CHAPTER 2.2.2.

DEVELOPMENT AND OPTIMISATION OF ANTIGEN DETECTION ASSAYS

INTRODUCTION

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

The detection and identification of an agent is confirmatory evidence of either infection with or disease caused by a particular pathogen. There are many different direct and indirect test methodologies available. Classical direct[1] agent detection assays include electron microscopy, light microscopy (*e.g. observation of unique histopathological or pathognomonic features, identification of parasites* in situ*, etc.*)*, virus isolation, bacterial culture and parasitic digestion techniques. Many direct techniques require secondary procedures to assist in the characterisation and identification of these agents* (*e.g. haemagglutination inhibition or virus neutralisation tests, special bacteria or fungal stains, etc.*)*. Representatives of indirect agent and in particular, antigen* (*Ag*) *detection tests include enzyme linked immunosorbent assays* (*ELISA*)*, immunofluorescence or immunoperoxidase assays, Western blotting techniques, micro-arrays, fluorescence activated cell sorting* (*FACS*) *and biosensors. Often direct and indirect methods are used in tandem, for example a direct technique may be used to isolate, enrich and/or extract the organism followed by indirect techniques to characterise and identify the agent.*

Practical matters in selecting an assay format

- Is high-throughput essential? Will the test be automated?
- Will the test be used as a herd test or for testing individuals?
- What is the anticipated turnaround time? Is that suitable?
- What level of sophistication is needed to run the assay?
- What skills are required to interpret test results?
- Will the assay be feasible for use in my laboratory?
- Will it be easily transferrable to other laboratories?

The examples given above vary widely in terms of laboratory requirements. Pen-side tests (*including lateral flow devices*) *may also be antigen detection tests. They are easy to use and were developed for testing of animals in the field but may have applications in laboratories. Cost, laboratory infrastructure, biocontainment/safety, technical sophistication, interpretation skills, turn-around time, throughput capacity, diagnostic performance characteristics, repeatability and reproducibility are important parameters which need consideration when selecting the most appropriate assay. They also vary with respect to suitability for different diagnostic applications.*

1 Definitions: In the context of direct and indirect detection of an analyte, the term "direct" is relevant to the organism or its antigens, e.g. direct detection by microscopy. The term "indirect" is relevant to the host response to the organism, (e.g., indirect detection inferred by detection of antibodies to the organism). In the context of diagnostic methods that are used to detect infection, microscopy is a means of direct observation of the organism, whereas using a reagent that infers the presence of the agent (such as an enzyme conjugated to a purified antibody that is specific for the agent) is an indirect method.

In contrast to serological or antibody detection tests, Ag detection tests depend heavily on the time of clinical onset and pathogenesis of the disease and the concentration of the pathogen in certain tissues and/or fluids. Successful diagnosis depends on appropriate timing and selection of the sampling site (affected tissue/lesions, scrapings, swabs, blood, and other body fluids), storage conditions *and specimen integrity during transport. For certain applications, the testing of individual animals and/or samples may be appropriate (e.g. confirmatory testing); while for other purposes (e.g. screening), pooling of animals may be efficient and effective. Selection of the appropriate specimen requires good understanding of the disease and the effect of the sample matrix on the pathogen (e.g. cloacal or tracheal swabs for avian influenza).*

Nucleic acid detection (NAD) tests are increasingly replacing classical antigen detection systems. For development and optimisation of NADs, please refer to Chapter 2.2.3. Although these NAD tests may seem to be the tests of choice for many applications, they are not always the most practical or efficient. In most cases it is still necessary and prudent, at least for an index case, to culture the agent on selective media or in susceptible cell lines or eggs to facilitate further characterisation and identification. While genotyping is an important consideration, especially in molecular epidemiology, other means of agent characterisation such as serotyping, pathotyping or biotyping are also important. Cultured and preserved agents have tremendous historical value and are also an important source of reference materials.

