INTRODUCTION

Validation is a process that determines the fitness of an assay, which has been properly developed, optimised and standardised, for an intended purpose. Validation includes estimates of the analytical and diagnostic performance characteristics of a test. In the context of this chapter, an assay that has completed the first three stages of the validation pathway (see Figure 1 below), including performance characterisation, can be designated as “validated for the original intended purpose(s)”\(^1\). To maintain a validated assay status, however, it is necessary to carefully monitor the assay’s daily performance, primarily through repeatability estimates of internal controls, to ensure that the assay, as originally validated, consistently maintains its performance characteristics. Should it no longer produce results consistent with the original validation data, the assay may be rendered unfit for its intended purpose. Thus, a validated assay is continuously assessed to assure it maintains its fitness for purpose – as determined by assessments of the assays validity in each run of the assay.

Assays applied to individuals or populations have many purposes, such as aiding in: documenting freedom from disease in a country or region, preventing spread of disease through trade, eradicating an infection from a region or country, confirming diagnosis of clinical cases, estimating infection prevalence to facilitate risk analysis, identifying infected animals toward implementation of control measures, and classifying animals for herd health or immune status post-vaccination. A single assay may be validated for one or several intended purposes by optimising its performance characteristics for each purpose, e.g. setting diagnostic sensitivity (DSe) high, with associated lower diagnostic specificity (DSp) for a screening assay, or conversely, setting DSp high with associated lower DSe for a confirmatory assay.

The ever-changing repertoire of new and unique diagnostic reagents coupled with many novel assay platforms and protocols has precipitated discussions about how to properly validate these assays. It is no longer sufficient to offer simple examples from serological assays, such as the enzyme-linked immunosorbent assay, to guide assay developers in validating the more complex assays. In order to bring coherence to the validation process for all types of assays, this chapter focuses on the criteria that must be fulfilled during assay development and validation of all assay types. The inclusion of assay development as part of the assay validation process may seem counterintuitive, but in reality, three of the validation criteria that must be assessed in order to achieve a validated assay, comprise steps in the assay development process. Accordingly the assay development process seamlessly segues into an assay validation pathway, both of which contain validation criteria that must be fulfilled. This chapter also provides guidance for evaluation of each criterion through provision of best scientific practices contained in the chapter’s appendices. The best practices are tailored for each of several fundamentally different types of assays (e.g. detection of nucleic acids, antibodies, or antigens).

DIRECT AND INDIRECT METHODS THAT REQUIRE VALIDATION

The diagnosis of infectious diseases is performed by direct and/or indirect detection of infectious agents. By direct methods, the particles of the agents and/or their components, such as nucleic acids, structural or non-structural proteins, enzymes, etc., are detected. The indirect methods demonstrate antibodies induced by
exposure to infectious agents or their components. The most common indirect methods of infectious agent
detection are antibody assays such as classical virus neutralisation, antibody enzyme-linked immunosorbent
assay (ELISA), haemagglutination inhibition, complement fixation, and the recently appearing novel methods,
such as biosensors, bioluminometry, fluorescence polarisation, and chemoluminescence.

The most common direct detection methods are isolation or in-vitro cultivation of viable organisms, electron
microscopy, immunofluorescence, immunohistochemistry, antigen-ELISA, Western immunoblotting, and nucleic
acid detection systems (NAD). The NAD systems include nucleic-acid hybridisation (NAH), macro- and
microarrays and the various techniques of nucleic acid amplification, such as the polymerase chain reaction
(PCR), or the isothermal amplification methods, such as nucleic acid sequence-based amplification (NASBA),
and invader or loop-mediated isothermal amplification (LAMP). NAD assays are rapidly becoming
commonplace and, in many cases, replacing virus isolation and bacteria cultivation, particularly for the
detection of agents that are difficult or impossible to culture. NAD tools are also used as a secondary means
for highly specific identification of strains, groups, or lineages of organisms following isolation or culture of
viruses, bacteria and parasites. Molecular diagnostics, such as PCR, do not require: a) the presence of
replicating organisms, b) expensive viral isolation infrastructure, c) up to several weeks to achieve a diagnosis,
or d) special expertise, which is often unavailable in many laboratories – all practical advantages. These
methods have become relatively inexpensive, safe and user-friendly (1–6, 13, 17, 18). Various real-time PCR
methods, nucleic acid extraction robots, and automated workstations for NAD, antibody, antigen, and agent
detection have resulted in a large arsenal of high throughput, robust, very rapid and reliable assays. Although
NAD systems often have the advantage of a greater diagnostic sensitivity and analytical specificity than
infectious agent recovery or antigen-capture ELISA procedures, that advantage usually carries with it a greater
challenge for validation of such assays.

PRELIMINARY CONSIDERATIONS IN ASSAY DEVELOPMENT AND VALIDATION

The first consideration is to define the purpose of the assay, because this guides all subsequent steps in the
validation process. By considering the variables that affect an assay’s performance, the criteria that must be
addressed in assay validation become clearer. The variables can be grouped into three categories: (a) the
sample – individual or pooled, matrix composition, and host/organism interactions affecting the target analyte quantitatively or qualitatively; (b) the assay system – physical, chemical, biological and operator-related factors affecting the capacity of the assay to detect a specific analyte in the sample; and (c) the test result – the capacity of a test result, derived from the assay system, to predict accurately the status of the individual or population relative to the analyte in question.

