Your assay has changed – is it still 'fit for purpose'? What evaluation is required?

P.D. Kirkland [(1)*] & K.M. Newberry [(2)]

(1) Elizabeth Macarthur Agriculture Institute, Woodbridge Road, Menangle, NSW 2568, Australia

(2) Australian Centre for Disease Preparedness, CSIRO, 5 Portarlington Road, East Geelong, VIC 3220, Australia

*Corresponding author: peter.kirkland@dpi.nsw.gov.au

Summary

A reliable laboratory assay is an essential tool for the diagnosis or surveillance of most animal diseases. Before routine use, assays should be appropriately validated to ensure that they have performance characteristics that provide reliable results and can be used for the intended purpose. It is inevitable that, over time, changes will need to be made to assay reagents, to the assay format, to test a different species or for implementation in a new laboratory. Whenever there is a change (whether it be components, application or location), it is essential to establish whether the new circumstances affect the biological basis and properties of the assay. If the modifications do not affect the biological basis of the assay, the changes might be considered minor and a verification study can be conducted to confirm that the performance characteristics have not been adversely affected. Major changes require a new validation to be carried out. A method comparability study, where original and modified assays are run concurrently to test the same sample panel, provides an extremely robust comparison. However, comparability studies are not always an option, especially for the introduction of a method to a new laboratory. Access to original validation data and suitable reference sample panels then becomes essential to provide evidence that the assay remains 'fit for the intended purpose'.

Keywords

Method comparison – Performance characteristics – Reference sample panels – Sensitivity – Specificity – Validation – Verification.

OIE Pre-print

Introduction

The terms validation and verification can sometimes be confusing as the objectives are broadly the same – to demonstrate that a test or method is 'fit for purpose'. Further, some of the steps and procedures involved can also be the same or very similar. In brief, when a new assay is developed, it is subjected to a wide range of assessments to establish its performance characteristics and to determine that these characteristics are of a sufficiently high quality to confirm that the test can be used for the intended purpose. This overall process is known as 'validation' and should also identify the limitations of a method, as well as factors that influence its accuracy and precision and to what extent they do so.

The definition of 'verification', and when such a process is an acceptable alternative to validation, is more problematic. Apart from detailed validation studies, it could be considered that any other evaluations that are undertaken are done so to verify that the assay remains 'fit for purpose'. In this manuscript, verification will be considered in this broad context. The international accreditation standard used by most veterinary laboratories states, 'The laboratory shall verify that it can properly perform methods before introducing them by ensuring that it can achieve the required performance. If the method is revised by the issuing body, verification shall be repeated to the extent necessary' (1). The World Organisation for Animal Health (OIE) *Manual of Diagnostic Tests for Terrestrial Animals* (*Terrestrial Manual*) states, in Chapter 2.2.8., 'Verification represents evidence that the performance characteristics, e.g. accuracy and precision of a validated assay, are comparable when used in another laboratory' (2). However, this chapter of the OIE *Terrestrial Manual* also uses the term 'comparability' for a process to demonstrate that a test that has undergone 'minor' changes remains fit for purpose. In general, comparative studies are only appropriate for minor changes (2, 3). Another chapter of this thematic issue (3) provides a comprehensive review of minor changes to assays that would only require a comparative evaluation.

In contrast, any modifications that would alter the biological basis of the assay are regarded as major (2, 3). For example, the use of a test for a different animal species or specimen type, or, in the context of a polymerase chain reaction (PCR) assay, a change in the primers and probes for different targets in different regions of the genome would constitute major changes and would require a comprehensive validation rather than a limited verification study.

The most important decisions for verification studies (regardless of their type) are practical in nature and relate to study design. The resources and requirements for the introduction of a validated assay into a new laboratory will be different from those in which there has been a minor change in an assay component. Nevertheless, in each instance, the starting point must be the

OIE Pre-print

identification of a benchmark against which performance assessments will be judged. If a new assay, for which the laboratory has no current capability, is being established, the laboratory will be dependent on either published information or information provided by the kit manufacturer. While there are no specific steps documented, some guidance is provided by the OIE (4, 5). Chapter 1.1.6. of the OIE *Terrestrial Manual* states the following:

> 'If a laboratory is considering the use of a validated commercial kit or a candidate assay based on published literature with validation data, some form of verification will be required to determine whether the assay complies with either the kit manufacturer's or the author's assertions, with respect to Stage 1 validation criteria, in the context of the intended application. This may require a limited verification of both Analytical Specificity and Analytical Sensitivity using available reference materials, whether they be external and/or locally acquired from the target population. Once the laboratory is confident that the assay is performing as described from an analytical perspective, then proceeding to a limited Stage 2 validation should be considered in the context of the intended application and target population before the assay is put into routine diagnostic use' (4).

These steps can be used as a guide to identify the aspects that may need to be considered during a verification study. If the assay is the subject of a validation that has been accepted by the OIE and listed on the OIE Register of diagnostic kits, access to data relating to performance criteria should not be an impediment, with a summary of essential data available on the OIE website (6). However, in some instances, one of the challenges may be the availability of suitable reference samples. After an assay has been established as a new capability for a laboratory, parallel testing in collaboration with another laboratory and participation in a proficiency-testing programme are valuable ways to expand the data panel to demonstrate 'fitness for purpose'.

