CHAPTER 2.2.4.

MEASUREMENT UNCERTAINTY

INTRODUCTION

The WOAH Validation Recommendations provide detailed information and examples in support of the Chapter 1.1.6 Validation of diagnostic assays for infectious diseases of terrestrial animals*. Estimation of measurement uncertainty* (*MU*) *is a requirement for testing laboratories based on international quality standards such as ISO/IEC 17025-2017 General requirements for the competence of testing and calibration laboratories. The measurement process for detection of an analyte in a diagnostic sample is not entirely reproducible and hence there is no exact value that can be associated with the measured analyte. Therefore, the result is most accurately expressed as an estimate with an associated imprecision level. This imprecision is the measurement uncertainty* (*MU*). *MU is limited to the measurement process of quantitative tests. The approach described here is known as "top-down" or "control sample" because it uses a weak positive control sample and expresses the MU result at the diagnostic threshold, where it most matters. It is not a question of whether the measurement is appropriate and fit for whatever use to which it may be applied. It is not an alternative to test validation but is rightly considered a component of that process* (*see chapter 1.1.6 Section B.1.1* Repeatability).

A. THE NECESSITY OF DETERMINING MU

To assure compliance with ISO/IEC 17025-2017 requirements, national accreditation bodies for diagnostic testing laboratories require laboratories to calculate MU estimates for accredited test methods that produce quantitative results, e.g. optical densities (OD), percentage of positivity or inhibition (PP, PI), titres, cycle threshold (CT) values, etc. This includes tests where numeric results are calculated and then expressed as a positive or negative result at a cut-off value. For the purpose of estimating MU in serology and reverse transcriptase polymerase chain reaction (RT-PCR), suitable statistical measures are mean target values ± 2 standard deviations (SD), which is an approximate 95%reference interval (RI), relative standard deviation (RSD = SD / mean of replicates) and coefficient of variation (CV = RSD × 100%). Examples provided below assume normal distribution of data. Alternative methods are available that are less sensitive to both that assumption and to the presence of outliers; they are not illustrated here. The concept of MU does not apply to strictly binary (qualitative) results (positive or negative).

1. Samples for use in determining MU

Repeatability is the level of agreement between results of replicates of a sample both within and between runs of the same test method in a given laboratory. During assay development, repeatability is estimated by evaluating variation in results of independent replicates from a minimum of three (preferably five) samples representing analyte activity within the operating range of the assay (see Chapter 1.1.6 *Validation of diagnostic assays for infectious diseases of terrestrial animals*, Sections A.2.5 *Robustness* and B.1.1 *Repeatability*, and Chapter 2.2.6 *Selection and use of reference samples and panels*, Section A.4.2). Typically, the variation in replicate results is expressed as RSD or CV. The significant feature is that repeatability studies can be used to define the expected precision of the assay in the detection of a range of analyte concentrations.

The use of internal quality or process controls over a range of expected results has become part of daily quality control and quality assurance operations of accredited facilities (see chapter 1.1.6, Sections A.2.6 *Calibration of the assay to standard reagents* and B.5.1 *Monitoring the assay*, and Chapter 2.2.6, Section C.1). These results provide a continuous monitor relative to different aspects of repeatability, e.g. intra- and inter-assay variation, intra- and inter-operator variation and intra- and inter-batch variation, which, when subjected to statistical analysis, provide an expression of the level of robustness (precision) of a test procedure. The monitoring of assay quality control parameters for repeatability provides evidence that the assay is or is not performing as expected. For control

samples to provide valid inferences about assay precision, they should be treated in exactly the same way as test samples in each run of the assay, e.g. including sample preparation such as extraction steps or dilution of serum samples for an antibody enzyme-linked immunosorbent assay (ELISA).

The variation of the results for control samples can also be used as an estimate of those combined sources of uncertainty and is called the "top-down" approach. This approach recognises that the components of precision will be manifest in the ultimate measurement. So monitoring the precision of the measurement over time will effectively show the combined effects of the imprecision associated with component steps.