Due to its worldwide application, the Ag ELISA is used as an example for application of best practices in antigen detection assays in this chapter. Most of the basic processes used to validate other types of antigen detection assays will become evident by extension of those used to validate ELISAs. Because of the many conceptual similarities between antigen and antibody detection assays this chapter frequently cross-refers to Chapter 2.2.1 on antibody detection.

A. ANTIGEN DETECTION ASSAY DEVELOPMENT PATHWAY

1. Intended purpose(s) of the antigen assay

The intended purpose of the test is a key factor which will guide decisions in the selection and early development of the candidate assay. Given the WOAH-defined 'fitness for purpose' categories in Table 1, Ag detection systems may be appropriate for certain applications. Support for disease eradication or surveillance programmes generally require testing of high numbers of samples, with an emphasis on diagnostic sensitivity and throughput capacity. In contrast, confirmation of clinical cases does not entail high numbers to be tested, but diagnostic specificity and turn-around-times become very important. At the outset, the questions posed in the text box, above, should be carefully considered.

Table 1. Determinants of an antigen assay's fitness for its intended purpose

Assay characteristics	*Determinants of fitness for purpose*						
	*1**		*2**	*3**	*4**	*5**	*6**
	a	b					
Diagnostic sensitivity (DSe)		+++	+++	+++	+++		
Diagnostic specificity (DSp)		+	+	+	+++		
Positive predictive value (PPV)		+	+	+	+++		
Negative predictive value (NPV)		+++	+++	+++	+++		
Throughput capacity		+++	++	+	–		
Turn-around time of test		+	+	+	+++		

Assay characteristics	*Determinants of fitness for purpose*						
	*1**		*2**	*3**	*4**	*5**	*6**
	a	b					
QA capability		+++	+++	+++	+++		
Reproducibility		+++	+++	+++	+++		
Repeatability		+++	+++	+++	+++		

Other characteristics such as the technical sophistication of the assay, and the skill required for interpretation will be related to the disease or infection under investigation.
NB NAD tests may also be used for agent detection, and are considered in Chapter 2.2.3.
Symbols: +++ = essential; + = of less importance; – = not important.
*Basic purposes for which an assay may be deemed fit: 1. Contribute to the demonstration of freedom from infection in a defined population. 2. Certify freedom from infection or presence of the agent in individual animals or products for trade/movement purposes; 3. Contribute to the eradication of disease or elimination of infection from defined populations; 4. Confirmatory diagnosis of clinical cases (includes confirmation of positive screening test); 5. Estimate prevalence of infection or exposure to facilitate risk analysis; 6. Determine immune status in individual animals or populations (post-vaccination)

1.1. Purpose 1

For disease freedom categories as given in purpose (Table 1), antibody screening tests of high diagnostic sensitivity (DSe) are usually the tests of choice provided that a detectable immune response is a significant indicator of infection. However, there may be certain disease situations where the humoral immune response can be misleading and pathogen detection may be more appropriate (e.g. mycobacterial or trypanosomal infections). In these cases, Ag detection tests may be more effective. The antigen screening test should demonstrate a high DSe. Tests of high DSe demonstrate low false negative (FN) rates and when applied to low prevalence populations, the NPV is at its highest level. As DSe and diagnostic specificity (DSp) are usually inversely related, a decrease in DSp will result in an elevated false positive (FP) rate. Therefore, all screening test positive results should be subjected to some form of confirmatory testing to evaluate their true status. Confirmatory tests characteristically have high DSp and therefore a low FP rate. These tests are often more sophisticated and require enhanced interpretive skills.

1.2. Purpose 2

For a number of diseases included in the *Terrestrial Manual*, agent identification is listed as the preferred test method for the purpose of determining "Individual animal freedom from infection prior to movement". Although in many cases this involves culture of the organism or detection of nucleic acid, there may be situations where an antigen detection test is appropriate for this purpose. To avoid risk of disease spread through trade, a test offering high DSe is to be preferred.

