SADCWater PT Chemistry workshop 2018 - Part 1: **Traceability & Method Validation**

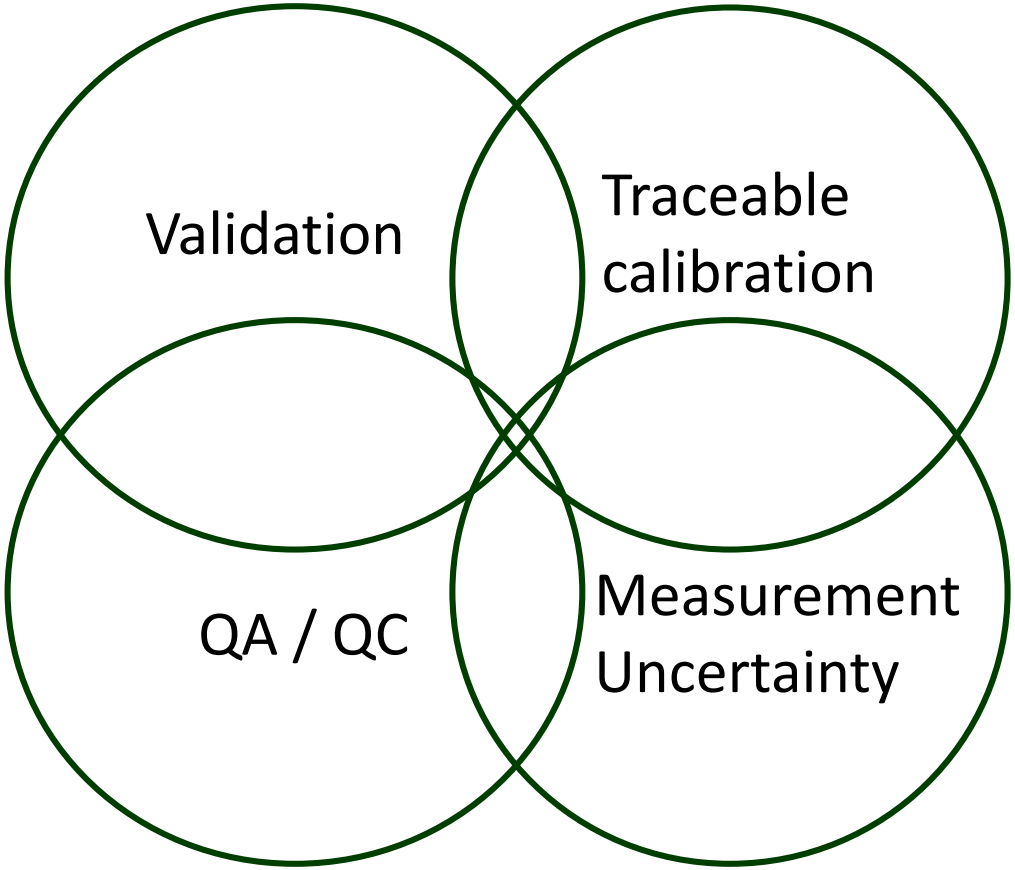
Maré Linsky

26-27 November 2018

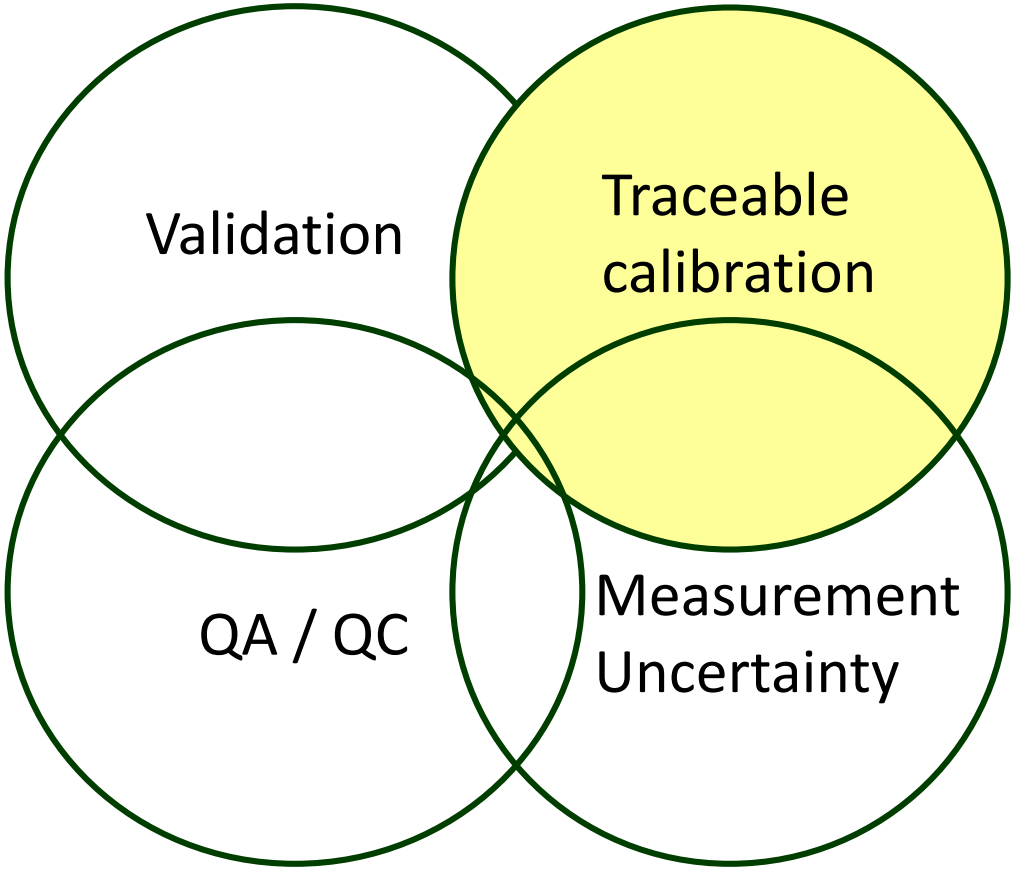
Importance of Analytical Measurement

- Everyday millions of tests and measurements performed in thousands of laboratories around the world
  - Trade - Value of product
  - Quality of drinking water, food and feed
  - Healthcare
  - Forensics
  - Environmental analysis
- High costs associated with these measurements
- Impact of these measurements could be far-reaching
  - Health
  - Cost (fines)
  - Legal
- Ensuring the reliability of these measurements is the responsibility of the Analytical Chemist

