Plant Health Review

Principles of Diagnostic Assay Validation for Plant Pathogens: A Basic Review of Concepts

Kitty Cardwell,† Oklahoma State University, Entomology and Plant Pathology, Stillwater; **Geoffrey Dennis,** USDA Animal and Plant Health Inspection Service (APHIS), Plant Protection and Quarantine (PPQ), Riverdale, MD; **Andrew R. Flannery,** PathSensors, Inc., Baltimore, MD; **Jacqueline Fletcher,** Oklahoma State University, Entomology and Plant Pathology, Stillwater; **Doug Luster,** USDA Agricultural Research Service, Foreign Disease–Weed Science Research Unit, Fort Detrick, MD; **Mark Nakhla,** USDA-APHIS-PPQ, Riverdale, MD; **Anna Rice,** Envirologix, Inc., Portland, ME; **Pat Shiel,** USDA-APHIS-PPQ, Center for Plant Health Science and Technology, Raleigh, NC; **James Stack,** Kansas State University, Plant Pathology, Manhattan; **Coilin Walsh,** Agdia Inc., Elkhart, IN; and **Laurene Levy,** USDA-APHIS-PPQ, Riverdale, MD (deceased)

Accepted for publication 22 August 2018.

Abstract

Diagnostic test results are used for many purposes and are heavily relied upon to prevent movement of pathogens from one country or region to another; to clear plants as free from disease for commerce; and to surveil against natural, accidental, or unintentional pathogen introductions and new and reemerging diseases. Diagnostic test results are also used by plant producers to make critical management decisions. The level of confidence in each assay is defined by a set of metrics that describe the performance of the assay under defined conditions. Collectively, these metrics are called "validation" of the assay. There are hundreds of diagnostic plant disease assays used every day in the United States that either are not validated or are validated in an ad hoc way. An endemic pathogen in one region can be an exotic pathogen in another. Therefore, this multitude of diagnostic assays for which performance criteria are not standardized impairs communication about confidence in the test outcome. We propose to create a framework for standardizing assay validation language and definitions within the United States across existing plant diagnostic networks. The long-term goal is to have operable standards, understanding the "trueness" of assay results, and sustained communications between diagnostic laboratories that use and those that develop plant disease diagnostic assays.

Validation is a scientific process that defines the requirements of an assay for the intended purpose (fitness for purpose), which includes optimization and standardization (Van der Vlught et al. 2007). The results from assay validation can be used to express the expectations about the quality, reliability, and consistency of analytical results. There is research involved in assay validation, but the process itself is distinct from the research that is understood and done by most pathologists. The intent of this review is to elevate the principles and language of diagnostic assay validation for the plant pathology community.

All diagnostic assays should be validated by taxon of host and pathogen for which it is developed. Validation includes estimates of the analytical and diagnostic performance characteristics of a test. Validation is an ongoing process to be sure that the assay continues to be effective and maintains its performance characteristics (OIE 2017). This review will describe generic concepts because of the evolving repertoire of new and unique diagnostic assays. This article focuses on the validation criteria for all types of assays, focusing on processes that can relate to any assay type. Note: for further assistance, a glossary of definitions can be found at https://www.apsnet.org/edcenter/intropp/Pages/AssayValidationGlossary.aspx.

†Corresponding author: Kitty Cardwell; E-mail: kitty.cardwell@okstate.edu

Funding: This work was supported financially by Agdia Inc., PathSensors Inc., and the American Phytopathological Society.

Capabilities for diagnosis (determination of the presence of disease agents) in plant pathology are being strengthened with new tools and techniques for pathogen detection, identification, and discrimination at the genus, species, strain, and even individual level. Use of plant disease diagnostic technologies is essential to the work of diverse groups of professional practitioners, including diagnosticians who provide plant disease answers to producer clients, diagnosticians who support plant health regulators at the federal, state, and local levels, customs officials, academic researchers, industry personnel, and agricultural producers. Plant disease diagnosticians such as those in the National Plant Diagnostic Network (Stack et al. 2014), the National Clean Plant Network, and the National Seed Health System as well as state departments of agriculture are adapting and applying a new array of assays to meet the specific needs of the plants grown in their region. Plant health regulators and bioforensic investigators depend on accurate detection, identification, and discrimination among endemic and exotic strains of pathogens (Budowle et al. 2005; Fletcher et al. 2006; Kingsolver et al. 1983). Customs officials, facing an ever-increasing barrage of imported plant and food materials, also require rapid, reliable, and accurate tools to determine the presence or absence of destructive, highly pathogenic organisms (Kingsolver et al. 1983). Researchers in law enforcement, regulatory, academic, and private sector laboratories are using next-generation sequencing, melting-point discriminators, rapid primer development, and other new technologies to develop new detection and diagnostic tools that address the needs of each of these stakeholder groups, providing more precise evidence for criminal

investigation, for making regulatory decisions, for improving the quality and depth of research data, and for facilitating crop management decisions.

