

ISTA GMO Proficiency Tests: Rating System for Quantitative Results

By **Jean-Louis Laffont**, ISTA Statistics Committee Member (picture)
Pioneer Génétique, Chemin de l'Enseigne, 31840 Aussonne, France, jean-louis.laffont@pioneer.com

Sylvain Grégoire, ISTA Statistics Committee Chair
GEVES, La Minière, 78285 Guyancourt Cedex, France, sylvain.gregoire@geves.fr

Kirk Remund, ISTA Statistics Committee Vice-Chair
Life Sciences Company, 800 North Lindbergh Blvd, St. Louis Q2A, Missouri, USA, kirk.m.remund@monsanto.com



Introduction

ISTA performs multi-laboratory proficiency tests (PT) with the goal to detect seeds with specified trait(s). In the PTs performed to the present date, laboratories were asked to provide the following information on the samples they received:

- presence or absence of GM seeds (qualitative test);
- quantification of the GM seed level using either a sub-sampling quantification method (also named semi-quantitative test) or a quantitative test (most often by PCR).

The same ISTA PT overall rating procedure, as used for Purity of Germination, is used here as described in Table 1.

Depending on the results obtained by a laboratory on a given PT, the laboratory is rated A, B, C or BMP. The definitions below show

the philosophy of the system which can be applied to many different types of tests:

- A: no problem has been detected in this test.
- B: there are small problems, but no specific investigation or action is suggested to the participant.
- C: problems, ISTA may indicate there are things the lab needs to explain or correct.
- BMP (Below Minimum of Performance): ISTA indicates by a letter that the results were poor and the laboratory has to explain and correct things.

Each rating communicates a number of points, the bigger the number, the better the results. The decrease from A to C and the "0 points" for BMP, are another way to understand the meaning of these ratings.

The rating for a given PT is an indication to

the laboratory on its performance in this test.

In the context of GMO detection, two rating systems are defined depending on the nature of the test ; qualitative to detect presence/absence, and semi-quantitative or quantitative test to quantify the presence. A computational procedure is used to establish the laboratory rate, and the final decision is left to appointed experts for the PT. In this article, we describe the computational procedure used in the rating system to quantify the presence of GM seeds.

General overview

The decision tree in Figure 1 exhibits the general definitions used for defining A, B, C and BMP ratings in GM quantification.

These general definitions have been translated into computational terms as exhibited in Figure 2.

The system is based on two main quantities: the true levels and the z-scores. They are described in the next two sections. Samples with a zero spiking level are not used in quantification rating, but they are considered in the rating system for qualitative tests as described in Seed Testing International N° 128 pages 8-10 October 2004. The computation

Table 1: ISTA rating system for 6 PTs based on the in-round rating values.

One test rating	One test Score Value	Overall rating on 6 tests	Range on 6 tests
A	5 points	A	28 – 30 points
B	4 points	B	21 – 27 points
C	3 points	C	16 – 20 points
BMP	0 points	BMP	below 16 points

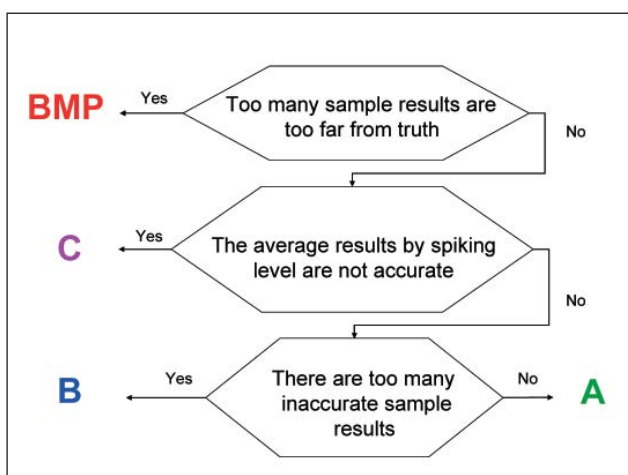


Figure 1: Decision tree defining A, B, C and BMP ratings in semi-quantitative/quantitative PTs