1.3. Purpose 3

If the purpose of the test is the eradication of disease or elimination of infection from defined populations, antigen screening tests, like antibody screening tests, of moderate to high DSe, are the tests of choice. However, the rationale is slightly different in that the testing will likely be done at the herd or compartment level. At the beginning of the eradication campaign, when the disease prevalence is high, moderate DSe and DSp are suitable as both FP and FN rates are less relevant at this juncture and a moderate level of error is tolerable. Depending on the nature of the disease and rapidity of spread, high throughput and fast turn-around-times may become critical. Usually decisions are made without confirmatory testing at this point.

In the latter stages of the campaign, a higher DSe is warranted as the FN rate becomes the more critical factor. Much like Purpose 1, positive reactors will need to be subjected to some form of confirmatory testing to evaluate their true status. In these latter stages, antigen and/or nucleic acid detection systems are critical in the detection of sub-clinical cases, shedders and possibly, latent carriers.

1.4. Purpose 4

Although antibody tests of high DSp are often the tests of choice for confirmatory diagnosis of clinical cases, antibody tests may not always be the tests of choice, especially if clinical signs appear before an immune response is mounted. A prime example would be highly pathogenic avian influenza infections where mortality may occur before an immune response is even detectable. Antigen and/or nucleic acid detection tests are usually a better choice for confirmation of clinical cases provided that they offer a fast turn-around-time. In these cases, the idea is to maximise DSp and thereby minimise any potential FP reactions. For some clinical cases, e.g. vesicular diseases in terrestrial animals, several tests may be required to quickly rule out select pathogens that present similar clinical signs. In this category of test, fast turn-around-times are extremely critical in identifying potential outbreaks.

1.5. Purpose 5

For estimates of prevalence of infection and/or shedding to facilitate risk analysis, e.g. for health surveys, herd health status and to monitor disease control measures, antigen detection tests of moderate DSe and DSp are the tests of choice. In general, this would balance both FN and FP rates and result in a more accurate estimate of the true prevalence in the target population. However, if accurate estimates of both DSe and DSp have been established, statistical approaches can be used to minimise bias attributable to FN and FP rates (see Chapter 2.2.5 *Statistical approaches to validation*).

1.6. Purpose 6

Not applicable to antigen detection assays.

2. Assay development – experimentation

2.1. Reference materials, reagents and controls

2.1.1. Test samples

Samples required for antigen detection assays should be handled as described in Chapter 1.1.2 *Collection, submission and storage of diagnostic specimens*. Sample matrixes for antigen detection assays can be very heterogeneous (e.g. blood, faeces, milk, skin, semen, saliva, blisters, vesicles or swabs from affected tissues such as oropharynx (Probang), trachea, genitals, cloaca, oesophagus, etc.). The ideal specimen is the one that is easy to obtain and with a high concentration of the analyte. In many cases blood or swabs are the specimens of choice but depending on the pathogen, other tissues or fluids are needed, e.g. skin, organs such as brain (rabies, transmissible spongiform encephalopathies [TSE]), lymphatic organs, such as spleen and lymph nodes (classical swine fever), kidney, liver, heart, parts of respiratory tract (avian influenza), digestive tract (parvovirus), milk, faeces, semen, saliva, tumour material (enzootic bovine leukosis), etc.

As stated in Chapter 1.1.2, the usual considerations apply to limit bacterial and fungal contamination of specimens. The use of preservatives and fixatives is not usually recommended and samples should be sent with minimal delay and under refrigeration to a diagnostic laboratory. During transport and storage it is important to be aware of the physical and chemical requirements of the pathogen (e.g. foot and mouth disease [FMD] virus is highly labile at low pH and requires equal amounts of glycerol and phosphate buffer to maintain a pH over 7).

If samples are to be tested as pools, experiments need to be undertaken to demonstrate that the assay is fit for that purpose (e.g. that the analytical sensitivity is sufficiently high to detect one infected animal in a pool of 5, 10, 50 or more samples from uninfected animals).

2.1.2. Reference Standards

See Chapters 2.2.1 *Development and optimisation of antibody detection assays* and 2.2.6 *Selection and use of reference samples and panels*.

