Method Validation / Fitness for purpose of an Analytical method

Maré Linsky
mlinsky@nmisa.org
5 December 2017

nmisa
National Metrology Institute of South Africa

Overview

- Basic Principles
- Method Validation Parameters
  - Trueness
  - Precision
  - Limit of Detection (LOD) & Limit of Quantification (LOQ)
  - Working range (Linearity)
  - Selectivity / Specificity
  - Robustness (ruggedness)
  - Measurement Uncertainty
  - Metrological Traceability
- Quality Control
  - External processes
  - Internal processes (Control samples & Control charts)

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

Importance of Analytical Measurement

- Impact of these measurements (or of the decisions made based on these results) could be far-reaching
  - Health
  - Cost (fines)
  - Legal
- Ensuring the reliability of these measurements is the responsibility of the Analytical Chemist

What is a measurement?

- A set of operations to determine the value of a quantity
- Measurements are made using (a) measuring instrument(s) according to a specific method / procedure

Measurand

Quantity intended to be measured

Measurement procedure

Detailed description of a measurement according to one or more measurement principles and to a given measurement method, based on a measurement model and including any calculation to obtain a measurement result

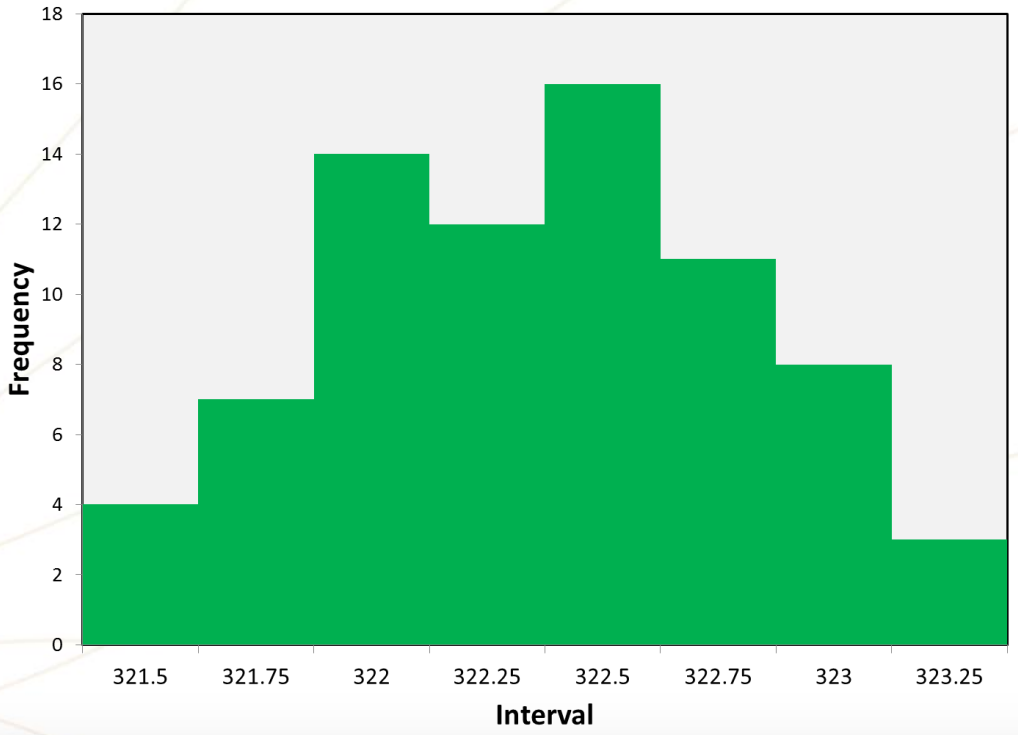
Result of a measurement

Set of quantity values being attributed to a measurand together with any other available relevant information

Analytical Results Vary

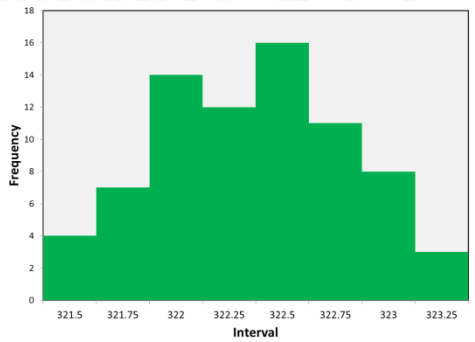
Variations are always present.

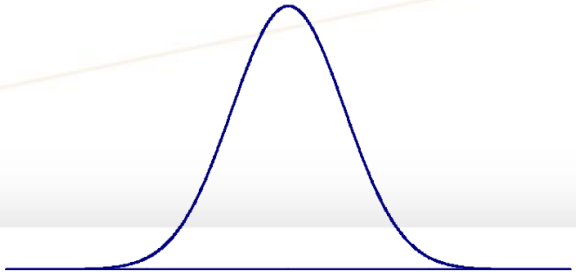
322,29
322,49
322,28
322,17
321,67
321,76
321,75
322,17
321,58
322,36
322,40



Population vs. Sample

Sample	Population
A selection of 1000 inhabitants of a town	All inhabitants of a town
Any number of measurements of lead in samples from Lake Malawi	*Not possible*

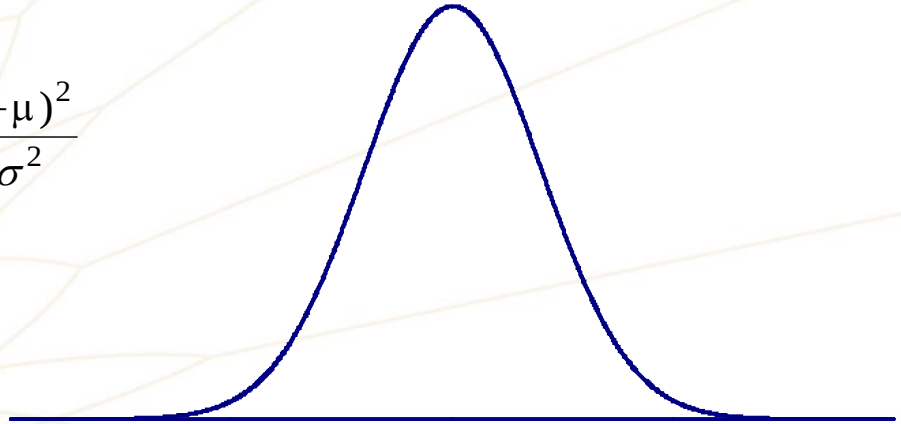




Normal Distribution

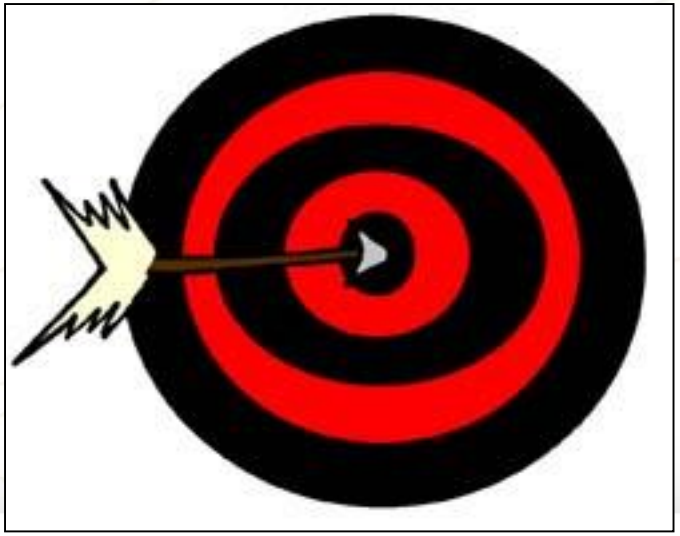
- The curve is symmetrical about μ
- Completely determined by μ and σ
- The greater the value of σ the greater the spread of the curve

$$y = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{(x-\mu)^2}{2\sigma^2}}$$



Trueness

Closeness of agreement between a test result or measurement result and the true value



True value

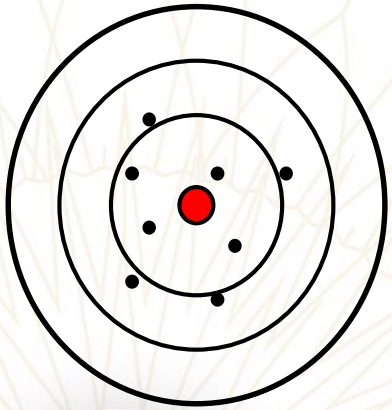
Quantity value consistent with the definition of a quantity.

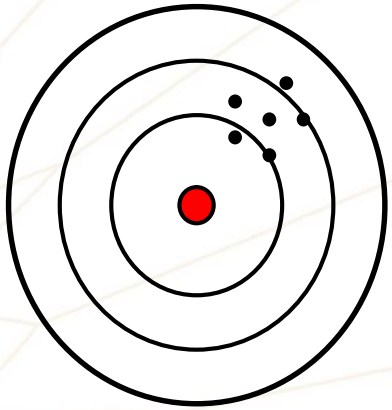
Conventional true value

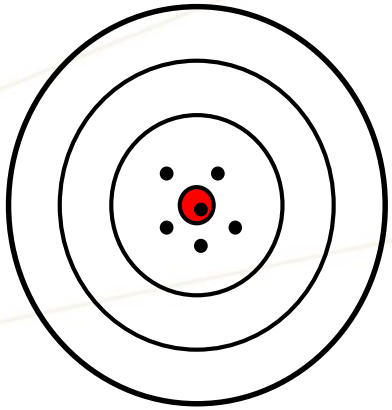
A value attributed to a particular quantity and accepted, sometimes by convention, as having an uncertainty appropriate for a given purpose.

