Evaluation of Proficiency Testing results

Maré Linsky

28 November 2016

SADCWater PT Evaluation Workshop 2016

Tanzania

Overview

- Background
- Evaluation of performance
  - Assigned value & Acceptable range
  - Performance statistics
  - Interpretation of results
    - Short term
    - Long term
  - Investigation
- Conclusion

What is Proficiency Testing?

- A regular independent assessment of the technical performance of a laboratory is necessary to assure the validity of measurements of tests and should be part of an overall quality strategy.
  - External quality assurance
    - Proficiency testing (PT)
    - Interlaboratory comparisons (ILC)
  - Complements Internal quality assurance measures
    - *Method development / validation*
    - Regular use of (certified) reference materials
    - Comparison of analysis by independent techniques
    - Control charts
    - Replicate tests
    - Intermediate checks on measuring equipment

Proficiency tests: Goals

- Compare its performance at a particular time against an external standard of performance

  How accurate is the data?

- Compare performance over a period of time

  Is it getting better or worse?

- Compare performance with that of other laboratories at a particular time

  Within peer group - how well does laboratory perform?

- Enable the organisers (regulatory authorities) to identify participants whose performance is unsatisfactory and establish whether there is a general improvement in performance in time

  Is the PT scheme doing its job of improving the quality of chemical measurements?

nmisa
National Metrology Institute of South Africa

Proficiency tests
Other goals

- Operated for the benefit of the participating laboratories, but data may be used from other organisations to aid understanding the capabilities and competence of the laboratories:
  - Accreditation bodies
  - Regulatory authorities
  - Customers of analytical services

Evaluation of performance

- Evaluation of performance against:
  - Assigned value
  - Acceptable range
- Determination of assigned value
  - Formulation
  - Certified reference value / Reference value
  - Consensus value:
    - Expert laboratories
    - Participants
- Determination of acceptable range / standard deviation
  - Prescribed value / Perception
  - Model, e.g. Horwitz curve
  - Precision experiment (standard method)
  - Standard deviation from participants' results

Evaluation of performance

Determination of assigned value

- Formulation
  - Spiking with known amount, concentration of an analyte to a base material containing none.
- Certified reference value (CRM)
- Reference value

For a reference value to be suitable its associated uncertainty should be 5 to 10 times better (smaller) than the uncertainties of the participants

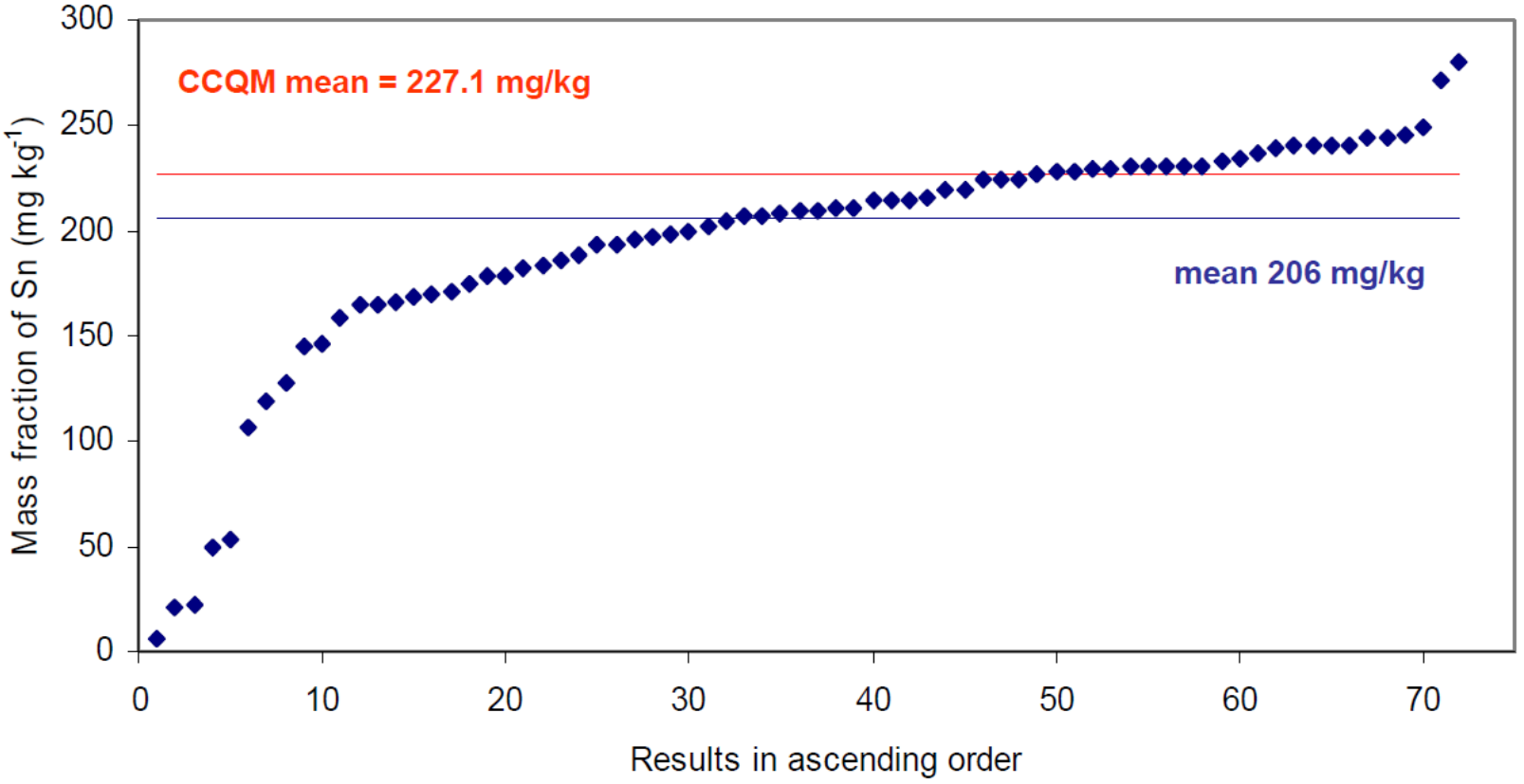
$$u_x \leq 0.3\hat{\sigma}$$

nmisa
National Metrology Institute of South Africa

Evaluation of performance

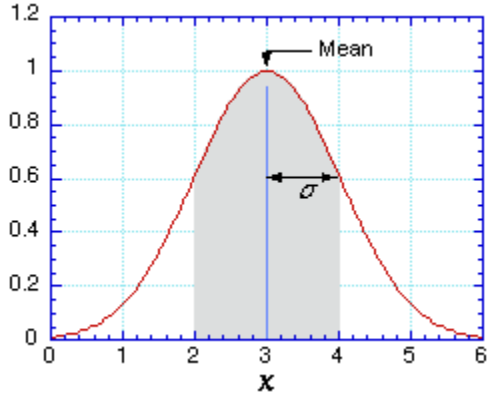
Determination of assigned value

- Consensus value:
  - Expert laboratories
    - Use of certified value of an analyte produced by group of expert or referee laboratories

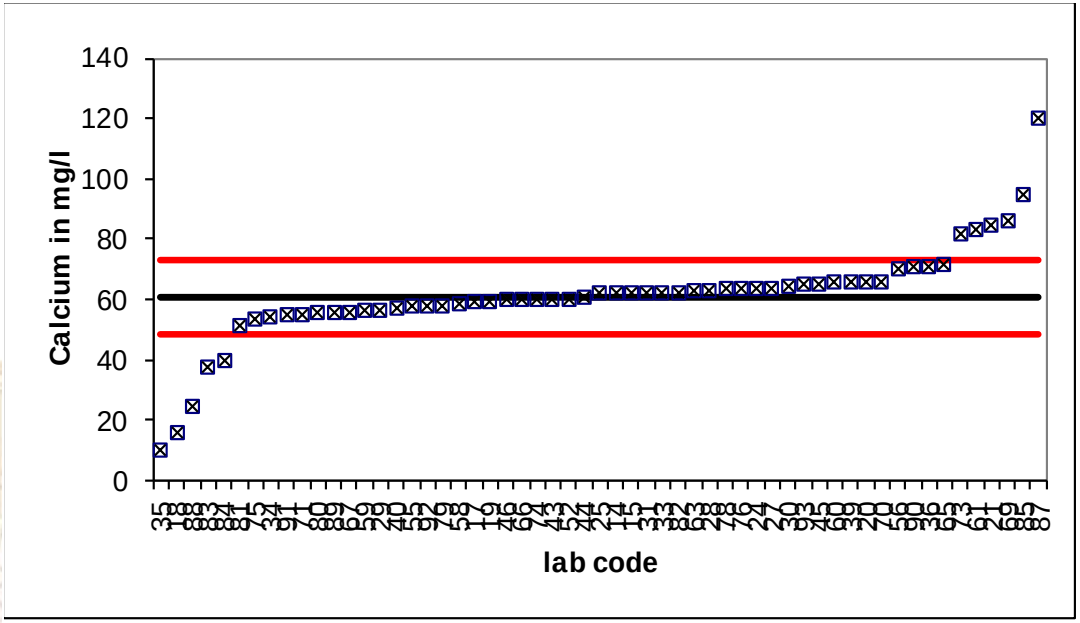


Evaluation of performance

Determination of assigned value



- Consensus value:
  - Participants
    - Parametric approach:
      - Normal distribution data
      - Arithmetic mean
      - Standard deviation - estimate of spread
      - Sensitive for deviating results, requires the use of outlier tests

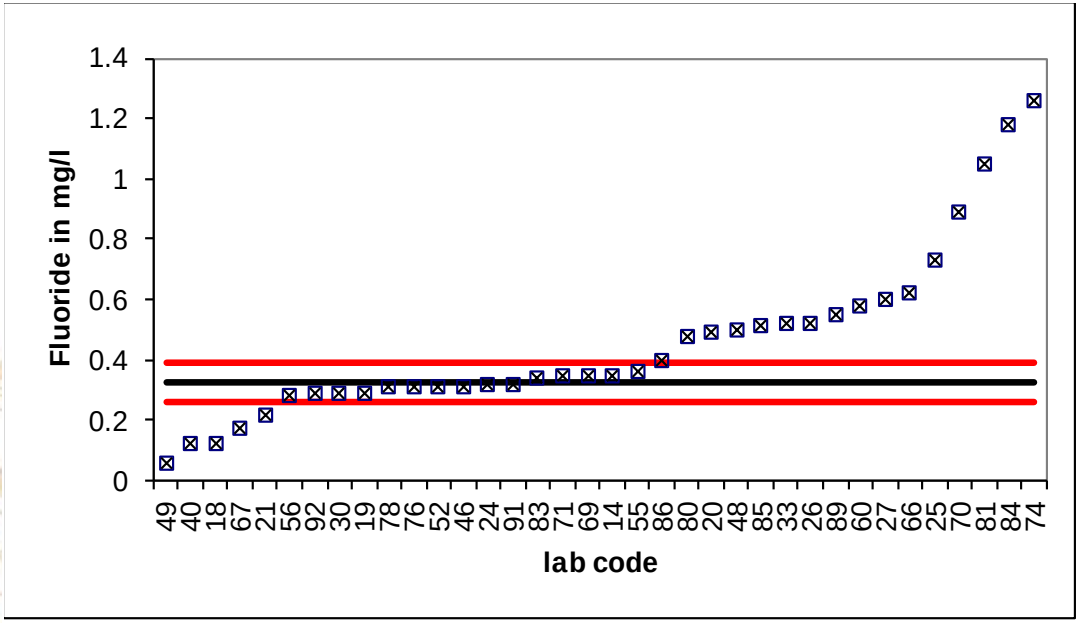


values:	59
removed:	2
mean:	61.38
ref.-value:	60.74
recovery:	101.1%
std:	7.336
rstd:	12.1%
std limit:	10%
upper limit:	72.89
lower limit:	48.59
too high:	6
too low:	7
outside limits:	13

Evaluation of performance

Determination of assigned value

- Consensus value:
  - Participants
    - Non-parametric approach:
      - Normal distribution of data is not required
      - Median
      - Median absolute deviation
      - Outlier tests are not required