Thus, diagnostic assays are common to widely divergent professional applications having a range of objectives and requirements for stringency and confidence. When the work of these different professional groups overlaps, such as when a diagnostic laboratory receives a sample that is subject to regulatory policies, or when the investigation of an agricultural crime brings together law enforcement, regulatory personnel, and plant diagnosticians, effective interactions depend upon understanding a common language related to assay development, validation, application, and reporting.

Validation is the process that (i) assesses the ability of a procedure to get reliable results under specific conditions, (ii) defines the conditions needed for specific results, (iii) determines limitations, (iv) identifies sources of variability that must be controlled, and (v) forms the basis for interpretation (Budowle et al. 2008). Terms such as specificity, inclusivity, exclusivity, sensitivity, accuracy, and precision are metrics of robustness and confidence pertaining to a diagnostic assay. These metrics are based upon well-defined calculations.

Confidence requirements may be different across a continuum of need from law enforcement forensic tracing of an unusual outbreak of foodborne illness to regulatory rejection of commodities in food and ornamental plant trade to successful management of a local or regional plant disease outbreak caused by an endemic species. The ability to ascribe a level of confidence to characteristics of a diagnostic assay considered acceptable to a specific field is critical for its development and use. The appropriate metrics can be based on whether the determination is quantitative, qualitative, or taxon-mediated and/or the exigencies of a specific outbreak event. Those concerned about the consequences of misdiagnoses are seeking concurrence on a framework within which to discuss, understand, and ascribe appropriate confidence levels over different applications with agreed stringencies, a standardized but flexible scientific base supporting the metrics utilized, and agreed-upon terminology and research methods for validating diagnostic assays.

Fitness for Intended Purpose

Fitness for purpose refers to an ideal level of confidence in the results of a diagnostic assay for which it is designed. The amount of validation is subjective or situational and can be defined in terms of the goals of the users and circumstances surrounding the testing needs. Validation of a diagnostic assay, then, is the testing of the assay in a variety of settings and circumstances relevant to the intended application (e.g., forensic investigation, regulatory confirmation, general diagnostics, research, etc.) and analysis of the test data to demonstrate that the assay is fit for its intended purpose. The types of assay "fitness for purpose" are described in this way by the World Organisation for Animal Health (OIE 2018):

1. To confirm diagnoses of exotic or high-consequence pathogens for critical regulatory decisions
2. To demonstrate population "freedom" from infection or reestablishment of freedom after an outbreak
3. To demonstrate freedom from infection for trade purposes
4. To demonstrate efficacy of eradication policies
5. To estimate prevalence of infection for risk analysis (survey, trace forward, trace back)

Additional uses relevant to plant pathology may include the following:

1. Research such as environmental and/or product sampling
2. Attribution in legal or forensic applications

Examples of fitness requirements of assay validation for different applications are described below.

Forensic diagnostics for attribution. In the post-9/11 world, the potential threats of use of biological agents for terroristic or criminal purposes led to establishment in the United States of dedicated national-level microbial forensic programs to address the threats. The National Bioforensic Analysis and Countermeasures Center (NBFAC), a branch of the National Biodefense Analysis and Countermeasures Center, was established by presidential directive (HSPD10: Biodefense for the 21st Century) to serve as the lead federal facility for analysis of material recovered from sites of suspected actual or planned use of biological agents. The mission of NBFAC is to identify the biological agents in the recovered materials, providing information that can be used as evidence supporting the attribution of a crime. In this case, the purpose of the diagnostic assay is to identify an organism associated with an event, distinguishing it with confidence from organisms of the same or similar species. Real-time polymerase chain reaction (PCR), a rapid and relatively inexpensive method for identification of biological agents, is frequently requested as part of the analysis for regulatory decision making and for evidentiary materials (Bonants et al. 2003; Budowle et al. 2005; Fletcher et al. 2006). To establish that they are fit for use, the real-time PCR assays used by NBFAC are validated in-house, in compliance with a quality assurance program requiring that all assays have detailed tracking of the reagents, personnel, equipment, and samples used. For plant pathogens, NBFAC turned to the National Institute of Microbial Forensics in Food and Agricultural Biosecurity (NIMFFAB) for validation of assays for plant select agents (Fletcher et al. 2006). Validations are documented through validation plans, reports, and periodic data review. Laboratory personnel are tested on their ability to correctly perform the assays and complete all documentation before any assay is performed on evidentiary material.

