Proficiency testing and ring trials

P. Johnson [(1)] & L. Cabuang [(2)]

(1**)** Animal and Plant Health Agency, Quality Assurance Unit, The Elms, Sutton Bonington, Leicestershire, LE12 5RB, UK

(2) External Quality Assessment Schemes, National Serology Reference Laboratory, Melbourne, Australia

*Corresponding author Paula Johnson:

Summary

The validation of diagnostic methods (and the subsequent results generated by a laboratory) are improved through participation in inter-laboratory comparisons (IC), such as proficiency-testing (PT) programmes and other exercises referred to as 'ring tests' or 'ring trials' (RTs). This is a requirement to comply with international quality standards. Validating a method is a continuous process and taking part in ongoing PT programmes supports the management of a method's life cycle, providing continuing assessment of fitness (sometimes referred to as the 'validation retention status'). Proficiency-testing panel designs ensure that the methods used, particularly diagnostic specificity and sensitivity, are suitably challenged. Appraising PT results over time can illustrate whether the laboratory's performance is stable, improving or worsening, and proficiency tests also highlight variations in the performance of assays.

The development of new proficiency tests can support the implementation of novel diagnostics technologies, such as whole genome sequencing and point-of-care testing, and assist in cross-sectoral partnerships focusing on One Health approaches, which are high on the agenda for infectious disease control.

For example, the rapid design and distribution of emergency exempted assays for severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) means that these assays were not as rigorously evaluated as assays for established infectious diseases. Therefore, participation in PT programmes for SARS-CoV-2 is essential to understand the performance of these assays.

While other mechanisms help to underpin laboratory activities, PT has been, and should remain, an integral part of laboratory quality assurance. Resources must be directed towards increasing and improving the quality of PT (for example, availability and accessibility of suitable biological and reference materials are essential for a PT provider to execute its duties), to support established and novel methods such as genomic and point-of-care tests.

Keywords

Method validation − Proficiency testing − Verification

Introduction

The validation of diagnostic tests will be strengthened through ongoing inter-laboratory comparisons and these, in turn, will form the basis for continuous monitoring of diagnostic test validity. International Standardization Organisation (ISO)/International Electrotechnical Commission (IEC) Standard 17025, 'General requirements for the competence of testing and calibration laboratories', specifies that a laboratory shall have quality control procedures for monitoring the validity of tests (1). In addition to a laboratory's internal quality control (IQC) procedures, this monitoring may include participation in inter-laboratory comparisons, proficiency testing (PT) or ring trials (RTs). Generally, PT programmes are often thought of as more formal: they are external, independent and ongoing to monitor performance at regular intervals over time. Ring trials tend to be infrequent exercises, sometimes used for test development and validation, and are usually reviewed independently as a discrete event. Inter-laboratory comparison, as the term implies, compares the results obtained by different laboratories on the same set of samples.

The nomenclature for PT and similar activities is not universally harmonised, and nor are their definitions. As shown in Table I, terms used to describe testing schemes are frequently interchanged (2). Irrespective of the nomenclature, PT is always a prescribed assessment of the competence of work by comparing it to either or both the results of other participants and reference standards. Standardisation of terms would require cooperation and agreement from the relevant official international organisations across all sectors to provide definitions of the stringency of different quality assurance activities. In this paper, the term 'PT' is generally used to cover all similar activities, as defined in Table I.

The PT process (including its use as part of a method validation procedure) is shown in Figure 1. Proficiency testing provides an independent, confidential and impartial verification of testing processes and should always be an honest representation of a laboratory's work. Proficiency testing is, however, only a snapshot of a single point in the testing framework. For this reason,

conclusions drawn from one PT exercise should not be taken as representative of a laboratory's overall capability. Not all proficiency tests assess the full testing process. Certain procedures before (pre-analytical) and after (post-analytical) the actual test measurement are not always applicable to PT samples. In these cases, laboratories should establish alternative procedures to evaluate any potential sources of error.

The standard to which PT providers must conform to be accredited is ISO/IEC 17043. Note that accreditation is applied to PT programmes rather than the entirety of the PT provider (2), while ISO standard 13528 (3) supports and is complementary to ISO/IEC 17043 for the statistical analysis of PT.

Participation in a PT programme is not a substitute for IQC and daily in-house quality assurance procedures, due to PT's intermittent nature. Nevertheless, PT has numerous diverse uses, as shown in Table II. Those directly relevant to the test method are especially important when PT is used to substantiate the validation of a method.

Developing and validating a method is a continuous process, involving testing suitable samples, data capture, analysis, modification of conditions and reassessment of fitness for purpose (4).

Method validation involves challenging the method to determine its performance capabilities and usefulness in the laboratory. Once these are established, there should be method life cycle management within the laboratory. This process has been referred to as 'the validation retention status' (4) and involves re-verifying the validated method to ensure that the test continues to function as originally intended (5). It is a continuous multi-component process accompanying IQC monitoring, and is normally less rigorous than the initial validation process. Participation in PT strengthens the validation retention status. The accumulation of data (and possibly cumulative statistical analysis) helps to monitor precision and accuracy (where appropriate), and verify the degree of ongoing diagnostic sensitivity (DSe) and diagnostic specificity (DSp).

Consequently, modifications can be made to methods on their subsequent revalidation, where necessary. Table III shows the supporting role of PT in method validation and the similarities between the two processes (6). Proficiency testing contributes data for a test validation dossier and can form part of a reproducibility assessment, but individual organisations should determine whether PT alone can replace them – for example, when few samples are available from naturally infected animals.

Methods

In order for a PT provider to evaluate participants' results, determine the 'true' positive and set evaluation criteria, they should be aware of the design and capabilities of the method types used by participants. Normally, PT involves laboratories performing the same test on the same samples and comparing the results.

There are, however, exceptions to this in which proficiency tests consist of two or more method types. The test(s) used to verify the samples before issue must be specified by the PT provider and it would not be appropriate to use the least sensitive method to apply the assigned values. The values assigned to PT samples are likely to depend on the method used, and it is not uncommon for these methods to give different results.

Proficiency-testing results are usually grouped by method, enabling intra-method comparisons to be made. It is important that the participants consider the uses of the test and the actions taken, based on the results generated by the methods. For example, the importance of a result might vary, depending on whether the test is used for international trade purposes or as a screening test.