Ensuring valid Analytical Measurements



Ensuring valid Analytical Measurements

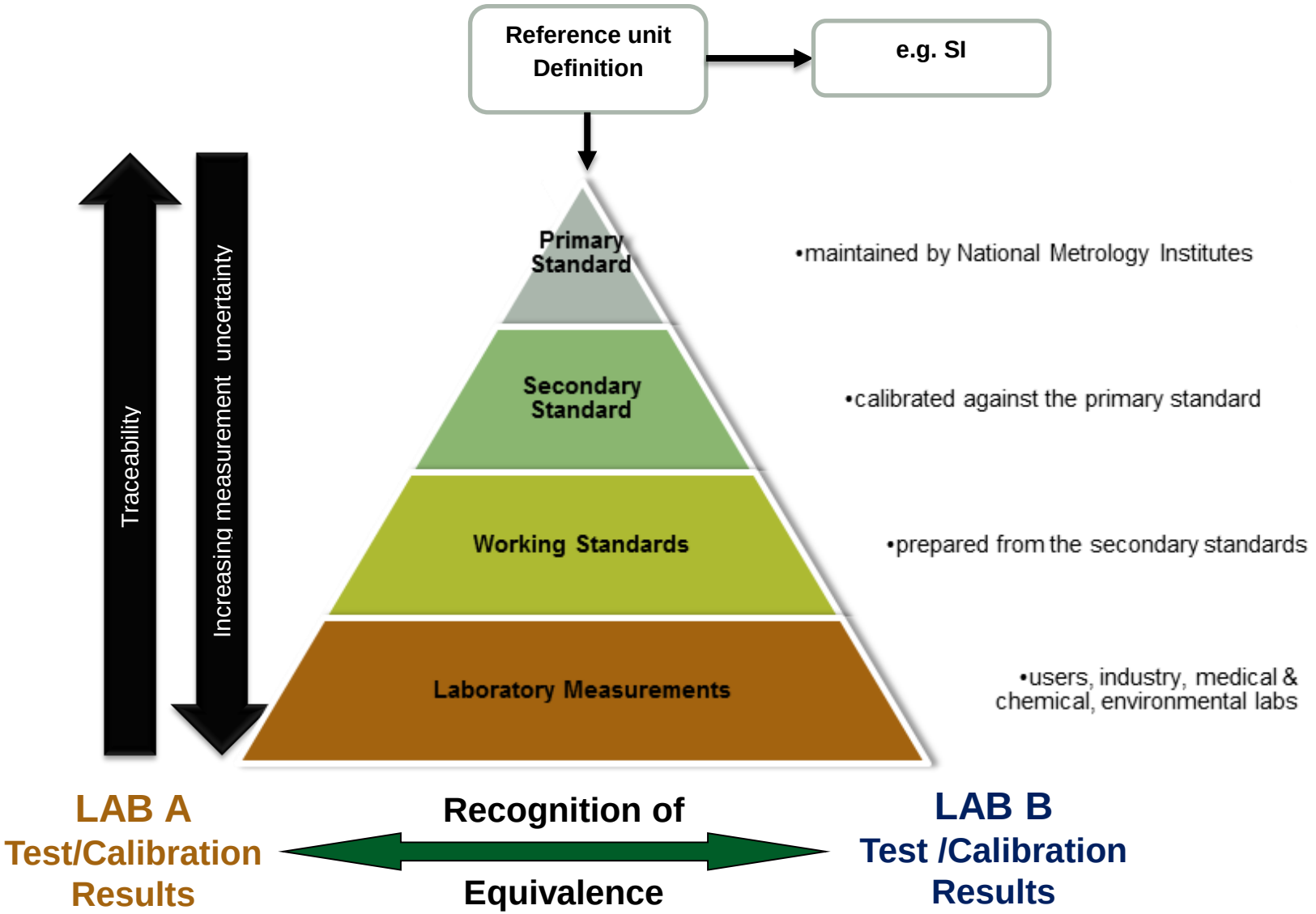


Metrological Traceability

- Good analytical results are essential to ensure reliable decisions
- Result can be related to a reference through a documented unbroken chain of calibrations, each contributing to the measurement uncertainty
- Comparability through traceability to consistent and agreed set of measurement units and scales, i.e. SI

OR

Establishing measurement traceability



Metrological Traceability

- For every step in the traceability chain:
  - Documented, unbroken chain of calibrations
    - Traceable to appropriate references
  - Performed according to appropriate method
  - Measurements by technically competent laboratories
  - Measurement uncertainty determined according to agreed methods
- Establish traceability for:
  - Each parameter in mathematical model
  - Each of the specified conditions
    - *NOTE: Essential for critical values in measurement*

Establishing Metrological Traceability in Analytical Chemistry

- Physical measurements, e.g. mass, volume, temperature
  - Uncertainties are typically not significant compared to those in analytical measurements
- Confirmation of Identity
  - Certified pure material
  - Authentic samples from reputable source
  - Reference data, e.g. reference wavelength spectra
- Amount of substance / Concentration
  - Pure Materials
  - Other reference materials, e.g. single- or multi-element standards, etc.
  - Certified refence materials

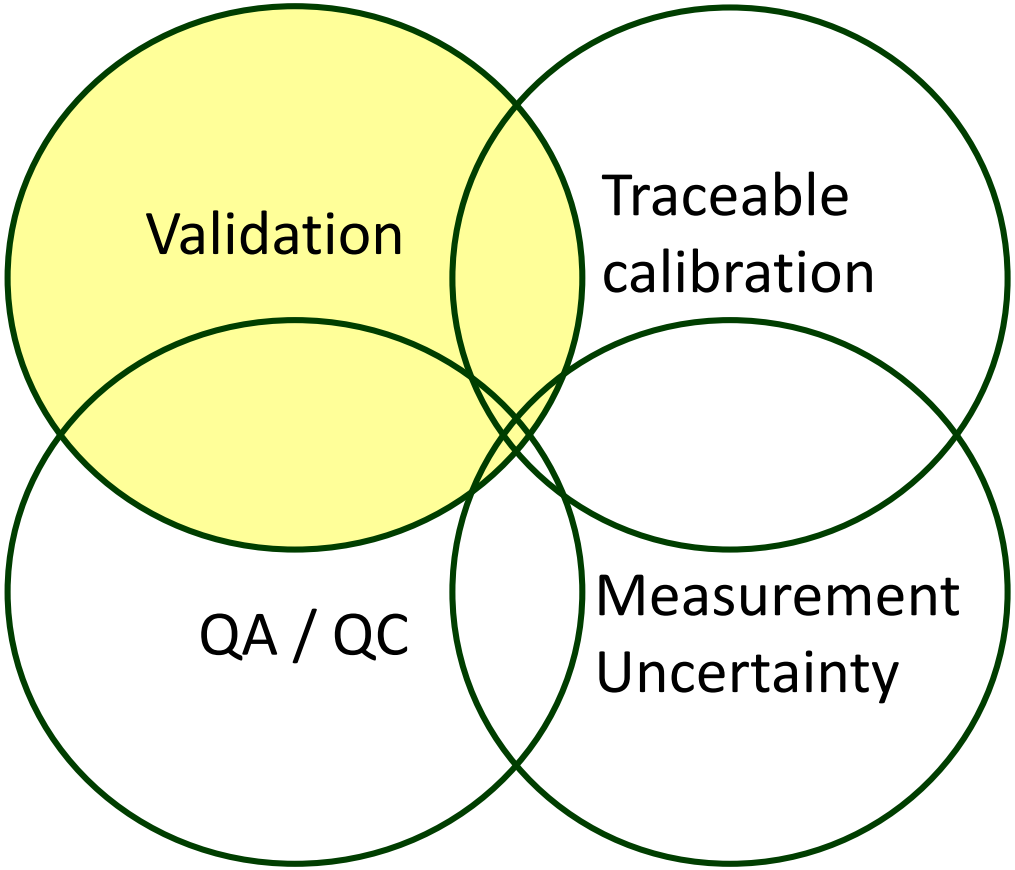
Metrological Traceability

- Confirming metrological traceability of calibration standards, CRMs, etc.:
  - Accreditation of manufacturer to ISO 17034 and ISO 17025
  - Useful information stated on Certificate:
    - Specification of measurand
    - Measurement unit
    - Characterisation methods
    - Specifications for sample handling
    - Measurement uncertainty
- Confirming metrological traceability of calibration certificates, e.g. for mass balance
  - Accreditation of calibration laboratory to ISO 17025
    - Evaluation of trueness - Correction factors
    - Measurement Uncertainty

Traceability in Analytical Chemistry

- To conclude:
  - Metrological Traceability – basis for establishing comparability of measurement results
    - Calibrated equipment, e.g. mass balance
    - Certified calibration standards
    - Validated methods
  - Uncertainty is part of the definition
    - Uncertainty of a traceable result = Uncertainty (reference) + Uncertainty (measurement)

Ensuring valid Analytical Measurements



Method validation

- Method validation is required to establish the fitness for purpose of a method for the specific requirements of customers when applied to a specific laboratory
- Method validation studies produce data on the overall performance or individual influence quantities associated with the results of a method in normal use in the laboratory

Method Validation Approaches

- Data on overall method performance parameters are obtained from:
  - Interlaboratory studies
  - Single laboratory: In-house validation protocols
    - Validation
    - Verification

