (eGFR) given the uncertainties in the various input quantities to the eGFR when using either the MDRD equation [25] or the CKD-EPI equation [26, 27]. In a similar manner to the propagation of measurement uncertainties, the propagation of biological variation in the eGFR equation has also been described using a procedure derived from the GUM [27].

In summary, both total error and GUM-style MU procedures contribute to an understanding of uncertainty in measurement as it relates to medical testing laboratories. In many respects, they provide similar information which may be utilized to determine analytical method performance and to demonstrate the quality of laboratory results for accreditation purposes. However, whatever method is chosen to provide an uncertainty estimate, GUM outlines the manner in which this can be described in the form of a standard deviation.

Author contributions: All the authors have accepted responsibility for the entire content of this submitted manuscript and approved submission.

Guarantor: IF.

Research funding: None declared.

Employment or leadership: None declared.

Honorarium: None declared.

Competing interests: The funding organization(s) played no role in the study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the report for publication.

References

1. Oosterhuis WP, Bayat H, Armbruster D, Coskun A, Freeman KP, Kallner A, et al. The use of error and uncertainty methods in the medical laboratory. *Clin Chem Lab Med* 2018;56:209–19.
2. Westgard JO. Error methods are more practical, but uncertainty methods may still be preferred. *Clin Chem* 2018;64:636–8.
3. International Organization for Standardization. Third edition, ISO 15189:2012, Medical laboratories – requirements for quality and competence. <http://www.iso.org/iso/cataloguedetail?csnumber=56115>. Accessed: May 2018.
4. Bureau International des Poids et Mesures. JCGM 100:2008, Evaluation of measurement data – guide to the expression of uncertainty in measurement (GUM). <http://www.bipm.org/en/publications/guides/gum.html>. Accessed: May 2018.
5. Farrance I, Badrick T, Sikaris KA. Uncertainty in measurement and total error – are they so incompatible? *Clin Chem Lab Med* 2016;54:1309–11.
6. Joint Committee for Guides in Metrology, International vocabulary of metrology – basic and general concepts and associated terms (VIM). JCGM 200:2012. <http://www.bipm.org/en/publications/guides/vim.html>. Accessed: May 2018.
7. Westgard S. Quality requirements and standards – quality goals at the crossroads: growing, going or gone? <http://www.westgard.com/gone-goals-gone.htm>. Accessed: May 2018.
8. Westgard JO, Westgard S. Making Quality Great Again! I'm with Quality! Some dismay, but less disagreement on MU vs. TE. <https://www.westgard.com/make-quality-great-again.htm>. Accessed: May 2018.
9. International Organization for Standardization. Draft international standard ISO 15189.2:2002. Medical laboratories – particular requirements for quality and competence.
10. International Organization for Standardization. ISO 15189:2003, Medical laboratories – particular requirements for quality and competence.
11. International Organization for Standardization. Second edition, ISO 15189:2007, Medical laboratories – particular requirements for quality and competence.
12. Magnusson B, Naykki T, Hovind H, Krysell M. Handbook for calculation of measurement uncertainty in environmental laboratories. Nordtest technical report TR537, edition 1.2, 2003.
13. European co-operation for accreditation. EA guidelines on the expression of uncertainty in quantitative testing. EA-4/16, 2003.
14. 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 1/2006.
15. White GH, Farrance I. Uncertainty of measurement in quantitative medical testing – a laboratory implementation guide. *Clin Biochem Rev* 2004;25:S1–S24.
16. Magnusson B, Naykki T, Hovind H, Krysell M, Sahlin E. Handbook for calculation of measurement uncertainty in environmental laboratories. Nordtest technical report TR537, edition 4, 2017. <http://www.nordtest.info/index.php/technical-reports/item/handbook-for-calculation-of-measurement-uncertainty-in-environmental-laboratories-nt-tr-537-edition-4.html>. Accessed: May 2018.
17. Frenkel RB, Farrance I. A statistical procedure for the assessment of bias in analytical methods using conditional probabilities. *Accred Qual Assur* 2017;22:265–73.
18. Koerbin G, Sikaris KA, Jones GR, Ryan J, Reed M, Tate J, et al. Evidence-based approach to harmonized reference intervals. *Clin Chim Acta* 2014;432:99–107.
19. Westgard JO, Westgard S. Quality control review: implementing a scientifically based quality control system. *Ann Clin Biochem* 2016;53:32–50.
20. Padoan A, Antonelli G, Aita A, Sciacovelli L, Plebani M. An approach for estimating measurement uncertainty in medical laboratories using data from long-term quality control and external quality assurance schemes. *Clin Chem Lab Med* 2017;55:1696–701.
21. Joint Committee for Traceability in Laboratory Medicine (JCTLM). Database of higher-order reference materials and measurement methods. <https://www.bipm.org/jctlm>. Accessed: May 2018.
22. Farrance I, Frenkel R. Uncertainty of measurement: a review of the rules for calculating uncertainty components through functional relationships. *Clin Biochem Rev* 2012;33:49–75.
23. Theodorsson E. Uncertainty in measurement and total error, tools for coping with diagnostic uncertainty. *Clin Lab Med* 2017;37:15–34.
24. Clinical and Laboratory Standards Institute (CLSI). Expression of Measurement Uncertainty in Laboratory Medicine; Approved Guideline. CLSI document EP29-A, CLSI, Wayne, PA, 2012.
25. Farrance I, Frenkel R. Uncertainty in measurement: a review of Monte Carlo simulation using Microsoft Excel for the calculation of uncertainties through functional relationships, including uncertainties in empirically derived constants. *Clin Biochem Rev* 2014;35:37–61.
26. Frenkel RB, Farrance I. Uncertainty in measurement: procedures for determining uncertainty with application to clinical laboratory calculations. *Adv Clin Chem* 2018;85:149–211.
27. Farrance I, Badrick T, Frenkel R. Uncertainty in measurement: a review of the procedures for determining uncertainty in measurement and its use in deriving the biological variation of the estimated glomerular filtration rate. *Pract Lab Med* 2018 in press, <https://doi.org/10.1016/j.PLABM.2018.e00097>.