

## Introduction

### Validation of tests for OIE-listed diseases as fit-for-purpose in a world of evolving diagnostic technologies

I.A. Gardner\* <sup>(1, 2)</sup>, A. Colling <sup>(2, 3)</sup>, C.G. Caraguel <sup>(2, 4)</sup>, J.R. Crowther <sup>(5)</sup>, G. Jones <sup>(2, 6)</sup>, S.M. Firestone <sup>(2, 7)</sup> & C. Heuer <sup>(2, 8)</sup>

(1) Atlantic Veterinary College, University of Prince Edward Island, 550 University Avenue, Charlottetown, Prince Edward Island, C1A 4P3, Canada

(2) OIE Collaborating Centre for Diagnostic Test Validation Science in the Asia-Pacific Region, CSIRO Newcomb, CSIRO Australian Centre for Disease Preparedness, 5 Portarlington Road, Newcomb, Victoria 3219, Australia

(3) Australian Centre for Disease Preparedness, Commonwealth Scientific and Industrial Research Organisation, 5 Portarlington Road, East Geelong, Victoria 3219, Australia

(4) School of Animal and Veterinary Sciences, The University of Adelaide, Roseworthy Campus, Roseworthy, South Australia 5371, Australia

(5) Joint United Nations Food and Agriculture Organization/International Atomic Energy Agency Division of Nuclear Techniques in Food and Agriculture, Department of Nuclear Sciences and Applications, Vienna International Centre, PO Box 100, A1400 Vienna, Austria

(6) School of Fundamental Sciences, Massey University, Palmerston North 4410, New Zealand

(7) Melbourne Veterinary School, Faculty of Veterinary and Agricultural Sciences, The University of Melbourne, Victoria 3010, Australia

(8) School of Veterinary Science, Massey University, Tennent Drive, Palmerston North 4410, New Zealand

\*Corresponding author: [iagardner@upei.ca](mailto:iagardner@upei.ca)

## Summary

The World Organisation for Animal Health (OIE) has made leading contributions to the discipline of test validation science by providing standards and guidelines that inform the test validation

process in terrestrial and aquatic animals. The OIE *Manual of Diagnostic Tests and Vaccines for Terrestrial Animals*, and the *Manual of Diagnostic Tests for Aquatic Animals* describe the test validation pathway in the context of fitness for purpose, elaborate on the importance of diagnostic sensitivity (DSe) and specificity (DSp) as measures of test accuracy, and designate additional factors (e.g. test cost, laboratory throughput capacity and rapidity of test results) that influence choices of a single test over others or the inclusion of a new test in a diagnostic process that includes multiple tests. This paper provides examples of each of the six main testing purposes listed in the *Terrestrial Manual* and describes additional metrics such as ruggedness and robustness that should be included in the validation of point-of-care tests. Challenges associated with new diagnostic technologies and platforms are described. Validated tests with estimates of DSe and DSp are needed to measure confidence in test results for OIE-listed diseases, to facilitate risk assessments related to animal movement, to estimate true prevalence, and for certification of disease freedom and use in epidemiological (risk factor) studies.

### Keywords

Animal diseases – Diagnostic accuracy – Fitness for purpose – Likelihood ratios – Repeatability – Reproducibility – Sensitivity – Specificity – Test validation.

### Introduction and historical perspectives

Over the last 30 years, the World Organisation for Animal Health (OIE) through its reference laboratories and commissions has provided leadership in advancing the discipline of test validation science, including the setting of international validation standards. The overarching standard is the chapter entitled ‘Principles and methods of validation of diagnostic assays for infectious diseases’ in both the *Manual of Diagnostic Tests and Vaccines for Terrestrial Animals (Terrestrial Manual)* and the *Manual of Diagnostic Tests for Aquatic Animals (Aquatic Manual)* (1, 2) which describes a four-stage pathway for test validation with measures of analytical (see Bowden *et al.*, this issue [3]) and diagnostic accuracy (see Cullinane and Garvey [4], and Gifford *et al.* [5], this issue) and test reproducibility (see Waugh and Clark, this issue [6]). The principles described in the chapter apply to tests for infectious diseases in animals, regardless of whether they are listed by the OIE or not. Modifications to address the many validation challenges for infectious disease in wildlife are described in Chapter 2.2.7. of the *Terrestrial Manual* (7) and elsewhere (8) and (see Michel *et al.*, this issue [9]).

During the same period, there have been improvements in diagnostic technologies, including the use of multiplex assays and next generation sequencing of important pathogens (10, and see Halpin *et al.*, this issue [11]); increased development of pen-side (point-of-care) tests for rapid diagnosis

of important animal diseases (see Halpin *et al.*, this issue [11], and [12]) and adoption of latent class models for estimation of the diagnostic sensitivity (DSe) and specificity (DSp) when the reference test or standard is imperfect (13, and see Cheung *et al.*, , this issue [14]). As an example of new initiatives, the OIE established a registry of diagnostic test kits that have been evaluated analytically and diagnostically and certified as ‘valid’ by the OIE for specific purposes, and similar initiatives have been established in several countries (see the section of this paper entitled ‘Test approval and certification’, and Gifford *et al.*, this issue [5], on the Secretariat for the Registration of Diagnostic Kits [SRDK]). End-user and producer responsibilities for validation of test kits have been described in a prior publication (15). In addition, the OIE approved a Collaborating Centre for Diagnostic Test Validation Science in the Asia-Pacific region in 2017 as a joint initiative of the Australian Centre for Disease Preparedness (previously named the Australian Animal Health Laboratory) of the Commonwealth Scientific and Industrial Research Organisation (CSIRO), the Faculty of Veterinary and Agricultural Sciences of the University of Melbourne and the EpiCentre, Institute of Veterinary and Biomedical Sciences, Massey University, New Zealand. The Centre’s mission is ‘to generate new knowledge and techniques that improve the use and interpretation of diagnostic tests used in human and animal health and to promote dissemination of that knowledge to the wider medical and veterinary communities’ (<https://fvas.unimelb.edu.au/research/centres/oie-dx>; see also Brown *et al.*, this issue [16]).

