International Seed Federation
Seed is Life

Technical Guidelines for the Organisation of Comparative Tests for the Validation of Seed Health Tests

DECEMBER 2024, VERSION 1

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 design of comparative tests for the validation of seed health tests.

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

A comparative test (CT) is an evaluation of an assay or method that involves running replicate subsamples in multiple laboratories with pre-determined conditions and within a defined time frame to show an assay or method's reproducibility. Reproducibility is one of the method validation parameters, as described in the International Seed Health Initiative's (ISHIs) Guidelines for the validation of seed health methods. It assesses the degree of similarity in results when the assay or method is performed across laboratories. It is evaluated after the assessment of the validation criteria analytical specificity, analytical sensitivity, selectivity, repeatability, and diagnostic performance (in the case that a reference method is available).

These technical guidelines have been developed to ensure uniformity in the execution of CTs within ISHI and their successful outcome. The principles, factors, and tools that should be considered and deployed when organizing a CT for methods detecting seed-borne pathogens are described in this document.

The different steps involved in the preparation and execution of a CT are summarized in Figure 1 and detailed in the present document.

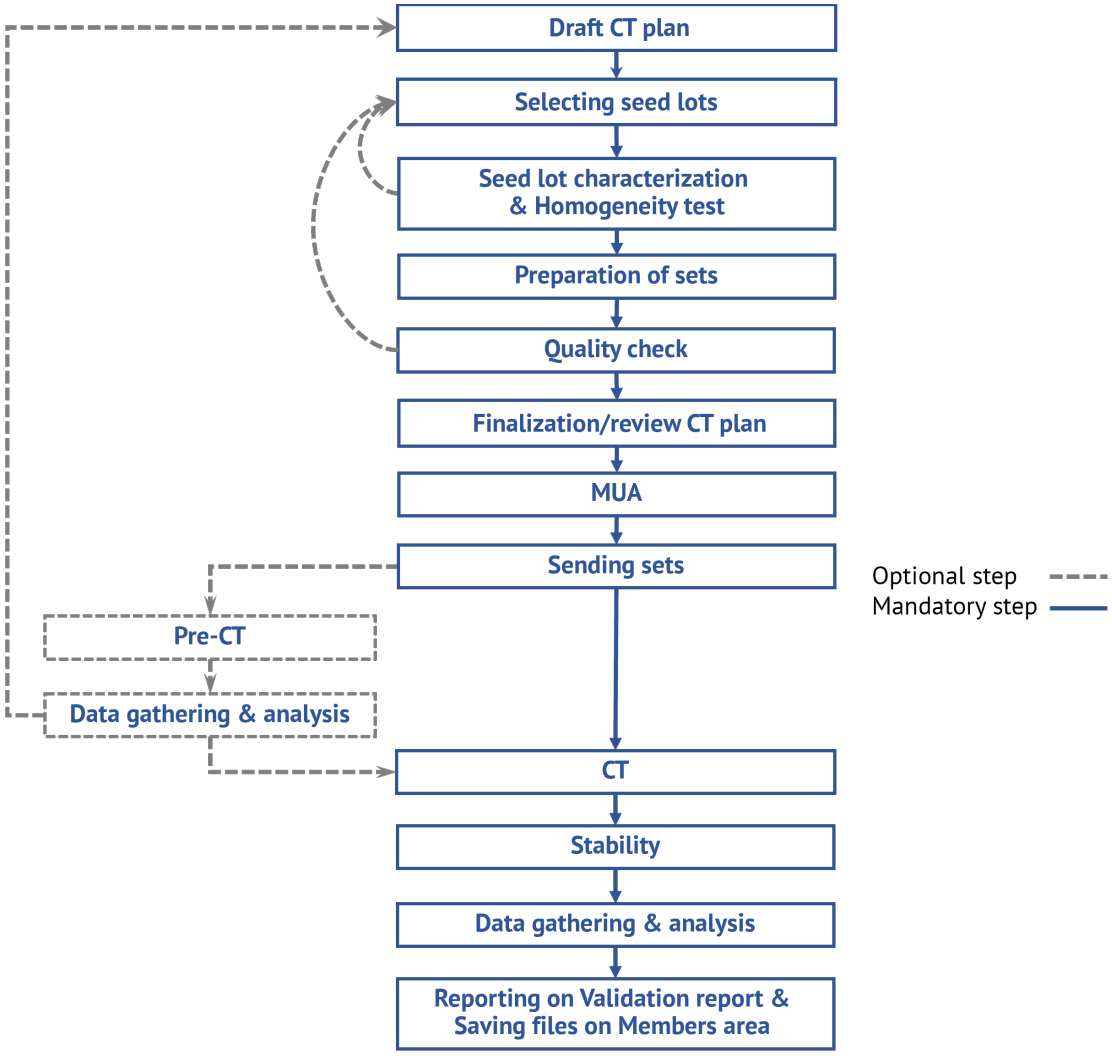

Figure 1. Comparative test process flow.

1 CT PLAN

The first step when preparing for a CT is to draft a CT plan. Essential elements of this CT plan are the organization and scope of the CT, its objective(s), material, and method required, a data submission template and information on how the data will be analysed. When a pre-CT will be performed, the CT plan should also define how the pre-CT is conducted. The CT plan should not describe the subsample details, such as the level of infection of targets or non-targets. The CT plan will be reviewed by the members of the ITG corresponding to the crop species. When the CT plan deviates from the original validation plan or when the minimum requirements from these guidelines cannot be met, then the MVT reviewers should also review the CT plan. The CT plan should be shared and presented to the CT participants before the onset of the (pre-)CT.

1.1. CT participants

It is recommended to include six to eight participants from at least two countries, who should ideally be experienced in performing the method. The minimum number of participating laboratories, for statistical analysis purposes, is three. Over eight participants could be included to increase statistical power or in case that some of the participating laboratories would need to be excluded (e.g., due to problems with the correct execution of the method).