2.1.3. Positive and negative reference panels

Samples, containing concentrations of antigen over the intended operating range of the assay, should be used as controls during the development and standardisation of an antigen detection assay. They may be obtained from field specimens or produced in the laboratory as spiked samples (see Chapter 2.2.6). Negative samples should be obtained from known uninfected animals and this same matrix should be used when spiked samples are produced.

2.1.4. Purified and crude antigens for antibody production

In general, antigens to be used for the production of immunological reagents should be as natural as possible in terms of conformation to ensure that the presentation of epitopes mimics the orientation on the live organism. Therefore, isolation and/or purification methods used should preserve the antigenic integrity of the agent as much as possible.

For very large pathogens such as poxviruses, bacteria and protozoal parasites, protein microarrays may be useful to identify and select antigens which elicit strong immune responses.

2.1.5. Monoclonal and polyclonal antibodies for indirect antigen detection assays

Monoclonal antibodies demonstrate unique analytical specificities and are very useful in the detection of agent-specific epitopes at the group, strain or sub-strain levels. As such, they need to be considered carefully with respect to the application at hand and the desired specificity of the assay. Polyclonal antibodies, by nature, tend to demonstrate a broader range of specificities. Purified or semi-purified polyclonal antibodies are often the reagent of choice for trapping complex antigens because they usually demonstrate higher affinities than their monoclonal counterparts.

2.2. Design of test method

2.2.1. Choice of test

The prologue to a proper test design requires careful consideration of many variables in the context of performance requirements. The choice of assay must be coupled with its intended application, which usually necessitates consensus between the assay developer, statisticians, and other stakeholders such as epidemiologists and regulatory bodies. If the purpose is to develop a screening test (e.g. during a post-outbreak surveillance period) the emphasis will be on high DSe, high throughput, low cost, technical simplicity, low interpretative skill, etc. If the purpose is to develop a confirmatory test (e.g. for the confirmation of clinical cases or confirmation of positive screening test reactors), a different set of priorities will come into play including high DSp, fast turn-around-times, technical sophistication and interpretative skills. There are a growing number of point-of-care or pen-side tests that have their own set of additional robustness and ruggedness requirements given the variable conditions of the environment in which they will be used and the skill level of the operator who will performing and interpreting the test.

Aspects affecting choice of test

- Is the assay to be used for screening or confirmatory purposes, or both?
- Will the assay be used as a laboratory or field based test?
- Will it be used for one or more species? Which ones?
- Will the test be used for typing organisms to group, serotype or strain-specific levels?
- Will the test be applied nationally or internationally?

The antigen ELISA is conceptually the same method employed for the antibody ELISA (Chapter 2.2.1), with the exception that antigen is the targeted analyte, and antibodies are the primary reagents used for capture and detection of antigens. Depending on whether the antigen is adsorbed directly on the microplate or captured by antibodies on a solid phase, along with subsequent detection steps, different formats are available.

It is not always apparent which assay format should be used to best fit the intended purpose. Availability of reagents and the limit of detection of the assay for a particular application may be significant factors in limiting the choice. Since many of the systems now target highly specific antigens, the choice of antibody for both trapping and/or detection becomes critical.

Preparation of the test sample is also a critical consideration depending on the test format being used. The use of trapping or capture antibody in sandwich-type assays enhances selectivity and reduces potential matrix effects. For assays requiring direct application of the analyte to the solid phase, preparatory extraction, centrifugation or filtration methods may be necessary to remove extraneous material. For a more in depth discussion of these different assay formats, please see Crowther (2001).

Of critical importance is the size and complexity of antigen and the availability of relevant reagents, such as capture antibodies (e.g. antigens to be detected in sandwich assays must have at least two unfettered epitopes) which limits this assay type to relatively large antigens or whole pathogens. The affinities of the immunological reagents come into play as the stability of the resulting antibody-antigen complexes in the microplate or on beads will affect the performance characteristics of the assay.