The matrix in which the targeted analyte may reside (serum, faeces, tissue, etc.) may contain endogenous or
exogenous inhibitors that prevent enzyme-dependent tests such as PCRs or ELISAs from working. Other
factors that affect the concentration and composition of analyte (mainly antibody) in the sample may be mainly
attributable to the host and are either inherent (e.g. age, sex, breed, nutritional status, pregnancy, immunological responsiveness) or acquired (e.g. passively acquired antibody, active immunity elicited by vaccination or infection). Non-host factors, such as contamination or deterioration of the sample, also affect the ability of the assay to detect the specific targeted analyte in the sample.

Factors that interfere with the analytical performance of the assay system include instrumentation, operator
error, reagent choice (both chemical and biological) and calibration, accuracy and acceptance limits of assay
controls, reaction vessels and platforms, water quality, pH and ionicity of buffers and diluents, incubation
temperatures and durations, and error introduced by detection of closely related analytes. It is also important
that biological reagents are free of extraneous agents.

Factors that may negatively impact diagnostic performance of the assay are primarily associated with choice of
reference sample panels from known infected/exposed or known uninfected animals selected for evaluating the
diagnostic sensitivity (DSe) and diagnostic specificity (DSP) of the assay. This is particularly difficult because the
degree to which the reference animals represent all of the host and environmental variables in the
population targeted by the assay has a major impact on the confidence of test-result interpretation. For example, experienced serologists are aware that an assay, validated by using serum samples from northern European cattle, may not give valid results for the distinctive populations of cattle in Africa. Diagnostic performance of the assay is further complicated when sample panels of known infection status are not available, often because they are impossible to obtain. In this situation, DSe and DSP can be estimated by use of latent class models (9, 11 and Appendix 1.1.2.5).
Chapter 1.1.2. — Principles and methods of validation of diagnostic assays for infectious diseases

**THE CRITERIA OF ASSAY DEVELOPMENT AND VALIDATION**

It is apparent that assay performance is affected by many factors that span from the earliest stages of assay development through the final stage of performance assessments when the test is applied to targeted populations of animals. An assay, therefore, cannot be considered validated unless a specific set of essential validation criteria (see accompanying box) have been tested and affirmed or fulfilled, either quantitatively or qualitatively (for detail on terms, see glossary in reference 19). Lack of attention to any one of these criteria will likely reduce the level of confidence that an assay is fulfilling the purpose(s) for which it is intended. The first four of these criteria typically are addressed during development of the assay (the Development Pathway), and the remaining eight are evaluated during the first three stages of assay validation (the Validation Pathway) as described below.

![Assay validation criteria](image)

**A. ASSAY DEVELOPMENT PATHWAY**

1. **Definition of the intended purpose(s) for an assay**

   The OIE *Standard for Management and Technical Requirements for Laboratories Conducting Tests for Infectious Diseases* (20) states that test methods and related procedures must be appropriate for specific diagnostic applications in order for the test results to be of any relevance. In other words, the assay must be ‘fit for purpose’². The capacity of a positive or negative test result to predict accurately the infection and/or exposure status of the animal or population of animals is the ultimate consideration of assay validation. This capacity is dependent on development of a carefully optimised (Section A.2.d), and standardised (Section A.2.g) assay that, through accrual of validation data, provides less biased and more precise estimates of DSe and DSp. These estimates, along with evidence-based data on prevalence of infection in the population being tested, are the basis for providing a high degree of confidence in the predictive values of positive and negative test results. That ultimately is the culmination of the assay validation process: to provide assurance that the test method is validated and, along with evidence of proper maintenance of the validation criteria, quality control, and quality assurance programmes, that test results provide useful diagnostic inferences about the animal or population infection/exposure status. Figure 1 shows the assay validation process, from assay design through the development and validation pathways to implementation, deployment, and maintenance of the assay.

   As outlined in the background information in *Certification of diagnostic assays* on the OIE website (www.oie.int), the first step is selection of an assay type that likely can be validated for a particular use.

   The most common purposes are to:

1) Demonstrate freedom from infection in a defined population (country/zone/compartment/herd) (prevalence apparently zero):
   1a) ‘Free’ with and/or without vaccination,
   1b) Re-establishment of freedom after outbreaks

2) Certify freedom from infection or presence of the agent in individual animals or products for trade/movement purposes.

3) Eradication of disease or elimination of infection from defined populations.

4) Confirmatory diagnosis of suspect or clinical cases (includes confirmation of positive screening test).

5) Estimate prevalence of infection or exposure to facilitate risk analysis (surveys, herd health status, disease control measures).

6) Determine immune status of individual animals or populations (post-vaccination).

² This is a specific interpretation of the more generally stated requirements of the ISO/IEC 17025:2005 international quality standard for testing laboratories (15). The OIE Standard further states that in order for a test method to be considered appropriate, it must be properly validated and that this validation must respect the principles outlined in the validation chapter of the Terrestrial Manual (from which this chapter has been adapted).
These purposes are broadly inclusive of many narrower and more specific applications of assays (see Appendices for each assay type for details). Such specific applications and their unique purposes need to be clearly defined within the context of a fully validated assay.