Note: As the number of participants increases, the complexity of the logistics and analysis may also increase.

1.2. Pre-CT

A pre-CT is optional, it is recommended when it is a new assay or method, major evolution of an existing assay or method, or lack of proficient participants, for CT participants to gain proficiency in the method. When performed, its description should be included in the CT plan.

1.3. Protocol

The protocol should be included in the CT plan. It is important to include which deviations are allowed and the critical control points on which deviations are not tolerant because of their significant impact on the outcome of the CT. These deviations should be assessed and decided by the project team before the onset of the CT.

2 SEED LOT SELECTION

Both healthy and infected seed lots should be identified for the CT. Selection of suitable seed lots is crucial to conduct a successful CT. For seed lot selection, the following points should be taken into consideration:

- Select untreated seed lots (unless treated seeds are in the scope of the project). Preferably select healthy negative seed lots produced in a pest-free location.
- Naturally infected seeds are recommended since they provide a realistic representation of the pathogen's stability and seed localization. When no naturally infected seed lots have been identified, seed lots can be artificially produced or spiked with naturally infected seeds/pathogen. In this latter case, the method of production of infected seed should be described/cited in the CT plan.

- Preferably select seed lots that are homogenously infected, with all subsamples detected as infected. If not available, heterogenous seed lots can be used, for which several subsamples are infected (see Section 3.1).
- If selectivity shows negligible or accepted influence of the crop on the results, for methods that are being validated on multiple crops, the CT can be done on a single crop. If, however, during the selectivity evaluation, crops have been shown to influence the test in terms of protocol parameters these variables should be included in the CT.

3 SEED LOT CHARACTERISATION

It is critical to thoroughly characterize the selected seed lots before conducting the CT. Understanding the homogeneity, level of infection, and pathogen viability, when applicable, of the selected lots is needed to determine the number of seed lots and subsamples to be included in the CT to meet the statistical analyses requirements.

Previously characterized seed lots, seed lots used in earlier stage of the method validation or new seed lots may be used for the CT. Previously known test results may assist seed lot selection. However, the infection level needs to be confirmed with the newly developed method/assay. All work on seed lots should be performed on subsamples prepared according to good laboratory practices (e.g., ISTA, 2024). This recommendation is valid for the preparation of subsamples for the seed lot characterisation, homogeneity test and preparation of the subsample sets steps.

