International Seed Federation
Seed is Life

Technical Guidelines for the Validation of Seed Health Methods

JUNE 2025, VERSION 4

Developed by
International Seed Health Initiative (ISHI) of the International Seed Federation (ISF)

TECHNICAL GUIDELINES AND THEIR PURPOSE

The following technical guidelines have been developed to facilitate in the validation of seed health methods. These guidelines are not intended to serve as a legal reference. They are not binding on ISHI nor ISF members.

DISCLAIMER

This document has been developed based on the current technical state of the art. ISF cannot be held liable for any possible claims associated with the application of these guidelines.

CONTENTS

INTRODUCTION

The International Seed Health Initiative (ISHI) aims to secure the delivery of healthy seed to customers by developing and validating methods for seed health testing that are internationally recognized as reference methods and accepted as industry standards. ISHI primarily focuses on developing methods for pathogens that are transmitted through seed. Because of their impact on seed trade, ISHI occasionally develops methods for organisms for which the pathway for introduction and establishment via seed has not been proven.

Note that a method is a collection of one or more assays that together show the presence, viability, and pathogenicity of a pest (See ISHI Glossary of Terms). In this document, the term assay will only be used, when something is not applicable to a method consisting of multiple assays.

1. METHOD VALIDATION

Validation is a requirement of method development and is a process that determines the fitness of a method for its intended purpose. ISHI has identified six performance characteristics to make this assessment (see Section 2), which are in line with those used by other regulatory bodies in seed health.

The scope of the method and its performance characteristics should be properly defined and documented in a validation plan before the experimental studies commence. The minimum requirements set for each performance characteristic should also be established and justified. The assistance of a statistician when designing the experimental studies may be required to ensure that the experiments proposed in the validation plan are statistically sound and appropriate for the analysis.

Data published in scientific, and peer reviewed journals can be used in method validation when they assist in demonstrating that the method is fit for purpose. Data generated during the developmental phase of the method can also be used.

Factors limiting the ability to validate the method according to these guidelines, such as deviations in the number of isolates tested due to the availability of infected material in the case of non-culturable, new, or emerging pathogens, should be explained.

For the validation, well-characterized seed samples should be used. If naturally infected seed samples are not available, artificially infected seed lots may be produced and used instead. The pathogen processed on artificially contaminated seed should ideally behave similarly to naturally infected seed.

After a method has been developed and is in use, ISHI encourages laboratories to continue monitoring its performance and report unexpected findings to the ISF Secretariat.

2. PERFORMANCE CHARACTERISTICS

2.1. Analytical specificity

Definition: *Analytical specificity is the ability of an assay to detect the target pathogen(s) (inclusivity) while excluding non-target pathogens (exclusivity).*

Analytical specificity is evaluated, preferably, using pure cultures of representative isolates. When a pathogen cannot be cultured (for example, virus, viroid and non-culturable bacteria), alternatives such as infected host material or synthetic DNA/RNA representing the target of interest can be used.

An assay should be evaluated with isolates that represents the known phenotypic and genotypic diversity of the pathogen. In addition, the assay should be evaluated with a diverse, characterized collection of relevant non-target isolates, such as those originating from the same crop species and location, and closely related species.

The recommended size of a characterized collection is a minimum of 20 targets, as well as a minimum of 20 non-targets. For culturable pathogens or when using synthetic DNA/RNA, the concentration of the isolates should be determined and set to ensure a consistent detection and be similar between all isolates.

Note: For nucleic acid-based assays, a complementary in silico *analysis is recommended for the validation, as described in the Technical Guidelines for PCR assay development, but should not replace the experimental validation.*

2.2. Analytical sensitivity

Definition: *Analytical sensitivity of an assay is the smallest amount of the target pathogen that can be detected with a given degree of confidence, commonly expressed as the limit of detection (LOD).*

Analytical sensitivity is evaluated using a minimum of three independently replicated dilution series of preferably a pure culture or a synthetic nucleic acid fragment of a representative isolate (e.g., well-characterized isolate selected from the analytical specificity experiment) of the pathogen. To this end, a known concentration of the target pathogen is spiked in extract from at least one seed sample. When a pathogen cannot be cultured (for example, virus, viroid and non-culturable bacteria), synthetic controls such as G-Blocks, RNA fragments or ssDNA oligonucleotides can be used as spike (for PCR based assays; Annex A).

The dilution series of the target pathogen in seed extract should be prepared with a maximum of a 10-fold difference between dilutions and with a final concentration of at least two orders of magnitudes lower than the expected detection limit based on experiences gained during assay development (e.g., 100 times below the LOD when using a 10-fold dilution series). Subsequently, the assay is performed with at least three technical replicates for (at least) the lower three concentrations (higher dilutions; Figure A.1 in Annex A).