When there has been a minor change to an assay, verification that the assay remains 'fit for purpose' can take the form of a comparability or methods comparison study. With adequate forward planning, this evaluation should produce a robust outcome because both assays (the original before the change and the new, modified format) can be run concurrently, with only a single parameter changed. The original assay becomes the benchmark against which the impact of the change is assessed. If there is a need to change more than one component, then the changes should be made in a stepwise manner, with only one change being evaluated at a time. An excellent series of examples of the application of a methods comparison study has been provided by Reising *et al.* (7).

Achieving similar performance characteristics after a change provides evidence that the test remains 'fit for that purpose'. On the other hand, when a change is made that has sufficient

OIE Pre-Print

impact that the performance of the test does not meet the original criteria, it may be necessary either to undertake a validation study or, for changes in test components, to reject that batch of reagents and source new materials.

Ultimately, the extent of the assessment that is required must be judged by appropriately skilled scientists. The flow chart in Appendix 1 may be a useful guide to determine whether a verification process will be adequate or whether validation is required to ensure that the test remains 'fit for a defined purpose'.

Reference sample panels

Whether the goal is to establish satisfactory performance of a new commercially produced kit, to confirm the performance of a published test or to confirm the ongoing fitness of an established 'in-house' test after minor changes, there is a need to have access to suitable reference materials to use in a verification study. These contribute to the benchmark against which assay performance is judged. Depending on the nature of the test, sourcing and maintaining a stock of appropriate reference materials can be challenging, particularly for infectious diseases. These materials are likely to be used over a long period of time. Ideally, this should be a selection that was used in the original assay validation, though this can be difficult when verifying the performance of a commercially manufactured test kit or for assays developed in other countries. For some diseases, national or international reference samples are available, but these are usually limited in both number and volume and cannot provide the full range of reactivity needed for a comprehensive verification.

When sourcing samples to establish a reference panel for infectious diseases, it is important to consider the biology of the disease and time line for decay of the analyte to ensure that samples are of a reliable status for the test. For example, while it is important to have samples with varying concentrations of the analyte, such as antibodies, selecting samples early in the course of an infection or in animals with declining analyte levels (e.g. maternally derived antibodies) may give more variable results and with greater variation in isotypes, especially immunoglobulin G (IgG) and immunoglobulin M (IgM). It is essential that reference panels are stored in a manner that maximises stability to ensure reliable performance over a long period of time. Depending on the stability of the reference material, this may involve preparing a large number of single-use aliquots that are stored frozen, freeze dried or treated to allow long-term storage under refrigeration. There are no definitive guidelines on the number of samples that should be included in a reference collection. However, the collection should cover the spectrum of reactivity that is routinely encountered for both positive and negative samples in the assay, and should be of sufficient volume to allow replication during testing. To support long-term standardisation, each

OIE Pre-print

sample in a collection should have sufficient volume to support its use over many years. Further considerations for the development and use of biobanks and reference collections are also described in other chapters of this thematic issue (see Watson *et al.* and Ludi *et al.)* (8, 9).

Although it is difficult to specify the absolute number of samples needed for a verification study, panels consisting of about 30 samples from infected animals and a similar number of uninfected animals, representing the target population, will usually provide an acceptable degree of rigour. The number of samples will also be influenced by other factors mentioned elsewhere, such as the significance of the disease, purpose of the test and stage of infection. Within the sample collection, if the main purpose of testing is for surveillance, there may be a trend towards weaker reacting samples, whereas, for routine diagnostic testing, the samples may cover the spectrum of reactivity encountered for the disease under investigation.

When establishing reference sample panels for either validation or verification studies for infectious diseases, the value of a collection of samples from a longitudinal study cannot be overstated. Most of the limitations with experimentally infected animals can be overcome if a natural route of infection is used. With careful planning, it is possible to generate frequent and critically timed collection of an appropriate range of samples. These can be invaluable, supporting first the evaluation of methods for agent detection (whether by direct antigen detection methods, PCR or culture) and, later, the detection of antibodies. Since the time of exposure to the pathogen is precisely defined, accurate data can be established for the onset and duration of nucleic acid and/or antigen detection, as well as the length of time during which the infectious agent may be detected (the size of the 'diagnostic window'). Samples collected later in the course of the infection underpin determination of the onset of an antibody response, the identification of samples in which IgM is likely to be detected and changes in antibody titres over time. In some instances, the IgM response can be strong at the onset of disease and rapidly decline. As an example, if an IgM enzyme-linked immunosorbent assay (ELISA) for West Nile virus was being verified, samples collected at three weeks after the onset of disease would not be suitable, as there is only infrequently a detectable IgM response three weeks after the first observation of disease (10).

Although usually not the primary purpose of an experimental study, with appropriate storage, these samples become a very valuable resource. However, it is extremely important to include a selection of negative and positive samples from the target population to verify assay performance. Further, while serially collected samples from a longitudinal study are invaluable for assessing or optimising test performance over time, such samples cannot be used to determine

OIE Pre-print

diagnostic sensitivity (Dse) or diagnostic specificity because they do not fulfil the statistical requirement for independent observations (4).

When is a verification study required?

