DOI: 10.1111/epp.12889

EPPO STANDARD ON DIAGNOSTICS

PM 7/122 (2) Guidelines for the organization of interlaboratory comparisons by plant pest diagnostic laboratories

Specific scope: This EPPO Standard specifies:

- The requirements for a plant pest diagnostic laboratory to be able to organize interlaboratory comparisons.
- The requirements for the development and operation of interlaboratory comparisons for plant pest diagnostics.

This EPPO Standard was developed taking into account elements included in ISO 17043 *Conformity assessment – General requirements for proficiency testing* (ISO, 2010) and generally follows the structure of this Standard, however, accreditation for the organization of interlaboratory comparison activities is not covered.

This Standard should be used in conjunction with PM 7/*147 Guidelines for the production of biological reference material* and PM 7/76 *Use of EPPO diagnostic protocols*.

Specific approval: First approved in 2014–09. Revised version approved in 2022–09. The revision was based on the outcome of the EU funded H2020 project VALITEST.[1] Authors and contributors are given in the Acknowledgements section.

1 | INTRODUCTION

Interlaboratory comparisons (ILC), i.e. proficiency testing (PT) and test performance studies (TPS), the latter being also referred to as ring tests or collaborative trials, have become an essential aspect of laboratory practice in all areas of testing and their use is increasing internationally. Proficiency tests may be organized at the request of different stakeholders such as NPPOs, National Reference Laboratories, Regional reference laboratories (such as the European Union Reference Laboratories for EU countries), research institutes, private companies for their laboratories, producers associations. Participation in PT is often a requirement in the accreditation process according to the ISO 17025 (ISO/IEC, 2017). TPS provide added value to the validation process.

Typical objectives for interlaboratory comparisons are presented in Table 1 with a distinction made between those relevant for PT and TPS.

Characteristics of interlaboratory comparison in diagnostics for quarantine pests

The number of laboratories performing diagnostics for quarantine pests is relatively low in EPPO member countries compared to other areas of testing (e.g. chemical analysis). In addition, the number of combinations (target/matrix) is huge (quarantine pests may have many hosts and be present in different parts of a plant or in other material) and in many cases, several tests (e.g. morphological, molecular) are used for a reliable diagnosis. The regional capacity to organize interlaboratory comparison for a specific target/matrix/test combination is consequently limited.

In most ILC the samples are simultaneously distributed to participants for concurrent testing. After completion of the testing, the results are returned to the organizer of the interlaboratory comparison to be evaluated.

Some ILC are sequential and involve the sample (e.g. a mounted slide for an insect) being circulated successively from one participant to the next (i.e. sequential participation), and occasionally circulated back to the organizer for rechecking. This takes place when it is not possible to prepare enough homogenous samples (e.g. limited number of specimens).

All test results in laboratories performing tests for quarantine pests are given in qualitative terms (test positive or negative or inconclusive). It is recognized that some tests will generate quantitative data (e.g. optical density for ELISA, number of cells for IF, Ct values for real-time PCR,[2] measurements for morphological features, etc.). However, such quantitative data is used to assign a qualitative value to the test result (positive/negative/inconclusive) (see section 3.3.4.).

The statistical procedures which may be applied to qualitative data are somewhat limited but recent advances have been made and are presented in this Standard (see section 3.6). The choice of statistical analysis will in part be determined by the type of data generated by the test in question.

Unlike, ISO 17043 *Conformity assessment – General requirements for proficiency testing* (ISO, 2010),

[1]Validation of diagnostic tests to support plant health (Grant Agreement number 773139).

[2]It should be noted that Ct, Cq or Cp, are equivalent.

TABLE 1 Objectives for interlaboratory comparisons.

	PT	TPS
(a) Evaluation of the accuracy of the results produced by laboratories for specific tests and monitoring laboratories' continuing performance	X	(X)[a]
(b) Identification of problems in laboratories which should lead to the initiation of actions for improvement. For example, these may be related to inadequate test procedures, effectiveness of staff training and supervision, or verification of equipment	X	
(c) Establishment of the comparability of tests	(X)	X[b]
(d) Provision of additional confidence to laboratory customers	X	(X)
(e) Identification of interlaboratory differences	X	(X)
(f) Education of participating laboratories based on the outcomes of such comparisons	X	
(g) Establishment of uncertainty levels		X
(h) Evaluation of the performance characteristics of a test[c]	(X)[d]	X

Note: X, main objective; (X), possible objective.

[a]It should be noted that in a TPS, participant laboratories are assumed to be competent to undertake the test in question. However, participation in a TPS may provide independent demonstration of laboratory competence, provided that the results obtained by the laboratory are accurate.

[b]When several tests are included in the TPS.

[c]It should be noted that the performance characteristics obtained for a test are influenced by the composition of the panel of samples. In addition, in interlaboratory comparisons, the number of samples is usually limited, and it is not always possible to evaluate all performance characteristics. Some performance characteristics can be obtained during preliminary studies performed by the TPS organizer.

[d]When several laboratories use the same test in a PT and have been shown to be proficient.

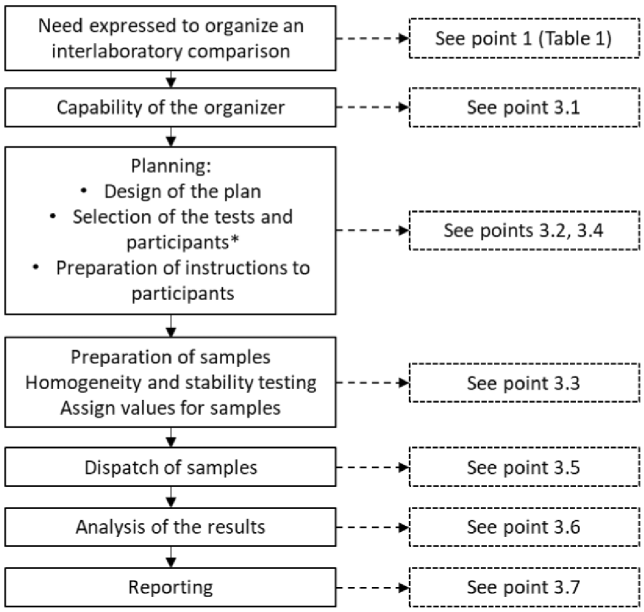

FIGURE 1 Different steps to be followed when organizing an interlaboratory comparison. *Generally for TPS only.

confidentiality requirements are not included in a separate paragraph but underlined in 'Personnel' (Section 3.1.2.), 'Elements to be included in a plan for an interlaboratory comparison' (Section 3.2.1.) and 'Reports' (Section 3.7.).

The flow diagram presented in Figure 1 summarizes the different steps to be followed when organizing an interlaboratory comparison.