Furthermore, results may also differ between the assays used for a single type of method. This is where an independent PT provider (for example, one that does not have reference laboratory status) must exercise caution, as there may be implications if they are seen to either endorse or devalue one assay in regard to another. It is not the PT provider's role to 'police' a laboratory's activities, but rather to provide them with the tools to make informed decisions about their testing practices.

Sample matrix and panel composition

A 'matrix', in this context, consists of the components of a PT sample other than the target analyte. They may be simple (serum or milk) or complex (faeces or tissue) and can potentially affect the method or the analyte. Components of the matrix can affect the procedure's ability to accurately measure an analyte, causing interference or cross-reactivity.

An example is heat-treating serum over 60°C, which can denature the analyte of interest (antibodies). A further example occurred in a bovine viral diarrhoea virus (BVDV) polymerase chain reaction (PCR) proficiency test, in which the sample matrix was heparinised blood (contained anticoagulant, used to help maintain sample integrity). This can be an inhibitor that adversely affects the efficacy of RNA extraction or the real-time reverse-transcription PCR assay, causing false-negative results (7). This was noticed across all dilutions of the positive samples for BVD antigen (Ag) in the panel. In such cases, a serum is more appropriate for those PCR platforms

in order to successfully detect BVD Ag. These cases demonstrate the importance of internal controls to check extraction and amplification processes.

The matrix of PT samples can also influence whether a testing process is entirely or partially challenged. In a molecular method, a bacterial isolate will challenge more of the testing process than providing a DNA extract.

The configuration of PT panels must be blinded to prevent collusion and falsification (except where specific samples might be used as part of a test validation procedure). A survey with 75 respondents revealed that the suitability of the PT samples was rated as the second-highest criterion for laboratories when selecting a PT programme. A carefully selected panel of samples that is fit for purpose can provide the necessary challenges, despite variability in the design and performance of assays. Increasing the number of sera samples in a panel from animals of known infection status can reduce random errors in the estimates of DSe and DSp (8). A panel consisting of a limited range of 'mock' or spiked samples, as opposed to 'real' representative diagnostic samples, may reduce the ability to evaluate the DSe and DSp of an assay. It is here that a laboratory must decide whether the proficiency test's limitations in classifying an intended weak positive are acceptable. Otherwise, misguided conclusions may be drawn on its performance by the test commentator and participants.

Results from PT samples should reflect clinical samples taken at any stage of an immune response, allowing for diversity in testing. For example, serological tests, such as the enzyme-linked immunosorbent assay (ELISA), complement fixation test (CFT) and agar gel immunodiffusion (AGID), used to detect bovine herpesvirus-1 (BHV-1), should detect all latently affected cattle (9), as required for international trade activities. However, low antibody concentrations are rarely seen when monitoring poultry flocks for certain diseases post-vaccination. Here, the sensitivity of the method is not of primary importance in PT, but this is not the case when validating a method before implementation.

In BHV-1, if the PT panel excludes samples that approach the method's limit of detection, the PT will only demonstrate reproducibility at the extremes of analyte concentration or reactivity. This will compromise the results reported for the target population and thus the correct classification of the infection status of the animal (8). Therefore, performance around the cut-off value or limit of detection is essential to challenge a method and for method validation. Proficiency-testing panels could include a 'reference sample', i.e. a secondary material calibrated against an international reference preparation or a well-characterised sample known to demonstrate consistently 'weak' reactivity over various methods.

There is encouragement from some accreditation bodies for PT providers to issue PT samples that challenge a test method's limit of detection. It is essential that, where a participant reports a negative result for a low-positive PT sample (with a value below a reference standard), the participant must understand the limitations for the test method in use, and consider the impact of reporting a false negative for a weak PT sample. For example, is the laboratory located in a region where the target disease is endemic or exotic? Is the laboratory conducting tests for screening and surveillance purposes or is it running confirmatory testing? The magnitude of DSe is of greater concern when a participant finds the stronger positive samples to be negative.

Variations in the performance of assays have been highlighted by VETQAS® proficiency tests. One such case, involving one PT programme, is illustrated in Figure 2. The PT covered three method types. For one of the method types, all participants used an assay from the same manufacturer. Two distributions of samples were issued about three months apart, and a shift in the DSe of the assay was highlighted. This coincided with participants moving over to a new assay batch. The results for the other method types were as expected, with values remaining consistent with those reported in previous distributions. The intra-methodology comparisons, along with the known status of the samples and the use of serial dilutions of the sera, supported the anomalous results. A new batch had been released but it was not expected that the change would affect the DSe of the assay. This highlights the need for laboratories to be made aware of manufacturers' changes in assay design so that they can carry out further performance verification assessments or validation studies to assess performance.

Table IV provides examples of sample types that may represent a population of animals from which samples are routinely taken. Also shown is the impact of these sample types on challenging diagnostic assays.

Statistical analysis

International Standardization Organization Standard 13528 describes the procedures used to evaluate PT data sets and make objective judgements to assess the fitness of laboratory tests (3). These laboratory tests may be established independently of participant results or, in most cases, derived from participant results. To gain maximum benefit from the PT, participants must understand:

a) the statistical model used for scoring results

b) the criteria for acceptable limits

c) the evaluations made by the PT provider.

This helps to avoid misinterpretation of results and, consequently, a false impression of performance.

Scoring protocols can differ between PT providers, particularly for interpretive proficiency tests. Many proficiency tests convert participants' results into a z-score, an applied statistical measurement of a result's association to the assigned value, which can be used to identify outliers from the data set.

A z-score of ≤2 means the result is satisfactory. A z-score of >2 and <3 means the results are questionable and so testing will need monitoring. A z-score > 3 means the result is unsatisfactory and an investigation to identify the cause is necessary.

Outliers, and the reasons for them, are variable, causing difficulties when producing meaningful statistics. Where participant numbers are low, the statistical evaluation of results is more uncertain, making it difficult to identify the distribution of the results and detect outliers reliably. The inclusion of a reference material in the panel will help to give confidence in the accuracy of the results, as will making comparisons with the results of samples tested using a 'gold standard' method. Document EA4-21 provides advice about proficiency tests with small numbers of participating laboratories (10).

A Youden plot (or Youden analysis) is a useful and effective means to make comparisons and indicate whether any deviating results in a PT exercise are caused by systematic error (bias) and/or by relatively large random errors. This can be achieved with little effort by the inclusion, for each parameter, of two PT samples with similar or identical measurands.