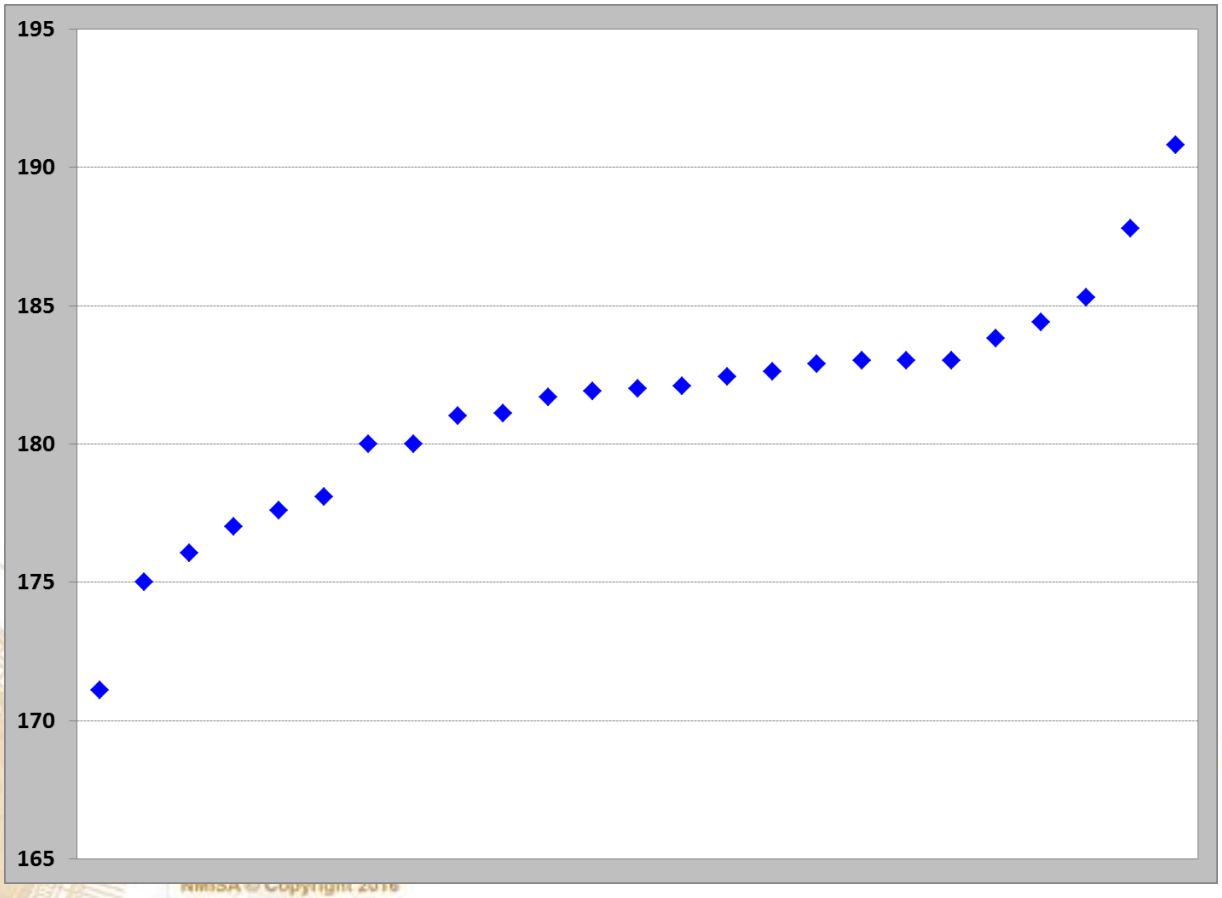


values:	**39**
removed:	**3**
mean:	**0.41**
ref.-value:	**0.33**
recovery:	**126.8%**
std:	**0.207**
rstd:	**63.5%**
std limit:	**10%**
upper limit:	**0.39**
lower limit:	**0.26**
too high:	**19**
too low:	**5**
outside limits:	**24**

Evaluation of performance

Determination of assigned value

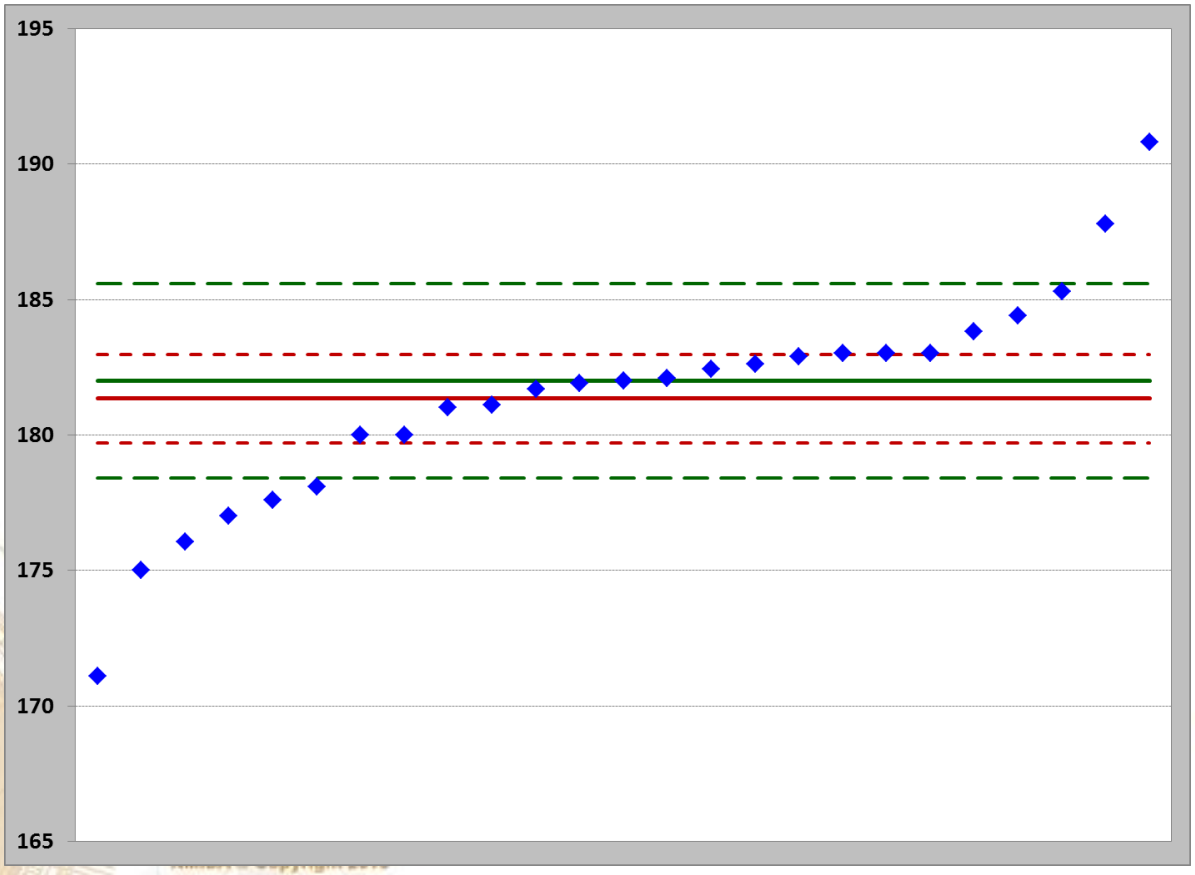
- Consensus value:
  - Participants
    - Parametric approach: **Zn in geological material**



Evaluation of performance

Determination of assigned value

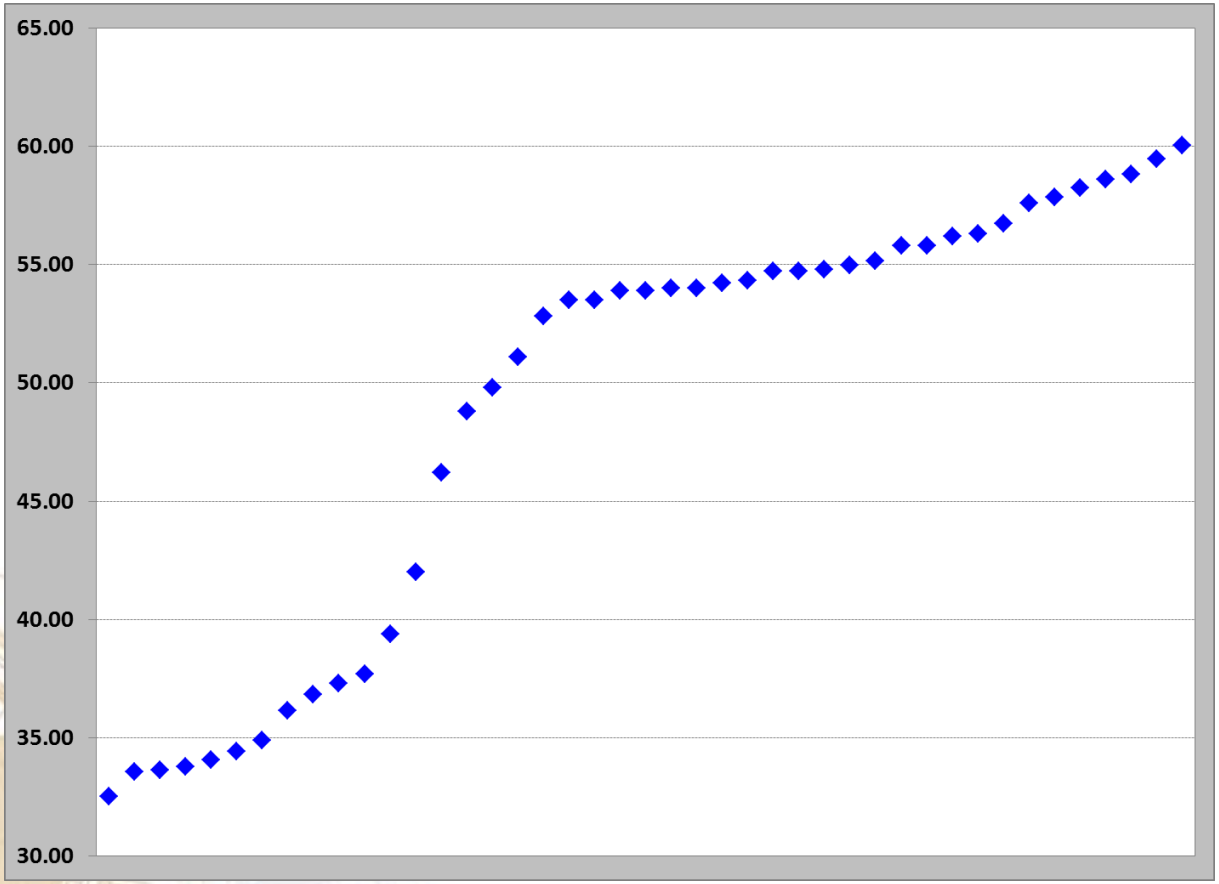
- Consensus value:
  - Participants
    - Parametric approach: **Zn in geological material**



