World Organisation for Animal Health

Template for the validation report form for tests recommended in the WOAH *Terrestrial Manual*

Section 1. Guide for contributors to the *Terrestrial Manual*

Information to fill out in this form

Reference Laboratory experts contributing to *Terrestrial Manual* chapters are invited to use this template to post their validation data online. In this way a repository of validation reports for tests recommended in the *Terrestrial Manual* will be created for anyone seeking the validation data available for a given test. Before filling in this form, contributors should consult the WOAH *Manual of Diagnostic Tests and Vaccines for Terrestrial Animals* (*Terrestrial Manual*) Chapter 1.1.6 'Principles and Methods of Validation of Diagnostic Assays for Infectious Diseases' and supporting chapters 2.2.1., 2.2.2., and 2.2.3., which provide information for validation of fundamentally different assays such as for the detection of antibodies, antigens and nucleic acid. An up-to-date compilation of the relevant validation standards (WOAH and non-WOAH) and guidance documents for all stages of diagnostic test validation and proficiency testing, including design, analysis, interpretation as well as clear reporting and case studies are provided in the WOAH *Scientific and Technical Review* issue Diagnostic Test Validation Science (Vol. 40, April 2021).

Ideally information should be provided as shown in Figure 1 from chapter 1.1.6 of the WOAH *Terrestrial Manual* pertaining to intended purpose(s), optimisation and standardisation, analytical sensitivity (ASe) and analytical specificity (ASp), repeatability, cut-off, diagnostic sensitivity (DSe) and diagnostic specificity (DSp), reproducibility, and conclusion about fitness for purpose. However, what is important is not how much data can be inserted under each heading in the template but rather that validation data are presented to demonstrate that the performance of the assay has been evaluated and that it is proven fit for the chosen purpose in Table 1. It is important that validation information supports the specific purpose, e.g. a screening test would need to show high DSe and a confirmatory test high DSp to conclude fitness for purpose.

Section 2. General information

2.1. Information about the contributor

Contact person	
Job title	
Organisation	
WOAH Reference Laboratory	
Address	
Phone	
E-mail	

Organisation	
Address	
Phone	
Contact person	
Job title	
E-mail	

2.1.1. Accreditation of the test to ISO 17025 or equivalent

Double click on a check box to indicate accreditation or certification status of applicant. Select 'Checked' ☒ to indicate test is accredited. Select 'Not Checked' ☐ to indicate test not yet accredited.

Please provide link to accreditation certificate:

2.1.2. Publication(s) relating to validation of test or relevant use of test

List publications here.

2.2. Name and purpose of the test

2.2.1. Type of method

Indirect or competitive ELISA, conventional or real-time PCR, etc.

2.2.2. Commercial name (if applicable)

2.2.3. Intended purpose(s) of the test

Please select the specific purpose(s) of the test from the list of intended purposes provided below. If none of these purposes are suitable, please select "Other", and describe the purpose. Suitable data need to be provided to substantiate fitness for each selected purpose in the application.

Double click on a check box to indicate the purpose of the test. Select 'Checked' ☒ to indicate Yes. Select 'Not Checked' ☐ to indicate No. Please cross reference with Table 1 of the disease specific chapters in the Terrestrial Manual.

Intended Fitness for Purpose

☐ 1 Demonstrate freedom from infection in a defined population (country/zone/compartment/herd)

- ☐ 1a "Free" with vaccination
- ☐ 1b Historical freedom
- ☐ 1c Re-establishment of freedom after outbreaks

☐ 2 Certify freedom from infection or agent in individual animals or products for trade/movement purposes

☐ 3 Eradication of infection from defined populations

☐ 4 Confirmatory diagnosis of clinical cases (includes confirmation of suspect cases and a positive screening test)

☐ 5 Estimate prevalence of infection to facilitate risk analysis (surveys/herd health schemes/disease control)

☐ 6 Determine immune status in individual animals or populations (post-vaccination)

☐ 7 Other [please specify] for example differentiation of vaccinated from infected animals (DIVA)_____

2.3. Test description and requirements

2.3.1. Protocol of the test

Include your detailed working protocol here to allow users to set up the test in their laboratory and interpret the results. Please include targets in analytical terms (e.g. antibody isotype, gene sequence etc.), the species and specimens that can be examined and the positive and negative controls included, Include definitions of test positivity cut-offs or results categories.