The imprecision or uncertainty of the measurement process associated with a test result becomes increasingly more important the closer the test value is to the diagnostic cut-off value. This is because an interpretation is made relative to the assay threshold regarding the status of the test result as positive, negative, or inconclusive (as will be described in the following example). In this context, weak positive samples, like those used in repeatability studies or as the weak positive control, are most appropriate for estimation of MU. The rationale being that MU, which is a function of assay precision, is most critical at decision-making points (i.e. thresholds or cut-offs), which are usually near the lower limit of detection for the assay. In this chapter, the application of MU with respect to cut-off (threshold) values, whether recommended by test-kit manufacturers or determined in the diagnostic laboratory, is described.

MU, using the top-down approach, ideally requires long-term accumulated data from a weak positive control sample after multiple test runs over time, with multiple operators and variable conditions. The examples given below are based on 10 data points but higher numbers will increase robustness.

2. Example of MU calculations in ELISA serology

For most antibody detection tests, it is important to remember that the majority of tests are measurements of antibody activity relative to a threshold against which a dichotomous interpretation of positive or negative is applied. This is important because it helps to decide where application of MU is appropriate. In serology, uncertainty is frequently most relevant at the threshold between positive and negative determinations. Results falling into this zone are also described as intermediate, inconclusive, suspicious or equivocal (see chapter 1.1.6, Section B.2.4 *Selection of a cut-off* (*threshold*) *value for classification of test results*).

A limited data set from a competitive ELISA for antibody to avian influenza virus is used as an example of a "top-down" approach for serology. A weak positive control sample was used to calculate MU at the cut-off level[1].

2.1. Method of expression of MU

As the uncertainty is to be estimated at the threshold, which is not necessarily the reaction level of the weak positive control serum, the relative standard deviation (RSD), or coefficient of variation (CV), if expressed as a percentage, provides a convenient transformation:

$$RSD\ (X) = SD\ (X) / \text{mean}\ (\bar{X})$$

X represents the set of replicates

To simplify assessment, a suitably transformed result (such as sample-to-positive ratio, per cent inhibition, or background-corrected optical density) is regarded as the assay output result, which is then averaged across the number of replicates ($\bar{X}$). In the case of this example, a competitive ELISA, results are "normalised" (as defined in chapter 1.1.6, Section A.2.7 *'Normalising' test results to a working standard*) to a working standard by forming a ratio of all optical density (OD) values to the OD result of a non-reactive (negative) control (OD_N). This ratio is subtracted from 1 to set the level of antibody activity on a positive correlation scale; the greater the level, the greater the calculated value. This adjusted value is expressed as a per cent and referred to as the percentage inhibition or PI value. So for the weak positive control serum (OD_W), the transformation to obtain the per cent inhibition values for the weak positive control (PI_W) is:

$$PI_W = 100 \times [1 - \{OD_W / OD_N\}]$$

1 The Australian Government, Department of Agriculture, Fisheries and Forestry, has compiled worked examples for a number of diagnostic tests Available online at: https://www.agriculture.gov.au/agriculture-land/animal/health/laboratories/tests/measurement-uncertainty (accessed 22 June 2023)

The relative standard deviation becomes:

$$RSD\ (PI_W) = SD\ (PI_W)/\ mean\ (PI_W)$$

2.2. Example

A limited data set for the AI competitive ELISA example is shown below. In the experiment, the operator tested the weak positive control serum ten times in the same run. Ideally in the application of this "top down" method, a larger data set would be used, which would enable accounting for effects on precision resulting from changes in operator and assay components (other than only the control serum).

Table 1. Top-down or control sample approach for an influenza antibody C-ELISA

Test	*PI(%)*
1	56
2	56
3	61
4	64
5	51
6	49
7	59
8	70
9	55
10	42

Mean PI = 56.3; Std Dev (SD) = 7.9; Assays (*n*) = 10

2.3. Calculating uncertainty

From the limited data set,

$$RSD\ (PI_W) = SD/Mean\ PI = 7.9/56.3 = 0.14\ (or\ as\ coefficient\ of\ variation = 14\%)$$