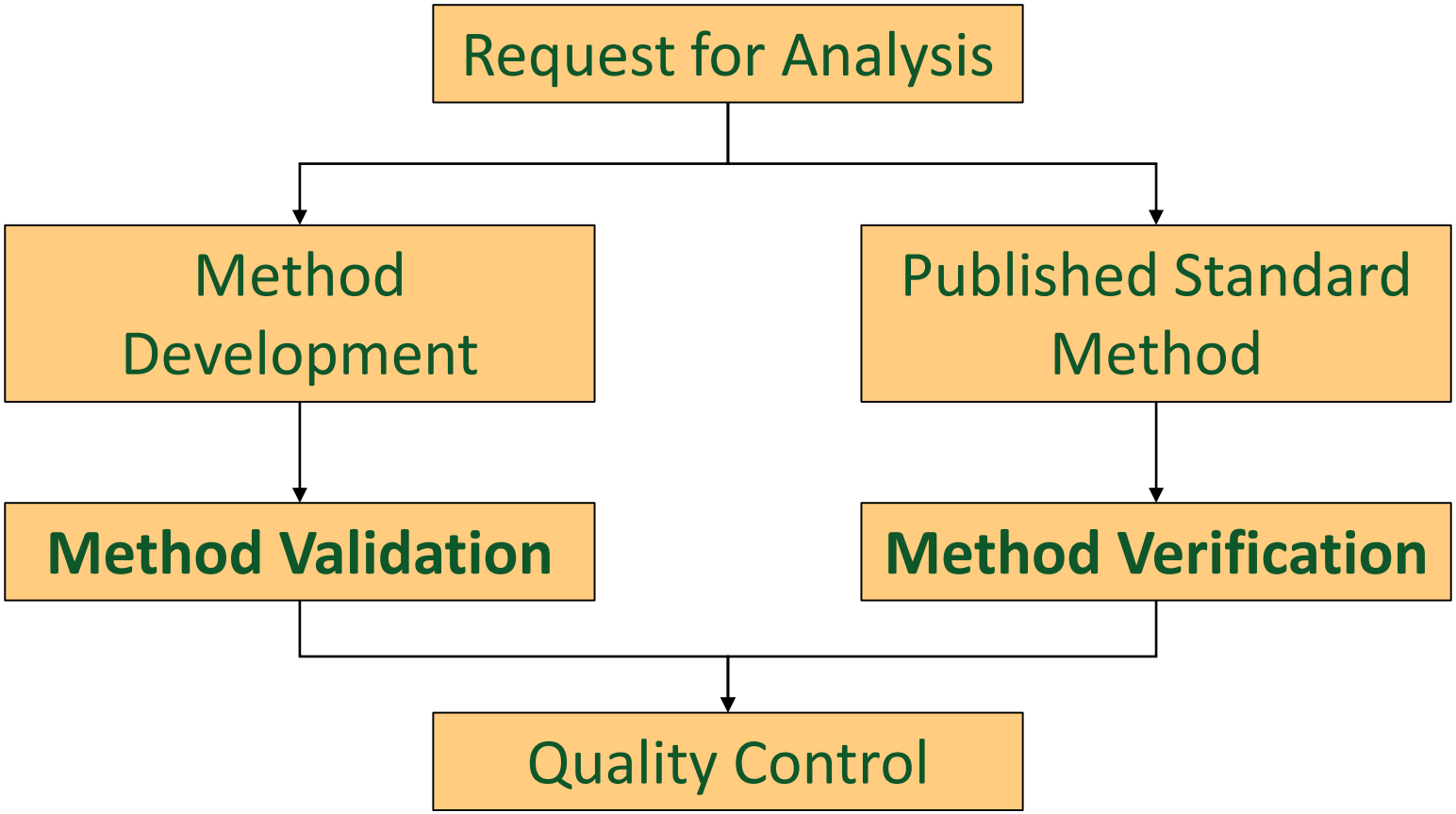
Method Validation Approaches

- Interlaboratory
  - Published standardised procedure, e.g. ISO, ASTM
  - Validated, employing interlaboratory comparisons, according to international protocols (e.g. ISO 5725 standards)
  - Laboratory's responsibility to confirm that analytical performance can be matched. Typically only for:
    - Precision
    - Bias
    - Note: Robustness, Selectivity covered in Standard

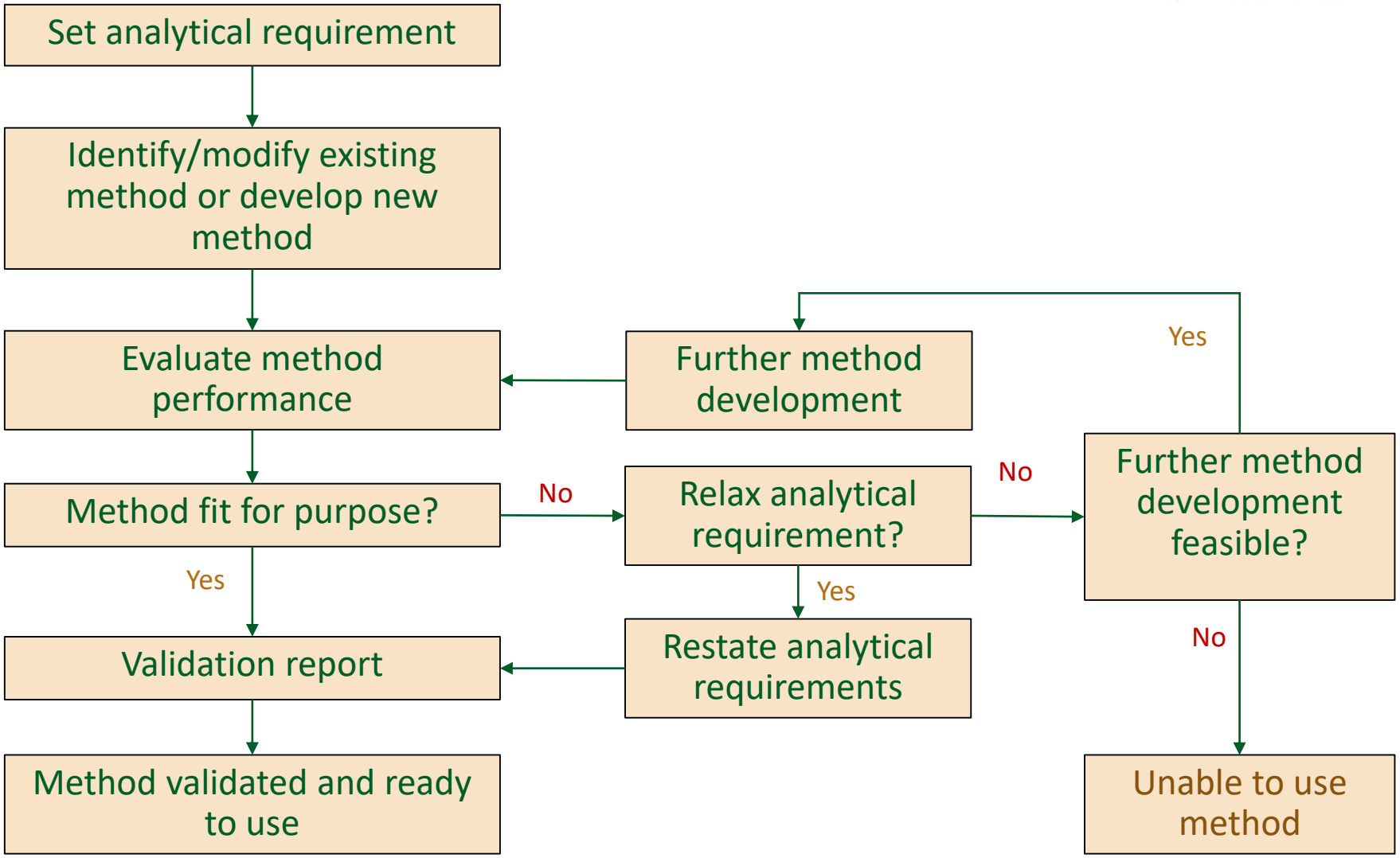
Method Validation Approaches

- Single laboratory
  - Method developed in-house
  - Standard method used outside it's scope
  - Evaluate all relevant performance criteria, e.g.
    - Selectivity; LOD, LOQ; Working Range; Trueness; Precision; Ruggedness; Uncertainty
  - Validation is a balance between costs, risks and technical possibilities (routine vs ad hoc)

Approaches:
Validation vs. Verification



Method Validation Process



Method Validation Approaches

- Emphasis is on identifying and removing or reducing significant effects, i.e. continue with method development if method performance is not satisfactory
- It is the responsibility of the laboratory to ensure that a method is fit for its intended use.

Overview: Method Validation

- Performance Criteria
  - Precision
  - Trueness
  - Limit of Detection (LOD) & Limit of Quantification (LOQ)
  - Working range (Linearity)
  - Selectivity / Specificity
  - Robustness (ruggedness)
  - Measurement Uncertainty
  - Metrological Traceability

General method validation requirement

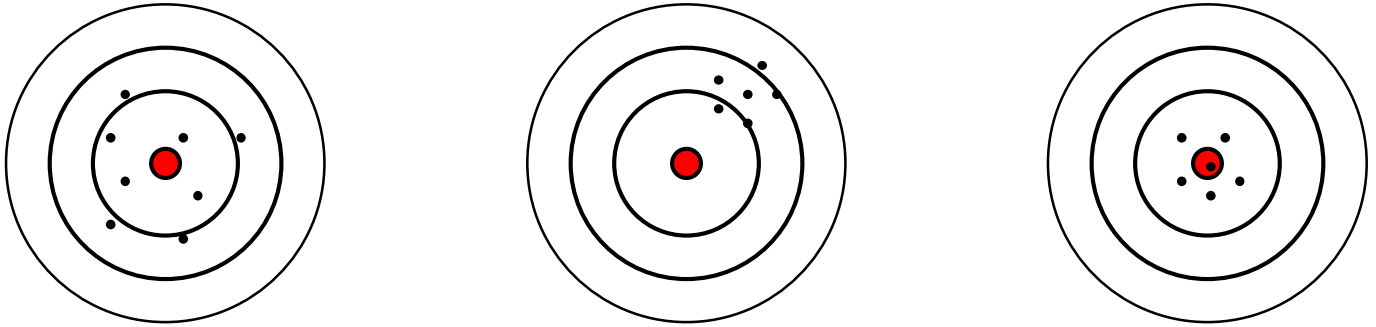
- Representativeness
- Representative variation