Validation for quarantine and regulatory diagnostics. When the USDA Animal and Plant Health Inspection Service (APHIS) uses diagnostic assays for regulatory purposes, it must comply with the U.S. laws, regulations, and policies that authorize regulatory actions, especially the U.S. Plant Protection Act (Public Law 106-224, 2000). These statutes inform the assay's fitness by requiring that regulatory decisions made from assay results use sound science and are sufficiently documented to withstand legal challenges, both in a U.S. court of law and by international bodies. The validation components that establish the assay's fitness are adopted from well-established U.S. governmental regulatory bodies such as the Food and Drug Administration (FDA) and the Environmental Protection Agency. As the FDA states in its *Guidelines for the Validation of Analytical Methods for the Detection of Microbial Pathogens in Foods and Feeds* (FDA 2015), "As a regulatory agency tasked with ensuring the safety of the nation's food supply, it is imperative that the laboratory methods needed to support regulatory compliance, investigations and enforcement actions meet the highest analytical performance standards appropriate for their intended purposes. Development of standardized validation requirements for all regulatory methods used to detect chemical and radiological contaminants, as well as microbial pathogens, used in our laboratories is a critical step in ensuring that we continue to meet the highest standards possible."

Some test requirements in the regulatory realm are similar to those in the forensic realm, but there are differences as well. For example, usually it is necessary to establish that an organism is identified as the precise species or subspecies listed in the

regulation, and that the organism was viable and pathogenic at the time of sampling, in order for a response action to be initiated (i.e., treatment, confiscation, removal, or re-exportation of regulated materials). Identification of the key test performance measurements required for an assay's intended use will define what test data will be needed for validation and demonstrate that the assay is ready for implementation.

Guidance for determining a range of validation "stages and tiers" needed for regulatory use has been developed by multidisciplinary teams in cross-network laboratory testing forums such as the Integrated Consortium of Laboratory Networks (https://www.icln.org/) and strategic coalitions of regulatory plant diagnostic agencies such as the Quadrilateral Scientific Collaboration in Plant Biosecurity ("the Quads"). A framework for diagnostic assay development and validation that identifies stages and tiers of calibration of an assay for use by assay developers and diagnosticians is developed/proposed in Figure 1. Each tier lists basic test performance measurements that establish suitability for purpose. Tiers can be adjusted for tests employed in circumstances ranging from specific emergency use to use in a sustained multiyear national regulatory program. Even after an assay is fully validated, continuous monitoring identifies the assay's strengths and areas for possible improvement.

Confirmatory assays for regulatory applications are generally more rigorously assessed for accuracy and detection specificity than are those assays for screening and surveying. Nevertheless, in some emergency situations assay validation may be incomplete, the minimum requirement being known analytical specificity, sensitivity, and repeatability as established in tier 1 validation (Fig. 1). Over time, in such situations, validation research will continue to enhance the available metadata about an assay, so that greater certainty is generated (tier 2 validation). Accuracy, precision, and robustness of an assay must be demonstrated in at least two to three laboratories for tier 3, and an assay is fully validated for national use by tier 4, when it has been used by multiple laboratories (Fig. 1).

Validation of commercial assays. The private sector has a vibrant plant pathogen diagnostic industry that serves a wide range of markets—from voluntary, home-based testing employed by orchid hobbyists to regulatory/quarantine programs distributed across the United States. This industry's diversity presents a myriad of unique challenges for commercial assay development and validation. As with other plant diagnostic testing, validation of commercial diagnostic assays uses "fitness of purpose" as the driving force to determine the appropriate validation requirements. In addition, developers of commercial assays cover traditional (prerelease) validation as well as approaches to the often-overlooked process of postrelease (ongoing) validation to ensure that each test stays fit for use over time in an ever-changing market.

Steps to commercialize a diagnostic assay are an iterative process, wherein the end user is continually engaged. Requirements of