Appraising the results from each single distribution and the cumulative results over time, despite some limitations, can help to establish whether incorrect results can be attributed to imprecision, systematic error or random error. Provided that the data from subsequent rounds are comparable (e.g. the same standard deviation is used to generate a score), such an appraisal illustrates whether the performance is stable, improving or worsening, as well as the stability of the analytical test (11). Comparing results between participants also helps to gauge whether any unsatisfactory result is an isolated incident or if multiple unsatisfactory results were caused by a problem with the measurement process or by a PT sample.

Irrespective of how scores are derived, they should be put into perspective by participating laboratories, handled according to their quality systems and any diagnostic treatments applied.

Improvements through participation in proficiency testing

The overall performance of laboratories has been shown to improve with increased participation (12). Where there are known testing problems, small-scale exercises can be tailored and issued to a group of laboratories, either for performance monitoring or method validation. It is in such cases that 'bespoke' or 'closed' proficiency tests have been instrumental in driving the improvement in a laboratory's performance. Figure 3 demonstrates how two different countries have, with regular participation in slightly different *Salmonella* proficiency tests, improved their performance.

A small-scale, bespoke proficiency test was issued by VETQAS® to a single country where there were some suspected issues with a well-established test method. The results of the first proficiency test confirmed this, with 73% of the results correct. A collaborative investigation between laboratories revealed the use of incorrect methods and difficulties with interpreting the test. Remedial measures including training were implemented across the network. Subsequent proficiency tests were then issued to the same laboratories, with correct results of 93% and 99% reported.

A further example demonstrates that PT is not just an educational accomplishment for participating laboratories, but also provides noteworthy method comparison data.

Taylorella equigenitalis (*TE*) is the causative agent of contagious equine metritis (CEM), a sexually transmitted disease of horses and donkeys. Notifiable to the OIE, it has been identified in many countries. In 1997/98, a new, closely related organism, *Taylorella asinigenitalis* (*TA*) was identified from animals that showed no clinical signs of disease. This species is difficult to differentiate from *TE* and there are reports of *TA* being incorrectly identified as *TE* (13).

The inclusion of a *TA* isolate in the VETQAS® PT panel was suggested by the OIE CEM Reference Laboratory to raise awareness of this emerging potential pathogen. Fifty-nine laboratories participated in the PT, using either or both the culture method and PCR. Fourteen (~24%) laboratories reported false positives, misidentifying the *TA* isolate as *TE* (the target pathogen). Of the 14 laboratories, 12 used the culture method. The other two laboratories reporting false positives did not state the method used. An evaluation of the performance in the proficiency test included recommendations on further confirmatory tests post-culture, to help distinguish between *TA* and *TE*.

Laboratories participating in a subsequent proficiency test recorded 100% correct results, again using either the culture method or PCR or both. The proficiency test raised awareness of this emerging organism, and lessons were learned and acted upon by participants. Furthermore, the

conclusions derived from the proficiency tests supported the advantages of the PCR test for CEM organisms, as noted in Chapter 3.5.2. of the OIE *Manual of Diagnostic Tests and Vaccines for Terrestrial Animals* (14).

Challenges for proficiency testing

For some animal diseases, the availability of negative and positive biological material for use by PT providers is limited. The exchange of this material is often hampered by country-specific regulations and complex logistics. The risks of biosecurity and bioterrorism must be managed. However, the introduction of a global tariff code for international PT shipments, including import permits, could help, although this requires cooperation from all relevant authorities. Furthermore, a shortage of reference materials and lack of harmonised methods presents the PT provider with difficulties for the provision and design of proficiency tests.

Predicting the next animal disease, outbreak or pandemic is not easy. So how does a PT provider react to a novel disease and how does an outbreak influence demand for the next proficiency test? This can be illustrated by the current severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) pandemic. While PT programmes are also a necessity in medical diagnostics, it is helpful to understand the differences between veterinary and medical diagnostics.

Medical diagnostic laboratories accredited under ISO 15189 must also participate in PT programmes to maintain accreditation, with PT providers being accredited under ISO 17043 (2, 15). The aims for PT in medical and veterinary diagnostics do not differ greatly. The testing environments differ, as all methods used to obtain results that are then reported to clinicians must be accredited with the relevant regulators for medical devices. This also applies to methods designed in-house.

Therefore, participants select PT programmes that best reflect their reporting processes. The PT programmes need to be generalised enough that a single programme can accommodate the evaluation of testing algorithms, different test methods and analytes.

All participants subscribed to the programme presented in Table V would be evaluated for submitted patient status, irrespective of the methods used. Method-specific analysis would be performed at the analyte level, grouping participants' testing in a specific method for analysis in their 'peer group'. Programmes are designed to review aims, much like those programmes used in veterinary diagnostics (reproducibility, sample carry over).

Medical PT programmes are frequently designed with multiple organisms that cause similar symptoms in patients. (Although this is not conventional in veterinary PT, panels for some swine

diseases and crustacean diseases are available in which laboratories follow a diagnostic algorithm to either 'rule in' or 'rule out' diseases. The same applies to panels used in veterinary microbiological PT programmes.)

Since many respiratory viruses manifest symptoms like those caused by SARS-CoV-2, diagnosis of COVID-19 was made by exclusion before the development of SARS-CoV-2-specific assays. The global cooperation of laboratories to share their knowledge and materials has led to the rapid design and distribution of emergency exempted assays for SARS-CoV-2, as well as access to COVID-19 PT programmes. However, the nature of the current SARS-CoV-2 pandemic means that these assays have not been as rigorously evaluated as assays for established infectious diseases, such as influenza, considering the regulatory requirements for diagnostic tests. Therefore, participation in PT programmes for SARS-CoV-2 contributes to our understanding of assay performance (for both commercially available and in-house-designed assays), while meeting accreditation requirements. In future, SARS-CoV-2 could be incorporated into PT programmes for other respiratory viruses, especially in the manufacture of multiplex assays designed to speed up differential diagnoses.

A One Health approach could improve the gains made in SARS-CoV-2 medical diagnostics, perhaps through the use of similar assays being used on sentinel animals. This could best be applied, in the first instance, for influenza viruses. Design considerations for different sample types (human versus swine versus avian influenza), and new technologies, such as molecular typing and ways of determining drug resistance, could help to identify strains with pandemic potential. Laboratory preparedness would improve as veterinary and medical diagnostic laboratories with similar capabilities could be deployed to meet surge capacity in an outbreak. Cross-sectoral partnerships that focus on fostering One Health concepts for PT, reference materials, and the quality and availability of demographic data are high priorities for infectious disease control.