3.1. Homogeneity testing

Homogeneity testing is conducted to estimate the distribution of the positive/negative subsamples in each seed lot. It needs to be performed by the CT organizer close to the time of performing the CT and with all the assays included in the validation scope. The minimum number of subsamples to be tested is either 10 subsamples, or three times the minimal recommended sample size of the method. The CT organizer needs to choose the option that tests the higher number of subsamples (e.g., for a method with a sample size of 3,000 seeds and a subsample size of 250 seeds, 36 subsamples per seed lot should be tested for homogeneity testing. In case the subsample size is 1,000 seeds, 10 subsamples per seed lot should be tested).

NOTE: In case of low seed availability or high costs of running the method, less than the minimal recommended subsamples may be tested, with the approval of the ITG and MVT reviewers. Interpretation of homogeneity test results is presented in Table 1.

Table 1. Analysis of homogeneity testing results according to the infection status of the seed lots.

Obtained results	Health status
Negative for all subsamples	Healthy seed lot (homogenous)
Positive for all subsamples	Infected homogenous seed lot
Positive and negative subsamples	Infected heterogenous seed lot

3.2. Level of infection

Data of the homogeneity testing is used to evaluate and confirm the infection level of a seed lot.

Four infection levels are recognized: High/Medium/Low/Healthy (Table 2). The level of infection is based on the quantitative results of the assay and is related to the sensitivity and limit of detection (LOD) of the assay. The level of infection does not imply that a seed lot is homogenously or heterogeneously infected and is determined by evaluating the following:

- Plating assay: number of CFU/mL
- Agar/blotter assay: number or percentage of infected seeds
- PCR assay: Cq-values
- ELISA: OD values
- Bioassay: number of lesions
- Grow-out: number of infected seedlings.

Table 2. Description of the different levels of infection of seed lots.

Infection level	Description
High	All subsamples should be detected positive with relatively high values compared to the LOD.
Medium	All subsamples should be detected positive with intermediate values compared to the LOD. Note that when homogeneous infected seed lots are not available, it is possible to use heterogeneously infected seed lots in which only a part of the subsamples tested are positive with relatively intermediate/high values from the LOD.
Low	All subsamples should be detected positive with values that are just above the LOD. Note that low infection subsamples will only be used at the pre-CT stage, to confirm that the participants are able to perform the method. As such, it is required for the seed lot to be homogenously infected or if not possible, artificially infected to ensure that all subsamples will be detected as infected.
Healthy	All subsamples should be detected negative (below the LOD)

When no or insufficient infected seeds are available, the project team can characterize the seed lot by spiking infected seeds in a negative subsample. Note that artificially infected seed lots are prone to pathogen decay and a decrease in the level of infection may be observed over time. As such, a seed lot may evolve from homogeneously to heterogeneously infected.

4 SUBSAMPLE SET PREPARATION

4.1. Number of sets

The total number of subsample sets required for the pre-CT and the CT is determined by adding up the number of sets needed for the following activities:

- Number of participants (one set per participant),
- Stability test (one set),
- Quality test (one set),
- Back-up sets in case of unforeseen issues such as lost shipment, issues during execution by participants that require repetition (two sets).

Each participant should be coded uniquely to ensure anonymous participation. The set and participant corresponding information is only known by the CT organizer. When discussing results with the project team or with the participants, only codes will be shared.

4.2. Set composition based on health status

Testing seed lots from different known infection levels will help assessing different aspects of the reproducibility. In addition, the set composition will give information in the risk of cross contamination, detection ability, and a potential difference in sensitivity among CT participants. It is recommended to include healthy seeds and at least two infection levels, medium and high (low level should not be used as it is prone to variations caused by participant proficiency and laboratory differences). If not available, it is essential for a CT to contain healthy seeds and the medium infection level.

4.3. Number of subsamples in the sample set

The number of subsamples in the sample set to be included in a CT differs with the use of homogenous and/or heterogenous seed lots.

For the healthy and high infected seed lot, it is recommended to use the same number of subsamples as for the homogenous seed lot with a medium infection level. However, it is essential to include a minimum of three subsamples.

4.3.1. Homogenous seed lots with a medium infection level

The minimum number of subsamples is determined by the number of participants, as described in Table 3. The lower the number of participants, the higher the number of subsamples. When possible, it is strongly recommended to increase the number of participants to reduce the number of subsamples and to ensure a reasonable workload per participant. However, it is recommended to include a higher number of subsamples that is the same as the number of participants minus two. Recommended and minimum numbers of subsamples per participant are presented in Table 3, and were calculated using the Langton-Accordance and Concordance Tool by International Seed Testing Association (ISTA). For example, in case that CT includes eight participants, the number of subsamples corresponding to six participants should be chosen. By doing so, the removal of a participant due to unforeseen circumstances will not jeopardize the statistical power of the CT.