Realistic survey of the number and range of effects during normal use of the method, especially concentration ranges and sample types

Precision

- How close independent results are to each other under specified conditions
- Determine typical variability, not minimum variability, i.e. ensure all operational conditions that would typically vary during routine operation are varied

$$s(x_k) = \sqrt{\frac{1}{n-1}\sum_{k=1}^{n}\left(x_k - \bar{x}\right)^2}$$



Precision

- Repeatability standard deviation s_r
  - Smallest variation in results
  - Single analyst performing analysis on the same equipment in 1 laboratory, over a short timescale (e.g. 1 day), using a single set of standards and reagents

- Reproducibility standard deviation s_R
  - Largest variation in results
  - Variability associated with different laboratories employing the same method

Precision

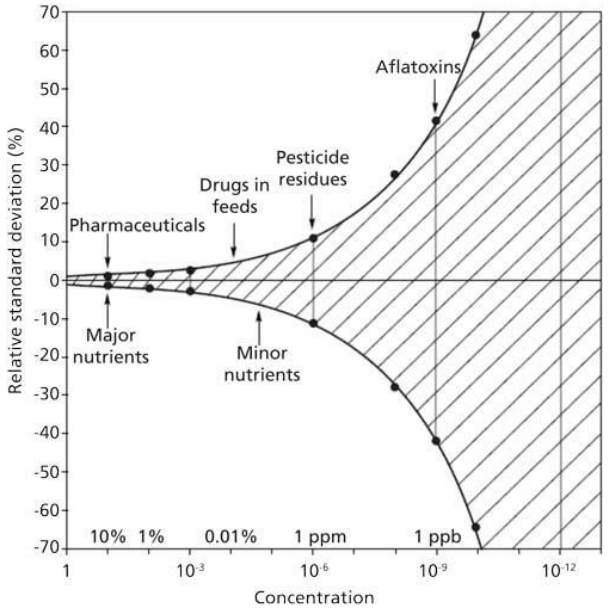
- Within-laboratory reproducibility, s_{Rw}
  - Largest variation that can be associated with results obtained in a single laboratory, i.e. *within-laboratory reproducibility*
  - Should represent typical variation that may be expected under routine operating conditions, e.g. different analysts performing analysis on the different equipment on different days, using independent sets of standards and reagents
- Intermediate precision s_i
  - Variation of specific variables

Repeatability < Intermediate Precision < Reproducibility

Precision

- Experimental determination of precision
  - Repeated analysis of test samples (or CRM) at concentration levels covering the working range of the method – precision is generally dependent on analyte concentration
  - n= 6-15 recommended

$$s(x_k) = \sqrt{\frac{1}{n-1}\sum_{k=1}^{n}\left(x_k - \bar{x}\right)^2}$$



Example:

- A concentration of calibration standard for Cd (mg/L) was analysed 6 times on a single day to determine the instrument precision:

0.231	0.235	0.236
0.224	0.230	0.229

Repeatability:

$$s(x_k) = \sqrt{\frac{1}{n-1}\sum_{k=1}^{n}\left(x_k - \bar{x}\right)^2}$$

$$= 0.004\ mg/L$$

$$\%s_x = \frac{s_x}{\bar{x}} \times 100$$

$$= \frac{0.004}{0.231} \times 100$$

$$= 1.9\%$$

Example:

- The following data was collected from a control chart for Ca-concentration (mg/L) in a water control sample over a period of 3 months. Calculate the % within-laboratory reproducibility of the method.

55.4	54.7	54.8	55.2
53.1	52.0	56.1	55.1

Intermediate precision

$$s(x_k) = \sqrt{\frac{1}{n-1}\sum_{k=1}^{n}\left(x_k - \bar{x}\right)^2}$$

$$= 1.34\ mg/L$$

$$\%s_x = \frac{s_x}{\bar{x}} \times 100$$

$$= \frac{1.34}{54.55} \times 100$$

$$= 2.5\%$$

Example:

- In an interlaboratory comparison for Pb in drinking water, 23 laboratories participated employing the same ISO standard method. The consensus value was 0.0461 mg/L, with the standard deviation for all the participants being 0.0027 mg/L.

% Reproducibility

$$\%s_x = \frac{s_x}{\bar{x}} \times 100$$
$$= \frac{0.0027}{0.0461} \times 100$$
$$= 5.9\%$$

Comparing precision

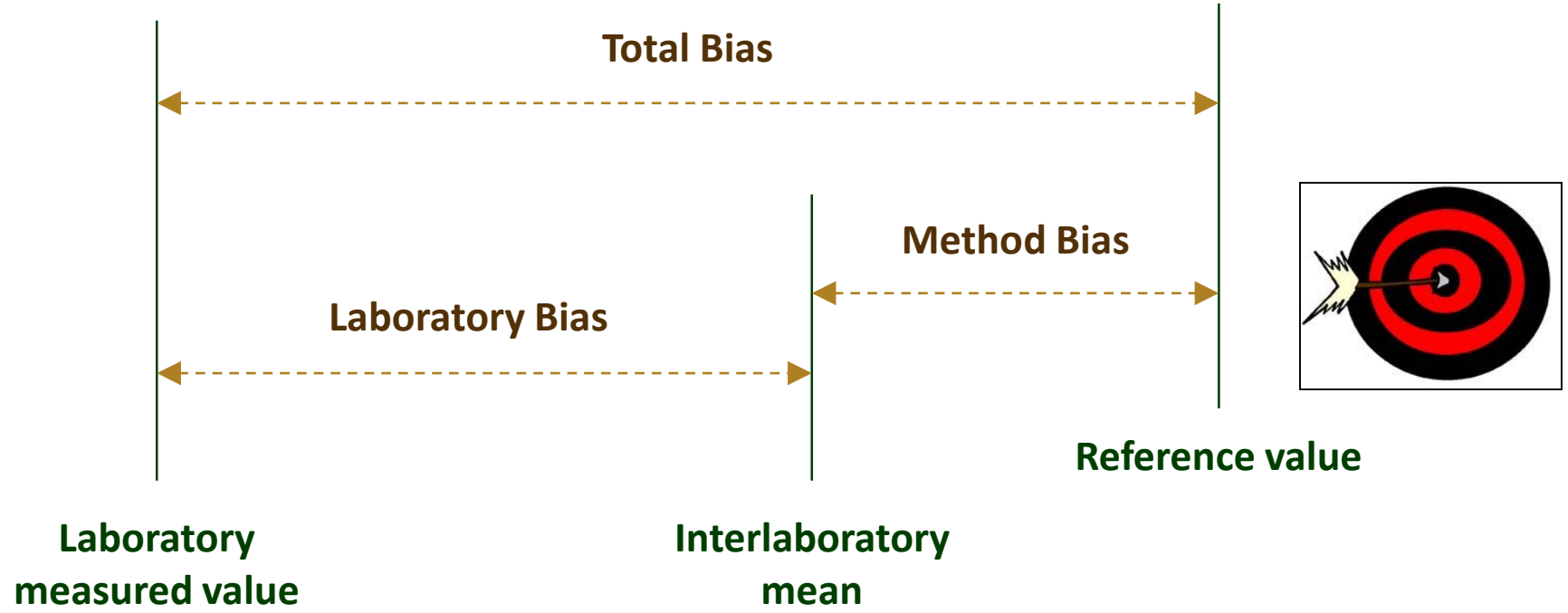
- F-test:
  - Comparison of methods' precision (standard deviation)

$$F = \frac{s_1^2}{s_2^2}$$

  - Where F ≥ 1
  - $F_{calc} < F_{crit}$: No significant difference between two methods' variances

Trueness

- Closeness of a number of measurements to the “true” value, i.e. evaluation of potential systematic error
  - Compare measurement mean to reference value



Trueness

- Establish a reference / trueness through e.g.:
  - The analysis of Certified reference material
  - The analysis of Spiked samples
  - Employing an Alternative method
  - Participation in an Interlaboratory intercomparison / Proficiency Testing scheme
- Criteria for selection of reference material / samples
  - Representative
    - Sample types/matrices
    - Concentration levels of measurand
  - Independent from calibration standards