The purpose of this introductory paper is to provide an overview of progress in the validation of tests for OIE-listed diseases in terrestrial, and wild animals in the context of fitness for purpose (see Cullinane [4], and Michel *et al.* [9], this issue, for terrestrial, and wildlife, respectively). A clearly defined testing purpose coupled with knowledge of disease pathogenesis guide the selection of appropriate animals, target specimens and analytical targets for inclusion in the validation process and it should be emphasised that these selected targets have a profound effect on the accuracy of most diagnostic assays. This is also true for pre-laboratory steps such as sample collection, use of preservatives, transportation times and handling of specimens in the receiving laboratory, where cross-contamination of samples in the field and in the receiving laboratory may lead to false-positive results (1, 7). Laboratories performing tests for OIE-listed diseases should be operating under a quality management system (17, and see Newberry and Colling [18], this issue) and should ideally participate in proficiency testing or ring trials for transboundary diseases of major international concern (see Waugh and Clark [6], this issue and Johnson and Cabuang [19], this issue). Design, implementation and reporting of test validation studies are described in Gardner *et al.* (20) and in other chapters in this thematic issue (see Kostoulas *et al.*[21], and Heuer and Stevenson [22], this issue).

## Validation and fitness for purpose

The OIE considers a test ‘validated’ for a targeted pathogen or disease once its analytical and diagnostic accuracy and precision (repeatability and reproducibility) are deemed fit for its intended purpose(s) in a target species based on the specimens routinely used for testing, and diagnostic systems in which tests are used. The need for validation applies equally to tests used for individual animals or populations and to testing of pooled samples from multiple animals. No single test can fit all required purposes perfectly (some tests are better suited for some specific purposes than others). For example, tests used in animals are imperfect in either DSe, DSp or both parameters, but tests do not need to be perfect to be useful especially when they can be combined with one or more additional tests that target a different analyte (e.g. serum antibodies versus nucleic acid detection). Although DSe, DSp, repeatability, and reproducibility are the primary evaluation criteria, ancillary considerations such as test cost, laboratory throughput capacity, level of operator skill and equipment to reliably perform the test and time to provide results, are relevant to end users. In addition, how a candidate test fits in with existing tests as part of a diagnostic process should be considered when assessing whether a validated test is likely to be used or not. Even though a test might be validated as fit for purpose by a developing laboratory or a test kit manufacturer, verification studies (see Kirkland and Newberry [23], this issue) of the accuracy and precision of a validated assay, should be done in a new laboratory which begins to use the test.

Consequences of test errors (false-negative or false-positive results) can vary from minor to major depending on aspects of the testing context, including the epidemiological unit of interest (e.g. individual animal, farm, region or country), the prevalence of disease, and whether effective mitigations are available. For instance, a test frequently yielding false-positive results would be unfit for demonstrating freedom from infection in an animal population but may be acceptable when selecting individual animals for movement providing that the test’s DSe is close to perfect. For test results measured on a quantitative scale such as enzyme-linked immunosorbent assay (ELISA) and quantitative polymerase chain reaction (qPCR), the cut-off (threshold) value may be adjusted to decrease false-positive results (i.e. higher DSp) at the cost of lower DSe (24). Alternate solutions include use of a test of near perfect individual specificity (i.e. >99.9%) or use of two or more tests with in-series interpretation of results (i.e. all test results need to be positive to consider an individual positive, and negative otherwise).

## OIE testing purposes

The OIE lists six overall purposes for use of diagnostic tests in its *Terrestrial Manual* (Table I) and the OIE Register of Certified Tests (see Gifford *et al.* [5], this issue, on SRDK) allows for an open option for description of purpose, for example the checkbox option 7 ‘others’ could be used

for ‘control of infection in stallions and mares at the start of the breeding season – *Taylorella equigenitalis* IFAT test’. In addition, testing in wild mammal populations is often focused on disease management and conservation purposes (8). The use of a test for prognostic purposes such as ‘prediction’ of a health outcome (e.g. survival or death) are important in human health and companion animals but are not considered in the OIE context.

It is important to note that a test may be fit for more than one purpose if appropriate evaluation data are provided for each identified purpose. Even if a test only has a single intended purpose (e.g. confirmation of clinical suspect cases), it is crucial that animals representative of the source population are sampled correctly for testing of tissues and fluids in which the target analyte is likely to be present. As an example, for purposes of movement of live animals, a test that is highly sensitive to the detection of infected animals not showing observable clinical signs is most relevant as clinically affected animals have a considerably lower probability of being moved or traded. For many diseases, the DSe of the same diagnostic test will be lower in animals without clinical signs compared to those with signs consistent with the disease of interest. For example, a diagnostic accuracy study of a RT-qPCR that detects the gene encoding core protein VP7 for African horse sickness (AHS) virus was made on blood samples from horses presenting clinical signs consistent with AHS (i.e. pyrexia and one more typical signs of AHS) and had a DSe of 97.8% and a DSp of 99.9% (29). At the time of the study there were only a few subclinical AHS-infected horses (all with AHSv vaccination histories) that had been detected in South Africa. Hence, a subsequent quantitative risk assessment of live horses exported from South Africa based on the RT-qPCR evaluated by Guthrie *et al.* (29) used a reduced DSe which had a median of 91.4% (95% probability interval, 65.7 to 99.4%) to ensure that there was a realistic range of DSe values in subclinically infected horses when used for certifying freedom from AHSv infection in horses moved internationally (30). However, without validation of the assay in subclinically infected horses the assumed range might have been overly optimistic.