Examples of some of the circumstances and changes or factors that can affect the performance of an assay and should establish the need for a verification study or perhaps, in some instances, limited validation studies include:

- introducing a well-validated test into a different laboratory;
- changes in the environmental conditions in which the test is performed. In some instances, even the move to a new facility on the same site has affected assay performance due to temperature variation and differences in water quality or equipment;
- changes in batches of test reagents or controls for an assay developed 'in house';
- changes to a method of production for reagents – for example, a different purification process for an antigen used in a serological test. In contrast, a change from an 'intact' or whole agent compared to the use of a purified protein or subunit, or from infectious to inactivated organisms or viruses, would trigger the need for a validation study;
- use of a test in a different geographical region or animal population from the one in which it was originally developed. Under these circumstances, the new laboratory needs a suitable collection of locally collected samples of known status for the pathogen of interest. This will inevitably require the use of other assays to establish the status of the donor animals. This may be challenging for agents with many different strains or serotypes. Depending on the availability of suitable material, validation may be required, rather than a verification study;
- for 'pan-reactive' assays, confirmation that the emergence or evolution of new strains or taxonomic species does not compromise assay performance. Long-term reliable detection of different strains or subtypes of infectious agents may be challenging;
- a change in the dilution factor for samples, particularly in serological and antigen detection assays;
- depending on the type of test, a change in the animal species may require either verification or validation. The impact of a change in species is unlikely to impact on an agent detection test such as PCR or a cultural method and a verification study would probably be acceptable. In contrast, a serological test to detect antibodies in a different animal species is likely to need a validation study;

OIE Pre-print

- the need to test for the same agent in a different sample matrix. The acceptability of a verification study may vary, depending on the class of test. A verification study may be appropriate for an agent detection assay (e.g. PCR) to test plasma instead of serum, whereas employing the same assay to test semen samples would require a validation study.

When undertaking a verification study, it is important to confirm that the essential performance characteristics can be achieved. If the verification is being undertaken as the result of a minor change that, based on experience, is unlikely to result in a change in performance, the evaluation may be less rigorous than in a situation where the consequences are unclear. For example, confirming the performance of a new batch of master-mix for a real-time PCR assay from a commercial supplier does not need to be as exhaustive as evaluating the introduction of a commercially prepared product from a new supplier in the market. The extent of a verification study must be guided by scientists and technicians with relevant expertise with the class of test involved. The expertise and requirements for an assay to detect antibodies in an unusual sample matrix will differ from those needed for a real-time PCR assay, and those for an antigen detection test will differ again. Professional judgement will be needed to determine whether the change is considered to be minor or major. This, in turn, influences the scope of the verification process or dictates that a validation study is required. In making these decisions, factors that must be considered include:

- the extent to which the test has previously been validated
- the availability of appropriate documentation supporting prior validation studies
- the availability of suitable reference panels and standards
- the purpose for which the test or method is being used.

Most importantly, it is essential to demonstrate that an assay remains ‘fit for purpose’. The rigour that is required to achieve this may be guided by the purpose for which the assay is being used and, in turn, the significance and implications of the results. The implications will inevitably be greater when a test for which the results have serious consequences returns a sub-optimal verification outcome. For example, the poor performance of a test for a multi-species zoonotic pathogen, a test for international trade or for a proof-of-freedom survey after an exotic disease outbreak will be viewed very differently from poor performance in a test being used for routine surveillance of an endemic disease.

In planning the scope of a verification study, the available data from an initial test validation should be reviewed by a scientist with expertise with the class of test involved, as well as the

OIE Pre-print

ability to assess the quality and suitability of the original validation study. At a minimum, it is likely that the performance characteristics to be assessed will include:

- analytical sensitivity
- analytical specificity
- accuracy
- precision
- repeatability.

Depending on the extent of verification needed, studies will perhaps include testing:

- representative healthy animals from the target populations (selectivity)
- animals with known infection to related agents/disease (exclusivity)
- a determination whether the test will detect infection with all strains/types of interest (inclusivity).

Examples of verification studies

Validation and verification of assays and methods used for applications in the biological sciences tend to be inherently more complex than those used for situations in the physical sciences. The examples that follow are limited to situations arising from the verification of assays that have been used for infectious diseases. It is hoped that these will illustrate some of the complexities and perhaps unexpected outcomes that have been encountered. It is important to have a strong understanding of the 'biology' of an infectious disease, especially the interactions between the host, pathogen and environment. However, there can be a wide range of other factors that may have an influence on the reliable performance of a test. Even though a test system might perform reliably, 'extrinsic' factors, such as sample matrix and specimen processing, can be important. While there are many different classes of tests and the range of applications is almost endless, the examples that follow provide an insight into situations that have been encountered with assay types that are in frequent use for infectious disease diagnosis.

Verification of specificity

When confirming the analytical specificity of a test for antibodies, it is relatively common to test samples from animals that have not been exposed to the pathogen of interest, either by virtue of their age or geographical location (e.g. they originate from an area where the disease of interest has never been detected or, for a vector-borne agent, an area where the insect vector has never been detected). However, other factors, such as the varying responses of different animal species' immunological factors or extraneous influences, such as the circulation of cross-reactive agents,

OIE Pre-print

may also be relevant. During verification of the specificity of a commercially available ELISA for assessment of antibodies against avian metapneumovirus (aMPV), the test was found not to be fit for the purpose of serosurveillance in a population of free-range 'village' chickens from a country known to be free from aMPV disease. The ELISA had been used for many years to assess sentinel chickens as part of an import protocol for quarantine purposes. The intended use of the test, as stated by the manufacturer, was to monitor aMPV type A, B and C antibodies from vaccinated chickens and turkeys, as well as in naïve (specific-antigen-free or SPF) flocks. However, when an unvaccinated population of free-range 'village' chickens, free from aMPV disease (population 2) were tested, the outcomes were remarkedly different from those gained from testing a population of sentinel chickens for quarantine and import purposes (population 1). The mean of sample-to-positive (S/P) ratio (sample optical density [OD]/positive control OD) for population 1 was 0.01 and for population 2 was 0.39. The negative/positive cut-off ratio of the kit was 0.2 S/P. Testing of the two populations was done in the same laboratory, under similar conditions, by the same technician. Commercial kit batches that were used varied. However, the in-house controls used between batches had similar test outcomes and both tests passed internal quality controls. Positive ELISA results were verified, using OIE-recommended standard immunofluorescence assay (IFA) and virus neutralisation test (VNT) methods, as false positives. Figures 1 and 2 show a frequency distribution of the ELISA results for both populations, indicating how differently two expected 'negative' populations may manifest themselves.