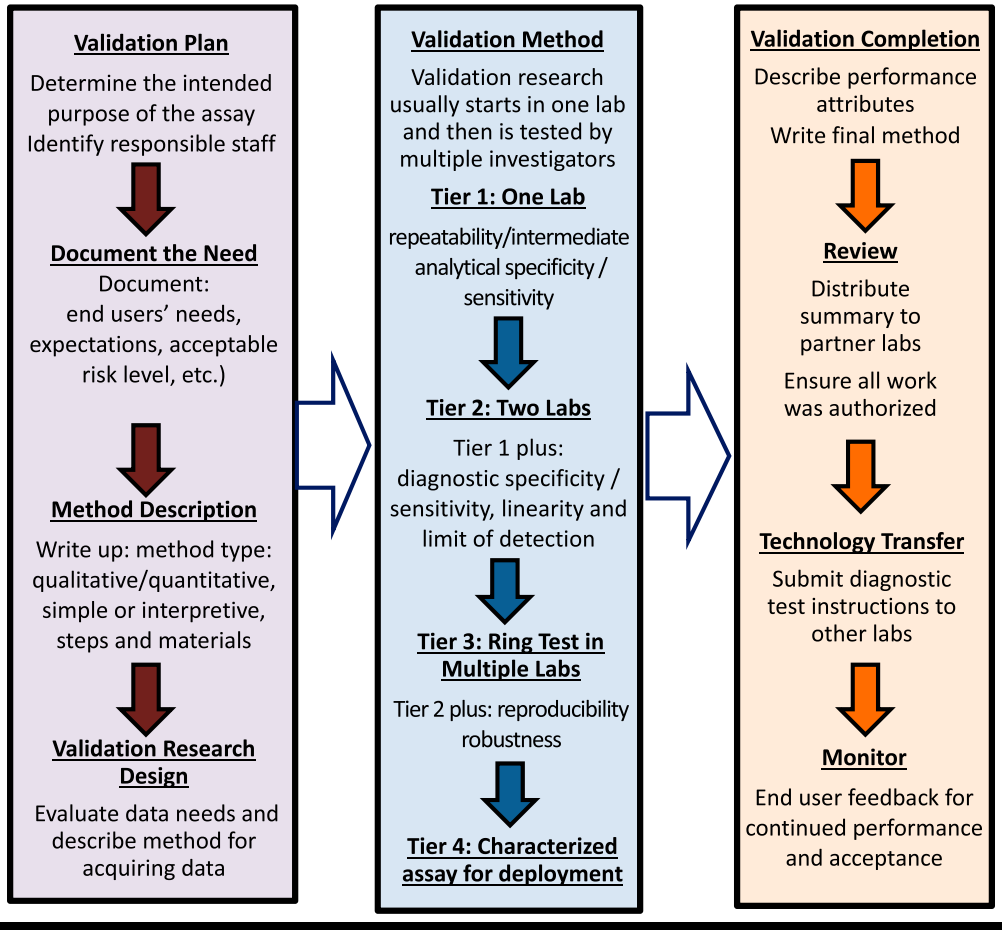

FIGURE 1

Overview of the assay validation process moving from establishing characteristics desired for validation to postvalidation activities. Each step is reviewed against predefined criteria before proceeding to the next. Validation planning (maroon), validation activities (blue), and validation and postvalidation work (orange) are also considered when planning the validation study.

the end user, sourcing of reference materials, and externally validating efficacy are critical to a successful validation. Assay validation metrics are established and results analyzed to ensure that the assay meets fitness of purpose. Once the product is launched, ongoing validation processes, listening to customer feedback, monitoring industry changes, and compiling and analyzing data are continued, and, if needed, revalidation is undertaken. Reasons to continually validate are that pathogen populations change, knowledge expands, and novel varieties are constantly being developed. For example, a new ornamental variety with a red pigmentation can cause a false positive on a lateral flow immunoassay that otherwise produces reliable results for other varieties in the same species that are not pigmented. To continually validate, there is a real-time stability (shelf-life) program with established monitoring processes. It is important to maintain two-way communications with customers, which requires that there are processes to document feedback, take appropriate action, and communicate actions taken to customers. Commercial test kit providers commonly maintain ongoing testing programs and industry relationships (collaborators from extension offices, university researchers, the American Phytopathological Society [APS], etc.)

Fitness for Pathogen Type

Fitness for intended purpose has been evaluated from the perspective of who will use the assay and the performance characteristics required by different users. However, there may be significantly different verification requirements when considering plant pathogens across the spectrum of taxa (e.g., Janse 2005; van Schadewijk et al. 2011). Minimum assay performance characteristics would likely differ when validating diagnostic assays for viral, phytoplasma, bacterial, stramenopile, and fungal plant pathogens, for example. Additionally, well-established versus emerging pathogens may have differing degrees of resources available for validation. When dealing with a recent discovery or taxonomic placement of an emerging pathogen, assay development and validation are often conducted with less information, fewer isolates, and more unknown factors, compared with established pathogens. New and emerging pathogens are not often represented with type specimens in culture collections, reference standards may not be developed or available, and biochemical, genomic, and proteomic data, and so on, may be marginal or lacking. Progressing from lower (viruses) to higher (fungi) taxa, with associated increasing genomic and structural complexity, criteria and requirements for validation become more challenging and complex. Lower taxa such as viral and bacterial pathogens often have significantly more genomic and biochemical information available, as well as defined near neighbors, to exploit in development and validation of diagnostic assays. Pathogens at lower taxonomic levels and associated lower genomic complexity thus may be more amenable to new technologies such as next-generation sequencing, with lower costs associated with validation. Conversely, pathogens with more complex genomes and lifestyles may present other challenges in development and validation.

