NEPTUNE AND COMPANY
Improving the quality of environmental decision making

ISO 13528 – Criteria for Evaluation of PT Performance

Kelly Black
kblack@neptuneinc.org

Evaluation of PT Participants

- Third-party assessment of laboratory performance
- PT schemes have different objectives.
- Evaluation of PT Participants:

  Provide an assessment of laboratory results against **pre-established criteria** that align with the objectives of the scheme.

Pre-Established Criteria for Evaluation

Section 8 of ISO 13528 gives guidance on methods to determine a criteria for evaluation of performance.

The basic approach for all purposes is to compare laboratory results with an assigned value….and to give some allowable deviation from that value.

The allowable deviation is commonly defined as:

$$3 \times \sigma_{pt}$$

where σ_{pt} = the standard deviation for proficiency assessment.

What are the options?

- Repeatability and reproducibility known about the only allowed measurement method.
- Use of information from previous rounds of the PT scheme.
- Regulatory or expert statement of acceptability
- Repeatability based on a general model.
- Data from the same round of the PT scheme.

Choosing a Method…

…depends on the objectives of the scheme and what information is available.

Externally Derived Criteria

Historical Data – Other Participants

Historical Data – Method Measurement Uncertainty

Emerging Contaminant – new PT scheme

Choosing a Method…

…depends on the objectives of the scheme and what information is available.

Externally Derived Criteria

Historical Data – Other Participants

Historical Data – Method Measurement Uncertainty

Emerging Contaminant – new PT scheme

Fitness for Purpose

Repeatability and Reproducibility of a Measurement Method

What you have

- Historical data – Method Measurement Uncertainty
- A standardized measurement method

What you get

- Performance scores that are not subject to variation from reported results
- Consistent interpretation of performance scores
- "Garbage in – garbage out"

Repeatability and Reproducibility of a Measurement Method

When the measurement method used for PT is standardized and information on σ_R and σ_r is available, σ_{pt} may be calculated by:

$$\sigma_{pt} = \sqrt{\sigma_R^2 - \sigma_r^2 \, (1 - 1 / m)}$$

- $\boldsymbol{\sigma_R}$ = information on reproducibility
- $\boldsymbol{\sigma_r}$ = information on repeatability
- $\boldsymbol{m}$ = # of replicate measures per round per participant

Repeatability and Reproducibility of a Measurement Method

Old Family Recipe – comes out consistently delicious every time

Data from Previous PT Rounds

What you have

- Historical Data – Other Participants

What you get

- Criteria will not vary round to round
- Empirically based results

Data from Previous PT Rounds

- From previous rounds of PT with the same measurand and comparable property values, σ_{pt} can be determined.
- This is a useful approach when there is no agreement among experts about fitness for purpose.
- A lab with improved performance may fail.

Data from Previous PT Rounds

Traditional Recipe – good, but may not keep up with current tastes

Regulatory Defined Limit

What you have

- Externally Derived Criteria

What you get

- Performance scores that are not subject to variation from reported results
- Consistent interpretation of performance scores
- Fitness for purpose defined externally

Regulatory Defined Limit

A regulatory requirement may be set for the maximum allowable measurement error.

In this case, a percentage or raw number may be given as the Regulatory Defined Limit, and σ_{pt} can be determined from it (e.g., divide by 3).

Regulatory Defined Limit

Mother-in-Law's Recipe – don't dare to change anything

Expert Perception

What you have

- Expert Opinion

What you get

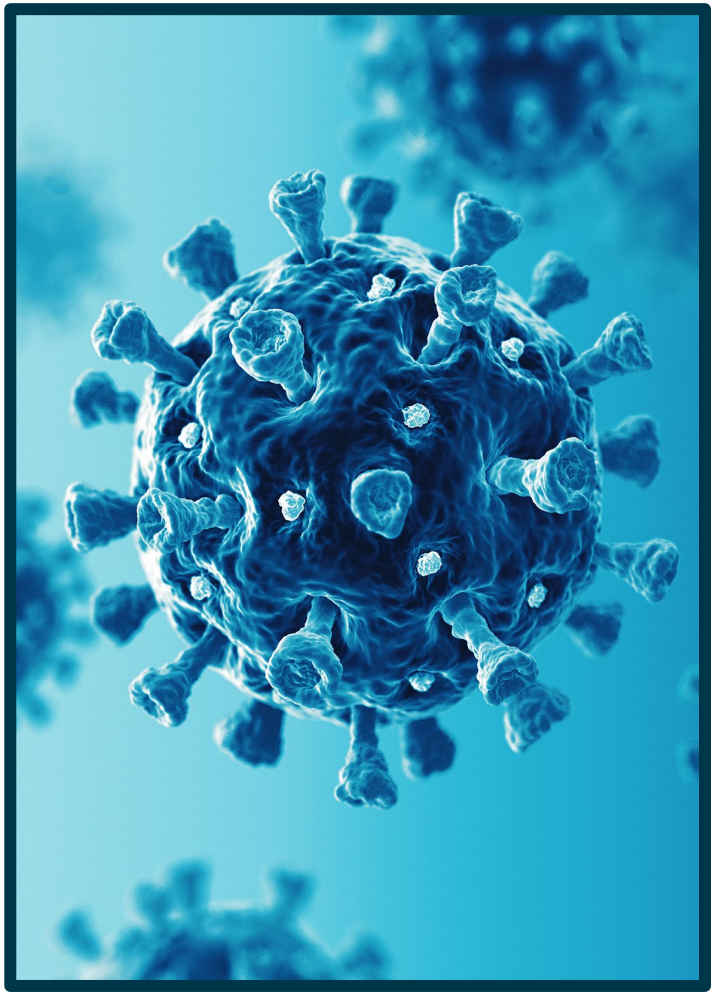
- A way to evaluate new PT schemes
- Potentially no agreement among experts

Expert Perception

σ_{pt} may be set by internal technical expert, a panel of external experts, a regulatory authority, or accreditation body.

This method is especially useful for emerging contaminants.

The evaluation criteria may change with each round as additional knowledge is gained.



Expert Perception

I hope the experts know what they're saying – it may be delicious, it may be awful

General Model – e.g., Horwitz / Thompson

What you have

- Single allowed test method
- Known method reproducibility

What you get

- Performance scores that are not subject to variation from reported results
- Consistent interpretation of performance scores
- Empirically based evaluation
- Reproducibility may lead to wide acceptance criteria

General Model

- Horwitz estimates reproducibility based on the concentration of the PT item.
- E.g., using Horwitz, lead in drinking water at 10 µg/L would have acceptance limits of +/- 2.2 µg/L.
- σ_{pt} derived from acceptance limits, perhaps 2.2/2 = 1.1 µg/L.

- General models are based on observations from collaborative trials of many parameters over an extended time period.
- The model should be fit for purpose.

General Model

Cake that may, or may not, fit the occasion

Data from Same PT Round

What you have

- Nothing
- Possibly an emerging contaminant

What you get

- See next slide

Data from Same PT Round

σ_{pt} is calculated from the results of participants in the same round of the proficiency testing scheme.

Main Advantages	Simplicity	Conventional acceptance	Feasibility
Main Disadvantages	Large variability	Potentially unreliable σ_{pt}	No useful interpretation

Data from Same PT Round

Variety – impacted by the results from each round

Conclusion

- There is a ~~cake~~ σ_{pt} for every occasion.
- Fitness for purpose is the guiding principle.