Estimation of uncertainty in measurements in the clinical laboratory

Accreditation of clinical laboratories aims at pushing quality of operation towards a level specified in international standards for laboratory medicine, mainly the EN-ISO 15189 [1]. A key issue is the requirement to specify the uncertainty in measurements according to another internationally agreed document, the JCGM “Guide to the expression of uncertainty in measurement”, commonly known as GUM [2]. The concept of uncertainty has been met with some skepticism in the profession, due to the difficulties in identifying procedural sources of uncertainty. Ramamohan et al. [3] describe an approach to resolve this difficulty by a generally applicable method. Some background discussion of quality assessment in the laboratory may be justified before commenting on this approach.

Quality control in laboratory medicine and, in particular, clinical chemistry has a long history. In principle there are two aspects that are routinely monitored, the imprecision or random error, and the bias, or systematic error, of the results. Essentially, internal quality control (IQC) addresses imprecision whereas external quality assessment (EQA), or proficiency testing (PT), aims at assessing bias.

The publication by Belk and Sunderman [4, 5] in 1947 is often regarded as the starting point of EQA whereas the IQC has developed gradually over the years since the 1970s when Westgard, Groth and de Verdier [6, 7] introduced simulations as a scientific tool to evaluate and design control rules.

The performance of the laboratory can thus be satisfactorily described in terms of random and systematic errors. The metric “total error” (TE), defined as the “net effect of method bias and imprecision” includes both types of error in a single metric [8, 9]. A common objection to the TE is that if a known bias is included, why keep it? Furthermore, bias has a sign whereas the imprecision is a characteristic of a distribution. Therefore the quantities included in the TE are not really comparable.

The RiLiBAEK [10], the mandatory IQC system in Germany, relies on the root mean square (RMS) which can be transferred to a single metric, including experimental variance and bias.

GUM, first published in 1993 and successively updated is now freely available on the Internet [2]. The document is a result of the collaboration of the seven most influential international professional organizations in pure and applied chemistry and physics including a representative of the accreditation bodies. The GUM has been embraced by accreditation standards, e.g., EN/ISO 17025 and EN/ISO 15189 [1] which require that the uncertainty in the measurements (MU) in accredited laboratories is estimated according to GUM and made available to the end-user. Several practical guidelines have been published [11–15] to assist laboratories to estimate the MU.

The GUM requires that the variance of each the input quantities is calculated using statistical (Type A) methods or estimated (Type B) based on experience or literature information. The result obtained by the measurement is the “best estimate” of the quantity (measurand) and the attached uncertainty indicates an interval within which the “true” result is anticipated to be, at a certain level of confidence.

The “standard uncertainties” of the input quantities are shown in an “uncertainty budget” and eventually combined according to a “measurement function” or model, using mathematical rules of error propagation to obtain the “combined uncertainty” of the procedure.

MU requires that the bias is estimated and reduced, eliminated or – if judged insignificant – disregarded. The elimination of a bias is accompanied by an uncertainty which shall be included in the combined uncertainty. Details of how this is accomplished is described – with worked examples – in the CLSI document C51 [11], the CITAC Guide 4 [12], a technical report from Eurolab [13] and experience from the Australian group [14, 15].

Elimination of the bias may represent a paradigm shift in metrology since it requires, and allows, a recalculation of results according to some algorithm based on comparison of samples, recognized as recalibration. Although the traceability and quality of calibrators, much thanks to IFCC, has improved, it does not seem to solve the problem of obtaining comparable results of measurements in laboratory medicine. Therefore, harmonization of results, i.e., recalibration, in addition to the standardization of
measurement methods may become necessary. Comparisons are typically based on regression analysis and application of the regression function either to the calibrators or to the individual results. Since patient samples are used it is likely that commutability of the test material can be assumed. Care must be taken to retain traceability to certified reference materials.