2 | DEFINITIONS

Only definitions specifically relevant to this Standard are included. Other definitions are included in PM 7/76.

Participant: laboratory, organization or individual that receives samples for interlaboratory comparison and submits results for review by the organizer of the interlaboratory comparison.

3 | TECHNICAL REQUIREMENTS

3.1 | Capability of the organizer

3.1.1 | General

The feasibility (e.g. technical and logistical) of the ILC should be evaluated by the organizer.

Appropriate technical competence is required for organizers of an interlaboratory comparison as described below:

- *Organizers of PT* should have appropriate technical competence on the organism and the test(s) performed or access to such competence. Outsourcing of such competence should be documented;
- These requirements are also valid for *organizers of TPS*, with the exception that the organizer should not outsource technical competence on the organism and the test(s) performed;
- In both cases (PT and TPS) complementary competence may be required for statistical analysis; in such

situations the organizer should have documented access to such competence;
- In both cases, organizers should have a quality assurance system established according to PM 7/84.

3.1.2 | Personnel

Members of the organizer's staff and, when relevant, sub-contractors, should be designated as responsible for specific tasks on the basis of their competence. A list of the main tasks for the organization of an interlaboratory comparison is given below (more details on these tasks are provided in subsequent sections of this Standard):

Coordination of the interlaboratory comparison:

- Plan the testing schemes (including the distribution of information);
- Supervise the technical operations and the statistical analysis;
- Analyse the results of the interlaboratory comparison and prepare a report;
- Assure confidentiality to the participants;
- Ensure communication with the participants;
- Authorize the issuance of the report.

Technical operations:

- Select appropriate samples and when relevant equipment and reagents;
- Prepare, handle and distribute samples and when relevant equipment and reagents;
- Operate specific equipment (e.g. lyophilizer);
- Conduct measurements to determine stability and homogeneity, as well as assigned values and estimate the measurement uncertainty.

Statistical analysis:

- Select the appropriate statistical approach;
- Operate the data processing system;
- Conduct statistical analysis;
- Evaluate the performance of participants (PT), or of the test(s) (TPS).

Whenever possible tasks should be assigned to more than one person. If the organizer (and when relevant a sub-contractor) of the interlaboratory comparison is also a participating laboratory, the staff involved in the test preparation should not participate in the interlaboratory comparison unless appropriate precaution is taken (e.g. randomization and labeling of the samples are done by another person in the laboratory).

Requirements for competence and training of staff are included in PM 7/84 *Basic requirements for quality management in plant pest diagnostic laboratories* (EPPO, 2021a).

3.1.3 | Equipment and facilities

The organizer should ensure that conditions are appropriate for the preparation and operation of the interlaboratory comparison. This includes facilities and equipment:

- To prepare, test (homogeneity, stability see Section 3.3.3. 'Homogeneity and stability'), handle, store and dispatch samples and when relevant reagents;
- To process data, communicate, and retrieve materials (e.g. stability samples, slides) and records.

The organizer should ensure that cross-contamination between test samples is avoided:

- Environmental conditions in the laboratory and related to equipment that may compromise the operation of the interlaboratory comparison should be monitored (e.g. monitoring insects, spores);
- The work area should be appropriately decontaminated as required;
- Areas where incompatible activities take place are clearly separated, see PM 7/84 *Basic requirements for quality management in plant pest diagnostic laboratories* (EPPO, 2021a) section accommodation of environmental conditions.

The organizer should ensure that the equipment required for the operation of the interlaboratory comparison is appropriately maintained, controlled (e.g. temperature) and calibrated (see PM 7/84). Particular care should be taken for equipment that is used to confirm homogeneity and stability, to define assigned values and to store samples/materials.

Where potentially hazardous samples, chemicals and materials are used, facilities should be adequate to ensure their safe handling, decontamination and disposal.

3.1.4 | Handling of quarantine pests

Acquisition, preparation, handling, distribution (see also 'Data analysis and evaluation of the results of the interlaboratory comparison') and disposal of quarantine organisms should be carried out in accordance with the relevant regulatory requirements (e.g. specific authorization to move regulated material). Specific guidance on handling quarantine organisms has been developed (see Table 1 in EPPO Standard PM 3/64 *Intentional import of organisms that are plant pests or potential plant pests* (EPPO, 2006)) and specific regulations may apply in countries, e.g., Regulation (EU) 2016/2031 (EU, 2016).[3] These conditions also apply to samples that are produced/collected at sites away from the organizer's

[3]The relevant articles in this Regulation are 60 to 64.

permanent facility (e.g. test fields for producing inoculated samples, soil).

3.2 | Planning an interlaboratory comparison

The plan should describe the objectives and provide a detailed description of the interlaboratory comparison including the timing and foreseen duration of the ILC.

When developing such a plan the organizer should identify the critical processes which directly affect the quality of the interlaboratory comparison, and document them in the plan.

If necessary, specific requests from clients of the diagnostic laboratories can be considered when developing a plan (e.g. period or frequency of organization of PT, inclusion of samples at specific level of infestation, etc.).

When developing a plan the organizer can also take into account recommendations or requirements expressed by laboratories involved in plant pest diagnostics (e.g. minimum/maximum number of samples that the laboratory can accept; the period of the year the test(s) should be performed taking into account the seasonal activity of the laboratories, etc.). National regulations may also need to be taken into account.

3.2.1 | Elements to be included in a plan for an interlaboratory comparison

a. Scope of the interlaboratory comparison
- A description of the objective of the interlaboratory comparison in particular to clearly identify if the interlaboratory comparison is a PT or a TPS and to provide relevant details (pest, matrix, test(s), etc.); The scope should include whether the evaluation covers the whole or part of a test (e.g. when nucleic acid is provided, nucleic acid extraction is not part of the evaluation).

b. Elements concerning the organizer
- The name and address of the organizer;
- When relevant, the name, address and affiliation of the coordinator and other personnel involved in the design (e.g. those in charge of delivery of reference material, of not yet published primer sets);
- When relevant, the activities to be subcontracted and the names and addresses of subcontractors involved, if needed. Requirements for proof of competency from the subcontractors should be defined and evidence of competency should be recorded;