For tests that are not already described in the disease-specific chapter of the *Terrestrial Manual,* please complete Sections 2.3.2–2.3.5.

2.3.2. End-user requirements (laboratory or field use)

Describe minimum laboratory requirements for optimal test performance; include laboratory biocontainment/biosafety requirements, environmental conditions (temperature and humidity ranges), equipment, etc.

2.3.3. Computational requirements (if applicable)

Describe hardware and software requirements for assay operation and data processing.

2.3.4. Test kit format (*if applicable*)

For commercial tests, outline the number of samples that can be tested with one kit. Describe any flexibility in kit formats that would accommodate various test throughput volumes (e.g. multi-well plate vs *strip formats).*

2.3.5. General precautions/ safety aspects/disposal of reagents

List potential health hazards and safety precautions.

Section 3. Development and validation of the assay

3.1. Assay development pathway or Assay modification/optimisation (as applicable: where this is an existing assay please insert reference[s])

3.1.1. Design, development, optimisation and standardisation of the assay for the intended purpose

Assay design, development, optimisation and standardisation, For guidance, refer to Chapters 2.2.1 'Development and optimisation of antibody detection assays', 2.2.2 'Development and optimisation of antigen detection assays', and 2.2.3 'Development and optimisation of nucleic acid detection assays' of the WOAH Terrestrial Manual *relevant to the type of test and also to "In Diagnostic test validation science: a key element for effective detection and control of infectious animal diseases".*

3.2. Validation pathway stage 1 – analytical characteristics

3.2.1. Stage 1. Repeatability data

Repeatability is the level of agreement between replicates of a sample both within (intra-assay) and between (inter-assay) runs of the same test method in a single laboratory. Repeatability is estimated by evaluating variation in results of replicates. The number of replicates should preferably be determined in consultation with a statistician with a suggested minimum of three different samples representing analyte activity within the operating range of the assay. Within or intra-assay variation can be assessed using three or more replicates of each sample in one run (one operator). Intra-assay and inter-assay variation can be assessed by testing the panel of samples over several days, using two or more operators.

The data/detail provided should include:

- *i) the number of different isolates use, ideally minimum of three covering analytical range of test (strong/moderate/weak)*
- *ii) the number of replicates per sample for intra-assay and inter-assay analysis*
- *iii) The number of runs performed*
- *iv) the number of different operators used at a single site*

For repeatability of PCR, all replicates are treated like individual diagnostic samples subject to individual extraction and testing. Therefore, for PCR the WOAH Terrestrial Manual *recommends using virus and including independent extraction step in each measure of repeatability, so that repeatability assesses both extraction and PCR (WOAH* Terrestrial Manual *Chapter 2.2.3.). For tissue it is recommended that aliquots of the same homogenised specimen are individually extracted and tested.*

For serology, all replicates are treated like individual diagnostic samples including preparation of any working dilutions. Also for serology, a minimum of three different sera covering analytical range (e.g. negative, weakly positive, and strongly positive control samples) should be assessed (WOAH Terrestrial Manual *Chapter 2.2.1.).*

Include statistical data where applicable, e.g. coefficients of variation or upper and lower ranges.

