CHAPTER 2.2.1.

DEVELOPMENT AND OPTIMISATION OF ANTIBODY 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.*

Detection of antibodies that are elicited in response to infectious agents or their components constitutes an indirect means of laboratory-based disease diagnosis. The most common antibody detection methods are classical virus neutralisation test (VNT), enzyme-linked immunosorbent assay (ELISA), haemagglutination inhibition (HAI) and the complement fixation test (CFT). Other, less common, antibody detection tests are the agar gel immunodiffusion (AGID), the indirect fluorescent antibody test (IFAT), the serum agglutination test (SAT), the latex agglutination test (LAT), and the microscopic agglutination test (MAT). More recent novel methods include biosensors, bioluminometry, fluorescence polarisation, chemoluminescence and lateral flow devices also known as point of care or pen-side tests. Other immunological assays that use antibodies in antigen detection tests are described in Chapter 2.2.2.

When considering a candidate assay type for disease diagnosis, one should include antibody detection assays because of their practicality, ease of sample collection and preparation, generally good diagnostic performance characteristics, suitability for automation (high-throughput), low cost and fast turn-around time. They are particularly useful for processing large numbers of samples in epidemiological and population studies, or for mass diagnosis and surveillance programmes. Antibody assays are also widely used for export, import and trade of animals, and still represent the majority of WOAH recommended tests for international trade.

A characteristic of antibody assays is their capacity to indicate prior exposure to an infectious agent in the absence of detectable organisms or their analytes. They are also adaptable to a variety of matrices, such as serum, plasma, whole blood, milk, lacrimal secretions and saliva. Immunoglobulin isotype or subclass-specific test systems may selectively target early or late immune responses, e.g. IgM and IgG, respectively. Specifically designed detection systems allow differentiation between responses to vaccine and field strains and are available as commercial kits, e.g. the detection of antibodies to classical swine fever virus in pigs. Competitive or blocking formats allow use of the same basic assay for a variety of animal species while other formats are species specific. Many types of chemical or physical indicators are used to indicate the presence of specific antibody in a specimen (chromogens, fluorochromes, agglutinins, among many others). Because of the large number of antibody detection methods available, it is not possible to describe the best practices for validation of each of these assay types in this chapter. The most widely used antibody detection system, the ELISA, will therefore be used as an example for application of best practices in antibody assays. Most of the basic processes used to validate other types of assay systems will become evident by extension of those used to validate ELISAs.

A. ANTIBODY DETECTION ASSAY DEVELOPMENT PATHWAY

1. Intended purpose(s) of the antibody assay

The first consideration in assay development is to define clearly the specific purpose and application of the test to be developed. Many decisions in developing assays will be based on these first considerations. For antibody detection assays (hereafter in this chapter designated as "antibody assays") such as ELISA, such knowledge will guide the selection of the most appropriate type of antibody detection system to achieve the intended purpose. Many factors related to the assay's intended purpose, use, and suitability need to be taken into account (see the WOAH Validation Standard for other possible purposes).

The six basic intended purposes for diagnostic assays are stated in the WOAH Validation Standard, and listed in the footnote to Table 1 below. Because antibody assays have such a broad range of applications, and can be configured for very specific purposes, it is useful to consider and evaluate several parameters when establishing the specific purpose(s) for the candidate assay. Table 1 summarises characteristics of antibody assays when applied for different purposes. Consideration of these characteristics will provide guidance in establishing the specific purposes for which the candidate assay will be fit.

Note – The reader is advised to read Section B.4. *Programme implementation*, as a primer for the following discussions. That section describes the inter-relationships between diagnostic sensitivity and specificity, false positive and negative test errors, and positive and negative predictive values. For a more in-depth discussion of predictive values as a function of prevalence, see Jacobson, 1998.

Table 1. Determinants of an antibody 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	+	+	+	+	+++	–	+
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.

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 purposes 1a and 1b (Table 1), antibody screening tests of high diagnostic sensitivity (DSe) are the tests of choice. As indicated in the purposes above, these tests would be applied to populations that have an apparent prevalence of zero. Tests of high DSe demonstrate low false negative (FN) rates and when applied to low prevalence populations, the negative predictive value (NPV) is at its highest level. However, DSe and diagnostic specificity (DSp) are usually inversely related and as such, a decrease in DSp will result in an elevated false positive (FP) rate. Other considerations, if

this is to involve a continuous volume of surveillance samples, would include high throughput, low cost and technical simplicity. 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, more costly and may require enhanced interpretive skills.

