**Annex 6. Item 5.1. – Chapter 2.2.4 Measurement uncertainty**

2 MEETING OF THE WOAH BIOLOGICAL STANDARDS COMMISSION

3 **Paris, 4–8 September 2023**

4

**USA COMMENTS IN RED FONT**

5

**CHAP T E R 2 . 2 . 4 .**

# 6 M E A S U RE M EN T U N CE RT A I N T Y

## 7 INTRODUCTION

1. *The WOAH Validation Recommendations provide detailed information and examples in support of*
2. *~~WOAH Validation Standard that is published as~~ Chapter 1.1.6* ~~Principles and methods of~~ Validation of
3. diagnostic assays for infectious diseases of terrestrial animals *~~this~~* ~~Terrestrial Manual~~*~~, or Chapter 1.1.2 of~~*
4. *~~the~~* ~~Aquatic Manual~~*. ~~The Term “WOAH Validation Standard” in this chapter should be taken as referring to~~*
5. *~~those chapters.~~*
6. *Estimation of measurement uncertainty* (*MU*)*~~, sometimes termed measurement imprecision,~~ is a*
7. *requirement for testing laboratories based on international quality standards such as ISO/IEC 17025-~~2005,~~*
8. *2017 General requirements for the competence of testing and calibration laboratories* ~~(~~*~~ISO/IEC 17025~~*~~)~~. *The*
9. *measurement process for detection of an analyte in a diagnostic sample is not entirely reproducible and*
10. *hence there is no exact value that can be associated with the measured analyte. Therefore, the result is*
11. *most accurately expressed as an estimate ~~together~~ with an associated ~~level of~~ imprecision level. This*
12. *imprecision is the measurement uncertainty* (*MU*). *MU is limited to the measurement process of quantitative*
13. *tests. The approach described here is known as “top-down” or “control sample” because it uses a weak*
14. *positive control sample and expresses the MU result at the ~~cut-off~~ diagnostic threshold, where it most matters.*
15. *of whether the measurement is appropriate and fit for whatever use to which it may be applied. It is not an*
16. *alternative to test validation but is rightly considered a component of that process* (*see ~~the WOAH Validation~~*
17. *~~Standard,~~ chapter 1.1.6 Section B.1.1* Repeatability)*.*

**RATIONALE:** Replaced cutoff with threshold for consistency with verbiage used later in the document.

## 25 A. THE NECESSITY OF DETERMINING MU

1. To assure compliance with ISO/IEC 17025-~~2005~~ 2017 requirements, national accreditation bodies for diagnostic testing
2. laboratories require laboratories to calculate MU estimates for accredited test methods that produce quantitative results,
3. e.g. optical densities (OD), percentage of positivity or inhibition (PP, PI), titres, cycle threshold (CT) values, etc. This
4. includes tests where numeric results are calculated and then ~~are~~ expressed as a positive or negative result at a cut-off
5. value. For the purpose of estimating MU in serology and reverse transcriptase polymerase chain reaction (RT-PCR),
6. suitable statistical measures are mean target values ± 2 standard deviations (SD), ~~which is approximately equal to a 95%~~
7. ~~confidence interval (CI),~~ relative standard deviation (RSD = SD / mean of replicates) and coefficient of variation (CV = RSD
8. × 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
9. (qualitative) results (positive or negative).

**RATIONALE:** This is not approximately equal to a confidence interval, even for normally distributed data. (A confidence interval is based on the standard error, not standard deviation.) Rather, this is an approximate 95% reference interval. See discussion in Sec. 15.5 of Martin Bland (2000), An Introduction to Medical Statistics, 3d ed. (Oxford University Press). Note also that ISO/IEC Guide 98-3, Guide to the Expression of Uncertainty in Measurement (2008), Sec. 6.2.2, states that they deliberately do not use the term "confidence" in relation to this interval.

Alternative methods are less reliant on distributional assumptions, and better handles the presence of outliers. See for example Maronna, Martin, Yohai, and Salibian-Barerra (2019), Robust Statistics: Theory and Methods, 2d ed. (Wiley).