3.2.2. Stage 1. Analytical specificity data (as appropriate for the test type and disease)

Analytical specificity is the degree to which the assay distinguishes between the target analyte and other components in the sample matrix; the higher the analytical specificity, the lower the level of false positives. The assessment of analytical specificity is qualitative, and the choice and sources of sample types, organisms and sequences for the assessment should reflect test purpose and assay type in addition to sample source and test matrix. Analytical specificity is further characterised by determining:

- *a) selectivity, which is the extent to which a method can accurately quantify the targeted analyte in the presence of interferents, for example, a) of matrix components such as inhibitors of enzymes in the reaction mix, b) degradants (toxic factors), c) nonspecific binding of reactants to a solid phase, e.g. conjugate of an ELISA absorbed to well of microtiter plate, d) antibodies to vaccination that may be confused with antibodies to active infection*
- *b) exclusivity, laboratory testing should be carried out to determine the capacity of the assay to detect an analyte or genomic sequence that is unique to a targeted organism and exclude all other known organisms that are potentially cross-reactive. This would also define a confirmatory assay, for example an assay to detect avian influenza virus (AIV) H5 subtypes should be assessed for cross-reaction with non-H5 AIV subtypes. Specificity testing should also include other organisms that cause similar clinical signs.*
- *c) inclusivity, which is the capacity of an assay to detect several strains or serovars of a species, several species of a genus, or a similar grouping of closely related organisms or antibodies thereto. It characterises the scope of action for a screening assay (Bowden & Wang, 2021; Ludi et al., 2021).*

Include statistical data where applicable, e.g. coefficients of variation or upper and lower ranges.

3.2.3. Stage 1. Analytical sensitivity data

Analytical sensitivity is synonymous with 'Limit of Detection', the smallest detectable amount of analyte in a specified matrix that would produce a positive result with a defined certainty. An analyte may include antibodies, antigens, nucleic acids, or live organisms. The WOAH Terrestrial Manual *Chapter 2.2.1. suggests each dilution in the series should be tested in 10 replicates, however, three to five replicates are acceptable. The dilution series must extend to at least one dilution past end-point (negative/not detectable). Criteria for end-point dilution must be established, e.g. the end-point is the last dilution for which all replicates are positive. A precise estimate of ASe is often not available for assays for infectious diseases, except for PCR where it is possible to calculate the threshold number of copies of a target nucleic acid sequence that can be detected by the assay. Alternatively, it is possible to compare the limit of detection between the test under evaluation and the reference test to obtain a relative estimate for ASe (Chapter 2.2.5., Bowden & Wang, 2021).*

Include statistical data where applicable, e.g. coefficients of variation or upper and lower ranges.

3.2.4. Stage 1. Standard of comparison

For a preliminary evaluation, the standard method(s) of comparison (reference standard) should, if possible be run in parallel on a small but select group of highly characterised test samples representing the linear operating range of the new method(s). However, it is acceptable to use one sample and dilute it in a suitable matrix. Identify and cite the reference method(s) and protocol(s) used in the study (Bowden & Wang, 2021).

3.3. Stage 2 – Diagnostic performance of the assay

3.3.1. Study design(s)

Ideally, study design(s) should be done with the assistance of a statistician to ensure that the sample size and the experimental approach are valid. Give an overview of the chosen approach used for determination of diagnostic specificity and sensitivity estimates. Include rationale for statistical design, choice of populations, animals or animal models, numbers of animals used to generate confidence intervals for sensitivity and specificity etc. Field samples are preferable to samples from experimental infection studies which may not be representative of natural infection and often result in overestimation of Dse and DSp. Host variables in the target population must be represented and recorded. The true status (positive/negative, etc.) of the reference animal populations should be independently verified by a different technique. Bayesian latent class model (LCM) Chapter 1.1.6 can be used to account for imperfect reference tests and test accuracy data from animals of unknown status. LCMs do not assume that the reference test is a perfect "gold standard" but estimate the accuracy of the reference test and the test under evaluation with the joint test results. Assumptions and statistical basis for study design need to be disclosed (Cheung et al., 2021, Kostoulas et al., 2021).

3.3.2. Stage 2. Reference animal populations

It is imperative that estimates of DSe and DSp are as accurate as possible. Ideally, they are derived from testing a panel of samples from reference animals, of known history and infection status relative to the disease/infection in question and relevant to the country or region in which the test is to be used.

Such panels or guidance on where they may be sourced, are available from WOAH Reference Laboratories.