c. Elements concerning the participants
- A description of the acceptance criteria to be met for participation (see section 3.4 for more details). For a TPS, it is critical that participants can provide evidence that they are competent to undertake the test in question;
- For PTs information as to whether participation is mandatory or voluntary.
- The number and type of expected participants in the interlaboratory comparison. The number of participants will depend on e.g. the number of laboratories currently performing the test proposed in the interlaboratory comparison, the availability of the matrix/pest to be tested. TPS require a minimum number of participating laboratories (see section 3.6.4);
- The name and address of the participants where the samples should be sent (e.g. the direct address of the laboratory that will perform the tests);
- A description of the procedure for coding the participants in order to keep their identity confidential and known only to persons involved in the operation of the interlaboratory comparison, unless the participant waives confidentiality;
- For continuous proficiency testing schemes,[4] a description of the frequency or dates upon which test samples are to be distributed to participants, the deadlines for the return of results by participants and, where appropriate, the dates on which testing or measurement is to be carried out by participants;
- Information on methods or procedures which participants need to use to prepare the test material and perform the tests or measurements; e.g. which methods have to be used, which materials have to be provided by the participants themselves (sera, media, plants for bioassay, primer sets etc.);
- A description of the information which is to be supplied to participants and the time schedule for the various phases of the interlaboratory comparison; e.g. what will be supplied, when the samples will be sent, how the samples should be handled during the testing, measures that participants need to undertake after finishing the interlaboratory comparison (e.g. destruction of samples, sending the set of samples back to the organizer), etc.;

d. Preparation of samples
- A description of the target(s) or characteristic(s) of interest, including information on what the participants are to identify, measure, or test for in the interlaboratory comparison (e.g. specific morphological features of the pest, counting of bacteria cells, assessment of severity of signs or symptoms caused by the pest, measurement of DNA, presence/absence of PCR products);
- A description of the range of values or characteristics, or both, to be expected from participants for

[4]ISO 17043 has defined a proficiency testing scheme as proficiency testing designed and operated in one or more rounds for a specified area of testing, measurement, calibration or inspection.

the test samples, e.g. 100% correctness expected above the limit of detection or a defined level of detection;
- A list of potential major sources of errors involved in the preparation of interlaboratory comparison, if relevant (e.g. natural infested material with unknown level of infestation, incorrect storage of the sample, time delay in testing procedure, source of ingredients for agar media);
- A description of the requirements for the production, quality control, storage and distribution of samples, e.g. special attention is necessary if the tests are based on isolation of living organisms;
- A description of the procedure to be used for coding and labelling the samples and randomly assigning the numbered samples to the laboratories;
- Description of the test or measurement methods to be used for the homogeneity and stability testing and, where applicable, to determine their biological viability; e.g. time scale for different tests, storage conditions for the materials during the tests, instructions for counting, measuring, isolation of the organisms and confirmation;

Guidelines for most of these descriptors can be found in PM 7/147 *on Guidelines for the production of biological reference material*.

e. Dispatch of samples
- An identification of requirements related to the packaging and delivery of samples to participants that may affect the interlaboratory comparison (e.g. stability, biosafety, specific regulations)
- A description of the actions to be taken in the case of losing or damaging samples;

f. Analysis of ILC results
- Preparation of any standardized results forms to be used by participants, e.g. tables with the dates of arrival of the samples, conditions of the samples at arrival, conditions of storing the samples until testing, dates of starting the tests, dates of finishing the test (e.g. bioassays); sources of consumables (if relevant for the results), results of the tests. Example of a standardized results form for TPS organization is available in Vučurovič et al., (2022).
- A description of information to be reported by the participants about how they performed the test (ease of use, equipment, deviations to the test under evaluation);
- A description of the statistical analysis to be used;
- A description of the procedure to be followed if the same sample is consistently not giving the expected result during the interlaboratory comparison (e.g. investigations to be initiated and actions to be taken, number of participants failing).

g. Reporting
- A description of the extent to which participant results, and the conclusions that will be based on the outcome of the interlaboratory comparison, are to be made public.

3.2.2 | Elements only applicable for PT

When organizing a PT, where different tests can be used, it is strongly recommended that the organizer ensures that:

a. The organizer has a procedure regarding comparison of results obtained by different tests;
b. The organizer is aware of which different tests are technically equivalent, and
c. The organizer will assess participants' results using these tests accordingly.

3.2.3 | Elements only applicable for TPS

When organizing a TPS, the selection of the test(s) to include in the TPS should be based on a list of criteria to be fulfilled by a test based on the scope of the TPS. Example of a list of criteria is available in Vučurovič et al., (2022). It may also include the analysis of available validation data and if needed the generation of new validation data in preliminary studies (see Vučurovič et al., 2022 for an example). The test selection process should be documented.

3.3 | Preparation of samples and when relevant equipment and reagents

3.3.1 | General considerations for the preparation of the samples

Guidelines for the production of reference material can be found in PM 7/147 (EPPO, 2021b).

Sufficient material needs to be prepared in order to allow for:

- The number of participants (as established at the planning phase see Section 3.2. 'Planning an interlaboratory comparison');
- Performance of homogeneity and stability testing (as established at the planning phase see section 3.2. 'Planning an interlaboratory comparison');
- Replacement of materials that are lost or damaged during distribution/handling (preparation of a number of extra sample set(s)).

It may also be useful to prepare material to be used after the interlaboratory comparison e.g. as training aids, reference material or controls.

3.3.2 | Composition of the panel of samples

The composition of the panel of samples (e.g. the type of samples, the number and diversity of target and non-target samples, the concentrations and the number of replicates) depends on many constraints/conditions:

- Type of material to be used: an important aspect of ILC is that the sample matches (as closely as possible) the materials encountered in routine testing. This includes the matrix (e.g. host), target (e.g. pest) and concentration (e.g. infection level). For some pest/host combinations this may be a challenge (e.g. phytoplasmas). Examples of types of samples are provided in Table 2.

- The selection of samples is also linked to the scope of the interlaboratory comparison and to the statistical requirements needed to evaluate the performance criteria of interest. The panel of samples may include:
  - Target samples to evaluate the ability of the test(s)/laboratories to detect different specimens/strains/isolates/populations of the target pest, if relevant in a specific or different matrices and low level of infestation: these may cover the diversity of the target and different concentrations.
  - Non-target samples to evaluate the ability of the test(s)/laboratories not to detect the pest in samples which are free of the target pest: these may include different matrices or non-target organisms (known to cross react with the target pest).
  - If relevant, replicates to evaluate the ability of the test(s)/laboratories to deliver repeatable results.

Note that a sample can be used to assess several performance criteria. For TPS, for a meaningful statistical analysis the panel of samples should ideally include at least 2 replicates per sample for the evaluation of analytical sensitivity and, if relevant, five dilution points in triplicate. Examples of a panel of samples for a PT and a TPS are provided in Tables 3 and 4.

- Finally, the selection of the samples may also be influenced by knowledge on the pest (survival, detectability), resources (financial or personnel) and technical limitations (e.g. availability of reference material).

3.3.3 | Homogeneity and stability

The organizer should take care that the samples prepared and used for the interlaboratory comparison are as homogenous and as stable as possible, because this can affect the evaluation of the participants or the test performance.

The assessment of homogeneity and stability should be performed by the same laboratory (generally the organizer) using the same analytical method and measuring the same characteristic of the samples. The test used for homogeneity and stability testing should be a standardized (or validated) test that can be implemented in the laboratory.