![Assay Development Pathway](image)

While this chapter deals with validation and fitness for purpose from a scientific perspective, it should also be noted that other practical factors might impact the relevance of an assay with respect to its intended application. These factors include not only the diagnostic suitability of the assay, but also its acceptability by scientific and regulatory communities, acceptability to the client, and feasibility given available laboratory resources. An inability to meet operational requirements of an assay also may make it unfit for its intended use. Such requirements may include performance costs, equipment availability, level of technical sophistication and interpretation skills, kit/reagent availability, shelf life, transport requirements, safety, biosecurity, sample throughput, test turn-around times, aspects of quality control and quality assurance, and whether the assay can practically be deployed to other laboratories.

2. Assay development – the experimental studies

a) Essential prerequisites: factors that impact assay validation

i) Quality Assurance: Whether developing assays in the laboratory or performing analyses of clinical material, the objective is to produce data of high quality. This requires that key requirements have to be fulfilled within the laboratory (see Chapter 1.1.1 of this Aquatic Manual). The establishment of quality
assurance (QA) and quality control (QC) systems is essential, i.e. a set of quality protocols, including the use of assay control samples that ensure that the system is working properly and confirms data reproducibility and quality. QA and QC systems, together with trained and competent personnel, have already been established in many laboratories world-wide.

ii) Equipment selection: Equipment that is not maintained and calibrated can be a major impediment to achieving a quality assay. Apparatus (freezers, heating blocks, incubators, refrigerators, optical colorimeters, thermocyclers, plate washers, pipettes, etc.) must be calibrated according to the laboratory's quality assurance protocols. Examples of this need include robotics used for automation of entire assays, or parts thereof, for routine diagnostic processing. It is not sufficient to assume that robotic extraction of nucleic acid, for example, is equivalent to previously used manual extraction methods or that an automated ELISA plate washer provides uniform washing among wells of the plate. The instrument must be calibrated and the protocol validated to confirm performance efficiency and to ensure that cross-contamination in NAD systems or inadequate washing of some wells in a plate does not occur. (See Appendices on best practices for more details.)

b) Reference materials
Selection, collection, preparation and management of samples are critical variables in designing, developing, and validating an assay. Other variables such as transport, chain of custody, tracking of samples, and the laboratory information management system are also critical sources of variation/error that become especially important when the assay is implemented for routine testing. Integrity of experimental outcomes during assay development and validation is only as good as the quality of the samples used in experimentation or routine diagnostic testing. Anticipating the factors that can negatively impact sample quality must precede launching an assay validation effort. Reference samples used in assay development and validation should be in the same matrix used in the assay (e.g. serum, tissue, whole blood) and representative of the species to be tested by the resulting assay. The reference materials should appropriately represent the range of analyte concentration to be detected by the assay. Details on proper sample collection, preparation, management, and transport are available in Part 2. Recommendations applicable to specific diseases: general introduction, and Chapters 2.2.0, 2.3.0 and 2.4.0 of this Aquatic Manual, Chapter 1.4 of the Aquatic Code and the OIE Guide for Aquatic Animal Health Surveillance (2009).

c) Design of test method
Considerable thought and planning needs to go into designing all steps of a new assay destined for validation, or an existing assay that is being modified. Assistance is offered in Appendices to this chapter, which cover best practices for development and validation of assays for detection of various analytes (e.g. antibody, antigen, and nucleic acid detection).

d) Optimisation
The labour-intensive process of optimising an assay is fundamental and critical to achieving a quality assay. Scientific judgment and use of best scientific practices provided in accompanying Appendices to this chapter are the best assets to guide optimisation of all elements of an assay. The approach outlined provides a solid foundation for development of an assay that fulfils the criteria for 'ruggedness' when used over extended periods of time within a laboratory or when implemented in other laboratories. Optimisation starts with support structure and functions, all of which need to be optimised and include appropriate facilities designed for contamination curtailment, (i.e. for PCR), a laboratory-wide quality assurance programme, equipment maintenance and calibration under QC, and sample handling specifications that minimise negative impacts on the targeted analyte.

Factors to be addressed in optimisation of the assay include variability in instrumentation, reagent choice (chemical and biological) and methods to assess them, water quality, pH, ionicity, and concentrations of buffers, diluents, critical chemicals, primers and probes, incubation temperatures and durations, and other added assay components, such as blockers like milk solution or bovine serum albumin, and technical error, etc.

In some assay types, a correct assay result is fully dependent on getting a particular step in the testing process correct, requiring special attention during optimisation. A case in point is nucleic acid extraction from the sample. Both commercial (robotic, spin columns, and magnet-based extractions, etc.) and standard chemistry-based methods are used for DNA or RNA extraction. It is crucial to determine the
most reproducible and efficient extraction method as part of the optimisation exercise. Extraction needs to be optimised for every type of tissue that may be targeted by the assay. If the method of extraction is changed, at a minimum, equivalent efficiency of extraction should be demonstrated (see Section 6 below and associated Appendix 1.1.2.3 for additional information on establishing equivalency when reagents are changed).

A variety of analyte reference samples and other process controls that are routinely included in any assay system are identified in the next sections. These provide critical assay monitoring functions that require special attention during assay optimisation. In addition, proper preparation and storage of all biological reagents and reference materials must be heeded to ensure stability.