Future directions for proficiency testing

Future directions for PT include supporting new technologies, such as molecular typing, tackling drug resistance, and anticipating emerging diseases. Whole genome sequencing (WGS) techniques are becoming the methods of choice for identifying and characterising pathogens (e.g. typing, virulence and antimicrobial resistance) to support disease outbreak investigations, surveillance and research. There are three distinct components within a testing process that uses WGS: the wet processes of library preparation and sequencing; and the bioinformatics pipeline using *in silico* methods for WGS data analysis (the dry bench process). Each stage can be assessed independently using PT but the whole testing process must function correctly to provide a valid result. Moreover,

the diversity between laboratory processes for each stage makes methods-based PT more suited to the *in silico* stage of the assays than the traditional analyte-specific PT programmes (16).

Zamperin *et al.* (17) conducted a proficiency test for the application of WGS in a veterinary virology context, using real data to evaluate the bioinformatics pipeline for consensus sequence generation (irrespective of the range of techniques used for library production and sequencing). Results showed that the analytical process of the bioinformatics pipeline had a critical effect on assay reproducibility. Proficiency testing is key in comparing and assessing sequence data using a variety of methods, with harmonisation as its goal.

Proficiency tests for such technology in the veterinary sector are in a state of infancy. The complexities of handling large amounts of data, coupled with a variety of analytical methodologies, cause difficulties in agreeing how best PT can be delivered. However, this is a continually evolving process. From a One Health perspective, there is much focus on developing PT exercises in an attempt to collate information and share metadata to aid the standardisation of 'wet' laboratory and bioinformatics analyses, so that laboratories within all sectors can achieve comparable characterisation of isolates.

Point-of-care tests (POCTs) face the same pre-analytical, analytical and post-analytical concerns as any other tests and so proficiency tests that focus on the whole testing process to check the validation retention status are necessary. The requirement for PT and IQC is actually enhanced by the simplicity of the technology; however, PT for POCTs is challenging for PT providers for several reasons, including but not limited to:

– the fact that PT assesses only a single POCT device (within-batch variation is possible);

– the end-users of the tests are the participants and they may not understand PT. Hence, the resulting data must be presented in a useful and actionable format;

– the scarcity of stable, commutable samples and reference materials;

– peer laboratory target values are yet to be established.

Other issues have been discussed elsewhere (18). Consequently, there are few PT programmes and, in their absence, POCT results are generally assessed by comparing the results from a patient sample analysed using a POCT device and by a conventional laboratory method. Infectious diseases that affect livestock and companion animals and zoonoses are high-value targets as veterinary POCT diagnostics continue to expand. For this reason, more PT programmes aimed at veterinary POCTs are needed (19).

Conclusions

A PT programme should assess a laboratory's testing capability and provide assurance that the performance characteristics of an assay are consistently maintained, i.e. the validation retention status. The information about a PT programme must be well defined so that a laboratory can select a PT programme that is fit for purpose. The clinical usefulness of a test (i.e. whether the information gained from the result may determine the effective treatment or preventative strategies implemented for a patient) (20) may have some bearing on the selection of PT. Accreditation bodies and websites such as the European Proficiency Testing Information System (EPTIS) (www.eptis.bam.de), an international database of PT schemes, provide details of PT programmes.

The European Accreditation paper EA-4/18 INF: 'Guidance on the level and frequency of proficiency testing participation' suggests that laboratories should take a risk-based approach (21). For example, when a laboratory might be bound by financial constraints but has several test methods, one proficiency test for each type of method can generally be justified to an accreditation body. Selecting a PT scheme that is compliant with ISO/IEC 17043 will satisfy some critical points in the selection process (2). Supporting this are the results of a VETQAS® survey, indicating that the majority of respondents rated the use of PT providers that offer accredited PT schemes as the most important factor when selecting a proficiency-testing programme. Indeed, a laboratory may have to explain a decision to use a non-accredited PT scheme to its accreditation body. Where a suitable PT scheme is not available, the accreditation body and laboratory should agree on a suitable alternative means by which performance can be assessed.

A laboratory should be mindful of the limitations as well as the benefits of PT, as illustrated in Box 1 (22). The quality of the laboratory testing process is generally judged by its performance when testing the sample (i.e. the testing phase). However, when selecting a PT scheme, a laboratory should take account of the pre-analytical, analytical, and post-analytical phases of the testing process. A medical laboratory in Italy reported that 62% of its testing errors were due to events in the pre-analytical phase, before the specimen reached the laboratory bench, and a further 23% occurred after testing was completed (23).

Proficiency-testing schemes with a small number of participants, using less common methods, are deemed inappropriate for meaningful comparisons. More information about this can be accessed in Document EA4-21 (10). In addition, there is likely to be greater variation in the data when the PT is composed of more than one test method for the same measurand, and is therefore less likely to detect diagnostically relevant errors.

Irrespective of how a laboratory selects an appropriate PT programme, the laboratory must formulate, document and regularly review a plan for its PT activities.

The challenges for PT providers and laboratories will remain significant, with the focus set to continue on the standardisation of methods (including those for WGS and POCTs) and the sourcing of biological material for the development of PT samples and reference materials. The cross-sectoral One Health approach of sharing information, particularly on PT, will benefit all. As evident from the lessons learned during the rapid development of SARS-CoV-2 assays, the educational value of PT should not be underestimated.

Acknowledgements

Thanks are due to Dr S.L.R. Ellison of LGC, the United Kingdom, for his kind permission to use Table III in this paper.

Essais d'aptitude et essais comparatifs inter-laboratoires

P. Johnson & L. Cabuang

Résumé

Les procédures de validation des méthodes de diagnostic (et les résultats obtenus par la suite par les laboratoires) peuvent être améliorées en participant à des comparaisons inter-laboratoires, sous forme notamment de programmes d'aptitude inter-laboratoires ou d'autres exercices désignés collectivement sous l'appellation d'essais comparatifs inter-laboratoires (*ring trials* en anglais). Cette participation constitue une obligation au regard de la conformité avec les normes internationales de qualité. La validation d'une méthode est un processus continu et la participation à des programmes d'aptitude inter-laboratoires offre des garanties quant à la gestion du cycle de vie de la méthode considérée, car elle se traduit par une évaluation continue de l'aptitude à l'emploi du test (également désignée comme la « conservation de son statut de validation »). Les panels des essais d'aptitude sont conçus de manière à garantir que les méthodes utilisées font l'objet d'essais appropriés, en particulier concernant leur spécificité et sensibilité diagnostiques. L'appréciation des résultats des essais d'aptitude dans le temps permet d'établir si les performances d'un laboratoire restent stables, s'améliorent ou se dégradent ; les essais d'aptitude mettent également en lumière les éventuelles variations dans les performances d'un essai.