Table 3. The minimum number of subsamples needed for homogeneous seed lots with a medium infection level according to the number of CT participants.

Number of participants	Minimum number of subsamples	Recommended number of subsamples
3	14	14
4	10	14
5	8	14
6	7	10
7	6	8
8	5	7
9	5	6
10	4	5

4.3.2. Heterogenous seed lots

Firstly, estimate the percentage of infection by entering your homogeneity test results in the "Qual Impurity Estimation" sheet from the ISTA Seedscalc8 tool (Figure 2).

- Enter the number of subsamples (i.e., "# of Seed Pools"), number of seeds per subsample (i.e., "# of Seeds per Pool") and the number of positive subsamples (i.e., "# Deviants Pools") in the relevant cells. The desired confidence level recommended is 95%.
- The output "Computed % in sample" is the estimated infection rate expressed in percentage of the seed lot.

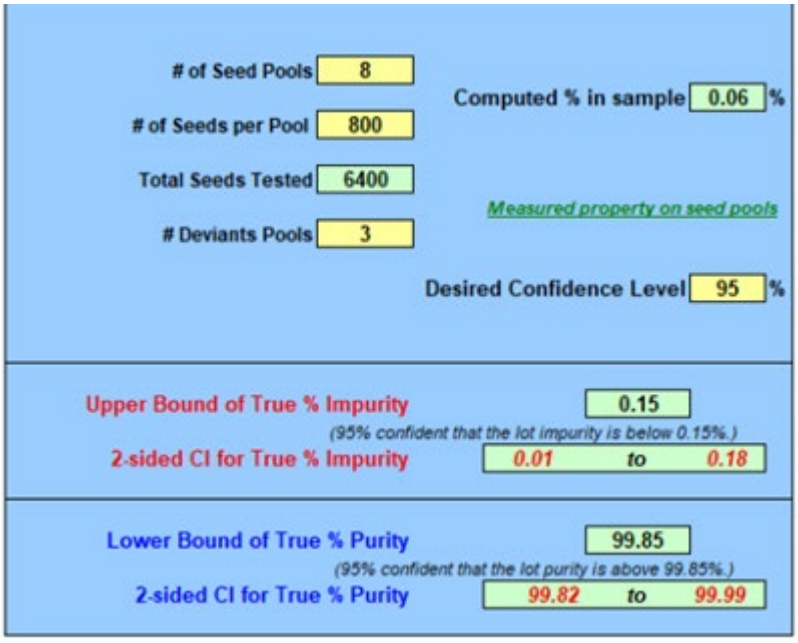

Figure 2. Seedcalc8 "Qual Impurity Estimation" sheet results for the calculation of the percentage of infection.

Secondly, define the number of subsamples to use with the Probability of k positive samples out of n tool by ISTA (Figure 3).

- Enter the number of seeds per subsample (i.e., "# of seeds per sample (m)"), the infection rate, which is the "Computed % in sample", calculated by Seedcalc8 "Qual Impurity Estimation" (see Figure 2) sheet (i.e., "True contamination rate (π)") and the number of subsamples ("Total # of samples (n)") to obtain the "prob (%) of k positive out of n" (Figure 3).
- Adjust (by increasing or decreasing) the number of subsamples (i.e., "Total # of samples (n)") until reaching the lowest number in which the probability of k=2 is below 5% (which represents the likelihood to identify three or more positive subsamples with a confidence rate of minimum 95%) (Figure 3).
- This number is the minimal number of subsamples to be prepared per participant in the CT.