The procedures for the assessment of homogeneity and stability should be documented.

TABLE 2 Example of types of samples.

	Virology	Bacteriology	Mycology	Nematology	Entomology	Botany	Phytoplasmology
Matrix	A	A	A	A	A	A	A
Substrate (e.g. soil and water)	A	A	A	A	A	A	NA
Cultures	NA	A	A	NA	NA	NA	NA
Specimens	NA	NA	NA	A	A	NA	NA
Vectors	A	A	A	A	NA	NA	A
Slides	NA	A	A	A	A	A	NA
DNA/RNA	A	A	A	A	A	A	A

Note: A, applicable; NA, not applicable.

TABLE 3 Example of a panel of 10 samples for the evaluation of accuracy, rate of true positives and rate of true negatives in a PT. (adapted from ANSES, FR).

Type of samples	Number of biological samples	Number of replicates	Dilution	Note	Performance criteria evaluated
Non-target sample	5	1	Not applicable	Different matrices and/or non-target organisms. For a meaningful statistical analysis, negative samples should be independent from each other (i.e. not an aliquot of the same sample)	Accuracy, rate of true negatives
Target sample	5	1	Different concentrations above LOD	For a meaningful statistical analysis, positive samples should be independent from each other (i.e. not an aliquot of the same sample)	Accuracy, rate of true positives

TABLE 4 Example of a panel of 25 samples for the evaluation of analytical sensitivity, diagnostic sensitivity, diagnostic specificity, repeatability and reproducibility in a TPS. This panel of samples, proposed in the framework of the VALITEST project, was considered as a good balance between statistical power and practicality for TPS organizers (see Brotaux et al., 2021; Massart et al., 2022; Vučurovič et al., 2022).

Type of samples	Number of biological samples	Number of replicates	Dilution	Note	Performance criteria evaluated
Non-target sample	3	2	Not applicable	For a meaningful statistical analysis, negative samples should be independent from each other (i.e. not an aliquot of the same sample)	Diagnostic specificity, repeatability, reproducibility
Target sample	2	2	One positive sample has a medium concentration and the other one has a low concentration of the target pest (i.e. close to the limit of detection)	For a meaningful statistical analysis, positive samples should be independent from each other (i.e. not an aliquot of the same sample)	Diagnostic sensitivity, repeatability, reproducibility
Target sample	1	3	Five dilution points (the range of the dilution points should include dilutions above and below the limit of detection)	For a meaningful statistical analysis, positive sample used for the dilution series should be independent to the other positive samples.	Analytical sensitivity, repeatability, reproducibility

In some cases, it is not feasible for samples to be subjected to homogeneity and stability testing. Such cases would include, for example, when limited material is available to prepare samples.

Samples which are demonstrated to be not sufficiently homogeneous or stable during the homogeneity and the stability testing respectively, should not be used to evaluate the performance of the laboratory (for a PT) or the performance of the test (for a TPS).

3.3.3.1 | Homogeneity testing

The assessment of homogeneity should generally be performed after the samples have been packaged in the final form and before distribution to participants. Homogeneity can be demonstrated prior to packaging where no influence of packaging is reasonably expected. On some occasions, homogeneity testing cannot be carried out prior to dispatch for practical, technical or logistical reasons (e.g. tests that require isolation of viable organisms by incubation, when documented evidence is available from previous homogeneity testing on similar samples prepared by the same procedure). On such occasions parallel testing of extra samples should be conducted and participants should be informed when invited to participate in the interlaboratory comparison (section 3.4).

3.3.3.1.1 | Types of samples. In plant pest diagnostics the evaluation of homogeneity depends on the type of pest/matrix to be tested. Examples are described in Table 5 (for numbers of samples see Section 3.3.3.1.2 'Number of samples to be included in homogeneity testing').

3.3.3.1.2 | Number of samples to be included in homogeneity testing. Current available guidelines (e.g. ISO 13528, ISO, 2015) recommend to test a minimum of 10 randomly chosen samples for each independent sample provided within the panel (i.e. for each pest/matrix/infestation level, including negative samples). Some laboratories use the square root (rounded up) of the total number of samples. Based on current experience and depending on the method used it is recognized that this is not always feasible because of the multiple pest/matrix/infestation level combinations. Therefore, the number of samples included in the homogeneity testing may be reduced if suitable data are available from previous homogeneity testing on similar samples prepared by the same procedures or according to the expertise of the organizer.

The choice (including the rationale) of the number of samples should be documented.

3.3.3.1.3 | Statistical analysis of homogeneity testing. When quantitative data is generated (e.g. optical density for ELISA, number of cells for IF, Ct values for real-time PCR, measurements for morphological features), statistical analysis of homogeneity testing is commonly performed by comparing the sample mean and the coefficient of variation obtained during the homogeneity testing to expected levels. Other statistical approaches, such as comparison of the standard deviation of the samples (potentially expanded with F test) to the proficiency testing standard deviation (ISO 13528, ISO, 2015), or F test if sufficient data is available (Guide ISO 35, ISO, 2017), may be used depending on the data available and the experience of the laboratory.

When quantitative data is unavailable, the statistical analysis is limited to the qualitative results. In any case (with or without quantitative data), all qualitative results obtained during the homogeneity testing should correspond to the assigned value.

TABLE 5 Evaluation of homogeneity depending on the type of pest/matrix to be tested.

Type of pest/matrix to be tested	Examples and exceptions of how homogeneity should be tested
Slides and mounted specimens of e.g. fungal structures	Homogeneity can be ensured by documented verification that each diagnostic feature of the pest is present on each slide before dispatch.
DNA/RNA	The process for the preparation of RNA/DNA solutions and aliquots thereof ensures homogeneity; a number of samples should be tested before dispatch.
Spiked matrix (e.g. potato extract spiked with known dilutions of bacteria, soil spiked with nematodes, wheat grain spiked with one bunted kernel for *Tilletia indica*)	The process of preparation of samples generally ensures homogeneity; a number of samples should be tested before dispatch whenever technically possible.
Artificially inoculated matrix (e.g. *Phytophthora ramorum* on leaves, mechanical inoculation of test plants)	A number of samples should be tested before dispatch whenever technically possible.
Ground or mixed freeze dried material (e.g. plant material infested with viruses/viroids/phytoplasmas)	The process of sample preparation aims to ensure homogeneity; a number of samples should be tested before dispatch whenever technically possible.
Naturally infested material	Samples should be taken from a known infested source and randomized. A number of samples should be tested before dispatch whenever technically possible.
Samples that mimic infested material	In cases where naturally infested material is difficult to obtain or homogeneity difficult to ensure (e.g. *Chalara fraxinea* and ash twigs), laboratories can combine in one sample the matrix and the pest (e.g. a plug of a pure culture associated with the matrix in a tube). A number of samples should be tested before dispatch whenever technically possible.
Cultures (e.g. on growing media, extracts, suspensions)	A number of samples should be tested before dispatch whenever technically possible.
Traps (pheromone traps for insects)	Homogeneity is less of an issue for these types of samples because the objective is to determine if the target is present or not. Consequently the only requirement is that the location of the target on each trap should be documented before dispatch.