La mise au point de nouveaux essais d'aptitude peut encourager l'utilisation de technologies de diagnostic innovantes, par exemple le séquençage du génome entier et les tests utilisables sur le

lieu des soins, et faciliter les partenariats intersectoriels axés sur des approches Une seule santé, qui figurent parmi les grandes priorités en matière de lutte contre les maladies infectieuses.

Par exemple, la conception et la distribution rapides de tests de détection du coronavirus 2 du syndrome respiratoire aigu sévère (SARS-CoV-2) suivant un protocole d'exception imposé par l'urgence signifient que ces tests n'ont pas fait l'objet d'une évaluation aussi rigoureuse que les tests de détection de maladies infectieuses bien établies. Par conséquent, la participation à des programmes d'essais d'aptitude pour le SARS-CoV-2 est essentielle pour déterminer précisément les performances de ces tests.

S'il existe d'autres mécanismes permettent d'étayer les activités d'un laboratoire, les essais d'aptitude ont été et doivent continuer à être une partie intégrante de l'assurance qualité des laboratoires. Il convient d'affecter des ressources à l'accroissement du nombre d'essais d'aptitude et à l'amélioration de leur qualité (il est notamment essentiel que les fournisseurs de tests d'aptitude puissent se procurer facilement des matériels biologiques et réactifs de référence appropriés afin de mener à bien leur tâche), de manière à soutenir aussi bien les méthodes classiques que celles qui procèdent d'innovations comme les tests génomiques et ceux utilisables sur le lieu des soins.

Mots-clés

Test d'aptitude – Validation des méthodes de diagnostic – Vérification.

Pruebas de competencia y pruebas interlaboratorios

P. Johnson & L. Cabuang

Resumen

La validación de métodos de diagnóstico (y de los subsiguientes resultados obtenidos por un laboratorio) mejora con la participación en procesos de comparación entre laboratorios, como pueden ser los programas de pruebas de competencia u otros procesos denominados globalmente "pruebas interlaboratorios". Se trata de un requisito de obligado cumplimiento según dictan las normas internacionales de calidad. La validación de un método es un proceso permanente y, en este sentido, el hecho de tomar parte en programas continuos de pruebas de competencia ayuda a gestionar un método de prueba durante todo su ciclo de vida, aportando en todo momento una evaluación de su nivel de idoneidad (lo que a veces se denomina también el "estado de retención de la validación"). El diseño de paneles de pruebas de competencia sirve para garantizar que los

métodos empleados, y en particular su especificidad y sensibilidad de diagnóstico, sean convenientemente evaluados. El análisis de los resultados de pruebas de competencia a lo largo del tiempo puede indicar si el laboratorio se mantiene en niveles estables de rendimiento o si este mejora o empeora. Las pruebas de competencia también pueden poner de manifiesto variaciones en el rendimiento de un ensayo.

La creación de nuevas pruebas de competencia puede contribuir a la implantación de novedosas tecnologías de diagnóstico, como la secuenciación de genoma completo o las pruebas practicadas en el punto de consulta, y ayudar a trabajar en alianzas intersectoriales desde planteamientos en clave de "Una sola salud", extremo este de lo más prioritario en los planes de lucha contra las enfermedades infecciosas.

Valga como ejemplo la celeridad con que se han concebido y distribuido ensayos aplicables al coronavirus del síndrome respiratorio agudo severo de tipo 2 (SARS-CoV-2), con las exenciones propias de un procedimiento de urgencia, lo que supone que estos ensayos no hayan pasado por un proceso de evaluación tan riguroso como el que se aplica a patógenos infecciosos más antiguos. Por ello, en el caso concreto del SARS-CoV-2, la participación en programas de pruebas de competencia es fundamental para aprehender el rendimiento que ofrecen estos ensayos.

Si bien hay otros mecanismos que ayudan a reforzar el trabajo de laboratorio, las pruebas de competencia han sido, y deben seguir siendo, parte integrante de la garantía de calidad que ofrece un laboratorio. Para potenciar el uso de métodos ya arraigados o novedosos, como puedan ser la genómica o las pruebas en el punto de consulta, es preciso dedicar recursos a la realización de un mayor número de pruebas de competencia de mejor calidad (la existencia de material biológico y de referencia adecuado y la facilidad de acceso a él, por ejemplo, son sendos factores básicos para que un proveedor de pruebas de competencia pueda cumplir su cometido).

Palabras clave

Pruebas de competencia − Validación de métodos – Verificación.

References

1. International Organization for Standardization/International Electrotechnical Commission (ISO/IEC) (2017). – ISO/IEC 17025: General requirements for the competence of testing and

calibration laboratories. ISO, Geneva, Switzerland, 30 pp. Available at: http://imed.ir/userfiles/files/11/ISO-IEC%2017025-2017.pdf (accessed on 7 January 2021).

2. International Organization for Standardization/International Electrotechnical Commission (ISO/IEC) (2010). – ISO/IEC 17043: Conformity assessment − General requirements for proficiency testing. ISO, Geneva, Switzerland, 39 pp. Available at: www.iso.org/standard/29366.html (accessed on 6 March 2021).

3. International Organization for Standardization (ISO) (2015). – ISO 13528: Statistical methods for use in proficiency testing by interlaboratory comparison. ISO, Geneva, Switzerland, 89 pp. Available at: www.iso.org/standard/56125.html (accessed on 6 March 2021).

4. 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.

5. Public Health England (2017). – UK standards for microbiology investigations Q1: evaluations, validations and verifications of diagnostic tests. Public Health England, London, United Kingdom, 45 pp. Available at: www.gov.uk/government/publications/smi-q-1-commercial-and-in-house-diagnostic-tests-evaluations-and-validations (accessed on 15 November 2020).

6. Ellison S.L.R. (2016). – The role of PT in method validation. An adaptation from the presentation: Trends in inter-laboratory method validation, presented at the Eurachem International Conference, 9−10 May, Ghent, Brussels.

7. World Organisation for Animal Health (OIE) (2019). – Bovine viral diarrhoea. Chapter 3.4.7. *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/3.04.07_BVD.pdf (accessed on 30 January 2021).

8. 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, 8th Ed. OIE, Paris, France. Available at: www.oie.int/index.php?id=2439&L=0&htmfile=chapitre_validation_diagnostics_assays.htm (accessed on 24 February 2021).

