



## Organizing and analyzing results of the Seed Health Proficiency Tests

Any hardcopy of this document is not controlled copy

CREATED	REVIEWED	APPROVED
07.02.2019	10.08.2021	10.09.2021
VALERIE GRIMAULT, JEAN-LOUIS LAF-FONT, MARK BUIMER AND ISABELLE SERANDAT (SHC)	VALERIE GRIMAULT SEED HEALTH COMMITTEE CHAIR	VALERIE GRIMAULT SEED HEALTH COMMITTEE CHAIR

## SCOPE

The ISTA Seed Health Committee organizes proficiency tests (PT) to monitor the ability of accredited laboratories to detect seed borne pathogens using the Annexe to Chapter 7 Seed Health Testing methods. Several different methods have been used for analysing results and rate laboratories. Different types of results are expected based on the method (individual seeds or pooled seeds) used to detect the pathogen.

- For detection of fungi, in most cases seeds are tested individually on media or blotters and results are generally expressed as the number of infected seeds per the number of seeds tested or as a percentage infection. These are quantitative results.
- For detection of bacteria or viruses, seeds are tested using pools containing X number of seeds per pool (subsample). Results are expressed as the number of infected subsamples (or positive subsamples) detected within a total number of subsamples tested. These are qualitative results.

It has been difficult for organizers to decide which statistical methods to use for these two scenarios. Therefore, the ISTA SHC decided to work with the ISTA Statistics Committee to define some guidelines to help PT organizers to set up and analyse their tests.

The following text establishes the second version of guidelines for statistical analysis for PTs.

## RESPONSIBILITY

The Seed Health Committee is responsible for organising Proficiency Tests and inter-laboratory tests for validation of methods and the evaluation of their results.

## ABBREVIATIONS

BMP: Below Minimum Performance

PT: Proficiency Test

StatCommittee: Statistic Committee

SHC: Seed Health Committee

## PROCESS DESCRIPTION

PT rounds are prepared by PT leaders. The PT leaders:

- are defined in the 3-year PT Programme Plan
- are members of the PTC
- belong or have access to a laboratory for technical preparation; this laboratory must be an ISTA accredited laboratory for the PT species, for sampling in the laboratory and for the major expected tests
- are approved by the PTC Chair in consultation with the Accreditation and Technical Department; and apply the current version of PT-P-02-Proficiency Test Sample Preparation (confidential document).

## SCHEDULE

Regular PT rounds are performed three times per year. Each round is made up of three samples for each test (or group of tests) and for each species. The samples undergo purity analysis, other seed determination, germination testing, moisture content determination, viability testing, vigour testing and/or thousand-seed weight testing, as applicable. PTs for tests other than those mentioned may be organised by the TCOM, but are, for the time being, not yet part of the rating.

## PROGRAMME PLAN

### Seed lots:

Seed lots used for PT can be either naturally infected or artificially contaminated.

Artificial contamination can be achieved:

- For viruses for ELISA testing: by mixing healthy seeds with ground infected dried tissues or with infected seeds.
- For bacteria by using a vacuum to infiltrate a bacterial suspension into individual seeds (with or without damaging the teguments). These seeds should be added to healthy seeds with a normal saprophytic background to mimic natural seed lots.
- For fungi, by immersing seeds in a spore suspension or plating seeds on a fungus colony under restriction of water (use of Mannitol to decrease water availability). These seeds should be added to healthy seeds with a normal saprophytic background to mimic natural seed lots.

Stability of the pathogen is often lower in case of artificially contaminated seeds and should be checked carefully in those applications where viability and infectivity are an issue.

A pre-test will be necessary to determine the level of contamination, i.e., the expected results of subsamples.

**Samples/subsamples:** subsamples will be prepared by the organizers and sent to each participant to avoid variation in subsampling in laboratories and to be certain of what is sent. **No subsampling should be done in the participating laboratories.**

### Levels of infection

- Levels of infection are determined based on pre-tests. To have a good estimation, it is advised to test a sufficient number of samples during the pre-test.
- Healthy, no presence of pathogen (with saprophytes if possible): all subsamples should be negative. It should have a saprophyte load equivalent to routinely tested commercial seed lots.
- Low and/or medium infection level (combination of low and/or medium).
  - For quantitative results, the organizer needs to be rather certain, based on pre-tests, that all low-level subsamples are contaminated (at least one infected seed in subsample analysed) and that contamination is over the detection limit of the method. For example, a minimum of 5% contamination for fungi is required to be sure that all subsamples distributed will most likely contain at least one contaminated seed.
  - For qualitative results, the organizer will determine the % of contamination of the seed lot based on Seedcalc program provided in the STATCOM webpage (tools), and then determine the probability to find a number of contaminated samples over the number tested with the “probability tool for qualitative results” provided on the SHC webpage (tools). For example, in figure 1, a 0.07% infection (computed % in sample) has been determined because on 8 subsamples of 1000 seeds tested, 4 were positive. In this case, (Fig 2) for a 0.07% of infection with 1000 seeds tested, if 3 subsamples are tested in the PT, at a probability >5%, it is possible to detect from 0 to 3 positive subsamples, which is not very informative for a future PT.