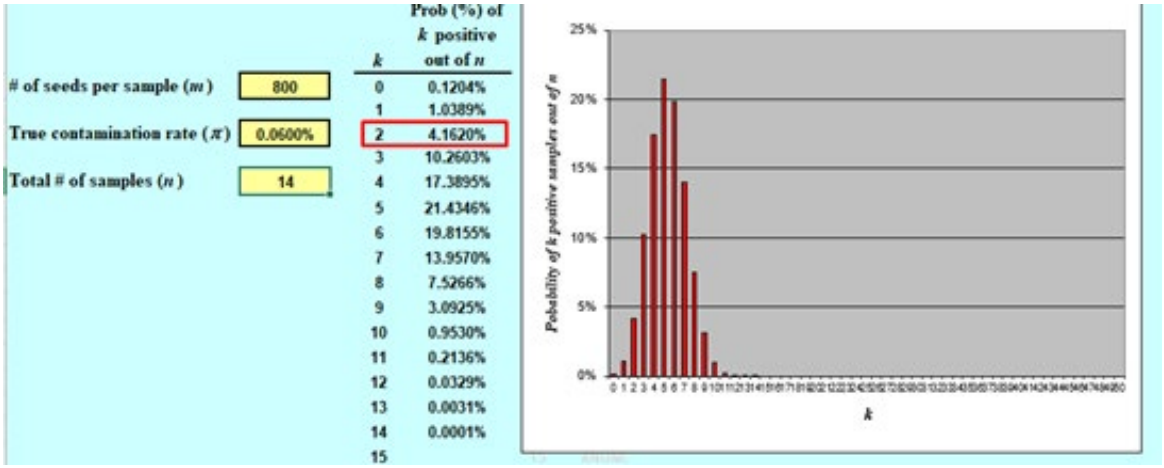

Figure 3. Determination of the minimal number of subsamples from heterogenous seed samples to be included per CT participant per infection level with a minimum confidence rate of 95% using the Probability of k positive samples out of n tool by ISTA.

The figure shows the probability of k positive samples out of n results if 14 subsamples are included per CT participant per infection level. Red square indicates that when testing 14 samples, k=2, as the probability obtained is <5%.

4.4. Preparation of the subsamples - Coding and randomizing

To ensure a blind test and mimic a realistic situation of sample testing in a seed health laboratory, the subsamples for the CT should be randomized. This can be achieved by using randomizing tools such as the Excel functions RAND or RANDBETWEEN. Ensure that the different infection levels are spread evenly in the subsamples and not clustered. In case they are clustered, randomize the subsamples again. The same subsample numbering should be applied to all sample sets.

To minimize variation between sampling sessions, the total number of subsamples defined in the previous section should be sampled and prepared at the same time by the organizer, before performing any tests.

All subsamples for each seed lot should be sampled using good laboratory practices (e.g., ISTA, 2024) to obtain representative samples based on the thousand seed weight (TSW) or by counting and packaged in suitable bags (e.g., plastic bag with internal aluminium coated, paper bags) or packages.

Controls

The controls should be defined in the CT plan including their scoring and expected outcomes. The controls should be characterized (defined quantitative value), homogenous (qualitative positive or negative) and included in the quality and stability tests.

The negative process control (NPC) and positive process control (PPC) should always be provided by the CT organizer.

In molecular methods, it is recommended to also include:

- A positive amplification control (PAC) provided by the CT organizer to give indication of the participated (technical) laboratory performance.
- An in-house PAC and NTC of the CT participant.
- An internal amplification control (IAC) provided by CT organizer. In case specific regulations prohibit this action, then it is the responsibility of the participants to provide one which should be validated in the pre-CT.

4.5. Quality check

This quality check is done by the CT organizer on a full set of subsamples, randomly chosen from the prepared sets. A quality check should be performed after sampling and before shipment to participants using all methods under the validation scope. This check is important to exclude possible errors during the subsample set preparation and packing.

The results of the quality test will be compared to the expected outcomes of the infection levels based on the previously conducted homogeneity testing and subsample determination. In case of contradictory results between a seed lot expected result and the quality check, it will be the decision of the CT organizer to:

- Re-estimate the seed lot homo-/heterogeneity by consulting the MVT and or the statistician, and adjust the number of subsamples accordingly, if possible.
- Cancel the seed lot.
- Use another characterized seed lot.

5 PRE-CT

When a pre-CT is conducted, the results can be used to identify participant performances issues. Based on the pre-CT results a project team can suggest (some) participants to optimize the method or even decide to exclude participants from the CT due to weak performances. In rare cases, a re-evaluation of the method itself may be warranted. Requirements of pre-CT subsamples are listed here:

- The pre-CT set should include all the controls applied in the CT.
- Inclusion of low infection level subsamples (just above LOD) in the pre-CT is important to provide information about laboratory performance.