The fraction of replicates generating a positive test result of the total number of replicates performed for the dilution, gives an estimation of the detection confidence (Figure 1). The concentration of the target pathogen, where ⩾95% of the replicates are consistently detected, is considered the LOD of the assay. Figure A.2 in Annex A introduces suggestions on determining the LOD.

$$\frac{Replicates\ with\ positive\ result\ for\ specific\ dilution}{Total\ replicates\ for\ specific\ dilution} \times 100\% = Detection\ confidence\ in\ \%$$

Figure 1. Calculation for the confidence detection for specific dilution.

For ELISA or in cases where quantification of the target pathogen is impossible due to its non-culturable state, or the impossibility to be isolated, sensitivity can be established by assessing the capacity of the assay to detect a single infected seed in a seed sample. To this end, the assay is performed on samples consisting of a range of preferably naturally infected seeds (for example, using the weight equivalent of 5, 3, 1 and 0.5 seeds) added to seeds from a healthy lot. Artificially infected seeds may be used in case no naturally infected seeds are available.

2.3. Selectivity

Definition: *Selectivity is the effect of different seed matrices on the ability of the assay to detect the target pathogen(s).*

Selectivity is evaluated by analysing the effect of the matrix variations by spiking a known concentration of the target pathogen in a relevant collection of different seed matrices.

Seed matrix variation may result from differences in crop species or varieties. The varieties may differ in saprophytic species and concentrations, seed treatments, year of production, or geographical origin and their effect on the final result should be verified.

The different seed matrices can be spiked with either pure cultures, or with seeds or leaf material infected with the target pathogen. The spike should consist of a suitable concentration in the medium range, as defined based on the LOD determined during the evaluation of the analytical sensitivity, with an expected detection of 100%. Spiking should result in a homogeneous contamination level. All crop species for which the protocol is under validation must be included.

A minimum of three and ideally five seed batches per crop species should be tested with at least one of the variables, such as variety, location, production year, etc., per matrix for qualitative and quantitative analysis.

2.4. Repeatability

Definition: *Repeatability of the assay is the degree of similarity in results of replicates of the same seed subsamples when the assay is performed with minimal variations in a single laboratory.*

Repeatability of an assay is evaluated by the same technician assessing technical replicates of seed subsamples (infected and healthy) using the same reagents and equipment and processed consecutively within a short time frame.

At least five seed subsamples with a range of infection levels should be tested, including a minimum of one healthy subsample, three subsamples with an infection level just above the LOD, and one subsample with a medium/high infection level. The infected subsamples should be uniformly infected. At least three technical replicates should be tested from each seed subsample and tested at three different timepoints, resulting in a minimum of nine data points for the healthy and

medium/high infected subsamples and a minimum of 27 datapoints for the subsamples with an infection level just above the LOD.

For qualitative results, the repeatability of the assay can be estimated by calculating the accordance, which represents the chance that identical test materials analysed by the same laboratory under standard repeatability conditions will both yield a positive or negative result (Langton et al., 2002). This is based on the probability of finding the same test results for identical test materials by the same laboratory. Accordance of at least 90% is required (ISTA, 2021).

2.5. Reproducibility

Definition: *Reproducibility is the degree of similarity in results when the assay is performed across laboratories.*

Reproducibility of an assay is evaluated in a comparative test (CT), in which seed subsamples are tested in multiple laboratories with the pre-determined conditions described in the protocol and within a defined timeframe.

For the CT, the assay should be performed by experienced laboratories on replicate seed subsamples. It is recommended to include six to eight participants in each CT to ensure reliable and robust reproducibility data, although the minimum requirement is to have three participants for statistical purposes. The seed subsamples should be drawn from seed lots with a range of infection levels. The CT technical guidelines present the process for determining the number of seed lots per infection level. Replicates should preferably be infected uniformly, and this is assessed by homogeneity testing of the seed lots prior to the CT. Additionally, the CT organizer should check the stability of the pathogen in the seed subsamples to assess the effect of time and conditions during shipment on the pathogen and the impact they may have on the CT test results. A detailed description of CT organization is provided in the CT technical guidelines.

For qualitative results, the reproducibility of the assay can be estimated by calculating concordance, which indicate the rate at which two identical subsample sets tested in different laboratories will give identical results (Langton *et al.*, 2002). This is based on the probability of finding the same test results for identical test materials between laboratories. A concordance of at least 90% is considered to indicate good reproducibility (ISTA, 2021).

2.6. Diagnostic performance

Definition: *Diagnostic performance is an evaluation of the ability of the assay to discriminate between positive and negative seed samples.*

To assess diagnostic performance, the results of the assay should be compared to the results obtained with a reference assay. The two assays should be performed concurrently on preferably the same seed subsamples to circumvent any bias in the analysis due to potential deterioration of the seed subsamples over time.