Trueness

- Considerations when selecting an approach to evaluate trueness:
  - Certified matrix reference material
    - Ideal
    - Exact matching of matrix and analyte concentration may be difficult to achieve
    - Expensive
  - Spiking
    - Behaviour of added measurand is probably different from naturally incurred measurand (e.g. bound to matrix)
    - Unrealistically high recoveries can be expected

Trueness

- Considerations when selecting an approach to evaluate trueness:
  - Alternative method
    - Uncertainty (Reference Method) < Uncertainty (Candidate Method)
    - Tested on real samples
  - Interlaboratory Comparison / PT
    - Exact matching of matrix and analyte concentration may be difficult to achieve
    - Reputability of study, e.g. number of participants, reliability of reference value

Trueness

- General considerations
  - Experimentally
    - At 3 concentration levels
      - Close to limit of detection
      - Mid-range
      - Upper concentration limit
    - n=9 recommended
  - Expected to be negligible or accounted for

Trueness

- Statistical evaluation of data
  - Outlier testing
    - Grubbs
    - Dixon
- Performance criteria
  - % Recovery
  - % Bias
  - t-test
  - E_n or ζ-score

Trueness: Outlier testing

- View results graphically
- Grubbs' test
  - Small number of samples
  - ISO recommended

$$G = \frac{|SuspectValue - \bar{x}|}{s}$$

  - Outlier if $G_{calc} > G_{crit}$

- Dixon's test (Q-test)
  - Sample sizes: n = 3 – 7

$$Q = \frac{|SuspectValue - NearestValue|}{Largest Value - SmallestValue}$$

  - Outlier if $Q_{calc} > Q_{crit}$

n	Q_{crit} CL at 90%	Q_{crit} CL at 95%	Q_{crit} CL at 99%
3	0.941	0.970	0.994
4	0.765	0.829	0.926
5	0.642	0.710	0.821
6	0.560	0.625	0.740
7	0.507	0.568	0.680
8	0.468	0.526	0.634
9	0.437	0.493	0.598
10	0.412	0.466	0.568

Dixon Test

The Dixon test can be used to test for a single outlier in a univariate data set. This test is primarily used for small data sets (Dataplot limits the sample to be between 3 and 30).

Specifically, given a set of ordered observations Y1, Y2, ..., YN, the Dixon test is computed as follows:

Sample Size	Test for Minimum	Test for Maximum
$3 \le N \le 7$	$(Y_2-Y_1)/(Y_N-Y_1)$	$(Y_N-Y_{N-1})/(Y_N-Y_1)$
$8 \le N \le 10$	$(Y_2-Y_1)/(Y_{N-1}-Y_1)$	$(Y_N-Y_{N-1})/(Y_N-Y_2)$
$11 \le N \le 13$	$(Y_3-Y_1)/(Y_{N-1}-Y_1)$	$(Y_N-Y_{N-2})/(Y_N-Y_2)$
$14 \le N \le 30$	$(Y_3-Y_1)/(Y_{N-2}-Y_1)$	$(Y_3-Y_1)/(Y_{N-2}-Y_1)$

https://www.itl.nist.gov/div898/software/dataplot/refman1/auxillar/dixon.htm

Dixon Test

Dixon's Q test, or just the "Q Test" is a way to find outliers in very small, normally distributed, data sets. Small data sets are usually defined as somewhere between 3 and 7 items

$$Q_{exp} = \frac{x_2 - x_1}{x_n - x_1}$$

Finding the Q statistic for different sample sizes (n) of between 8 and 30 (in Step 2 above):

8< n >10: use R_{11}: $$r_{11} = \frac{x_2 - x_1}{x_{n-1} - x_1}$$

11< n >13: use R_{21}. $$r_{21} = \frac{x_3 - x_1}{x_{n-1} - x_1}$$

14< n >30: use R_{22}. $$r_{22} = \frac{x_3 - x_1}{x_{n-2} - x_1}$$

https://www.statisticshowto.datasciencecentral.com/dixons-q-test/

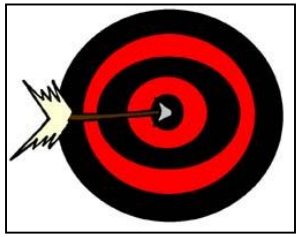
Trueness

- % Recovery
  - Expected to be close to 100%
    - Depending on application field and concentration levels

$$\%\,Recovery = \frac{C_{meas}}{C_{Ref}} \times 100$$

- % Bias
  - Expected to be close to 0%
    - Depending on application field and concentration levels

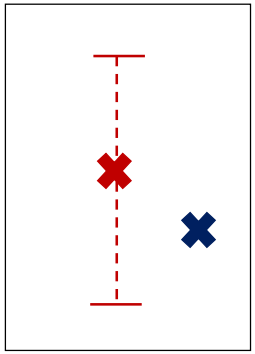
$$\%\,Bias = \frac{C_{meas} - C_{Ref}}{C_{Ref}} \times 100$$



Trueness

- t-test
  - Comparison of mean with reference value
  - Uncertainty of reference value not considered
  - Null hypothesis (H_0): No significant difference between measured and "true" value
    - H_0: $t_{calc} < t_{crit}$: No significant bias

$$t = \frac{(\bar{x} - \mu)\sqrt{n}}{s}$$

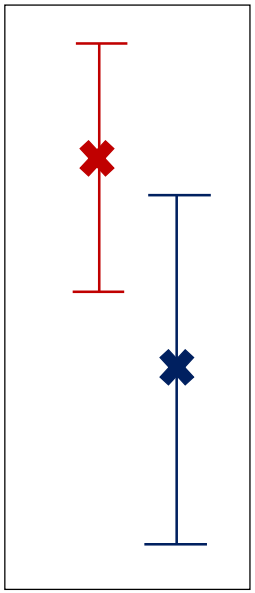


Trueness

- E_n-score / ζ-score
  - A measure of agreement between the reference value (X) and the method's result (x), i.e. are they within their respective uncertainty ranges?

$$E_n = \frac{x - X}{\sqrt{U(x)^2 + U(X)^2}} \qquad \zeta = \frac{x - X}{\sqrt{u(x)^2 + u(X)^2}}$$

En Score	ζ-score	
$\lvert E_n \rvert \leq 1$	$\lvert \zeta \rvert \leq 2$	Satisfactory
$\lvert E_n \rvert > 1$	$\lvert \zeta \rvert > 2$	Unsatisfactory



Trueness
- Example

- During method validation for analysis of alloy samples a CRM was analysed 10 times for Zn:

  C_{CRM} = 2.013 ± 0.034 g/kg

  C_{meas} = 2.035 ± 0.054 g/kg

- Calculate the % Bias and % Recovery for this method.

$$\%Bias = \frac{C_{meas} - C_{CRM}}{C_{CRM}} \times 100$$

$$= \frac{2.035 - 2.013}{2.013} \times 100$$

$$= 1.09\%$$

$$\%\,Recovery = \frac{C_{meas}}{C_{CRM}} \times 100$$

$$= \frac{2.035}{2.013} \times 100$$

$$= 101.09\%$$

Trueness

- Example

- During method validation for analysis of alloy samples a CRM was analysed 10 times for Zn:

  C_{CRM} = 2.013 ± 0.034 g/kg

  C_{meas} = 2.035 ± 0.054 g/kg

- Calculate if there is a significant difference between the mean and the consensus true value, employing the t-test.

$$t_{calc} = \frac{(\bar{x} - \mu)\sqrt{n}}{s} = \frac{(2.035 - 2.013)\sqrt{10}}{0.054}$$

$$= 1.29$$

$$t_{crit} = 2.26$$

H_0: $t_{calc} < t_{crit}$: No significant bias

Trueness
- Example

- During method validation for analysis of alloy samples a CRM was analysed 10 times for Zn:

  C_{CRM} = 2.013 ± 0.034 g/kg

  C_{meas} = 2.035 ± 0.054 g/kg

- Calculate if there is a significant difference between the mean and the consensus true value, employing the E_n-score.

$$E_n = \frac{x - X}{\sqrt{U_x^2 + U_{ref}^2}} = \frac{2.035 - 2.013}{\sqrt{(0.054)^2 + (0.034)^2}} = 0.34$$

$|E_n| \leq 1$: No significant bias

Trueness
- Example

- During method validation for analysis of alloy samples a CRM was analysed 10 times for Zn:

  C_{CRM} = 2.013 ± 0.034 g/kg

  C_{meas} = 2.035 ± 0.054 g/kg

- Calculate if there is a significant difference between the mean and the consensus true value, employing the ζ-score.