During the lifetime of a test, there may have been a change of purpose. For example, at the start of avian influenza (AI) detection the TaqMan A quantitative PCR assay was used to confirm AI virus infection in clinically suspect birds but over time the test has been also used for detection of infection in apparently healthy birds. The question is whether the TaqMan A remained fit for purpose, for example is the confidence in a positive finding (i.e. low chance of false-positive result) high enough when screening birds in flocks with low or extremely low prevalence? For this example, the answer was ‘yes’ because the assay did produce a very low frequency of cross reactions resulting practically in no false-positive test results and at the same time reacted broadly enough with different AI strains.

Examples and case studies are provided in the aforementioned papers in this thematic issue and in a paper on the application of serological and molecular methods to study epidemiological aspects of bluetongue and African horse sickness (see Mayo *et al.* [31], this issue). The case studies illustrate OIE testing purposes and emphasise diagnostic performance parameters such as DSe, DS<sub>p</sub> and positive and negative predictive values (PPV, NPV) during different pre-clinical, clinical and post-clinical/recovery phases of an outbreak, including the use of multiple tests. Examples include ways to deal with inconclusive and indeterminate results, including retesting of samples and resampling, as specified/allowed in laboratory testing protocols or standard operating procedures during routine testing of samples, or undertaking sensitivity analyses considering equivocal results both ways, as positive or negative as recently undertaken in an assessment of serological evidence for wobbly possum virus in Australia (32).

### Test accuracy parameters

DSe and DS<sub>p</sub> (or likelihood ratios [LRs] which range from 0 to infinity) and NPVs and PPVs, which are prevalence dependent, are parameters that help to determine whether the assay is fit for the intended purpose for a population with a 'known' prevalence of infection (Box 1). For tests measured on a continuous scale, the area under the receiver-operating curve (AUC) and likelihood ratios provide alternative metrics of test accuracy (33, and see Caraguel and Colling, this issue [34]).

The effects of prevalence on predictive values (NPV and PPV) are shown in the following example and in Figure 1. During an effective disease eradication campaign, a progressive decrease of disease prevalence can be expected. This will cause an increased risk of false-positive test results among all positive test results (i.e. lowering PPV) and should be expected unless the DS<sub>p</sub> is 100%. A test with a high DS<sub>p</sub> or using two or more test results interpreted in series (where only individuals are categorised as positive, if all test results are positive) or the use of other information (e.g. clinical history) that increases the prevalence is useful to increase the PPV. The NPV is the percentage of truly-negative results among all test-negative results and its value is not appreciably affected when the prevalence of infection is low (e.g. <10%). High DSe or the use of two or more tests' results interpreted in-parallel (where any positive-test result makes the individual positive) can help to increase the NPV. These principles which apply to individual animals can be extended to multiple epidemiological units (i.e. herds and clusters of herds in the same geographical area) (35).

At the individual animal level, the fitness of a particular test is driven by the consequence of a false-positive or false-negative result according to the context of intended purpose. For instance, an exporting country would be more concerned about the consequences of a false-positive result

when demonstrating freedom from disease in its animal population (Purpose 1a) and would favour a diagnostic test or algorithm with high DSp or LR+ (i.e. trying to maximise the PPV). On the other hand, importing countries want to minimise false-negative results (i.e. maximise DSe). At the population level, to demonstrate and maintain freedom status, a suitable test should facilitate widespread geographical coverage of at-risk populations and be able to deliver rapid results cost-effectively. It is important to note that there is no OIE-designated acceptable threshold for the accuracy of any of these tests (DSe, DSp, LR+, LR-, PPV and NPV) that would make one test preferable to another, and there is additional uncertainty attributable to the limited precision of DSe and DSp estimates from a single evaluation study (Table I, 20).

A frequent limitation during diagnostic test validation is a lack of reference samples from confirmed infected or non-infected animals especially for diseases that occur in a restricted geographical area, occur infrequently or where access to animals and specimens is regulated (e.g. endangered wild animals). Hence, estimates of DSe and DSp often have wide confidence intervals and therefore have limited robustness from a statistical perspective. The OIE defines ‘provisional recognition’ as an assay with satisfactory estimates for analytical sensitivity and specificity (ASe, ASp) and repeatability and preliminary estimates for DSe and DSp and reproducibility. For example, attaining statistically robust numbers of sera from Hendra virus (HeV)-infected horses is problematic because affected horses usually die or are euthanised before developing a serum antibody response. Consequently, an antibody ELISA was evaluated using sera from 19 infected and 477 non-infected horses (36). In addition, six panels of well-characterised sera were tested to estimate the ASe, ASp, repeatability and reproducibility in a network approach, and network laboratories provided additional baseline data from HeV-negative horses, which resulted in robust estimates for DSp in non-infected populations. Eventually the test was approved by the Subcommittee on Animal Health and Laboratory Standards (SCAHLS), the national regulator in Australia. The use of validation panels using well-characterised samples in network laboratories also provides useful information on test accuracy and precision (see Ludi *et al.* [37], and Watson *et al.* [38], this issue).

## **Evaluation and use of point-of-care tests and new technologies such as multiplex PCR, next generation sequencing (NGS), and biomarkers (microRNAs)**

### **Point-of-care tests**

Typically, point-of-care tests (POCTs), also known as animal or pen-side tests, are applied in the field under conditions of variable temperature, humidity and light using specimens from sick or dead animals by operators with a diverse range of experience, training and proficiency.