Additionally, it can be a challenge to collect sufficient samples for specificity testing, particularly when such samples are only available from collaborators in foreign countries. This reinforces the need for postdeployment monitoring of assay performance and the potential need for revalidation as more information and specimens become available.

Across all taxa, commonalities exist for development of validation protocols:

- The need for type cultures and reference standards such as antigens, barcode sequences, and databank genomes
- Access to multiple isolates, strains, pathovars, or genotypes and outgroup species for inclusivity and exclusivity testing
- The need for solid taxonomic status before identifying subjects for inclusivity/exclusivity testing
- The need for known positive samples to establish metrics of specificity and sensitivity
- The need for partner laboratories to use the assay and provide feedback to the developer on efficacy and reliability

Validation research planning for diagnostic assay development. Validation is a scientific study to determine the performance characteristics of a test. Decisions about the metrics of performance of a diagnostic assay can be subjective, depending on perceived risks and consequences of an incorrect diagnosis. The selection of the type of validation study is dependent on the targeted level of confidence in the test results and on the stringency level of the validation study. The required level of stringency of the validation study will vary based on the scope of study and the intended use of the diagnostic test. Other factors such as the availability of resources, urgency, and technological limitations will also affect the selection of level of validation. The number of replicates required in the validation process is determined by the scope of the study and the desired level of confidence in the test results. The minimum numbers of required replicates for each metric of confidence such as specificity, sensitivity, specificity, exclusivity, and robustness might be different (Fig. 1).

Assay Performance Characteristics

Once the purpose and initial preparations for the development of an assay are complete, the next steps are to define the performance characteristics for deployment.

Diagnostic sensitivity and specificity. Diagnostic sensitivity and specificity, the main properties of any clinical assay, are measurements of the test's capability to effectively distinguish the presence or absence of a specific pathogen. They are the basis for calculation of other parameters and impact the kinds of inferences that can be made about the test results. Evaluating specificity and sensitivity metrics during diagnostic assay development is essential because they suggest a potential concentration for the method's limit of detection and its strength to discriminate among similar taxa, providing consistent measures of test performance. Therefore, it is important that these two characteristics are as accurate as possible. Ideally, they are derived by testing a series of reference samples with known infection status.

Sensitivity predicts the test's capability to yield a positive result when the targeted characteristic (nucleotide, protein, organism, etc.) is present. When assessed using the slope of a concentration curve, its shift can be used to evaluate reduction in sensitivity. Specificity is the ability to distinguish background, or nontargeted organisms, from those containing the target organism.

Specificity. When designed and developed well, PCR methods can be rapid, specific, and sensitive; PCR is frequently used for plant pathogen detection and identification. In most cases, PCR assays are performed with pathogen-specific primers because nonspecific primers can crossreact with closely related species, resulting in false positive test results. Comprehensive specificity validation includes properly designed inclusivity and exclusivity panels of microbes, defining the probabilities of false negative and false positive results. In general, members of exclusivity panels should include pathogens from the same niche, from closely related species and subspecies, and from environmental samples.

To minimize the probability of false negative results during testing, representative isolates of the targeted pathogen collected

from different geographical regions, time periods, and hosts should be used in the inclusivity panel. Selecting isolates for inclusivity panels requires having background information on the biology, ecology, and population genetics of the isolates, which helps to evaluate the importance of each isolate for the validation.

Although the best practices for assay design are tailored for each of several fundamentally different types of assays (e.g., detection of nucleic acids, antigens, or analytes), it is important to first consider the intended purpose for all diagnostics. A single assay may be designed for one or more intended purposes by optimizing its performance characteristics separately for each purpose: for example, setting sensitivity high with associated lower diagnostic specificity for a screening assay or, conversely, setting specificity high with associated lower sensitivity for a confirmatory assay.

