PT Evaluation Procedures

August 28, 2025
Version 3.0

PTC | Proficiency Testing Canada

Table of Contents

1.0 Introduction

This document summarizes the data evaluation procedures performed by Proficiency Testing Canada Inc. (PTC) in the course of a PT Round.

General overview of the PTC PT Scheme with information on design and implementation can be viewed in *PAR03 PT Scheme*.

2.0 Procedures

The following section details the chronologic steps in the conduct of a PTC PT Round.

2.1 POSTING PT SCHEDULE

By November of each year, a schedule of shipping dates, reporting dates and change deadlines is posted on the PTC website for the following year. Any changes made to the schedule after initial posting will be highlighted in red. This schedule is arrived at in consultation with the PTC subcontractors.

2.2 NOTIFICATION OF PT SAMPLE SHIPMENT

At least eight weeks prior to a scheduled PT round, participants are notified by email of the pending shipment and sent the PT invoice, which is also available for download from the PTC Portal.

2.3 SUBCONTRACTOR NOTIFICATION, WEBSITE UPDATE AND SHIPMENT TRACKING

Two weeks before the scheduled PT round, Instruction sheets and ship-lists are produced and provided to the subcontractors. These are used by the subcontractor to package shipments for participants.

Subcontractors ship samples by FedEx unless otherwise specified by the participant. Before each shipment the subcontractors provide an EXCEL sheet with all tracking numbers for the shipment. PTC uses the online tracking systems for each courier to track each shipment. The delivery date is recorded on this sheet. For any shipment that is unduly delayed, the subcontractor is notified for investigation.

A list of the PTC subcontractors who prepare, handle and distribute proficiency testing items can be viewed in *PAR05 List of PT Subcontractors*.

2.4 PROCESSING PARTICIPANT DATA

2.4.1 Review of PTC Portal

Following the reporting deadline, and prior to the evaluation of participant data, the comments entered by participants during reporting are reviewed.

2.4.2 Flag All Non-Detects and Non-Reports

All results that are qualified (<, >) are temporarily removed from the initial set of raw data for the purpose of calculating the summary statistics.

2.4.3 Outlier Removal

For the purpose of calculating summary statistics, outliers are temporarily removed from the individual data-sets. Statistical applications for the identification of outliers is not used as assigned values and SDPA are estimated using robust procedures that minimize the impact of outlying data. Only very obvious outliers are flagged for removal. This would include data reported in the wrong units or other gross errors.

2.4.4 Assess Homogeneity and Stability

The uncertainty associated with sample homogeneity and sample stability should not contribute significantly to the overall uncertainty of the PT evaluation.

For each Test Group/Analyte/Sample combination, regression analysis is conducted on participant reported result against bottling order. A possible homogeneity problem will display a trend in the data (increasing or decreasing) where the slope of the regression is significantly different than zero at $\alpha = 0.05$ and the ratio of the maximum deflection of the regression curve to the standard deviation is greater than 1.

As well, PTC assesses stability by plotting the reported result against the date of analysis.

A possible stability problem will display a trend in the data (increasing or decreasing) where the slope of the regression is significantly different than zero at $\alpha = 0.05$ and the ratio of the maximum deflection of the regression curve is greater than 1.

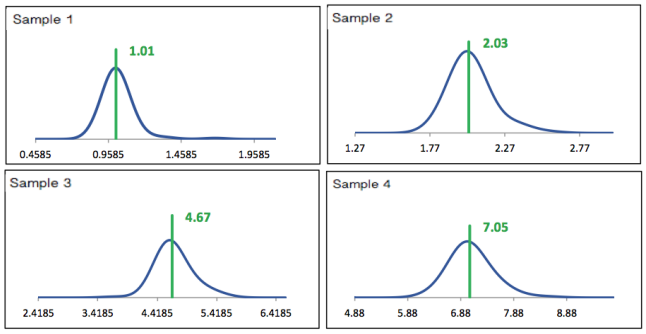
When homogeneity or stability is flagged as a possible problem, the standard deviation of proficiency assessment is increased to the point at which the ratio mentioned above is ≤ 1.