Figure 1: Calculation of % infection using seedcalc

**Impurity Estimation & Confidence Intervals (Assay measures impurity characteristic)**

(Number of seed sampled should not exceed 10% of total number in population)

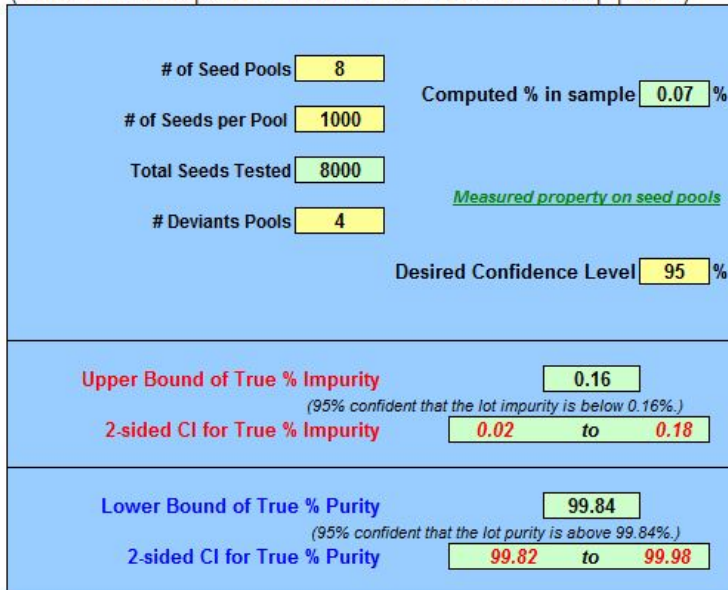
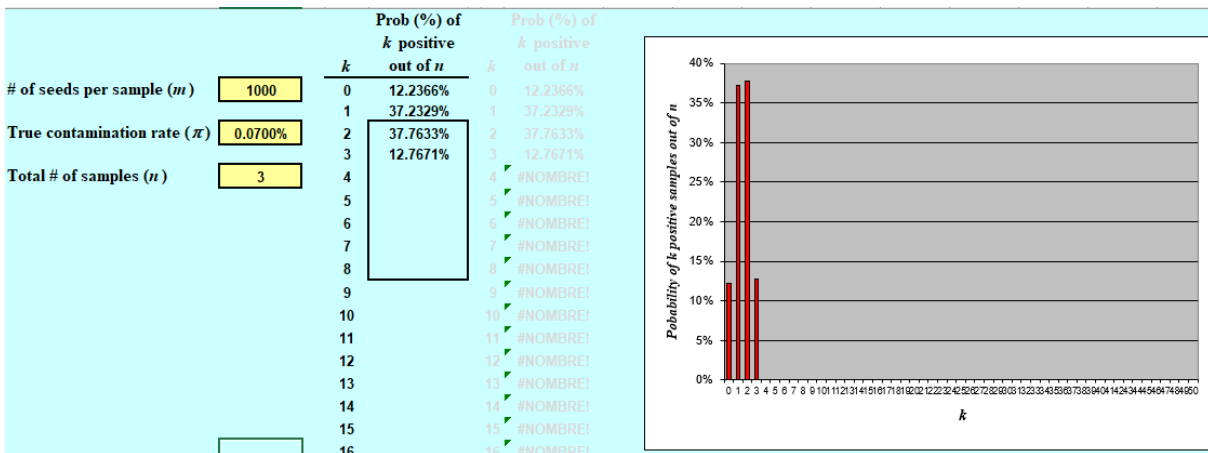


Figure 2: Probability tool used to determine expected number of positive subsamples on total subsamples tested based on %infection, at 5% confidence



- High infection level, all subsamples should be positive

**Number of subsamples per infection level:**

- For plating or blotter tests (fungi), use the same number of subsamples per infection level (i.e., 3 for healthy, 3 for low, 3 for high), and if possible, an extra subsample of one of the infection level to avoid participants to know what result to expect.
- For testing of seed pools (bacteria and viruses), subsample numbers depend on infection/contamination level with more focus on low/medium levels i.e., more subsamples need to be tested at lower levels where there is more chance to have a false negative result. The numbers of pools (subsamples) to test for a low/medium level will depend on the % infection/contamination of the sample and the number of seeds per pool. The number of subsamples will need to allow the organizing lab to rate the participating labs. In the previous example testing 3 subsamples would not allow rating of the labs as at this % of infected seeds, there is a probability to detect between 0 and 3 subsamples when 3 subsamples are tested. So, every lab would obtain the expected result whatever its result. The Figures 3 and 4 below show that 8 subsamples will be

necessary to rate the laboratories as for an A rating between 2 and 6 positive subsamples expected at a 5% probability and it is possible to assign a B rating (see bullet point “rating of lab” for more explanation).