Expanded uncertainty (U) is the statistic defining the interval within which the value of the measure and is believed to lie within a specified level of confidence, usually 95%. Expanding the uncertainty is done by multiplying the RSD (PI_W) by a factor of 2; this allows the calculation of an approximate 95% reference interval around the threshold value (in this case at PI = 50%), assuming normally distributed data. If data are not normally distributed they must be transformed to fit a normal distribution using a log scale.

$$U\ (95\%RI) = 2 \times RSD = 0.28$$

This estimate can then be applied at the threshold level

$$95\%\ RI = 50 \pm (50 \times 0.28) = 50 \pm 14\%$$

2.4. Interpretation of the results

A sample with a PI between 36% and 64% is within the MU surrounding the threshold value, and thus its diagnostic status is less certain than those of samples with results further from that threshold. This zone of lower confidence may correlate with the "grey zone" or "inconclusive/suspect zone" for interpretation that should be established for all tests (Greiner *et al.*, 1995).

3. Example of MU calculation in molecular tests

3.1. Example

For real-time PCRs, replicates of positive controls with their respective cycle threshold (Ct) values can be used to estimate MU using the top-down approach (Newberry & Colling, 2021). The method of expression follows the same formula as for the ELISA example above. This example uses data from replicate runs of a weak positive control sample (10 runs) of an equine influenza hydrolysis probe assay.

Table 2. Top-down or control sample approach for an equine influenza TaqMan A assay

Test	*Ct value*
1	33.60
2	33.20
3	33.96
4	33.18
5	33.96
6	32.72
7	33.57
8	33.45
9	32.80
10	33.20

Mean = 33.36; Std Dev (SD) = 0.43; Assay n=10

3.2. Calculating uncertainty

From the limited data set,

$$\text{RSD } (PI_W) = \text{SD/Mean } 0.43/33.36 = 0.0128 \text{ (or as coefficient of variation = 1.28\%)}$$

Expanded uncertainty (U) is the statistic defining the interval within which the value of the measure and is believed to lie within a specified level of confidence, usually 95%. Expanding the uncertainty is done by multiplying the RSD (PI_W) by a factor of 2; this allows the calculation of an approximate 95% confidence interval around the threshold value (in this case at Ct value = 37), assuming normally distributed data.

$$U\ (95\%RI) = 2 \times RSD = 0.0255$$

This estimate can then be applied at the threshold level

$$95\%\ RI = 37 \pm (37 \times 0.0255) = 37 \pm 0.94$$

The mean cycle threshold (Ct) value after 10 runs is 33.36 and the standard deviation is 0.43. The relative standard deviation is 0.0128. The expanded uncertainty (95% RI) is 2 × the relative standard deviation = 0.0255. Measurement of uncertainty (MU) is most relevant at the cut-off (Ct = 37) and can be applied by multiplication (37 × 0.0255 = 0.94). Subtraction from the threshold (37-0.94) provides the lower 95% reference limit (Ct = 36.06) and addition (37+0.94) the upper 95% reference limit (Ct = 37.94).

3.3. Interpretation of the results

A sample with a Ct between 36 and 38 is within the MU surrounding the threshold value, and thus its diagnostic status is less certain than those of samples with results further from that threshold.

B. OTHER APPLICATIONS

The top-down approach should be broadly applicable to a range of diagnostic tests including molecular tests. For the calculation of tests using a typical two-fold dilution series for the positive control such as virus neutralisation, complement fixation and haemagglutination inhibition tests geometric mean titre (i.e. geomean and expanded [SD] of log base 2 titre values) of the positive control serum should be calculated. Relative standard deviations based on these log scale values may then be applied at the threshold (log base 2) titre, and finally transformed (by antilog) to represent the uncertainty at the threshold. However, in all cases, the approach assumes that the variance about the positive control used to estimate the RSD is proportionally similar at the point of application of the MU, for example at the threshold. If the RSD varies significantly over the measurement scale, the positive control serum used to estimate the MU at the threshold should be selected for an activity level close to that threshold. The Australian Government, Department of Agriculture, Fisheries and Forestry, has compiled worked examples for a number of

diagnostic tests (see footnote 1). Other approaches and variations have been described, i.e. for serological tests (Dimech *et al.,* 2006; Goris *et al.,* 2009; Toussaint *et al.,* 2007). Central documents to MU are the Guide to the expression of uncertainty in measurement (GUM), ISO/IEC Guide, 1995 and Eurachem/CITAC Guide, 2012 CG 4: Quantifying uncertainty in analytical measurement.