If demonstration of freedom from infection is to be achieved after an outbreak, in which vaccination has been used for disease control, then screening of massive numbers of sera is often required. In addition to the considerations above, this also necessitates an antibody detection test which is able to distinguish between infected and vaccinated animals (i.e. a DIVA [differentiation of infected from vaccinated animals] test). At the same time an antigen or nucleic acid detection test may be warranted in some situations to prove that shedding and/or circulation of the infectious agent has ceased.

1.2. Purpose 2

If the purpose is to qualify individual animals for international movement, antibody screening tests of high DSe are again the tests of choice. The same rationale as stated above applies with respect to the NPV. Again, all positive reactors will need to be subjected to some form of confirmatory testing to evaluate their true status or may be excluded from shipment without further testing. In cases where borderline positives are observed, it may be wise to request a repeat sampling of the animal(s) at a suitable time interval to ensure that herd/flock has not been very recently infected.

1.3. Purpose 3

If the purpose of the test is the eradication of disease or elimination of infection from defined populations, 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 herd or compartment level. At the beginning of the 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 test 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 Purposes 1 and 2, positive reactors will need to be subjected to some form of confirmatory testing to evaluate their true status. In these latter stages, antibody detection tests are often applied in conjunction with antigen and/or nucleic acid detection systems to detect subclinical cases and possibly, latent carriers.

1.4. Purpose 4

For the confirmatory diagnosis of clinical cases, antibody tests of high DSp are the tests of choice. In these cases, the idea is to minimise the FP rate and enhance the PPV of the test. As a general rule, infection is well established and the immune response is usually well underway. In some situations it may be preferable to carry out a screening test of high DSn but a lower DSp, then following up positives with a high DSp confirmatory test. For some clinical cases, e.g. vesicular diseases in terrestrial animals, several tests may be required to rule out select pathogens that present similar clinical signs. In some cases, antigen and/or nucleic acid detection tests may be a better choice for confirmation of clinical cases provided that they offer a fast turn-around-time. A prime example would be highly pathogenic avian influenza infections where mortality may occur before an immune response is even detectable.

1.5. Purpose 5

For estimates of prevalence of infection or exposure to facilitate risk analysis, e.g. for health surveys, herd health status and to monitor disease control measures, antibody tests of moderate DSe & DSp are the tests of choice. In general, this would balance both FN & FP rates and result in a more accurate estimate of the true prevalence of infection 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 & FP rates (see Chapter 2.2.5 *Statistical approaches to validation*).

1.6. Purpose 6

For the determination of the immune status in individual animals or populations, e.g. post-vaccination, antibody tests of high DSp are required. Such tests have very low FP rates and as such provide a high degree of confidence in the PPV of the result. For use in individual animals, the use of virus neutralisation (VN) tests in cell culture for the detection of vaccine-induced neutralising antibodies against rabies virus in dogs and cats would be a prime example of a test with high DSp used for expression of titres in international units. However, these tests are technically sophisticated, expensive to maintain and run, and require strict biosafety procedures. For larger volume applications, such as monitoring regional vaccination programmes, ELISA-based tests would be more applicable, given their simplicity, cost effectiveness and high throughput. The same DSp considerations should be applied to these types of tests.

The experience of laboratory diagnosticians is not only essential in the choice of an appropriate test that will achieve the desired purpose, but is also required to determine reliably the scientific limitations of an assay and practical considerations such as cost, equipment and reagent availability, throughput capacity of the laboratory and test turn-around-times.

2. Assay development – experimentation

2.1. Reference materials, reagents and controls

Designing the method

- Has the design been shaped by the intended purpose of the assay?
- What is the specific application?
- What are the types and statistically relevant numbers of samples to be tested? (See Chapter 2.2.5)
- Will the test be field or lab based?

2.1.1. Test samples

Samples to be tested in antibody assays should be handled as described in Chapter 1.1.2 *Collection, submission and storage of diagnostic specimens*. The sample matrix in which antibodies are usually detected is serum, but may also include plasma, whole blood, milk, meat juice, egg yoke, lacrimal secretions and saliva.

2.1.2. Reference Standards

Antisera directed against the reference strain of a pathogen are known as reference sera or reference standards (Wright *et al.,* 1993; WOAH Validation Standard; Section 1.4 of Chapter 2.2.6 *Selection and use of reference samples and panels*). Such sera containing antibody of known concentration/activity are useful in the initial development of an assay. For a number of WOAH listed diseases, (e.g. avian influenza, foot and mouth disease, classical swine fever, etc.) international reference standards are available through WOAH Reference Laboratories and Collaborating Centres. When not available from other sources, it may be necessary to produce in-house reference standards against which working standards (process or quality controls) are calibrated.