Diagnostic sensitivity and diagnostic specificity are measures of the test's ability to discriminate between the presence and absence of the target pathogen. Diagnostic sensitivity is a measure of the confidence that a positive result is a true positive, while diagnostic specificity is a measure of the

confidence that a negative result is a true negative. Evaluating these metrics as accurately as possible is essential, as they are measures of confidence in the test results (Cardwell *et al*., 2018).

A minimum of three infection levels should be tested including healthy seed, medium and high infection levels. At least three seed subsamples should be tested for each infection level (i.e., a total of nine seed subsamples). In the case of a lack of naturally infected seeds, spiking in three different seed backgrounds should be considered.

The two measures of diagnostic performance are obtained by a calculation based on true and false results (Table 1).

Table 1. Assessment of diagnostic sensitivity and diagnostic specificity by comparing the new assay results against the reference assay results.

		Reference assay results[a]	
		Positive	**Negative**
Test results	**Positive**	True positive (TP)	False positive (FP)
	Negative	False negative (FN)	True negative (TN)
		DIAGNOSTIC SENSITIVITY = (TP / (TP + FN)) × 100%	**DIAGNOSTIC SPECIFICITY = (TN / (FP + TN)) × 100%**

[a] Possibility to use data from a comparative test, when no reference assay is available.

The required rates for diagnostic sensitivity and diagnostic specificity should be defined in the validation plan for each assay. Seed samples should be drawn from relevant origins, production years and infection levels among others that have been listed and selected based on the most important sources of variations.

If a reference assay is not available, the results of a comparative test (CT) can be used to determine the diagnostic performance. The results of the CT participants are compared to the CT organiser's results and used to calculate the diagnostic sensitivity and diagnostic specificity using the formulas presented in Table 1.

REFERENCES

Cardwell, K., Dennis, G., Flannery, A.R., Fletcher, J., Luster, D., Nakhla, M., Rice, A., Shiel, P., Stack, J., Walsh, C., and Levy L. (2018). Principles of diagnostic assay validation for plant pathogens: A basic review of concepts. *Plant Health Progress,* **19**(4), 272-278. https://doi.org/10.1094/PHP-06-18-0036-RV.

ISTA (2021). Validation of seed health methods and organisation and analysis of interlaboratory comparative tests (CT). Version 2.0. https://www.seedtest.org/api/rm/XS74H86BYJ87N33/validation-of-seed-health-methods-and-organisation.pdf. Last accessed on 02/04/2024.

Langton, S.D., Chevennement, R., Nagelkerke, N. and Lombard, B. (2002). Analysing collaborative trials for qualitative microbiological methods: Accordance and concordance. *International Journal of Food Microbiology*, **79**(3), 175-181.

ANNEXES

Annex A. Additional information for analytical sensitivity

Synthetic controls for PCR based assays

Synthetic controls are fragments or (near) full genomes of RNA or DNA (single or double stranded) that are synthesized in vitro or are a product from cloning a specific gene target of interest. Synthetic controls are an alternative to using material infected with a (pathogenic) virus or viroid, when an organism is difficult to culture, or for laboratories that lack appropriate biosafety containment facilities. Furthermore, the RNA or DNA fragments in the synthetic control can be quantified opening avenues for standardization of the controls. When working with synthetic controls, their nucleotide sequence should match the sequence of the PCR target and at least encompass the amplicon length from the genome of the (representative) pathogen. Markers can be included in the synthetic control (e.g., an additional probe sequence) to distinguish the synthetic control from the true natural target, to rule out contamination by the synthetic control.

Testing of dilution series

Dilution series are prepared by first making a serial dilution series of the target pathogen in extraction buffer followed by spiking each dilution to different aliquots of seed extract. Alternatively, a dilution (with a known concentration) of the target pathogen is initially suspended in seed extract followed by serial dilutions in seed extract.

Three technical replicates need only to be performed for (at least) the lowest three concentrations (higher dilutions) to avoid an undesirable excess of analyses to be performed for the higher concentrations (lower dilutions; Figure A.1). The latter applies provided that the higher dilutions yield a positive result. For example, if three dilutions series are prepared and three technical replicates are tested in the three lowest concentrations (higher dilutions), a total of nine technical replicates are performed. If cut-off values are applied to determine the qualitative test result (positive/negative) of an assay, such a cut-off value needs to be justified and explained, by providing data from the assay for negative (healthy) samples and for samples with a (ideally) low infection level. If negative results are not obtained, further dilutions of the target pathogen should be tested until the first negative results are obtained.