As population 2 came from free-range chickens residing in a country which is officially declared to be free of avian pneumovirus (APV), confirmation was required. The use of group-reactive aMPV IFAs and specific VNTs for all serotypes of aMPV (types A, B and C) gave negative results, suggesting that the ELISA results were false positives. The outcome of the verification was that the ELISA, when used alone, is not fit for the purpose of surveillance testing of free-range (village-type) chickens, but is fit for the purpose of testing sentinel chickens used during the monitoring of imported chickens in quarantine. If the test is to be used for surveillance purposes, the MedCalc software (Ostend, Belgium) suggests a cut-off at 0.81, giving the highest combined diagnostic sensitivity (100%) and diagnostic specificity (91%) using the data from Figure 2. The most likely explanation for the reduced diagnostic specificity is the difference in age and geographic location of each population. The sentinel chickens are considered to be almost SPF, due to their age and the environment in which they were raised. The free-range chickens were much older and more likely to have been exposed to a range of pathogens in the area where they were raised.

Care should be taken when selecting animals for specificity studies to ensure that they are truly representative of the target population. Using the correct representative populations to determine

OIE Pre-print

selectivity (representative, healthy populations), exclusivity (known infection to related agents) and inclusivity (detecting all strains/types of interest) is important. Results obtained when testing samples from young SPF chickens may be 'fit for purpose' when testing SPF chicken populations, but are likely to be quite different from those obtained when testing adult birds in commercial flocks, where exposure to other pathogens that may increase the level of 'background' (non-specific) reactivity is frequently encountered. Vaccination of test animals for diseases other than the target pathogen should also be considered as this can increase the potential for 'non-specific' reactivity and loss of analytical specificity. As well as having a potential impact on analytical specificity, the response to an unrelated antigen (for example, cell-culture proteins from a vaccine with a semi-purified virus preparation) can also impact on analytical sensitivity. Vaccinating horses with an equine herpesvirus-1 vaccine interfered with the equine arteritis VNT and had the potential to reduce test sensitivity (11).

Factors other than assay 'cut-off' values may also need to be considered. When testing sheep and cattle for antibodies against bluetongue viruses, to achieve an optimal balance between sensitivity and specificity with a blocking ELISA, it was necessary to test sheep sera at a dilution of 1/10 whereas cattle sera were tested at a 1/5 dilution.

Verification of sensitivity

Where there are multiple antigenically or genetically different strains or subtypes of pathogens, broadly reactive ('pan-reactive') assays can be extremely valuable for screening purposes, for either surveillance or disease investigation. Typical examples include bluetongue and influenza viruses and bacteria from the genus *Leptospira*. When using a well-designed, broadly reactive, real-time PCR assay, failure to detect the agent in appropriate samples from acutely affected animals can (sometimes falsely) exclude the target agent as a cause of the disease. Similarly, during surveillance, negative results across a suitable number of animals from a population in a broadly reactive serological assay would suggest that the agent is probably not circulating in the region. However, maintaining an appropriate level of sensitivity and specificity can be demanding due to the broad spectrum of reactivity that these assays need to cover. Some of the challenges that can be encountered during verification studies are highlighted by the following examples for a bluetongue group-reactive blocking antibody ELISA and a real-time PCR assay for pestiviruses.

As a result of the difficulties and costs associated with producing test kits for infectious diseases, there are usually a limited number of commercial manufacturers. When test kits are validated, it is common to test samples from a region near the location of the manufacturer. However, this can introduce bias and ultimately affect either sensitivity or specificity. For example, during the

OIE Pre-print

verification of blocking ELISA kits for antibodies against bluetongue virus in cattle, one kit produced in Australia had a specificity of 99.9% and sensitivity of 99.5% when testing panels of sera of known status collected from Australia, Canada, the People's Republic of China, Italy, New Zealand and the United States of America (USA). On the other hand, although kits produced in Europe and the USA both had high specificity when testing the international sample panel, the sensitivity was inferior (90%–95%), despite those kits having had acceptable sensitivity when testing locally collected sera. This difference was considered to be due to the original selection of test monoclonal antibodies against the limited range of virus serotypes that were available locally.

During the investigation of a disease outbreak in sheep, serology results suggested that a pestivirus, most likely border disease virus (BDV), could be involved. Testing samples from aborted foetuses resulted in very weak reactivity of a single sample in an 'in-house', pan-reactive, real-time reverse transcription PCR (RT−PCR) for the detection of pestiviruses. All other samples gave negative results. This assay had been thoroughly validated and published and was in widespread use internationally. Further testing of blood samples from ten newborn lambs gave some positive results in the pan-reactive PCR (Table I). In order to confirm the specific virus that was involved, the same blood samples were then tested using a published BDV-specific assay. The results obtained showed that the pan-reactive assay lacked sensitivity. There were markedly higher cycle threshold (Ct) values when compared to the BDV-specific assay, with three samples giving a negative result.