Fig 3: Example of probability tool results used with 6 subsamples at a 0.07% infection

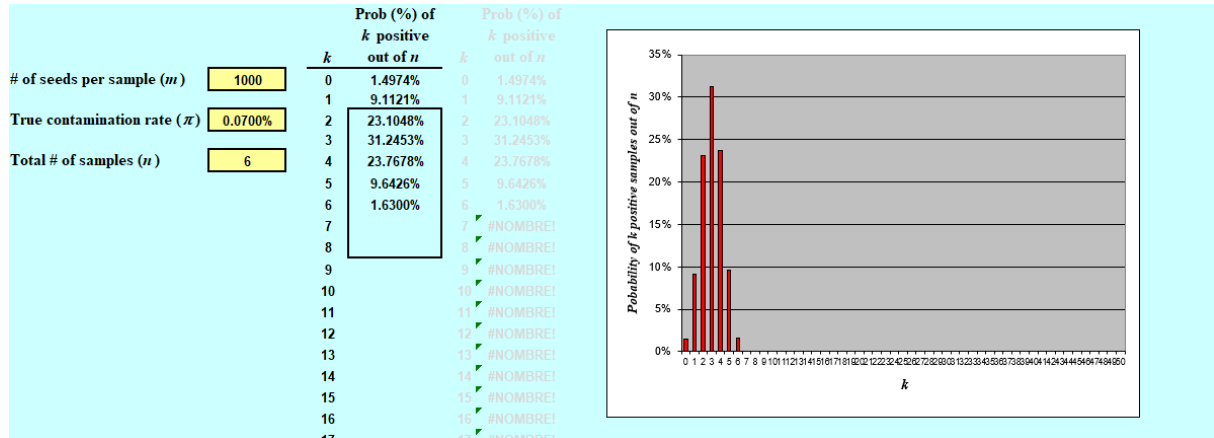
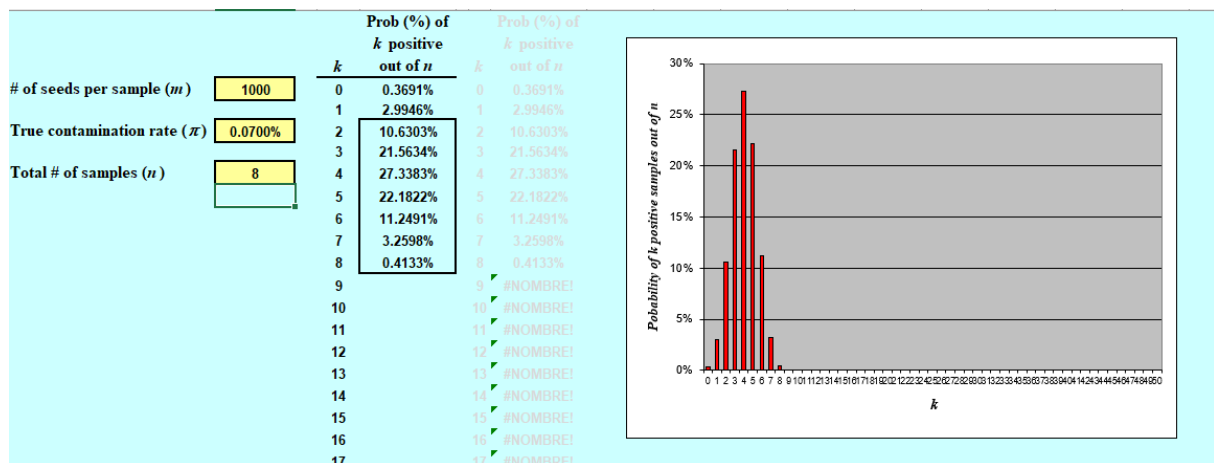


Fig 4 Example of probability tool results used with 8 subsamples at a 0.07% infection



- In all cases, a minimum of 3 subsample per level will be needed
- 1 result will be reported per subsample

**Size of subsample:**

- The organizer should use the subsample size indicated in the method.

**Homogeneity test of subsamples:**

- Done to ensure a homogeneous distribution of pathogen in samples and enables a PT organizer to know if sample is acceptable, i.e., all subsamples distributed to participants will give the expected result.
- Done for each pathogen if multiple pathogens are included in the same PT.
- Completed using the method to be used in the PT
- Done by the organizing lab on a minimum of 8 subsamples (for each level and each pathogen), preferably 10 subsamples,
- Done just before shipping, when subsampling is completed, and subsamples are packed
- For methods giving a quantitative result (fungi, grow out tests), the excel tool to calculate homogeneity of samples is used to check homogeneity. It can be found on SHC web page (toolbox). An example is given below. The result is given in red: “OK” for homogeneous sample, “Out” for non-homogeneous sample.