This approach has the benefit of taking conditions during sample shipping into account.

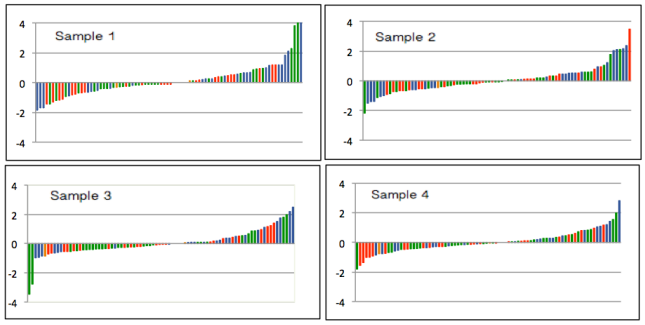
The only samples that are not assessed for homogeneity and stability is C05B Microbiology-presence/absence. However, these samples are prepared in the same way, using the same bacterial strains, as C05A for which homogeneity and stability are assessed. As well, serious homogeneity or stability problems would be reflected in the failure rate which, for this PT, rarely exceeds one or two per round.

2.4.5 Examination of Data Distribution

The statistical assumption of most PTC PT Schemes is that the data reported does not show bi-modality and that it approximates a normal distribution. This is examined through the use of kernel density plots. This type of plot is good for identifying modality and skewing problems.



As well, ranked *z*-score plots are produced. For further investigatory value, the bars are colour-coded to denote the method of analysis used by the participant. Whenever one method is found to be obviously and recurrently biased relative to other methods, laboratories using this method may be excluded from the evaluation and participants notified that the PT is not appropriate for that method. When this is done, a note is also included in PAR02 –Catalogue.



2.4.6 Examination of Summary Statistics

The last step is to review the summary statistics for the analyte. This table will include things like the number of participants, median, robust mean, robust standard deviation, uncertainty of the assigned value and if an analyte was flagged for possible homogeneity and/or stability concerns. As well, a break-down of the number of results reported by the most common methods is displayed.

Summary Statistics

Statistic	C01A-1	C01A-2	C01A-3	C01A-4
N	127	127	127	127
Median mg/L	125	84.0	63.7	164
Robust Mean mg/L	124	83.9	63.9	165
U mg/L	0.438	0.327	0.314	0.530
Robust Standard Deviation mg/L	3.95	2.95	2.83	4.78
Regression Standard Deviation mg/L	9.32	6.29	4.79	12.4
Stability Flag				
Homogeneity Flag				
Standard Deviation Used (SDPA) mg/L	9.32	6.29	4.79	12.4
Outliers	0	0	0	0
|z|>3.0	0	0	2	0
2<|z|<3	3	4	1	0

Methods Used

Method	C01A-1	C01A-2	C01A-3	C01A-4
TITR (Blue)	77	77	77	77
CLR-A (Red)	4	4	4	4
COLOR (Green)	8	8	8	8
TITR-A (Orange)	38	38	38	38

2.4.7 Exclusion of Sample from Evaluation

At any step during this examination, one or more of the samples can be excluded from the evaluation. When this occurs, the final evaluation is based on the remaining samples.

2.4.8 Estimation of Assigned Value and SDPTA

Prior to the issuing of any PT reports, the data sets for all Test Group/Analyte/Sample combinations are examined to ensure that the evaluations being made are statistically valid. When participation levels are less than 11, or if there are a large number of non-detects, the data will be examined to ensure that the assigned value is an acceptable estimate. If there is any doubt, the sample will not be evaluated.

When availability allows, the data sets for all Test Group/Analyte/Sample are reviewed by the Executive Director and Program Officer. In this case, plots comparing the assigned values and SDPTAs obtained by each is graphically compared. Whenever there is a significant difference, the data is examined in greater detail to determine the cause of the difference.

2.5 PRODUCE AND DISTRIBUTE PRELIMINARY REPORTS

Following the procedures detailed in Appendix I, Preliminary PT Reports are prepared in EXCEL format and emailed to participants. This is done within one week of the reporting deadline.