To verify the unexpectedly poor results in the published pan-pestivirus, real-time RT−PCR, a methods comparison study was undertaken. All samples were subsequently tested using an antigen detection ELISA and a modified, pan-pestivirus, real-time PCR. Collectively, the results confirmed that each of these animals was infected with BDV. The data showed that the original, pan-pestivirus, real-time RT−PCR had both poor analytical and diagnostic sensitivity. The analytical sensitivity of the pan-reactive RT−PCR was approximately 100-fold less than that of the modified pan-pestivirus assay (a mean Ct difference of 7.2). The modified pan-reactive assay was also about ten-fold less sensitive than the BDV-specific assay (mean Ct difference of 3.1). The published, pan-pestivirus, real-time RT−PCR also had lower diagnostic sensitivity (only detecting 7/10 infected animals), although it might be argued that this assay could be 'fit for purpose' if the objective was to achieve a flock diagnosis. In contrast, the antigen ELISA and both the modified pan-pestivirus and BDV-specific real-time RT−PCR assays had acceptable diagnostic sensitivity at both the flock and individual animal levels. It was also concluded that the established pan-reactive PCR assay was not 'fit for purpose' for the diagnosis of BDV in sheep.

OIE Pre-print

Extrinsic factors affecting assay results

During verification studies, with the wide range of assay classes that are available, it is also important to consider how samples were processed during an original validation and, in some cases, whether there was any exposure of the target agent to any other substances. Recent experiences with testing for severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) have clearly shown that a sample transport medium that has been designed for testing by culture methods may not be suitable for testing by molecular methods (12). Some of the virus transport media investigated rapidly destroyed the target RNA while others contained extraneous agents as the result of the addition of foetal bovine serum to the medium. A similar situation has been reported (13) when detecting viral antigens and nucleic acid in tissue samples. Substances used to elute the antigens from tissues rapidly degraded the nucleic acid to give false-negative results. If these aspects are not considered during assay validation and verification, the conclusions could be quite erroneous.

Availability and use of reference sample panels

There are many examples of published experimental studies that could provide excellent samples for both assay validation and verification (see King *et al.*, this issue) (9). For example, a longitudinal study of equine influenza in naturally infected horses (14), in which nasal swabs and serum samples were collected daily, has been used to verify the performance characteristics of two real-time PCR assays and two blocking ELISAs for antibody detection. This sample panel was extremely useful for comparative testing by demonstrating the onset, progression and cessation of virus excretion, as well as the development of the immune response. In many instances, a study that has been conducted for a completely different purpose (such as control animals from vaccine trials where there is a challenge with the pathogen of interest) can provide a valuable set of samples.

Maintenance of assay performance

The scope of a verification study will be influenced by the purpose for which an assay or method is used and the nature of the change that has occurred, and should be commensurate with the scale of the original validation. Once an assay has been introduced for routine use, it should be subjected to ongoing 'verification'. Although it may not always be recognised as part of the verification process, participation in external proficiency testing and monitoring the performance of test control samples both contribute to the continuing demonstration that an assay remains 'fit for purpose' (see Waugh and Clarke, this issue) (15).

OIE Pre-print

Scientists in both diagnostic and research settings have a need for assays and methods that are 'fit for purpose'. Perhaps the most essential component that is needed for either assay validation or verification is the availability of well-characterised sample panels appropriate for the intended purpose, and in sufficient quantity to last for many years. It is important for all scientists to be vigilant in identifying and retaining suitable materials when they become available.

Conclusions

Verification studies are essential whenever a test is used in a new laboratory and if there has been a change in a component of an assay or the method. It is important to review the nature of the change to ensure that it is of a minor nature and does not affect the biological basis and properties of the method. Otherwise, a full validation study would be required to assess the impact of the change. As there are no rigid guidelines for verification studies, each situation requires careful appraisal by suitably qualified and experienced scientific staff to see that an appropriate evaluation is carried out to ensure that the assay remains 'fit for purpose'.

OIE Pre-print

Votre essai a été modifié – est-il toujours « apte à l'emploi » ? Comment faut-il l'évaluer ?

P.D. Kirkland & K.M. Newberry

Résumé

Les essais de laboratoire fiables constituent des outils essentiels pour le diagnostic ou la surveillance d'une majorité de maladies animales. Avant d'être utilisés en routine, les essais doivent faire l'objet d'une validation appropriée, destinée à s'assurer qu'ils possèdent les caractéristiques de performance nécessaires pour générer des résultats fiables et à déterminer qu'ils peuvent être utilisés pour la finalité prévue. Au fil du temps, il est inévitable que certains changements soient apportés aux réactifs de l'essai ou à son format, visant par exemple à appliquer le test sur autre espèce animale ou dans un nouveau laboratoire. À chaque changement introduit (qu'il s'agisse d'une composante du test, de sa modalité d'application ou du lieu où il est conduit), il est essentiel de déterminer si ces nouvelles circonstances affectent la base biologique et les propriétés de l'essai. Si les modifications n'affectent pas la base biologique de l'essai, les changements peuvent être considérés comme mineurs et une étude de contrôle des performances pourra être réalisée pour confirmer que les caractéristiques de performance n'ont pas subi d'altération indésirable. En cas de changement majeur, une nouvelle étude de validation devra être réalisée. L'étude de comparabilité de méthodes, qui consiste à réaliser simultanément l'essai original et l'essai modifié sur un même panel d'échantillons fournit une comparaison

extrêmement robuste. Néanmoins, dans certaines situations les études de comparabilité ne sont pas une option, notamment lorsqu'il s'agit d'introduire la méthode modifiée dans un nouveau laboratoire. Il devient alors indispensable de pouvoir accéder aux données de validation originales et de disposer de panels d'échantillons de référence appropriés afin de fournir la preuve que l'essai est toujours « apte à l'emploi » qui lui a été assigné.