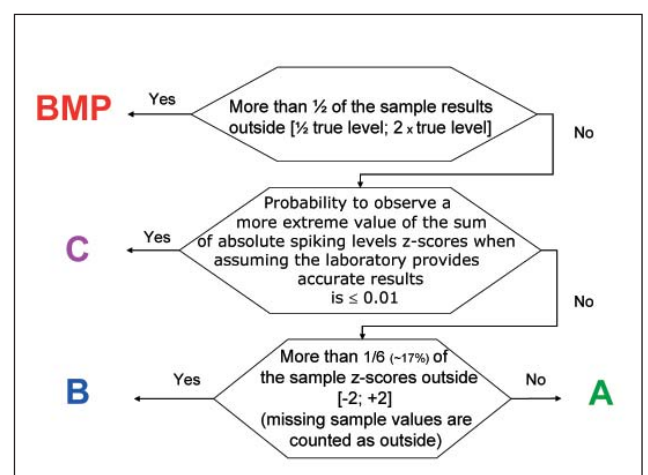


Figure 2: Analytical definition of A, B, C and BMP ratings in semi-quantitative/quantitative PTs

associated with each decision box in the decision tree in Figure 1 and 2 is illustrated with a graphic: each laboratory is plotted on the y-axis using an artificial number for the identification and a letter indicating the test used by the laboratory (S for semi-quantitative tests, Q for quantitative tests). An artificial number is used to prevent identification of laboratories by non appointed persons. There is one panel for each non-zero spiking level. Examples of these graphics are provided, using the 4th PT (PT4) for which 3 samples for each of 3 non-zero spiking levels were sent to the laboratories resulting in a total of 9 non-zero samples.

True level definition

In order to accommodate various measurement units, three true levels for each sample sent to the laboratories have been defined:

- Ratio of the number of GM seeds to the total number of seeds in the sample;
- Ratio of the weight of GM seeds to the weight of the sample;
- For a given spiking level, median of the sample results reported by the participating laboratories.

This results in three sub-rating systems and thus three ratings are available to the experts.

Z-scores

Z-scores are useful to establish rules from distributions with different means and/or different standard-deviations. They are already used in ISTA PT rating systems, on Germination and Purity for example.

The definition of a z-score is as follows:

Consider a value x from a distribution with mean μ and standard-deviation σ . The formula for converting x into its corresponding z-score is:

$$z = \frac{x - \mu}{\sigma}$$

This value indicates how far and in what direction x deviates from μ , in units of σ .

The distribution's mean and the distribution's standard-deviation of the z-scores are equal to 0 and 1 respectively. When the distribution of reference is normal, the z-scores distribution is also normal and thus the probability to have a z-score in the interval $[-2 ; +2]$ is approximately 0.95. This property is used in the PT rating.

BMP rating

The rule is:

If more than half of the sample results are outside the acceptance interval defined by $[\frac{1}{2}$ true level; $2 \times$ true level], then the rating is BMP.

This rule assigns a BMP rate when too many sample results are too far from the true level. As the variance of the results is an increasing function of the true level, the acceptance interval decreases with the true level as exhibited in Figure 3.

Figure 4 and Figure 5 visualize the results for PT4 when two different units are used for the true level, i.e. percent of GM seeds by number and percent of GM seeds by mass respectively.

In Figure 4, the two vertical purple lines for a given spiking level define the acceptance intervals. These intervals are the same across laboratories whereas they are unique for each sample in Figure 5 as the true level defined in percent seed by mass is unique for each sample.

When all the 9 sample results are available from a laboratory, the BMP rating is assigned if the number of samples out of the acceptance intervals is superior to 4. This limit might be different for laboratories not reporting for some reasons results for all the samples: for example, a laboratory might provide results for 6 samples only and thus, the limit is equal to 3 for this particular laboratory.

Laboratories with a BMP rating are identified in the top of the graphic.