#### 1. Samples for use in determining MU

1. Repeatability is the level of agreement between results of replicates of a sample both within and between runs of the same
2. test method in a given laboratory. During assay development, repeatability is estimated by evaluating variation in results
3. of independent replicates from a minimum of three (preferably five) samples representing analyte activity within the
4. operating range of the assay (see ~~the WOAH Validation Standard,~~ Chapter 1.1.6 *Validation of diagnostic assays for*
5. *infectious diseases of terrestrial animals*, Sections A.2.5 *Robustness* and B.1.1 *Repeatability*, and Chapter 2.2.6 *Selection*
6. *and use of reference samples and panels*, Section ~~3.1~~ A.4.2). Typically, the variation in replicate results is expressed as
7. RSD or CV. The significant feature is that repeatability studies can be used to define the expected precision of the assay
8. in the detection of a range of analyte concentrations.
9. The use of internal quality or process controls over a range of expected results has become part of daily quality control
10. and quality assurance operations of accredited facilities (see ~~the WOAH Validation Standard,~~ chapter 1.1.6, Sections A.2.6
11. *Calibration of the assay to standard reagents* and B.5.1 *Monitoring the assay*, and Chapter 2.2.6, Section ~~1.4~~ C.1). These
12. results provide a continuous monitor relative to different aspects of repeatability, e.g. intra- and inter-assay variation, intra-
13. and inter-operator variation and intra- and inter-batch variation, which, when subjected to statistical analysis, provide an
14. expression of the level of robustness (precision) of a test procedure. The monitoring of assay quality control parameters
15. for repeatability provides evidence that the assay is or is not performing as expected. For control samples to provide valid
16. inferences about assay precision, they should be treated in exactly the same way as test samples in each run of the assay,
17. e.g. including sample preparation such as extraction steps or dilution of serum samples for an antibody enzyme-linked
18. immunosorbent assay (ELISA).
19. The variation of the results for control samples can also be used as an estimate of those combined sources of uncertainty
20. and is called the “top-down” approach. This approach recognises that the components of precision will be manifest in the
21. ultimate measurement. So monitoring the precision of the measurement over time will effectively show the combined effects
22. of the imprecision associated with component steps.
23. The imprecision or uncertainty of the measurement process associated with a test result becomes increasingly more
24. important the closer the test value is to the diagnostic cut-off value. This is because an interpretation is made relative to
25. the assay threshold regarding the status of the test result as positive, negative, or inconclusive (as will be described in the
26. following example). In this context, ~~low~~ weak positive samples, like those used in repeatability studies or as the low weak
27. positive control, are most appropriate for estimation of MU. The rationale being that MU, which is a function of assay
28. precision, is most critical at decision-making points (i.e. thresholds or cut-offs), which are usually near the lower limit of
29. detection for the assay. In this chapter, the application of MU with respect to cut-off (threshold) values, whether
30. recommended by test-kit manufacturers or determined in the diagnostic laboratory, is described.
31. MU, using the top-down approach, ideally requires long-term accumulated data from a weak positive control sample after
32. multiple test runs over time, with multiple operators and variable conditions. The examples given below are based on 10
33. data points but higher numbers will increase robustness.

#### 2. Example of MU calculations in ELISA serology

1. For most antibody detection tests, it is important to remember that the majority of tests are measurements of antibody
2. activity relative to a threshold against which a dichotomous interpretation of positive or negative is applied. This is important
3. because it helps to decide where application of MU is appropriate. In serology, uncertainty is frequently most relevant at
4. the threshold between positive and negative determinations. Results falling into this zone are also described as
5. intermediate, inconclusive, suspicious or equivocal (see ~~the WOAH Validation Standard,~~ chapter 1.1.6, Section B.2.4
6. *Selection of a cut-off* (*threshold*) *value for classification of test results*).
7. A limited data set from a competitive ELISA for antibody to avian influenza virus is used as an example of a “top-down”
8. approach for serology. A ~~low~~ weak positive control sample was used to calculate MU at the cut-off level [32](#_bookmark107).

##### 2.1. Method of expression of MU

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

\_

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-](https://www.agriculture.gov.au/agriculture-land/animal/health/laboratories/tests/measurement-uncertainty) [uncertainty](https://www.agriculture.gov.au/agriculture-land/animal/health/laboratories/tests/measurement-uncertainty) (accessed 22 June 2023)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| RSD | (X) | = | SD | (X) | / mean(X), |

1. where X represents the set of replicates. To simplify assessment, ~~the~~ a suitably transformed result (such as sample-to-positive ratio, percent inhibition, or background-corrected optical density) is regarded as the assay output result, which is then averaged across
2. the number of replicates (�X). In the case of this example, a competitive ELISA, results are “normalised” (as defined
3. in ~~the WOAH Validation Standard,~~ chapter 1.1.6, Section A.2.7 *‘Normalising’ test results to a working standard*) to a
4. working standard by forming a ratio of all optical density (OD) values to the OD result of a non-reactive (negative)
5. control (ODN). This ratio is subtracted from 1 to set the level of antibody activity on a positive correlation scale; the
6. greater the level, the greater the calculated value. This adjusted value is expressed as a per cent and referred to as
7. the percentage inhibition or PI value. So for the low weak positive control serum (ODL), the transformation to obtain
8. the per cent inhibition values for the ~~low~~ weak positive control (PIL) is:

##### RATIONALE: X needs to be defined, and the equation has been corrected. Clarified what is meant by transformed above, and the equation below.