9. Perrin B., Calvo T., Cordioli P., Coudert M., Edwards S., Eloit M., Guerin B., Kramps J.A., Lenihan P., Paschaleri E., Perrin M., Schon J. & Van Oirschot (1994). –Selection of European

Union standard reference sera for use in the seriological diagnosis of infectious bovine rhinotracheitis. *Rev. Sci. Tech. Off. Int. Epiz.,* **13** (3), 947–960. doi:10.20506/rst.13.3.810.

10. European Accreditation–Eurolab–Eurachem Proficiency Testing Working Group (EEE–PT) (2018). – EA-4/21: Guidelines for the assessment of the appropriateness of small interlaboratory comparisons within the process of laboratory accreditation. European Accreditation, Paris, France, 12 pp. Available at: https://european-accreditation.org/wp-content/uploads/2018/10/ea-4-21-inf-rev00-march-18.pdf (accessed on 15 November 2020).

11. Mann I. & Brookman B. (2011). – Selection, use and interpretation of proficiency testing (PT) schemes, 2nd Ed. Eurachem, London, United Kingdom, 52 pp. Available at: www.eurachem.org/images/stories/Guides/pdf/Eurachem_PT_Guide_2011.pdf (accessed on 12 October 2020).

12. Pitkälä A., Gindonis V., Wallin H. & Honkanen-Buzalski T. (2005). – Interlaboratory proficiency testing as a tool for improving performance in laboratories diagnosing bovine mastitis. *J. Dairy Sci*., **88** (2), 553–559. Available at: www.sciencedirect.com/science/article/pii/S002203020572717X (accessed on 15 November 2020).

13. Herbert L., Moumen B., Pons N., Duquesne F., Breuil M.F., Goux D., Batto J.M., Laugier C., Renault P. & Petry S. (2012). − Genomic characterization of the *Taylorella* genus. *Plos One,* **7** (1), e29953. doi:10.1371/journal.pone.0029953.

14. World Organisation for Animal Health (OIE) (2019). – Contagious equine metritis. Chapter 3.5.2. *In* Manual of Diagnostic Tests and Vaccines for Terrestrial Animals, 28th Ed. OIE, Paris, France. Available at: www.oie.int/fileadmin/Home/eng/Health_standards/tahm/3.05.02_CEM.pdf (accessed on 16 January 2021).

15. International Organization for Standardization (ISO) (2012). – ISO 15189: Medical laboratories — requirements for quality and competence. ISO, Geneva, Switzerland, 89 pp. Available at: www.iso.org/standard/56115.html (accessed on 7 March 2021).

16. Duncavage E.J., Abel H.J. & Pfeifer J.D. (2017). – *In silico* proficiency testing for clinical next-generation sequencing. *J. Molec. Diagn*., **19** (1), 35–42. doi:10.1016/j.jmoldx.2016.09.005.

17. Zamperin G., Lucas P., Cano I., Ryder D., Abbadi M., Stone D., Cuenca A., Vigouroux E., Blanchard Y. & Panzarin V. (2019). – Sequencing of animal viruses: quality data assurance for NGS bioinformatics. *Virol. J.,* **16**, 140. doi:10.1186/s12985-019-1223-8.

18. Stavelin A. & Sandberg S. (2017). – Essential aspects of external quality assurance for point-of-care testing. *Biochem. Med. (Zagreb)*, **27** (1), 81–85. doi:10.11613/BM.2017.010.

19. Cummins B.M., Ligler F.S. & Walker G.M. (2016). – Point-of-care diagnostics for niche applications. *Biotechnol. Adv*., **34** (3), 161–176. doi:10.1016/j.biotechadv.2016.01.005.

20. Lesko L., Zineh I. & Huang S.-M. (2010). – What is clinical utility and why should we care? *Clin. Pharmacol. Therapeut.,* **88** (6), 729–733. doi:10.1038/clpt.2010.229.

21. European Accreditation–Eurolab–Eurochem Proficiency Testing Working Group (EEE–PT) (2010). – EA-4/18: Guidance on the level and frequency of proficiency testing participation. European Accreditation, Paris, France, 15 pp. Available at: https://european-accreditation.org/wp-content/uploads/2018/10/ea-4-18-inf-rev00-june-2010-rev.pdf (accessed on 18 February 2021).

22. European Commission (EC) (2001). – Regulation (EC) No 999/2001 of the European Parliament and of the Council of 22 May 2001 laying down rules for the prevention, control and eradication of certain transmissible spongiform encephalopathies. *Off. J. Eur. Union,* **L 147**, 31/05/2001, 1–40. Available at: https://eur-lex.europa.eu/eli/reg/2001/999/oj (accessed on 24 February 2021).

23. De la Salle B., Marrington R. & Mackenzie F. (2019). – Developing pre- and post-analytical error monitoring in laboratory medicine. *Clin. Chim. Acta,* **493** (Suppl.), S526–S527. doi:10.1016/j.cca.2019.03.1108.

_____

Table I

An explanation of the more common proficiency-training schemes of a similar nature, which have varying nomenclature

Scheme	Utility	Organisation	Strengths	Limitations
Internal Quality Assura	Routine inter-assay assessment performance	Internal provider	Cost-effective routine test performance evalu If quantitative, can be used to e measurement uncertainty	Data limited to in-house Not always blind
External Quality Assur	Routine external inter-as assessment of test perfo	External provider or interna different laboratory in the s	External/ internal assessment o performance Routine test performance evalu	Target values may be knov May or may not be blind Cost
Ring trial (RT)	Inter-laboratory assessm used for the developmer or unusual samples	External provider	Provides comparison between laboratories May or may not be blind	Not always compliant with Often treated as a one-off
Proficiency testing (PT	Ongoing periodic assess laboratory test performa other participants	External provider	Often run according to ISO/IEC Enables comparison of data be participants External feedback on performa Samples with undisclosed cont	Probably the most costly o Limited and set number of per year

Note: there are other similar 'local or regional' adaptations of the example schemes shown in the table. In Australia, the Laboratories for Emerging Animal Disease Diagnosis and Response produce national quality control standards (NQCs). These standards operate alongside Australia's external quality assurance and are a hybrid between internal and external quality assurance, using relevant, well-characterised field samples to monitor ongoing assay performance in state laboratories