During the experimentation to optimise the assay, take note of assay characteristics that have a narrow range in which they are optimal, as these are the critical points that may affect an assay’s robustness (see Section A.2.g)

i) **Analyte reference samples**

Development of all assays is dependent on analyte reference samples that reflect the target analyte in the population for which the assay is intended. The reference samples may be sera, fluids or tissues that contain the analyte of interest or a genomic construct consistent with the target analyte. These reference materials are used in experiments conducted throughout the development process and carried over into the validation of the assay. It is useful to select four or five well-defined reference samples representing the analyte ranging from high positive to negative. These samples ideally should represent known infected and uninfected animals from the population that eventually will become the target of the assay once it is validated. Obtaining such reference samples, however, is not always possible, particularly for nucleic acid and antigen detection assays. The alternative of preparing reference samples spiked with cultured agents or positive sera is inferior because the matrix of field samples may be very different from spiked-sample matrix. But, when no other alternative exists, spiking a sample with a known amount of the analyte derived from culture, or diluting a high positive serum in negative serum of the same species may be all that is available. In either case, it is imperative that the matrix, into which analyte is placed or diluted, is identical to, or resembles as closely as possible the samples that ultimately will be tested in the assay. Ideally, reference samples have been well characterised using at least two alternate methodologies. These samples can be used in experiments to determine if the assay is able to distinguish between varying quantities of analyte, and for optimising the reagent concentrations and perfecting the protocol. In principle, for all assay types, it is highly desirable to prepare and store a sufficient amount of each reference sample in aliquots for use in every run of the candidate assay as it is evaluated through the entire development and validation process. Switching reference samples during the validation process introduces an intractable variable that can severely undermine interpretation of experimental data and, therefore, the integrity of the development and validation process.

| Analyte reference samples, containing the analyte of interest in varying concentrations, are useful in developing and evaluating the candidate assay’s validation criteria. |

| Linear operating range of an assay: an interval of analyte concentrations over which the method provides suitable accuracy and precision. |

e) **Inhibitory factors in sample matrix**

Generally, for antibody detection in serum, assays are rather resistant to inhibitory factors, with the exception of certain assays, e.g. toxic factors in viral neutralisation assays, or when endogenous substances found in certain sample types inhibit enzymatic reactions in ELISAs and PCRs. For nucleic acid detection, sample matrices including blood, serum, body tissues, and swab samples allow for easy extraction of target nucleic acids, while faeces, autolysed tissues and semen samples can be more difficult to handle.

f) **Operating range of the assay**

During optimisation of the assay, the lower and higher detection limits are established. To formally establish this range, a high positive concentration of analyte is serially diluted to extinction in an analyte-negative matrix that is of the same constitution as the sample matrix for samples from animals in the population targeted by the assay. The results are plotted in the form of a ‘response-curve’, with analyte concentration (or dilution) as a function of indicator (such as optical density values, cycle threshold, etc.). The curve sometimes referred to as a ‘dose–response curve’ as in pharmacological applications, establishes the linear range of assay values that are valid for diagnostic test samples subjected to the assay.
16 Manual of Diagnostic Tests for Aquatic Animals 2009

Chapter 1.1.2. — Principles and methods of validation of diagnostic assays for infectious diseases

1. Definition of a validated assay.

“Validation” is a process that determines the fitness of an assay that has been properly developed, optimised and standardised for an intended purpose. Validation includes estimates of the analytical and diagnostic performance characteristics of a test. In the context of this document, an assay that has completed the first three stages of the validation pathway (Figure 1), including performance characterisation, can be designated as “validated for the original intended purpose(s)”.

To retain the status of a validated assay, however, it is necessary to assure that the assay as originally validated consistently maintains the performance characteristics as defined during validation of the assay (see Section 6 below). This can be determined by carefully monitoring the assay’s daily performance, primarily through repeatability estimates of internal controls. Should the assay cease to produce results consistent with the original validation data, the assay would be rendered unfit for its intended purpose. Thus, a validated assay must be continuously assessed to assure it maintains its fitness for purpose.

2. Stage 1 — Analytical performance characteristics

a) Repeatability

Repeatability is estimated by evaluating results from a minimum of three samples representing analyte activity within the linear operating range of the assay. At a minimum, duplicates of these samples are treated as though they are individual test samples containing analyte in the same matrix as test samples from the population that ultimately will be targeted by the assay. It is, therefore, unacceptable to prepare a working dilution of a sample in a single tube from which an aliquot of the dilution is pipetted into three or four reaction vessels, or to use replicates from one extraction of nucleic acid to create replicates. Such ‘samples’ do not constitute valid replicates. Rather individual aliquots of a sample are pipetted into separate tubes to create the

B. ASSAY VALIDATION PATHWAY

g) Robustness

An assay should tolerate slight changes in reagents and physical parameters, while still providing the expected result. If not, the assay is said to lack robustness.