## CALCULATION OF CONFORMITY FOR HOMOGENEITY TEST

Homogeneity test - seed health										
Sample size	400									
Subsample	1	2	3	4	5	6	7	8	9	10
nb infected seeds	2	5	1	3	6	5	1	8	3	4
percentage of non infected seed	99.5	98.75	99.75	99.25	98.5	98.75	99.75	98	99.25	99
Sample Mean	99.05									
Tolerance	1.41									
H value	0.35									
Homogeneity check	OK									

Taille d'échantillonnage										
Taille d'échantillonnage	400									
Echantillons	1	2	3	4	5	6	7	8	9	10
nb. sem. infectées	51	52	41	53	36	42	46	29	40	45
pourcentage de semences saine	87,25	87	89,75	86,75	91	89,5	88,5	92,75	90	88,75
Moyenne	89,13									
Tolérance	1,41									
Valeur de H	9,28									
Réponse d'homogénéité	Out									

- As this tool is currently difficult to use in cases of some fungi or with different computers, an alternative is to use Hampels method to detect outliers: if no outliers are detected, the sample is considered as homogeneous. An example is given below.

MS Excel Hampels Outlier Test Example					
Lab	Lab Values (Xi)	Xi - M	Status		
1	12,75	1,875	OK	Median (M):	10,875
2	13,00	2,125	OK	MAD:	1,375
3	10,25	0,625	OK	5.2 X MAD	7,150
4	13,25	2,375	OK		
5	9,00	1,875	OK		
6	10,50	0,375	OK		
7	11,50	0,625	OK		
8	7,25	3,625	OK		
9	10,00	0,875	OK		
10	11,25	0,375	OK		

- For methods giving a qualitative result,
  - Homogenous distribution means that all subsamples tested in homogeneity test are positive for infected samples and negative for healthy samples
  - In some cases, for naturally infected seeds (generally low/medium levels), the distribution of the pathogen will not be homogeneous i.e.: all subsamples tested will not have the expected positive result. In this case, the sample is not homogeneous, but homogeneity test can be used to determine the level of contamination (in % using seed calc) and then an expected number of positive samples out of the total sent to participants (see fig 2-4).
- Homogeneity test will be used to validate the expected results of subsamples
- Samples are valid for further analysis if homogeneous.
- In case of lack of homogeneity, it will be decision of the organizer to:
  - Use another seed lot or
  - Cancel this level of infection or

- Not use the subsamples corresponding to this level of infection in the statistical analysis / not rank the laboratories on these subsamples or
- Use the subsamples corresponding to this level of infection and adapt the statistical analysis (for example for qualitative results for low/medium lots, see fig 2-4)

#### **Stability test of subsamples:**

- Performed to be sure that the infection of the subsample does not change during the course of the PT. It is particularly important for artificially infested/contaminated seeds where the level of infection can decrease in time.
- Done with the method used in the PT
- Done on a minimum of 3 subsamples (per level per pathogen), by the organizing lab
- Done just after the deadline for participants to complete the test.
- It must give the expected result (healthy for healthy subsamples, infected for infected seed samples).
- The same statistical tools as homogeneity test can be used to verify stability of samples
- If results show that the sample is not stable: this is an indication that unexpected or unpredicted results of PT may be due to problems with the sample and not to errors or mistakes by the lab. In this case, it is the decision of the organizer to:
  - Not use the subsamples in the statistical analysis / not rank the laboratories
  - Use the subsamples and adapt the statistical analysis

#### **Number of participants:**

- For quantitative results, when comparing results to a mean, a minimum of 8 participants are required. If less than 8 participants or a high variability of results of participants, the results are compared to the expected value.
- For qualitative results, when each participant is compared to the expected value and because the probability tool is used to analyse results, it is not necessary to have a specified number of participants.
- The number of samples to be prepared will depend on the number of participants. Samples for homogeneity test, for stability test and five extra sample sets as replacements need to be added to the total number.

#### **Recording of results:**

- Templates for notation sheets are provided, in order to record all necessary information about lab, samples tested, raw results and final result per sample of the lab. Templates will need to be adapted depending on method used.

#### **Statistical analysis:**

Statistical analysis for **quantitative methods** (fungi: % contamination, number of contaminated seeds):

- The analysis is done using Z scores (see rating of laboratories)

Results will be analysed with the tool developed by STATCOM on results without outliers (unexpected results, results too far from expected ones) determined by Hampel's rule for detecting outliers.

In cases of low number of participants or high variability of results, results can be analysed with a box plot and compared to expected values determined with homogeneity and stability tests.

Statistical analysis for **qualitative methods** (bacteria, viruses, some fungi's results):

- For qualitative results (subsample identified as positive or negative for viruses or bacteria), the counting of colonies or number of lesions or the optical density will be indicative (for a level of contamination for example). It is not expected to do statistical analysis on this aspect. Only the number of positive and negative subsamples will be used for statistical analysis.

- For homogeneous subsamples (generally healthy and high levels). The norm NF EN ISO 16140 will be used to determine the performance criteria of the method through sensitivity, specificity, and accuracy. The analysis consists of a comparison between the expected result (known samples, validated by homogeneity test to be determined by organizer) and the obtained result in participating laboratories. This comparison records positive and negative agreement or positive and negative deviation (cf. table 1).