3.3.3.2 | Stability testing

Samples should be demonstrated to be sufficiently stable to ensure that they will not undergo any significant change throughout the conduct of the interlaboratory comparison, including storage and transport conditions. When required, stability testing should be conducted in conditions that mimic transport and storage conditions. As an alternative, samples can be sent to the participant with the most challenging environmental or transport conditions and returned unopened for testing. For TPS critical reagents are usually also provided to participants in addition to samples. Stability of those reagents which have an influence on the outcome of the test should be verified following the same procedures.

A stability check may be performed on samples held by the organizer. This should be done after the deadline for performing analyses by the participants, in order to verify that the stability of samples has been maintained throughout the interlaboratory comparison. Depending on the type of sample, additional stability testing may also be needed between the dispatch of samples and the deadline for performing the analysis by the participants.

Some pest stages are known to be stable over long periods (e.g. *Globodera* spp. cysts, or fungal spores) in such case stability testing is not needed.

3.3.3.2.1 | Types of samples. In plant pest diagnostics the evaluation of stability depends on the type of pest/matrix to be tested. Examples are described in Table 6.

3.3.3.2.2 | Number of samples to be included in stability testing. Current available guidelines (e.g. ISO 13528, ISO, 2015) recommend testing a minimum of 2 randomly chosen samples for each independent sample provided within the panel (i.e. for each pest/matrix/infestation level including negative samples). Based on current experience and depending on the method used, it is recognized that it is not always feasible because of the multiple pest/matrix/infestation level combinations.

Therefore, the number of samples included in the stability testing may be reduced if suitable data are available from previous stability testing on similar samples prepared by the same procedures or according to the expertise of the organizer.

TABLE 6 Evaluation of stability depending on the type of pest/matrix to be tested.

Type of pest/matrix to be tested	Examples and exceptions of how stability should be tested
Slides and mounted specimens of e.g. fungal structures	Testing for stability is usually not needed for correctly mounted slides except if some staining procedures are used (e.g. immunofluorescence) and a number of slides should then be examined during or after the interlaboratory comparison. Integrity has to be ensured by appropriate packaging.
DNA/RNA, spiked matrix, artificially inoculated matrix, freeze-dried infested material ground/mixed with naturally infested material, samples that mimic infested material, and cultures	A number of samples should be tested at least after the interlaboratory comparison.
Traps (pheromone traps for insects)	Stability should be ensured by appropriate packaging to avoid damage to the specimens, and minimizing transport time. A number of traps should be examined during or after the interlaboratory comparison.

The choice (including the rationale) of the number of samples should be documented.

3.3.3.2.3 | Statistical analysis of stability testing. When quantitative data is generated (e.g. optical density for ELISA, number of cells for IF, Ct values for real-time PCR, measurements for morphological features), statistical analysis of stability testing is commonly performed by comparing the sample mean and the coefficient of variation obtained during the stability testing to expected levels. Other statistical approaches may also be used depending on the data and the experience of the laboratory, such as difference between stability and homogeneity means based on size relative to proficiency testing standard deviation (ISO 13528, ISO, 2015), or t test of stability vs. homogeneity means if sufficient data is available.

When quantitative data is unavailable, the statistical analysis is limited to the qualitative results. In any case (with or without quantitative data), all the qualitative results obtained during the stability testing should correspond to the assigned value.

3.3.4 | Assigned values for samples

The organizer has to define/establish assigned values for samples, i.e. value attributed to a particular property of an interlaboratory test sample.

In the plant health field, assigned values correspond to the expected result of the test (pest present or absent, concentration of the pest, morphological characteristics of the specimen, etc.). In some cases the assigned value may be declared as inconclusive (e.g. samples yielding an OD between positive and negative threshold in ELISA, specimens presenting overlapping morphological characters).

The organizer should have previously described in a procedure how this value would be assigned e. g.:

- Known values (spiked sample) this is the most common situation, usually qualified as low/medium/high or present/absent;
- Known values (naturally infested plants, samples) in cases where the infestation status of a plant (e. g. nursery tree) is known and the target pest was previously confirmed by an expert, using a validated test considered to be highly accurate and comparable to tests in use, usually qualified as low/medium/high or present/absent;
- Reference values obtained from comparison with reference material
- Consensus values from expert participants/reference laboratories, provided they demonstrated their expertise in diagnosis of the target pest, using a validated test considered to be highly accurate and comparable to tests in use (for a TPS, this approach may be used);
- Consensus values from participants in the PT. Rules for definition of these values from the participants results should be defined: statistical methods, outliers effect (e.g. in virology interlaboratory comparison may assign the values this way).

Uncertainty of assigned values should be defined. For known values uncertainty is low or almost 0 (depending on homogeneity and stability), but for consensus values from participants, the uncertainty might be high depending of the competence of the participating laboratories.

The organizer should have a policy regarding the disclosure of assigned values. The policy should ensure that participants cannot gain advantage from early disclosure. Disclosure should happen as late as possible.

3.4 | Instructions for participants

Announcement

To announce an invitation for participation in an interlaboratory comparison test, the organizer may:

- Consult relevant professional network(s),
- Consult the EPPO Database on diagnostic expertise in order to target relevant laboratories,
- Use social media.

Information required in the invitation.

The organizer should make possible applicants fully aware of all conditions and demands for participation in the interlaboratory comparison test offered. This includes:

- General information:
  - Whether the interlaboratory comparison is organized for PT or TPS,
  - The test organism(s) with a description of the scope (sample type, matrix),
  - If relevant, the method(s) and/or test(s) to be evaluated,
  - The timeline.
- The acceptance criteria to be met for participation:
  - Possible phytosanitary requirements such as official documents required (e.g. import permit, letter of authority) and facilities required to avoid unintended spread of the test organism.
  - Possible selection criteria between applicants (see Vučurovič et al., 2022 for an example of a list of criteria for selection of TPS participants), and any geographic limitation for participation or any pre-qualification requirements involved (see Section 3.2.1. 'Elements to be included in a plan for an interlaboratory comparison').
  - Equipment and other facilities required.
  - Chemicals involved (for evaluation of national/local safety and transport regulations). It should be possible to judge whether the samples may be legally sent to participants and are not listed in the export control list.
- Information related to the contract and technical information:
  - Description of the test(s) to be used; in the case of a TPS including the full protocol (or link to the reference) with contact details for clarification if needed and the policy on possible changes to any part of the protocol during the test period (see Vučurovič et al., 2022 for an example of TPS technical sheet).
  - Definition of the rights and obligations of the parties involved, and detailed description of the timelines and the conditions of participation for the interested laboratories (see Vučurovič et al., 2022 for an example of TPS participants' contract).
  - Any item (e.g. portable thermocycler) which may be sent with the samples with essential details.
  - Any fees and other expenses involved as well as costly chemicals to be purchased by the participants and any excessive delay to be expected for the receipt of specialized equipment or chemicals required for the study.
  - It should also be mentioned if the remainder of the sample can be kept and use by the participants and to what end (in relation with property rights and/or Nagoya Protocol restrictions) or if any of the items/samples need to be returned to the organizer and who will pay for this.
  - Information about whether the report/results will be made public and to which extent.
  - The consent to the participation in the interlaboratory comparison (including the requirement to provide the direct address of the laboratory that will perform the tests).
  - Confidentiality agreement, if necessary.

3.5 | Packaging, labelling and distribution of interlaboratory comparison test samples

When distributing the samples for the interlaboratory comparison precautions should be taken to maintain the quality and integrity of the material.

For example:

- Care should be taken that there is no contamination or mixing of samples.
- When distributing biological material such as cysts or bacteria or infested material (naturally or artificially), extremes of hot and cold temperatures should be avoided if these were shown to be detrimental to the sample. In such situations appropriate distribution means should be selected.
- The time of the year should be taken into account e.g. summer or winter, public holidays, in order to make allowances for extremes of temperature and packages waiting for long periods before being unpacked.

There is specific packaging material for sending out quarantine material (bio hazard – symbol) or DNA/RNA solutions.

A reliable parcel service that guarantees the delivery of the samples to the participants in a similar timeframe should be used.

In addition the organizer should ensure that:

- The interlaboratory comparison samples are clearly labelled in such a way that the label will remain legible throughout the interlaboratory comparison;
- The package contains all the relevant paperwork e.g. instructions, plant health licence, unique laboratory identifier and samples identifiers;
- The package and its content conform to international safety and transport requirements.

A procedure should be followed to enable confirmation of delivery of interlaboratory comparison samples.

3.6 | Data analysis and evaluation of the results of the interlaboratory comparison

Data analysis methods should be appropriate to the nature of data (see section 1).

3.6.1 | Exclusion of data (for TPS)

Some results may introduce bias and affect the outcome of the interlaboratory comparison. Such results should be identified and documented. Consideration should be given to exclude them from the analysis of the interlaboratory comparison depending on the reasons for such results and the objective of the TPS. Exclusion may concern single data points or the whole data set obtained from a laboratory or for a given sample.

Examples of TPS data that may need to be excluded are given below:

- Datasets for which the results of the controls were not as expected;
- Missing or inconclusive results;
- Results obtained by laboratories that have deviated from the original protocol if the deviation invalidates the results.

Outlier results may also be excluded, i.e. results that fall outside of the expected ranges or that appear to be inconsistent with the other results observed in a specific dataset. However, it may be difficult to identify the cause of outliers (problem with a laboratory or with the test not performing well) and care should be taken when excluding them.

3.6.2 | Comparison of participants' results with assigned values

The standard way of analysing qualitative results for each participant is based on the numbers of true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN), as presented in Table 7. The rules for the interpretation of inconclusive results may vary depending on the interlaboratory comparison and should be clearly described (e.g. whether they are excluded from the analysis or interpreted as false positives or false negatives).

The values from Table 7 may be used to calculate different performance criteria in order to evaluate the performance of the laboratory (for PT) or the performance of the test (for TPS) (Table 8). Appendix 1 describes how to calculate the different performance criteria and the associated confidence intervals which are used to assess the quality of the estimation of each parameter.

3.6.3 | Evaluation of the performance of laboratories (PT)

In PT, the performance of the laboratory is evaluated in terms of the levels required for the performance criteria (e.g.: 100% for accuracy). These levels are defined independently of the PT results.

3.6.4 | Evaluation of the performance of (a) test(s) (TPS)

In TPS, the performance of the test(s) is evaluated either by the simple characterization of the performance criteria of the test, or by comparison to other tests, or by comparison to the performance characteristics obtained for the test during a single laboratory validation.

Note that the number of laboratories participating in a TPS affects the estimation of the reproducibility of the test(s) and the robustness of the calculation of other performance criteria and of their confidence intervals. Therefore, TPS require a minimum number of participating laboratories (ideally a minimum of 10 valid laboratory

TABLE 7 Definition of true positive (TP), true negative (TN), false positive (FP) and false negative (FN). True or false (positive/negative) may also be described as (positive/negative) agreement or deviation with assigned values, particularly in validation studies see PM 7/98 (appendix 6 table A9) (EPPO, 2021b).

Participant result	Assigned value = Positive	Assigned value = Negative
Result obtained is positive	TP = True positive	FP = False positive
Result obtained is negative	FN = False negative	TN = True negative

TABLE 8 Performance criteria commonly used to analyse the results of ILC.

	PT	TPS
Accuracy / % of conforming results	X	X
Rate of true positives	(X)	X – also called diagnostic sensitivity
Rate of true negatives	(X)	X – also called diagnostic specificity
Repeatability	(X)	X
Reproducibility	NA	X
Analytical sensitivity	NA	X

Note: X: commonly used performance criteria; (X): performance criteria not commonly used; NA., not applicable.

data sets). The inclusion of more laboratories will increase the reliability of the estimations of the performance characteristics. However, it is recognized that this may be a constraint in plant pest diagnostics (in particular because of limited available resources). If conclusions are based on the results of <10 laboratories, they should be presented with caution and appropriate warnings. This is because the resulting performance characteristics will be subject to an increased uncertainty with possible incorrect estimations of the confidence intervals due to the limited number of laboratories (Brotaux et al., 2021; Massart et al., 2022).

3.7 | Reports

The reports should be clear and comprehensive and include data covering the aggregated results of all participants as well as individual results for PT. The data are anonymized but each laboratory should receive information allowing identification of its results.