In Figure 4, BMP rating is assigned to 6 labo-

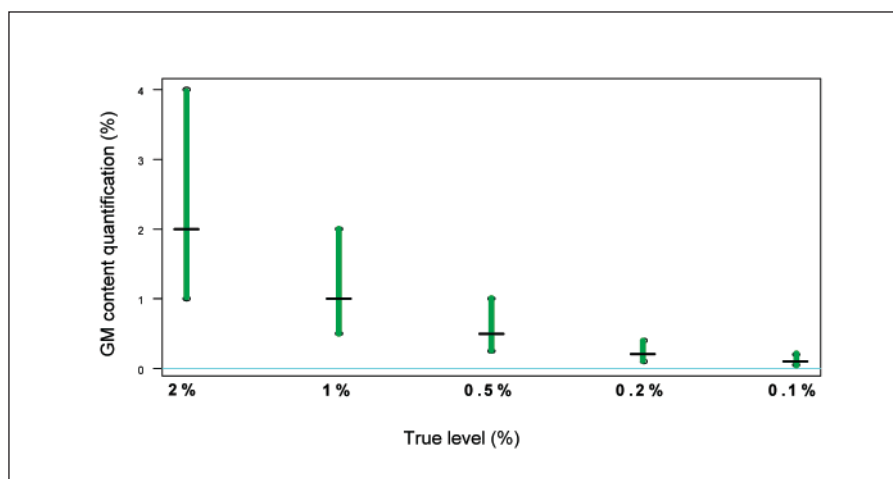


Figure 3: Size of the “acceptance” intervals (in green) as a function of the true level

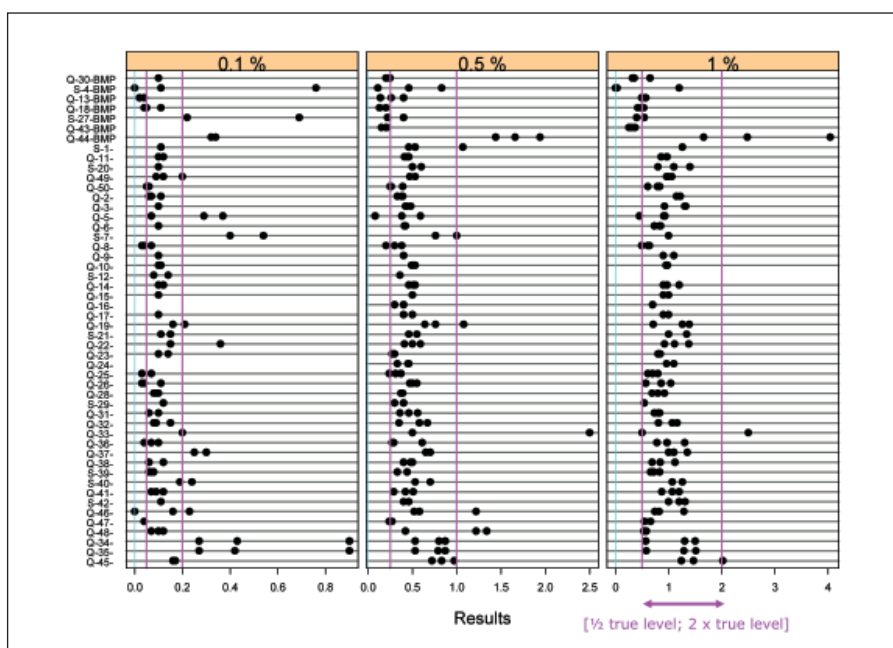


Figure 4: Graphic used for the BMP rating in PT4 when the true level is expressed in percent seed by number