Consequently, in addition to general validation requirements, point-of-care-specific parameters such as robustness and ruggedness need to be defined and addressed (see Halpin *et al.*, this issue [11]). Point-of-care tests are likely to perform with a higher degree of variation attributable to varied competence and experience of operators and on samples of lower quality than might be submitted to diagnostic laboratories. For example, if a POCT has been evaluated with oropharyngeal swabs but if the operator can only take cloacal swabs the result for this specimen will not be reliable. The same issues apply to swabbing for COVID-19 testing because many operators have minimal training, multiple swab types are used in different countries and jurisdictions, and the swabbing area may differ (e.g. nasal versus oropharyngeal sites). Similarly, storage conditions and the lifetimes of kit reagents and platforms, including maintenance, may be subject to higher variability than in an accredited laboratory. Another key variable, not specific to POCTs, but important and commonly requiring consideration is the stage of clinical disease in the animal under investigation or when samples are taken post-mortem, the time since death. This is well illustrated by the validation of the anthrax immunochromographic test (ICT) for use in the field to guide diagnostic and control decisions whilst awaiting laboratory confirmation. The anthrax ICT has been demonstrated to have very high DSe and DSp when used on carcasses within 48 hours of death. DSe has been demonstrated to decrease from 93% on samples collected from animals that were <48h post-mortem to 56% in samples from animals >48 post-mortem (39) leading to recommendations on how the test is used in the field (40). In summary, irradiation, exposure to direct sunlight, humidity, temperature, dust, soiling, sampling procedures and other influential physical variables need to be assessed in a POCT validation study as they define the kit's fitness for purpose. If it is likely that a POCT will be used after its expiry date or on samples of compromised quality, it is important to know details about its robustness to changes in climatic and other parameters and the likely consequences of these changes.

During the validation of POCT, intra- and inter-operator variation needs to be assessed under a range of realistic conditions. Results will help to detect robustness (repeatability) against internal variation and ruggedness against external conditions such as climatic conditions and various levels of proficiency (reproducibility). To assess and monitor reliability, it is important to include internal quality controls to confirm the basic functioning of the device, for example a weak-positive sample (to confirm DSe of the device and avoid false-negative results), a negative control (to confirm that test reagents are not contaminated producing false-positive results), and in case of a nucleic acid detection assay an internal control to assess the presence of matrix inhibitors (12).

During the recent COVID-19 outbreak, the Johns Hopkins University created a webpage with information about evaluation of diagnostic POCTs. Most of the kits tested were combined IgG/IgM POCTs with limited validation (i.e. performance was compared to established molecular

tests). As the majority of serological tests do not start to react until days 5–7 post-infection, their results may be negative when molecular tests are positive because of the different targets for each assay. In contrast, an elevated number of apparently ‘false-positive’ serological test results can be expected (i.e. > 2 weeks after onset of clinical signs) because molecular tests may give negative results when serological are positive. In this situation, the serological test results should be viewed as evidence of exposure to COVID-19. Furthermore, description about use and purpose of POCTs indicates that potential users have a limited understanding of diagnostic principles and the epidemiology of COVID (41).

### **Other technologies**

Technologies and applications such as next generation sequencing (NGS), PCR in a cartridge, loop-mediated isothermal amplification (LAMP), gene knock-out technology (CRISPR), micro-RNAs etc. are poised to bridge the gap between the research and diagnostic laboratory. It is important to define their potential diagnostic value and purposes of modern diagnostic methods in veterinary diagnostic laboratories and explore their compatibility with the existing validation principles and methods of the OIE. Critical parameters for evaluation and quality control (quality metrics) need to be fit-for-purpose (10). As many of these modern technologies have higher accuracy and precision than existing methods the use of a reference standard for validation is not recommended. In these circumstances, Bayesian latent class models (BLCMs) are more appropriate because they do not rely on a perfect reference standard (12, and see Halpin *et al.* [11], and Cheung *et al.* [14], this issue).

### **Test approval and certification**

Regulatory bodies that deal with registration processes of diagnostic kits exist at the national, regional and global levels. For example, the OIE has a science-based validation, certification and registration process (SOP) through which 14 kits have been registered (see Gifford *et al.* [5], this issue). The use of validation templates facilitates guidance, transparency and objective assessment of data. Australia has a similar test approval process but uses test-specific templates for molecular and serological tests through which 17 tests have been approved. The Friedrich Loeffler Institute in Germany uses application forms for the licensing of batches of reagents and diagnostic kits (list of certified products). Kit producers suffer from a lack of mutual recognition between regulatory bodies considering the use of tests in different regions or countries and variation of exigency and costs. Regulatory requirements for human and veterinary diagnostic testing require further harmonisation, especially for zoonotic diseases (42).

When changes are made to an assay (e.g. replacement of depleted reagents with a new batch, technical and biological modifications and enhancements), a method comparison study may be required to provide evidence that test performance is not negatively affected (43). A test kit that has been developed and validated requires verification if used in a new laboratory to provide assurance that specifications are met (23).