$$\zeta = \frac{x - X}{\sqrt{u(x)^2 + u(X)^2}} = \frac{2.035 - 2.013}{\sqrt{(0.027)^2 + (0.017)^2}}$$

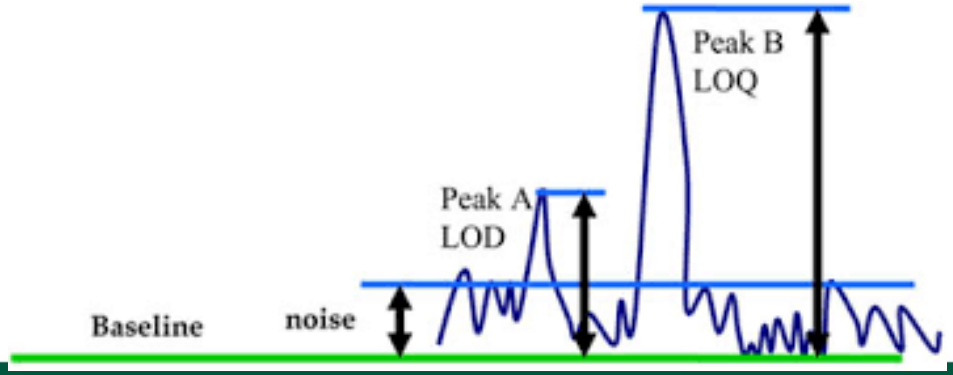
$$= 0{,}69$$

$|\zeta| \leq 2$: No significant bias

Validation parameters

Limit of Detection & Quantification

- Limit of Detection (LOD):
  - Lowest concentration that can be reliably detected, but not quantified.
- Limit of Quantification (LOQ):
  - Lowest concentration that can be accurately quantified / *at which performance is acceptable for typical application*.



Validation parameters
LOD & LOQ

- Limit Of Detection (LOD)

$$LOD = 3 \cdot s'_0$$

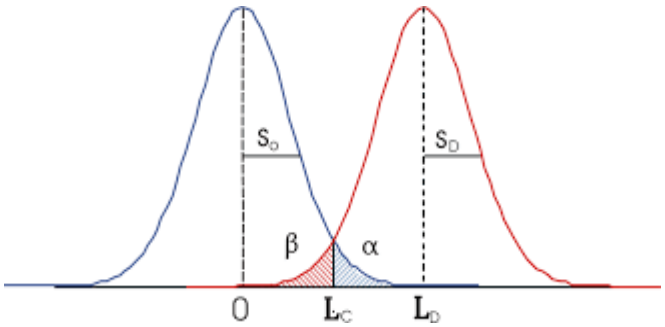
- Limit of Quantification (LOQ)

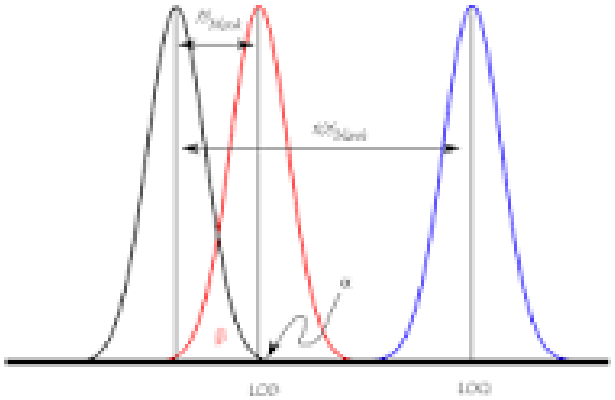
$$LOQ = k_Q \cdot s'_0$$

With:

- k_Q = 10 (corresponding to 10% RSD)

- Note: s'_0 should be in concentration units





Validation parameters
LOD & LOQ

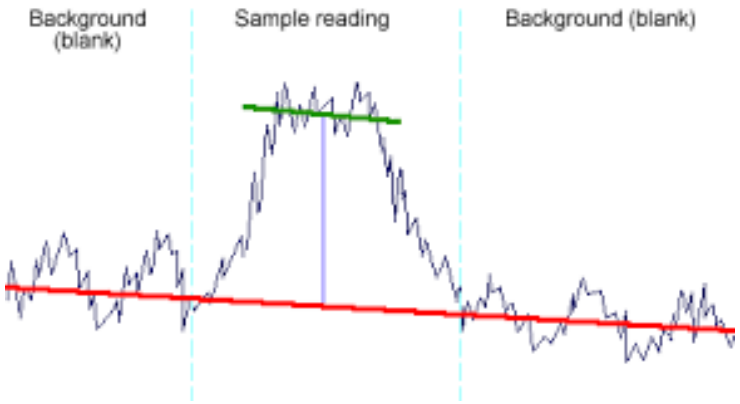
- Experimentally:
  - Sample:
    - Blank
      - Reagent blank
      - Sample blank - matrix
    - Samples with concentration @ LOD
  - Number of measurements
    - A reliable estimate of standard deviation requires 6-15 measurements.
    - In practice, typically 10
  - Calculate Standard deviation, s'_0

Validation parameters
LOD & LOQ

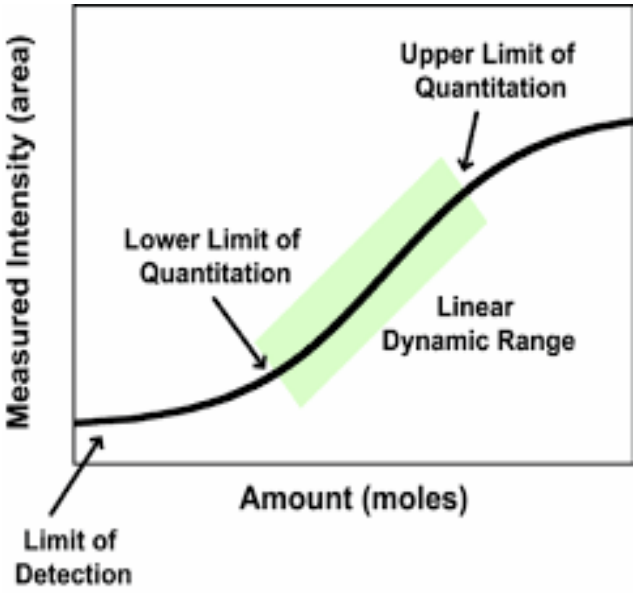
- Instrument LOD:
  - Sample blank / low concentration sample directly analysed on instrument, i.e. no sample preparation.
- Method LOD:
  - Sample blank / low concentration sample taken through complete sample preparation procedure.
  - Results calculated as stipulated in measurement procedure (i.e. corrected for dilution effects)

Validation parameters
LOD & LOQ

- Considerations regarding reliability
  - If blank values varies significantly from day-to-day:
    - s'_0 should be intermediate precision, rather than repeatability
  - Samples concentrations expected to be close to LOD:
    - LOD/LOQ should be monitored regularly – estimates of standard deviation are inherently variable



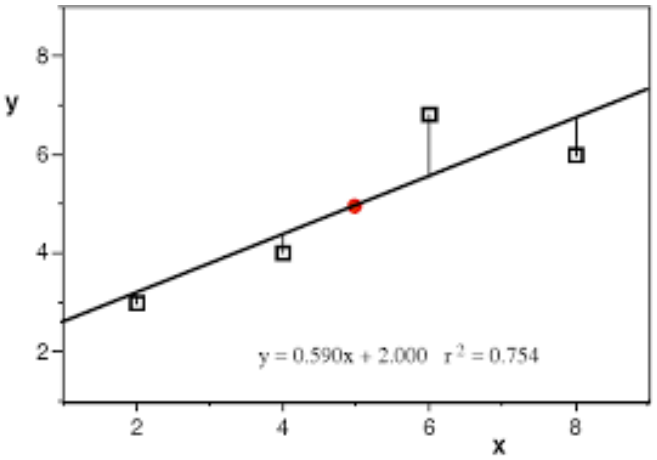
Working range



- Interval over which method provides results with an acceptable uncertainty.
  - Lower end:
    - LOQ, or
    - Minimum expected concentration in samples
  - Upper end:
    - Concentrations at which anomalies in analytical sensitivities become apparent _ *Linearity*

Working range

- Linearity Evaluation
  - Visually inspect
  - Regression statistics
    - e.g. $r^2 > 0.9995$
  - Residuals plot
    - Random distribution about zero: Linear
    - Systematic trends: Non-linear
- Non-linearity corrected for by:
  - Restricted operating range
  - Non-linear calibration