Method comparisons using patient samples have been described by CLSI [16] and recently Haeckel et al. [17] published an appraisal of available regression functions. All regression analyses are a kind of average and the algorithms will not necessarily apply to individual results. However, the same objection can be made to the calibration itself, particularly if the calibrator is not fully commutable with the patient samples. Even if a bias is not eliminated by recalibration it is likely to reduce the distribution of results from different measuring systems.

It might not be possible to identify and quantify all input quantities and their uncertainties in all measuring systems. The Type B approach then offers a possibility to estimate the MU and this can be applied to the entire measurement procedure as well as to individual input quantities. Therefore, the MU of a procedure can be based on the uncertainty estimated from repeated measurements of suitable materials arranged in a model fit for purpose, e.g., records of IQC data. This is commonly recognized as the “top-down” procedure, different from the “bottom-up” which requires the detailed specification of uncertainty contributions from the input quantities. Measurement uncertainties by Type A and Type B are treated equally in the uncertainty budget.

Elimination of the bias and substituting it with an uncertainty is the major difference compared to what has been done traditionally in the laboratories. As a result, the best estimate, in theory, will be unbiased. The accuracy is thus expressed in a single metric that is easily explained and understood.

Ramamohan et al. [3] describe an interesting approach to satisfy the need for establishing a bottom-up uncertainty budget with a potential to simplify the procedure. A measuring system includes a pre-analytical, an analytical and a post-analytical step. GUM and also the authors focus on the analytical step with a view to estimate the capability of instruments in delivering a certain level of quality based on the specifications provided by the manufacturer. Many measurement procedures in the clinical laboratory can be classified according to a measurement principle defined as “phenomenon serving as a basis of a measurement” [18], e.g., those based on substrate assays and those of ion-selective electrode assays (ISE). Therefore, input quantities, i.e., the sources of uncertainty, can be assumed to be similar in these measurement procedures and only vary in size and importance. The input quantities are the calibrator and calibration, volumes of samples and reagents, and the spectrometer or electrode. The uncertainties used in the study are specified by the manufacturer and therefore limited to errors introduced into the measurement procedure by the physical or chemical components. The quantified input quantities are combined in a model that seems equivalent to the measurement function described in GUM. Thus, a combined uncertainty can be achieved and allow the user to simply plug-in values of the input quantities into a common measurement function. Any contributions from the pre- and post-analytical phases need to be considered separately. Although considered major sources of uncertainty of the result their inclusion is a major problem in the MU and GUM concepts.

The presented model does not consider the bias but it is claimed to be accommodated by changing the assumed distribution of the uncertainties. This is not illustrated, which is unfortunate, considering that the elimination of bias is one of the major features of the uncertainty concept.

The final verification of the model is summarized in a table describing the calculated uncertainties of nine substrate methods and three ISE methods. These are compared with those calculated from QC data of the corresponding quantities measured at the Mayo Clinics in Rochester USA, i.e., a top-down estimation. In some cases the results are of the same magnitude but often they are clearly deviating, and it is not obvious which one is the better approximation.

The authors conclude that the verification in the study is essential but incomplete. The input quantities, as defined in the flowcharts, are assumed to vary among instruments and manufacturers of calibrators and reagents and may not be readily available. Furthermore, those in the document will not necessarily be valid in any other measuring system. The reader is therefore left with an unanswered question what the model can contribute to the clinical laboratory in addition to a top-down or a bottom-up procedure based on Type B assumptions of the size of the uncertainties of input quantities, an uncertainty budget and a measurement function.

Conflict of interest statement

Authors’ conflict of interest disclosure: The authors stated that there are no conflicts of interest regarding the publication of this article.

Research funding: None declared.

Employment or leadership: None declared.

Honorarium: None declared.
References

1. ISO 15189:2012 Medical laboratories — Requirements for quality and competence.

Anders Kallner, Department of Clinical Chemistry, Karolinska University Hospital, 17176 Stockholm, Sweden, E-mail: anders.kallner@ki.se