Precision

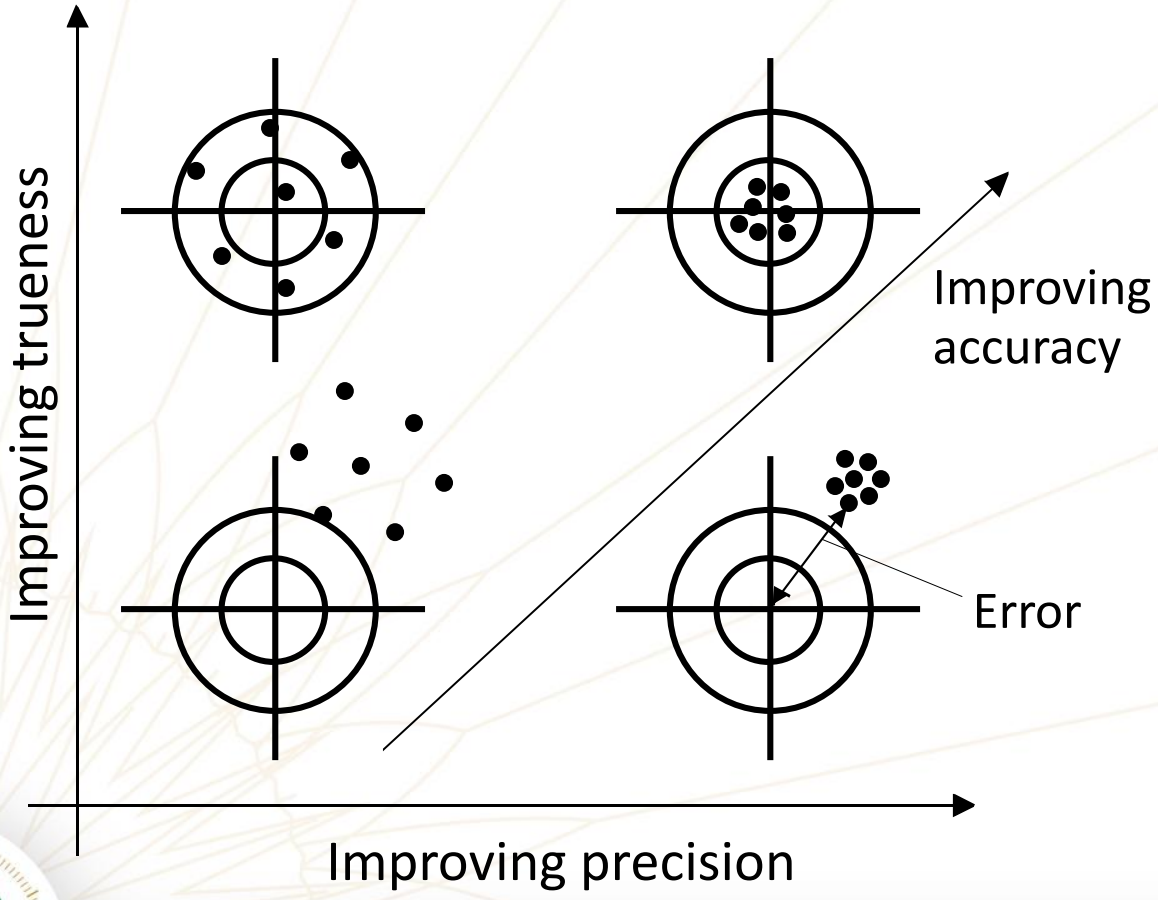
Closeness of agreement between independent test / measurement results obtained under stipulated conditions







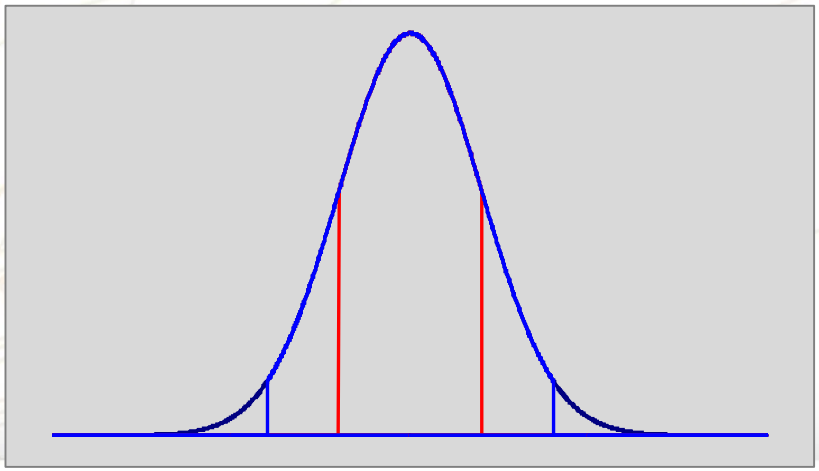
Accuracy: Trueness & Precision



from: LGC, vamstat II

Normal Distribution:
Important Properties

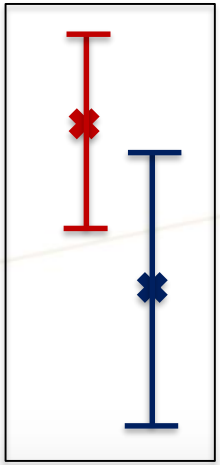
- Approximately 68% (68,27%) of the data lie within $\mu \pm 1\sigma$
- Approximately 95 % (95,45%) of the data lie within $\mu \pm 2\sigma$
- Approximately 99,7 % (99,73%) of the data lie within $\mu \pm 3\sigma$



Uncertainty of Measurement (UoM)

Parameter, associated with the result of a measurement, that characterises the spread of values that could reasonably be attributed to the measurand

$$m = 1000.00250 \pm 0.00050\ g$$



What is uncertainty of measurement?

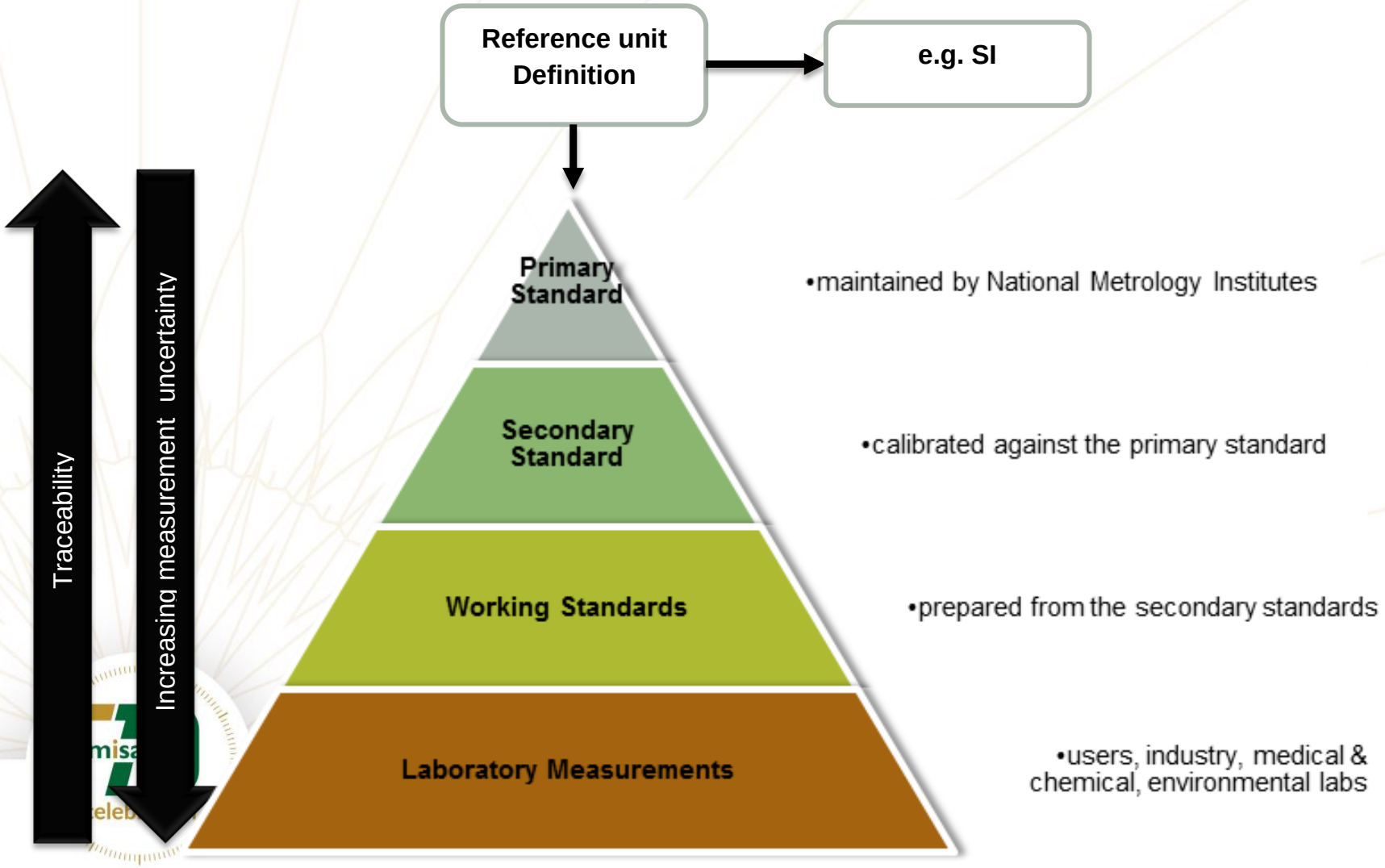
- It tells us something about how much you can trust the measurement i.e. the quality of the measurement
- We need two numbers to quantify uncertainty:
  - The width of the margin of doubt, the confidence interval, and
  - The confidence level, how sure we are that the true value is within the margin of doubt.