Table 1:

	expected result +	expected result -
Obtained result +	positive agreement +/+ (PA)	positive deviation -/+ (PD)
Obtained result -	negative deviation +/- (ND)	negative agreement -/- (NA)

These definitions allow the calculation of sensitivity, specificity, accuracy and reproducibility according to the following mathematical formulas:

$$\text{Sensitivity} = \frac{\Sigma PA}{(\Sigma PA + \Sigma ND)} \times 100$$

$$\text{Specificity} = \frac{\Sigma NA}{(\Sigma NA + \Sigma PD)} \times 100$$

$$\text{Accuracy} = \frac{(\Sigma NA + \Sigma PA)}{(\Sigma PA + \Sigma NA + \Sigma PD + \Sigma ND)} \times 100$$

PA = positive agreement

ND = negative deviation

NA = negative agreement

PD = positive deviation

A sensitivity of 100% shows that the method / laboratory always detects the target pathogen: there were no false negatives.

A specificity of 100% shows that the method / laboratory does not give a positive result for a sample that does not contain the target pathogen: there were no false positives.

An accuracy of 100% shows that the target pathogen is always detected; there were no false negatives or false positives.

- For non-homogeneous subsamples, generally low/medium levels. Analysis cannot be performed according to the norm NF EN ISO 16140. In this case, the analysis will be performed based on the % of infection of the samples:
  - Non-homogeneous seed samples are identified during pre-tests and homogeneity test (see page 6).
  - Homogeneity test, or stability test if the result is different from homogeneity test, will be used to give precision on determination of the expected value i.e., the % of infection of the sample. This % will be determined with seed calc. Statistical analysis will be done on a probability to determine the number of positives samples / total analysed with the "probability tool for qualitative results" provided on the SHC webpage (tools).
    - On the previous example with a contamination level of 0.07% and 8 subsamples of 1000 seeds tested, we expect participants to detect from 2 to 6 subsamples with a >5% probability.



## Rating of laboratories:

For SH testing, rating of laboratories is now in the evaluation phase, it has no consequences on the accreditation of the laboratories yet. It is based on an A, B, C and BMP rating like basic testing PTs (see: [PT-P-01 ISTA Standard Proficiency Test](#), for more details on basic tests rating).

Only the end result is used for rating.

- **Quantitative results:**

Use median values of all labs without outliers. Outliers, per level and per pathogen, are identified using Hampel's rule for detecting outliers.

Results of labs will be analyzed using Z scores and automatically calculated by the excel sheet available on ISTA website, SHC page, tools. The paragraphs below explain how the calculations are done and can be exceptionally changed for healthy sample.

### Healthy:

**A:** limit is 0 contaminated seeds. In exceptional cases, the prescreening can show a negative result while the homogeneity test can show 1 positive seed in one subsample among the 10 or more tested. In this case, if possible, the lot is cancelled from the PT. But if no other negative seed lots are available, it can be exceptionally allowed to use this lot for the PT. In this case 1 positive seed in one subsample from the healthy sample will be allowed for an A rating. The % infection corresponding to 1 seed in one subsample will be calculated to fill in the spreadsheet and the corresponding z score will be chosen as the limit.

**B:** accept 1 false positive on total tested and enter the z-score value of 1 false positive as the limit. In the exceptional cases described for A, 1 positive seed in the healthy sample more than for A rating will be allowed for a B rating and the corresponding z score will be chosen as limit.

**C:** accept 2 false positive on total tested and enter the z-score value of 2 false positive as the limit. In the exceptional cases described for A, 1 positive seed in the healthy sample more than for B rating will be allowed for a C rating and the corresponding z-score will be chosen as limit.

### Contaminated samples:

Limits of acceptable Z scores for A, B, C and BMP will be dependent on limits given by quantile of a normal distribution:

- Limit for A: 0.67 which corresponds to 0.75 quantile of a normal distribution
- Limit for B: mean between A and C
- Limit for C: 2.33 which corresponds to 0.99 quantile of a normal distribution

It will give the values already filled on the sheet: (in absolute values of Z scores)

- A: < 0.67
- B: >0.67 and <1.5
- C: > 1.5 and <2.33
- BMP: > 2.33

## Rating for quantitative SH PTs

Change any value in a yellow cell					
	Max abs(z-score) for the Healthy lot		Max abs(z-score) for the Medium lot		Max abs(z-score) for the High lot
Minimum requirements for A rating :	0.08	and	0.67	and	0.67
Minimum requirements for B rating :	0.22	and	1.5	and	1.5
Minimum requirements for C rating :	0.33	and	2.33	and	2.33