3.3.2.1. Negative reference animals/samples

True negative samples, from animals that have had no possible infection or exposure to the agent, may be difficult to locate. It is often possible to obtain these samples from countries or zones that have eradicated or have never had the disease in question. Provide details of the source of the samples and the tests used to define status of animals and a description of the source population ideally with similar characteristics as the target population where the assay will be applied (Heuer & Stevenson, 2021).

3.3.2.2. Positive reference animals/samples

Positive refers to known exposure to, or infection with, or presence of analyte or the agent in question (case definition). Provide details of the tests used to define status of animals or prevalence within population. It is generally problematic to find sufficient numbers of true positive reference animals as determined by isolation of the pathogen. It may be necessary to resort to samples from animals that have been identified by another test of sufficiently high accuracy, such as a validated nucleic acid detection assay. The test under evaluation is applied to these reference samples and results (positive and negative) are cross-classified in a 2 × 2 table. This has been called the "gold standard model" as it assumes the reference standard is perfect (*Heuer & Stevenson, 2021*; *Chapter 1.1.6.*,*Table 1 and 2*).

3.3.3. Stage 2. Samples from animals of unknown status

A way to overcome the problem of an imperfect reference standard is to perform a latent class analysis of the joint results of two tests assuming neither test is perfect. LCMs estimate the accuracy of the test under evaluation and the reference standard with the joint test results. If Bayesian latent class analysis was used, please describe sampling from the target population(s), the characteristics of other tests included in the analysis, the appropriate choice of model and the estimation methods based on peer-reviewed literature (*Chapters 1.1.6. and 2.2.5.*; *Cheung* et al., *2021*).

3.3.4. Stage 2. Experimentally infected or vaccinated reference animals

In cases when the near-impossibility of obtaining suitable reference samples from naturally exposed animals necessitates the use of samples from experimental animals for validation studies, the resulting DSe and DSp measures should be considered as less than ideal estimates of the true DSp and DSe. Multiple serially acquired pre- and post-exposure results from individual animals are not acceptable for establishing estimates of DSe and DSp because the statistical requirement of independent observations is violated. Single time-point sampling of individual experimental animals can be acceptable but the strain of organism, dose, and route of administration to experimental animals are examples of variables that may introduce error when extrapolating DSe and DSp estimates to the target population. Please provide complete description: age, sex, breed, etc. Immunological status. Type of exposure, inoculation, source, aerosol, contact, sampling plan and procedures, etc.

3.3.5. Stage 2. Threshold determination

Complete description of method used to determine thresholds (*cut-off*[*s*]) *used to classify animals as test positive, negative, or indeterminate* (*if relevant*). *Include statistical calculations, frequency distributions, receiver operating characteristic* (*ROC*) *analysis, etc., as applicable. See* WOAH Terrestrial Manual *Chapter 1.1.6. and Chapter 2.2.5*.

Complete either Section 3.3.6 if defined reference animals were used, or Section 3.3.7 if a latent class model was used.

3.3.6. Stage 2. Diagnostic sensitivity and specificity estimates – with defined reference animals

Diagnostic sensitivity is the proportion of known infected reference animals that test positive in the assay; infected animals that test negative are considered to have false-negative results. Diagnostic specificity is the proportion of known uninfected reference animals that test negative in the assay; uninfected reference animals that test positive are considered to have false-positive results. Please include a 2×2 table and confidence intervals for estimates of these parameters.

For quantitative diagnostic tests, a useful adjunct to estimates of diagnostic sensitivity and specificity is an estimate of the area under the receiver operating characteristic (ROC) curve. Please include this information where relevant (Chapter 2.2.5.; Heuer & Stevenson, 2021).