Reports should include the following:

- The name and contact details of the organizer
- The name of staff involved and respective responsibility
- Identification of person(s) authorizing the report
- The date of issue and status (e.g. preliminary, interim, or final) of the report
- Page numbers and total number of pages of the report
- A statement of the extent to which results are confidential
- Unique identification of the report and the interlaboratory comparison
- A clear description of the test samples used, including necessary details of the sample preparation and homogeneity and stability assessment
- The participants' results (when possible, otherwise a summary of the results, e.g. in tabulated or graphical form, can be supplied). This is required for PT.
- When relevant statistical data and summaries, including assigned values and range of acceptable results and graphical displays
- Procedures used to establish any assigned value
- When relevant details of the metrological traceability and measurement uncertainty of any assigned value, e.g. for quantitative analysis or qualitative analysis including measurements
- Procedures used to establish the standard deviation for proficiency assessment, or other criteria for evaluation
- Assigned values and summary statistics for the test(s) evaluated (TPS) or for the tests used by each group of participants (if different tests are used by different groups of participants) (PT), when relevant
- Comments on participants' performance by the organizer and technical advisers
- Information about the design and implementation of the interlaboratory comparison
- Procedures used to statistically analyse the data
- Advice on the interpretation of the statistical analysis and
- Comments or recommendations, based on the outcomes of the interlaboratory comparison.

Reports should be made available to participants within planned within the planned timescales. The organizer should have a policy for the use of reports by individuals and organizations (for TPS especially, results might be for example the property of a research project).

4 | FEEDBACK ON THIS DIAGNOSTIC STANDARD

If you have any feedback concerning this Diagnostic Standard, or any of the tests included, or if you can provide additional validation data for tests included in this protocol that you wish to share please contact diagnostics@eppo.int

5 | PROTOCOL REVISION

An annual review process is in place to identify the need for revision of diagnostic protocols. Protocols identified as needing revision are marked as such on the EPPO website. When errata and corrigenda are in press, this will also be marked on the website.

ACKNOWLEDGEMENTS

The revision of this Standard was drafted by an Expert working group composed of Dreo T (NIB, SI), Ferreti L (CREA, IT), Inghelbrecht S (ILVO, BE), Mouaziz H (ANSES, FR), Raaymakers T (NRC-NPPO, NL). The Standard was reviewed by the Panel on Diagnostics and Quality Assurance.

REFERENCES

Agresti A, Coull BA (1998) Approximate is better than "Exact" for interval estimation of binomial proportions. *American Statistician* 52, 119–126.

Brotaux Y, Chabirand A, Lebas B, Massart S (2021) VALITEST deliverable 2.1 Guidelines for the revision of the EPPO Standards PM 7/98 and PM 7/122 for validation studies (including Test Performance Studies) available at https://doi.org/10.5281/zenodo.7143214

EPPO (2021a) PM 7/84 Basic requirements for quality management in plant pest diagnostic laboratories. *Bulletin OEPP/EPPO Bulletin*, 51, 457–467.

EPPO (2021b) PM 7/98 (5) Specific requirements for laboratories preparing accreditation for a plant pest diagnostic activity. *Bulletin OEPP/EPPO Bulletin*, 51: 468–498.

EPPO (2006) PM 3/64 Intentional import of organisms that are plant pests or potential plant pests. *Bulletin OEPP/EPPO Bulletin*, 36, 191–194.

EPPO (2018) PM 7/76 (5) Use of EPPO Diagnostic Standards. *Bulletin OEPP/EPPO Bulletin*, 48: 373–377.

EU (2016) Regulation (EU) 2016/2031 of the European Parliament of the Council of 26 October 2016 on protective measures against pests of plants, amending Regulations (EU) No 228/2013, (EU)

No 652/2014 and (EU) No 1143/2014 of the European Parliament and of the Council and repealing Council Directives 69/464/EEC, 74/647/EEC, 93/85/EEC, 98/57/EC, 2000/29/EC, 2006/91/EC and 2007/33/EC.

ISO (2010) *ISO 17043 Conformity Assessment – General Requirements for Proficiency Testing*. ISO/IEC, International Organization for Standardization, Geneva (CH), 39 pp.

ISO/IEC (2017) Standard 17025. General requirements for the competence of testing and calibration laboratories, Available at www.iso.org [accessed 26 September 2022].

ISO (2017). *ISO Guide 35: Reference Materials – General and Statistical Principles for Certification*, 3rd Edn. International Organization for Standardization, Geneva (CH).

ISO (2015) *ISO 13528 Statistical Methods for Use in Proficiency Testing by Interlaboratory Comparisons*. ISO/IEC, Geneva (CH).

Langton SD, Chevennement R, Nagelkerke N, Lombard B (2002) Analysing collaborative trials for qualitative microbiological methods: accordance and concordance. *International Journal of Food Microbiology*, 79: 175–181.

Massart S, Brostaux Y, Brabarossa L, Batlle A, César V, Dutrecq O, Fonseca F, Guillem R, Komorowska B, Olmos A, Steyer S, Wetzel T, Kummert J, Jijakli M H (2009a) Inter-laboratory evaluation of two reverse-transcriptase polymeric chain reaction-based methods for the detection of four fruit tree viruses. *Annals of Applied Biology*, 154: 133–141.

Massart S, Brostaux Y, Brabarossa L, César V, Cieslinska M, Dutrecq O, Fonseca F, Guillem R, Laviña A, Olmos A, Steyer S, Wetzel T, Kummert J, Jijakli MH (2009b) Inter-laboratory evaluation of a duplex RT-PCR method using crude extracts for the simultaneous detection of Prune dwarf virus and Prunus necrotic ringspot virus. *European Journal of Plant Pathology*, 122: 539–547.

Massart S, Lebas B, Chabirand A, Chappé A-M, Dreo T, Faggioli F, et al. (2022) Guidelines for improving statistical analyses of validation datasets for plant pest diagnostic tests. *EPPO Bulletin*, 52, 419–433.

Wehling P, Labudde RA, Brunelle SL & Nelson MT (2011) Probability of Detection (POD) as a statistical model for the validation of qualitative methods. *Journal of AOAC International* 94, 1–11.

Vučurović A, Mehle N, Anthoine G, Dreo T, Ravnikar M (2022) *Critical points for the organisation of test performance studies in microbiology: plant pathogens as a case study*. Volume 12. Springer, Cham, Switzerland. Available on https://link.springer.com/book/10.1007/978-3-030-99811-0

How to cite this article: (2022) PM 7/122 (2) Guidelines for the organization of interlaboratory comparisons by plant pest diagnostic laboratories. *EPPO Bulletin*, 52, 604–618. Available from: https://doi.org/10.1111/epp.12889

APPENDIX 1 - PERFORMANCE CRITERIA USED TO EVALUATE TESTS AND/OR LABORATORIES

Table A1 lists the performance criteria that were recommended for TPS in the framework of the VALITEST project (Brotaux et al., 2021) and that are commonly used by diagnostic laboratories. Accuracy, rate of true positives and rate of true negatives are also relevant for the analysis of PT results (see Table 8). Other performance criteria may be used depending on the needs of the interlaboratory comparison organizer (Brotaux et al., 2021). Methods of calculation, proposed in the framework of VALITEST for each performance criterion, are reported in Table A1 and in sections 1 to 3 below but other methods may be used to evaluate those criteria (e.g. see PM 7/98).