Scope and limitations of the top-down approach

Methods for quantifying uncertainty (addressing MU) for tests vary. When estimating MU for quantitative, biologically based diagnostic tests, where variations in the substrate or matrix have large and unpredictable effects, a top-down approach is recommended (Dimech *et al.,* 2006; Eurachem 2012; Goris *et al.,* 2009; ISO/IEC Guide 98-3:2008; Newberry & Colling, 2021; Standards Council of Canada, 2021; and footnote 1). The advantage of this method is that quality control data are generated during normal test runs and can be used to estimate the precision of the assay and express it at the cut-off. The application at the cut-off depends on the performance of the test at different analyte concentrations, e.g. variation is likely to increase at higher diluted samples. The top-down approach does not identify individual contributors to measurement uncertainty but rather provides an overall estimate. Measurement uncertainty does not replace test validation; however, the validation process includes assessments of repeatability through quality control samples which facilitate calculation of MU.

REFERENCES

DIMECH W., FRANCIS B., KOX J. & ROBERTS G. (2006). Calculating uncertainty of measurement for serology assays by use of precision and bias. *Clin. Chem*., **52**, No. 3, 526–529.

ELLISON S.L.R. & WILLIAMS A. (EDS) (2012). Eurachem/CITAC Guide CG 4: Quantifying uncertainty in analytical measurement, 3rd Edition. Eurachem, Teddington, United Kingdom, Section 3. Available at: www.eurachem.org/images/stories/Guides/pdf/QUAM2012_P1.pdf (Accessed 10 June 2023).

GORIS N., VANDENBUSSCHE F., HERR, C., VILLERS, J., VAN DER STEDE, Y. & DE CLERCQ K. (2009). Validation of two real-time PCR methods for foot-and-mouth disease diagnosis: RNA-extraction, matrix effects, uncertainty of measurement and precision. *J. Virol. Methods,* **160**, 157–162.

GREINER M., SOHR D. & GOEBEL P.A. (1995). Modified ROC analysis for the selection of cut-off values and the definition of intermediate results of serodiagnostic tests, *J. Immunol. Methods*, **185**, 123–132.

ISO/IEC 17025:2017. General Requirements for the Competence of Testing and Calibration Laboratories. International Organization for Standardization (ISO), www.iso.org.

ISO/IEC GUIDE 98-3:2008 (E). Uncertainty of measurement – Part 3 Guide to the expression of uncertainty in measurement (GUM:1995), International Organization for Standardization (ISO), www.iso.org (accessed 10 June 2023).

NEWBERRY K. & COLLING A (2021). Quality standards and guidelines for test validation for infectious diseases in veterinary laboratories. *Rev. Sci. Tech. Off. Int. Epiz*., 40, 227–237. Available at https://doc.woah.org/dyn/portal/index.xhtml?page=alo&aloId=41245 (accessed 22 June 2023)

SCC- STANDARDS COUNCIL OF CANADA. Requirements and Guidance for method Verification and Validation in Testing Laboratories (RG-MVVT), 2021-05-24. http://www.scc.ca/ (accessed 10 June 2023).

TOUSSAINT J.F., ASSAM P., CAIJ B., DEKEYSER F., KNAPEN K., IMBERECHTS H., GORIS N., MOLENBERGHS G., MINTIENS K., & DE CLERCQ K. (2007). Uncertainty of measurement for competitive and indirect ELISAs. *Rev. sci. tech. Off. int. Epiz.,* **26**, 649–656.

*
* *

NB: There is a WOAH Collaborating Centre for
Diagnostic Test Validation Science in the Asia-Pacific Region (please consult the WOAH Web site:
https://www.woah.org/en/what-we-offer/expertise-network/collaborating-centres/#ui-id-3).
Please contact the WOAH Collaborating Centre for any further information on validation.

NB: FIRST ADOPTED IN 2014. MOST RECENT UPDATES ADOPTED IN 2024.