CHAPTER 2.2.4.

MEASUREMENT UNCERTAINTY

INTRODUCTION

The WOAH Validation Recommendations in Section 2.2 Validation of diagnostic tests *of this* Terrestrial Manual *provide detailed information and examples in support of the WOAH Validation Standard that is published as Chapter 1.1.6* Validation of diagnostic assays for infectious diseases of terrestrial animals. *The Term "WOAH Validation Standard" in this chapter should be taken as referring to that chapter.*

Estimation of measurement uncertainty (*MU*)*, sometimes termed measurement imprecision, is a requirement for testing laboratories based on international quality standards such as ISO/IEC 17025-2005, General requirements for the competence of testing and calibration laboratories* (*ISO/IEC 17025*)*. 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 together with an associated level of imprecision. This imprecision is the measurement uncertainty* (*MU*)*. MU is limited to the measurement process. 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 the WOAH Validation Standard, Section B.1.1*).

A. THE NECESSITY OF DETERMINING MU

To assure compliance with ISO/IEC 17025-2005 requirements, national accreditation bodies for diagnostic laboratories require MU estimates for 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 are expressed as a positive or negative result at a cut-off value. For the purpose of estimating MU in serology and RT PCR, suitable statistical measures are mean target values ± 2 standard deviations (SD), which is approximately equal to a 95% confidence interval (CI), relative standard deviation (RSD = SD / mean of replicates) and coefficient of variation (CV = RSD × 100%). The concept of MU does not apply to strictly binary 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 the WOAH Validation Standard, Sections A.2.5 and B.1.1, and Chapter 2.2.6 *Selection and use of reference samples and panels*, Section 3.1). 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 the WOAH Validation Standard, Sections A.2.6 and B.5.1, and Chapter 2.2.6, Section 1.4). 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. In order 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, low positive samples, like those used in repeatability studies or as the low 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) 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.

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 the WOAH Validation Standard, Section B.2.4).

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 low positive control sample was used to calculate MU at the cut-off level.

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 low positive control serum, the relative standard deviation (RSD), or coefficient of variation (CV), if expressed as a percentage, provides a convenient transformation:

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

To simplify assessment, the transformed result is regarded as the assay output result, which is the averaged across the number of replicates ($\bar{X}$). In the case of this example, a competitive ELISA, results are "normalised" (as defined in the WOAH Validation Standard, Section A.2.7) 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 low positive control serum (OD_L), the transformation to obtain the per cent inhibition values for the low positive control (PI_L) is:

$$PI_L = 100 \times [1 - \{OD_L / OD_N\}]$$

The relative standard deviation becomes:

$$RSD\ (PI_L) = SD\ (PI_L) / (PI_L)$$

2.2. Example

A limited data set for the AI competitive ELISA example is shown below. In the experiment, the operator tested the low 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).

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_L) = SD/Mean 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_L) by a factor of 2; this allows the calculation of an approximate 95% confidence interval around the threshold value (in this case at PI = 50%), assuming normally distributed data.

U (95%CI) = 2 × RSD = 0.28

This estimate can then be applied at the threshold level

95% CI = 50 ± (50 × 0.28) = 50 ± 14%

2.4. Interpretation

Any positive result (PI > 50%) that is less than 64% is not positive with 95% confidence. Similarly, a negative result (PI < 50%) that is higher or equal to a PI of 36 is not negative at the 95% confidence level. 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*).

B. OTHER APPLICATIONS

The top-down approach should be broadly applicable for 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. mean and 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) titre, and finally transformed 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 and Water Resources, has compiled worked examples for a number of diagnostic tests, which are available online at:

http://www.agriculture.gov.au/animal/health/laboratories/tests/worked-example-measurement

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.

Other approaches and variations have been described, i.e. for serological tests (Dimech *et al.,* 2006; Goris *et al.,* 2009; Toussaint *et al.,* 2007). Additional work and policy documents are available from the National Pathology Accreditation Advisory Group and Life Science. The central document to MU is the Guide to the expression of uncertainty in measurement (GUM), ISO/IEC Guide (1995).

REFERENCES

AMERICAN ASSOCIATION FOR LABORATORY ACCREDITATION (A2LA). Policy on estimating measurement uncertainty for life science testing labs https://portal.a2la.org/policies/A2LA_P103b.pdf (accessed 22 November 2018).

AUSTRALIAN GOVERNMENT, DEPARTMENT OF AGRICULTURE AND WATER RESOURCES. Worked of measurement uncertainty http://www.agriculture.gov.au/animal/health/laboratories/tests/worked-example-measurement (accessed 22 November 2018).

AUSTRALIAN GOVERNMENT DEPARTMENT OF HEALTH AND AGEING (2007). Requirements for the estimation of measurement uncertainty, National Pathology Accreditation Advisory Group. http://www.health.gov.au/internet/main/publishing.nsf/Content/B1074B732F32282DCA257BF0001FA218/$File/dhaeou.pdf (accessed 22 November 2018).

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.

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 (1995). Guide to the expression of uncertainty in measurement (GUM), ISO/IEC Guide 98:1995. International Organization for Standardization (ISO), www.iso.org.

ISO/IEC (2005). ISO/IEC 17025:2005. General Requirements for the Competence of Testing and Calibration Laboratories. International Organization for Standardization (ISO), www.iso.org.

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: FIRST ADOPTED IN 2014.