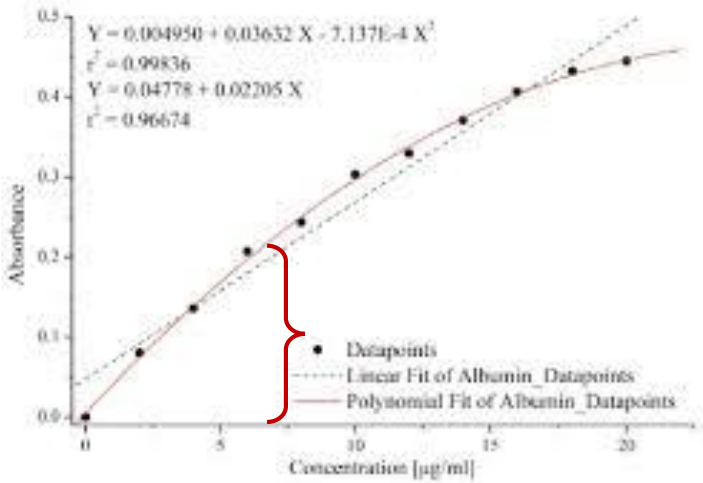
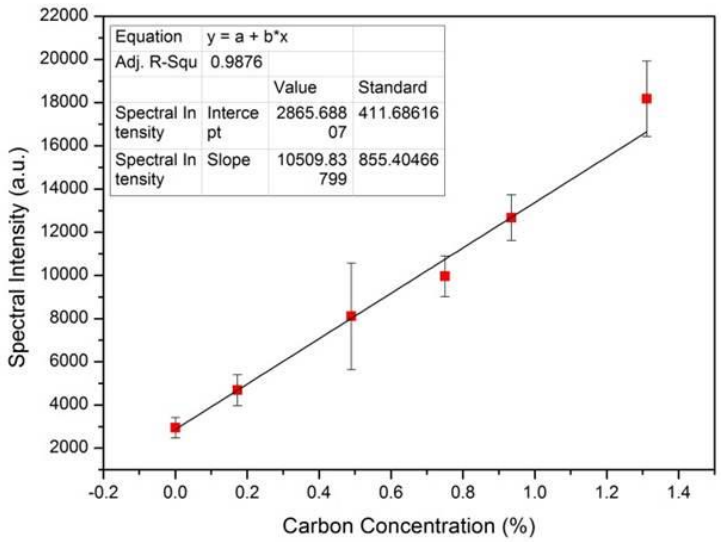
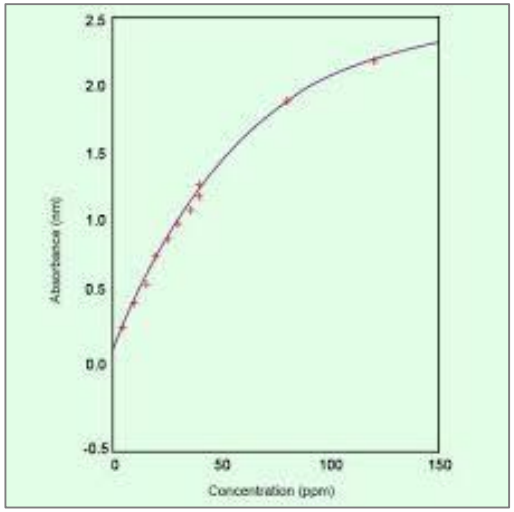


Figure 3 The essence of a linear model optimization for experimental data in Least Squares Method. The model-experiment differences are indicated by solid lines.

Working range

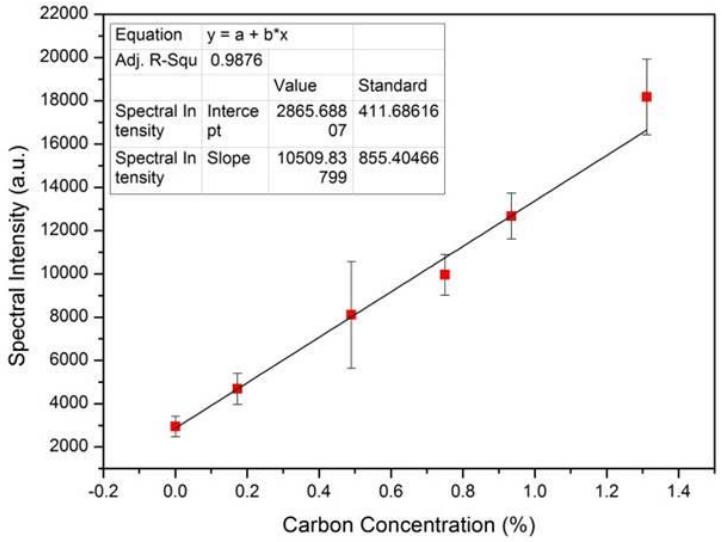
- Instrument calibration
  - Linear calibration (n ≥ 5)
  - Quadratic calibration (n ≥ 7)
  - Higher functions not advisable
  - Weighted fit (standard deviation proportional to concentration)
  - Transformation of values – e.g. log-normal calibration





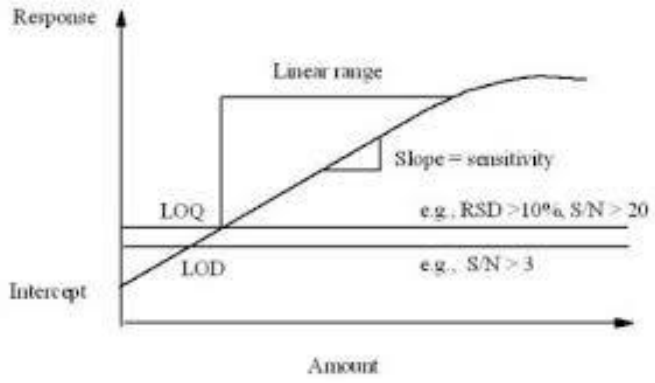
Working range

- Method working range
  - Instrument working range
  - Sample preparation restrictions
    - Minimum/maximum sample size
    - Dilution factors



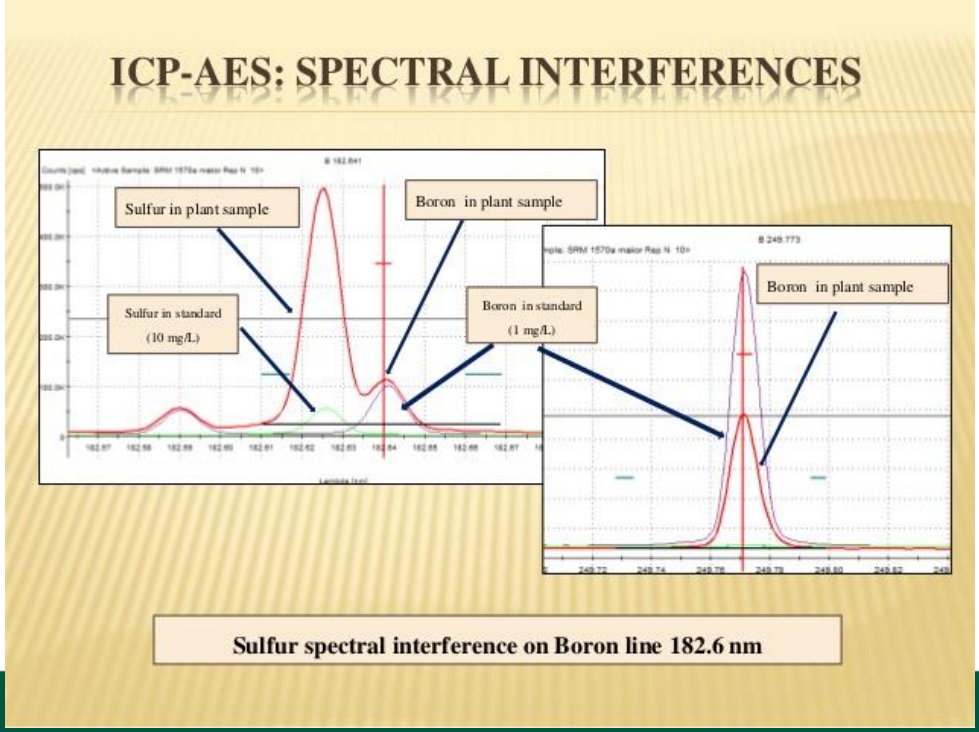
Sensitivity

- Change in instrument response corresponding to change in measured quantity
- As part of instrument quality assurance the sensitivity can be checked routinely to ensure that it doesn't fall below a minimum level



Selectivity/specificity

- The degree to which a method responds uniquely to the required analyte
- Interference may increase (enhance) or decrease (suppress) analyte signal