Evaluation of performance

Determination of assigned value

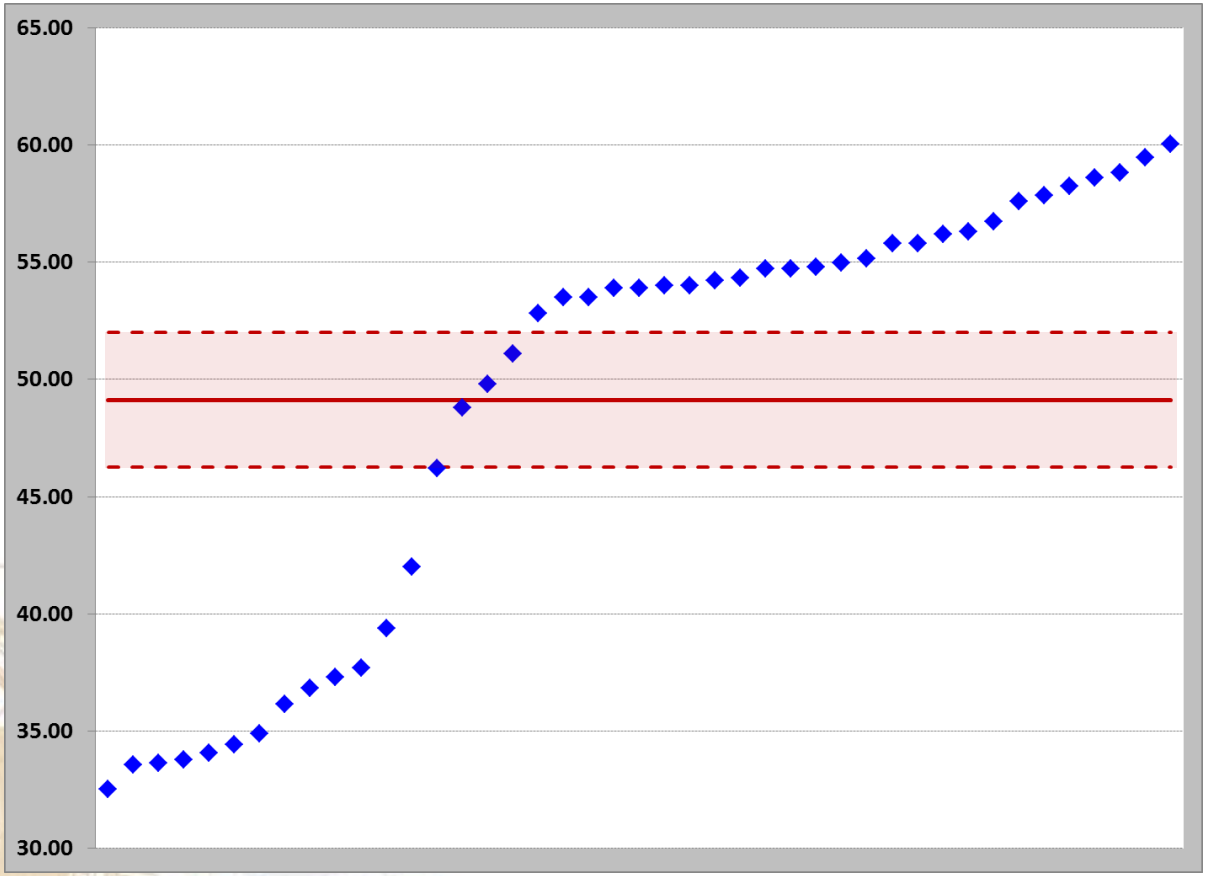
- Consensus value:
  - Participants
    - Non-parametric approach: **Ca in geological material**



Evaluation of performance

Determination of assigned value

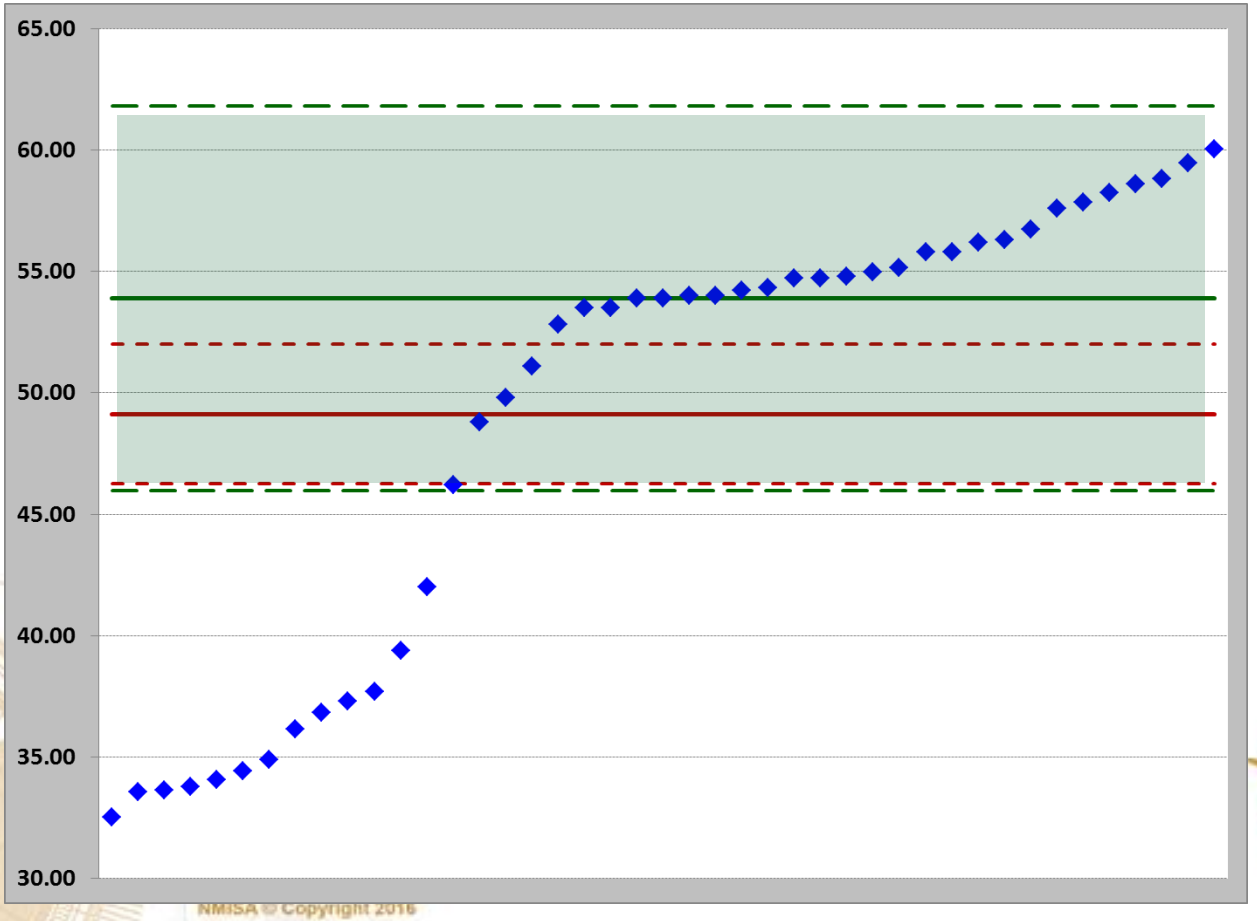
- Consensus value:
  - Participants
    - Non-parametric approach: **Ca in geological material**



Evaluation of performance

Determination of assigned value

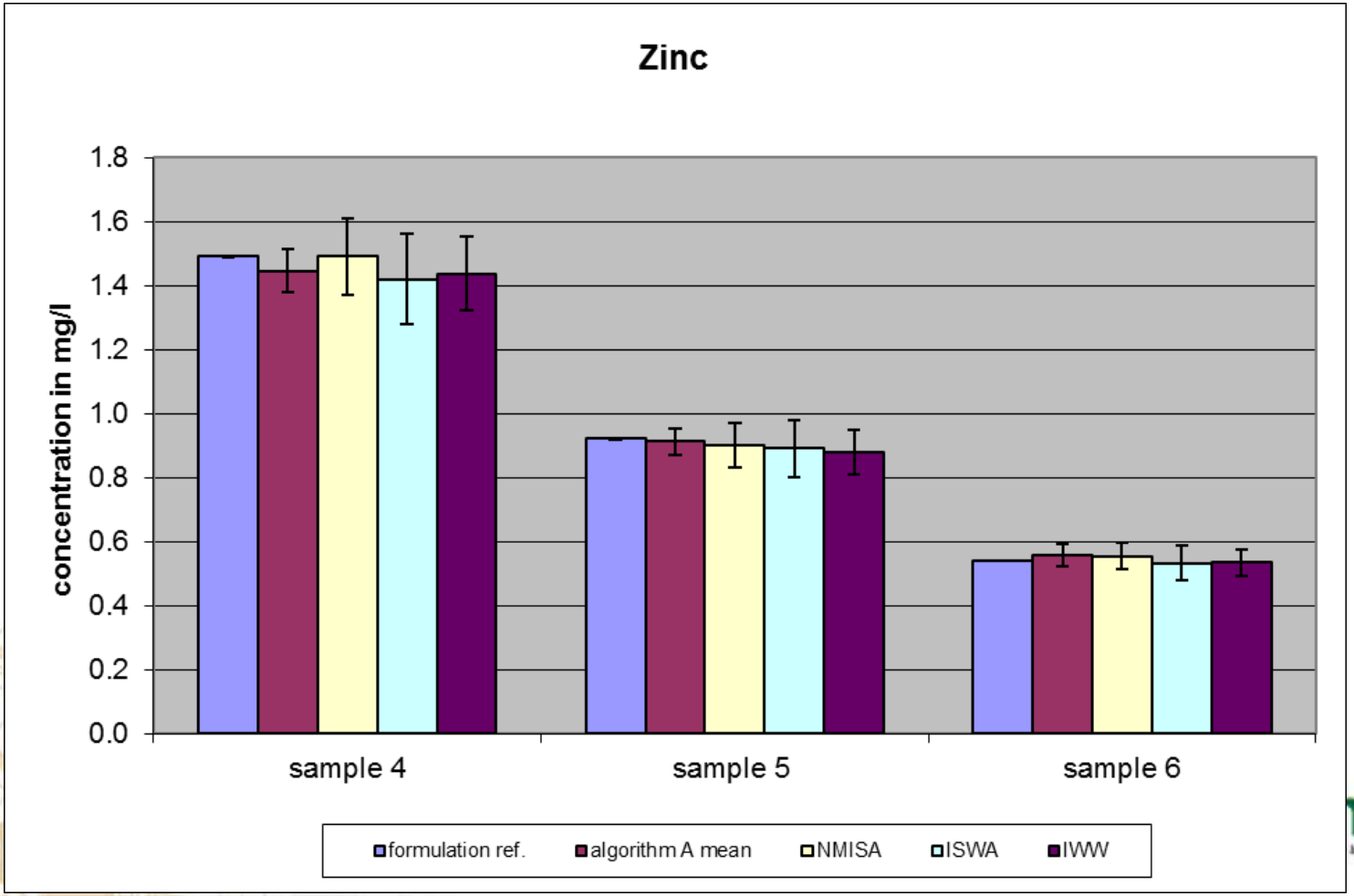
- Consensus value:
  - Participants
    - Non-parametric approach: **Ca in geological material**



Evaluation of performance

Determination of assigned value

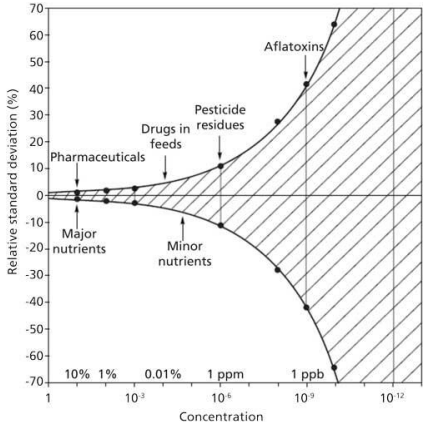
- SADCWater PT:



Evaluation of performance

Determination of acceptable range / standard deviation

- Prescribed value / Perception
  - Legislation
  - Fit for purpose (determined by coordinator / members)
- Model, e.g. Horwitz curve



- Precision experiment (standard method)
- Standard deviation from participants' results
  - Varies between rounds
  - 5% non-conforming results

