Uncertainty of measurement and conformity assessment

Uncertainty of measurement and conformity assessment

Silvana Stajkovic[1], Dragan Vasilev[1], Mirjana Dimitrijevic[1] and Nedjeljko Karabasil[1]

[1] University of Belgrade, Faculty of Veterinary Medicine, Department of Food Hygiene and Technology, Bulevar Oslobođenja 18, Belgrade, Republic of Serbia

E-mail: silvana@vet.bg.ac.rs

Abstract. Knowledge of the measurement uncertainty of test results is fundamentally important for laboratories, their customers and all parties using and interpreting these results. In conformity assessment, a measurement result is used to decide if an item of interest conforms to a specified requirement. Because of measurement uncertainty, there is always the risk of incorrectly deciding whether or not an item conforms to a specified requirement based on the measured value of a property of the item. Conformity assessment can be quite challenging when the entity measured is so close to the tolerance limits of the specification that its uncertainty, however estimated, critically affects decision-making. In such cases, different decision rules can be used to make statements of conformity. The aim of this paper is to provide a survey of methods for the evaluation of measurement uncertainty in testing, as well as to stress the need for appropriate estimation of measurement uncertainty. This paper also aims to assist testing laboratories in understanding the different decision rules used in conformity assessment and level of risk (such as false accept and false reject) associated with the decision rule employed.

1. Introduction

Credibility and reliability of analytical data has never caught the public eye more than today. They are used extensively by regulators for the public benefit in the provision of services that promote safe food, clean water, an unpolluted environment, energy, health and social care services [1]. Unreliable results bring a high risk of incorrect decisions and could lead to higher costs, health risks, and illegal practices. Hence, the goal of any analytical measurement is to obtain consistent, reliable, and accurate data. For this purpose, the results of laboratories accredited according to ISO/IEC 17025 are used [2].

The result of each real measurement is not perfect. It is only an estimate of the value of the test item's characteristic being measured. In order to make adequate decisions, through the conformity assessment, which demonstrates that specified requirements relating to a product, process, system, person or body are fulfilled [3], it is necessary that these data contain evaluated measurement uncertainty, a parameter associated with the result of a measurement, that characterizes the dispersion of the values that could reasonably be attributed to the measurand [4]. Proper consideration of uncertainty is imperative when testing a sample against legal/compositional limits, especially when the entity measured is close to the tolerance limits of the specification. The new edition of the standard ISO/IEC 17025 has introduced the concept of decision rules [2]. It is defined as "a rule that describes how measurement uncertainty will be accounted for when stating conformity with a specified requirement" [2]. Conformity with a requirement is inherently connected to the decision rule employed [1]. The introduction of this concept clarifies that no single decision rule can be applied to all conformity assessment to specifications,

because it significantly depends on the test area itself. Also, whether a measurement could result in a decision on conformity (acceptance) using one decision rule and rejection using a different decision rule needs to be considered. It is, therefore, expected that the decision rule is agreed before the measurements are taken [2]. When performing conformity assessment, there are probabilities related to two types of incorrect decisions, one for the supplier and one for the consumer [5].

Therefore, the aim of this study is to provide a survey of methods for the evaluation of measurement uncertainty in testing, and to stress the need for appropriate estimation of measurement uncertainty. This paper also aims to assist testing laboratories in understanding the different decision rules used in conformity assessment and level of risk (such as false accept and false reject) associated with the decision rule employed.

2. Measurement uncertainty

All measurements are affected by a certain error. The classical approach starts from the assumption that by measuring, we determine the true value of the measured quantity and its errors, which can be random or systematic in nature. The main difficulty is that neither the true value nor the measurement error can be perfectly known [6]. Hence, the new approach, the uncertainty approach, omits the term "true value" and, in accordance with the definition, considers the interval in which that value is. This interval, which includes the best estimate of the measured quantity, is in fact the measurement uncertainty. Reporting is required when information on uncertainty is relevant to the validity or application of the test results, when the client requires it, or when the uncertainty affects conformity with a specification limit [2]. ILAC-G17:01/2021 [7] recognized that there are situations where the requirement for reporting measurement uncertainty may not be obvious.