- In general, three subsamples (healthy/low/high), preferably from homogenous seed lot, are enough to achieve the pre-CT goal.
- The subsamples should not be blind, i.e., information regarding the pre-CT subsamples' level of infection should be provided to the participants of the pre-CT.
- The pre-CT subsamples can be shipped along with the CT subsamples but need to be clearly identified.

Once the pre-CT is concluded, analysed, and discussed, the CT can start.

6 MATERIAL USE AGREEMENT (MUA) AND SHIPMENT

Once the CT plan and the pre-CT and CT subsample sets are ready, the Material Use Agreement (MUA) and the subsample shipment need to be prepared.

- Prepare an ISHI standard MUA for the subsamples to be shipped and have all participants to fill and sign. The organizer should ensure that all MUAs are signed before shipment and shared with the ISHI technical coordinator for documentation purposes.
- Check with all participants the putative import/quarantine permits and/or phytosanitary certificate required to ensure good reception of the subsamples.
- Pack shipping cartons with proper labelling to avoid transportation delays and to minimize the possibility of breakage.
- For international delivery, indicate on the label "Laboratory samples-no commercial value" and other designation required by customs regulations of the country to which the package is being sent, such as the import permit, phytosanitary certificate.
- All participants should confirm by email to the CT organizer the date of receipt of the package and its status upon arrival.

7 STABILITY TEST

Since shipment and storage can influence the subsample conditions, it is important to ensure the infection of the subsamples is stable during the entire length of the CT. The stability test performed by the organizer should start after the last participant has initiated their tests. It includes a subsample set and uses the method as described in the CT plan.

- The subsample set to be used for the stability test is recommended to be shipped to a location and shipped back to the organizer to mimic the effect of shipment on the subsample set. The minimum requirements are to store the subsample set at room temperature until the last participant has received its subsample set.
- The stability test result should be used to analyse the CT results. If results show that the subsample is not stable, it is an indication that unexpected or unpredicted CT results may be due to the subsample and not due to the method, the shipment conditions or CT participant. In this case, it is the responsibility of the CT organizer to either consult the MVT, exclude the subsamples in the statistical analysis or adapt the statistical analysis (see Section 8).

In addition, a data logger may be placed in each shipment box to record the variations in temperatures during shipment.

8 RESULT ANALYSIS

Result analyses is performed per infection level.

8.1. Pre-CT data analysis

Since the goal of the pre-CT is to familiarize the participants with the method. The analysis of the results should be carried out at the qualitative level. However, values of the data gathered (e.g., Cq values, OD values etc.) can still be evaluated. Results of the pre-CT should be recorded in an Excel file and uploaded in the ISHI members area for information purposes only. Pre-CT results should be concluded based on the project lead's discretion.

8.2. Quality and Stability test data analysis

Data gathered in the quality and stability test should be comparatively analysed against the homogeneity test results.

For homogenous seed lots, all infected subsamples in the quality and stability tests should give a positive result, and all healthy subsamples a negative result. For heterogenous seed lots, the number of positive subsamples expected for the infected subsamples in the quality and stability test is calculated with the Probability of k positive samples out of n tool by ISTA. For each test, indicate the number of seeds tested per subsample (i.e., "# of seeds per sample (m)"), the estimated infection rate calculated with the homogeneity data (See Section 4.3.2.) (i.e. "True contamination rate (π)") and the number of subsamples tested in a single subsample set in the (i.e. "Total # of samples (n)"). The k values reaching a probability of k positive out of n above 5% are considered as expected. In the example in Figure 3, three to eight subsamples out of the 14 subsamples tested are expected to be detected as positive.

If the quality and/or stability test results deviate from the expected results, they should be considered for the CT statistical analysis (see Section 8.3.2).

8.3. CT data analysis

In order to understand the origin of the differences between participants, and to identify if these are significant or not, statistical tools need to be applied on the results. Statistical methods and tools will differ between homogeneous and non-homogeneous seed lots. The steps described below are applied on qualitative data only.

8.3.1. Homogeneous seed lots

In case of using homogenous seed lots in the CT, the Langton-accordance and Concordance Tool by ISTA is used to calculate the concordance.