Metrological traceability

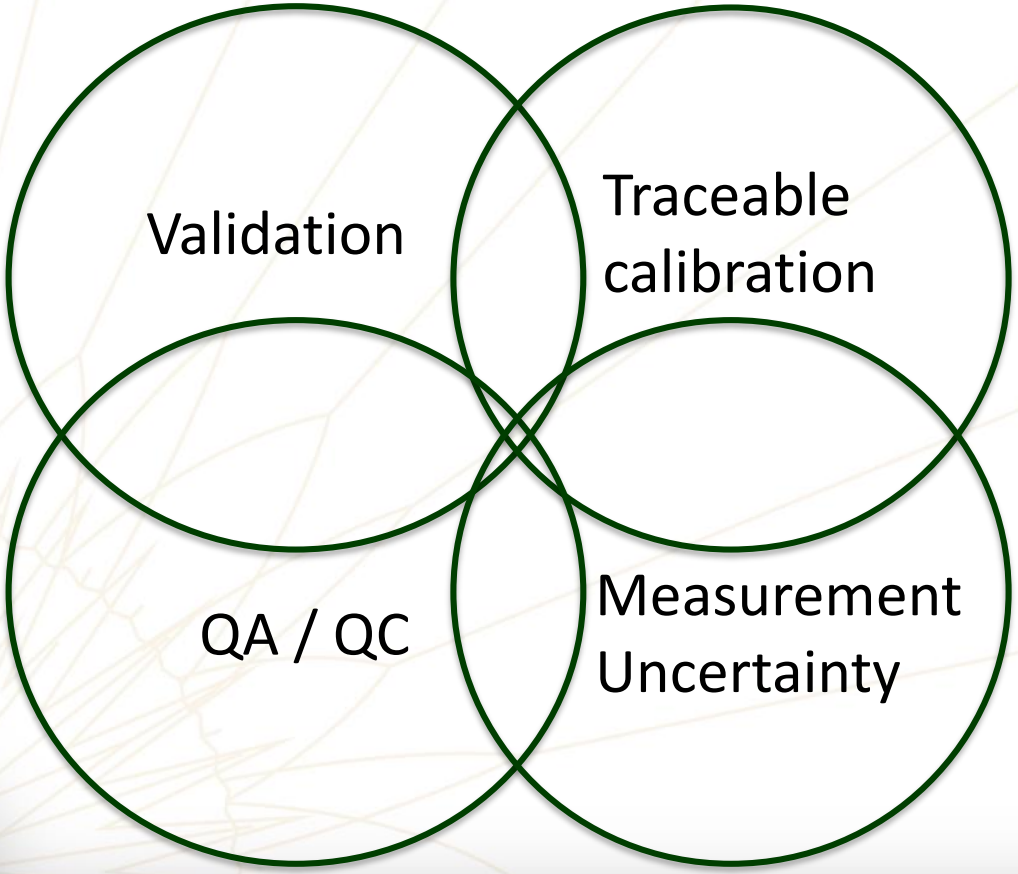
Result can be related to a reference through a documented unbroken chain of calibrations, each contributing to the measurement uncertainty

OR

Calibration hierarchy to establish measurement traceability



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

Performance Criteria

- Method Validation
  - Trueness
  - Precision
  - Limit of Detection (LOD) & Limit of Quantification (LOQ)
  - Working Range & Linearity
  - Sensitivity
  - Selectivity / Specificity
  - Measurement Uncertainty
  - Metrological Traceability

Method Validation Approaches

- Data on overall method performance parameters are obtained from:
  - Interlaboratory studies
  - Single laboratory: In-house validation protocols
- It is the responsibility of the laboratory to ensure that a method is fit for its intended use.

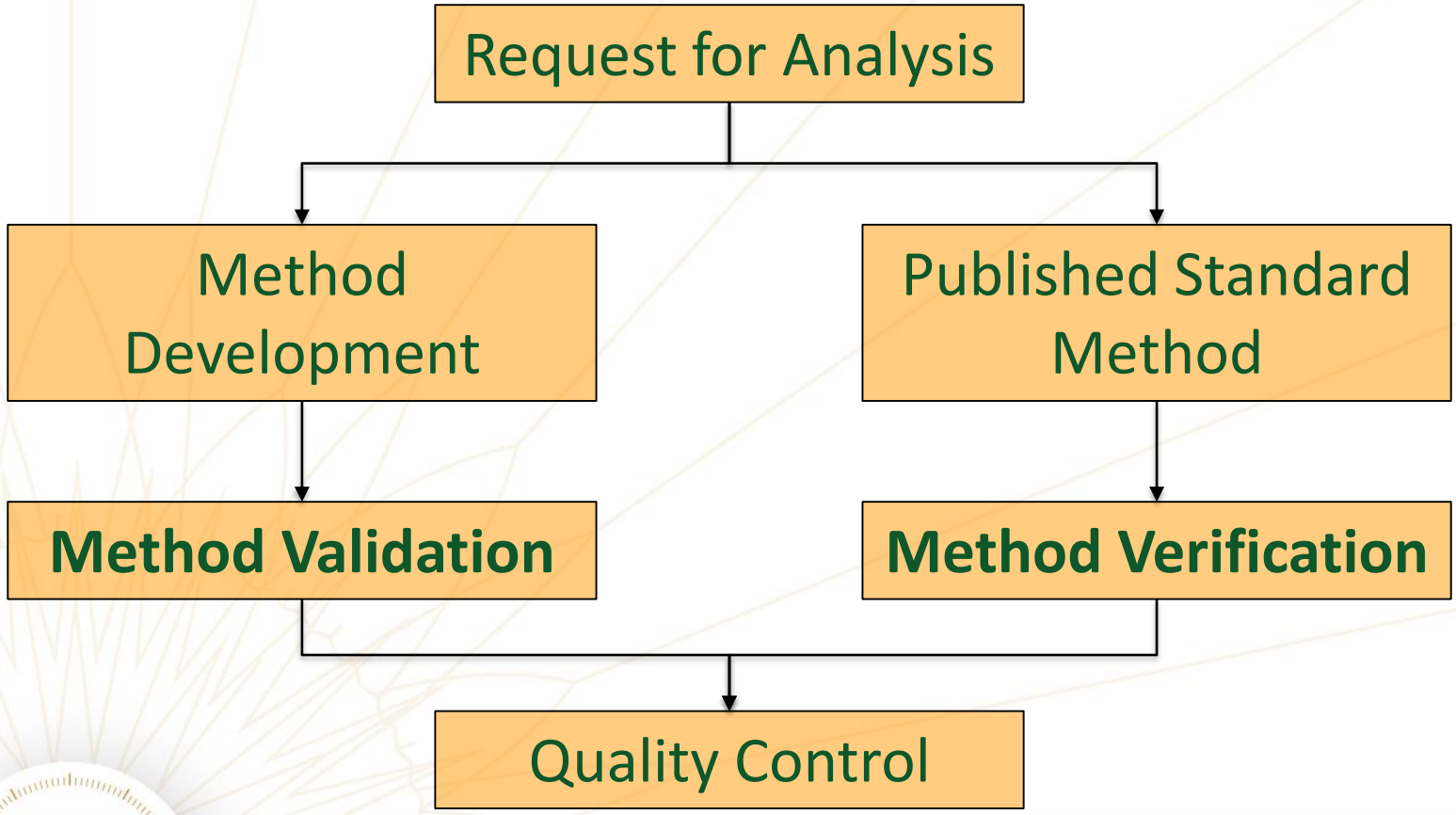
Method Validation Approaches

- 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
  - Validation is a balance between costs, risks and technical possibilities (routine vs ad hoc)
    - Selectivity; LOD, LOQ; Working Range; Trueness; Precision; Ruggedness; Uncertainty
  - Emphasis is on identifying and removing or reducing significant effects, i.e. continue with method development if method performance is not satisfactory

Validation vs. Verification



Validation plan

- Scope
  - Clearly define method, measurand, types of samples
- Criteria
  - Identification of relevant performance criteria
  - Acceptance criteria
- Performance evaluation
  - Evaluation of method's performance, employing suitable test sample, standard and blank materials
- Report

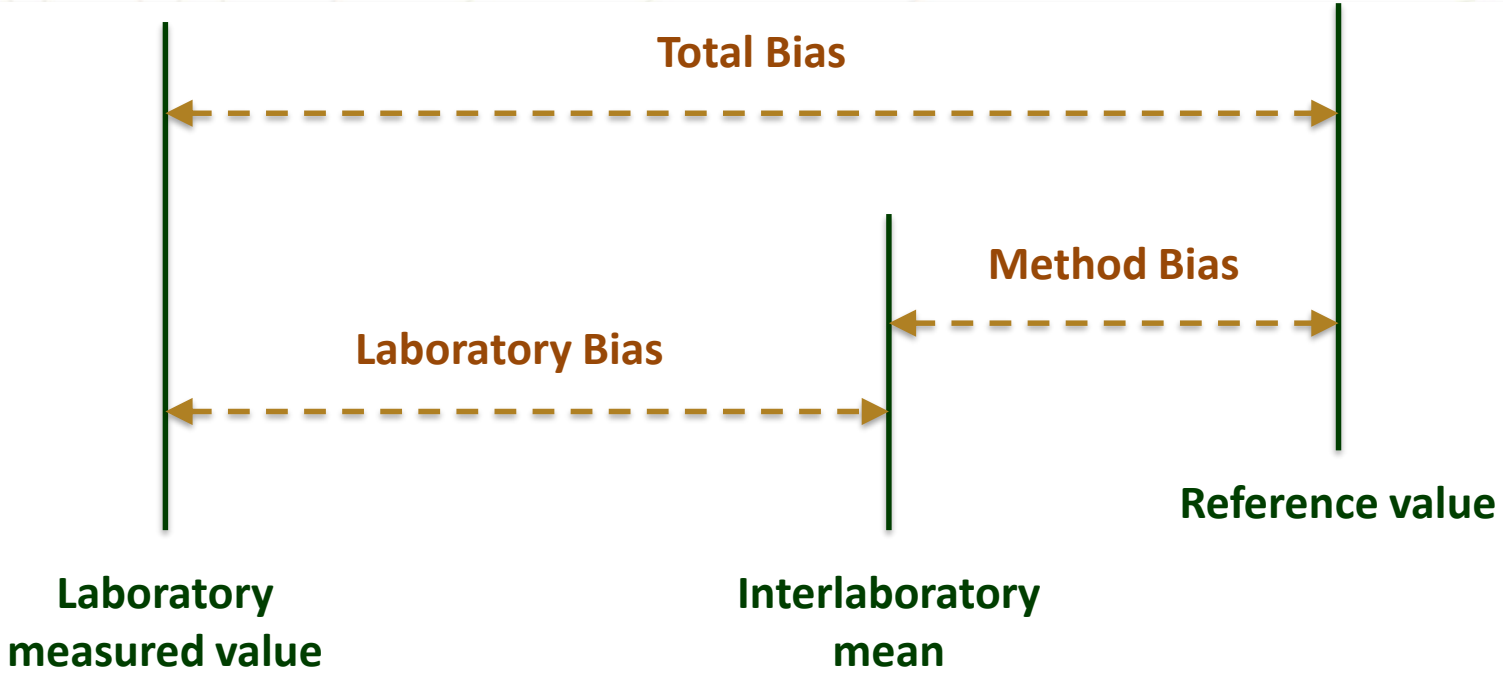
How to do validation studies?