There are many possible sources of uncertainty in testing, such as sampling, sample effects, storage conditions, instrument and operator effects, reagent purity, measurement conditions, assumed stoichiometry, computational effects, blank correction, random effects etc. [8, 9]. The required depth of the uncertainty estimations can be different in different fields. The contributions of all the above factors to the total value of the measurement uncertainty constitute the budget of the measurement uncertainty related to a particular test [8]. The basis for the evaluation is a measurement and statistical approach, where the different uncertainty sources are estimated and combined into a single value. Most of the information needed to evaluate the uncertainty is likely to be already available, like quality control charts, validation, proficiency testing, certified reference material, handbooks etc. [8, 9, 10].

2.1. Evaluating measurement uncertainty

Measurement uncertainty is assessed through two methods that are essentially just concepts for processing different types of measurement results: the type A method and the type B method. Both types of estimates are based on probability distributions. Type A standard uncertainty is calculated from a series of repeated observations and is equal to the square root of the statistically estimated variance. The type A method is called the standard deviation [8, 9]. However, the uncertainty component can also be determined without actual observations, through experience based on available information. Such an estimate is called a type B method, and the derived uncertainty is referred to as standard type B uncertainty [8, 9]. The pool of information could include previous measurement data, data provided in calibration certificates, manufacturer's specifications, uncertainty assigned to reference data taken from handbooks, experience with or general knowledge of the behaviour and properties of relevant materials and instruments etc. [8, 9].

Several approaches to obtaining an uncertainty estimate are described [8, 9, 10]. According to the EURACHEM concept, it is necessary to specify measured quantities, identify and group sources of measurement uncertainty, quantify uncertainty components (convert components to standard deviations), and finally calculate the combined, uc, and expanded, U, measurement uncertainty [8]. It should also be clear whether a sampling step is included within the procedure or not. If it is, estimation of uncertainties associated with the sampling procedure need to be considered [9]. In the Nordtest model, within laboratory reproducibility, standard deviation is combined with estimates of the method and

laboratory bias [10]. Alternatively, according to ISO 21748 [11], the combined standard uncertainty, uc, can be directly estimated from the reproducibility between laboratories, sR [10]. The Nordtest model covers all steps in the analytical chain from the arrival of the test sample in the laboratory to the reporting of the analytical result, but sampling, sample transportation, equipment and possible gross errors during data storage/retrieval are not included [10].

2.2. Reporting uncertainty

For most purposes in food analysis, an expanded uncertainty should be used, which is obtained by multiplying the combined standard uncertainty, uc, by a coverage factor, k. The choice of the factor k is based on the level of confidence desired. For an approximate level of confidence of 95%, k is 2 [8, 9, 12]. The test result, y, and the expanded uncertainty, U, should be reported as $y \pm U$ and accompanied by a statement of confidence [8, 10, 12].

Not all the uncertainty sources identified during an uncertainty evaluation will make a significant contribution to the combined uncertainty. Whether the contribution of uncertainty will be neglected depends on the relative size of the largest and smallest contributions, its impact on overall uncertainty, user requirements or regulations. Failure to properly consider all sources of uncertainty leads to a lower assessment of uncertainty.

3. Conformity assessment

Conformity assessment is any activity undertaken to determine, directly or indirectly, whether a product, system, process, body or person meets relevant standards and fulfils specified requirements [13]. It is a common activity performed in testing, inspection and calibration, being defined to establish confidence for consumers and for the safety and quality of life [14]. Over time, in practice, many statements of conformity to a specification or standard have been confused with opinions and interpretations, although these are two completely different terms. Statements of conformity can serve as a basis in the processes of giving opinions and interpretations. In conformity assessment, a measurement result is used to decide if an item of interest conforms to a specified requirement [13]. To establish procedures in order to perform conformity assessment in practical situations, objective criteria are needed in what is called "decision rule" [2, 14]. The development of a probabilistic approach in measurement, introducing measurement uncertainty as a parameter to express the variability of measurement, had a significant impact in decision process [14]. Due to measurement uncertainty, there is always the risk of an incorrect decision, especially when the result is close to the specification limit [13]. Such incorrect decisions are of two types. First, with false acceptance, an item accepted as conforming could actually be non-conforming. The probability of such an incorrect decision is called consumer's risk, because the cost associated with such a mistake is often borne by a consumer [13]. Second, with false rejection, an item rejected as non-conforming could actually be conforming. The probability of such an incorrect decision is called producer's risk, because the cost associated with such a mistake is often borne by a producer who cannot sell an item that has failed a test of conformity [13, 14]. These errors are also known as Type I (α) and type II (β) errors, respectively, meaning that conforming products are incorrectly rejected or non-conforming products are incorrectly accepted [14]. General rules for calculation of conformance probability and risk of false decision are given in ISO/IEC Guide 98-4:2012(E) [13]. The evaluated probabilities depend on the measuring system and the production process. Where the decision rule is prescribed by the customer, regulations or normative documents, a further consideration of the level of risk is not necessary [2].