Mots-clés

Caractéristiques de performance – Comparaison de méthodes – Contrôle des performances d'un test – Panel d'échantillons de référence – Sensibilité – Spécificité – Validation.

Si su ensayo ha cambiado, ¿sigue siendo "idóneo para el propósito previsto"? ¿Qué evaluación se requiere?

P.D. Kirkland & K.M. Newberry

Resumen

Un ensayo de laboratorio fiable es una herramienta básica para el diagnóstico o la vigilancia de la mayoría de las enfermedades animales. Antes de poder emplear de forma sistemática un ensayo es preciso validarlo debidamente, para tener la seguridad de que presente características de rendimiento adecuadas, que deparen resultados fiables, y de que se ajuste al propósito previsto. Con el tiempo, inevitablemente, será preciso modificar los reactivos y el formato del ensayo, con objeto de aplicarlo a una especie diferente o de practicarlo en un nuevo laboratorio. Siempre que haya un cambio en el ensayo (ya sea en sus componentes o en su modo o lugar de aplicación), será esencial determinar si las nuevas circunstancias influyen en su base biológica o en sus propiedades. Cuando las modificaciones no incidan en su base biológica, se podrá considerar que los cambios son de importancia menor y se podrá realizar un estudio de verificación para comprobar que las características de rendimiento no se han visto negativamente afectadas. Los cambios de mayor entidad exigen un nuevo proceso de validación. Los estudios de comparación de métodos, en los que paralelamente se aplican la técnica original y la modificada a un mismo panel de muestras, deparan una comparación sumamente robusta. Sin embargo, los estudios de comparabilidad no siempre son una posibilidad factible, sobre todo cuando se trata de empezar a aplicar un método en un nuevo laboratorio. En tales casos resultará fundamental tener acceso a los datos de validación originales y a paneles adecuados de muestras de referencia para asegurarse de que el ensayo siga siendo «idóneo para el propósito previsto».

Palabras clave

OIE Pre-print

Comparación de métodos – Características de rendimiento – Especificidad – Paneles de muestras de referencia – Sensibilidad – Validación – Verificación.

References

1. International Organization for Standardization (ISO) (2017). – ISO/IEC 17025: 2017: General requirements for the competence of testing and calibration laboratories. Section 7.2.1.5. ISO, Geneva, Switzerland. Available at: www.iso.org/standard/66912.html (accessed on 7 December 2020).

2. World Organisation for Animal Health (OIE) (2018). –Comparability of assays after changes in a validated test method. Chapter 2.2.8. *In* Manual of Diagnostic Tests for Terrestrial Animals. OIE, Paris, France, 11 pp. Available at: www.oie.int/fileadmin/Home/eng/Health_standards/tahm/2.02.08_COMPARABILITY_ASSAYS_AFTER_CHANGE0S.pdf (accessed on 7 December 2020).

3. 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.A. Gardner & A. Colling, eds). *Rev. Sci. Tech. Off. Int. Epiz.*, **40** (1), XX–YY. doi:10.20506/rst.40.1.XXXX.

4. World Organisation for Animal Health (OIE) (2018). – Principles and methods of validation of diagnostic assays for infectious diseases. Chapter 1.1.6. *In* Manual of Diagnostic Tests for Terrestrial Animals. OIE, Paris, France, 16 pp. Available at: www.oie.int/fileadmin/Home/eng/Health_standards/tahm/1.01.06_VALIDATION.pdf (accessed on 7 December 2020).

5. World Organisation for Animal Health (OIE) (2019). – Principles and methods of validation of diagnostic assays for infectious diseases. Chapter 1.1.2. *In* Manual of Diagnostic Tests for Aquatic Animals. OIE, Paris, France. Available at: www.oie.int/index.php?id=2439&L=0&htmfile=chapitre_validation_diagnostics_assays.htm (accessed on 7 December 2020).

6. World Organisation for Animal Health (OIE) (2021). – Register of diagnostic kits certified by the OIE as validated as fit for purpose. OIE, Paris, France. Available at:

OIE Pre-print

www.oie.int/scientific-expertise/registration-of-diagnostic-kits/the-register-of-diagnostic-kits/ (accessed on 20 March 2021).

7. Reising M.M., Schumann K.R., Crossley B.M., Glas P.S., Jacobson R.H., Toohey-Kurth K.L., Pedersen J.C., Siev D. & Martin B.M. (2014). – Methods comparison: an alternative approach for evaluating the impact of a modification to a validated assay. *J. Vet. Diagn. Invest.*, **26** (4), 480–487. doi:10.1177/1040638714535402.

8. Watson J.W., 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.A. Gardner & A. Colling, eds). *Rev. Sci. Tech. Off. Int. Epiz.*, **40** (1), XX–YY. doi:10.20506/rst.40.1.XXXX.

9. 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.A. Gardner & A. Colling, eds). *Rev. Sci. Tech. Off. Int. Epiz.*, **40** (1), XX–YY. doi:10.20506/rst.40.1.XXXX.

10. Read A.J., Finlaison D.S., Gu X., Hick P.M., Moloney B., Wright T. & Kirkland P.D. (2019). – Clinical and epidemiological features of West Nile virus equine encephalitis in New South Wales, Australia, 2011. *Aust. Vet. J.*, **97** (5), 133–143. doi:10.1111/avj.12810.