- Representativeness
- Representative variation

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

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

- Reference value:
  - Certified reference material
  - Spiked samples
  - Alternative method
  - *Interlaboratory intercomparison*
- Criteria for reference material / samples
  - Representative
    - Sample types/matrices
    - Concentration levels of measurand
  - Independent from calibration standards

Trueness

- Considerations
  - Spiking
    - Behaviour of added measurand probably different from naturally incurred measurand (e.g. bound to matrix)
    - Unrealistically high recoveries can be expected
  - Alternative method
    - Uncertainty (Reference Method) < Uncertainty (Candidate Method)
    - Tested on real samples

Trueness

- 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

- Outlier testing
  - Grubbs
  - Dixon
- % Recovery
- % Bias
- t-test
- E_n-score
- f-test

Trueness

Statistical evaluation

- Outlier testing
  - Small number of samples
  - Grubbs' test
    - ISO recommended

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

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

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

Trueness

Statistical evaluation

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

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

Trueness

Statistical evaluation

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

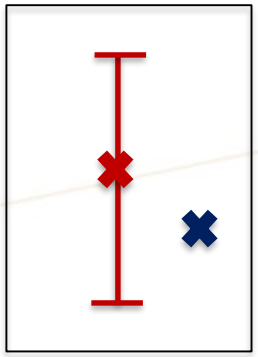
$$\%Bias = \frac{C_{meas} - C_{Ref}}{C_{Ref}} \times 100$$

Trueness

Statistical evaluation

- 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

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



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

Trueness

Statistical evaluation

- t-test (two means)
  - Samples from populations with equal standard deviations (F-test)

$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \qquad s^2 = \frac{(n_1 - 1)\cdot s_1^2 + (n_2 - 1)\cdot s_2^2}{(n_1 + n_2 - 2)}$$

  - Samples from populations with statistically different standard deviations (F-test)

$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{s\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

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

Trueness
Statistical evaluation

- 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

Statistical evaluation

- Paired t-test
  - Comparing results for different samples analysed employing 2 different methods

$$t = \frac{\bar{d} \cdot \sqrt{n}}{s_d}$$

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

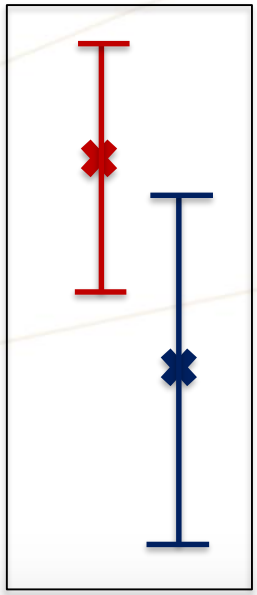
Trueness

Statistical evaluation

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

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

|En| ≤ 1	Satisfactory
$\|E_n\| > 1$	Unsatisfactory



Trueness - Example 1a

- 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 1b

- 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 1c

- 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.

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

$$= 0.34$$

|En| < 1: No significant bias

Trueness
- Example 2

t-test:

A method (method 1) for determining the concentration of selenium is compared with a reference method (method 2).

	Mean	s	n
Method 1	5,40	1,471	5
Method 2	4,76	2,750	5

Trueness
- Example 2

- f-test:

$$F_{calc} = \frac{s_2^2}{s_1^2} = \frac{(2.750)^2}{(1.471)^2} = 3.495$$

$$F_{calc} = 9.605$$

$F_{calc} < F_{crit}$: No significant difference

- Pooled standard deviation s_c:

$$s_c = \sqrt{\left(\frac{s_1^{\,2} \times (n_1 - 1) + s_2^{\,2} \times (n_2 - 1)}{n_1 + n_2 - 2} \right)}$$

$$= \sqrt{\left(\frac{1.471^2 \times (5-1) + 2.750^2 \times (5-1)}{5+5-2} \right)} = 2.205$$

Trueness
- Example 2

- Student t-test:

$$t = \frac{\overline{x_1} - \overline{x_2}}{s_c\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{5.40 - 4.76}{2.205\sqrt{\left(\frac{1}{5} + \frac{1}{5}\right)}} = 0.46$$

  - t_{crit} = 2.3 for 8 degrees of freedom, LOC = 95%
  - $t_{calc} < t_{crit}$: No significant bias

|0.46| < 2.3: No significant bias

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

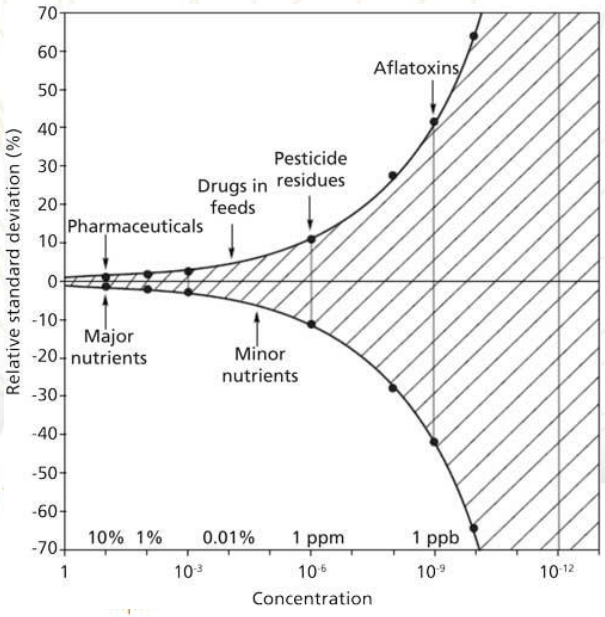
Precision

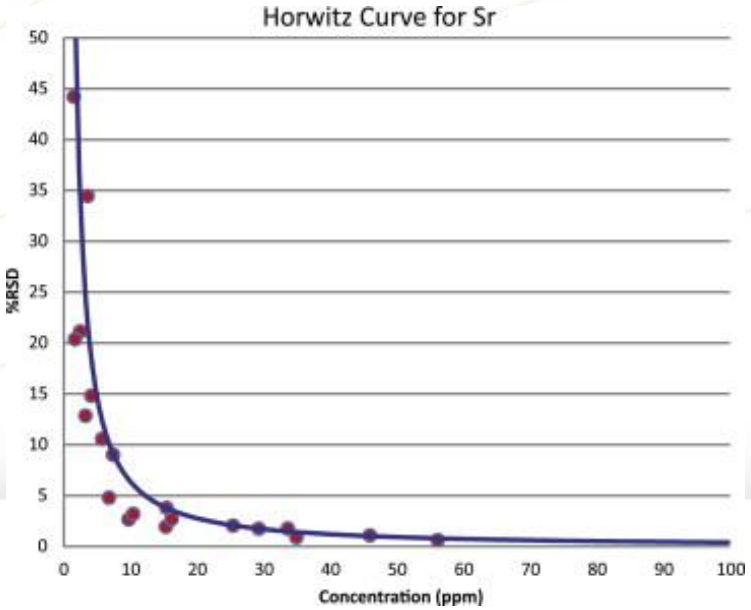
- Intermediate precision s_i
  - Largest variation that can be associated with results obtained in a single laboratory, i.e. *within-laboratory reproducibility*
  - Should represent typical variation that may be expected under routine operating conditions, e.g. different analysts performing analysis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

Repeatability < Intermediate Precision < Reproducibility

Precision

- Experimentally
  - 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





Precision - ANOVA

- Simultaneous determination of repeatability and reproducibility
- Subsamples analysed across a number of different runs, with maximum variation in conditions between the runs (e.g. different days, analysts, equipment)
  - n = 6 – 15 groups of duplicate measurements

Precision
- Example: Repeatability

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

  0.231, 0.235, 0.236, 0.224, 0.230, 0.229

$$\%s_r = \frac{SD}{\bar{x}} \times 100$$

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

$$= 1.9\%$$

Precision
- Example: Intermediate precision

- 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

$$\%s_i = \frac{SD}{\bar{x}} \times 100$$

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

$$= 2.5\%$$

Precision
- Example: Reproducibility

- 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 reproducibility standard deviation for all the participants being 0.0027 mg/L.

$$\%s_R = \frac{SD}{\bar{x}} \times 100$$

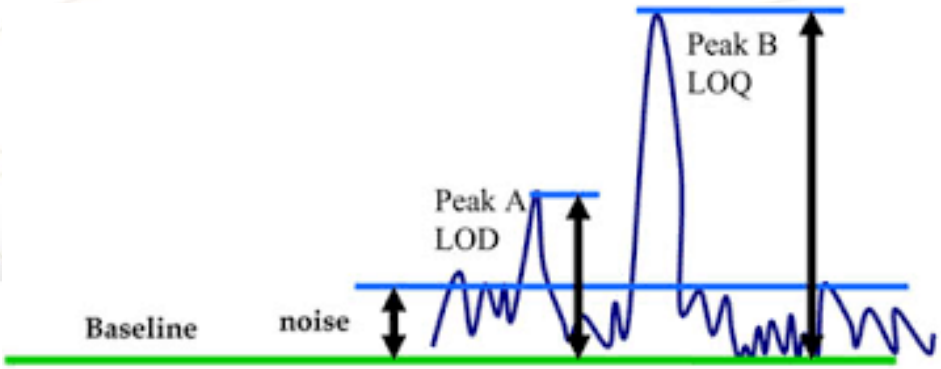
$$= \frac{0.0027}{0.0461} \times 100$$

$$= 5.9\%$$

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

- 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)

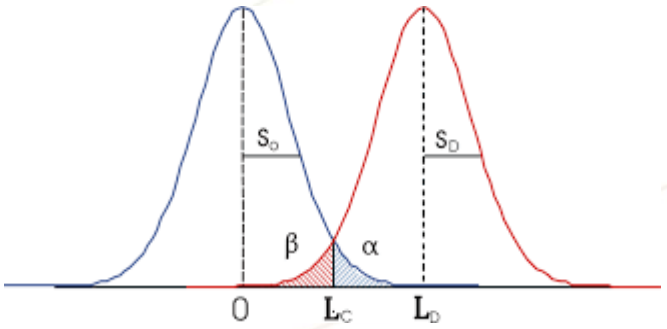
Validation parameters
LOD & LOQ

- 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

- Limit Of Detection (LOD)