When optimizing the assay's specificity, it is important to consider several factors inherent in the sample matrix. Because plant pathogens form intimate associations with their host or are present in the environment, diagnostic samples can consist of a variety of materials including plant tissues, soil samples, and aerial spore trap adhesive. Elements of the sample matrix that can affect testing may be grouped into (i) interferents, such as assay enzyme inhibitors, (ii) degradants, which could result in the destruction of target analytes, (iii) factors causing nonspecific binding, and (iv) common plant treatments that could alter the assay performance (pesticides, growth regulators, spray adjuvants). For these reasons, testing the sample matrix alone (negative control samples) is critical. In addition to matrix effects, it is important to know basic information about the target organism to help define an exclusivity panel. The target organism's niche may also contain resident nonpathogenic organisms that should be included in the panel to minimize the possibility of false positive results due to their presence. Furthermore, closely related but nontarget organisms should be included in the panel to minimize false positives. The level of specificity needed (i.e., genus versus species or pathovar) will help determine which organisms to include in the panel. It is also important to include organisms that may not be in the same niche but that are commonly associated with the test sample; these may include opportunistic organisms or contaminating organisms from the environment. Finally, it is important to determine if a specific stage in the life cycle is being queried or if distinguishing the organism as live/dead when sampled is important.

Diagnostic specificity, a measure of how sure one can be that a negative result is truly negative, is represented by the following equation (Emory University School of Medicine 2018):

$$\text{Specificity} = \frac{\text{number of observed negatives}}{\text{number of true negatives} + \text{false positives}} \times 100$$

Sensitivity. Sensitivity measures the effective accurate pathogen detection level. It is important to know the difference between diagnostic (rate of true positive detection) and analytical sensitivity (dose-response curve of the method) metrics (OIE 2017).

Diagnostic sensitivity is represented by the following equation:

$$\text{Sensitivity} = \frac{\text{number of observed positives}}{\text{number of true positives} + \text{false negatives}} \times 100$$

Analytical sensitivity (quantitative test) represents the smallest amount of substance in a sample that can be accurately measured by an assay (Saah and Hoover 1997) and is represented as limit of detection (LOD).

A test that has high analytical sensitivity may not necessarily yield the desirable diagnostic sensitivity. Such a test would be capable of detecting very low levels of the target analyte or organism but may produce a high rate of false positive results. Whereas diagnostic sensitivity is the percentage of diseased samples that are identified by the test as being positive for the disease, analytical sensitivity is defined as the lowest level of target analyte that can be measured accurately by the test. For qualitative assays, the analytical sensitivity is the lowest concentration consistently detected as positive under repeat testing. For quantitative assays, the analytical sensitivity measurement includes a level of accuracy by incorporating the standard deviation (or variation in the test result) into the calculation and is also known as LOD (Emory University School of Medicine 2018):

$$\text{LOD} = \text{mean of replicate tests of the same positive sample} \pm 2 \text{ standard deviations of the replicate}$$

It is common practice to check method accuracy during development using a limited number of samples. This approach is the most efficient pathway to the best designed method. Quantifying test performance during method development and ways to predict its performance require practice. Appreciating and using sensitivity metrics will benefit diagnosticians, providing consistent measures to evaluate test performance, as well as guide assay design to yield the most beneficial diagnostic tools.

Diagnostic precision, reproducibility, repeatability, and accuracy. One of the most fundamental measuring sticks for any test is its range of error. The closeness of agreement (distribution of variability) between a series of measurements provides a window into the *precision* and *accuracy* of an assay. Through assessments of the range of error at the beginning of assay development, the method developer can estimate the amount of work required for improvements, whether it has a chance to be more sensitive than other assays, and to begin to define the expected range of error for future testing. Following this range through the interim of development work, the developer can observe if changes made to the assay are decreasing or improving its effectiveness, and expanding or decreasing the expected range of error as new sources or error are introduced. Finally, once the method is ready to become validated and prepared for routine use, all the characteristics and circumstances tested during development are defined to provide specific expectations to those who will eventually use the assay. *Validation establishes the expectations for how the test will perform and how conditions affect the integrity of results.*

Sources of error contribute in different ways from different points. The more critical control points (CCPs) that have defined ranges of error, the faster an end user can determine the root cause of unexpected results. CCPs can be loosely defined as points at which there is a risk that the assay might suffer catastrophic failure. There are three categorically distinct areas to be defined, collectively referred to as *precision*. Precision is defined as a collective of three categories: *repeatability*, intermediate *precision*, and *reproducibility* (ICH 1994). Respectively, these define sources contributing to error from replicates, different days or conditions in the same laboratory, and/or different laboratories. These areas require some form of verification and definition from a validation study.

Assay *precision* has to do with the range of outcomes. A small range in variance, standard deviation, or coefficient of variation among a series of measurements means the assay is more precise than an assay with more variability. The assay has high precision if the outcomes are consistent within the same lab and operating conditions (repeatability), between multiple operators with varying levels of experience and different equipment (intermediate precision), and between laboratories (reproducibility).