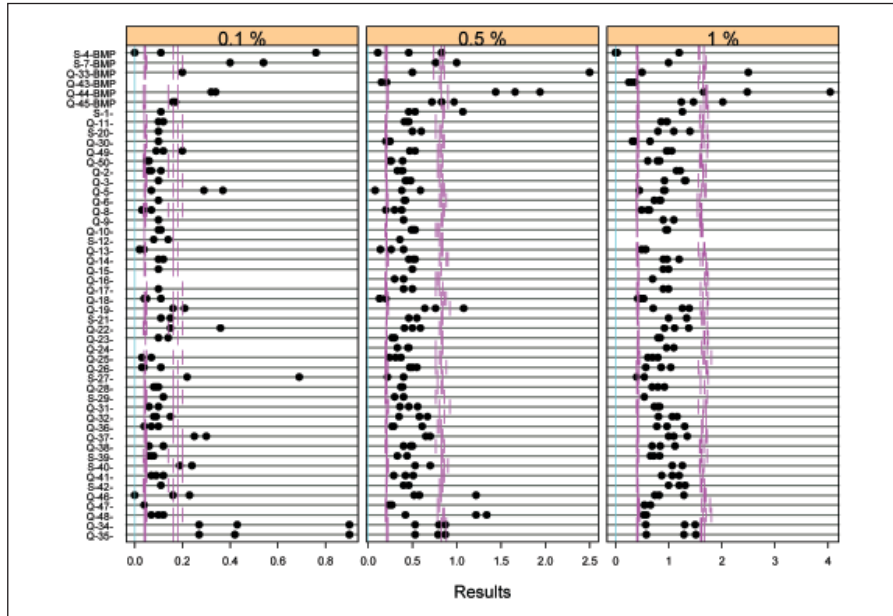


Figure 5: Graphic used for the BMP rating in PT4 when the true level is expressed in percent seed by mass

ratories. It is also assigned to 6 laboratories in Figure 5, 3 laboratories having a BMP rating for both units (Laboratory #4, #43 and #44). Using the percent seed by number is more appropriate when the sub-sampling strategy is used, using the percent by mass is more appropriate when quantification is made by a PCR method.

Reference intra-laboratory standard-deviation

Before going into the details of the C, B and A rating computations, let's define the reference intra-laboratory standard-deviation. Conceptually, this quantity represents the average intra-laboratory variation for a given spiking level. Computationally, we have used the following procedure to estimate it:

- For each laboratory k and each spiking level i , compute the variance of the sample results: $\hat{\sigma}_{ik}^2$
- Carry out the Cochran's test at the 95% level to see if the laboratory with the highest variance has an outlying spread of replicates.
- Estimate σ_i^2 by $\hat{\sigma}_i^2 = \text{mean of the } K \text{ variances } \hat{\sigma}_{ik}^2$ if no outlying variance has been identified with the Cochran's test, $\hat{\sigma}_i^2 = \text{mean of the } K-1 \text{ smallest variances otherwise.}$
- Estimate the reference intra-laboratory standard-deviation for spiking level i with: $\sqrt{\hat{\sigma}_i^2}$

C rating

The rule is:

If the probability to observe a more extreme value of the sum of absolute spiking levels z-scores when assuming the laboratory provides accurate results is ≤ 0.01 , then the rating is C.

This rule is stated using a terminology pertaining to statistical hypothesis tests. The idea here is to obtain an assessment of the overall accuracy of the average results by spiking levels and to rate C a laboratory for which this overall accuracy is not acceptable.

Computationally, this is done using z-scores and the reference intra-laboratory standard-deviation defined in the previous section. We first compute the following z-scores for each laboratory and each spiking level:

Notice that the reference laboratory standard-deviation above is divided by the square root of the number of sample results reported as this will constitute the reference standard-deviation of the mean of the sample results.

$$z = \frac{(\text{mean of the sample results}) - (\text{true level})}{(\text{reference lab stddev}) / \sqrt{(\text{number of reported samples})}}$$

Then we consider the theoretical distribution of the sum of these absolute independent z-scores. This distribution can be defined using simulations as described in Tattersfield publication (Seed Science and Technology, vol.7, No.2, pages 247-257) and will lead to the definition of the rejection regions at the 0.01 level to comply with the hypothesis testing statement in the rule. Table 2 provides these rejection regions when the number of spiking

levels varies from 1 to 5.

Table 2: Rejection region for C rating.

# of spiking levels	Rejection region at the 0.01 level
1	2.55
2	3.97
3	5.25
4	6.43
5	7.55

Thus, if a laboratory reported sample results for 3 spiking levels and if the sum of the absolute 3 z-scores defined above is superior to 5.25, a C rating is assigned.