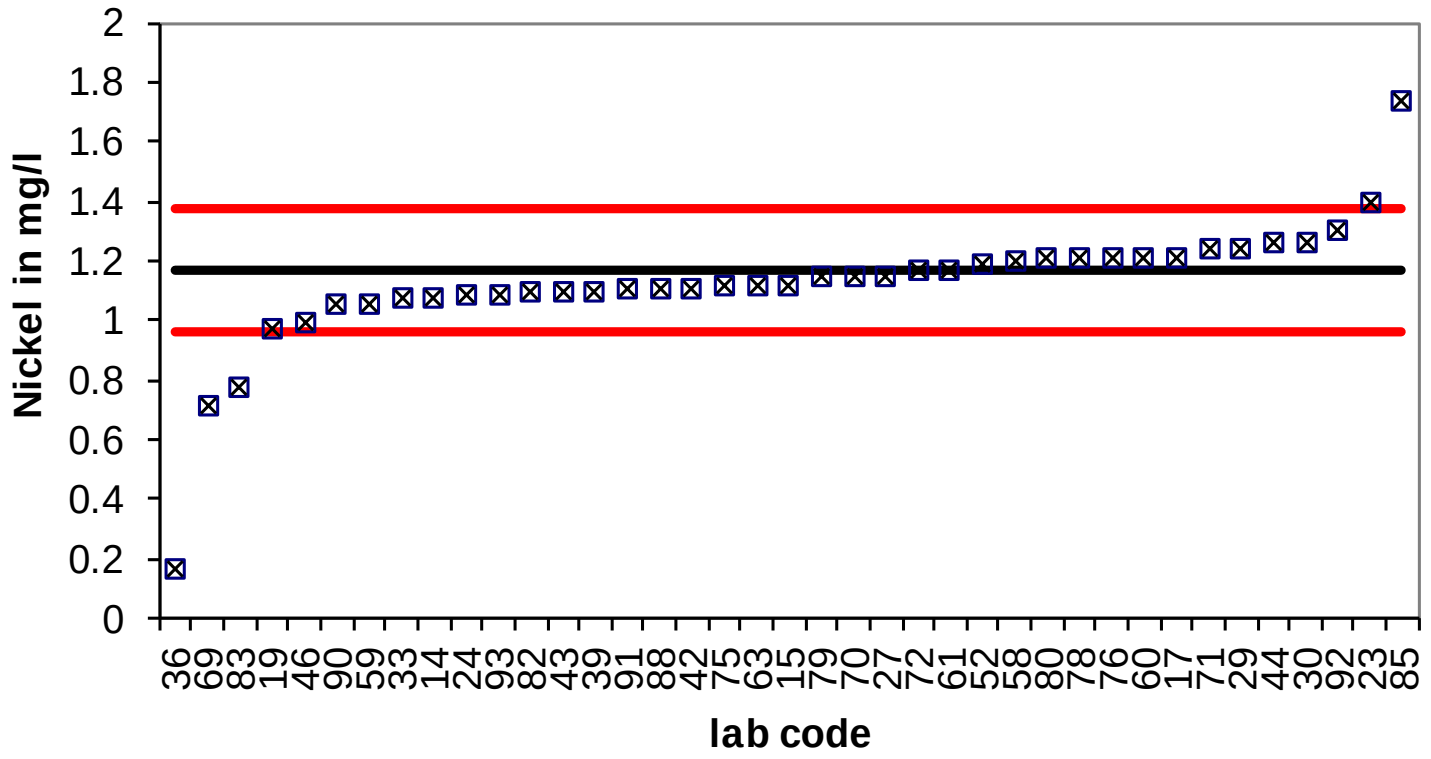
Limits for standard deviation 2015

	std limit
Sulphate	10 %
Chloride	10 %
Fluoride	10 %
Nitrate	10 %
Phosphate	10 %
TDS	10 %
Conductivity	10 %
Calcium	10 %
Magnesium	10 %
Sodium	10 %
Potassium	10 %

parameter	std limit
Iron	20 %
Manganese	20 %
Aluminium	20 %
Lead	20 %
Copper	20 %
Zinc	20 %
Chromium	20 %
Nickel	20 %
Cadmium	20 %
Arsenic	20 %
Cobalt	20 %

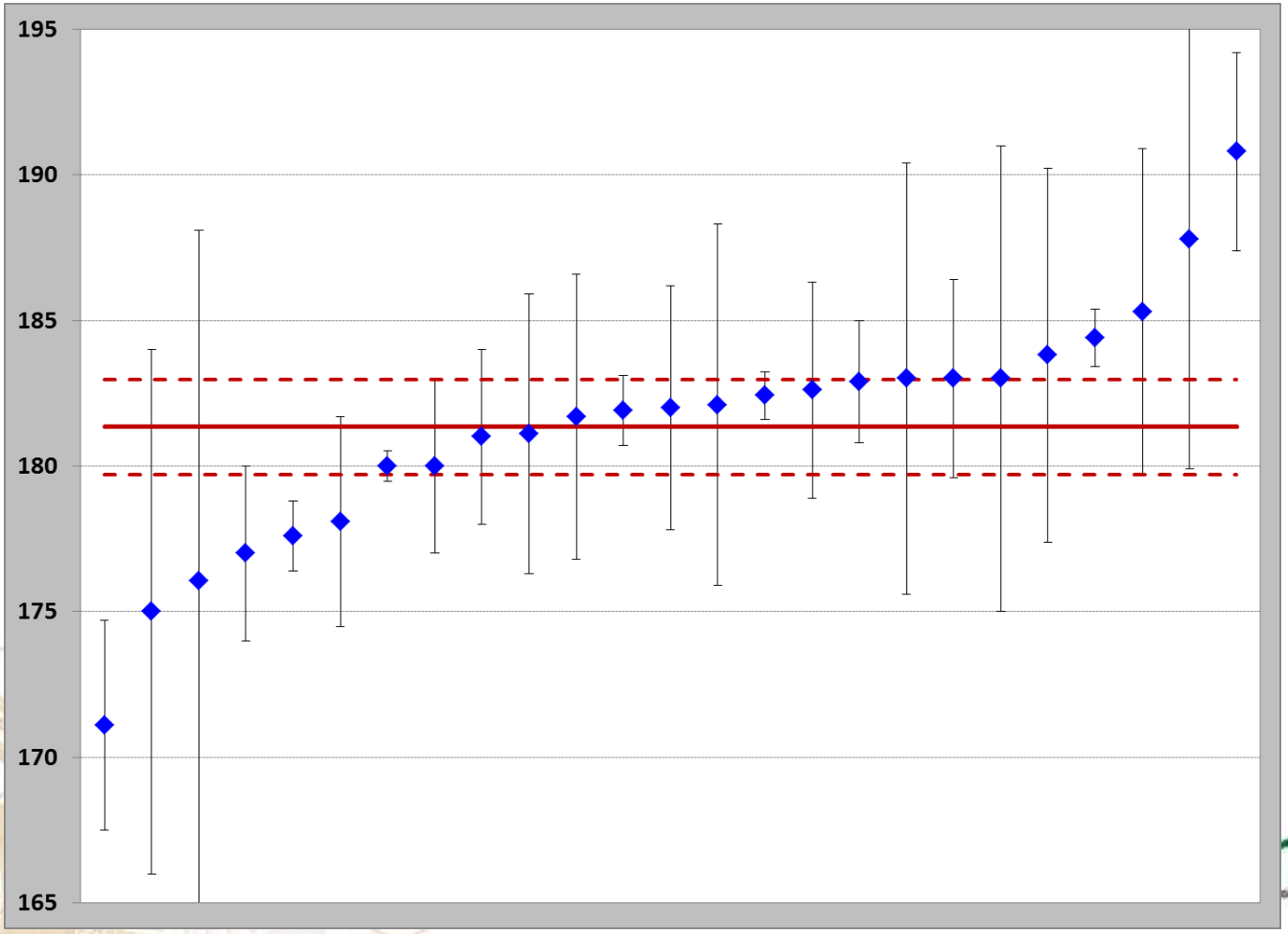
Presentation of PT/ILC results

- Normalised plot (concentration results)



Presentation of PT/ILC results

- Normalised plot (concentration results with uncertainties) _ Zn in bovine liver



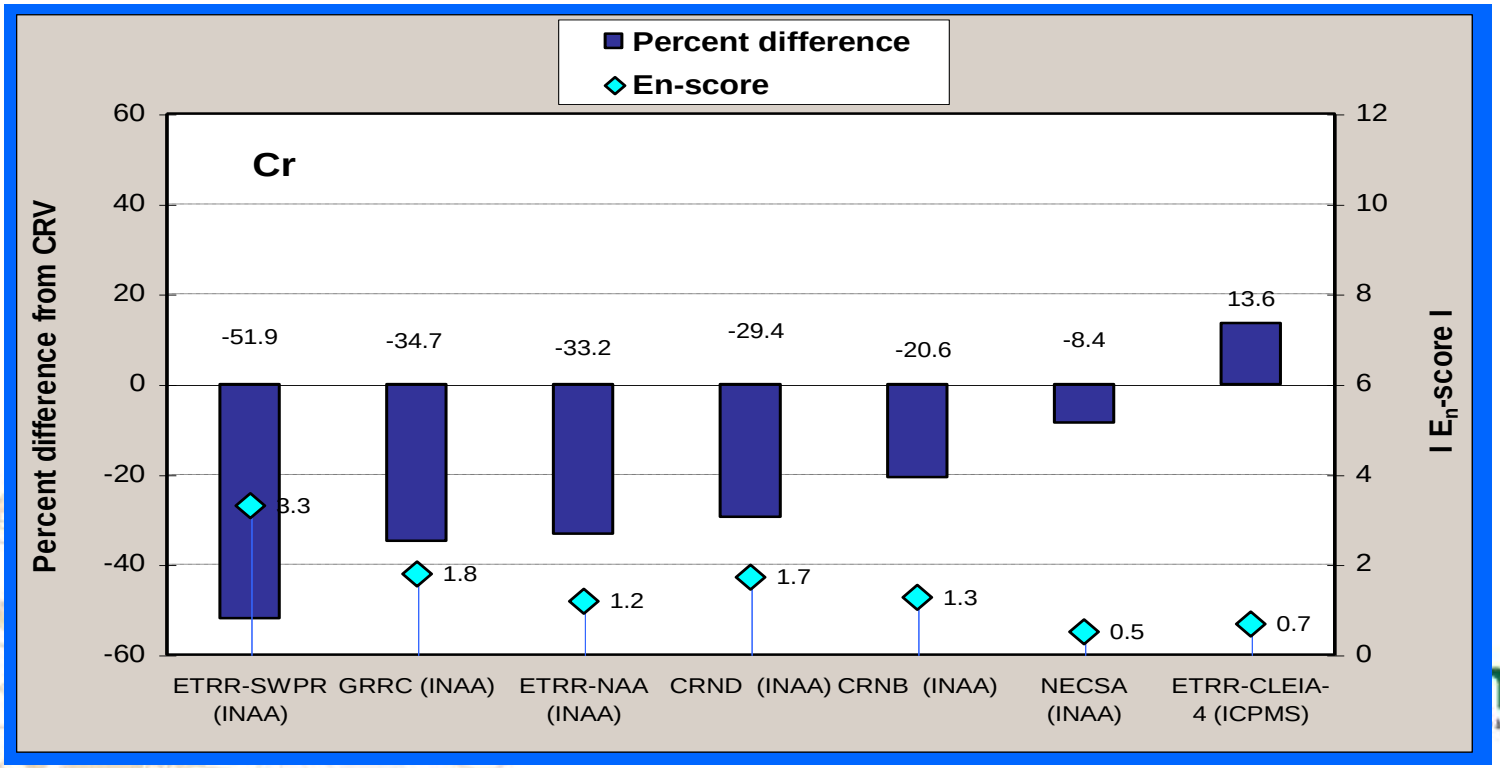
Performance evaluation / score

- Assign a normalised value that gives a score to each result, relative to the other results in the data set.
- Describes closeness of laboratory's result to consensus value
  - Bias / Percentage difference
  - z-Score
  - E_n-Score
  - ζ-score