2.6 PRODUCTION AND REVIEW OF TEST GROUP SUMMARY REPORTS

As part of the examination of data detailed above, a Test Group Summary Report is produced for each test group. Each report contains:

- Summary of evaluation procedure;
- Summary statistics;
- Sorted scatter plots;
- Colour-coded z-score plots;
- Kernel density plots to visualize the nature of the data distribution;
- Homogeneity and Stability regression plots;
- Box and Whisker plots; and,
- Plot of robust mean against robust standard deviation for the ten most recent rounds.

This examination is used to make decisions about the acceptability of the default evaluation procedure or if a modification is required. These reports are of value to PT participants investigating poor PT performance.

2.7 FINAL PT REPORTS

If there are any deviations from the published PT evaluation procedure (e.g., excluding an analyte or a sample) or when there is other information that might be of value to the participant, this is documented in the cover page of the Final Report.

An email with the attached PT Report is sent to each participant.

2.8 REVISED PT REPORTS

Whenever a non-conformance is identified after the Final Proficiency Testing Report has been issued, the non-conformance is investigated and, if necessary, results are re-evaluated and a revised Proficiency Testing Report is issued. The reports are clearly indicated as being Revised reports and the cover page(s) of the revised report will include a brief statement about the reason for the revision and what was revised.

Whenever a Revised Report is issued, the affected participants are notified by email along with the attached Revised Report. The Revised Report is always issued to all the participants in that PT scheme for that PT Round. Prior to issuing the Revised Report, an analysis of the potential impact on the other participants Robust mean is conducted. This analysis is to ensure there is no influence on the general performance of the other participants.

2.9 RECALLED PT REPORTS

If a non-conformance is identified after the Final Proficiency Testing Report has been issued and, following an investigation, PTC is not able to provide evaluation for 1 or more clients, then the original report sent is recalled and a notification issued.

3.0 History of Changes

Date	Rev. No.	Sections	Changes
12/18/2019	1.0		Initial publication
11/20/2020	2.0	2.4	Rewritten to reflect the new software used to process PT data and create the Test Group Summary reports.
		2.4.4	Statement added that homogeneity and stability assessments are not performed for C05B.
01/13/2021	2.1	Appendix I	Added clarification about estimating the SDPA from the fixed limits.
01/28/2021	2.2	AI.3	Corrected the equation.
03/28/2022	2.3	2.3	Expanded on the procedure for tracking PT shipments.
11/04/2022	2.4	2.9	Added 2.9 to depict actions taken if nonconformance is identified after PTC Final report is issued.
01/04/2023	2.5	2.4	Added additional information on who is involved in the Data Analysis process.
03/30/2023	2.6	1.0 & 2.3	Added reference to other PTC Documents related to the PTC PT Program.
		2.8	Added clarification about Revised PT Reports with regards to recipients.
12/18/2023	2.7		Updated to reflect new AMS system
06/06/2024	2.8		Replaced reference to Web-Data entry with PTC portal
08/26/2024	2.9	AI.5	Clarified the handling of non-detects.
08/28/2025	3.0	AIII	Added Appendix III for Evaluation of Fast Track Samples.

Appendix I: PT EVALUATION PROCEDURE

The following procedure is used for each test group except C05B microbiology by presence/absence and The PCB aroclors in C06B, C08 and C35. The details for these can be found in Appendix II.

AI.1 ASSIGNED VALUE ($\overline{X}$)

The default condition is to use the Robust Mean as described in Algorithm A of ISO 13528-*Statistical methods for use in proficiency testing by interlaboratory comparisons*.

AI.2 ROBUST STANDARD DEVIATION

The default condition is to use the Robust standard deviation as described in Algorithm A of ISO 13528-*Statistical methods for use in proficiency testing by interlaboratory comparisons*. Care must be taken when using this procedure for small data-sets containing numerous identical values. This will result in an unreasonably low estimation of the standard deviation or may result in a division-by-zero error. When this is observed, the arithmetic standard deviation will be used.