2.1.3. Positive and negative reference panel

These sera, containing concentrations of antibody over the intended operating range, (also known as dynamic range) of the assay, should be used throughout the development and standardisation of an antibody assay. It is recommended that they be prepared in sufficient quantities so that they may be used in various aspects of validation. These samples should represent known infected and uninfected animals from the population that eventually will become the target of the validated assay. They should preferably be derived from individual animals, but they may represent pools of samples from several animals (Chapter 2.2.6).

2.1.4. Monoclonal antibody reagents

The advent of monoclonal antibodies has greatly enhanced enzyme immunoassays. Whereas polyclonal anti-immunoglobulin conjugates are used in most indirect ELISAs, monoclonal antibody conjugates can be directed to specific immunoglobulin isotypes. Depending on the immunoglobulin epitope targeted, many of these monoclonals can be effectively used to detect

antibodies in related species, e.g. ruminants. Using monoclonal conjugates to either light or heavy chain epitopes can effectively modulate the DSp and DSe of the indirect ELISA.

Monoclonal antibodies are best known for their application in competitive or blocking ELISAs. In this case, the monoclonal specificity is directed to epitopes on the pathogen in question. Depending on the epitope targeted, the analytical specificity of the assay can be modulated.

Critical points to be addressed:

- Have you considered that concentrations of analyte in matrix significantly impact the lower limit of antibody detection and the operating range of the assay?
- Are the required antibody reagents (mono/polyclonal) available?
- Is available antigen sufficiently purified?
- Are reagents commercially available? If not, is it practical to produce them in-house?
- Are reference standard reagents available? If not, how are you going to resolve this deficiency? (See WOAH Validation Standard, Section A.2.6.)

Monoclonal antibodies can also be used in sandwich ELISAs, either for trapping antigen to the plate or for subsequently detecting antigens that have been trapped. Depending on the size and complexity of the antigen in question, it is sometimes preferable to use a polyclonal antibody preparation for trapping as they generally contain antibodies of high binding affinity.

2.1.5. Antigens

Aspects affecting choice of test

- Is the assay to be used for screening or confirmatory purposes, or both?
- Will it be used for one or more species? Which ones?
- Is the test intended for detection of early or late infection?
- Will the test be used to measure serotype- or subtype-specific antibodies?
- Will the assay be used to confirm sero-conversion after vaccination?
- Will it be a DIVA assay (differentiation of infected from vaccinated animals)?
- Will the test be applied to trade?

Antigens used in ELISAs are of critical importance to diagnostic performance given a particular application. Antigens expressing highly conserved epitopes, such as those found in some viral matrix or nucleoproteins, are generally useful in group-specific assays, such as ELISAs for the detection of responses to all Influenza A viruses. Other antigen epitopes can be used to restrict detection to certain serotypes. The choice of antigen must be carefully researched and considered.

Crude antigen preparations like cell lysates have had widespread use in the past, and are still deployed for some assays. However, antigens improved greatly as purification techniques advanced, e.g. affinity chromatography. Further improvements were achieved through the application of molecular cloning. Recombinant antigen technologies have greatly enhanced all aspects of ELISA performance, from analytical through diagnostic characteristics.

2.2. Design of test method

In designing a test, its intended application will influence the choice of assay format that is best suited for the task. For example, if its use is primarily for surveillance, then the type of ELISA needs to be conducive to achieving high DSe, as described in the 'Purposes' above. If, however, the screening assay's DSe is set so high that it generates many false positives, then a companion confirmatory test should also be considered at the same time. Many ELISA formats are available, each with their advantages and disadvantages that allow customisation of assays for very specific purposes (Table 2).

Practical matters in selecting an assay format

- Is high-throughput essential? Will it be automated?
- 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 the test?
- Will that assay be feasible for use in my laboratory?
- Will it be easily transferrable to other laboratories?

Important factors that influence the choice of an antibody assay format are availability of reagents and likely continuity of supply, not only for the design and optimisation stage but for operational scale application of the test. A limitation may be the unavailability of relevant antibody reagents for a particular format, e.g. competitive or blocking formats generally require antigen-specific monoclonal antibodies. Another example would be the need for an effective capturing antigen: a rather crude antigen may be acceptable for use in a sandwich-based ELISA screening assay, whereas a purified antigen would be necessary for a confirmatory assay. Other important considerations for choosing a particular ELISA format are which antibody isotypes, concentrations, avidities and antigenic specificities are diagnostically relevant; which antigen, and in particular which epitopes are relevant; and what is the desired operating range of the assay. All will play a large role in selecting a particular type of ELISA (Table 2). If it is anticipated that the test will be used in different species, including wildlife, a competitive/blocking format may be useful. Deciding on an assay format also requires that application of the assay be considered. Questions that should be addressed are detailed in the box above on "*Practical matters in selecting an assay format*" and practical questions in the boxes below Table 2. It is essential to deal with such questions at this point in assay development as they are essential to a positive outcome and application.