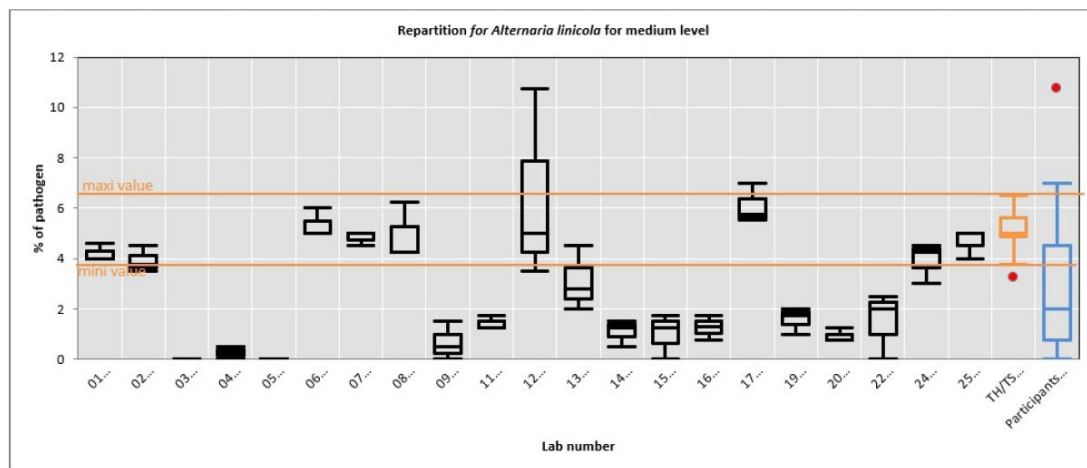
  

Lab	abs(z-score) for the Healthy lot	abs(z-score) for the Medium lot	abs(z-score) for the High lot	Rating
1	0.00	0.09	1.16	B
3	0.33	1.79	1.78	C
4	16.67	9.50	11.60	BMP
5	0.00	0.64	0.68	B
6	1.59	0.57	0.88	BMP
7	0.00	0.07	0.64	A
8	0.08	1.33	1.34	B
9	0.00	0.33	0.75	B
10	0.00	0.70	0.38	B
11	0.08	0.22	0.24	A
12	0.08	0.38	0.37	A

In this example, 1 contaminated seed was detected in the “healthy” seed lot, so the limits for A, B and C have been adapted to Z scores corresponding to 1, 2 and 4 contaminated seeds respectively (yellow cells).

In cases of low number of participants or high variability of results, results can be analysed with a box plot and compared to expected values determined with homogeneity and stability tests. In Fig 5, the max and min values have been determined based on Homogeneity (TH) and stability (TS) tests' results. The same rules for healthy sample apply. For the low medium A, B, C and BMP are determined visually by the organizer based on min/max values.

Fig 5: Analysis of results using box plot



- **Qualitative results:**

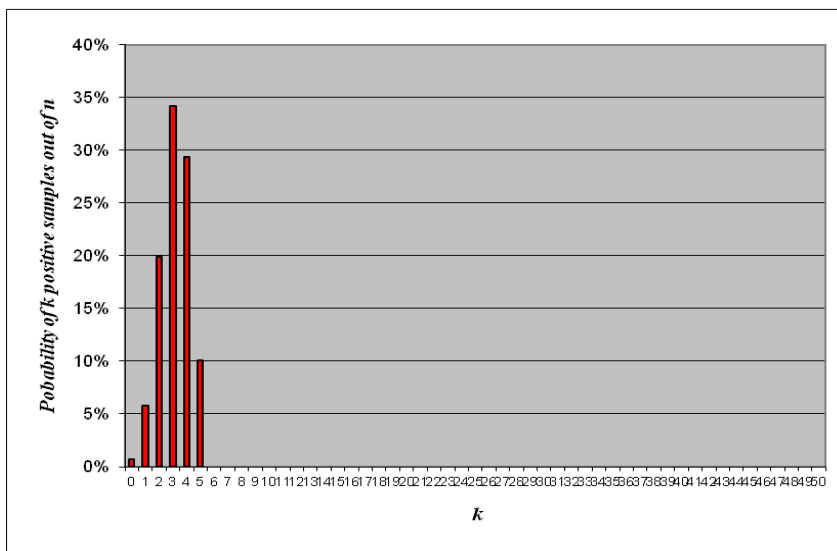
- **Healthy lot:** if homogeneity test shows 0% contamination, we expect 0 contaminated samples, so no flexibility for A B C. Then, if 1 or more contaminated samples are detected in a healthy seed lot the laboratory will be rated BMP.
  - In exceptional cases, the prescreening can show a negative result while the homogeneity test can show 1 positive subsample among the 10 or more tested. In this case, if possible, the lot is cancelled from the PT. But if no other negative seed lots are available, it can be exceptionally allowed to use this lot for the PT. In this case 1 positive subsample from the healthy sample will be allowed for an A rating. If more than 1 subsample is detected, the rating will then be BMP.