Suitable methods to achieve robustness vary depending on sample and organism type or analyte that is targeted. However, the primary parameter for measuring robustness is repeatability of internal quality control samples included in each run of the assay. Their use, statistical analysis of results and suitable charting methods are described elsewhere (8).

h) Calibration of the assay to reference analyte standards

i) National and international reference analyte standards

Ideally, international reference standards, containing a known concentration of analyte, are the reagents to which all assays are standardised. Such standards are prepared and distributed by international reference laboratories. National reference standards are calibrated by comparison with an international standard reagent whenever possible; they are prepared and distributed by a national reference laboratory. These standards are highly characterised through extensive analysis, and preferably the methods for their characterisation, preparation, and storage have been published in a peer-reviewed publications.

ii) Working reference samples (controls) for adjusting test results to a standard metrological unit

Working reference samples (controls) can be included in an assay for purposes of ‘normalising’ results (not to be confused with transformation of data to achieve a ‘normal’ [Gaussian] distribution), which adjusts raw test results (e.g. ODs, concentrations, dilutions etc.) to values of the working reference reagent. The resultant values may be expressed in many ways, such as a percent of a positive control (e.g. in an ELISA), or as a concentration of an analyte or number of targeted genomic copies derived from a standard curve of Ct (cycle threshold) values for real-time PCR. It is good practice to use this adjustment during development and validation because this provides a statistically valid means for direct comparison of results between runs of an assay. Automated assay systems make interpretation of test results easier by fitting raw data to a standard curve, or by reporting the cycle number at which the cycle threshold is exceeded as in real-time PCR. For more information, see Appendices 1.1.2.1 and 1.1.2.3.

16 Manual of Diagnostic Tests for Aquatic Animals 2009
working dilutions, or into separate extraction chambers, which in turn are transferred to their own reaction vessels. Between-run variation is determined by using the same samples in each run of a minimum of 20 runs (total), involving two or more operators, preferably at least on 5 separate days (see Appendix 1.1.2.4 on measures of uncertainty for assessments of repeatability).

b) Analytical specificity (ASp)

Analytical specificity initially should be assessed by testing a panel of samples from animals either infected with or exposed to organisms that are closely related to the assay’s targeted organism. ASp applies to both direct and indirect methods of analyte detection. It is desirable to obtain field samples from animals infected with genetically-related organisms, but this may prove difficult or even impossible. In such cases, cultivated organisms can be used in direct detection methods, or serum from animals exposed experimentally by natural routes for indirect detection methods. Acceptable cross-reactivity is largely dependent on the intended purpose of the test and the prevalence of the cross-reactive organisms/analytes in the target population samples – which must be determined for each case (see Appendix 1.1.2.5 for more detail). For PCR, it is useful to perform ‘in silico’ (computer simulation) studies as an adjunct to laboratory assessment of ASp; however, such studies are not sufficient by themselves to estimate ASp.

A factor unique to viral antibody detection is the possible antibody response of animals to carrier proteins found in vaccines. If such proteins are also present in the solid phase antigen on ELISA plates, they may bind antibody developed to vaccine carrier proteins and give false-positive results. Use of vaccine preparations as antigens in ELISAs is, therefore, not recommended. See Appendix 1.1.2.1 for specific practices to determine ASp.

c) Analytical sensitivity (ASE)

Analytical sensitivity is synonymous with the lower limit of detection of an analyte in an assay. For direct-detection assays, this may be expressed as the number of genome copies, infectious dose, colony-forming units, plaque forming units, etc. of the agent that can be detected and distinguished from the result of a matrix background. Most commonly, this is expressed as a number of copies, complement-fixing units or plaque-forming units giving 50% positive results for a specified sample volume or weight. For indirect detection assays, it is the least amount of antibody detected, usually, the penultimate dilution of sample in which the analyte is indistinguishable from the activity of a sample matrix control.

If the intended purpose is to detect low levels of analyte or subclinical infections and it is difficult to obtain the appropriate reference materials, for example samples from early stages of the infection process, it may be useful to determine comparative analytical sensitivity by running a panel of samples on the candidate assay and on another independent assay.

Where a new, more sensitive test is developed, it may be necessary to undertake serial tests taken from infected animals from early after infection through to the development of full infection/serology and to run these in parallel to previously used tests to demonstrate the increased sensitivity.

d) Standard test method for comparison with the candidate assay test method

There are situations where it is not possible to continue to Stage 2 of the Validation Pathway because appropriate samples from the target population are scarce and animals are difficult to access (such as for exotic diseases). In this case, provisional recognition of the assay can be made, provided that results of the candidate assay compare favourably with results of a standard test method or a known established and preferably published method (see Section B.2.g. below). A small but select panel of highly characterised test samples representing the linear operating range of the candidate method should be run in parallel on the candidate assay method and by the standard method if it exists.

e) Analytical accuracy of adjunct tests or procedures

Some test methods or procedures may be qualified for use as analytical tools in the diagnostic laboratory. These usually are secondary adjunct tests or procedures that are applied to an analyte that has been detected in a primary assay. The purpose of such analytical tools is to further characterise the analyte detected in the primary assay. Examples of such adjunct tests include virus neutralisation to type an isolated virus, or molecular sequencing to confirm a real-time PCR test result. Pathogenicity indices, haemagglutination inhibition, drug resistance determinations, etc. are other examples where adjunct tests or procedures are performed, either independent of, or as part of a primary assay.