Table 2. ELISA formats: advantages and disadvantages*

Type of ELISA	*Advantage*	*Disadvantage*
Indirect – bound Ab detected by anti-species conjugate or by Protein A/G conjugates	Use and availability of high variety of antispecies-specific conjugates often targeting particular antibody subsets, such as anti IgM, IgG1, IgG2, etc. Protein A and Protein G conjugates have a wide species specificity and may give lower background signals than anti-Ig reagents. Wide use for screening large numbers of samples	Variation in degree of nonspecific binding in individual sera To compensate for this problem high starting dilutions are required This can lead to a decrease in DSe in comparison to competitive/blocking formats Can only be used for one or a few species at a time
Sandwich – Ag presented on a solid-bound-phase capture antibody	The capture antibody on the solid phase can help to orient the antigenic molecule, which improves the chance that the sample antibody will bind. Unpurified antigen preparations can be used because capture antibody selectively binds crude antigen. Pre-coating with capture antibody can reduce the potential for subsequent binding of nonspecific proteins during the test.	Antigens must have at least two antigenic sites or epitopes which limits this type to relatively large antigenic complexes or more complex proteins Size and spatial relationship of epitopes can affect the assay
Competition (indirect and sandwich types) – Test antibody in sample mixed with pre-titrated detection antibody, then added to wells coated with capture antigen, either in direct or inhibition/blocking format.	Easy adaptation for use as antibody detection tests When highly specific MAbs are used the antigen does not have to be highly purified Can be used in different species for which no conjugated antibodies exist Advantage of competitive/blocking sandwich type relies on antigen capture Sera can be tested in low dilutions without risk of interference due to non-specific antibodies binding. This may contribute to a higher sensitivity of this format Different antibody concentrations can be used to favour either analytical sensitivity or specificity. This is particularly relevant for assays using polyclonal antibodies which are much more affected through the use of different dilutions of sera	Generally more steps and more optimisation may be needed, e.g. pre-titration and optimisation for liquid and solid phase reagents. Higher level of technical sophistication required

*Primary source is Crowther (2001).

2.3. Proof of concept experiments (feasibility studies)

After choice of an ELISA format, initial experiments are designed to determine if the proposed assay is viable. A reference panel such as described in Section A.2.1.3 should be tested in the prototype assay. If a reference standard is to be used for normalisation of test data, it should be selected and incorporated at this point in assay development. To provide continuity in data assessment throughout, both the reference panel and any reference standards should be included in all remaining aspects of the validation studies. The reference panel used in the feasibility study should span the entire anticipated operating range of the candidate assay and be run in replicates as a quick check for repeatability.

Proof of concept

- Was the feasibility study conducted with at least 4 to 5 samples spanning the operating range of the assay?
- Did you include one or more reference standards if required for data normalisation?
- Was separation of results between negative, low positive and high positive samples adequate?

The assay should achieve good separation in OD values, spanning the operating range of antibody activity. Adequate separation is particularly important between the negative and low positive samples. The lower OD range should be 0.1 or less for the negative control in indirect ELISAs, or for the strong positive control in competitive/blocking ELISAs. OD values at the upper end of the operating range should not exceed 2.0, as above this value plate readers become rather inaccurate. If the assay appears promising, optimisation is the next step.

2.4. Samples and data expression

2.4.1. Preparation and storage serum panels for optimisation studies

A best practice for antibody assays to select several (a minimum of four to five) serum samples that range from negative to high levels of antibodies against the infectious agent in question. These samples are initially used in experiments designed to demonstrate proof of concept. A large volume (e.g. a minimum of 10 ml) of each serum sample is acquired and divided it into 0.1 ml aliquots for storage at or below –20°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 serum with the same number of freeze–thaw cycles for all experiments (repeated freezing and thawing of serum can denature antibodies so should be avoided). Also, variation is reduced when the experimenter uses the same source of serum for all experiments rather than switching among various sera 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 may be suitable as a reference standard for data expression and the entire panel may be used for repeatability assessments both within and between runs of the assay (Jacobson, 1998). They may also serve as in-house working standards, i.e. quality or process controls given that their reactivity has been well characterised; such controls provide assurance that runs of the assay are producing accurate data (Wright *et al.,* 1993).

2.4.2. Normalisation of results and their expression

An optical density (OD) reading in ELISA is a measurement of colour development that is a function the amount of antibody present in a sample. Because colour development is a function of a reaction of enzyme and substrate in the presence of a chromogen, results from day to day are subject to variation attributable to external factors such as temperature, reaction time, etc. Comparison of OD results for the same samples between runs of an assay in the same laboratory, or between laboratories, lacks precision because of variation in results of reference standards included in each run of the assay. Therefore, OD results of test samples need to be adjusted as a function of the OD(s) of one or more reference standards in a specific assay run. This process is known as “normalisation” of ELISA results (see the WOAH Validation Standard, Section A.2.7 for

details). The method of normalisation and expression of data should be determined, preferably no later than at the end of the feasibility studies.