For TPS results of controls should not be used for the calculation of values for performance criteria.

To assess the quality of the estimation of performance characteristics, confidence intervals can be calculated. Confidence intervals are particularly useful to compare the performance of different tests, in particular when the estimations of the performance characteristics are based on a different number of datasets. Methods of calculation, proposed in the framework of VALITEST for each performance criterion (Brotaux et al., 2021), are reported in Figure A1 and in sections 1 to 3 below but other methods may be used.

1. *Accuracy, rate of true positive (diagnostic sensitivity) and rate of true negative (diagnostic specificity)*

The values from Table 7 may be used to evaluate accuracy, rate of true positive (diagnostic specificity) and rate of true negative (diagnostic sensitivity) using the formulas described in Table A1.

A test with a high diagnostic sensitivity strongly indicates a high probability of detecting a pest when it is present in a sample while a test with a high diagnostic specificity has a high probability of correctly giving a

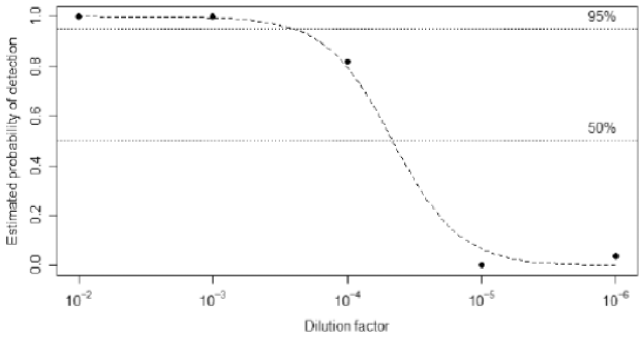

FIGURE A1 Estimation of the analytical sensitivity using probability of detection (y axis). The dotted lines indicate 50% and 95% probability of detection. A limit of detection of $10^{-3.6}$ is observed for a probability of detection of 95%.

TABLE A1 Performance criteria used to evaluate the results of ILC.

Performance criteria (EPPO Standard PM 7/98 (EPPO, 2021b))	Method of calculation of the parameter	Notes	Confidence interval's method
Accuracy (%)	$\frac{(TP+TN)}{N} \times 100$ see section 1		Agresti-Coull method (Agresti & Coull, 1998)
Rate of true positives (%) (Diagnostic sensitivity[a])	$\frac{TP}{(TP+FN)} \times 100$ see section 1	Defined in EPPO Standard PM 7/76 (EPPO, 2018) Note: the rate of false negative results obtained by the laboratory can be calculated as follows: $\left(1-\frac{TP}{(TP+FN)}\right) \times 100$	Agresti-Coull method (Agresti & Coull, 1998)
Rate of true negatives (%) (Diagnostic specificity[a])	$\frac{TN}{(TN+FP)} \times 100$ see section 1	Defined in EPPO Standard PM 7/76 (EPPO, 2018) Note: the rate of false positive results obtained by the laboratory can be calculated as follows: $\left(1-\frac{TN}{(TN+FP)}\right) \times 100$	Agresti-Coull method (Agresti & Coull, 1998)
Repeatability	e.g. using accordance see section 2	Defined in EPPO Standard PM 7/76 (EPPO, 2018)	Derived from bootstrap standard errors (Langton et al., 2002)
Reproducibility	e.g. using concordance see section 2	Defined in EPPO Standard PM 7/76 (EPPO, 2018)	Derived from bootstrap standard errors (Langton et al., 2002)
Analytical sensitivity	e.g. using probability of detection see section 3	Defined in EPPO Standard PM 7/76 (EPPO, 2018)	Generalized linear models

Note: Total number of samples (*N*), true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN).

[a]Diagnostic sensitivity and diagnostic specificity are often used to report the performance of tests.

negative result for a pest when it is truly absent. In addition, the calculation of diagnostic sensitivity and diagnostic specificity for each laboratory can be useful to spot laboratories with discordant results compared to the other participating laboratories, i.e. to identify outliers.

For the calculation of diagnostic sensitivity and diagnostic specificity, results obtained from independent samples should be used. Samples used for serial dilution which are not independent should not be used.

It should be noted that accuracy, diagnostic sensitivity and diagnostic specificity are strongly dependent on the choice of the target and non-target samples. In addition, accuracy estimation can yield misleading results if the data set is unbalanced between target and non-target samples. For example, if the sample panel contains only a small proportion of target samples, a very high accuracy can be obtained despite a very low diagnostic specificity. If the number of samples vary between laboratories and tests, it is particularly important to include confidence intervals in the analysis. Confidence intervals for accuracy, diagnostic sensitivity and diagnostic specificity can be calculated using the Agresti-Coull method (Agresti & Coull, 1998; Brotaux et al., 2021; Massart et al., 2009a, b Massart et al., 2022).

2. *Intra- and interlaboratory variations*

The organizer can calculate accordance, i.e. the probability of obtaining the same result (positive, negative or inconclusive) from replicate samples analysed in the same laboratory (Langton et al., 2002). For TPS, the accordance calculated for each test will provide information on repeatability. Accordance can be calculated per sample or per laboratory to identify the ones that give discordant results.

The organizer can also calculate concordance, i.e. the probability of obtaining the same result (positive, negative or inconclusive) from replicate samples analysed in different laboratories (Langton et al., 2002). For TPS, the concordance calculated for each test will provide information on reproducibility. Concordance can be calculated per sample to identify the samples leading to high discrepancies between laboratories.

Details on how to calculate accordance and concordance can be found in Langton et al. (2002), Brotaux et al., 2021, Massart et al. (2022).

Confidence intervals for accordance and concordance can be derived from bootstrap standard errors according to Langton et al. (2002).

3. *Analytical sensitivity*

Methods based on calculations of probability of detection (POD) can be used to compare the analytical sensitivity of different tests, in particular for TPS (Wehling et al., 2011). A binomial generalized linear model (bGLM) with a logit link function (log[p/(1-p] where p is the probability of detection) was recommended as POD model in the framework of VALITEST (Brotaux et al., 2021; Massart et al., 2022). This statistical method does not require any assumption on the number of technical and/or biological replicates so that those numbers can vary between samples and/or laboratories. The model proposed here requires a minimum of five dilution points to perform correctly.

The analytical sensitivity of different tests can be compared by looking at the dilutions corresponding to a specific probability of detection level (e.g. 95%, Figure A1). It should be noted that bGLMs assume that the probability of detection is decreasing when the dilution level increases. This hypothesis can be shown to be wrong in some cases, e.g. when all the samples show the same status whatever the dilution level or when the observed detection rate shows contradictory behaviour (e.g. decrease then increase again). When this happens, the model is not suitable.