AI.3 REGRESSION EQUATION STANDARD DEVIATION

The regression equation standard deviation is estimated using the regression equations estimated from historic studies or adopted from established fixed limits (see PROC11- *Regression Equations*).

Regression Equation Standard Deviation

$$stdev = m \times \bar{X} + b$$

where m = slope of regression equation
$\overline{X}$ = Assigned Value
b = intercept of regression equation

Fixed Limits Standard Deviation

$$stdev = (\#\#\% \times 0.5) \times \bar{X}$$

AI.4 STANDARD DEVIATION FOR PROFICIENCY ASSESSMENT (SDPA)

The standard deviation for proficiency assessment used to calculate the *z*-score is the greater of either the Robust Standard Deviation or the Regression Equation Standard Deviation.

If the assessment of homogeneity or stability has identified a concern, the SDPA is set as the maximum deflection of the homogeneity or stability regression curve, whichever is greater.

AI.5 *z* - SCORES

The outliers and non-detects are added back into the data-set in order to prepare the PT Reports for all the participants.

The Assigned Value and the Standard Deviation of Proficiency Assessment are rounded to the number of significant figures and decimal points that will appear on the final report. The *z*-score and absolute z-score for each reported result is calculated as:

$$z - score = \frac{(x_i - \bar{X})}{SDPA}$$

$$Absolute\ z - score = \frac{|(x - \bar{X})|}{SDPA}$$

If a laboratory has reported it's RDL in the PTC portal the following equations are used:

$$z - score = \frac{(x_i - \bar{X})}{\sqrt{SDPA^2 + (RDL/3)^2}}$$

$$Absolute\ z - score = \frac{|(x_i - \bar{X})|}{\sqrt{SDPA^2 + (RDL/3)^2}}$$

Where: x_i = reported result
$\bar{X}$ = assigned value
SDPA = standard deviation of proficiency assessment
RDL = the reporting detection level

Exception 1-Very High *z* Scores: If the calculated *z*-score is > 6.66 or < -6.66 then it is set to 6.66 or -6.66 respectively.

Exception 2-Non-detect Values: If the reported value is non-detect, the following rules apply:

- If reported non-detect value is less than the assigned value, the non-detect value is used to calculate a *z*-score;
- If reported non-detect value is greater than the assigned value, the *z*-score is left blank and not used in calculation of PT score;
- If all samples are accurately reported as non-detect, there is no PT score or evaluation calculated or displayed on reports.

Exception 3 - No Result Reported: If no result is reported, there is no PT score or evaluation calculated or displayed on reports.

AI.6 CALCULATE COMPOSITE PT SCORE

The average absolute *z*-score is calculated for each test group/analyte combination as

$$avgz = \frac{\sum absolutezscore}{N}$$

where absolute *z*-score = score as calculated above
N = number of samples per test group (generally 4).

Note: The N value will change if accurately reported non-detect values are reported.

The composite score for each test group/analyte is calculated as

$$PT\ Score = 100 + (-15 * avgz)$$

where *avgz* = average absolute z-score as calculated above.

AI.7 ESTIMATING AND FLAGGING BIAS

Biases are identified using the rescaled *z*-score procedure. Calculate the rescaled *z*-score as

$$RSZ = \frac{\sum z}{\sqrt{N}}$$

where *z* = the *z*-score as calculated above
N = the number of samples in the test group

Note: The N value will change if accurately reported non-detect values are reported.

Flags are assigned for each test group/analyte combination as follows:

- RSZ > -2 and < 2 no flag assigned;
- RSZ > 2 H (High);
- RSZ > 3 VH (Very High);
- RSZ < -2 L (LOW); and,
- RSZ < -3 VL (Very Low)

AI.8 STANDARD UNCERTAINTY OF THE ASSIGNED VALUE:

The standard uncertainty of the assigned value is estimated as:

$$u_x = 1.25 \times stdev / \sqrt{N}$$

Where stdev = inter-laboratory standard deviation as estimated in A1.2 above.
N = number of participants.

Note: The N value will change if accurately reported non-detect values are reported.