- **High level:** if homogeneity test shows all subsamples contaminated, we expect all subsamples to be contaminated, so there is no flexibility for A B C. If less than all subsamples are detected by the laboratory, it is rated as BMP.
- **Medium level:**
  - if homogeneity test shows all subsamples contaminated, we expect all subsamples to be contaminated so there is like for high level no flexibility for A B C. If less than all subsamples are detected by the laboratory, it is rated as BMP. Depending on results of the labs, because it is a medium level, some flexibility can be given for 1 negative subsample, in this case it is rated as B. More than 1 will be rated as BMP.
  - For non-homogeneous samples: calculate the % of contamination based on Seedcalc. The expected result is given by the probability to find a number of contaminated samples among the number tested with the “probability tool for qualitative results” provided on the SHC webpage (tools). The rating will be defined as follows:
    - $C > 1\%$   $B > 2.5\%$   $A > 5\%$  ;
    - C rating is not always possible depending on the number of subsamples. So, a minimum of subsamples is necessary to be able to rank the labs. Depending on the level of contamination, this C rating can be included or excluded from the rating (this occurs frequently for GMO testing, for example)
    - BMP is  $< 1\%$

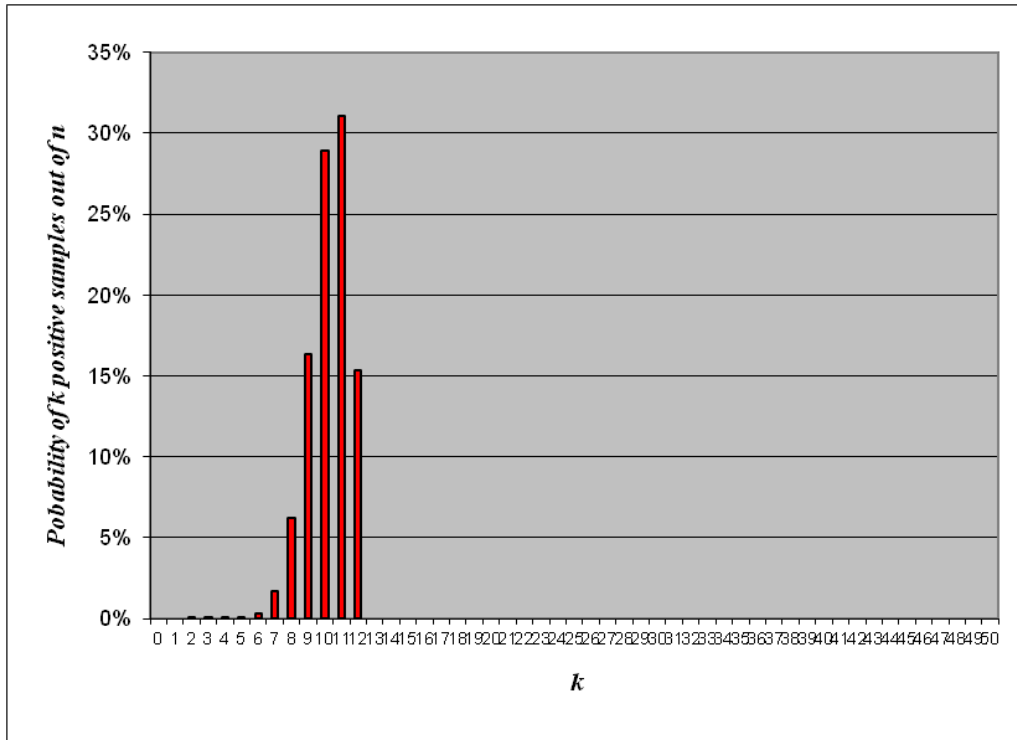
This process should be done while testing the homogeneity to know whether or not a sample is usable for rating the labs.

The A, B, C, BMP ratings are determined for the different infection levels; the worst rating is applied as the overall rating for the participating laboratory.

**Example 1: Xhc dilution plating PT 2003:** The lot chosen had a 0.02% infection. 5 subsamples of 10000 seeds were tested. In this case the lot chosen was not suitable for rating the labs, because there would be only an A and C.