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

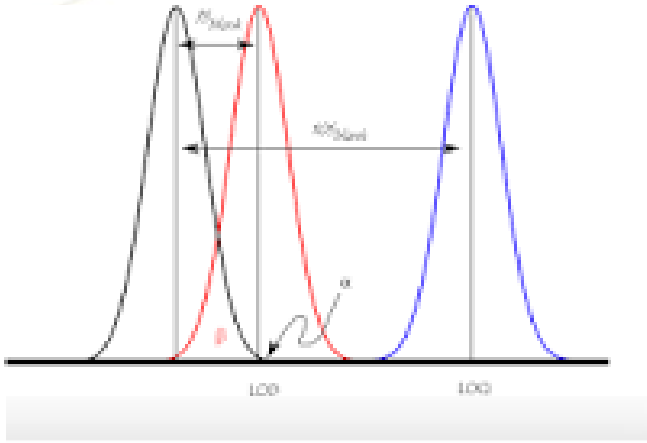


- Limit of Quantification (LOQ)

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

With:

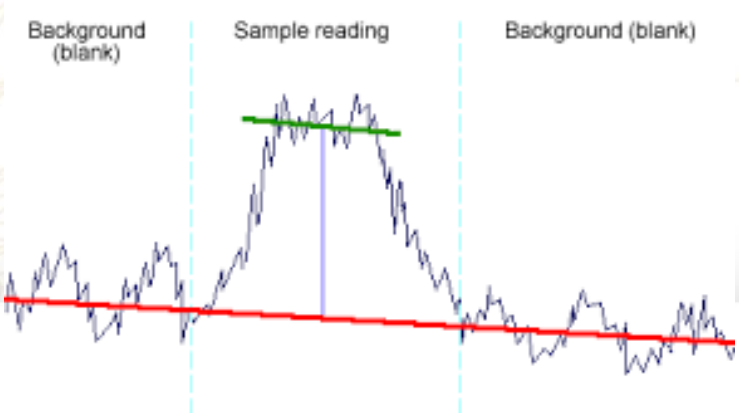
- k_Q = 10 (corresponding to 10% RSD)
- k_Q = 5 or 6 (corresponding to 20/17% RSD)



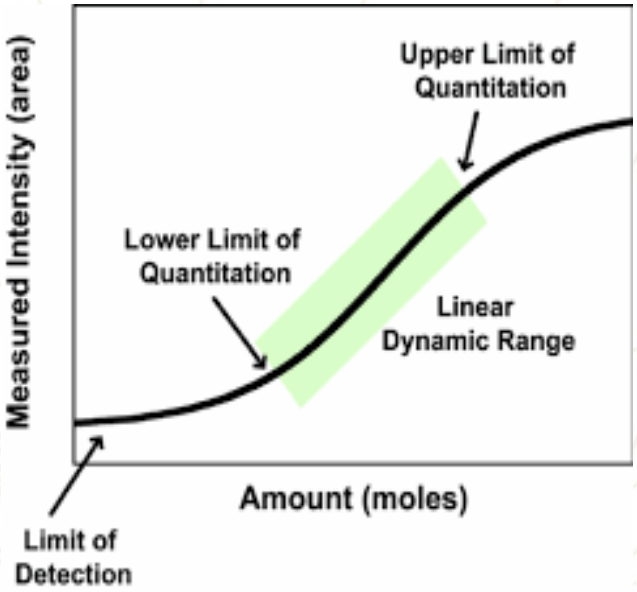
- Note: s'_0 should be in concentration units

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



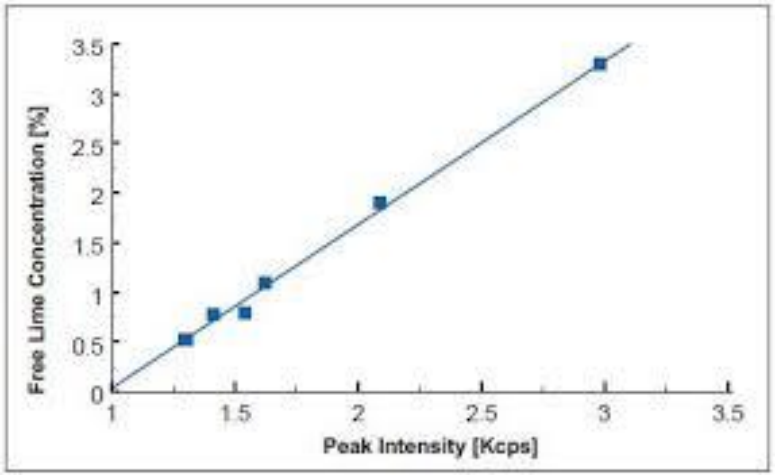
Working range



- 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*

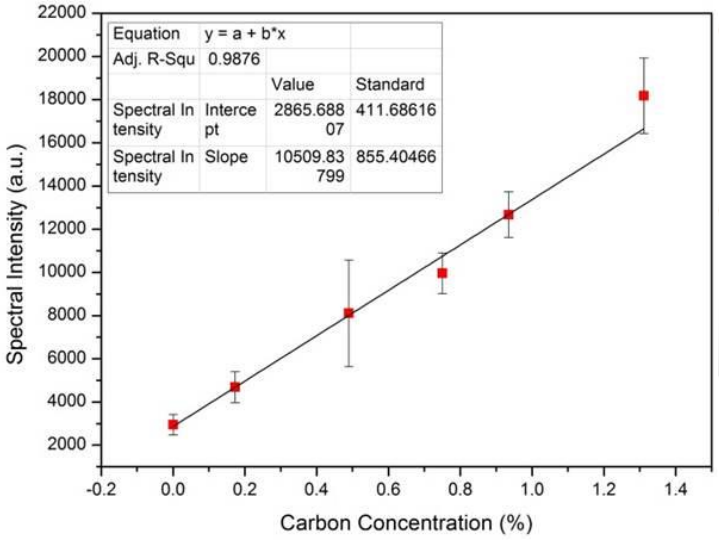
Working range

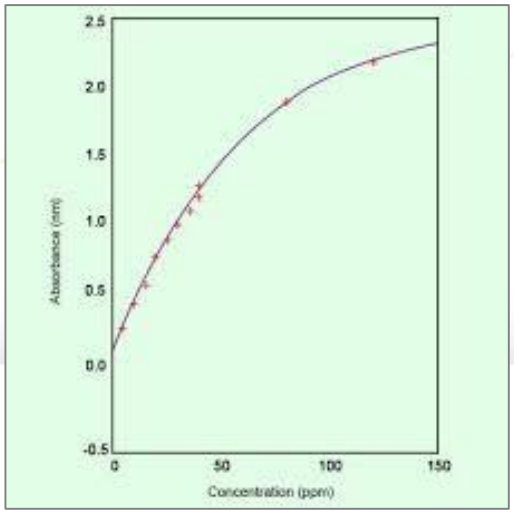
- Instrument working range
  - Limited by Instrument sensitivity and linearity
  - Determine using calibration standards
  - Select suitable calibration procedure
    - External calibration, standard addition, bracketing, etc.
    - Linear or quadratic?



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

- Linearity Evaluation
  - Visually inspect
  - Regression statistics
  - Residuals plot
    - Random distribution about zero - Linear
    - Systematic trends – non-linearity

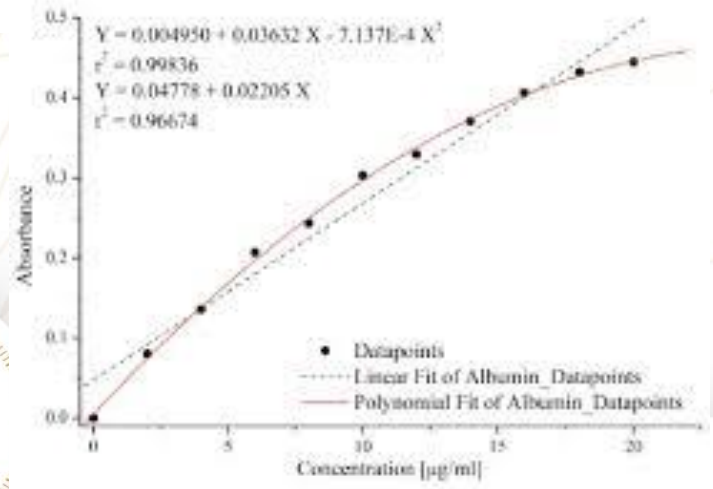
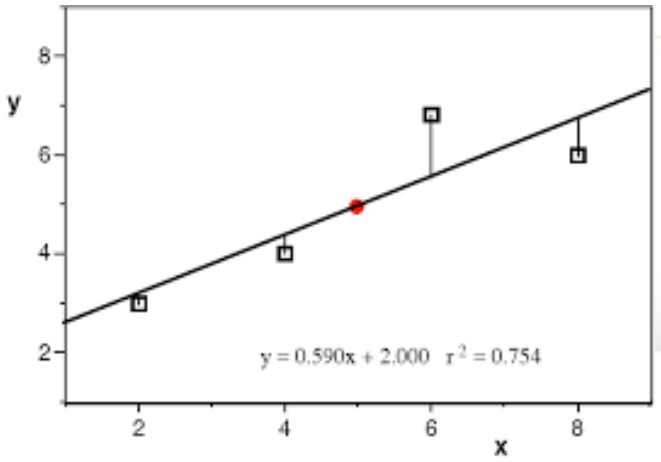
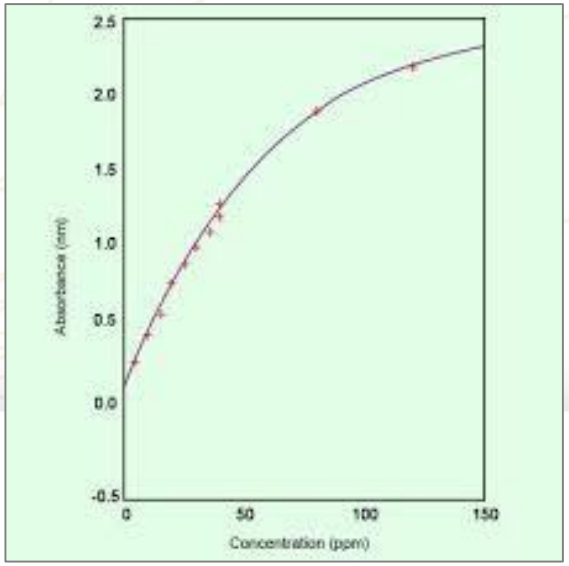


Figure 3 The essence of a linear model optimization for experimental data in Least Squares Method. The model-experiment differences are indicated by solid lines.