## References

1. World Organisation for Animal Health (OIE) (2019). – Chapter 1.1.6. Principles and methods of validation for diagnostic assays for infectious diseases. *In* Manual of Diagnostic Tests and Vaccines for Terrestrial Animals, 8th Ed. OIE, Paris, France. Available at: [www.oie.int/fileadmin/Home/eng/Health\\_standards/tahm/1.01.06\\_VALIDATION.pdf](http://www.oie.int/fileadmin/Home/eng/Health_standards/tahm/1.01.06_VALIDATION.pdf) (accessed on 6 October 2020).
2. World Organisation for Animal Health (OIE) (2019). – Chapter 1.1.2. Principles and methods of validation for diagnostic assays for infectious diseases. *In* Manual of Diagnostic Tests for Aquatic Animals, 7th Ed. OIE, Paris, France. Available at: [www.oie.int/index.php?id=2439&L=0&htmfile=chapitre\\_validation\\_diagnostics\\_assays.htm](http://www.oie.int/index.php?id=2439&L=0&htmfile=chapitre_validation_diagnostics_assays.htm) (accessed on 6 October 2020).
3. Bowden T.R., Crowther J.R. & Wang J. (2021). – Review of critical factors affecting analytical characteristics of serological and molecular assays. *In* Diagnostic test validation science: a key element for effective detection and control of infectious animal diseases (A. Colling & I.A. Gardner, eds). *Rev. Sci. Tech. Off. Int. Epiz.*, **40** (1), XXX–YYY. doi:10.20506/rst.40.1.XXXX.
4. Cullinane A. & Garvey M. (2021). – A review of diagnostic tests recommended in the World Organisation for Animal Health *Manual of Diagnostic Tests and Vaccines for Terrestrial Animals*. *In* Diagnostic test validation science: a key element for effective detection and control of infectious animal diseases (A. Colling & I.A. Gardner, eds). *Rev. Sci. Tech. Off. Int. Epiz.*, **40** (1), XXX–YYY. doi:10.20506/rst.40.1.XXXX.
5. Gifford G., Szabo M., Hibbard R., Mateo D., Colling A., Gardner I. & Erlacher-Vindel E. (2021). – Validation, certification and registration of certified tests and regulatory control of veterinary diagnostic test kits. *In* Diagnostic test validation science: a key element for effective detection and control of infectious animal diseases (A. Colling & I.A. Gardner, eds). *Rev. Sci. Tech. Off. Int. Epiz.*, **40** (1), XXX–YYY. doi:10.20506/rst.40.1.XXXX.

6. Waugh C. & Clark G. (2021). – Factors affecting test reproducibility among laboratories. *In* Diagnostic test validation science: a key element for effective detection and control of infectious animal diseases (A. Colling & I.A. Gardner, eds). *Rev. Sci. Tech. Off. Int. Epiz.*, **40** (1), XXX–YYY. doi:10.20506/rst.40.1.XXXX.
7. World Organisation for Animal Health (OIE) (2019). – Chapter 2.2.7. Principles and methods for the validation of diagnostic tests for infectious disease applicable to wildlife. *In* Manual of Diagnostic Tests and Vaccines for Terrestrial Animals, 8th Ed. OIE, Paris, France. Available at: [www.oie.int/fileadmin/Home/eng/Health\\_standards/tahm/2.02.07\\_WILDLIFE.pdf](http://www.oie.int/fileadmin/Home/eng/Health_standards/tahm/2.02.07_WILDLIFE.pdf) (accessed on 6 October 2020).
8. Jia B., Colling A., Stallknecht D.E., Blehert D., Bingham J., Crossley B., Eagles D. & Gardner I.A. (2020). – Validation of laboratory tests for infectious diseases in wild mammals: review and recommendations. *J. Vet. Diagn. Invest.*, **32** (6), 776–792. doi:10.1177/1040638720920346.
9. Michel A.L., van Heerden H., Prasse D., Rutten V., Al Dahouk S. & Crossley B.M. (2021). – Pathogen detection and disease diagnosis in wildlife: challenges and opportunities. *In* Diagnostic test validation science: a key element for effective detection and control of infectious animal diseases (A. Colling & I.A. Gardner, eds). *Rev. Sci. Tech. Off. Int. Epiz.*, **40** (1), XXX–YYY. doi:10.20506/rst.40.1.XXXX.
10. Van Borm S., Wang J., Granberg F. & Colling A. (2016). – Next-generation sequencing workflows in veterinary infection biology: towards validation and quality assurance. *In* Potential applications of pathogen genomics (P.R. Murcia, M. Palmarini & S. Belák, eds). *Rev. Sci. Tech. Off. Int. Epiz.*, **35** (1), 67–81. doi:10.20506/rst.35.1.2418.
11. Halpin K., Tribolet L., Hobbs E.C. & Singanallur N.B. (2021). – Perspectives and challenges in validating new diagnostic technologies. *In* Diagnostic test validation science: a key element for effective detection and control of infectious animal diseases (A. Colling & I.A. Gardner, eds). *Rev. Sci. Tech. Off. Int. Epiz.*, **40** (1), XXX–YYY. doi:10.20506/rst.40.1.XXXX.
12. Bath C., Scott M. [...] & Rodoni B. (2020). –Further development of a reverse-transcription loop-mediated isothermal amplification (RT-LAMP) assay for the detection of foot-and-mouth disease virus and validation in the field with use of an internal positive control. *Transbound. Emerg. Dis.*, preprint, 13 pp. doi:10.1111/tbed.13589.

13. Johnson W.O., Jones G. & Gardner I. (2019). – Gold standards are out and Bayes is in: implementing the cure for imperfect reference tests in diagnostic accuracy studies. *Prev. Vet. Med.*, **167**, 113–127. doi:10.1016/j.prevetmed.2019.01.010.
14. Cheung A., Dufour S., Jones G., Kostoulas P., Stevenson M.A., Singanallur N.B. & Firestone S.M. (2021). – Bayesian latent class analysis when there is an imperfect reference test. *In* Diagnostic test validation science: a key element for effective detection and control of infectious animal diseases (A. Colling & I.A. Gardner, eds). *Rev. Sci. Tech. Off. Int. Epiz.*, **40** (1), XXX–YYY. doi:10.20506/rst.40.1.XXXX.
15. 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** (3), 913–935. doi:10.20506/rst.25.3.1706.
16. Brown I., Slomka M.J., Cassar C.A., Mcelhinney L.M. & Brouwer A. (2021). – Diagnostic test validation science: a key element for effective detection and control of infectious animal diseases. *In* Diagnostic test validation science: a key element for effective detection and control of infectious animal diseases (A. Colling & I.A. Gardner, eds). *Rev. Sci. Tech. Off. Int. Epiz.*, **40** (1), XXX–YYY. doi:10.20506/rst.40.1.XXXX.
17. World Organisation for Animal Health (OIE) (2019). – Chapter 1.1.5. Quality management in veterinary testing laboratories. *In* Manual of Diagnostic Tests and Vaccines for Terrestrial Animals, 8th Ed. OIE, Paris, France. Available at: [www.oie.int/fileadmin/Home/eng/Health\\_standards/tahm/1.01.05\\_QUALITY\\_MANAGEMENT.pdf](http://www.oie.int/fileadmin/Home/eng/Health_standards/tahm/1.01.05_QUALITY_MANAGEMENT.pdf) (accessed on 8 October 2020).
18. Newberry K. & Colling A. (2021). – Quality standards and guidelines for test validation for infectious diseases in veterinary laboratories. *In* Diagnostic test validation science: a key element for effective detection and control of infectious animal diseases (A. Colling & I.A. Gardner, eds). *Rev. Sci. Tech. Off. Int. Epiz.*, **40** (1), XXX–YYY. doi:10.20506/rst.40.1.XXXX.
19. Johnson P. & Caguang L. (2021). – Proficiency testing and ring trials. *In* Diagnostic test validation science: a key element for effective detection and control of infectious animal diseases (A. Colling & I.A. Gardner, eds). *Rev. Sci. Tech. Off. Int. Epiz.*, **40** (1), XXX–YYY. doi:10.20506/rst.40.1.XXXX.
20. Gardner I.A., Colling A. & Greiner M. (2019). – Design, statistical analysis and reporting standards for test accuracy studies for infectious diseases in animals: Progress, challenges and recommendations. *Prev. Vet. Med.*, **162**, 46–55. doi:10.1016/j.prevetmed.2018.10.023.