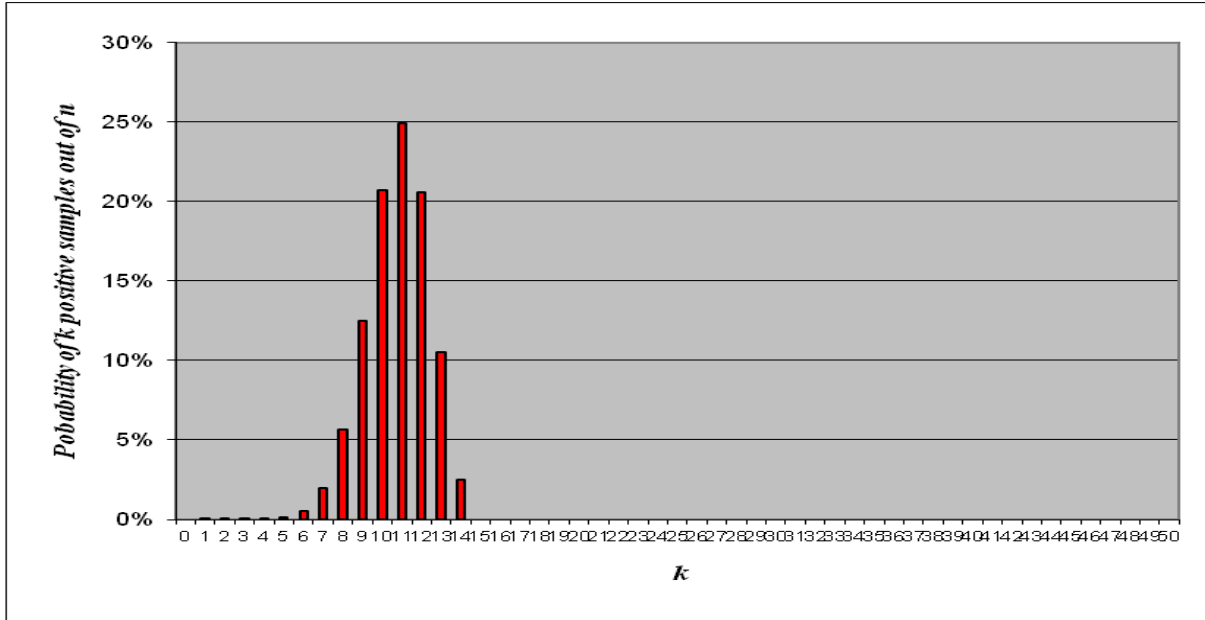


**Example 2: PepMV ELISA PT 2003:** The lot chosen had a 0.77% infection. 12 subsamples of 250 seeds were tested. In this case, there was more than 1% chance to find 7 contaminated samples, so we changed the requirement for B level at the probability for 7 contaminated samples which corresponds to 1.7%. No C rating is possible. See the theoretical example below to illustrate different ratings.



Minimum requirements for A rating :		Healthy lot	and	Min # of pos reps:	High level lot	and	Min prob for observing k pos:	Medium level lot
Max # of pos reps:		0			12			5.0%
Minimum requirements for B rating :		Healthy lot	and	Min # of pos reps:	High level lot	and	Min prob for observing k pos:	Medium level lot
Max # of pos reps:		0			12			1.7%
Minimum requirements for C rating :		Healthy lot	and	Min # of pos reps:	High level lot	and	Min prob for observing k pos:	Medium level lot
Max # of pos reps:		0			12			1.0%
		Healthy lot			High level lot			Medium level lot
							# of seeds/rep:	250
							True cont. rate:	0.7700%
							# of reps tested:	12
Rating	Lab	# of pos reps		# of pos reps		# of pos reps (k)	Prob for observing k pos	
BMP	1	0		11		12	15.31%	
BMP	2	4		12		12	15.31%	
BMP	3	1		12		12	15.31%	
A	4	0		12		12	15.31%	
BMP	5	2		12		5	0.05%	
BMP	6	0		12		6	0.33%	
B	7	0		12		7	1.69%	
A	8	0		12		8	6.23%	

**Example 3: Peronospora grow out PT 2006:** The lot chosen had a 1.45% infection. 14 subsamples were tested. In this case, there was more than a 1% chance to find 7 contaminated samples, so we changed the requirement for B level **at the probability** for 7 contaminated samples which corresponds to 2% and we put the requirement for C level at the probability for 6 contaminated samples which corresponds to 0.51% so changed to 1%. See the theoretical example below to illustrate different ratings.



**Rating for qualitative SH PTs**

Change any value in a yellow cell

Minimum requirements for A rating :		Healthy lot	High level lot	Medium level lot	
Max # of pos reps:	0	and	Min # of pos reps: 12	and	Min prob for observing k pos: 5.0%
Minimum requirements for B rating :		Healthy lot	High level lot	Medium level lot	
Max # of pos reps:	0	and	Min # of pos reps: 12	and	Min prob for observing k pos: 2.0%
Minimum requirements for C rating :		Healthy lot	High level lot	Medium level lot	
Max # of pos reps:	0	and	Min # of pos reps: 12	and	Min prob for observing k pos: 1.0%
	Healthy lot		High level lot	Medium level lot	
				# of seeds/rep: 100	
				True cont. rate: 1.4500%	
				# of reps tested: 14	
<b>Rating</b>	<b>Lab</b>	<b># of pos reps</b>	<b># of pos reps</b>	<b># of pos reps (k)</b>	<b>Prob for observing k pos</b>
BMP	1	0	1	13	10.49%
A	2	0	12	9	12.52%
BMP	3	1	12	12	20.61%
B	4	0	12	7	1.96%
BMP	5	0	0	6	0.52%

**Report:**

- Templates for results sheets generated for each participant are provided in order to record all necessary information: results compared to expected results, rating proposal.
- Report including description of the test and analysis of results available on request

## DISTRIBUTION LIST

Seed Health Committee

Statistical Committee

Accreditation and Technical Department

ISTA Website

## REVISION HISTORY

Version #	Changes
1.0	This document is partly replacing the document TCOM-TG-02-Organizing and Analysing PTs and CTs, but only for PTs. More precisions given for PT.
1.1	The document code has been changed from TCOM-TG-06 to PT-P-03
1.2	3 subsamples for stability tests has been changed to "a minimum of 3 subsamples"
2.0	Number of participants – sentence was added for clarification.
3.0	Layout changed Adjusting the link to the ISTA website Distribution list amended