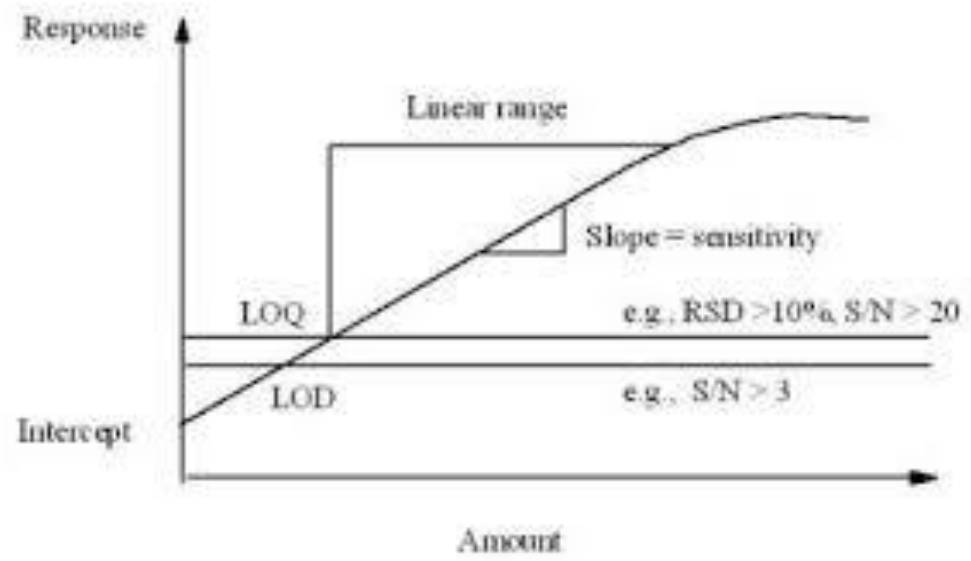
Working Range
Non-Linearity

- Non-linearity corrected for by:
  - Restricted operating range
  - Non-linear calibration
- Remaining deviations from linearity accounted for by overall precision estimates covering several concentrations, or through uncertainties associated with calibration.



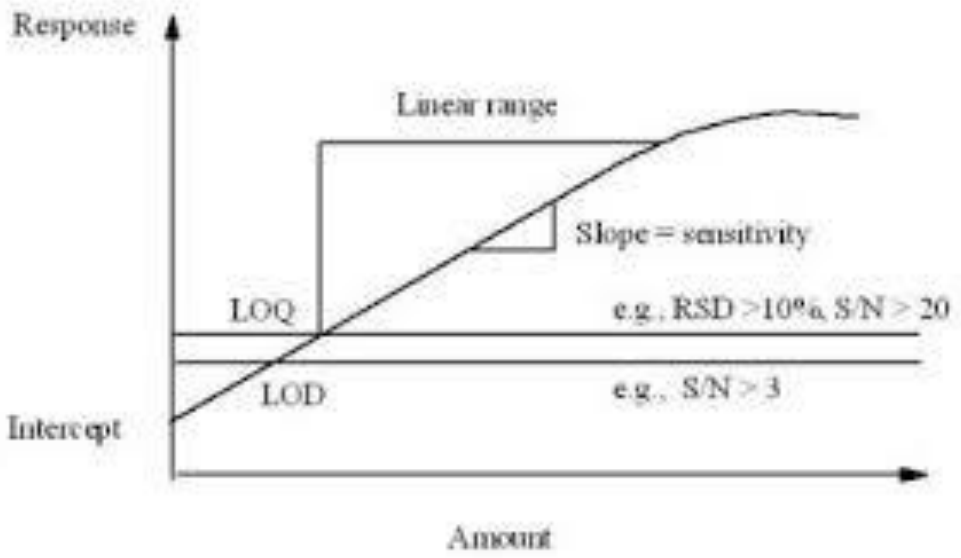
Working range

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



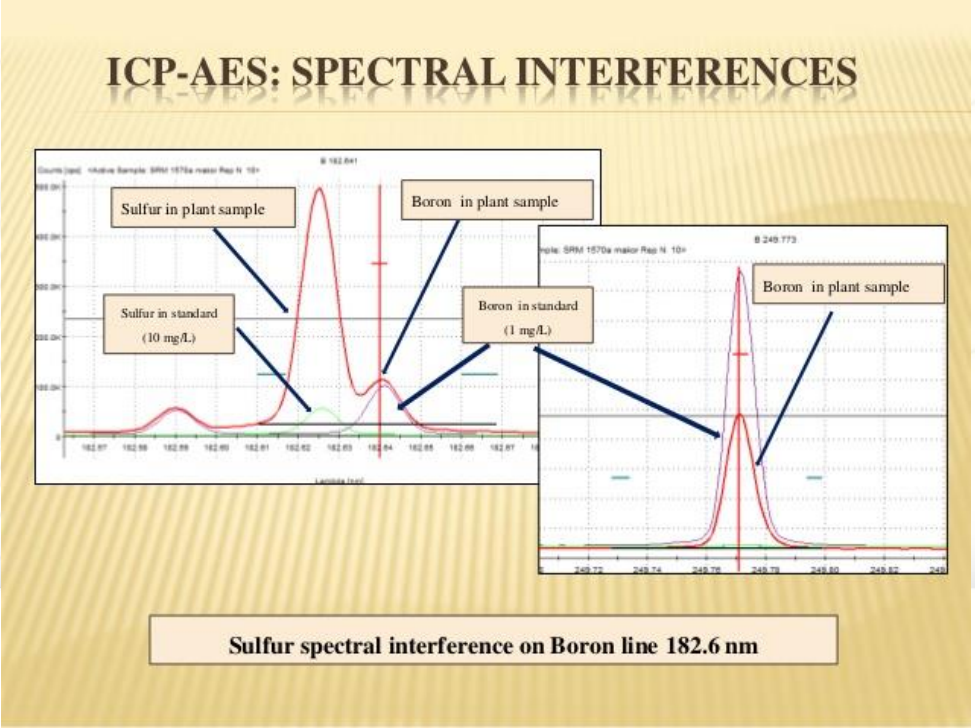
Sensitivity

- 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

- 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)

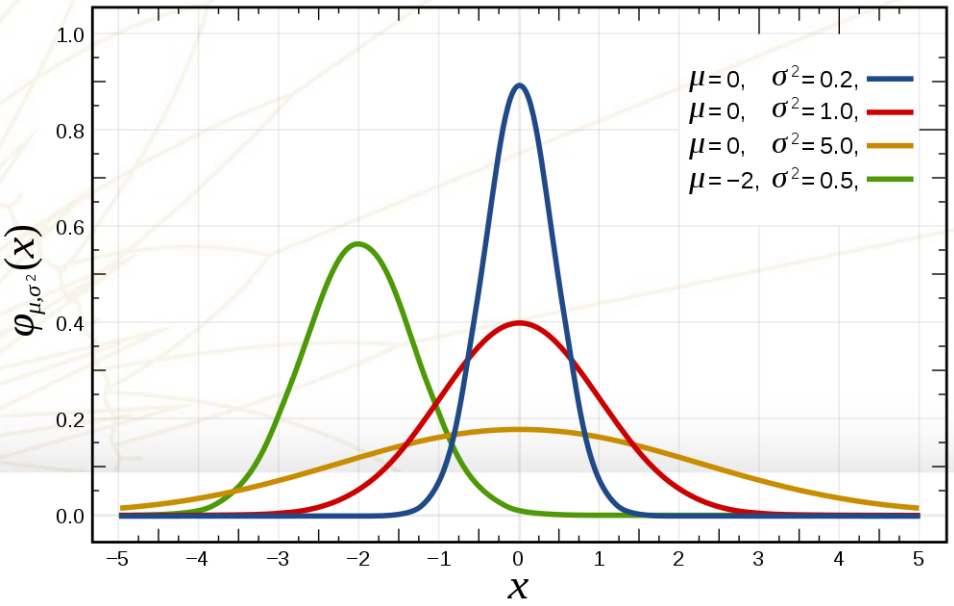
- Measure of a method's capacity to remain unaffected by small, but deliberate variations in method parameters
- Provides indication of method's reliability during normal use
- Required for:
  - In-house methods
  - Method developed from scientific literature
  - Standard methods used outside the method's 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

Uncertainty of measurement (UoM)

Parameter, associated with the result of a measurement, that characterises the spread of values that could reasonably be attributed to the measurand



What is uncertainty of measurement?

- It tells us something about how much you can trust the measurement i.e. the quality of the measurement
- We need two numbers to quantify uncertainty:
  - The width of the margin of doubt, the confidence interval, and
  - The confidence level, how sure we are that the true value is within the margin of doubt.

UoM: Basic concepts

- Standard Uncertainty

$$u(x_i) @ LOC = 68\%$$

- Combined Standard Uncertainty

$$u_c(y) = \sqrt{\sum (c_i \cdot u(x_i))^2}$$

- Expanded Uncertainty

$$U = k \times u_c(y)$$

k = Coverage factor associated with:

- Level of Confidence
- Degrees of Freedom

Evaluation: UoM

- Main contributions:
  - Calibration:
    - Equipment calibration, e.g. balance, volumetic glassware, thermometer
    - Standard Calibration materials
  - Long term precision (intermediate precision / reproducibility)
  - Bias and it's uncertainty (including uncertainty associated with bias measurements)
  - Significant other effects (robustness)