For each infection level, indicate the number of tested subsamples (i.e., total number of subsamples, identical for each participant) and the number of positive subsamples (i.e., participant dependent) in the table. It is expected for the concordance (used for reproducibility) to be ≥90%. Crude odd ratio (COR) should be below <1.3. Examples are presented in Figure 4.

Accordance and concordance worksheet

Clear data | Clear results | Run bootstrap

Confidence limits: 95
Bootstrap samples: 3000
Bootstrap method: representative

Laboratory	Number of samples	positives
1	13	0
2	13	0
3	13	0
4	13	0
5	13	0
6	12	0
7	13	0
8		
9		
10		
11		
12		

	Estimate	Bootstrap s.e.	Bootstrap 95% limits lower	upper
Within lab pairs	534			
Within lab matched pairs	534			
Accordance (propn)	1,000	0,0000	1	1
Accordance (percentage)	100,0%	0,00%	100,00%	100,00%
Total pairs	4005			
Total matched pairs	4005			
Between pairings	3471			
same between	3471			
Concordance (propn)	1,000	0,0000	1	1
Concordance (percentage)	100,0%	0,00%	100,00%	100,00%
COR	1,00		1	1

Above results are based on 3000 bootstrap samples of 7 labs using representative method

A

Accordance and concordance worksheet

Clear data | Clear results | Run bootstrap

Confidence limits: 95
Bootstrap samples: 3000
Bootstrap method: representative

Laboratory	Number of samples	positives
1	13	12
2	13	13
3	13	13
4	13	13
5	13	13
6	12	11
7	13	13
8		
9		
10		
11		
12		

	Estimate	Bootstrap s.e.	Bootstrap 95% limits lower	upper
Within lab pairs	534			
Within lab matched pairs	511			
Accordance (propn)	0,957	0,0262	0,9116466	1
Accordance (percentage)	95,7%	2,62%	91,16%	100,00%
Total pairs	4005			
Total matched pairs	3829			
Between pairings	3471			
same between	3318			
Concordance (propn)	0,956	0,0255	0,9111933	1
Concordance (percentage)	95,6%	2,55%	91,12%	100,00%
COR	1,02		0,961202	1,092999

Above results are based on 3000 bootstrap samples of 7 labs using representative method

B

Figure 4. Determination of concordance using data of A. healthy subsamples and B. infected subsamples.

Concordance for healthy subsamples is calculated to be A. 100% for healthy subsamples and B. 95.6% for infected subsamples, as presented in the red boxes.

8.3.2. Heterogenous seed lots

In case of using heterogenous seed lots in the CT, the estimated percentage of infection calculated using the homogeneity test results with the "Qual Impurity Estimation" sheet from the Seedscalc8 tool by ISTA (see Section 4.3.2, Figure 2) is used to calculate the expected number of positive subsamples using the Probability of k positive samples out of n tool (see Section 8.2 for calculation). However, when the quality and/or stability test results deviate from the expected results (see Section 8.2), the estimated percentage of infection should be calculated using the quality and stability test data, instead of the homogeneity test data (See Section 4.3.2 for calculation).

8.3.3. Excluding laboratories/results

The project lead and team, based on a set of well-defined criteria, may exclude certain data from the CT analysis. Reasons for data exclusion include:

- Controls not aligned with expectations,
- Cross-contamination,
- Known technical issues.

Note that the data from all CT participants should be included in the report and a clear explanation should be provided to explain data exclusion from the analysis.

REFERENCES

International Seed Testing Association. Chapter 2: Sampling (2024). Available: https://www.seedtest.org/en/publications/international-rules-seed-testing.html. Last accessed: 19 November 2024.

International Seed Testing Association. Probability of k positive samples out of n. Tool available: https://www.seedtest.org/en/services-header/tools/seed-health-committee.html. Last accessed: 27 November 2023.

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: 175-171. Available: https://www.sciencedirect.com/science/article/abs/pii/S0168160502001071.

Remund K., Simpson R., Laffont J.-L., Wright D., and Gregoire S. Seedcalc 8. Tool available: https://www.seedtest.org/en/services-header/tools/statistics-committee.html. Last accessed: 27 November 2023.

Seed is Life