Accuracy is the nearness of a test value to the expected value for a reference standard reagent of known activity or type organism of known titer. A reference standard, to which a type of specimen, reagent, or assay can be calibrated, is not available for many plant pathogens and may have to be simulated. Determining the accuracy of a test defines another form of error, bias. Not only do the measurements themselves have a range of error, but the nearness of a test value to the expected value for a reference standard of known activity or titer will produce a shared range between known and unknown means. When no reference standard exists for a pathogen of interest, it is incumbent upon the assay developer to establish the reference and submit to a known reference laboratory or identify source of sequence data to simulate the standard.

With so many different sources contributing to the error of a measurement, the developer must decide what range is (i) reasonable and (ii) predictable. The decision can be supported using *confidence intervals*, which rely on the standard deviation of a dataset, and defining *confidence levels* of those intervals, reliant on the number of data points observed. In general, a 95% confidence interval is a reasonable range to ascertain, and a confidence level between 85 and 95% carries decent predictability. The confidence interval measures a chosen percentage of measurements about the mean. However, that interval by itself is only applicable to that series of measurements. It may be reasonable for that dataset but lack predictability. The confidence level is a probabilistic prediction of consistency over time. Knowing the confidence level adds a defined probability that the interval observed represents the same range future measurements should fall into. In other words, the confidence interval is the range most likely to include the true mean.

Assay *robustness* defines how well an assay maintains precision when subject to variable factors (temperature, humidity, stability) likely to occur during use by many different diagnosticians in different laboratories. A *ring test* is an evaluation of assay performance and/or laboratory competence in the testing of defined samples by two or more laboratories; one laboratory may act as the reference in defining test sample attributes. Often, diagnostic laboratories are recruited into a ring test so that the assay developer can ascertain how robust the assay performance is.

Afterword

A one-day workshop titled "Principles of Validation for Plant Pathogen Diagnostic Assays" was held at the APS meeting on 5 August 2017 in San Antonio, TX, organized by the authors of this report, and attended by 65 participants across a spectrum of academic, federal/state government, industry, researchers, and students. The objectives of the workshop were to develop a common understanding among plant pathologists of the language and metrics of diagnostic assay validation and to explore the modalities of validation research planning. There were three sets of activities: presentations concerning fitness for purpose, exercises to demonstrate how to calculate metrics of validation, and finally group work on five classes of taxa to understand commonalities and distinctions in assay validation research, depending on whether the assay was for a bacterium, a virus, a fungus, a stramenopile, or an unculturable organism such as a phytoplasma. This review is a summary of the content of that workshop and is presented to further discussion on the science and practice of diagnostic assay validation, upon which are based the assays so critical to protecting plant health and safeguarding trade. An output of the workshop is a glossary of assay validation definitions, which can be found at https://www.apsnet.org/edcenter/intropp/Pages/AssayValidationGlossary.aspx.

Recommendations

Recommendations from the workshop participants included the following:

- For future workshops, include issues of sampling, sample size, and statistics and experimental design for diagnostic assay validation.
- Develop a standard glossary of diagnostic assay validation terms; publish online.
- Create and formalize a validation guide for phytopathogen diagnostic assays.
- Establish an editorial check list and standards for manuscripts on diagnostic assay validation in APS journals.
- Engage international diagnosticians on commonality of concepts and terminology.
- Engage seed health testing professionals with experience in validation and employ seed health examples/case studies in future activities.
- Develop didactic training materials/courses:
  - Develop interactive exercise examples in, for example, Excel worksheet format.
  - Prepare in-depth workshops on each of the validation metrics.
- Identify funding gaps and communicate to the APS Public Policy Board and stakeholders.

Acknowledgments

The workshop that led to the preparation of this review was sponsored by the APS Microbial Forensics Interest Group, Plant Pathogen and Disease Detection Committee, and Emerging Diseases and Pathogens Committee and supported financially by Agdia Inc., PathSensors Inc., and APS. Kathy Schneider (National Biodefense Analysis and Countermeasures Center), Pat Shiel (APHIS, PPQ), Jim Stack (Kansas State University), Coilin Walsh (Agdia, Inc.), and Mark Nakhla (APHIS, PPQ) provided introductory talks. The session on validation metrics was moderated by Pat Shiel, with assistance by M. Arif (University of Hawaii), A. Rice (Envirologix), A. R. Flannery (PathSensors, Inc.), F. Ochoa-Corona and K. F. Cardwell (Oklahoma State University, NIMFFAB), G. Dennis and V. Mavrodieva (APHIS, PPQ), and C. Walsh (Agdia, Inc.). The fitness for pathogen type session was chaired by Doug Luster and assisted by J. Fletcher (Oklahoma State University, NIMFFAB, emerita), J. Rascoe (APHIS, PPQ), C. Garzon (Oklahoma State University), I. Tzanetakis (University of Arkansas), J. Bienapfl (APHIS, PPQ), A. Espindola (Oklahoma State University, NIMFFAB), J. Stack (Kansas State University), W. Schneider (USDA, ARS), and S. Costanzo (APHIS, PPQ). The group also is grateful for the contributions of the late Laurene Levy as one of the organizers of this workshop and for her pioneering work in validation of diagnostic assays for USDA regulatory programs.