3.1. Guard bands

Tolerance limits, specified requirements for a measurand of interest, consist of limiting values, and separate intervals of permissible values of the measurand, tolerance intervals, from intervals of non-permissible values. Tolerance intervals can be either one-sided, with either a lower or an upper tolerance limit, or two-sided, with both lower and upper tolerance limits [13]. Most of this paper for simplicity deals with an upper tolerance limit. When the analytical result is close to the tolerance limit, acceptance

or rejection of an item can be an incorrect decision and lead to undesirable consequences. The use of guard bands can reduce the risks of incorrect accept/reject conformance decisions by defining an acceptance interval of permissible measured values of a measurand [13, 15]. They represent a safety factor built into the measurement decision process. The length of the Guard Band (w) is the Tolerance/specification Limit (*TL*) minus the Acceptance Limit (*AL*) or $w=TL - AL$ (Figure 1) [15].

Acceptance limits and corresponding decision rules are chosen in such a way as to manage the undesired consequences of incorrect decisions.

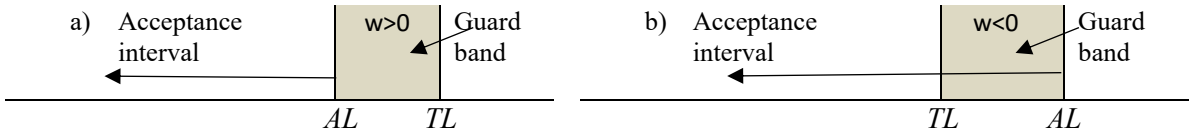

Figure 1. a) acceptance limit *AL* inside the tolerance interval *TL* and b) acceptance limit *AL* outside the tolerance interval *TL* according to ISO/IEC Guide 98-4:2012(E) [13].

The length parameter, w, is taken to be a multiple of the expanded uncertainty for a coverage factor k = 2, U = 2u. A common choice is r = 1, hence w = U. The probability of accepting a nonconforming item in that case is at most 2.3% (assuming a normal PDF for the measured quantity) [13]. There are cases where a multiplier other than 1 is more appropriate. Also, customers can define r [15].

It should be taken into account that the measurement could result in a decision on conformity (acceptance) using one guard band, or non-conformity (rejection) if a larger guard band is used. Therefore, compliance with the requirements is inherently related to the applied decision rule. Hence, the decision rule is expected to be agreed before measurements are made [2]. No single decision rule can address all statements of conformity across the diverse scope of testing and calibration.

3.2. Decision rules

An important and widely used decision rule is known as simple acceptance or shared risk. Under such a rule, where a guard band has a length equal to zero, w = 0, the producer and consumer (user) of the measurement result agree to accept as conforming (and reject otherwise) an item whose property has a measured value in the tolerance interval [13, 15]. When a measurement result is exactly on the tolerance limit (assuming a symmetric normal distribution of the measurements), the probability of being outside the tolerance limit could be as high as 50%. Hence, there would be a 50% chance of an incorrect decision [13, 15]. Either of these probabilities can be reduced by choosing acceptance limits that offset the tolerance limits. The risk of accepting a non-conforming item can be reduced by setting an acceptance limit *AL* inside the tolerance interval (guarded acceptance decision rule), and the risk of rejecting a conforming item can be reduced by setting an acceptance limit *AL* outside the tolerance interval (guarded rejection decision rule), as shown in Figure 1a and 1b, respectively. For a guarded acceptance decision rule, the guard band is w > 0, and for a guarded rejection decision rule, the guard band is w < 0. It is not possible to set the acceptance limits to minimize both the consumer's and producer's risks simultaneously. Decreasing one will increase the other [13]. Very importantly for the proper definition of a decision rule, the following question must be answered: What should be proved by the conformity assessment: compliance or non-compliance with a specification? Based on the answer, either the supplier's risk (α) or the consumer's risk (β) has to be specified [14]. Binary decision rules, acting to reduce the producer's risk (supplier's risk), will always increase the consumer's risk [15].

In conformity assessment, a binary decision rule exists when the result is limited to two choices (pass or fail), and a non-binary decision rule exists when multiple terms can express the result (pass, conditional pass, conditional fail, fail) [15].