Selectivity/specificity

- Experimentally:
- Investigate the effects of potential interferents:
  - Add potential interferent to blank
  - Add potential interferent to samples
  - Independent technique
  - Certified Reference Material
- Determine Trueness
  - Recovery / Bias
  - T-test
- Normally used to demonstrate insignificant effects

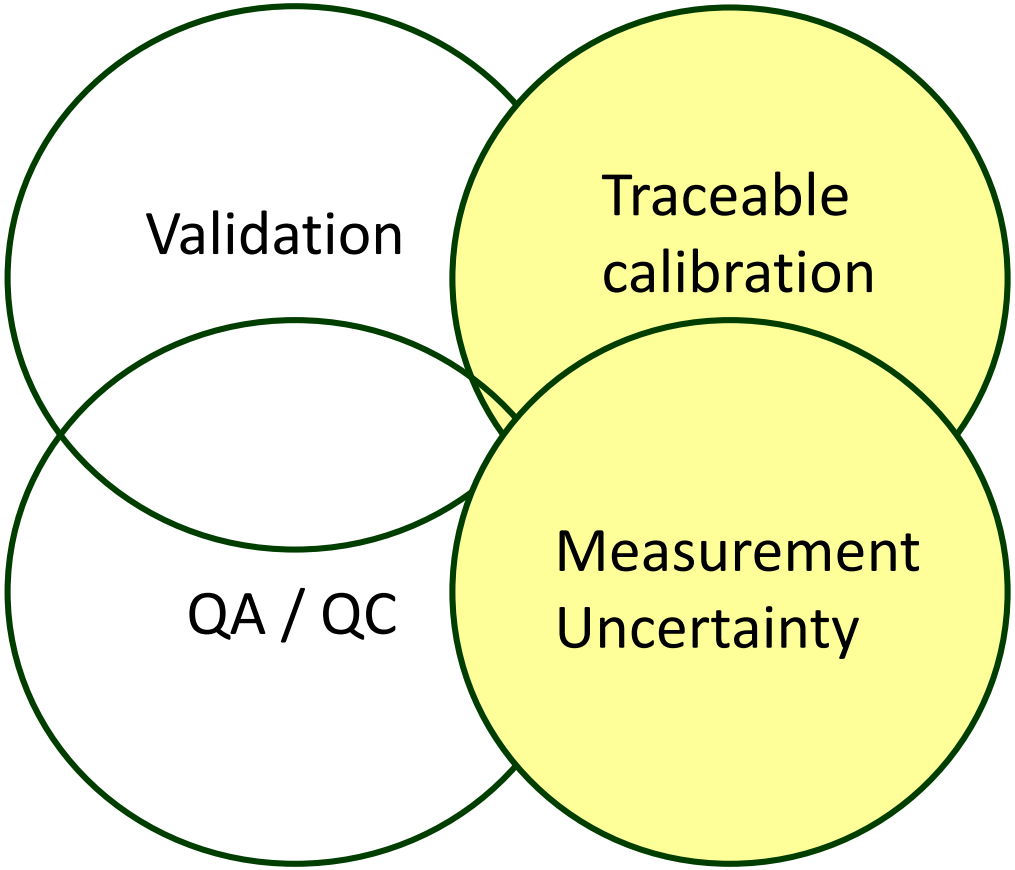
Robustness (ruggedness)

- Measure of a method's capacity to remain unaffected by small, but deliberate variations in method parameters
- Provides indication of method's reliability during normal use
- Required for:
  - In-house methods
  - Method developed from scientific literature
  - Standard methods used outside the method's scope
- *Not required:*
  - *Standard methods used within method's scope*

Robustness (ruggedness)

- Evaluation method:
  - Identify variables that could have significant effect on method
  - Make deliberate changes to variables identified to determine the effect of changes on the results
  - Significance testing to establish if statistically relevant
  - If the effect is significant:
    - Ensure that variable is effectively controlled when using the method
    - Improve method

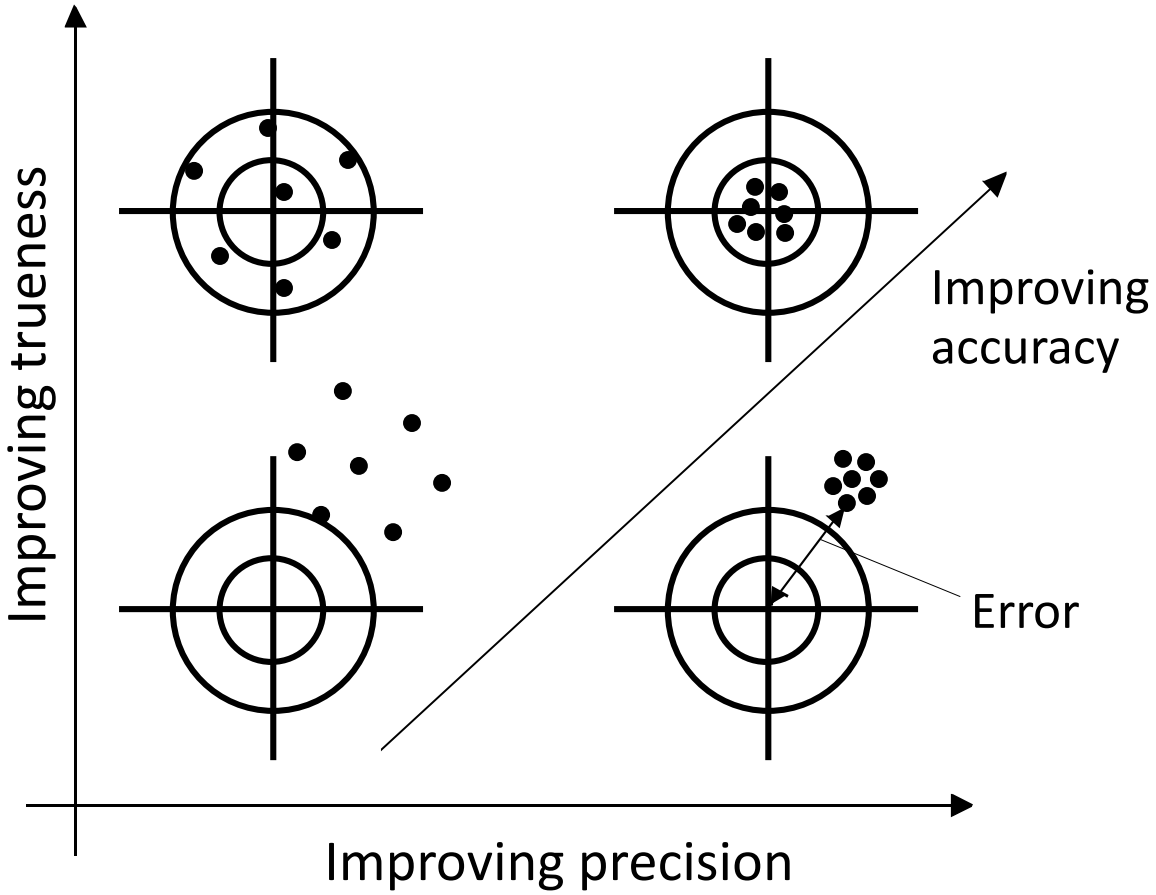
Ensuring valid Analytical Measurements



Fit for purpose..

Performance Characteristic	Type of Analytical Application		
	Qualitative	Quantitative: Impurity	Quantitative: Main comp.
Selectivity	✓	✓	✓
LOD			
LOQ		✓	
Working range, incl. Linearity		✓	✓
Trueness		✓	✓
Precision (s_r, s_i)		✓	✓

Accuracy:
Trueness & Precision



from: LGC, vamstat II

Conclusion

- For valid analytical measurements, we need
  - Measurement traceability through:
    - Calibrated equipment
    - Certified calibration standards
  - Validated methods that are:
    - Fit for purpose
    - Accuracy (trueness and precision) critical

- Also…
  - Quality control
  - Estimation of uncertainty of measurement

Sodium in Food: Proficiency testing scheme

Round 1	Savoury Stock Powder (Oct 2018)
Round 2	Bread (Dec 2018)
Round 3	Instant Noodles (Feb 2019)
Round 4	Fat Spread (Apr 2019)
Round 5	Flavoured Snack/Crisp (May 2019)
Round 6	Cured processed meat product (Jul 2019)

Sodium in Food: Proficiency testing scheme

Main parameter: Salt (sodium) content	
Optional Parameters:	
Iron and Zinc	Moisture
Protein	Fat
Total sugar	Cholesterol
Dietary fibre	Vitamins A, B1, B2, B3, B6, folic acid
Energy	

dprevoo@nmisa.org

nmisa
National Metrology Institute of South Africa

We measure what matters

Thank you

Maré Linsky
mlinsky@nmisa.org