Figure 6 visualizes the results for PT4 when the true level for a given spiking level is defined to be the median of the sample results reported by the participating laboratories.

In this graphic, the x-axis is the sum of the absolute z-scores for the mean of the result samples reported by the laboratories. Laboratories corresponding to a C rating are displayed in the bottom (17 laboratories). The rejection region is visualized with a vertical purple bar. There are 4 laboratories with a rejection region different to 5.25 (circled in blue): these two laboratories reported results for 2 samples only and thus the corresponding rejection region is 3.97. Note that a z-score can be positive or negative, but we use the absolute value to sum z-scores for a laboratory.

B and A rating

The rule is:

If more than 1/6 of the sample z-scores are outside $[-2; +2]$ (missing sample values are counted as outside), then the rating is B otherwise it is A.

In this rule, the z-scores for a given laboratory are computed for each spiking level by sample combination as follows:

$$z = \frac{(\text{sample result}) - (\text{true level})}{\text{reference lab stddev}}$$

As a z-score outside the interval $[-2; +2]$ has low probability (0.05) to occur for a laboratory providing accurate sample results, the interval $[-2; +2]$ is used to qualify the accuracy of a sample result.

Figure 7 visualises the results for PT4 when the true level is expressed in percent seed by

number.

Laboratories having an A rating are displayed on the top of the graphic, laboratories having a B rating in the bottom of the graphic. The two vertical purple lines for each spiking level visualize the interval [-2 ; +2] for the z-scores. A point outside these limits indicates an inaccurate sample result reported by the corresponding laboratory. Table 3 gives the maximum number of inaccurate sample results that can be accepted for an A rating.

Table 3: Maximum number of z-scores out of [-2 ; +2] that can be accepted for A rating.

# of samples with reported result	Max # of z-score-out of [-2;2]
1 to 5	0
6 to 11	1
12 to 17	2
18 to 23	3

Combining the different decisions

Figure 1 and 2 describe how the rating system is combining the different individual decisions that have been detailed in the previous sections. Figure 8 provides a summary of

the PT4 ratings for the 51 laboratories which provided quantitative results, using the three definitions of the true level.

The distributions of the PT4 rating are nearly identical for the different true level definitions.

The details of the different ratings for the different laboratories as well as all the graphics used to illustrate the different decisions and the comments provided by the laboratories are available to the appointed experts that make the final decision regarding the rating. ■

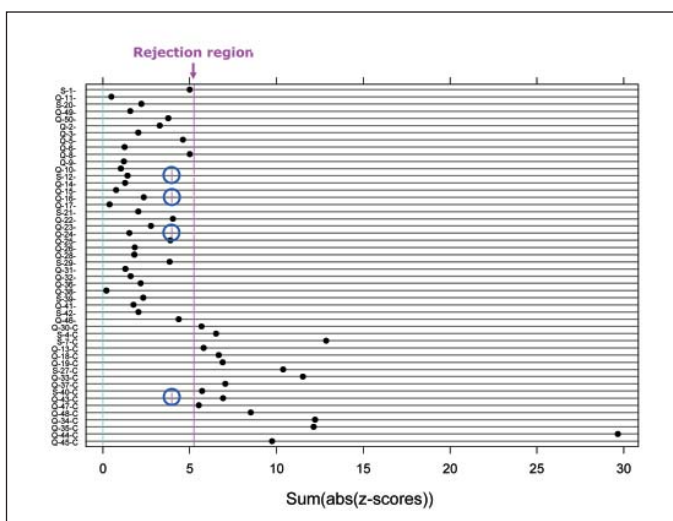


Figure 6: Graphic used for the C rating in PT4; the true level here is the median of the sample results.