Evaluation: UoM

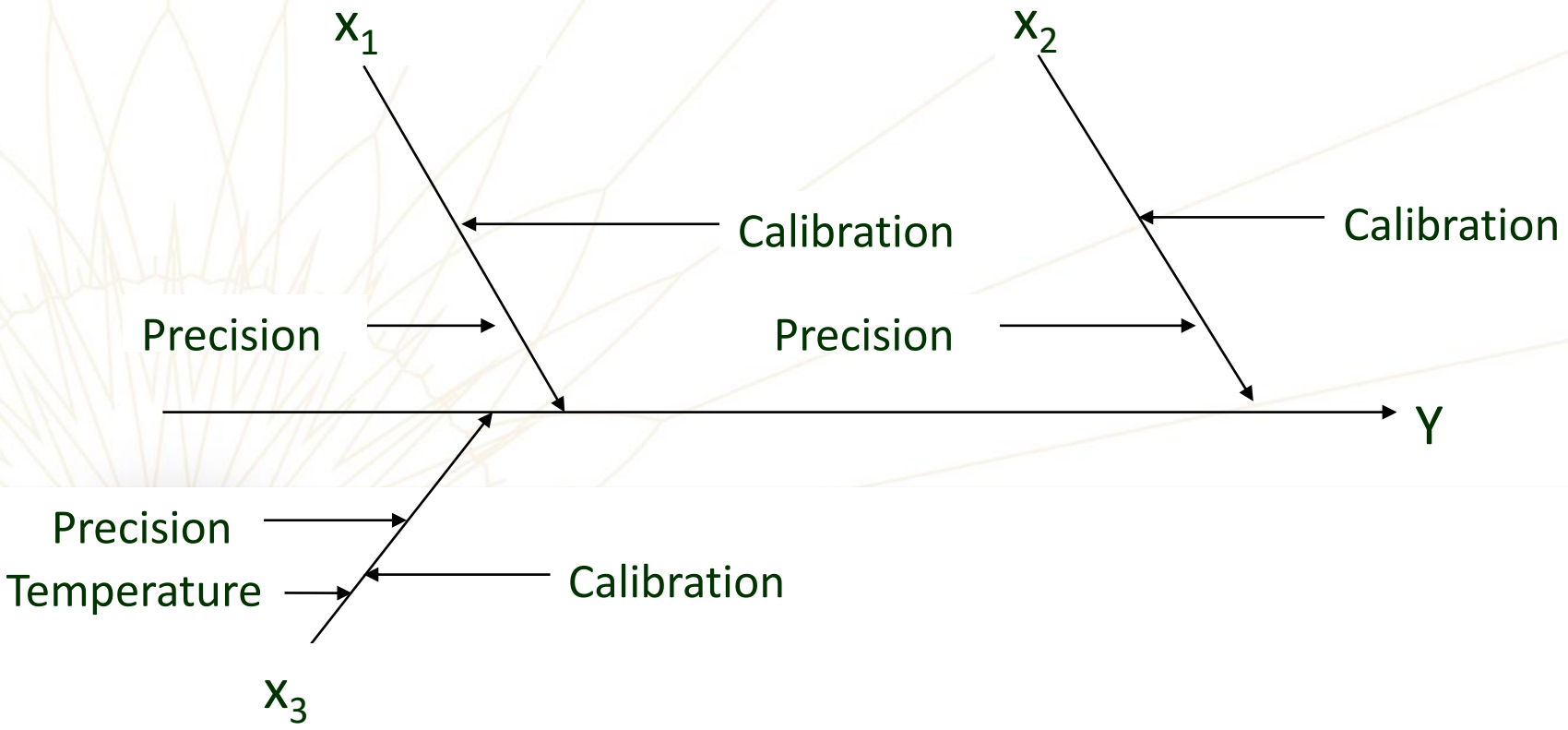
- Several approaches:
  - Bottom - Up:
    - Based on mathematical model describing the complete measurement procedure
    - Critical to identify all parameters during modelling
    - GUM
  - Top - Down:
    - Use of method validation and quality control data to estimate the uncertainty of measurement
    - No knowledge of model required
    - Approaches: EURACHEM / CITAC, Nordtest

Evaluation: UoM

- GUM approach
  - Step 1: Specification and Modelling
    - Identify Measurand, Matrix, Method and Model
    - Model: $y = f(x_1, x_2, .. x_i)$
    - Additional factors:
      - These factors ensure incorporation into uncertainty known effects that contribute to variability that do not occur in the model
      - Factors do not change the value of the result (e.g. typically f=1), but will increase the uncertainty
      - Examples: Digestion, Extraction, Blank correction

Evaluation: UoM

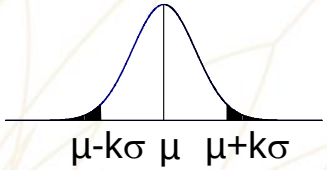
- GUM approach
  - Step 2: Identification of uncertainty sources



Evaluation: UoM

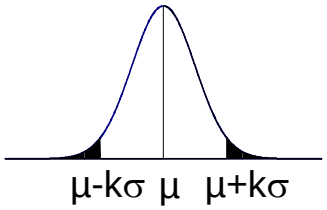
- GUM approach
  - Step 3: Quantification of individual uncertainties
    - Type A:

      Normal distribution:

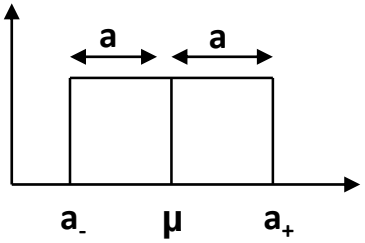


    - Type B:

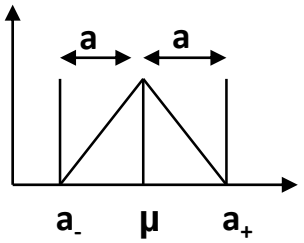
Normal distribution



Rectangular distribution



Triangular distribution



Evaluation: UoM

- GUM approach
  - Step 4: Combined standard uncertainty
    - Calculate sensitivity coefficients

$$c_i = \frac{\partial f}{\partial x_i}$$

    - Convert standard uncertainty to uncertainty contribution

$$u(y_i) = c_i \cdot u(x_i)$$

    - Calculate combined standard uncertainty

$$u_c(y) = \sqrt{\sum (c_1 \cdot u_1)^2 + (c_2 \cdot u_2)^2 + \ldots + (c_i \cdot u_i)^2}$$

Evaluation: UoM

- GUM approach
  - Step 5: Expanded uncertainty

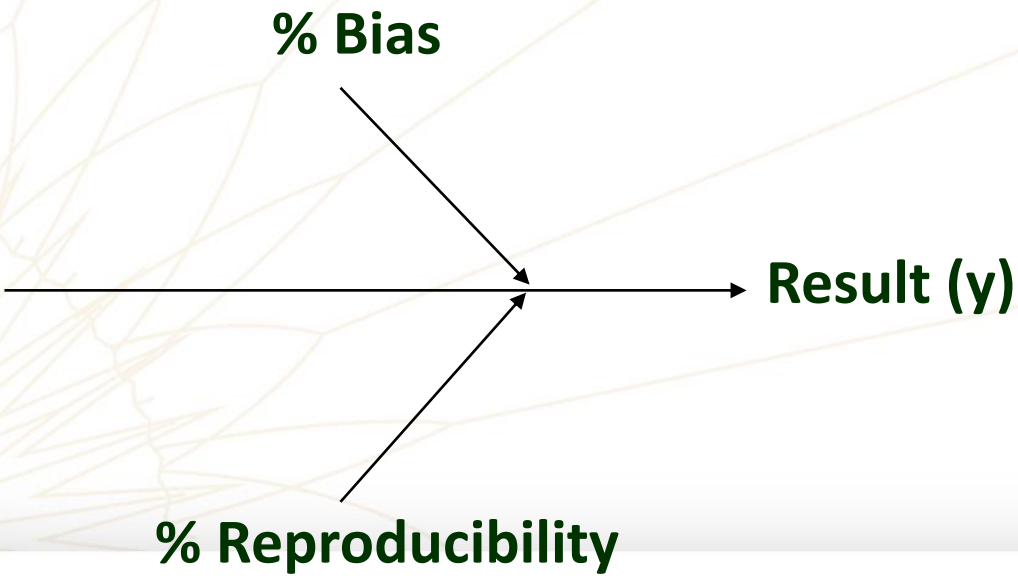
$$U = k \times u_c(y)$$

    - k = coverage factor chosen from the t-distribution table, depending on:
      - the desired level of confidence
      - the effective degrees of freedom

Evaluation: UoM

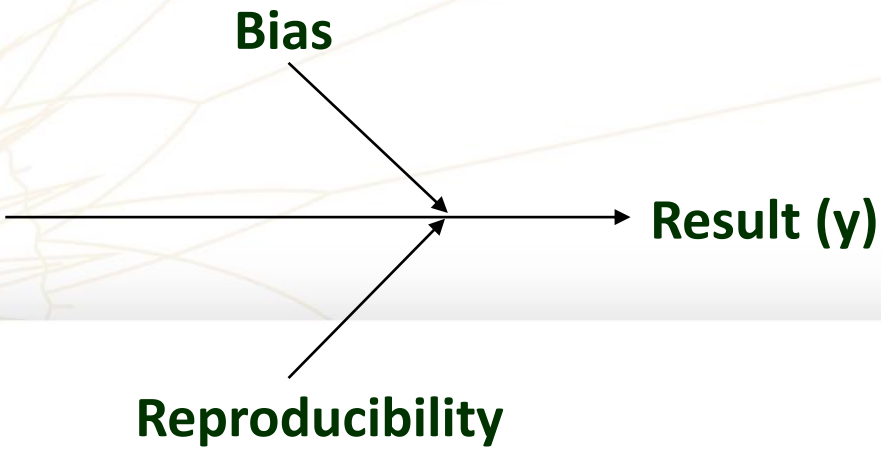
- Top Down:
  - % Reproducibility
  - % Method and Laboratory Bias

$$\%u = \frac{u(x_i)}{x_i} \times 100$$



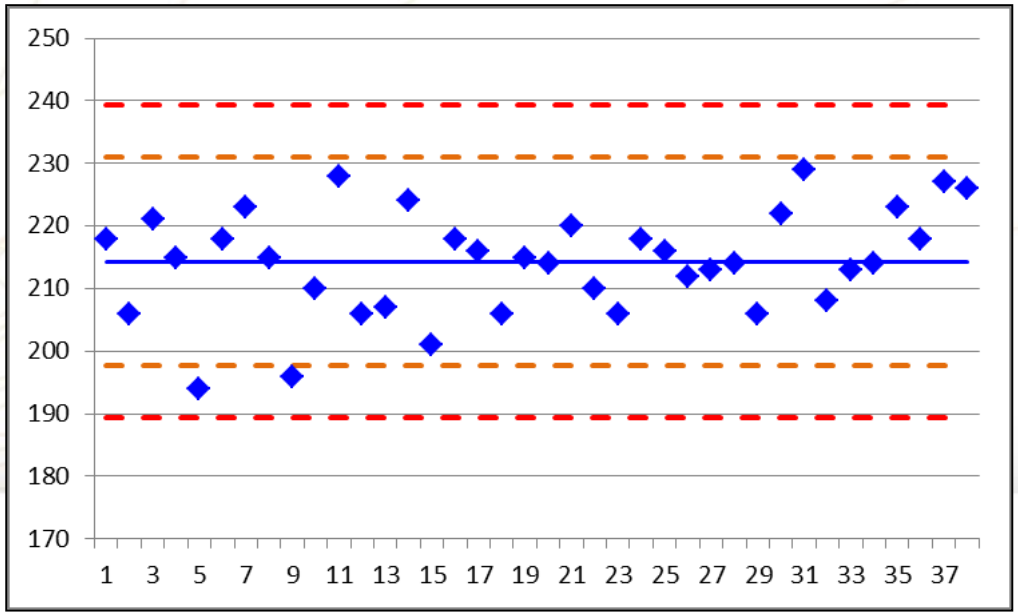
Evaluation: UoM

- Provided that reproducibility and bias data is representative:
  - Different stock standard solutions
  - Different batches of reagents
  - Re-calibration of instruments
  - Representative period of time – ideally 1 year
  - Minimum number of results: 50