90 PIL = 100 × [1– {ODL/ ODN}]

91 The relative standard deviation becomes:

92

##### 2.2. Example

RSD (PIL) = SD (PIL)/ Mean(PIL)

1. A limited data set for the AI competitive ELISA example is shown below. In the experiment, the operator tested the
2. low weak positive control serum ten times in the same run. Ideally in the application of this “top down” method, a
3. larger data set would be used, which would enable accounting for effects on precision resulting from changes in
4. operator and assay components (other than only the control serum).
5. ***Table 1.*** *Top-down or control sample approach for an influenza antibody C-ELISA*

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

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

##### 2.3. Calculating uncertainty

1. From the limited data set,
2. RSD (PIL) = SD/Mean = 7.9/56.3 = 0.14 (or as coefficient of variation = 14%)
3. Expanded uncertainty (U) is the statistic defining the interval within which the value of the measure and is believed
4. to lie within a specified level of confidence, usually 95%. Expanding the uncertainty is done by multiplying the RSD
5. (PIL) by a factor of 2; this allows the calculation of an approximate 95% ~~confidence~~ reference interval around the threshold value
6. (in this case at PI = 50%), assuming normally distributed data.
7. U (95%~~C~~RI) = 2 × RSD = 0.28
8. This estimate can then be applied at the threshold level
9. 95% ~~C~~RI = 50 ± (50 × 0.28) = 50 ± 14%

**RATIONALE:** Missing an equals sign. Changed confidence to reference based on previous comment above referencing Bland's textbook. Changed notation from CI to RI for reference interval. A confidence interval is based on standard error not standard deviation and is not appropriate here.

##### 2.4. Interpretation

1. ~~Any positive result (PI > 50%) that is less than 64% is not positive with 95% confidence. Similarly, a negative result~~
2. ~~(PI < 50%) that is higher or equal to a PI of 36 is not negative at the 95% confidence level.~~ 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
3. confidence may correlate with the “grey zone” or “inconclusive/suspect zone” for interpretation that should be
4. established for all tests (Greiner *et al., 1995*).

**RATIONALE:** This interpretation is too precise given the multiple approximations made and the nuances of the interpretation of a reference interval.

1. **3. Example of MU calculation in molecular tests**

##### 3.1. Example

1. For real-time PCRs, replicates of positive controls with their respective cycle threshold (CT) values can be used to
2. estimate MU using the top-down approach (Newberry & Colling, 2021). The method of expression follows the same
3. formula as for the ELISA example above. This example uses data from replicate runs of a weak positive control
4. sample (10 runs) of an equine influenza hydrolysis probe assay.
5. ***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 |

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

##### 3.2. Calculating uncertainty

1. From the limited data set,
2. RSD (PIL) = SD/Mean 0.43/33.36 = 0.0128 (or as coefficient of variation = 1.28%)
3. Expanded uncertainty (U) is the statistic defining the interval within which the value of the measure and is believed
4. to lie within a specified level of confidence, usually 95%. Expanding the uncertainty is done by multiplying the RSD
5. (PIL) by a factor of 2; this allows the calculation of an approximate 95% ~~confidence~~ reference interval around the threshold value
6. (in this case at Ct value = 37), assuming normally distributed data.
7. U (95% ~~C~~RI) = 2 × RSD = 0.0255
8. This estimate can then be applied at the threshold level
9. 95% ~~C~~RI = 37± (37 × 0.0255) = 37 ± 0.94
10. The mean cycle threshold (Ct) value after 10 runs is 33.36 and the standard deviation is 0.43. The relative standard
11. deviation is 0.0128. The expanded uncertainty (95% ~~C~~RI) is 2 × the relative standard deviation = 0.0255. Measurement
12. of uncertainty (MU) is most relevant at the cut-off (Ct = 37) and can be applied by multiplication (37 × 0.0255 = 0.94).
13. Subtraction from the threshold (37-0.94) provides the lower 95% ~~confidence~~ reference limit (Ct = 36.06) and addition (37+0.94)
14. the upper 95% ~~confidence~~ reference limit (Ct = 37.94).