11. Newton J.R., Geraghty R.J., Castillo-Olivares J., Cardwell J.M. & Mumford J.A. (2004). – Evidence that use of an inactivated equine herpesvirus vaccine induces serum cytotoxicity affecting the equine arteritis virus neutralisation test. *Vaccine*, **22** (29–30), 4117–4123. doi:10.1016/j.vaccine.2004.02.052.

12. Kirkland P.D. & Frost M.J. (2020). – The impact of viral transport media on PCR assay results for the detection of nucleic acid from SARS-CoV-2. *Pathology*, **52** (7), 811–814. doi:10.1016/j.pathol.2020.09.013.

13. Quinet C., Czaplicki G., Dion E., Dal Pozzo F., Kurz A. & Saegerman C. (2016). – First results in the use of bovine ear notch tag for bovine viral diarrhoea virus detection and genetic analysis. *PLoS One,* **11** (10), e0164451. doi:10.1371/journal.pone.0164451.

OIE Preprint

14. Read A.J., Arzey K.E., Finlaison D.S., Gu X., Davis R.J., Ritchie L. & Kirkland P.D. (2012). – A prospective longitudinal study of naturally infected horses to evaluate the performance characteristics of rapid diagnostic tests for equine influenza virus. *Vet. Microbiol.*, **156** (3–4), 246–255. doi:10.1016/j.vetmic.2011.10.031.

15. Waugh C. & Clarke 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.A. Gardner & A. Colling, eds). *Rev. Sci. Tech. Off. Int. Epiz.*, **40** (1), XX–YY. doi:10.20506/rst.40.1.XXXX.

OIE Pre-print

Table I

Comparison of reactivity of an assay specific for border disease virus with two pan-reactive, pestivirus, real-time polymerase chain reaction assays

Sample	Antigen ELISA	Real-time RT−PCR (Ct)			Real-time RT−PCR (Ct)		
		Pan-reactive pestivirus	Border disease-specific	Difference between pan-pestivirus- & BDV-specific	Modified pan-reactive pestivirus	Border disease-specific	Difference between modified pan-pestivirus- & BDV- specific
1	Positive	33.43	23.93	9.50	23.74	21.19	2.55
2	Positive	34.24	22.57	11.67	24.86	21.79	3.07
3	Positive	31.22	20.29	10.93	25.97	21.81	4.16
4	Positive	36.21	25.59	10.62	27.13	23.21	3.92
5	Positive	35.19	24.95	10.24	25.85	23.26	2.59
6	Positive	38.09	27.46	10.63	27.32	23.53	3.79
7	Positive	Negative	26.82	13.18	26.89	23.78	3.11
8	Positive	Negative	28.47	11.53	27.12	23.79	3.33
9	Positive	32.16	24.81	7.35	25.42	23.84	1.58
10	Positive	Negative	31.52	8.48	27.57	23.95	3.62
			Mean difference	10.41		Mean difference	3.17

BDV: Border disease virus

ELISA: enzyme-linked immunosorbent assay

RT−PCR (Ct): reverse-transcription polymerase chain reaction (cycle threshold)

OIE Pre-print

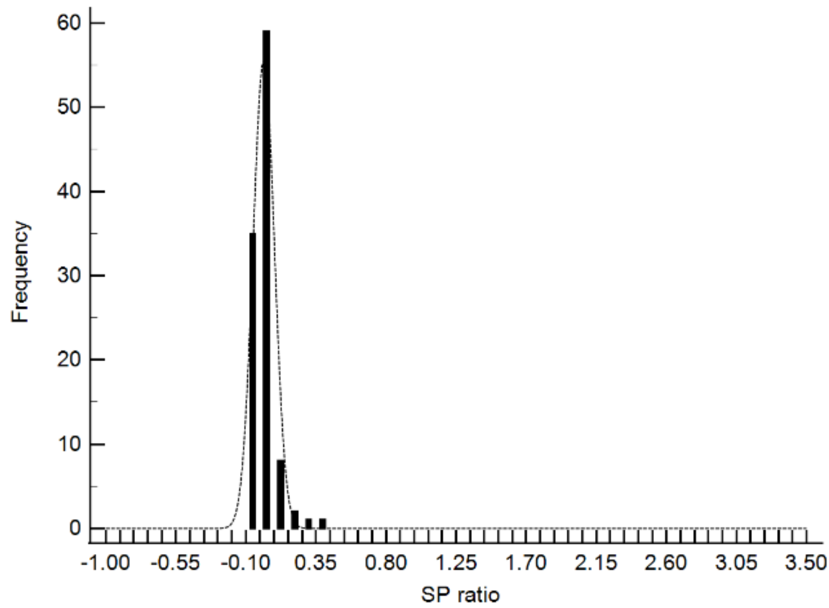


Fig. 1

A frequency distribution of the results, by enzyme-linked immunosorbent assay, of testing a population of sentinel chickens (population 1) for antibodies against avian metapneumovirus