Practical concerns are availability and use of antigen standards for quality control and assurance purposes, repeatability, reproducibility, throughput capacity, turn-around-time of a test result, cost and technical sophistication and interpretation skills.

Working with exotic and/or zoonotic agents requires particular attention to biosafety and biosecurity regulations (see also Chapter 1.1.4 *Biosafety and biosecurity: Standard for managing biological risk in the veterinary laboratory and animal facilities*).

2.3. Proof of concept experiments (feasibility studies)

The same types of experiments used in antibody detection tests are required for antigen detection tests (see Chapter 2.2.1).

2.4. Samples and data expression

2.4.1. Selection, storage and use of control samples for test development and validation studies

It is important to assess and monitor sensitivity and specificity of the test during development and validation. This is achieved by selecting several samples (4–5 is adequate) that range from negative to high levels of antigen of the analyte in question. These samples are used in experiments designed to optimise the assay. To achieve continuity of evidence requires care and forethought in preparation and storage of samples. A large volume (e.g. 10 ml) of each control sample is acquired and divided into 0.1 ml aliquots for storage at or below –80°C. One aliquot of each sample is thawed, used for experiments, and ideally then discarded. If it is impractical to discard the aliquot, it may be held at 4°C between experiments for up to about 2 weeks; however, there is a possibility of sample deterioration under these circumstances. Then, another aliquot is thawed for further experimentation. This method provides the same source of sample with the same number of freeze–thaw cycles for all experiments (repeated freezing and thawing of samples can denature antigen and/or facilitates the growth of other unwanted microorganisms and should be avoided). Also, variation is reduced when the experimenter uses the same source of sample for all experiments rather than switching among various samples between experiments. This approach has the added advantage of generating a data trail for the repeatedly run samples. After the initial stages of assay validation are completed, one or more of the samples can become reference reagent(s) that are the basis for data expression and repeatability assessments both within and between runs of the assay (Jacobson, 1998). They may also serve as in-house working standards if their activity has been predetermined; such standards provide assurance that runs of the assay are producing accurate data (Wright, 1998).

2.4.2. Normalisation of results and their expression

The same normalisation procedures as used for antibody detection assays are applicable to antigen detection assays (see Chapter 2.2.1 for details).

2.5. Optimisation

The aim of this stage is to finalise test parameters related to reagents, consumables and equipment that will lead to a fixed protocol and will be used during the assay validation pathway Part B. For details, see Chapter 2.2.1.

Optimisation and standardisation

- Is the test able to distinguish between samples with and without the analyte (what is the limit of detection = analytical sensitivity?)
- Does the test cross-react with non-target antigen in the sample or sample matrix (analytical specificity)?
- What is the noise or background activity in a negative sample?
- Has the repeatability been assessed for a range of control samples over a number of days
- Are sufficient positive and negative samples on hand to carry out the experiments for optimisation and validation
- If yes, are reference and control samples dispensed and stored properly to avoid introduction of sample related bias (sample deterioration)?
- Have all critical reagents been run against each other in checker board titrations?
- Did you find optimal concentration/dilutions for each reagent?
- Did you include reference reagents and working standards/controls and did you normalise the OD data to achieve the best possible comparative results?

2.6. Inhibitory factors in sample matrix

Due to a higher variety and complexity of specimens, antigen detection assays are more likely to be influenced by matrix factors than antibody detection assays, which normally detect antibodies in serum. Inhibitory substances are frequently found in complex matrixes such as pus, semen, tracheal/nasal/cloacal swabs and may have an impact on the test result. ELISA antigen detection systems are rather resistant to inhibitory factors, please see the WOAH Validation Standard, Section A.2.4, and Greiner *et al.* (1997) for descriptions of the type of inhibitors that could affect the assay. These references should be reviewed carefully to assure that all inhibitory factors are accounted for and controlled.

2.7. Calibration to reference sample and comparison to standard test method

Chapter 2.2.1 contains information relevant to this procedure.