21. Kostoulas P., Gardner I.A., Elschner M.C., Denwood M.J., Meletis E. & Nielsen S.S. (2021). – Examples of proper reporting for evaluation (Stage 2 validation) of diagnostic tests for diseases listed by the World Organisation for Animal Health. *In Diagnostic test validation science: a key element for effective detection and control of infectious animal diseases* (A. Colling & I.A. Gardner, eds). *Rev. Sci. Tech. Off. Int. Epiz.*, **40** (1), XXX–YYY. doi:10.20506/rst.40.1.XXXX.
22. Heuer C. & Stevenson M.A. (2021). – Diagnostic test validation studies when there is a perfect reference standard. *In Diagnostic test validation science: a key element for effective detection and control of infectious animal diseases* (A. Colling & I.A. Gardner, eds). *Rev. Sci. Tech. Off. Int. Epiz.*, **40** (1), XXX–YYY. doi:10.20506/rst.40.1.XXXX.
23. Kirkland P.D. & Newberry K.M. (2021). – Your assay has changed – is it still ‘fit for purpose’? What evaluation is required? *In Diagnostic test validation science: a key element for effective detection and control of infectious animal diseases* (A. Colling & I.A. Gardner, eds). *Rev. Sci. Tech. Off. Int. Epiz.*, **40** (1), XXX–YYY. doi:10.20506/rst.40.1.XXXX.
24. Caraguel C.G.B., Stryhn H., Gagné N., Dohoo I.R. & Hammell K.L. (2011). – Selection of a cutoff value for real-time polymerase chain reaction results to fit a diagnostic purpose: analytical and epidemiologic approaches. *J. Vet. Diagn. Invest.*, **23** (1), 2–15. doi:10.1177/104063871102300102.
25. Kirkland P.D. & Delbridge G. (2011). – Use of a blocking ELISA for antibodies to equine influenza virus as a test to distinguish between naturally infected and vaccinated horses: proof of concept studies. *Aust. Vet. J.*, **89** (S1), S45–S46. doi:10.1111/j.1751-0813.2011.00743.x.
26. Roest H., Tilburg J., Van der Hoek W., Vellema P., Van Zijderveld F., Klaassen C. & Raoult D. (2011). – The Q fever epidemic in The Netherlands: history, onset, response and reflection. *Epidemiol. Infect.*, **139** (1), 1–12. doi:10.1017/S0950268810002268.
27. Sellens E., Bosward K.L., Norris J.M., Wood N., Heller J., Graves S. & Gidding H.F. (2020). – *Coxiella burnetii* seroprevalence in unvaccinated veterinary workers in Australia: Evidence to support Q fever vaccination. *Zoonoses Pub. Hlth*, **67** (1), 79–88. doi:10.1111/zph.12658.
28. Muleme M., Campbell A., Stenos J., Devlin J.M., Vincent G., Graves S., Cameron A.R., Wilks C.R. & Firestone S.M. (2017). – A longitudinal study of *Coxiella burnetii* transmission dynamics in intensively-managed kid goats supports early use of vaccines. *Vet. Res.*, **48** (1), 50. doi:10.1186/s13567-017-0452-3