Performance evaluation / score

- Bias / Percentage difference
  - An estimate of laboratory bias adjusted for concentration

$$\%D = \frac{x - X}{X} \times 100$$



Performance evaluation / score

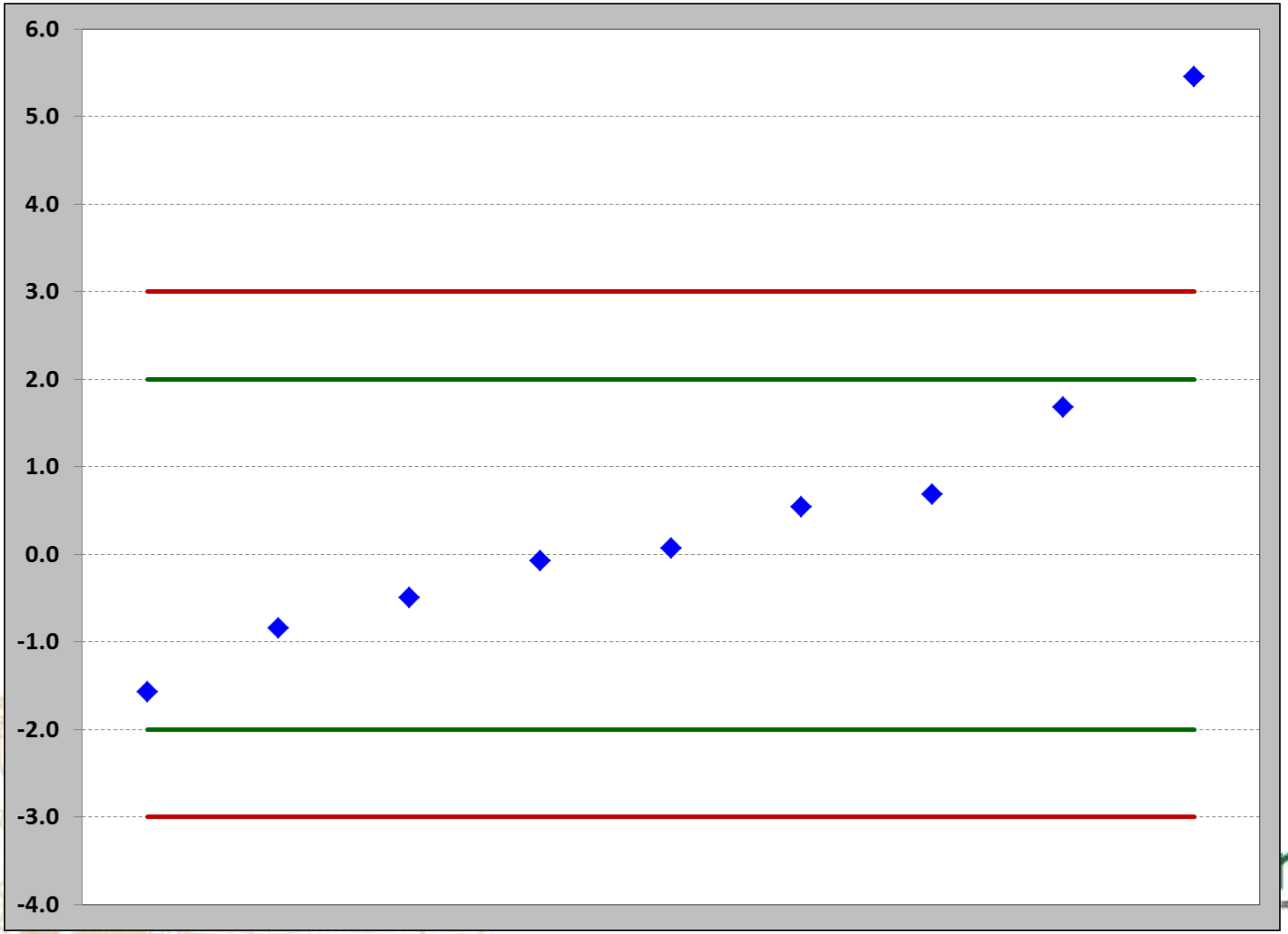
- z-Score
  - A measure of the deviation of the result from the assigned value for a particular measurand

$$z = \frac{x - X}{\hat{\sigma}} \quad \text{OR} \quad z = \frac{x - X}{s}$$

$\lvert z \rvert \leq 2$	Satisfactory	
$2 < \lvert z \rvert < 3$	Questionable	*Investigate possible causes to identify emerging or recurrent problems*
$\lvert z \rvert \geq 3$	Unsatisfactory	*Action signal indicating a need for corrective action*

Performance evaluation / score

- z-Score: **CO_2 in transformer oil**



Performance evaluation / score

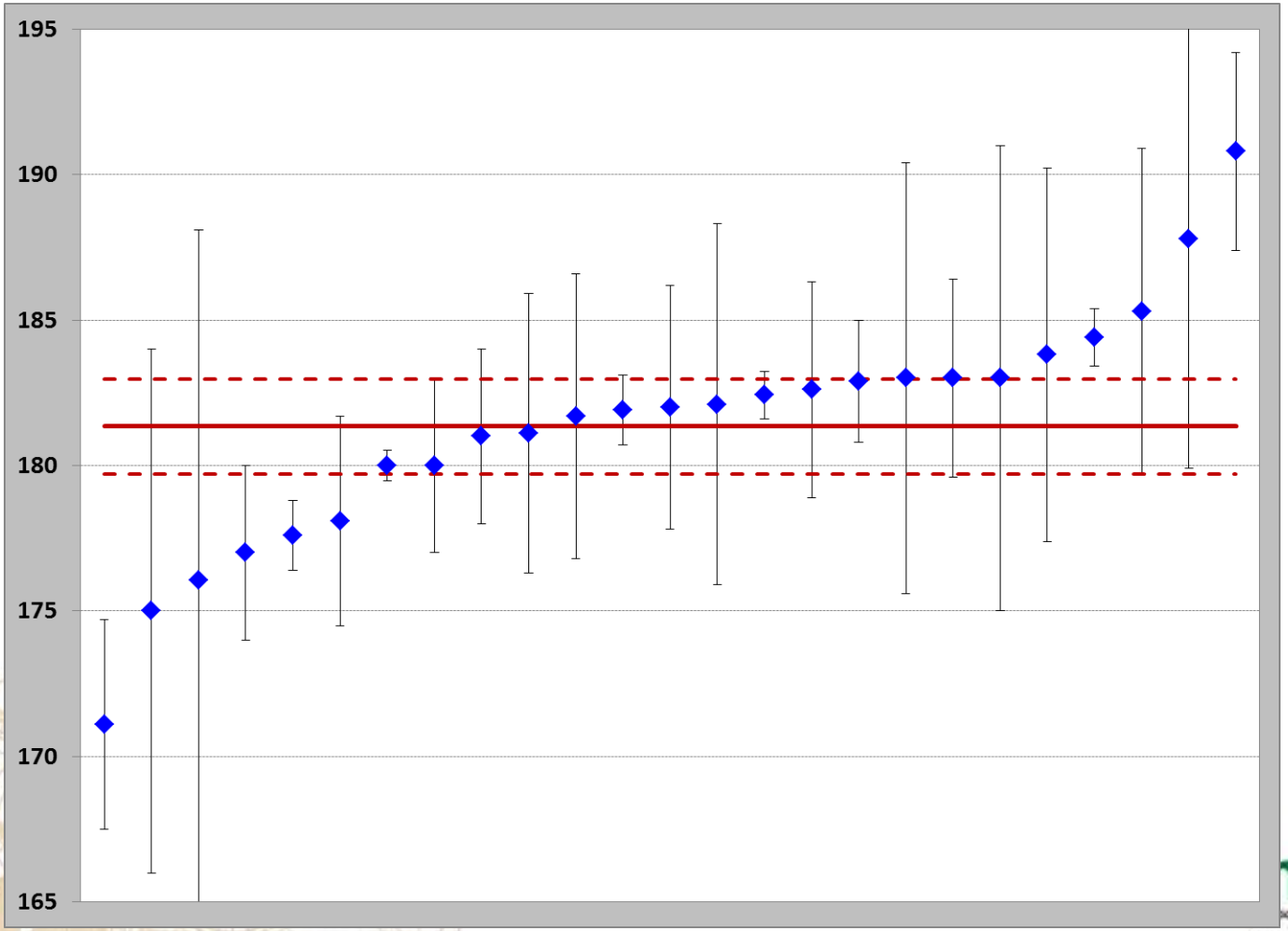
- E_n-Score
  - A measure of agreement between the assigned value and the participant's result within their respective uncertainties range

$$E_n = \frac{x - X}{\sqrt{U_x^2 + U_{ref}^2}}$$

$\lvert E_n \rvert \leq 1$	Satisfactory	
$\lvert E_n \rvert > 1$	Unsatisfactory	Action signal indicating a *need for corrective action*