OD values may be expressed in several ways (Wright *et al.,* 1993). A simple method is to express all OD values as a percentage of a single high-positive serum control that is included on each plate. For such calculations, this control must yield results that are in the linear segment of the operating range of the assay. A more rigorous normalisation procedure is to calculate results from a standard curve generated by plotting observed OD values against concentration (or dilution) of antibody for several serum controls that span the range of antibody activity of the assay. It requires a more sophisticated algorithm, such as linear regression, log-logit, or 4 or 5 parameter logistic regression analysis, among others. This approach is more precise because it does not rely on only one high-positive control sample for data normalisation, but rather uses several serum controls, adjusted to expected values, to plot a standard curve from which the sample value is extrapolated. This method also allows for exclusion of a control value that may fall outside expected confidence limits.

2.5. Optimisation

For ELISAs, the most important variables that need to be optimised are concentration/dilution of antigen adsorbed to the solid phase, test serum working dilution, enzyme molarity, antibody-conjugate dilution, and substrate solution concentration. These are evaluated through checkerboard assessments (each variable compared against all other variables within one run of an assay that is repeated several times). Other variables that need consideration are pH and ionicity of reagents, molecular factors such as valency and epitope density of antigens, isotype of targeted antibody and antibody affinity. Precision of test results can be graphically depicted or expressed numerically by various statistical methods (Crowther, 2001). ELISA studies require that instrumentation (plate washers and readers, etc.) must be properly calibrated prior to use – part of the laboratory's quality control programme.

Optimisation and standardisation

- Have all critical reagents been tested against each other in checker board titrations?
- Did you find optimal concentration/dilutions for each reagent?
- Did you incorporate quality or process control procedures and reagents?
- Did you incorporate methods for normalisation of test data?

2.6. Inhibitory factors in sample matrix

Although ELISA antibody detection systems are rather resistant to inhibitory factors, the WOAH Validation Standard, Section A.2.4, and Greiner *et al.* (1997) provide 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 standard sera

If international, national, or other-source reference sera are available, the assay should be calibrated to match the analytical sensitivity in terms of the metrological units ascribed to the calibration sera (Wright, 1998).

B. ASSAY VALIDATION PATHWAY

1. Stage 1 – Analytical performance characteristics

1.1. Repeatability

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?

Repeatability is the level of agreement between results of replicates of a sample, both within and between runs of the same method in one laboratory. The same or similar panel of samples used in the feasibility study is adequate. No less than three (preferably 5) samples covering the operating range of the assay, and of sufficient quantity for at least 20 runs of the assay over several days. Specifics of how the samples should be prepared and handled are provided in Chapter 2.2.6 and in the WOAH Validation Standard, Section B.1.1. It is valuable to include at least one reference sample in an indirect ELISA (a positive serum control) to which the test samples can be normalised by per cent of the positive control. The within run variation can be determined by the mean OD and coefficient of variation (CV) of the replicates of each sample. The CV should not exceed about 15% (with the possible exception of negative and very low positive samples which may have higher (and meaningless) CVs). If all of the samples have previously been calibrated to reference standards, and their expected ODs are thus known, the observed ODs for each sample in each run can normalised as a function of their expected ODs in linear regression analysis. This provides a correlation coefficient as evidence of closeness of fit to the expected value, and allows for normalised values to be plotted in control charts (Crowther, 2001).

1.2. Analytical specificity

Analytical specificity (ASp) is determined by testing sera from animals that are known to have been infected/exposed to all species/strains that the test should detect (Chapter 2.2.6, Section B.1). Cross reactivity with sera from animals infected with related species is used to evaluate the ASp. ELISAs are also subject to false positive results attributable to exogenous factors, such as nonspecific binding of serum or conjugate to the plastic surface that may require use of blocking agents. Care must be taken to eliminate this source of error. Blocking and competitive ELISAs may also suffer specificity problems due to stearic hindrance preventing proteins binding to their target sites.

1.3. Analytical sensitivity

Analytical sensitivity (ASe) is synonymous with the lower limit of detection (LOD) of antibody concentration in a sample. The different types of antibody assays vary considerably in their inherent limit in antibody detection. For instance, LODs for eight different types of antibody assays range from 1000 ng/ml (radial immunodiffusion) to 0.01 ng/ml (chemiluminescence) (Nielsen *et al.,* 1996). 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.