Table II

The diverse uses of proficiency testing

Uses of proficiency testing				
Examines and compares the perfor laboratories using the same method	Provides evidence of su participation	Evaluates performance of diagnostic methods	Method comparison studie proficiency testing consists one method)	Facilitates harmonisation, sta and rationalisation of test me
Helps to assess reproducibility and of a test	Checks the accuracy of (where reference sample	Highlights bias in a labor testing data	Provides confirmation of a uncertainty budget	Supports validation of diagnc
Provides assurance of method vali retention status	Characterises reference	Is a valuable laboratory risk management tool	Tests transfers between laboratories	Meets mandatory requireme trade agreements

Table III

The supporting role of proficiency testing when used in method validation (6)

	Validation study	**Proficiency testing to assess validatio status**
Sample characteristics	Stable, homogeneous samples distributed laboratories Replicate results returned to organiser	Stable, homogeneous samples distributed laboratories Single or replicate results returned to orga
Method diversity	Laboratories undertake replicate analysis u method	Laboratories undertake replicate analysis u method
Organiser	Can be 'internal' to the testing laboratory The organiser estimates repeatability and reproducibility	Usually external and independent of the te laboratory Organiser assigns value and 'scores' the la
Advantages	More controlled & easily performed	Comparisons with many methods for the s measurand Diversity in participating laboratories
Disadvantages	Fewer laboratories participate (either one la a network) No comparison with other methods	Typically provides single values (no replica Few samples per round Methodology detail not always collected

Adapted with the kind permission of S.L.R. Ellison, LGC, United Kingdom

Table IV
Representation of a selection of 'ideal' proficiency testing samples that may be included in a panel and the potential impact of their inclusion

Sample type	Impact of use
Samples from vaccinated animals and naturally infe	Challenges assays to determine whether they are able to differentiate between infe vaccinated animals
Samples from both specific-pathogen-free animals, known inoculum of specific concentration, taken at points during the course of infection, and field samp generally full of mixed populations. Cross-reactivity to be considered	Enables estimation of specificity (before inoculation) and sensitivity (after inoculatic Specificity may be overestimated due to the absence of potential cross-reacting pa Provides authentic reference samples but they are derived from experimental studi certain factors, such as the pathogen used, dose and route of inoculation and expo not be representative of those samples found in the field. Field samples from natur or non-infected animals are useful for realistic estimates of diagnostic sensitivity ar especially if they represent the target population to which the assay will be applied. Determination of true infection status may be challenging
Samples that contain similarities in their componen which potential cross- reactions may occur	Challenges assay specificity
Samples where the analyte concentration is close t	Challenges assay sensitivity
Pooled samples	Used when samples are scarce or screening tests are performed. In such cases, it to know the maximum number of samples that can be pooled, in which the test car the target analyte if just one animal is positive
A linear panel of samples using dilutions made dire same high-titre sample	Assesses the linearity of assay response versus dilution. Helps to highlight dilution allows the evaluation of accuracy and within method precision
Uses odd dilution factors	Generates a less predictable panel
Uses several dilutions of the same serum	Gives an indication of a laboratory's repeatability and if there is any change in the d sensitivity of the test
Replicates in a proficiency-testing panel	Checks intra-laboratory precision and any change in diagnostic sensitivity. In some using panels with an adequate number of weak positives and negatives (at a 30:70 to avoid bias, whilst maintaining a laboratory's ability to distinguish a weak-positive from no reaction

Panels with a small number of test items, compose or all negatives and one positive	Also helps to remove bias, but may not enable the participants to distinguish a wea result from a negative result. Can also be used to check if there is variation in the s
Using multiple samples from the same batch in a pa using samples from the same batch again in subse panels)	Enables monitoring of test performance over time to challenge the repeatability and reproducibility of a laboratory and analytical test methods
Reference strains & field strains	For microbiological reference cultures, field stains may demonstrate atypical chara and challenge a laboratory's ability to isolate and confirm the strain
The incorporation of educational specimens in a pr testing panel	Can be an effective mechanism for bringing interesting or important matters to the laboratories. Where an incorrect result is reported, it does not necessarily reflect po performance
Samples should be uniform in appearance	Avoids predicting the result

Table V

An example of a five-sample proficiency testing programme for hepatitis B virus serology

All samples provided in the challenge must be tested for the listed analytes, as accurate testing for each analyte affects the final patient status reported for the sample

Analyte				Patient status
Anti-	HBs	Anti-H	Anti-	
+	–	–	–	Vaccinated against HBV, no evidence of pas
+	–	+	–	Past resolved HBV infection
–	+	+	–	Current active infection with HBV
–	+	+	+	Chronic active infection with HBV
–	–	–	–	Not vaccinated, no evidence of past infectio

Anti-HBcT: total hepatitis B core (nucleocapsid) antibodies. Does not distinguish between antibody classes

Anti-HBe: hepatitis B pre-core protein antibodies. This protein is the precursor to the nucleocapsid

Anti-HBs: hepatitis B surface protein (envelope) antibodies

HBsAg: hepatitis B surface protein antigens. These assays detect the surface protein produced by the virus when actively replicating

HBV: hepatitis B virus

Box 1
Benefits and limitations of proficiency testing

Benefits of participating in proficiency testing
Quality
Aids compliance with ISO 17025: forms the basis for the requirements to gain or maintain laboratory accreditation (1)
Identifies technical issues and sources of error: instigates investigations into non-conforming work to implement corrective preventative actions
Valuable information can be gained from examining the follow-up actions taken after unsatisfactory performance in a profi testing (PT) exercise (is of particular interest to accreditation bodies and could help avoid the loss of accreditation)
Performance
Provides ongoing measurement of performance using known samples (reference materials) or unknown samples
Develops staff skills and knowledge
Surplus PT samples may be available from PT providers: acts as an 'unofficial' exercise that brings benefits similar to thos PT (particularly useful in follow-up work after poor performance)
Technical advice and support can be sought from PT providers
Methods
Assesses the suitability of the method from concept (validation) and verifies that it is fit for purpose throughout its life − checks validation retention status of a test method and enables adjustment of protocols if necessary
Provides valuable data to aid selection of appropriate methodology
Checks measurement uncertainties
Enables between-operator repeatability assessment (where PT permits results from multiple operators)
Results data can be used to estimate bias (where reference samples are used)
Enables laboratories to trial a new method or test kit, using samples with a 'known' value
Provides assurance of a successful test transfer between laboratories (verification)
Raises awareness of new technologies and methods
Other
Is a significant educational element for raising awareness of emerging and rare diseases and pathogens