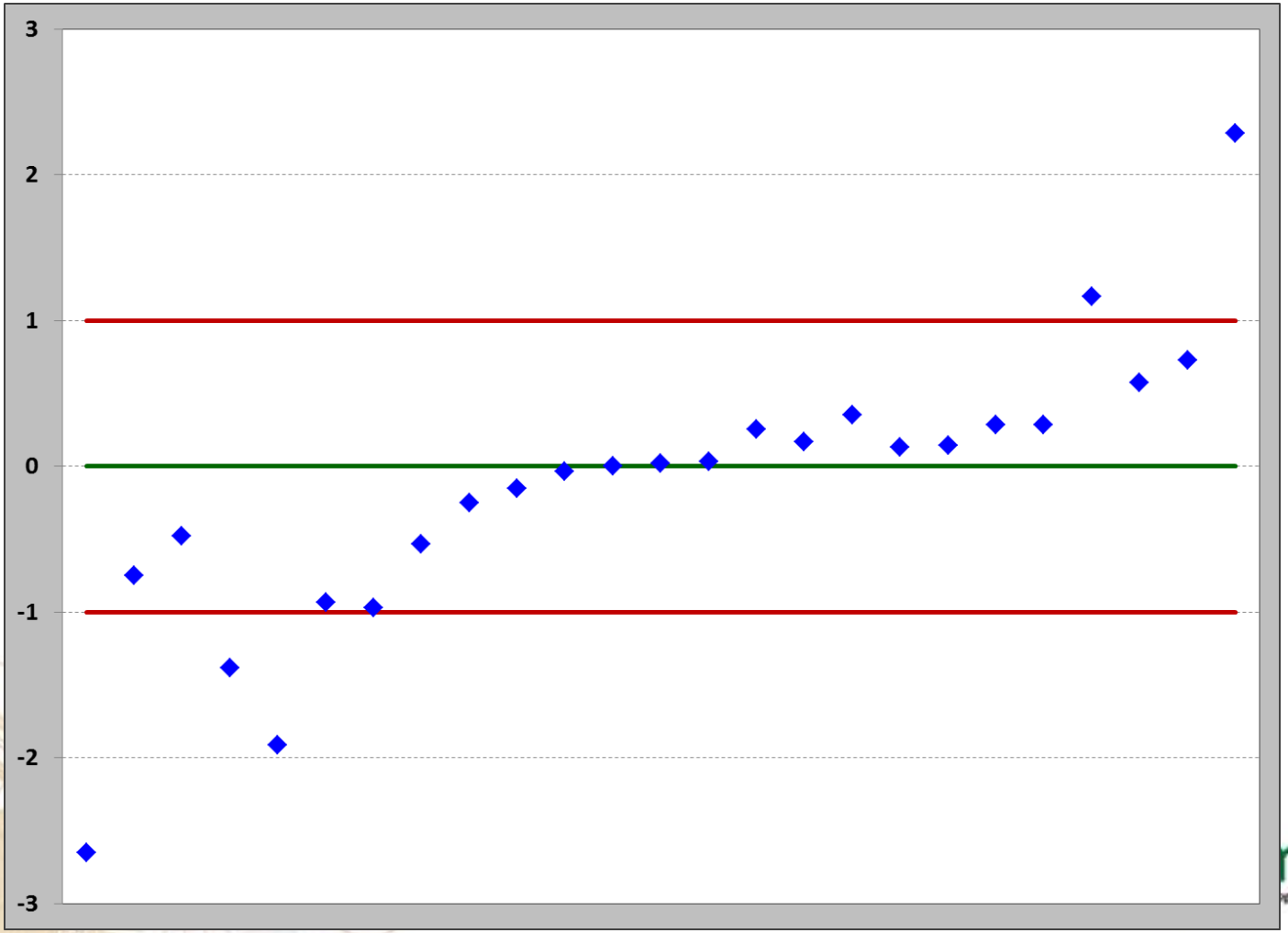
Performance evaluation / score

- E_n-Score
  - Zn in Bovine liver



Performance evaluation / score

- E_n-Score
  - Zn in Bovine liver



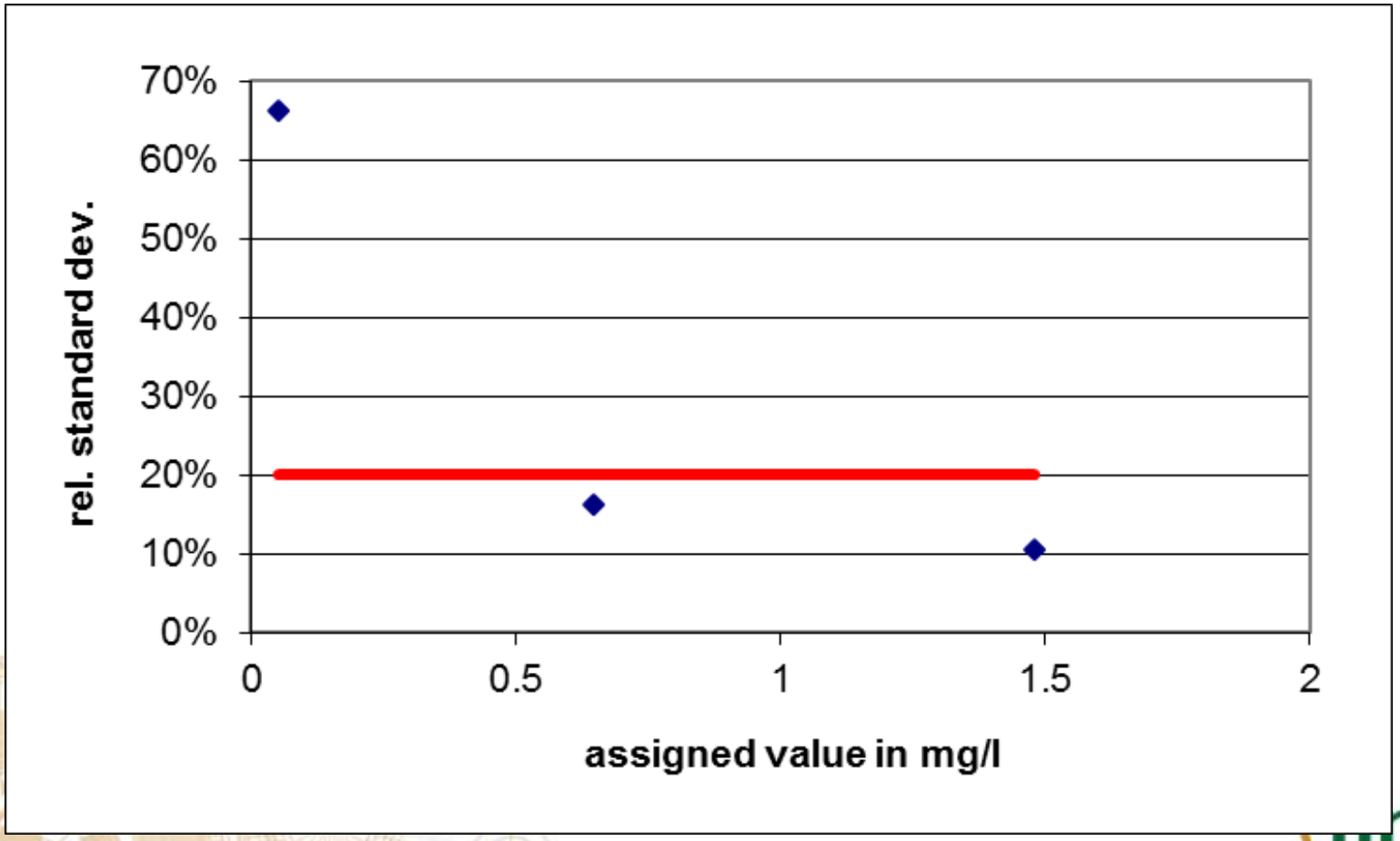
Performance evaluation / score

- ζ-score
  - Similar to En-score, using standard uncertainty instead of expanded uncertainty

$$\zeta = \frac{x - X}{\sqrt{u_x^2 + u_{ref}^2}}$$

$\lvert\zeta\rvert \leq 2$	Satisfactory	
$\lvert\zeta\rvert > 2$	Unsatisfactory	Action signal indicating a *need for corrective action*

Performance evaluation
Standard deviation vs selected limit (Pb)



How to effectively use PT results

Participating in a PT scheme is of limited value unless the laboratory takes advantage of its performance evaluation and the general information given in the PT scheme report.

- Read the PT report & review your performance:
  - How close to zero is the lab's z (or E_n) score?
  - Is the lab's result higher or lower than the consensus?
- Set own internal acceptance criteria
- Trend your performance
  - Spreadsheet
  - Graphically

nmisa
National Metrology Institute of South Africa

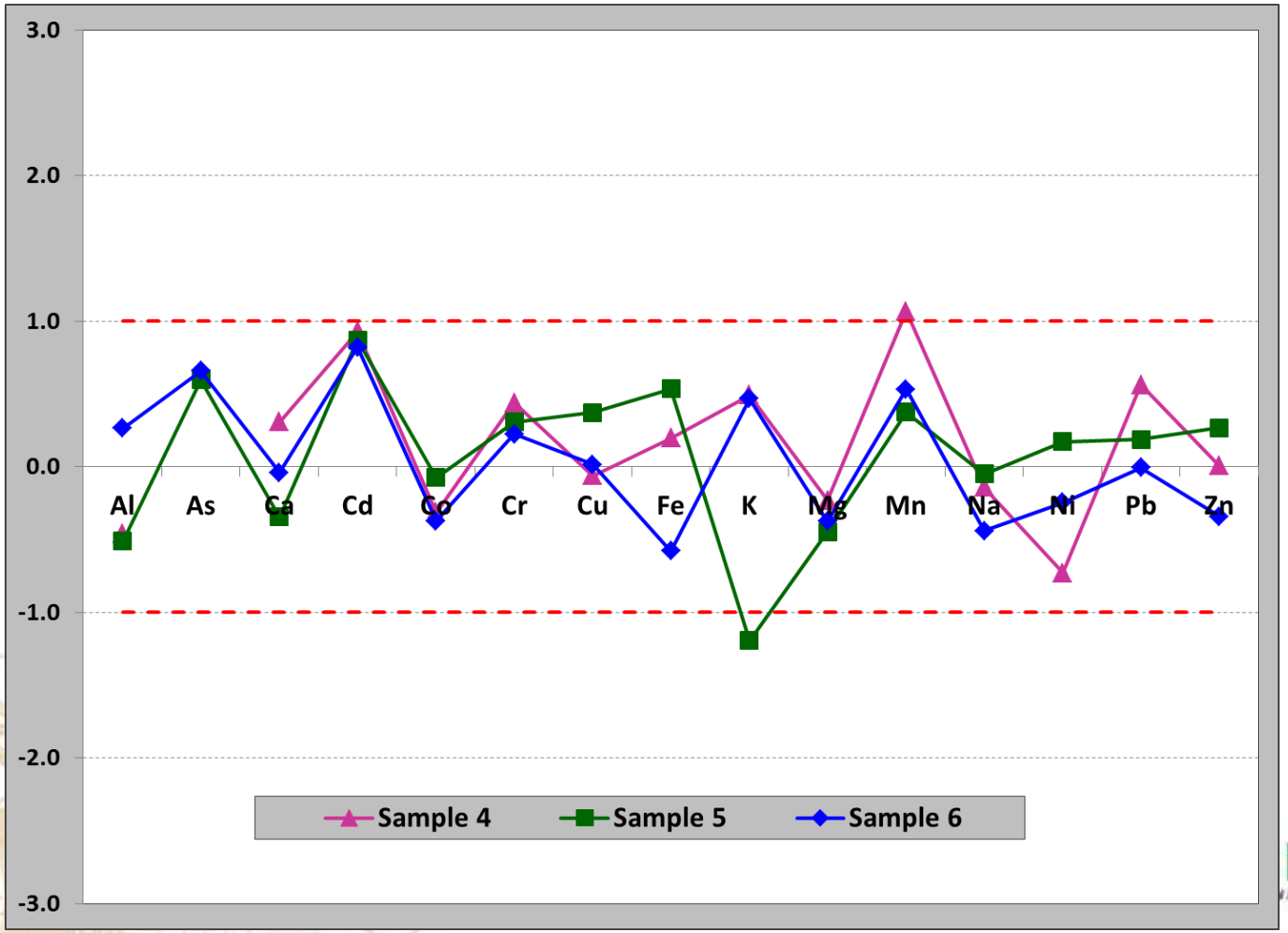
Interpretation of PT results

- Single PT round
  - Histogram of performance scores
  - Bar-plots of standardised laboratory bias
  - Standard deviation / repeatability / reproducibility
  - Normal probability plot

- Performance over time
  - Schewart control chart
  - Cusum control chart
  - Plot of standardized laboratory bias against laboratory average
  - Dot plot

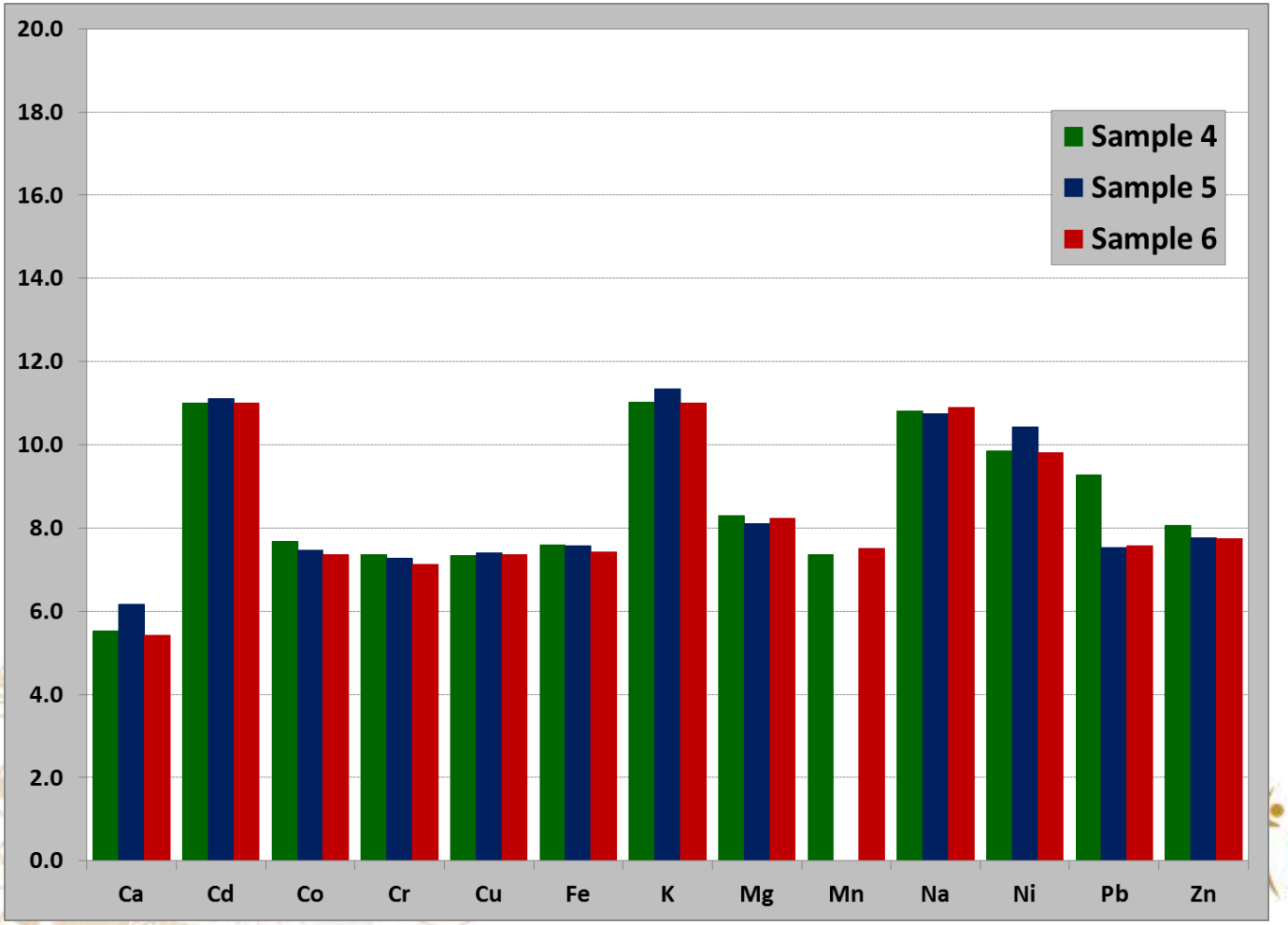
Interpretation of PT results
- Short term performance

- Schewart control chart (E_n-score)