3.3.7. Stage 2. Diagnostic sensitivity and specificity estimates – without defined reference animals

Complete description of LCM used (Bayesian or maximum likelihood). Describe rationale for use of this approach, and sources of priors (e.g. experts and published papers) for Bayesian models providing relevant, supporting data. Population selection criteria should be presented, including prevalence estimates. Other test methods evaluated should also include the standard method of comparison. The source data tables with cross-classified test results should be presented for each test population. Using best available priors, choose test populations with appropriate prevalences and select animals in sufficient numbers to generate estimates of sensitivity and specificity with an allowable error of ± 5% at a level of 95% confidence. If multiple laboratories are involved in the study design, data on reproducibility should be presented in Section 3.4.3. (Chapter 2.2.5., Cheung et al., 2021, Heuer & Stevenson 2021).

3.3.8. Stage 2. Comparison of performance between tests

For standard method(s) of comparison (reference methods) used in full field studies, indicate diagnostic sensitivity and specificity estimates as determined in either Section 3.3.6 or 3.3.7. The reference method could also be used to calculate relative DSe and DSp of the test under evaluation. Provide statistical measures of agreement between the reference method(s) and the new test being validated and suggest explanations for results not in agreement (Chapter 2.2.8.; Reising et al., 2021; Kirkland & Newberry, 2021).

3.4. Stage 3 – Reproducibility

Reproducibility is the ability of a test method to provide consistent results when applied to aliquots of the same sample tested by the same method in different laboratories. Where possible, the reproducibility assessment should include data from tests conducted at a WOAH Reference Laboratory or national laboratory. The panel should contain at least 20 samples and at least three laboratories should participate in the reproducibility testing. All steps of the testing protocol should be performed at each laboratory. Further information is available in the WOAH *Terrestrial Manual* Chapter 2.2.6 'Selection and use of reference samples and panels', Johnson & Cabuang, 2021; Waugh & Clark, 2021.

3.4.1. Stage 3. Laboratory identification

State the number of laboratories included (minimum of three), which should also include WOAH Reference Laboratories or Collaborating Centres, or national laboratories where they exist.

3.4.2. Stage 3. Evaluation panel

Description of test panel used for independent reproducibility study (interlaboratory comparisons), nature and number of samples and assessment of homogeneity and stability. Define acceptable limits. (2.2.6.).

3.4.3. Stage 3. Analysis of reproducibility (Johnson & Cabuang, 2021; Waugh & Clark, 2021)

Interpretation and statistical analysis of results.

3.5. Stage 4 – Monitoring the performance

Once a validated assay is in routine use it must be continually assessed to assure it maintains its fitness for purpose and maintains the performance characteristics as defined during validation of the assay. This can be determined in a quality assurance programme characterised by carefully monitoring the assay's daily performance, primarily through precision and accuracy estimates for internal controls, as well as outlier tendencies. Chapter 2.2.4. of the WOAH *Terrestrial Manual* provides guidance on the application of measurement uncertainty.

Reproducibility is assessed through external quality control programmes such as at least annual, proficiency testing. This is an essential requirement of ISO 17025 accredited laboratories.

At minimum a plan for monitoring the performance is essential for the initial acceptance as a WOAH validated test and implementation of that plan will be required to maintain its status as a WOAH validated test

3.5.1. Stage 4. Daily performance

Description of procedures in place or planned, to monitor the assay's daily performance within the applicant's laboratory. This monitoring differs from assessment of repeatability. Repeatability determined as part of the assay validation pathway, is the level of agreement between results of replicates of a sample, both within and between runs of the same method in one laboratory. Daily monitoring concerns the inclusion of quality control samples for example a strong and weak positive control, in the assay to confirm that the results fall within established limits which may be defined by the repeatability studies. Please include details of the controls included in the assay to determine whether it is performing as expected and the measurement of uncertainty.

3.5.2. Stage 4. Reproducibility

Description of external proficiency programme in place or planned, to monitor reproducibility. Provide details including frequency and source of proficiency panel.

3.5.3. Summary

Please include a summary of relevant parameters and a statement about whether the assay is fit for all purposes, its scope and limitations. Ideally this summary should include a succinct table inclusive of source population information as it relates to fitness for purpose, and also inclusive of numbers of animals in each relevant category (infected/uninfected).