B. ASSAY VALIDATION PATHWAY

1. Stage 1 – Analytical performance characteristics

1.1. Repeatability

Chapter 2.2.1.

1.2. Analytical specificity

Analytical specificity (ASp) is defined as the degree to which the assay distinguishes between the target analyte and other components that may be detected in the sample matrix (see Chapter 2.2.6, Section B.1). The higher the ASp, the lower the number of false positive results. ASp should be determined by testing well characterised samples from similar or related pathogens, which produce similar lesions as the target pathogen or are frequently found in samples containing the target pathogen. For example, to assess the ASp of a

Analytical performance characteristics

- Has repeatability been established for a range of positive and negative samples within and between runs of the assay?
- Have upper and lower control limits of the assay been established?
- Have you defined ASe and ASp for this assay?
- Does the candidate assay compare favourably with a standard test method, based on objective quantitative and qualitative criteria?

FMD antigen detection ELISA for one particular serotype (e.g. serotype O), it is necessary to assess its reactivity of all sub-strains within this serotype (e.g. O Campos, O Manisa, etc.) to assess inclusivity. At the same time it is important to show that the test does not cross-react with other serotypes such as A, Asia 1, C, SAT 1, 2 and 3. Finally, there is also a need to assess whether the test cross-reacts with agents from diseases which may produce similar signs, e.g. vesicular stomatitis, swine vesicular disease and swine vesicular exanthema. Another example is an ELISA designed to detect avian influenza virus: as a screening test the assay should detect the nucleoprotein or matrix antigen of all subtypes, e.g. H1-H16 and N1-N9. However it should not cross-react with viruses which cause similar clinical signs such as Newcastle disease or infectious bursal disease (diagnostic specificity) or with other non-specific components present in the matrix or on the solid phase. Some ELISAs may be subject to false positive results attributable to non-specific factors, such as non-specific binding of antibody conjugates to the plastic surface and may require use of blocking agents. Care must be taken to eliminate these types of errors.

1.3. Analytical sensitivity

Analytical sensitivity (ASe) is synonymous with the lower limit of detection (LOD) of antigen concentration in a sample. LODs are usually determined by endpoint dilution in which replicates (preferably 10) of each dilution in a $\log_2$ dilution series are run in the assay. The larger the number of replicates, the more precise the determination of the dilution at which the antigen is no longer detectable. Further information on LODs and ASe is given in the WOAH Validation Standard and in Chapter 2.2.5.

Screening assays or assays which are designed to detect sub-clinical infections or carriers should have a very high ASe. In these cases it may be difficult to obtain suitable samples and to determine the comparative ASe by running a panel of samples on the candidate assay and on another independent assay. If available, serial samples from experimentally infected animals could provide temporal information about the assay's capacity to detect antigen over the course of infection.

1.4. Standard test method comparison with the candidate test method

Chapter 2.2.1.

2. Stage 2 – Diagnostic performance characteristics

See also Chapter 2.2.1.

2.1. The challenge in establishing accurate estimates of DSe and DSp for antigen assays

For all antigen detection assays including ELISAs, particular consideration must be given to the timing when the sample is taken as the probability of detecting antigen or the pathogen itself is very closely linked to the stage of infection. The diagnostic window will likely be much smaller in antigen/pathogen detection assays than in antibody detection tests as immune responses can normally be measured over an extended period of time. Dynamics of infection, e.g. acute, persistent, sub-acute, chronic or carrier-state are important determinants for sampling recommendations. For example, during an acute viral infection, the sample should be taken as early as possible after the onset of clinical signs. In persistent, sub-acute, chronic or carrier animals there is a balanced relation between the pathogen and the host and the agent may be present in minute concentrations which may be very difficult to detect. During the course of pathogenesis, other organ systems may become involved and different tissues or fluids may be more appropriate target tissues for sampling as the disease progresses.