Interpretation of PT results
- Short term performance

- Bar chart (Expanded uncertainty)



nmisa
National Metrology Institute of South Africa

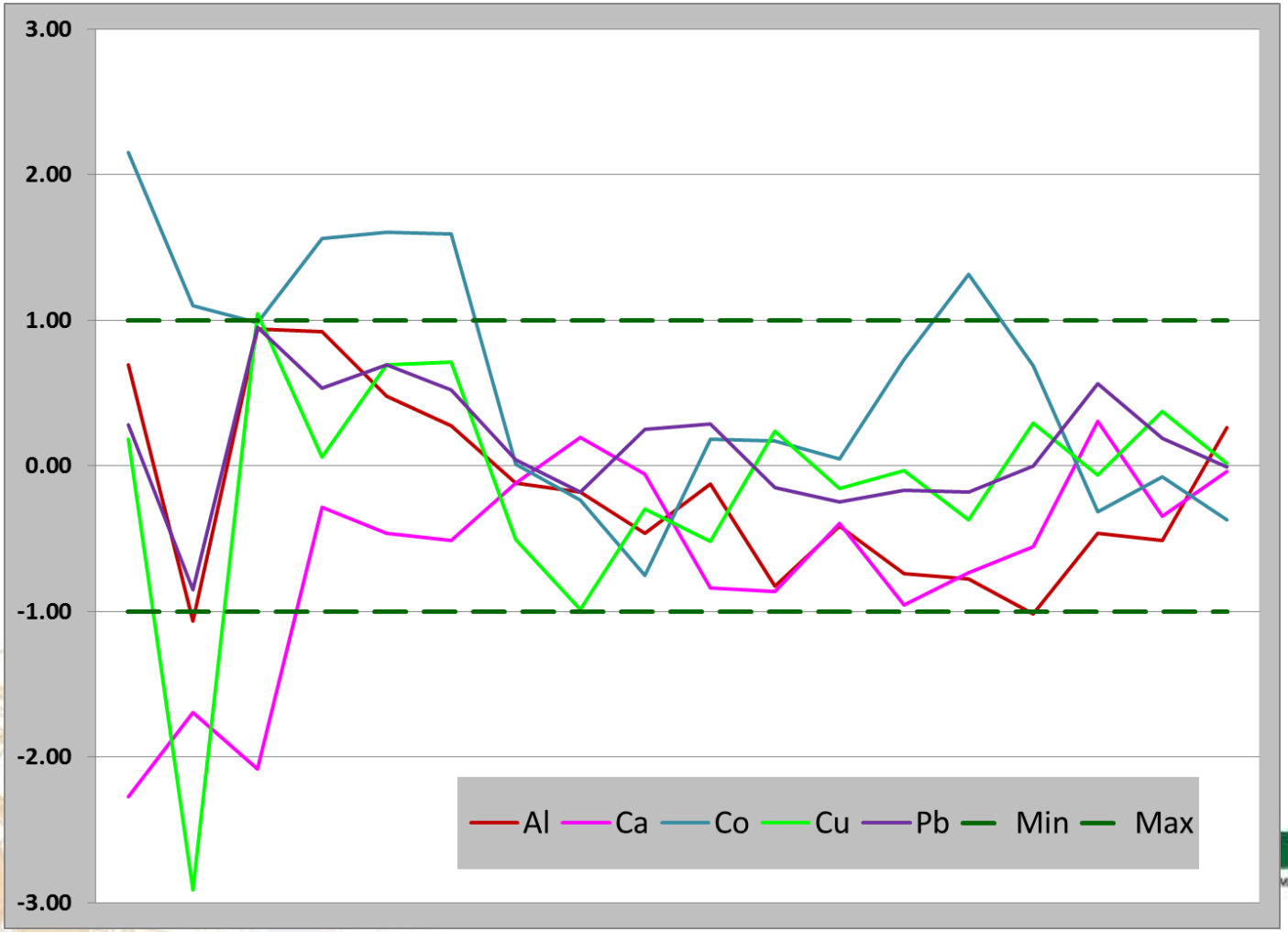
Interpretation of PT results

- Single PT round
  - Histogram of performance scores
  - Bar-plots of standardised laboratory bias
  - Standard deviation / repeatability / reproducibility
  - Normal probability plot

- Performance over time
  - Schewart control chart
  - Cusum control chart
  - Plot of standardized laboratory bias against laboratory average
  - Dot plot

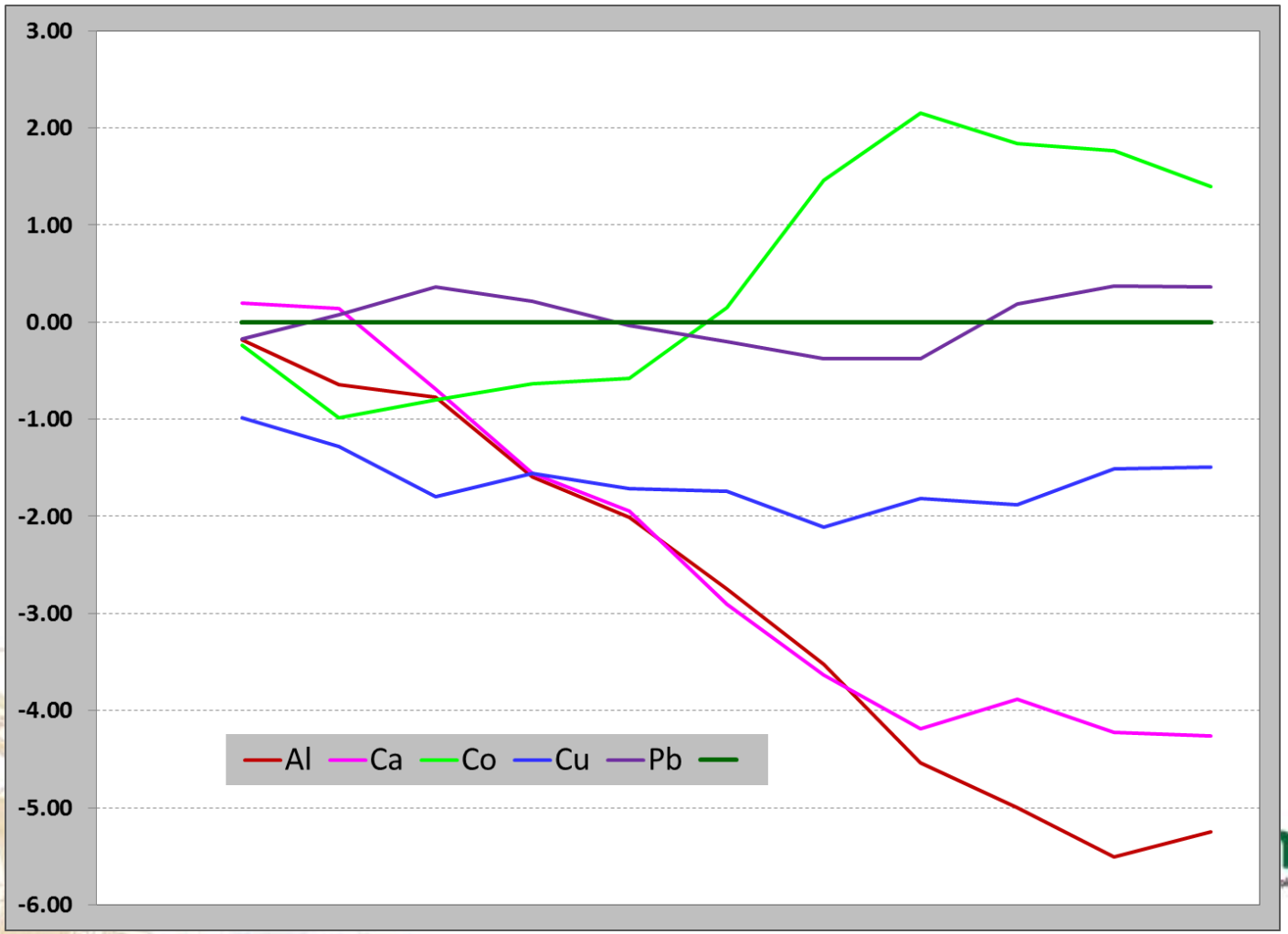
Interpretation of PT results
- Long term performance

- Schewart control chart (E_n-score)