3.2.1. Binary Statement for Simple Acceptance Rule (w=0). Reported as: Fail – the measured value is above the acceptance limit, $AL=TL$ and Pass – the measured value is below the acceptance limit, AL = TL [15].

3.2.2. Binary Statement with Guard Band. Reported as: Fail – the measurement result is above the acceptance limit, $AL = TL - w$ (rejection based on guard band); and Pass – the measurement result is below the acceptance limit, $AL = TL - w$ (acceptance based on guard band) [15].

3.2.3. Non-binary Statement with Guard Band. Reported as: Fail - the measured result is above the tolerance limit added to the guard band, $TL + w$; Pass - the measured result is below the acceptance limit, $AL = TL - w$; Conditional Fail - the measured result is above the tolerance limit but below the tolerance limit added to the guard band, in the interval $[TL, TL + w]$; Conditional Pass - the measured result is inside the guard band and below the tolerance limit, in the interval $[TL - w, TL]$ [15].

4. Conclusion

Awareness of the need to evaluate the measurement uncertainties is increasing in particular within the framework of accreditation. Evaluation of measurement uncertainty has still not matured equally well in all areas of testing, and many laboratories feel that uncertainty estimation is a laborious and intellectually challenging task to perform. Incorrect estimation the uncertainty associated with the result of conformity assessment measurement will lead to an incorrect accept/reject decision, so measurement uncertainty has to be considered and judged to be acceptable for the intended purpose. The need for statements of conformity with specifications has developed greatly, together with documents on the concept of decision rules used to make such statements. Decision rules need to be compatible with the customer, regulation or standard requirements. Where choices of decision rules are available, customers and laboratories will need to discuss levels of risk regarding the probability of false acceptance and false rejects associated with available decision rules. Customers are not sufficiently informed and involved in the choice of decision rules.

Acknowledgement

This research was supported by Ministry of Education, Science and Technological Development of the Republic of Serbia, Grant number TR31088 and TR31085.

References

[1] ILAC-G8:09:2019 Guidelines on Decision Rules and Statements of Conformity available from https://ilac.org/publications-and-resources/ilac-guidance-series/
[2] ISO/IEC 17025:2017 General requirements for the competence of testing and calibration laboratories
[3] ISO/IEC Guide 98-3:2008 Uncertainty of measurement — Part 4: Role of measurement uncertainty in conformity assessment, document is also available as JCGM 106:2012 Evaluation of measurement data – The role of measurement uncertainty in conformity assessment available from www.BIPM.org
[4] VIM International vocabulary of basic and general terms in metrology second edition 1993, International Organization for Standardization Geneva, Switzerland
[5] EUROLAB Technical Report No. 1/2017 Decision rules applied to conformity assessment available from https://www.eurolab.org
[6] Puydarrieux S, Pou J M, Leblond L, Fischer N, Allard A, Feinberg M, and Guennouni D E 2019 Role of measurement uncertainty in conformity assessment 19th International Congress of Metrology, 16003 available from https://doi.org/10.1051/metrology/201916003
[7] ILAC-G17:01/2021 Guidelines for Measurement Uncertainty in Testing available from www.ilac.org
[8] Ellison S L R and Williams A 3rd edition, 2012 Eurachem/CITAC guide: Quantifying

Uncertainty in Analytical Measurement, available from www.eurachem.org.
[9] Guide to the Expression of Uncertainty in Measurement. ISO, Geneva 1993 Reissued as ISO Guide 98-3 2008, also available from www.bipm.org as JCGM 100:2008 Guide to the Expression of Uncertainty in Measurement
[10] Magnusson B, Näykki T, Hovind H, Krysell M and Sahlin E (ed. 4) 2017 Handbook for calculation of measurement uncertainty in environmental laboratories, Nordtest Report TR 537 available from www.nordtest.info.
[11] ISO 21748:2017, Guide to the use of repeatability, reproducibility and trueness estimates in measurement uncertainty estimation. ISO, Geneva available from https://www.iso.org/standard/71615.html
[12] EA-4/16:2003 EA guideline on the expression of uncertainty in quantitative testing available from www.european-accreditation.org
[13] Joint Committee for Guides Metrology, ISO/IEC Guide 98-4:2012 Evaluation of measurement data – The role of measurement uncertainty in conformity assessment
[14] EUROLAB Technical Report No.1/2017 - Decision rules applied to conformity assessment available from https://www.eurolab.org
[15] ILAC. Guidelines on Decision Rules and Statements of Conformity. G8:09/2019 available from https://ilac.org/publications-and-resources/ilac-guidance-series