**RATIONALE:** Regarding the change from confidence to reference, see above comment citing Bland 's textbook.

##### 3.3. Interpretation of the results

1. A sample with a ~~PI~~ Ct between 36% and ~~64~~37% 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. ~~Any positive result (Ct < 37) that is higher than 36 Ct is not positive with 95% confidence. Similarly, any negative~~
2. ~~result (Ct > 37) that is less than 38 Ct is not negative with 95% confidence.~~

**RATIONALE:** This interpretation is too precise given the multiple approximations made and the nuances of the interpretation of a reference interval.

## 141 B. OTHER APPLICATIONS

1. The top-down approach should be broadly applicable ~~for~~to a range of diagnostic tests including molecular tests. For the
2. calculation of tests using a typical two-fold dilution series for the positive control such as virus neutralisation, complement
3. fixation and haemagglutination inhibition tests geometric mean titre (i.e. mean and SD of log base 2 titre values) of the
4. positive control serum should be calculated. Relative standard deviations based on these log scale values may then be
5. applied at the threshold (log) titre, and finally transformed (by antilog) to represent the uncertainty at the threshold.
6. **RATIONALE:** To clarify what transformation is expected.
7. However, in all

cases, the approach assumes that the variance about the positive control used to estimate the RSD is proportionally similar

1. at the point of application of the MU, for example at the threshold. If the RSD varies significantly over the measurement
2. scale, the positive control serum used to estimate the MU at the threshold should be selected for an activity level close to
3. that threshold. The Australian Government, Department of Agriculture, Fisheries and ~~Water Resources~~Forestry, has
4. compiled worked examples for a number of diagnostic tests (see footnote 1). ~~(DAFF, 2010), which are available online at:~~
5. [~~http://www.agriculture.gov.au/animal/health/laboratories/tests/worked-example-measurement~~](http://www.agriculture.gov.au/animal/health/laboratories/tests/worked-example-measurement)
6. ~~For quantitative real-time PCRs (qPCR) replicates of positive controls with their respective cycle threshold (CT) values can~~
7. ~~be used to estimate MU using the top-down approach.~~
8. Other approaches and variations have been described, i.e. for serological tests (Dimech *et al.,* 2006; Goris *et al.,* 2009;
9. Toussaint *et al.,* 2007). ~~Additional work and policy~~ Central documents ~~are available from the National Pathology~~
10. ~~Accreditation Advisory Group and Life Science. The central document~~ to MU ~~is~~are the Guide to the expression of
11. uncertainty in measurement (GUM), ISO/IEC Guide, (1995) and Eurachem/CITAC Guide, 2012 CG 4: Quantifying
12. uncertainty in analytical measurement.

#### Scope and limitations of the top-down approach

1. Methods for quantifying uncertainty (addressing MU) for tests vary. When estimating MU for quantitative, biologically based
2. diagnostic tests, where variations in the substrate or matrix have large and unpredictable effects, a top-down approach is
3. recommended (Dimech *et al.,* 2006; Eurachem 2012; Goris *et al.,* 2009; ISO/IEC Guide 98-3:2008; Newberry & Colling,
4. 2021; Standards Council of Canada, 2021; and footnote 1). The advantage of this method is that quality control data are
5. generated during normal test runs and can be used to estimate the precision of the assay and express it at the cut-off. The
6. application at the cut-off depends on the performance of the test at different analyte concentrations, e.g. variation is likely
7. to increase at higher diluted samples. The top-down approach does not identify individual contributors to measurement
8. uncertainty but rather provides an overall estimate. Measurement uncertainty does not replace test validation; however,
9. the validation process includes assessments of repeatability through quality control samples which facilitate calculation of
10. MU.