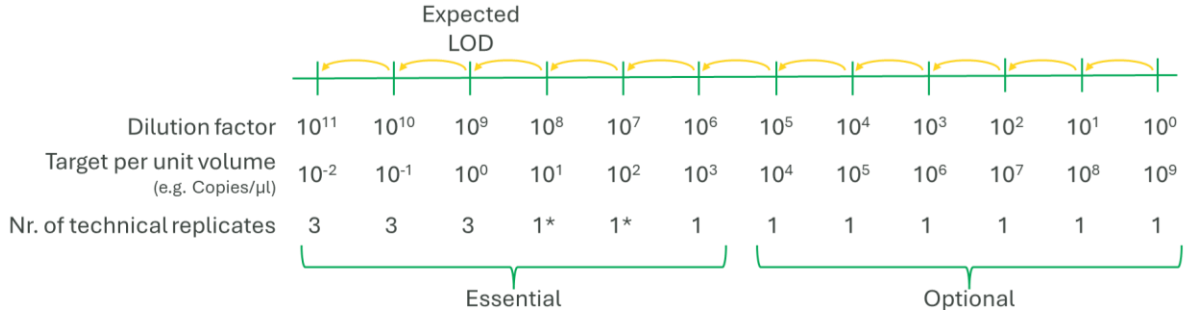

Figure A.1. Example of 10-fold dilution series to be made for analytical sensitivity assessment with the final concentration to be at least two orders of magnitudes lower than the expected detection limit. Three technical replicates are to be included for the lower three concentrations (higher dilutions) per dilution series. *: one technical replicate is essential, but three are recommended.

Inferring the LOD

One may consider performing additional replicates when only one of the replicates gives a negative result for a specific dilution and this is suspected to be an error. Similarly, additional replicates need to be performed when all dilutions for which multiple replicates are performed give a detection confidence of at least 95% (Figure A.2).

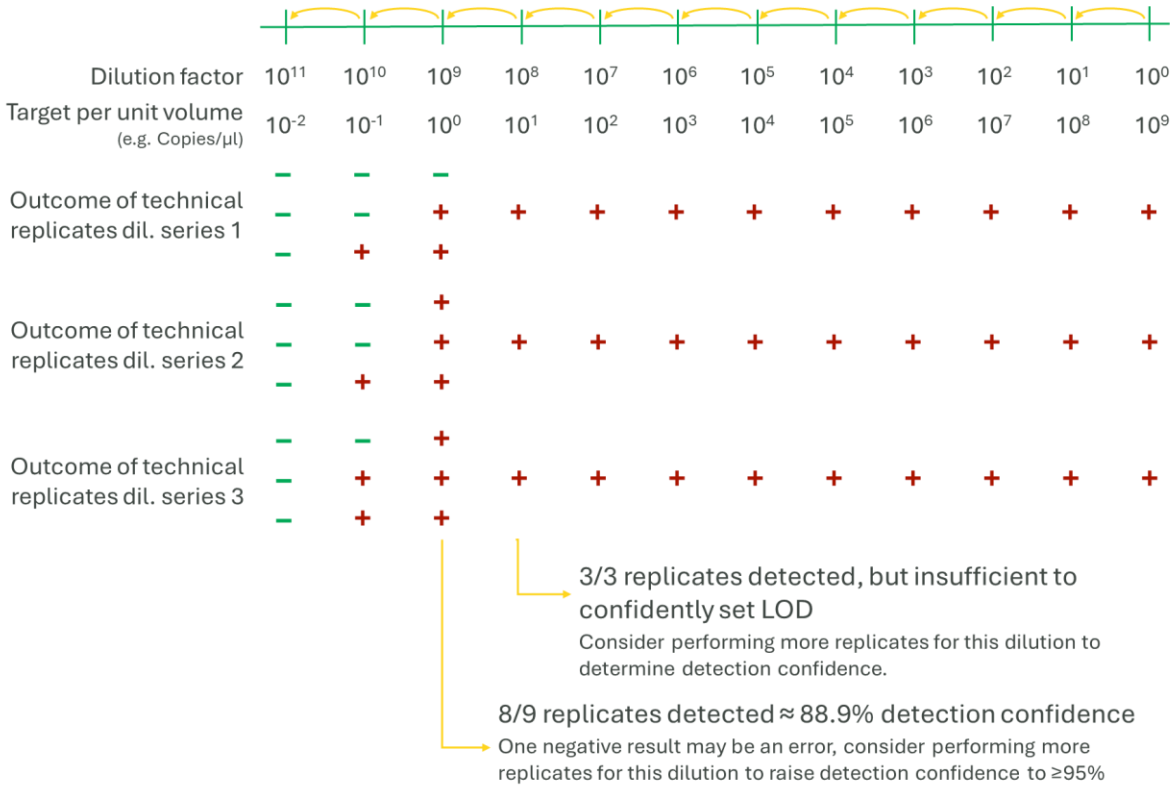

Figure A.2. Example of positive (+) and negative (-) results for (three replicates for the lower three concentrations (higher dilutions)) per dilution series and suggested actions to determine the LOD.

Seed is Life