Strengthens a laboratory's contribution to the quality and accuracy of animal disease surveillance and threat detection, thu preparedness to manage surge capacity for an outbreak response
Enhances a laboratory's reputation, thus helping to create international partnerships with improved collaboration and infor sharing between laboratories and countries
Facilitates trade by creating confidence in results, thus improving customer relations
Provides evidence of competence to help gain business
Demonstrates compliance with regulations, e.g. Regulation (EC) No 999/2001 2014 (Prevention, control and eradication o transmissible spongiform encephalopathies) (22)
Litigation – the use of a PT programme can help gain a significant advantage
Limitations of proficiency testing
Provides a snapshot of a single point in the testing framework as opposed to a full assessment of the testing process
Samples often do not reflect real samples
Some proficiency tests allow only one set of results for each sample
Proficiency testing for point-of-care tests only evaluates the quality of a single unit of the device used. Quality can vary bet unit devices, including within the same lot (to a certain extent, inter-batch variation can also apply to other diagnostics test

Fig. 1

The usual process of proficiency testing, showing the pre-analytical, analytical and post-analytical stages of the testing process

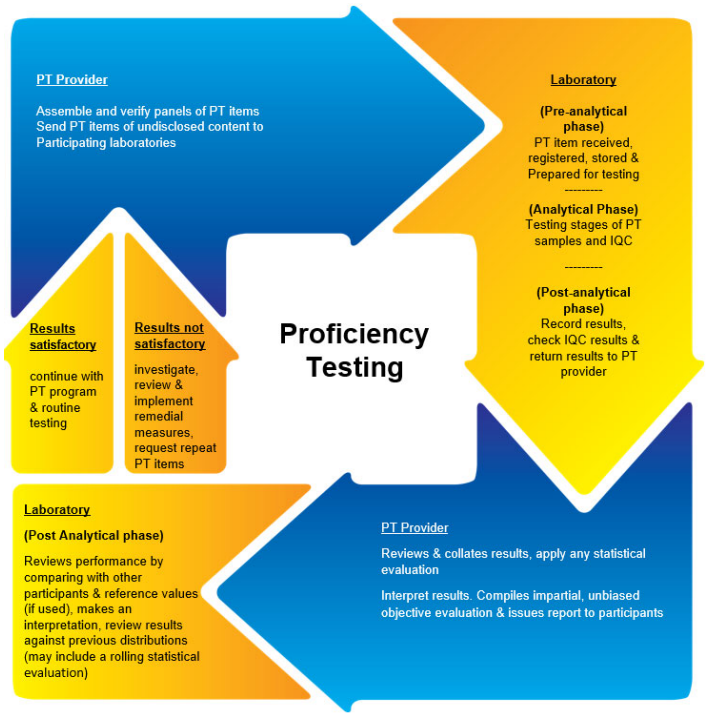

Fig. 2

A change in the diagnostic sensitivity of an assay, identified across two distributions, as the participants moved to a new assay batch

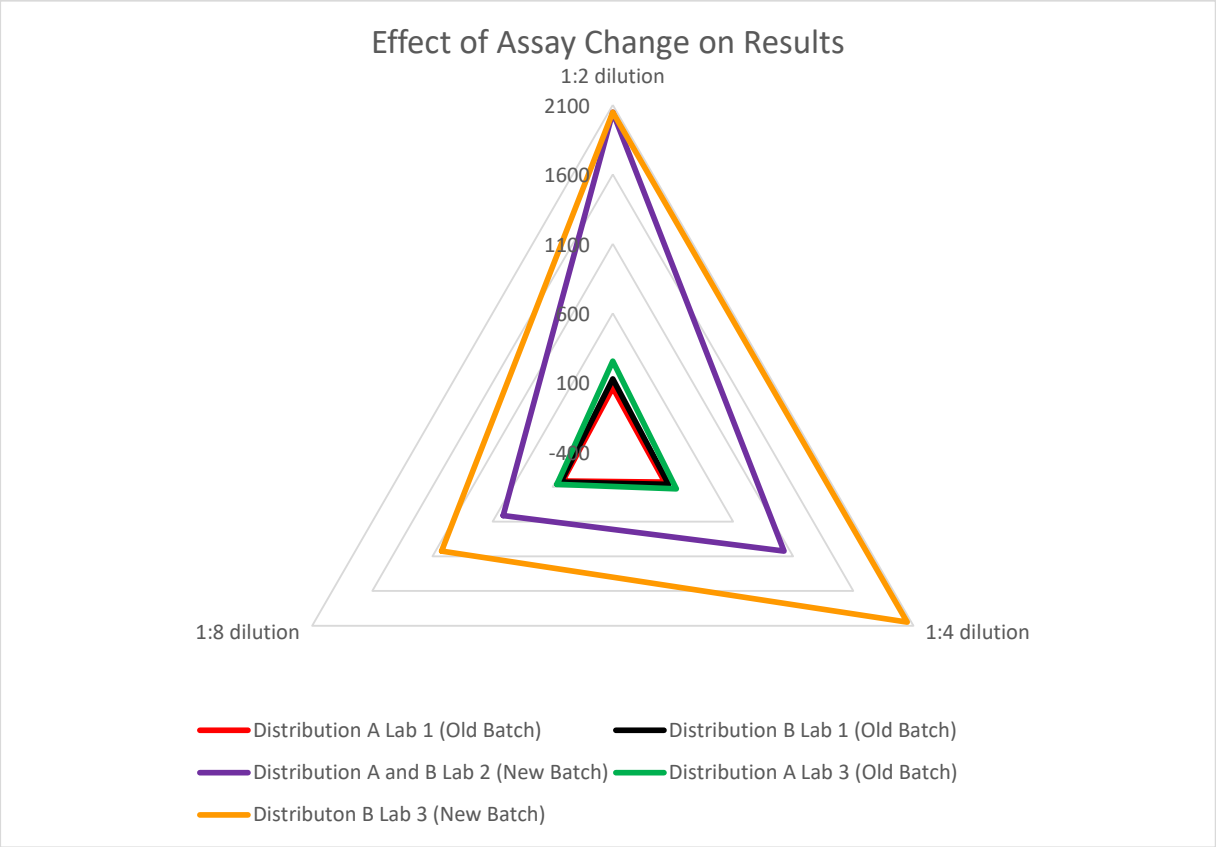

Fig. 3

A *Salmonella* isolation ring trial: results

This figure shows how regular participation in an annual *Salmonella* isolation ring trial has led to improvements in performance over time