### REFERENCES

1. ~~AMERICAN ASSOCIATION FOR LABORATORY ACCREDITATION (A2LA). Policy on estimating measurement uncertainty for life~~
2. ~~science testing labs~~ [~~https://portal.a2la.org/policies/A2LA\_P103b.pdf~~](https://portal.a2la.org/policies/A2LA_P103b.pdf) ~~(accessed 22 November 2018).~~
3. ~~Australian Government, Department of Agriculture and WATER RESOURCES.. Worked examples of measurement uncertainty~~
4. [~~http://www.agriculture.gov.au/animal/health/laboratories/tests/worked-example-measurement~~](http://www.agriculture.gov.au/animal/health/laboratories/tests/worked-example-measurement) ~~(accessed 22 November~~
5. ~~2018).~~
6. ~~AUSTRALIAN GOVERNMENT DEPARTMENT OF HEALTH AND AGEING (2007). Requirements for the estimation of measurement~~
7. uncertainty, National Pathology Accreditation Advisory Group.
8. [~~http://www.health.gov.au/internet/main/publishing.nsf/Content/B1074B732F32282DCA257BF0001FA218/$File/dhaeou.p~~](http://www.health.gov.au/internet/main/publishing.nsf/Content/B1074B732F32282DCA257BF0001FA218/%24File/dhaeou.pdf)
9. [~~df~~](http://www.health.gov.au/internet/main/publishing.nsf/Content/B1074B732F32282DCA257BF0001FA218/%24File/dhaeou.pdf) ~~(accessed 22 November 2018).~~
10. DIMECH W., FRANCIS B., KOX J. & ROBERTS G. (2006). Calculating uncertainty of measurement for serology assays by use
11. of precision and bias. *Clin. Chem*., **52**, No. 3, 526–529.
12. ELLISON S.L.R. & WILLIAMS A. (EDS) (2012). Eurachem/CITAC Guide CG 4: Quantifying uncertainty in analytical
13. measurement, 3rd Edition. Eurachem, Teddington, United Kingdom, Section 3. Available at:
14. [www.eurachem.org/images/stories/Guides/pdf/QUAM2012\_P1.pdf](http://www.eurachem.org/images/stories/Guides/pdf/QUAM2012_P1.pdf) (Accessed 10 June 2023).
15. GORIS N., VANDENBUSSCHE F., HERR, C., VILLERS, J., VAN DER STEDE, Y. & DE CLERCQ K. (2009). Validation of two real-time
16. PCR methods for foot-and-mouth disease diagnosis: RNA-extraction, matrix effects, uncertainty of measurement and
17. precision. *J. Virol. Methods,* **160**, 157–162.
18. GREINER M., SOHR D. & GOEBEL P.A. (1995). Modified ROC analysis for the selection of cut-off values and the definition of
19. intermediate results of serodiagnostic tests, *J. Immunol. Methods*, **185**, 123–132.
20. ISO/IEC ~~(1995). Guide to the expression of uncertainty in measurement (GUM), ISO/IEC Guide 98:1995. Internationa~~l
21. ~~Organization for Standardization (ISO),~~ [~~www.iso.org~~.](http://www.iso.org/)
22. ~~ISO/IEC (2005). ISO/IEC~~ 17025:~~2005~~2017. General Requirements for the Competence of Testing and Calibration
23. Laboratories. International Organization for Standardization (ISO), [www.iso.org.](http://www.iso.org/)
24. ISO/IEC GUIDE 98-3:2008 (E). Uncertainty of measurement – Part 3 Guide to the expression of uncertainty in measurement
25. (GUM:1995), International Organization for Standardization (ISO), [www.iso.org](http://www.iso.org/) (accessed 10 June 2023).
26. SCC- STANDARDS COUNCIL OF CANADA. Requirements and Guidance for method Verification and Validation in Testing
27. Laboratories (RG-MVVT), 2021-05-24. <http://www.scc.ca/>(accessed 10 June 2023) .
28. Toussaint J.F., Assam P., Caij B., Dekeyser F., Knapen K., Imberechts H., Goris N., Molenberghs G., Mintiens K., & De
29. CLERCQ K. (2007). Uncertainty of measurement for competitive and indirect ELISAs. *Rev. sci. tech. Off. int. Epiz.,* **26**, 649–
30. 656.
31. NEWBERRY K. & COLLING A (2021). Quality standards and guidelines for test validation for infectious diseases in veterinary
32. laboratories. *Rev. Sci. Tech. Off. Int. Epiz*., 40, 227–237. Available at
33. <https://doc.woah.org/dyn/portal/index.xhtml?page=alo&aloId=41245> (accessed 22 June 2023)

205 \*

206 \* \*

1. **NB:** There is a WOAH Collaborating Centre for
2. Diagnostic Test Validation Science in the Asia-Pacific Region (please consult the WOAH Web site:
3. <https://www.woah.org/en/what-we-offer/expertise-network/collaborating-centres/#ui-id-3>).
4. Please contact the WOAH Collaborating Centre for any further information on validation.
5. **NB:** First adopted in 2014.