1.4. Standard test method comparison with the candidate test method

The candidate test method should be run in parallel with an WOAH or other accepted reference test method, using the same panel of samples on both, to determine whether the candidate method exhibits the same quantitative and qualitative characteristics as the standard method. Favourable comparability lends strength to the belief that candidate method will be a successful substitute for the reference method (see also methods comparison studies, Chapter 2.2.5).

2. Stage 2 – Diagnostic performance characteristics

DSe and DSp are the primary performance indicators of the validation process. Antibody assays are subject to the same general procedures to achieve estimates of DSe and DSp as required of all other assay types (see the WOAH Validation Standard, Section B.2 for essential details). The number of samples needed to establish these estimates for a particular antibody assay require a sampling design that considers many variables. This includes creation of a sample panel that is tailored particularly for the intended purpose of the assay (e.g. a screening versus confirmatory test). It also requires predetermined desired levels of DSe and DSp (indicating acceptable levels of false negative and false positive results), allowable error in the estimates of such DSe and DSp, and the confidence level required for these estimates.

The number of animals required to establish acceptable DSe and DSp estimates is a function of the level of confidence desired in DSe and DSp estimates and the accepted allowable error. For instance, for a pathogenic disease like FMD, it is necessary to reduce the likelihood that infected animals will be misclassified as uninfected, which reduces allowable error in the test result which, in turn, increases the number of samples needed to establish a high level of confidence in the DSe estimates. Alternatively, for a confirmatory assay it is desirable to reduce the likelihood that uninfected animals will be classified as infected. A high DSp is then desired with minimal allowable error, requiring a larger sample size of uninfected animals. All of these general issues related to sample size, confidence intervals and allowable error in the DSe and DSp estimates are described in the WOAH Validation Standard, Section B.2, with additional detail and tables of sample numbers required available elsewhere (Jacobson, 1998).

It is often challenging to obtain a sufficient number of well characterised sera to achieve estimates of DSe and DSp that are sufficient for the intended purpose of the assay. Initially, it may be a compromise between what is statistically meaningful and practically feasible, resulting in an assay that is provisionally recognised (WOAH Validation Standard, Section B.2.6). However, over time, with accumulation of more well characterised samples, the estimates of DSe and DSp may be strengthened (see Section 5.4 below).

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

Antibody assays undergoing validation pose unique problems when attempting to assemble known positive and known negative samples in sufficient quantity to establish assay performance characteristics. Antibody is an indirect indicator of the presence of, or prior exposure to, an infectious agent or its components. Inferences from detection of antibody (or the lack thereof) depend on the host's qualitative and quantitative responses to the organism. Factors that affect the concentration and composition of specific antibody in serum samples are inherent to the host (e.g. age, sex, breed, nutritional status, pregnancy, immunological responsiveness) or acquired (e.g. passively acquired antibody, or active immunity elicited by vaccination or infection). Theoretically, samples from animals that represent all of these variables should be included in the panels used for establishing DSe and DSp estimates. Clearly, this becomes a daunting, if not impossible task. To surmount this problem, the initial sample panels should be representative of the majority of animals in the target population to achieve initial estimates of DSe and DSp. In reality, it is necessary to enhance DSe and DSp estimates after the assay has been implemented as more well characterised samples become available (see Section 5.4, below).

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?

Because it is often desirable to stretch the application of antibody detection assays to a huge number of animals spanning large geographical areas (e.g. as in screening assays for an entire continent), assembly of fully representative sample panels for such a large diagnostic window of variables may be nearly impossible. A useful alternative is to first establish DSe and DSp estimates for a rather homogeneous population of animals. If the assay is destined for use in disparate populations of animals, which may harbour a different infectious agent profile (with possibility of cross reactions not seen in the original targeted population), a reassessment of DSe and DSp may be necessary, drawing from data acquired using new sample panels that are representative of the population(s) targeted.

2.2. Reference animal populations

2.2.1. Animals of "known infection status"

Reference animals of known infected and known uninfected status are the ideal source of samples for determining DSe and DSp. However, such samples are rare and difficult to establish. The most familiar term for reference animals or samples used in establishing DSe and DSp is the so called "gold standard", a misnomer commonly used to classify almost any reference animal as infected/exposed or uninfected, with samples from such animals classified as positive or negative (see the WOAH Validation Standard, Section B.2.1–2.3).

