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

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

Information to fill out in this form

The purpose of this template is to provide pre-publication validation data so that the Aquatic Animal Health Standards Commission (the Aquatic Animals Commission) can consider inclusion of proposed tests in the WOAH *Manual of Diagnostic Tests for Aquatic Animals* (*Aquatic Manual*). Only tests that have been validated to level 2 will be considered for inclusion in the *Aquatic Manual*. If the proposed assay is accepted for inclusion in the *Aquatic Manual*, the completed template will be made available on the WOAH website.

Before filling in this form and submitting an application, applicants should consult Chapter 1.1.2 of the WOAH *Aquatic Manual*

WOAH *Terrestrial Manual* 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.

As shown in Figure 1, from the WOAH *Aquatic Manual* chapter 1.1.2, the following parameters have to be addressed: 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. 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

Organisation	
Address	
Phone	
Contact person	

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. 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.

1	Surveillance of apparently healthy animals	☐
2	Presumptive diagnosis of clinically affected animals	☐
3	Confirmatory diagnosis of a suspect result from surveillance or presumptive diagnosis	☐

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, 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.

2.3.2. Disease target/analyte target)
State targets in analytical terms.

Section 3. Development and validation of the assay

3.1. Assay development pathway

3.1.1. Design, development, optimisation and standardisation of the assay for the intended purpose
For guidance, refer to Section Assay development – the experimental studies of Chapter 1.1.2 of the WOAH Aquatic Manual.

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, e.g. for a total of 10–20 runs.

The data/detail provided must be clear, including:

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 different operators used at a single site*

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. 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, and d) antibodies to vaccination that may be confused with antibodies to active infection,*

b) *exclusivity, which is the capacity of the assay to detect an analyte or genomic sequence that is unique to a targeted organism, and excludes all other known organisms that are potentially cross-reactive. This would also define a confirmatory assay, and*

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.*

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 candidate test and reference test to obtain a relative estimate for ASe.*

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 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). Identify and cite the reference method(s) and protocol(s) used in the study.

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 candidate test with the joint test results.

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.

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.

3.3.2.2. Positive reference animals/samples

Positive refers to known exposure to, or infection with, the agent in question. 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 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.

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 candidate test 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.

<div style="border:1px solid #ccc; min-height:80px;"></div>

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.

<div style="border:1px solid #ccc; min-height:80px;"></div>

3.3.5. Stage 2. Stage 2. Cut-off (threshold) determination

The selection of the cut-off(s) should reflect the intended purpose of the assay and its application, and must support the required DSe and DSp of the assay. Options and descriptive methods for determining the best way to express DSe and DSp are available (Branscum et al., 2005; Georgiadis et al., 2003; Greiner et al., 1995; 2000; Jacobson, 1998; Zweig & Campbell, 1993; and Chapter 2.2.5 of the Terrestrial Manual).

<div style="border:1px solid #ccc; min-height:80px;"></div>

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

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

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.

<div style="border:1px solid #ccc; min-height:80px;"></div>

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

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

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.

<div style="border:1px solid #ccc; min-height:80px;"></div>

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 candidate test. Provide statistical measures of agreement between the reference method(s) and the new test being validated and suggest explanations for results not in agreement.

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 an 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. Further information is available in the WOAH *Terrestrial Manual*.

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.

3.4.3. Stage 3. Analysis of reproducibility

Interpretation and statistical analysis of results.

3.5. Stage 4 – Monitoring the performance

To retain the status of a validated assay it is necessary to assure that the assay as originally validated consistently maintains the performance characteristics as defined during validation of the assay.

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 an WOAH validated test and implementation of that plan will be required to maintain its status as an WOAH validated test.

3.5.1. Stage 4. Routine 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.

[blank shaded box]

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.

[blank shaded box]

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). Examples (from the Aquatic Manual):

Table 4.1. WOAH recommended diagnostic methods and their level of validation for surveillance of apparently healthy animals and investigation of clinically affected animals

Method [amend or shade in as relevant]	Surveillance of apparently healthy animals				Presumptive diagnosis of clinically affected animals				Confirmatory diagnosis1 of a suspect result from surveillance or presumptive diagnosis			
	Early life stages[2]	Juveniles[2]	Adults	LV	Early life stages[2]	Juveniles[2]	Adults	LV	Early life stages[2]	Juveniles[2]	Adults	LV

Table 4.4.1.1. Primers and probes (sequences) and cycling conditions for DIV1 real-time PCR

Target gene	Primer/probe (5'–3')	Concentration	Cycling parameters[a]
Method 1: Reference; GenBank Accession No.: xxx			
ATPase	F: AGG-AGA-GGG-AAA-TAA-CGG-GAA-AAC R: CGT-CAG-CAT-TTG-GTT-CAT-CCA-TG: FAM-CTG-CCC-ATC-TAA-CAC-CAT-CTC-CCG-CCC-TAMRA	500 nM 200 nM	40 cycles of 95°C/100 sec and 60°C/30 sec

For presumptive diagnosis of clinically affected animals

Test type	Test purpose	Source populations	Tissue or sample types	Species	DSe (*n*)	DSp (*n*)	Reference test	Citation

DSe = diagnostic sensitivity, DSp = diagnostic specificity, *n* = number of samples used in the study, PCR: = polymerase chain reaction.

For surveillance of apparently healthy animals

Test type	Test purpose	Source populations	Tissue or sample types	Species	DSe (*n*)	DSp (*n*)	Reference test	Citation

DSe = diagnostic sensitivity, DSp = diagnostic specificity, *n* = number of samples used in the study,
PCR: = polymerase chain reaction.