Such adjunct tests must be validated for analytical performance characteristics (Sections A.2 through B.2.d., above), but differ from routine diagnostic tests because they do not require validation for diagnostic
performance characteristics (Sections B.3 through B.5, below) unless the tool is built into a fully validated diagnostic test method. The analytical accuracy of these tools may be defined by comparison with a reference reagent standard, or by characteristics inherent in the tool itself (such as endpoint titration). In all of these examples, the targeted analyte is further characterised quantitatively or qualitatively by the analytical tool. These types of tools often require special reagents and/or equipment and are normally conducted by Reference Laboratories.

f) Preliminary evaluation of reproducibility

Preliminary reproducibility estimates of the candidate assay may be useful at this time in the validation process, where only a small panel of highly characterised samples is available. This panel could be used for a limited evaluation of reproducibility to enhance provisional acceptance status for the assay. The candidate test method is then duplicated in laboratories at one or more different institutes, and the panel of samples is evaluated using the candidate assay in each of these laboratories, using the same protocol, same reagents as specified in the protocol, and equivalent equipment. This is a scaled-down version of Stage 3 of assay validation.

g) Provisional assay recognition

Experience has shown that the greatest obstacle for continuing through Stage 2 of the Validation Pathway is the number of defined samples required to calculate DSe and DSp (see requirements for Stage 2, Diagnostic Performance Characteristics, below). The formula is well known and tables are available for determining the number of samples required to estimate various levels of DSe and DSp, depending on the amount of allowable error and the level of confidence in the estimates (16). But the formula assumes that the myriad of host/organism factors that may affect the test outcome are all accounted for. Since that assumption may be questionable, the estimated sample sizes are at best minimal. For a disease that is not endemic or widespread, it may be impossible initially to obtain the number of samples required, but over time, accrual of additional data will allow adjustment of the threshold or if no adjustment is needed, enhance confidence in the estimates.

Historical precedent would suggest that assays were generally the product of laboratory experiments with an emphasis on analytical sensitivity and analytical specificity, and evaluation of panels of field samples was nominal. Such bench validation for bovine spongiform encephalopathy (BSE) is a classical example where positive field samples were not available. Nevertheless, during extended periods of usage in diagnostic settings, such tests have undergone adjustments based on empirical evidence to reduce false-positive and false-negative results. For some of the BSE rapid tests, adjustments had to be made in the cut-off to reduce false positive results apparent in early implementation. Bench validation provided a level of confidence in diagnostic performance that was adequate for conditional diagnostic use as determined by national authorities. But, it must never become a replacement for a full field validation. Therefore bench validation of diagnostic assays can only offer provisional recognition with anticipation that full field validation will follow.

A provisional recognition of an assay by state or national authorities recognises that the assay has not been evaluated for diagnostic performance characteristics. As such, the laboratory should develop and follow a protocol for adding and evaluating samples, as they become available, to fulfill this requirement. Ideally, this process should be limited to a specific timeframe in which such an accrual would be directed toward fulfilling Stages 2 and 3 of the validation pathway. This concept of ‘on-going’ validation should be limited to emergency situations where rapid introduction of new tests is deemed essential by authorities, or where bi-lateral trade agreements, based on such a test, are mutually accepted. In exceptional cases for rare diseases where no other assay option exists, provisional recognition may be allowed, but reporting of results should include a statement of the provisional nature of the validation of the assay. In all cases, sound evidence must exist for preliminary estimates of DSp and DSe based on a small select panel of well-characterised samples containing the targeted analyte. (For additional perspectives on moving molecular diagnostics into routine use at veterinary laboratory agencies, see reference 18).

3. Stage 2 – Diagnostic performance of the assay

**Diagnostic Sensitivity and Diagnostic Specificity.** Estimates of DSe and DSp are the primary performance indicators established during validation of an assay. These estimates are the basis for calculation of other parameters from which inferences are made about test results (e.g. predictive values of positive and negative test results). Therefore, 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.
A sampling design must be chosen that will allow estimation of these parameters. It should contain known positive and known negative samples, which ideally should be in proportion to that of the population for which the assay is targeted. However, this is a difficult process complicated by logistical and financial limitations. It is also limited by the fact that reference populations and "gold standards" may be lacking (see Appendix 1.1.2.5 for further detail). The following are examples of reference populations and methodologies that may aid in determining performance characteristics of the test being validated.

a) Reference animal populations

 Ideally, selection of reference animals requires that important host variables in the target population are represented in animals chosen for being infected or exposed to the target agent, or have never been infected or exposed. The variables to be noted include but are not limited to species, age, sex, breed, stage of infection, vaccination history, and relevant herd disease history.

Negative reference samples: True negative samples, from animals that have had no possible exposure to the agent, may be difficult to locate. It is often possible to obtain these samples from countries that have eradicated or have never had the disease in question. Such samples are useful as long as the targeted population for the assay is similar to the sample-source population.

Positive reference samples: It is generally problematic to find sufficient numbers of true positive reference animals, as determined by isolation of the organism. It may be necessary to resort to samples from animals that have been tested by another test such as a nucleic acid detection system.

b) Reference animal status determined by other assays

Although the term 'gold standard' is commonly used to describe any standard of comparison, it should be limited to methods or a combination of methods that unequivocally classify animals as infected/exposed or uninfected. Some isolation methods themselves have problems of repeatability and analytical sensitivity, so are not truly gold standards particularly for purportedly negative samples.