Diagnostic performance characteristics

- Are the criteria used to determine the positive and negative reference populations legitimate?
- Do the reference samples fully represent the population targeted by the assay?
- Were there difficulties in obtaining a sufficient number of samples? If so, how was the problem addressed?

2.2. Reference animal populations

2.2.1. Animals of "known infection status"

Depending on the composition of positive and negative reference samples, the same test may have different estimates for DSe and DSp. Ideally the composition of reference samples and animals should match as closely as possible the samples which are expected from the target population for which the test was developed. It is important to have a clear case definition. This is a set of criteria used to decide whether an individual or group of animals is infected or not. The reference status must relate to the purpose of the testing. For example, if the purpose of the assay is to be used as a screening test to detect early infection of FMD (e.g. in free-ranging cattle with vesicles), then the majority of samples to determine DSe and DSp should be taken from this target population. Relevant information should be collected and summarised for all animals involved at this stage of validation (e.g. species, age, sex, breed), and information on other factors that are known to influence DSe and DSp (e.g. date and place of sampling, immunological status, vaccination and disease history, pathognomonic and surrogate tests used to define status of animals, prevalence within population and description how the reference status was derived).

A sample from a negative reference animal refers to lack of exposure to or infection with the agent in question. For example a classical swine fever negative population could be defined as a region with pig herds without confirmed clinical cases of the disease in recent years, supported by negative serological tests and negative virological test results of suspected cases. Samples from these animals fulfil the status of negative reference samples. The negative reference population should be chosen with care so as to ensure that it is representative and matches the positive reference population (e.g. as to the breed and exposure to environmental challenge).

If vaccination is carried out, it may interfere with antigen detection (e.g. vaccination with modified live viral vaccines). Samples from these animals should not qualify as negative reference samples.

The types and limitations of reference standards commonly used for evaluation of the performance characteristics of a new assay are listed here. An expanded description of each reference standard is provided in Chapter 2.2.1, Section B.2.2.1 and in Jacobson, 1998. The strengths and limitations of these reference standards must be considered carefully when using samples, derived from animals that fall into any of the following four categories, as sources for establishing DSe and DSp for a candidate assay (Jacobson, 1998).

i) An unequivocal reference standard

   An unequivocal reference standard: presence of the agent in the host or evidence of definitive (pathognomonic) histopathology.

ii) A composite reference standard

   A composite reference standard: verification of uninfected or unexposed animals.

iii) A relative reference standard

   A relative reference standard: reference animals that have been classified for their infection status by comparison with the test results of another assay for antigen or nucleic acid detection on the same samples. As with antibody assays, the estimates of DSe and DSp are useful only insofar as the reference test has documented, established and acceptable performance characteristics. A deficiency of relative reference standards is that they have their own established levels of FP and FN test results, which are sources of error that will be compounded in estimates of DSe and DSp for the new assay. Generally, however, the use of other well described test methods is regarded as good practice to determine the status of reference animals, but only if the inherent bias introduced by the relative reference standard is accounted for.

iv) An adjunct reference standard: experimental infection or vaccination

See the WOAH Validation Standard, Section B.2.3 for significant limitations of this type of standard. Note: In the context of antigen detection, "comparative" tests should include pathogen detection by isolation, culture, NAD tests or histopathology or other *in-situ* techniques.

2.2.2. Latent-class models for sample selection

For a discussion of this approach to sample selection, see the WOAH Validation Standard, Section B.2.5 and Chapter 2.2.6.

2.3. Threshold (cut-off) determination

It is important to clearly describe the method and the samples used for selecting a cut-off. It is strongly recommended to conduct a receiver operating characteristic (ROC) analysis to show the potential performance of the test in other epidemiological settings.

3. Stage 3 – Reproducibility and augmented repeatability estimates

Reproducibility assessments for antigen detection assays are not uniquely different from similar assessments for any other type of assay. Therefore the reader is directed to the WOAH Validation Standard, Section 3, for details on reproducibility analysis and for reference samples and panels to Chapter 2.2.6. WOAH provides guidelines for laboratory proficiency testing (https://www.woah.org/en/what-we-offer/expertise-network/reference-laboratories/#ui-id-5).