Assay developers should be aware of the advantages, and particularly the pitfalls, associated with various methods that are used to classify reference animals as infected or uninfected. The samples from such animals are deemed either positive or negative, and collectively become the *reference standard* upon which the candidate assay's DSe and DSp are based. It is, therefore, crucial to carefully consider the validity of various reference standards as exemplified in the following four examples:

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

If an infectious agent or definitive histopathological criterion is detected in an animal, this generally constitutes an unequivocal reference standard for that animal. Serum samples derived from such animals usually are considered to be unequivocal serum reference standards for determining DSe and DSp of the candidate assay. However, such samples may have their limitations. At the population level, a pathogen may be unequivocally present in some animals, but if the serum sample was taken from the animal early in the infection process, the immune response may not yet have produced detectable antibody. In this case, such serum samples used as reference standards would have been FN for the subset of animals in an early stage of infection. In contrast, for more chronic types of infection, using only reference animals that have confirmatory culture or histopathology may produce higher estimates of DSe than are realistic for the population targeted by the assay because the immune response will always be well established.

ii) A composite reference standard: verification of uninfected or unexposed animals

This standard is achieved by selecting reference animals from geographical areas where herd histories, clinical profiles, prior testing results and other parameters provide evidence suggesting the absence of the pathogen, and thus no specific host antibody response to the pathogen targeted by the candidate assay. These types of reference materials, their strengths and limitations are described elsewhere (Jacobson, 1998), and must be considered carefully when using samples from such sources for establishing DSe and DSp for a candidate assay.

iii) A relative reference standard: comparative serology

This standard is characterised by reference animals that have been classified for their infection status by comparison with the test results of another serological assay on the same samples. It often is the only practical source of reference material available for evaluation of a new serological test. If results of such a reference test are chosen as the standard for determining diagnostic performance characteristics of the candidate assay, the resultant estimates of DSe and DSp are useful only insofar as the reference test has documentable, 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.) In some cases, the only way to obtain positive samples is by experimental

infection. This approach is highly suitable to model the dynamics of the infection and to determine the 'diagnostic window' with the new assay. For example, it is possible to get estimates of the time interval between exposure to a pathogen and when antibody is first detectable, or when 25, 50, 75 and 100% of the infected animals return a positive test result. Nevertheless there are pitfalls in use of time-series data that must be avoided. Data representing repeated observations from the same animals cannot be used in calculation of DSe and DSe because the statistical models used to establish DSe and DSp require independent observations (only one sample from each anima). For statistically legitimate time-course studies, or when single samples are used from each of many experimental animals, the strain of cultured organism, route and dose of exposure, infection with other related, cross-reactive and non-related, non-cross-reactive organisms are variables which may produce quantitatively and qualitatively atypical responses which are not found in natural infections in the target population. Experimental conditions typically lead to an overestimation of sensitivity and specificity for example by artificially high challenge doses and by using specific pathogen free animals as negative controls.

The time point of sample collection (days post-infection) must be indicated. Sources and history of experimental animals should be described. The validation should not be based solely on experimental animals as they do not represent natural populations of animals subject to pathogens by natural exposure.

2.2.2. Latent-class models for estimation of DSe and DSp

For a discussion of this approach for estimation of diagnostic performance, see the WOAH Validation Standard, Section B.2.5 and Chapter 2.2.5.

2.3. Threshold (cut-off) determination

The procedures for establishing the cut-off between negative and positive results of antibody assays are as described in the WOAH Validation Standard, Section B.2.4.

3. Stage 3 – Reproducibility and augmented repeatability estimates

Reproducibility is the measure of precision of an assay when used in several laboratories located in distinct regions or countries using the identical assay (protocol, reagents and controls) to test the same panel of samples. Reproducibility assessments for antibody 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.

4. Stage 4 – Programme implementation

4.1. Interpretation of results and determination of predictive values

Best practices for programme implementation are general to all assay types (WOAH Validation Standard, Section B.4). However, as ELISA is often the assay of choice for surveillance programs to affirm absence of disease, or for eradication of disease or elimination of infection from defined populations, the issue of false positive results can be a significant problem even if the diagnostic specificity is very high.

A common misperception is that a test with 99% DSp and DSe will only mis-classify animals as FP or FN 1% of the time. The FN and FP rates vary depending on the prevalence of infection in the targeted population. False positive reactions in a disease eradication campaign can vary significantly from the beginning of the campaign when prevalence is relatively high (for example, 10%) to near the end of the campaign when it has decreased to 0.1%. The predictive values of test results then become very important. Predictive values are probabilities that a test result is truly positive or truly negative. In our example using an assay with 99% DSe and DSp for testing a population of animals with a 10% prevalence of disease, the predictive value of a positive test result (PPV) is 91.7%, meaning that there is a 91.7% probability that the animal is truly infected. The predictive value of a negative test result (NPV) is 99.9. When the prevalence drops to 5%, the PPV and NPV are 83.9% and 99.9%, respectively. However, if the prevalence drops further to 0.1%, by successfully removing infected animals from the population, the same test will produce a PPV of 9% and a NPV of 99.9%, meaning that there is only a 9% chance that a

positive test result is detecting a truly infected animal (of 1000 animals tested, only about 1 in 10 positive test results is indicative of an infected animal – the other 9 are false positive). So, if the test is intended for the purpose of eradication of a disease or elimination of infection from a population, the test developer is advised to consider moving the assay to a second cut-off that yields a higher DSp late in the campaign to reduce the probability of false positive reactions. It is instructive to examine a predictive value chart for assays of varying DSe and DSp, to visualise the effects of reduced prevalence on predictive values of an assay (WOAH Validation Standard, Table 2, and Jacobson, 1998).