NAD assays may be more sensitive and specific than existing ‘gold standard’ methods which render the established ‘gold standard’ as not suitable for use as a comparison. If the NAD is more sensitive than the ‘gold standard,’ an apparent lower relative specificity will be misleading. This problem may be partially resolved by assessing sample derivation, clinical history and sequencing of any PCR products to confirm analyte identity. The use of latent class models (see Section d below) may help resolve this.

c) Relative standard of comparison

Because a true gold standard may be lacking or is impossible to obtain, relative standards of comparison are often necessary; the most common of these include results from other assays. When only relative standards of comparison are available, estimates of DSe and DSp for the candidate assay may be compromised because the error in estimates obtained by comparison to the relative standard is carried over into those estimates for the candidate assay. Indeed, when using imperfect reference assays, the DSe and DSp performance estimates of the candidate assay will be flawed. It then is advisable to use a latent class model to gain more information toward establishing DSe and DSp (see next section).

d) Reference animals – status unknown

When it is not possible to assemble sera from animals of known infection status, it is possible to estimate DSe and DSp by no-gold standard methods, also known as latent class models (9, 14, and Appendix 1.1.2.5). Because these statistical models are complex, expert assistance may be needed to conduct and describe the 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.

e) Experimentally infected or vaccinated reference animals

Sera obtained sequentially from experimentally infected or vaccinated animals are useful for determining the kinetics of antibody responses or the presence/absence of antigen or organisms in samples from such animals. However, 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. Only single time-point sampling of individual experimental animals is acceptable. Also, for indirect methods of analyte detection, exposure to organisms under experimental conditions, or vaccination may elicit antibody responses that may not be quantitatively and qualitatively typical of natural infection in the target population (16). 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. For these reasons, validation of an assay should not be based solely on samples from experimental animals.
f) Threshold (cut-off) determination

To obtain DSe and DSp estimates of the candidate assay, the test results first must be reduced to categorical (positive, negative, or intermediate) status. This is accomplished by insertion of one or two cut-off points (threshold or decision limits) on the continuous scale of test results. Options and descriptive methods for determining the best way to express DSe and DSp are available. (10, 12, 16, 21 and Appendix 1.1.2.5 on statistical considerations). If considerable overlap occurs in the distributions of test values from known infected and uninfected animals, it is difficult to select a single cut-off that will accurately classify these animals according to their infection status. Rather than a single cut-off, two cut-offs can be selected that define a high DSe (e.g. inclusion of 99% of the values from infected animals), and a high DSp e.g. 99% of the values from uninfected animals). The intermediate values that fall between these percentiles may be classified as intermediate (see box), and would require testing by a confirmatory assay, retesting for detection of sero-conversion, or sequencing for identity.

The number of samples required to provide estimates of DSe and DSp at various confidence levels are available (16). For a thorough treatment of this issue, see Appendix 1.1.2.5 on best practices for statistical methods.

The main difficulty in establishing cut-offs based on diagnostic performance characteristics is the lack of availability of the required number of well-characterised samples. Alternatives are discussed in Section B.2.g. on provisional acceptance of an assay during accrual of data to enhance estimates of DSe and DSp.

4. Stage 3 – Reproducibility and augmented repeatability estimates

Reproducibility is an important parameter in establishing assay precision. It is determined in several laboratories located in distinct or different regions or countries using the identical assay (protocol, reagents and controls). Each of at least three laboratories test the same panel of samples (blinded) containing a minimum of 20 samples, with identical aliquots going to each laboratory (see Appendix 1.1.2.7 on panels of samples). This exercise generates preliminary data on the assay’s capacity to perform as expected when deployed to other laboratories, also known as ruggedness of the assay. In addition, within-laboratory repeatability estimates are augmented by the replicates used in the reproducibility studies. Measurements of uncertainty can be estimated, both for the reproducibility and repeatability data (see Appendix 1.1.2.4 on Measurement of Uncertainty for further explanation of the topic and its application).

5. Stage 4 – Programme implementation

a) Interpretation of test results.

i) Predictive values of test results. For various reasons, samples with and without the targeted analyte that are chosen for estimating DSe and DSp often are in a proportion that bears no relationship to the true prevalence of infection in the population targeted by the candidate assay. For instance if identical numbers of positive and negative reference sera are used, the artificial prevalence is 50%. However, the true prevalence may be only 10%. Determining predictive values for positive and negative test results using the artificial prevalence will not be useful. Rather, predictive values need to be based on the true prevalence. For screening assays used in surveillance of a ‘disease-free’ population, false positive results are a significant problem. For instance, an assay may have impeccable credentials (e.g. high precision and accuracy, 99% DSe and 99.9% DSp). But if the prevalence of disease is close to zero, and the assay has 1 false positive for every 1000 animals tested, false positive inferences are a problem (see reference 16 for PV tables for various DSe and DSp values). Hence, it may be critical to have perfectly specific confirmatory assays to determine whether screening assay results are true or false positives.

For nucleic acid assays, it may be necessary to confirm NAD-positive results by sequence analysis of the amplified product (an example of an assay to assist in resolving errors due to non-specific target or primer binding).

b) International recognition

Traditionally, assays have been recognised internationally by OIE when they are designated as prescribed or alternate tests for trade purposes. This has often been based on evidence of their usefulness on a national,
Chapter 1.1.2. — Principles and methods of validation of diagnostic assays for infectious diseases

For test kits that have completed the certification process, the final step is listing of the test in the OIE Register. Tests listed in the Register are certified as fit for a specific purpose if they have completed Validation Stages 1, 2 and 3. The Register is intended to provide potential test users with an informed and unbiased source of information about the test and its performance characteristics for an intended purpose. The Register is available on the OIE website at: http://www.oie.int/vcda/eng/en_vcda_register.htm

c) Deployment of the assay

Ultimate evidence of the usefulness of an assay is its successful application(s) in other laboratories and inclusion in national, regional and/or international programmes. Reference laboratories play a critical role in this process. In the natural progression of diagnostic and/or technological improvements, new assays will become the new standard method to which other assays will be compared. As such, they may progressively achieve national, regional and international recognition. As a recognised standard, these assays will also be used to develop reference reagents for quality control, proficiency and harmonisation purposes. These reference reagents may also become international standards.