4. Stage 4 – Programme implementation

4.1. Interpretation of results

See the Chapter 2.2.1, Section B.4.1.

5. Monitoring assay performance

5.1. Monitoring the assay

After performance characteristics of the new test have been established, on-going monitoring, maintenance and enhancement are required. It is important to continue monitoring the repeatability and reproducibility of the assay over time, on a per run basis, quality control samples have to fall within pre-established limits. If not, the test result is not valid and has to be repeated. Monitoring the assay controls' performance over time is an important way to detect changes or trends in the assay. Simple analysis of results, e.g. statistical assessment of mean values, standard deviation and coefficient of variation are useful in this process and results can be plotted in a control chart. Participation in external quality control or proficiency testing programs is useful to identify random and systematic errors and provide credibility in test results. Further information about this topic can be obtained from Crowther *et al.* (2006) and in the Chapter 2.2.1, Section B.5.1.

5.2. Minor modifications of the assay – replacement of depleted reagents

Over time changes in the test protocol may be necessary due to better or less costly reagents or because the target analyte has changed. Batch-to-batch variation of biological reagents is considered a major contributor to test variation. When reagents such as antibodies or antigen must be replaced they should be produced or procured using the same protocols or criteria as used for the original reagents. New biological reagents (e.g. control samples, antigen, capture or detection antibodies, conjugate, chemicals or consumables) need to be assessed for comparability. Chapter 2.2.8 *Comparability of assays after changes in a validated test method* gives an overview of acceptable comparability studies. A ground rule is never to change more than one reagent at a time in order to avoid the compound problem of evaluating more than one variable concurrently (see also Chapter 2.2.1, Section B.5.2).

5.3. Major modifications of the assay – changing to a new ELISA type

It is a major challenge of laboratory diagnosis to keep up with the evolving nature of infectious pathogens. Over time pathogens may change their antigenic characteristics and new strains may emerge, e.g. the emergence of Bluetongue virus serotype 8 in northern Europe in 2009. This change necessitates a full test development and validation study. Another major change is the use of a test in a different species than that for which it was originally validated, e.g. there may be a requirement to use an FMD Ag detection ELISA, validated in cattle, for testing camelids or buffalo in different geographical and climatic regions. Evaluation of reference samples that represent those populations of Stage 2 in Figure 1 of the WOAH Validation Standard will accomplish this requirement (see also Chapter 2.2.1, Section B.5.3).

5.4. Enhancing confidence in validation criteria

Due to the extensive set of variables that have an impact on the performance of antigen detection assays, it is useful to expand the number of reference samples when possible, due to the principle that error is reduced with increasing sample size (see also Chapter 2.2.1, Section B.5.4).

REFERENCES

CROWTHER J.R. (2001). The ELISA guidebook. *In:* Methods in Molecular Biology. Humana Press, Totowa, NJ, USA, 1–421.

CROWTHER J.R., UNGER H. & VILJOEN G.J. (2006). Aspects of kit validation for tests used for the diagnosis and surveillance of livestock diseases: producer and end-user responsibilities. *Rev. sci. tech. Off. int. Epiz.*, **25**, 913–935.

GREINER M., BHAT T.S., PATZELT R.J., KAKAIRE D., SCHARES G., DIETZ E., BÖHNING D., ZESSIN K.H. & MEHLITZ D. (1997). Impact of biological factors on the interpretation of bovine trypanosomosis serology. *Prev. Vet. Med.*, **30**, 61–73.

JACOBSON R.H. (1998). Validation of serological assays for diagnosis of infectious diseases. *Rev. sci. tech. Off. int. Epiz.*, **17**, 469–486.

WRIGHT P.F. (1998). International standards for test methods and reference sera for diagnostic tests for antibody detection. *Rev. sci. tech. Off. int. Epiz.*, **17**, 527–533.

*
* *

NB: FIRST ADOPTED IN 2014.