29. Guthrie A.J., Maclachlan N.J., Joone C., Lourens C.W., Weyer C.T., Quan M., Monyai M.S. & Gardner I.A. (2013). – Diagnostic accuracy of a duplex real-time reverse transcription quantitative PCR assay for detection of African horse sickness virus. *J. Virol. Methods*, **189** (1), 30–35. doi:10.1016/j.jviromet.2012.12.014.
30. Sergeant E.S., Grewar J.D., Weyer C.T. & Guthrie A.J. (2016). – Quantitative risk assessment for African horse sickness in live horses exported from South Africa. *PLoS ONE*, **11** (3), e0151757. doi:10.1371/journal.pone.0151757.
31. Mayo C.E., Weyer C.T., Carpenter M.J., Reed K.J., Rodgers C.P., Guthrie A.J., Mullens B.A., Barker C.M., Reisen W.K. & MacLachlan N.J. (2021). – Diagnostic applications of molecular and serological assays for bluetongue and African horse sickness. *In* Diagnostic test validation science: a key element for effective detection and control of infectious animal diseases (A. Colling & I.A. Gardner, eds). *Rev. Sci. Tech. Off. Int. Epiz.*, **40** (1), XXX-YYY. doi:10.20506/rst.40.1.XXXX.
32. Tolpinrud A., Firestone S.M., Diaz-Méndez A., Wicker L., Lynch S.E., Dunowska M. & Devlin J.M. (2020). – Serological evidence for the presence of wobbly possum disease virus in Australia. *PLoS ONE*, **15** (8), e0237091. doi:10.1371/journal.pone.0237091.
33. Gardner I.A. & Greiner M. (2006). – Receiver-operating characteristic curves and likelihood ratios: improvements over traditional methods for the evaluation and application of veterinary clinical pathology tests. *Vet. Clin. Pathol.*, **35** (1), 8–17. doi:10.1111/j.1939-165x.2006.tb00082.x.
34. Caraguel C.G.B. & Colling A. (2021). – Diagnostic likelihood ratio – the next-generation of diagnostic test accuracy measurement. *In* Diagnostic test validation science: a key element for effective detection and control of infectious animal diseases (A. Colling & I.A. Gardner, eds). *Rev. Sci. Tech. Off. Int. Epiz.*, **40** (1), XXX-YYY. doi:10.20506/rst.40.1.XXXX.
35. Christensen J. & Gardner I.A. (2000). – Herd-level interpretation of test results for epidemiologic studies of animal diseases. *Prev. Vet. Med.*, **45** (1–2), 83–106. doi:10.1016/s0167-5877(00)00118-5.
36. Colling A., Lunt R. [...] & Daniels P. (2018). –A network approach for provisional assay recognition of a Hendra antibody ELISA – test validation with low sample numbers from infected horses. *J. Vet. Diagn. Invest.* **30** (3), 362–396. doi:10.1177/1040638718760102.
37. Ludi A.B., Mioulet V., Bakkali Kassimi L., Lefebvre D.J., De Clercq K., Chitsungo E., Nwankpa N., Vosloo W., Paton D.J. & King D.P. (2021). – Selection and use of reference panels:

a case study highlighting current gaps in the materials available for foot and mouth disease. *In* Diagnostic test validation science: a key element for effective detection and control of infectious animal diseases (A. Colling & I.A. Gardner, eds). *Rev. Sci. Tech. Off. Int. Epiz.*, 40 (1), XX-YYY. doi:10.20506/rst.40.1.XXXX.

38. Watson J., Carlile G.A. & Williams D.T. (2021). – The value of virtual biobanks for transparency purposes with respect to reagents and samples used during test development and validation. *In* Diagnostic test validation science: a key element for effective detection and control of infectious animal diseases (A. Colling & I.A. Gardner, eds). *Rev. Sci. Tech. Off. Int. Epiz.*, 40 (1), XXX-YYY. doi:10.20506/rst.40.1.XXXX.

39. Muller J., Gwozdz J., Hodgeman R., Ainsworth C., Kluver P., Czarnecki J., Warner S. & Fegan M. (2015). – Diagnostic performance characteristics of a rapid field test for anthrax in cattle. *Prev. Vet. Med.*, 120 (3–4), 277–282. doi:10.1016/j.prevetmed.2015.03.016.

40. Agriculture Victoria (2020). – Significant disease investigation (SDI) program. Department of Jobs, Precincts and Regions, Victoria, Australia. Available at: <https://agriculture.vic.gov.au/biosecurity/animal-diseases/significant-disease-investigation-sdi-program> (accessed on 8 October 2020).

41. Center for Health Security (2020). – Serology-based tests for COVID-19. John Hopkins Bloomberg School of Public Health, Baltimore, United States of America. Available at: [www.centerforhealthsecurity.org/resources/COVID-19/serology/Serology-based-tests-for-COVID-19.html](http://www.centerforhealthsecurity.org/resources/COVID-19/serology/Serology-based-tests-for-COVID-19.html) (accessed on 8 October 2020).

42. Stevenson M., Halpin K. & Heuer C. (2021). – Detection of emerging infectious zoonotic diseases. *In* Diagnostic test validation science: a key element for effective detection and control of infectious animal diseases (A. Colling & I.A. Gardner, eds). *Rev. Sci. Tech. Off. Int. Epiz.*, 40 (1), XXX-YYY. doi:10.20506/rst.40.1.XXXX.

43. Reising M.M., Tong C., Harris B., Toohey-Kurth K.L., Crossley B., Mulrooney D., Tallmadge R.L., Schumann K.R., Lock A.B. & Loiacono C.M. (2021). – A review of guidelines for evaluating a minor modification to a validated assay. *In* Diagnostic test validation science: a key element for effective detection and control of infectious animal diseases (A. Colling & I.A. Gardner, eds). *Rev. Sci. Tech. Off. Int. Epiz.*, 40 (1), XXX-YYY. doi:10.20506/rst.40.1.XXXX.