5. Monitoring assay performance

5.1. Monitoring the assay

Once the assay is in routine use, internal quality control is accomplished by consistently monitoring the assay using quality control charts for assessment of repeatability and accuracy. Charts representing at least 30 runs will reveal trends or shifts in values of controls and standards. Lines representing the mean value of a control sample in at least 30 runs, plus/minus 3 standard deviations, are useful decision criteria for inclusion or exclusion of a run of the assay. The run is rejected if one control/standard exceeds ± 3 standard deviations (STD) or if 2 controls (or more) exceed ± 2 STD (Crowther, 2001). Decision criteria may need to be customised for a given assay because of inherent differences between assays attributable to the host pathogen system. Chapter 2.2.4 provides an example of how to apply measurement uncertainty for an antibody ELISA using a positive internal control sample.

Reproducibility of test results between laboratories should be assessed by External Quality Assurance at least once per year and is an essential requirement of ISO 17025 accredited laboratories. Membership in a consortium of laboratories that are interested in evaluating their output is valuable.

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

When quality or process control samples are nearing depletion, it is essential to prepare and repeatedly test the replacement samples. The replacement samples should be included in at least 10 routine runs of the assay, with their results normalised against the existing reference standard. The activity of the replacement control should be comparable to the replaced control. If the reference standard requires replacement, care must be taken to select a replacement that matches all of the original serum characteristics as closely as possible, thus allowing use of the replacement to normalise test results with comparable outcomes (see also Chapter 2.2.8 *Comparability of assays after changes in a validated test method*).

When other reagents such as antigen for capture of antibody, must be replaced they should be produced or procured using the same protocols or criteria as used for the original reagents. They need to be assessed using sera from routine submissions in 5–10 parallel runs that include the current and the new reagent(s). A panel of representative samples, such as a proficiency panel, is also a useful tool for assessing the comparability of the reagents (Chapter 2.2.6).

Monitoring assay performance

- Has the purpose of the assay changed?
- Has the epidemiology of the disease in question changed, e.g. prevalence, new serotypes or strains, etc.?
- Have critical reagents been changed, and if so, was comparability of the new reagents assessed?
- Are performance indicators included in day to day use of the assay (control charts, basic statistics)?
- Are upper and lower limits in control charts updated periodically as more experience with the control samples is achieved?
- Are test panels shared with other laboratories to assess reproducibility?
- Is proficiency testing included as part of continuing evaluation of the assay?

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

If the assay is to be changed from, say, a sandwich ELISA to a competitive/blocking format, the assay will require revalidation because of the many variables that may affect the performance characteristics of the assay. For an assay considered for implementation in another geographic region, e.g. from the northern to the southern hemisphere, it is essential to revalidate the assay by subjecting it to sera from populations of animals that reside under local conditions. Evaluation of reference sera that represent those populations is done by using stages 3–5 in Figure 1 in the WOAH Validation Standard. It is the only way to assure that the assay is valid for populations that are of different composition compared with the original population targeted by the assay.

5.4. Enhancing confidence in validation criteria

Due to the extensive set of variables that have an impact on the performance of serodiagnostic assays, it is useful to expand the number of reference sera wherever possible, recognising the principle that error is reduced with increasing sample size. An expanded reference serum bank should be accumulated with well characterised sera, and used periodically to update estimates for DSe and DSp for the population targeted by the assay.

REFERENCES

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

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

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.

NIELSEN K., GALL D., KELLY W., VIGLIOCCO A., HENNING D. & GARCIA M. (1996). Immunoassay Development: Application to Enzyme Immunoassay for the Diagnosis of Brucellosis, Copyright, Agriculture and Agri-Food Canada.

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.

WRIGHT P.F., NILSSON E., VAN ROOIJ E.M., LELENTA M. & JEGGO M.H. (1993). Standardization and validation of enzyme linked immunosorbent assay techniques for the detection of antibody in infectious disease diagnosis. *Rev. sci. tech. Off. int. Epiz*., **12**, 435–450.

*
* *

NB: FIRST ADOPTED IN 2014.