Literature Cited

Bonants, P. J., Carroll, G. C., De Weerdt, M., van Brouwershaven, I. R., and Baayen, R. P. 2003. Development and validation of a fast PCR-based detection method for pathogenic isolates of the citrus black spot fungus, *Guignardia citricarpa*. Eur. J. Plant Pathol. 109:503-513.

Budowle, B., Schutzer, S. E., Ascher, M. S., Atlas, R. M., Burans, J. P., Chakraborty, R., Dunn, J. J., Fraser, C. M., Franz, E. R., Leighton, T. J., Morse, S. A., Murch, R. S., Ravel, J., Rock, D. L., Slezak, T. R., Velesko,

S. P., Walsh, A. C., and Walters, R. A. 2005. Toward a system of microbial forensics: From sample collection to interpretation of evidence. Appl. Environ. Microbiol. 71:2209-2213.

Budowle, B., Schutzer, S. E., Morse, S. A., Martinez, K. F., Chakraborty, R., Marrone, B. L., Messenger, S. L., Murch, R. S., Jackson, P. J., Williamson, P., Harmon, R., and Velsko, S. P. 2008. Criteria for validation of methods in microbial forensics. Appl. Environ. Microbiol. 74:5599-5607.

Emory University School of Medicine. 2018. http://www.med.emory.edu/EMAC/curriculum/diagnosis/sensand.htm. Emory University, Atlanta, GA.

FDA. 2015. Guidelines for the Validation of Analytical Methods for the Detection of Microbial Pathogens in Foods and Feeds, 2nd Ed. U.S. Food & Drug Administration Office of Foods and Veterinary Medicine, Silver Spring, MD.

Fletcher, J., Bender, C., Budowle, B., Cobb, W. T., Gold, S. E., Ishimaru, C. A., Luster, D., Melcher, U., Murch, R., Scherm, H., Seem, R. C., Sherwood, J. L., Sobroal, B. W., and Tolin, S. A. 2006. Plant pathogen forensics: Capabilities, needs, and recommendations. Microbiol. Mol. Biol. Rev. 70:450-471.

ICH. 1994. Validation of analytical procedures: Text and methodology. http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Quality/Q2_R1/Step4/Q2_R1__Guideline.pdf. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use, Geneva, Switzerland.

Janse, J. D. 2005. Standardization, validation and approval of test methods for quarantine bacteria: Examples of harmonization in plant health laboratories in Europe. Phytopathol. Pol. 35:19-27.

Kingsolver, C. H., Melching, J. S., and Bromfield, K. R. 1983. The threat of exotic plant pathogens to agriculture in the United States. Plant Dis. 67: 595-600.

OIE. 2017. Manual of Diagnostic Tests for Aquatic Animals. http://www.oie.int/standard-setting/aquatic-manual/access-online/. Accessed 13 July 2017. World Organisation for Animal Health, Paris, France.

OIE. 2018. Registration of diagnostic kits. http://www.oie.int/en/our-scientific-expertise/registration-of-diagnostic-kits/background-information/. World Organisation for Animal Health, Paris, France.

Public Law 106-224. 2000. Agriculture Risk Protection Act of 2000, 106th Congress, Title IV—Plant Protection Act. United States Statutes at Large 114: 357-456.

Saah, A. J., and Hoover, D. R. 1997. "Sensitivity" and "specificity" reconsidered: The meaning of these terms in analytical and diagnostic settings. Ann. Intern. Med. 126:91-94.

Stack, J. P., Hammerschmidt, R., Bostock, R., Jones, J., and Luke, E. 2014. The National Plant Diagnostic Network: Partnering to protect plant systems. Plant Dis. 98:708-715.

Van der Vlught, R. A. A., Verbeek, M., and Bonants, P. J. M. 2007. Explanatory document for the validation of detection methods for plant pests. Plant Research International B.V. Report 135EN. https://library.wur.nl/WebQuery/wurpubs/fulltext/301443

van Schadewijk, A. R., Meekes, E. T. M., Verbeek, M., and Verhoeven, J. T. J. 2011. Validation of plant virus detection. Acta Hortic. 901:81-86.