**Table I**

**Examples of OIE-listed pathogens and tests used for testing purposes**

OIE purpos	Example
1a) Demonstration of freedom from infect vaccination	<ul style="list-style-type: none"> <li>• Testing to demonstrate evidence of 'historical freedom', e.g. <b>African swine fever (ASF)</b> virus surveillance testi had an ASF outbreak and all surveillance and survey testing results were negative. The country would also be risk of ASF introduction over the same time period.</li> <li>• <b>Equine influenza</b> - blocking enzyme-linked immunosorbent assay (ELISA) used in combination with specificall (canarypox-vectored) to enable differentiating infected from vaccinated animal (DIVA) testing and subsequent following incursion, large outbreak and eradication in 200 (25). This example also applies to 1(b).</li> <li>• Histopathological examination of the brain of animals with neurological signs to detect the modified prion protei <b>encephalopathy</b>.</li> </ul>
1b) Re-establishment of freedom after ou	<ul style="list-style-type: none"> <li>• 'Proof of freedom' testing after a disease outbreak of <b>white spot syndrome virus (WSSV)</b> in prawns in Austr developed and evaluated for detection of WSSV in apparently-healthy animals.</li> </ul>
2) Certify freedom from infection or prese animals or products for trade/movement	<ul style="list-style-type: none"> <li>• Serological tests to certify cats and dogs free for international travel such as <b>Ehrlichia canis</b> indirect fluorescer <b>Brucella canis</b> serum agglutination test (SAT), and <b>Hendra virus</b> serum neutralisation test (SNT) in horses.</li> </ul>
3) Contribute to the eradication of disease from defined populations	<ul style="list-style-type: none"> <li>• Serological and milk-tank testing for <b>Brucella abortus</b> in cattle, caudal-fold skin testing for <b>Mycobacterium bo species</b>, serological testing for <b>pseudorabies virus</b> in commercial swine herds, and serological and antigen he milk by ELISA and reverse transcription polymerase chain reaction (RT-PCR) to eliminate <b>Mycoplasma bovis</b></li> </ul>
4) Confirm diagnosis of suspect or clinica of positive screening test)	<ul style="list-style-type: none"> <li>• Confirmation (ruling in) of <b>foot-and-mouth disease (FMD)</b> in a herd of cattle with some individuals having clini salivation. If FMD was excluded because of negative test results (ruling out), testing may be done for other dise diagnosis list.</li> <li>• Use of a <b>Hendra virus</b> SNT to confirm a positive sol G HeV ELISA result in an unvaccinated horse.</li> <li>• <b>Bluetongue virus (BTV)</b> group antibody ELISA and subsequent testing of positive serum with serotype SNT.</li> </ul>

- 5) Estimate prevalence of infection or exposure (surveys, herd health status, disease control)
  - 6) Determine immune status of individual (vaccination)
  - 7) 'Other'  
Seasonal control of infection and epidemiology
- TaqMan A positive **avian influenza** (AI) virus swab and haemagglutinin (HA) and neuraminidase (NA) testing (sequencing).
  - **Bluetongue virus** (BTV) serology and molecular testing, including sequencing to estimate prevalence of infection and monitoring of spectrum (exotic serotypes) of BTV, including changes of geographical distribution to facilitate risk Monitoring Program, NAMP).
  - PCR on bulk tank milk used extensively in the Netherlands outbreak of **Q fever** from 2009–2010 to guide surveillance.
  - Testing of serum samples from commercial beef cattle in Australia by indirect immunofluorescent assay (IFA) for **Coxiella burnetii** exposure (27).
  - Fluorescent antibody virus neutralisation (FAVN) testing to assess the immune status of dogs and cats post-vaccination (i.e. for international travel [FAVN test > 0.5 IU to be categorised as protected for rabies and allowed to travel]).
  - Use of SNT titres in **Hendra virus** vaccinated horses to assess immune response after vaccination.
  - Haemagglutination inhibition testing for post-vaccination monitoring of antibody titres to **Newcastle disease virus**.
  - Anti-hemagglutinin competitive ELISA for antibody against **peste des petits ruminants** (PPR) virus to estimate response after vaccination.
  - Immunofluorescent assay testing for contagious metritis (infection with **Taylorella equigenitalis**) in stallions at breeding season.
  - Epidemiological research into the dynamics of disease, such as a longitudinal study of serological responses to **C. burnetii** and shedding at kidding among intensively managed goats to understand the optimal administration (28).
-

## Box 1

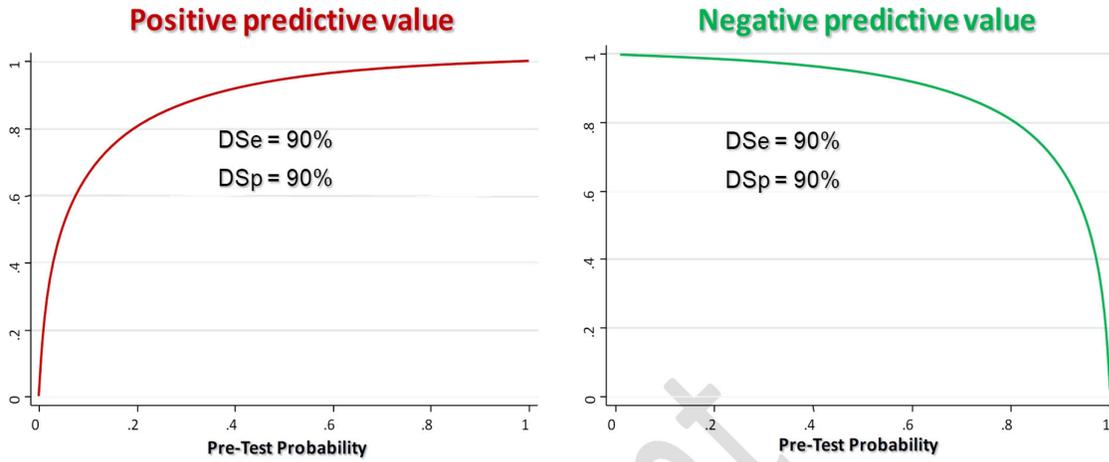
### Definitions for test accuracy parameters

**DSe** = percentage of truly-infected animals that would test positive

**DSp** = percentage of truly non-infected animals that would test negative

**LR+** = how many times more likely a positive test result is expected to occur in an infected compared with a non-infected animal:  $LR+ = DSe/(1-DSp)$

**LR-** = how many times more likely a negative test result is expected to occur in a non-infected compared with an infected animal:  $LR- = (1-DSe)/DSp$



DSe: diagnostic sensitivity

DSp: diagnostic specificity

**Fig. 1**

**Positive and negative predictive values for a test with DSe = DSp = 90% as a function of pre-test probability for an individual animal (or prevalence in the source population if the animal has no risk factor information, e.g. age or clinical signs, to distinguish its disease risk from other animals in the same population)**