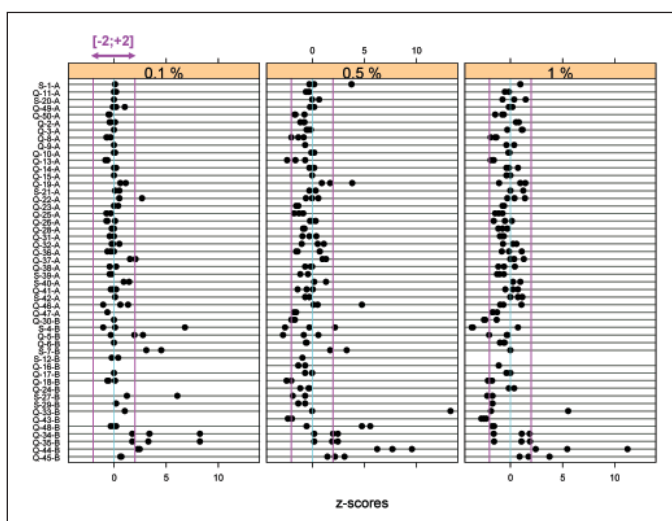


Figure 7: Graphic used for the A/B rating in PT4 when the true level is expressed in percent seed by number

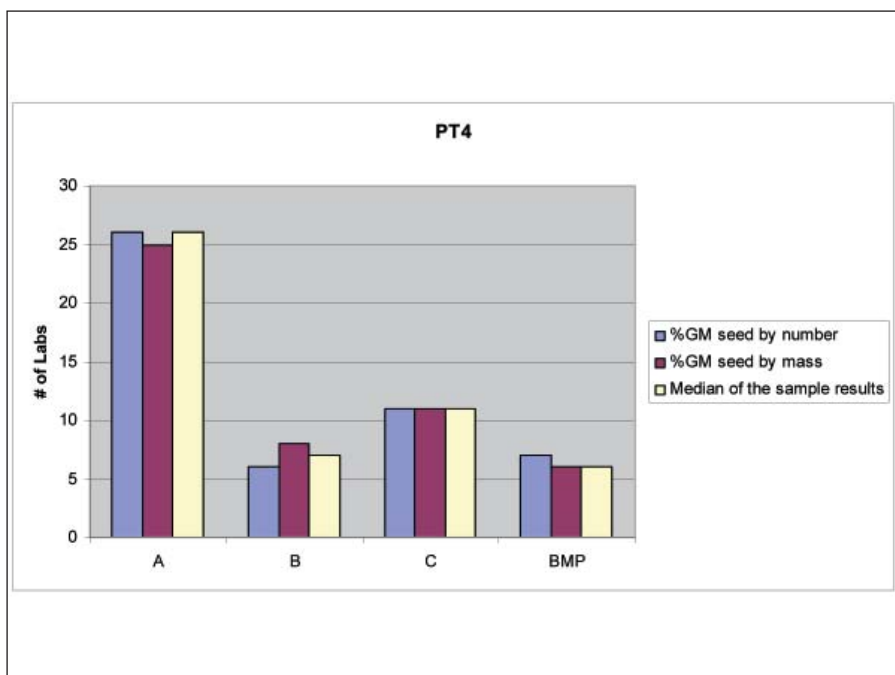


Figure 8: Barchart giving the number of the PT4 participating laboratories having rating A, B, C or BMP using three definitions of the true level

Announcement

6th ISTA Proficiency Test on GMO Testing on *Brassica napus* (GT73)

The aim of the proficiency test is to check the ability of individual laboratories to detect the presence or absence of GM seeds and to quantify their presence in samples of conventional seeds of *Brassica napus*.

Each participating laboratory will receive canola test samples. Some of the samples will be positive (i.e. contain GM seeds) and others will be negative (i.e. contain no GM seeds).

A qualitative test result can be either derived from the quantitative test result or from a separate test on the sample. The GM seed material in the positive samples can be quantified either by a sub-sampling quantification or by a quantitative test (e.g. real time PCR). An estimate of percentage GM seeds per positive test sample shall be reported.

Laboratories interested in participating should please contact the ISTA Secretariat:
 Email: ista.office@ista.ch
 Fax +41-44-838-6001

More details can be found on the ISTA Website at www.seedtest.org