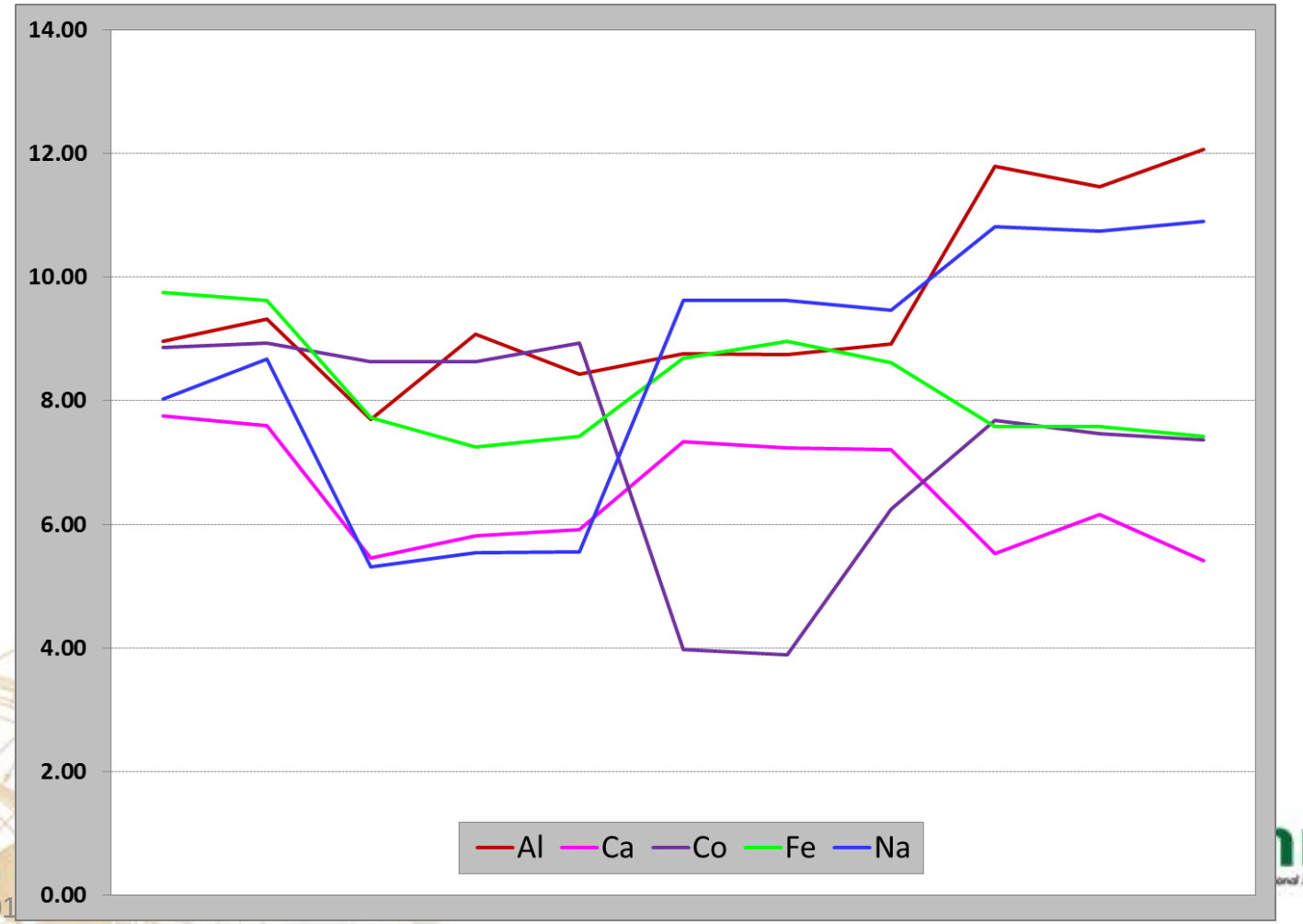


Interpretation of PT results
- Long term performance

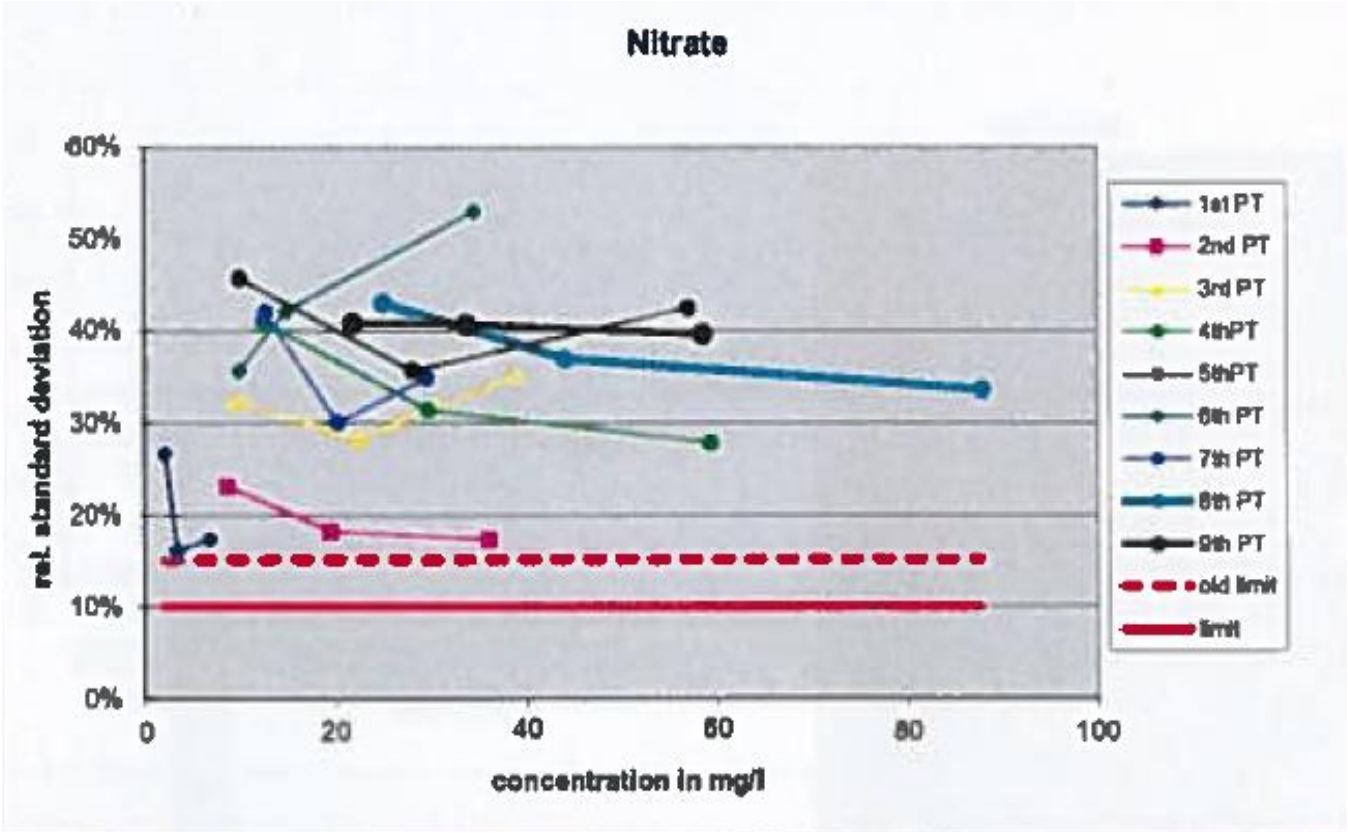
- Cusum control chart (E_n-score)



PT standard deviation:
Performance over time – Uncertainty / SD reported



PT standard deviation:
Performance over time vs limit



Interpretation of PT results
- Unsatisfactory results

- Root cause investigation
  - Raw data
  - Overall performance
  - Successive PT studies
  - Internal quality control data
- Typical causes
  - Clerical error
  - Technical problem
  - PT Scheme

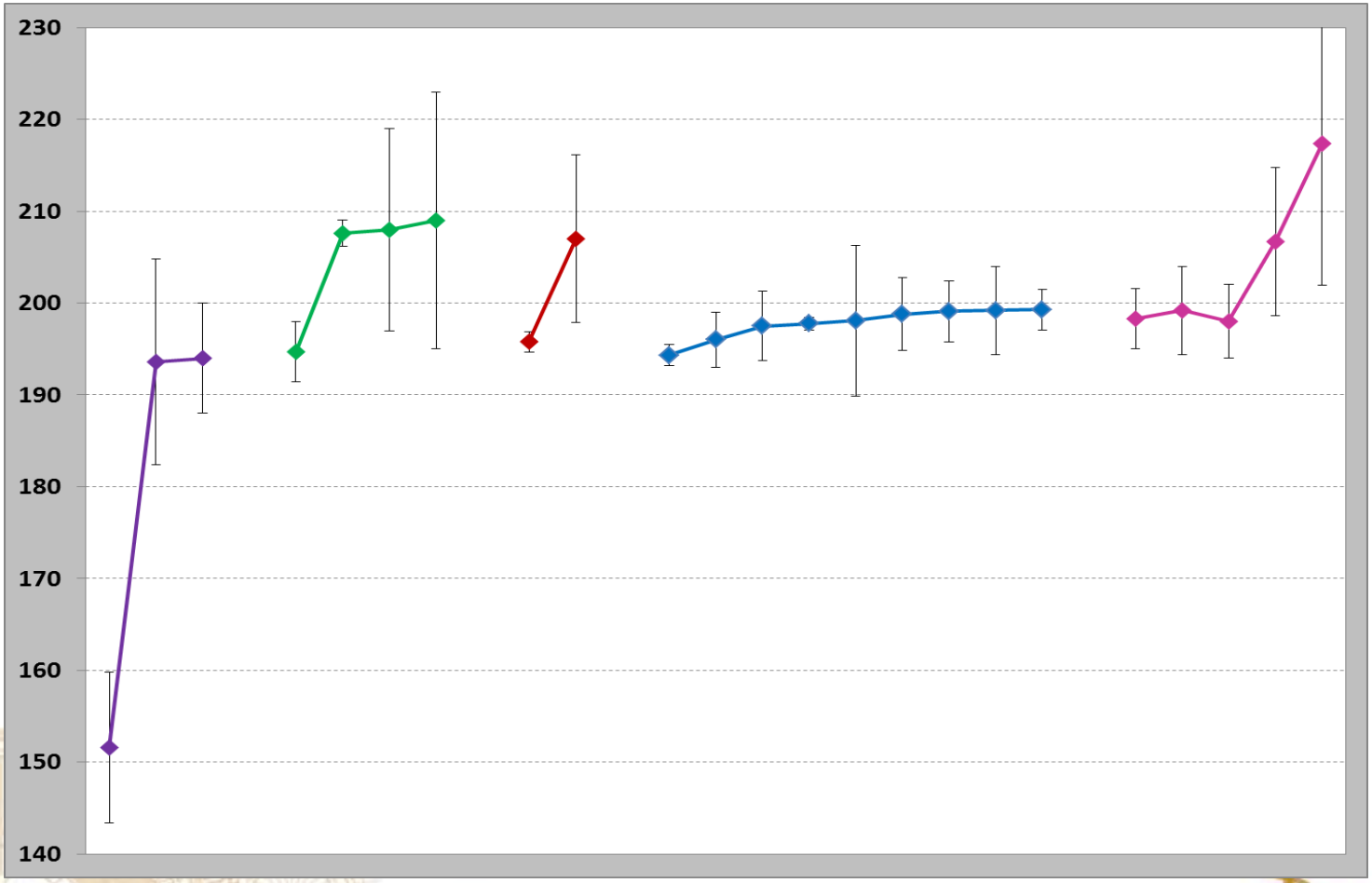
Interpretation of PT results
- Unsatisfactory results

- Typical causes
  - Clerical error
    - Transcription error
    - Labelling
    - Units
    - Decimal error

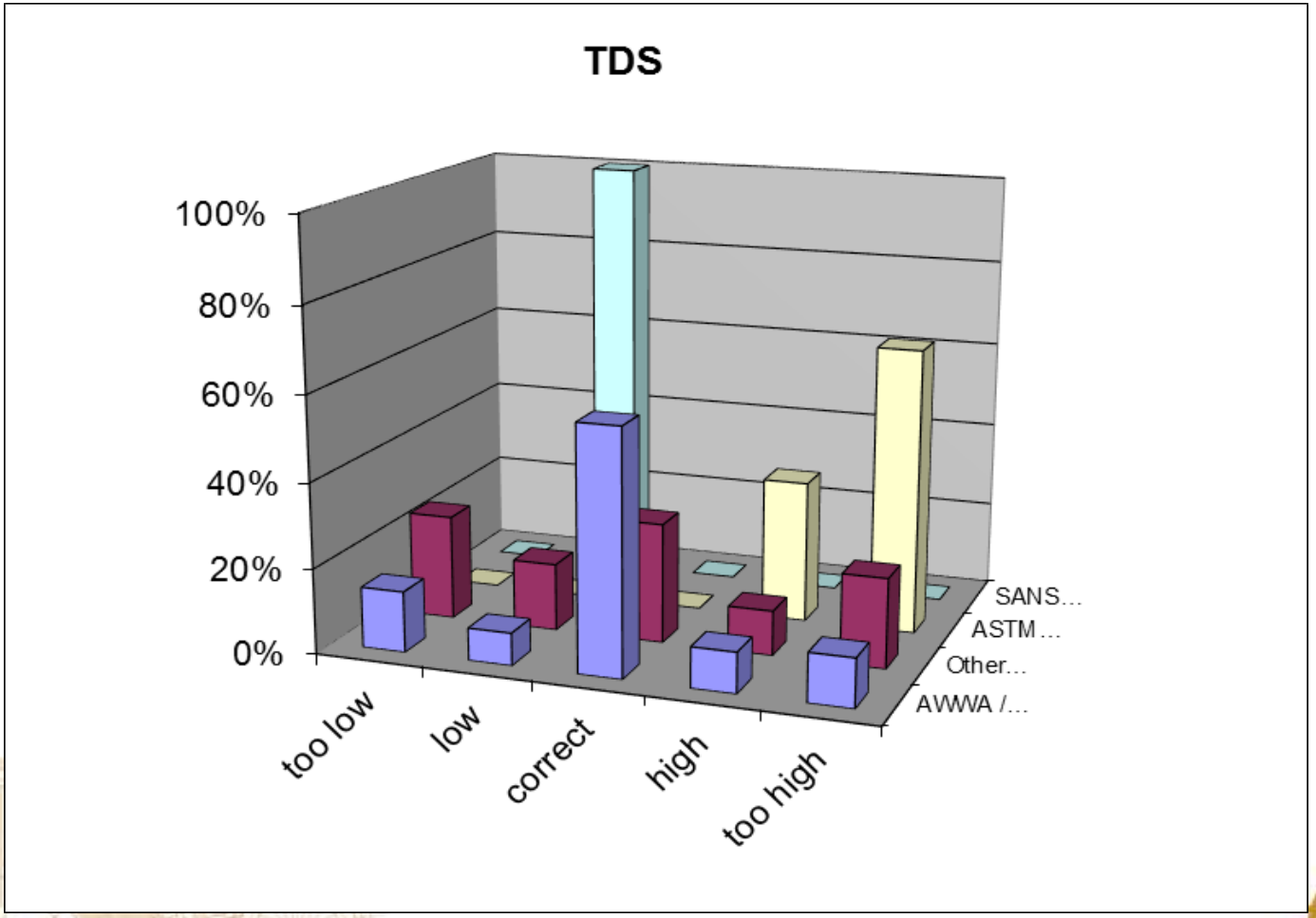
Interpretation of PT results
- Unsatisfactory results

- Typical causes
  - Technical problem
    - Storage / pre-treatment of sample
    - Method validity
    - Calibration:
      - Equipment
      - Standards
    - Reagent purity
    - Equipment
    - Environmental conditions
    - Staff competence

Interpretation of PT/ILC results
- Comparison of methods



Interpretation of PT/ILC results
- Comparison of methods



Interpretation of PT results
- Unsatisfactory results

- Typical causes
  - PT Scheme
    - Sample stability / homogeneity
    - Matrix differences
    - Concentrations outside scope of application
    - Inappropriate / inexperienced peer group
    - Incorrect assigned value / standard deviation

Investigation of PT results
- Unsatisfactory results

PT result ok?

- **No** → Long term performance ok?
  - Long term performance ok? → **Yes** → Short term problem:
    - Transcription, units, decimal error
    - Calibration standard, reagent purity (QC chart)
    - Environmental conditions
    - Staff changes?
  - Long term performance ok? → **No** → Long term problem:
    - Equipment maintenance
    - Method
    - Method validation
    - Staff competence / training
- **Yes** → Long term performance ok? → **Yes**

nmisa
National Metrology Institute of South Africa

Interpretation of the PT results

- PT is not about *“passing” or “failing”* a test - It is about taking part and learning from the results
- Consistent good performance is the goal.
  - One-time good performance does not necessarily make a laboratory good
  - One bad result in any round of PT does not make a laboratory bad; but it must be studied and lessons learned

Conclusion

- Proficiency Testing is a very powerful tool
  - Allows you to identify problems in testing and improve the performance of the laboratory
- A lot of information available to the laboratory. Onus is on the laboratory to use this.

Thank You

nmisa
National Metrology Institute of South Africa