AI.9 INTERPRETATION OF PT RESULTS

The specific Acceptable/Unacceptable status for each test group/analyte combination is assigned as:

- PT score ≥ 70 = Acceptable;
- PT score < 70 = Unacceptable.

AI.10 FINAL REPORT

If any of the evaluations are modified (e.g., eliminate one sample, eliminate entire analyte, etc.) the Final Report clearly identifies the modified evaluation. The reason for the modification is also included in a PT Notice located in the cover page(s) of the Final Report.

The Final PT Reports are stored as pdf files on the PTC Server.

In addition to the pdf reports, EXCEL files containing the same information is also made available to participants.

Each Final Report contains information about the participant as well as,

Sample ID	PT code
Laboratory Information	Analyte
Method	Units
Assigned Value	Reported Value
SDPTA	*z* - Score
Bias flags	PT score
Acceptable/Unacceptable status	

The Test Group Summary Reports are uploaded to the PTC website for participant access.

Appendix II: PT Evaluation for Microbiology (Presence/absence) and PCB Aroclors

AII.1 MICROBIOLOGY BY PRESENCE/ABSENCE

For the C05B microbiology samples, an unacceptable evaluation is assigned for any false positive or false negative.

AII.2 PCB AROCLORS

PCB aroclors are evaluated as a combination z-score and presence/absence procedure. Each of the four samples in the test group is spiked with a single aroclor. For the aroclors that are not spiked into the sample, a threshold concentration is estimated as a fraction of the spiked aroclor concentration. For each aroclor, if a laboratory reports a false positive at a concentration above the relevant threshold value, then the aroclor is assigned an UNACCEPTABLE evaluation regardless of any calculated z-scores. If there are no unacceptable false positives, then the PT score is evaluated based on the *z*-score(s) of the sample(s) that were spiked with the aroclor.

Appendix III: Evaluation of Fast Track Samples

The following procedure is used for Test Groups where Fast Track samples are available. This is an area that is very flexible depending on availability. Please reach out to PTC to determine if your sample of interest is available at that time.

AIII.1 ASSIGNED VALUE

The assigned value (x_{pt}) is determined from at least 3 expert laboratories; the Mean of their data is used (ISO 13528-*Statistical methods for use in proficiency testing by interlaboratory comparisons).* These expert laboratories are determined from Performance Analysis of all laboratories that participate in that Test Group with PTC.

AIII.2 STANDARD DEVIATION

The standard deviation is determined from examining the last two PT Studies Regression Data. The average percent of the Robust Standard Deviation divided by the Robust Mean is used for the Standard Deviation (SD)

AIII.3 *z* - SCORES

The Assigned Value (x_{pt}) and the Standard Deviation (SD) are rounded to the number of significant figures and decimal points that will appear on the final report. The *z*-score and absolute z-score for each reported result (x_i) is calculated as:

$$z - score = \frac{(x_i - x_{pt})}{SD}$$

$$Absolute\ z - score = \frac{|(x_i - x_{pt})|}{SD}$$

AIII.4 CALCULATE COMPOSITE PT SCORE

The average absolute *z*-score is calculated for each test group/analyte combination as

$$avgz = \frac{\sum absolutezscore}{N}$$

where absolute *z*-score = score as calculated above
N = number of samples per test group (generally 4).

Note: The N value will change if accurately reported non-detect values are reported.

The composite score for each test group/analyte is calculated as

$$PT\ Score = 100 + (-15 * avgz)$$

where *avgz* = average absolute z-score as calculated above.

AIII.5 INTERPRETATION OF PT RESULTS

The specific Acceptable/Unacceptable status for each test group/analyte combination is assigned as:

- PT score $\geq$ 70 = Acceptable;
- PT score < 70 = Unacceptable.

AIII.6 FINAL REPORT

The Final PT Reports are stored as pdf files on the PTC Server.

Each Final Report contains information about the participant as well as,

Sample ID	PT code
Laboratory Information	Analyte
Method	Units
Assigned Value	Reported Value
SD	*z* - Score
Acceptable/Unacceptable status	PT score