Population 1 comprises sentinel chickens for quarantine for import purposes

SP ratio: mean sample-to-positive (S/P) ratio

OIE Pre-print

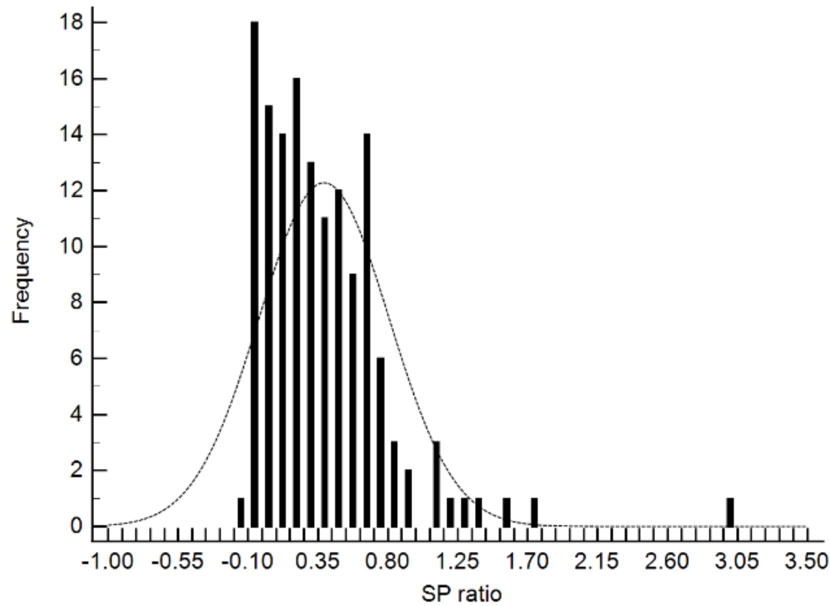


Fig. 2

A frequency distribution of the results, by enzyme-linked immunosorbent assay, of testing a population of free-range 'village' chickens (population 2) for antibodies against avian metapneumovirus

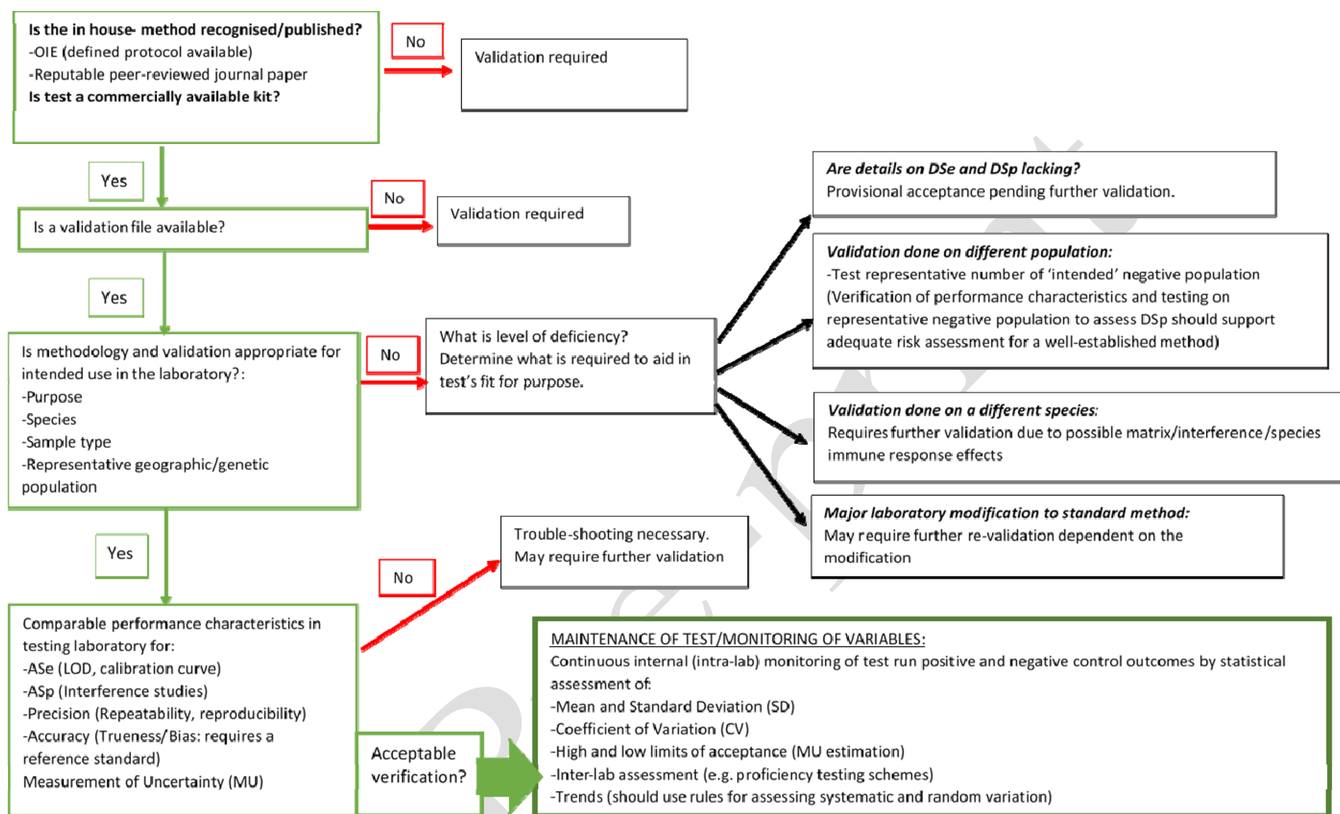
Population 2 comprises an unvaccinated avian-metapneumovirus-free population of free-range village chickens

SP ratio: mean sample-to-positive (S/P) ratio

OIE Pre-print

Appendix 1

Flow chart for assay verification testing



This chart was produced at the Australian Centre for Disease Preparedness in 2019 by K. Newberry, as part of a quality assurance verification template (initiated by R. Lunt in 2018). It is based on the World Organisation for Animal Health decision tree and ISO/IEC 17025:2017 (1)

ASe: analytical sensitivity

ASp: analytical specificity

DSe: diagnostic sensitivity

DSp: diagnostic specificity

LOD: limits of detection