An assessment of ruggedness should be repeated when the test is transferred from the development laboratory to the field and is used in pen-side applications. Predictable changes, e.g. extremes of temperature and levels of operator experience, should be assessed as additional sources of variation in assay results that may affect estimates of ruggedness (which is mostly derived from reproducibility estimates).

6. Monitoring assay performance after initial validation

a) Monitoring the assay

A validated assay in routine use needs to be consistently monitored for repeatability through process controls to evaluate possible temporal changes in test accuracy. These changes can be monitored graphically in control charts. Such monitoring is essential to maintain validation of the assay as it is implemented for routine use as a diagnostic tool. Ongoing evaluation of reproducibility is also essential through external quality control programme such as proficiency testing.

b) Diagnostic modifications - considerations for changes in the assay

Over time, modification of the assay will often be necessary to address changes in the analytes targeted (i.e., modification of the assay to adjust diagnostic performance) or technical modifications may occur to improve assay efficiency or cost-effectiveness.

If the assay is to be applied in another geographical region and/or population, revalidation of the assay under the new conditions is recommended. Lineages or sub-lineages of a virus, derived from animals in different geographic locations, are known to have different target sequences or primer sites requiring revalidation of the assay. This is especially true for NAD systems as it is very common for point mutations to occur in many infectious agents (i.e. RNA viruses). Mutations, which may occur within the primer or probe sites can affect the efficiency of the assay and even invalidate the established performance characteristics. It is also advisable to regularly confirm the target sequence at the selected genomic regions for national or regional isolates of the infectious agents. This is especially true for the primer and probe sites, to ensure that they remain stable and the estimates of DSe for the assay are not compromised.

A similar situation may occur with incursion of new viral lineages into countries or regions where that viral lineage did not previously exist. In these circumstances, existing NAD assays which did not target these novel lineages may need to be modified to include primers or probes targeting these new analytes. The same would be true for typing sera used in virus neutralisation assays.

i) Technical modifications and equivalency assessments

Technical modifications to a test such as changes in instrumentation, extraction protocols, and conversion of an assay to a semi-automated or fully automated system using robotics will typically not necessitate full revalidation of a test but rather an assessment of equivalency. Equivalency is demonstration that an assay/process and/or procedure performs equal to, or better than, the “reference” procedure, which has been validated for the intended purpose. Equivalency can be established by running the modified procedure and original procedure side-by-side, with the same control samples in both, over 20 runs and then comparing results to demonstrate no statistical differences in performance. A second approach is to determine equivalency of two NAD extraction procedures or platforms by regression analysis of Ct values for cohorts run on both platforms or subjected to different extraction procedures. For details, see Appendix 1.1.2.6 on equivalency testing and Appendix 1.1.2.7 on reference sample panels.
An equivalency assessment may be adequate if the test is applied to a different sample matrix, e.g. validated on blood and used on another tissue in the sample species. Revalidation may be necessary if the test is used in a new species.

ii) Replacement of depleted reagents

When a reagent such as a control sample is nearing depletion, it is essential to prepare and repeatedly test a replacement before such a control is depleted. The prospective control sample should be included in multiple runs of the assay in parallel with the original control to establish their proportional relationship. Whenever possible, it is important to change only one reagent at a time to avoid the compound problem of evaluating more than one variable.

c) Enhancing confidence in validation criteria

Because many host variables impact on the diagnostic performance of assays, it is highly desirable over time to increase the number of reference samples from animals of known infection status. This improves the precision of the overall estimates of DSe and DSp, and may allow calculations of DSe estimates by factors such as age, stage of disease, and load of organisms. New data should be included annually in relevant test dossiers.

REFERENCES


*  
*  
*  
APPENDIX 1.1.2.1.

DEVELOPMENT AND OPTIMISATION OF ANTIBODY DETECTION ASSAYS

IN PREPARATION

* * *

* * *
APPENDIX 1.1.2.2.

DEVELOPMENT AND OPTIMISATION OF ANTIGEN DETECTION ASSAYS BY IMMUNOLOGICAL MEANS

IN PREPARATION

*  *
*  *
APPENDIX 1.1.2.3.

DEVELOPMENT AND OPTIMISATION OF NUCLEIC ACID DETECTION (NAD) TESTS

IN PREPARATION

* *
* *
APPENDIX 1.1.2.4.

MEASUREMENT OF UNCERTAINTY

IN PREPARATION

*  *
  *  *
APPENDIX 1.1.2.5.

STATISTICAL APPROACHES TO VALIDATION

IN PREPARATION

* *
**
APPENDIX 1.1.2.6.

EQUIVALENCY

IN PREPARATION

* 
** 
***
APPENDIX 1.1.2.7.

SELECTION AND USE OF REFERENCE PANELS

IN PREPARATION

*  
* *