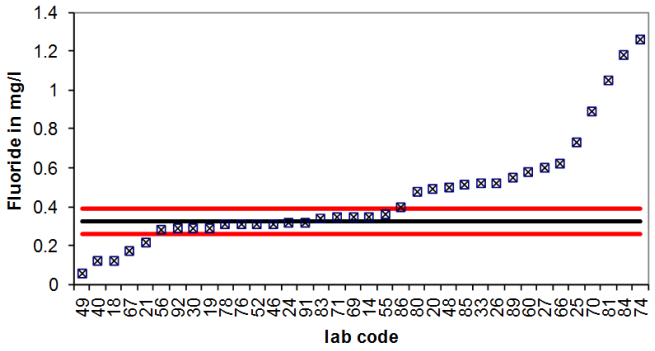
Evaluation: UoM

- Top Down:
  - Within-laboratory Reproducibility
    - Mean control chart
    - Sample duplicates analysed



Evaluation: UoM

- Top Down:
  - Method and Laboratory Bias
    - Combination of:
      - Bias, *% Bias*
      - Uncertainty associated with the reference value, *% u_c(RefVal)*
  - Experimentally:
    - Certified Reference Materials
    - Interlaboratory comparisons / Proficiency testing
    - Recovery experiments



Evaluation: UoM

- Top Down:
  - Calculate Combined Standard Uncertainty (u_c):
    - Combine Reproducibility and Bias components
      - Reproducibility (R_w): From control samples and other estimations
      - Bias (u_{bias}): From CRM, PT or recovery tests

$$\%u_c = \sqrt{\%u(R_w)^2 + (\%u_{bias})^2}$$

Evaluation: UoM

- Top Down:
  - Determine the expanded uncertainty (U):

$$U = k \times u_c(y)$$

    - Assume k=2 for an approximate level of confidence of 95% with assumed effective degrees of freedom > 30.

Which approach to use?

- Bottom Up (GUM)
  - Mathematical model needed
  - Complex calculations
  - Smaller uncertainties
- Top Down (Nordtest, Eurachem/CITAC)
  - No model needed
  - Simpler combination of data already available in accredited laboratory
  - Uncertainties are larger, but perhaps more realistic?
- Fit for purpose?

Metrological Traceability

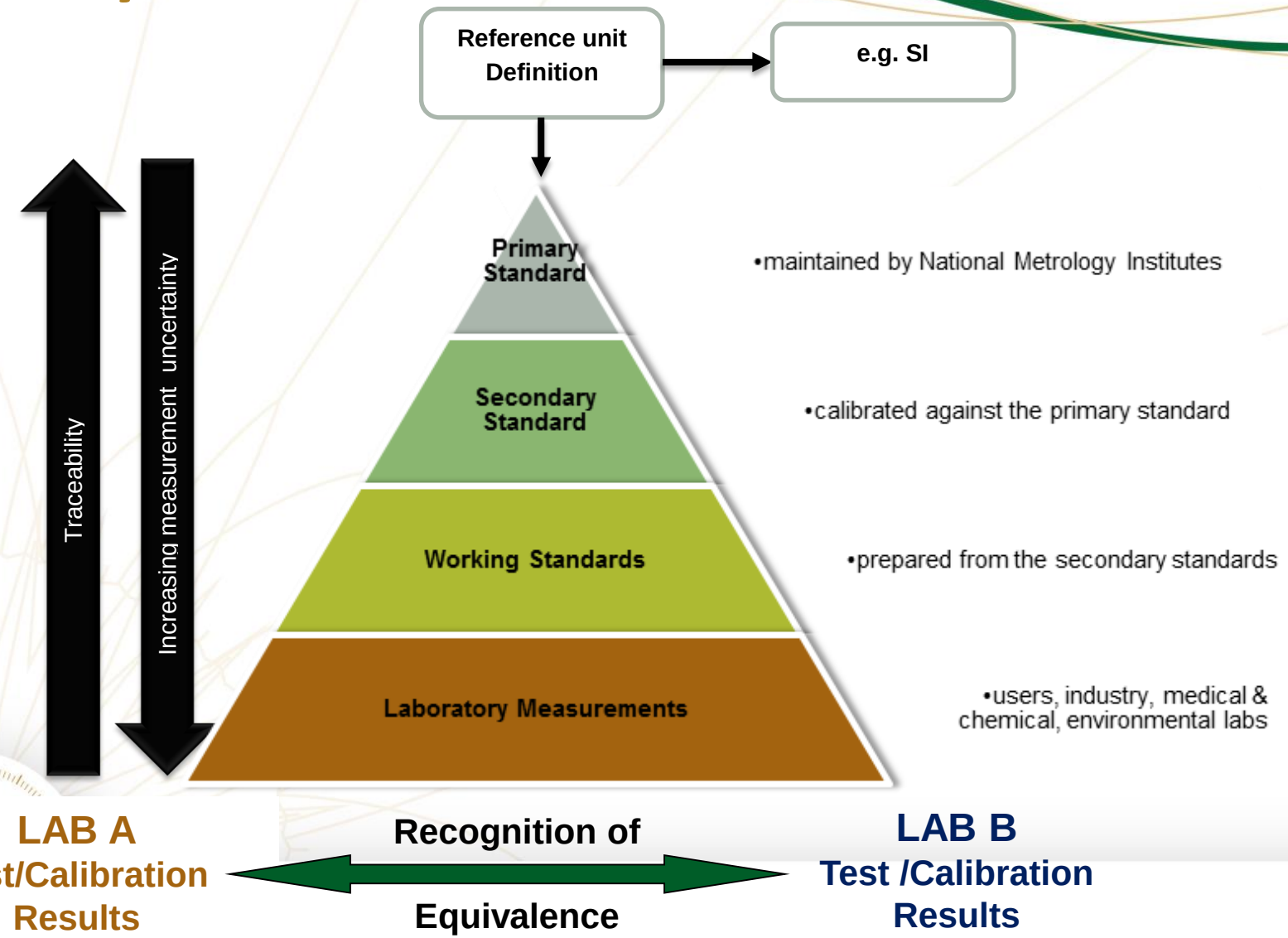
- Good analytical results are essential to ensure reliable decisions
  - Comparability through traceability to consistent and agreed set of measurement units and scales, i.e. SI

Metrological traceability

Result can be related to a reference through a documented unbroken chain of calibrations, each contributing to the measurement uncertainty

OR

Establishing measurement traceability to the reference unit



Metrological Traceability

- Specification of the measurand
- Documented, unbroken chain of calibrations
  - Traceable to appropriate references
- For every step in the traceability chain:
  - Performed according to appropriate method
  - Measurement uncertainty determined according to agreed methods
  - Measurements by technically competent laboratories

Establishing Metrological Traceability

- Method validation to establish optimised procedure, including:
  - Mathematical model / calculations
  - Set of measurement conditions
- Establish traceability through calibration for:
  - Each parameter in mathematical model
  - Each of the specified conditions
    - NOTE: Essential for critical values in measurement, not so for less critical values

Establishing Metrological Traceability

- Options to establish traceability:
  - Physical measurements, e.g. mass, volume, temperature
    - Uncertainties are typically not significant compared to those in analytical measurements
  - Confirmation of Identity against:
    - Certified pure material
    - Authentic samples from reputable source
    - Reference data, e.g. reference wavelength spectra

Establishing Metrological Traceability

- Calibration options to establish traceability:
  - Certified refence material
    - Demonstrable traceability to national or international standards
    - Statement of uncertainty
    - Matrix CRMs not recommended for calibration
      - High cost
      - Sufficiently good matrix matching is rare
      - Large uncertainties

Establishing Metrological Traceability

- Calibration options to establish traceability:
  - Certified refence material (cont)
    - Matrix CRMs not recommended for calibration
      - High cost
      - Sufficiently good matrix matching is rare
      - Large uncertainties
  - Pure Materials
    - Purity established through preparation procedure, impurity measurements, incorporating data such as density

Establishing Metrological Traceability

- Calibration options to establish traceability:
  - Other reference materials, e.g. multi-element standards, alloys, etc.
  - Reference data, e.g. reference spectroscopic data to calibrate wavelength scales

Metrological Traceability

- Selection of CRM:
  - Should match:
    - Measurand
    - Concentration range
    - Matrix match with potential interferences
    - Sample size
    - Homogeneity and stability
    - Measurement uncertainty

Metrological Traceability

- Confirming metrological traceability of CRM:
  - Ideally, accreditation of manufacturer to ISO 17034 and ISO 17025
  - Information stated on Certificate:
    - Specification of measurand
    - Measurement unit
    - Characterisation methods
      - Measurement methods
      - Certification approach
    - Specifications for sample handling
    - Measurement uncertainty

Traceability in Analytical Chemistry

- 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)

Method validation Plan

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

Method validation Report

- Introduction
- Method & Scope
- Acceptance criteria / specific requirements
- Procedure
  - Including samples / materials analysed
- Performance characteristics
  - Results
- Conclusion:
  - Statement of validity

Quality control

- Measures to ensure that a validated method remains “in control”
- External quality control
  - Proficiency testing
    - Reproducibility & Bias checks
- Internal quality control
  - Analysis of Quality Control samples
    - Frequency of analysis depend on nature, criticality, batch size, frequency with which method is employed and complexity of the method
      - Typically 5%
      - Lower for high sample throughput
      - 20-50% possible for complex procedures or non-routine analysis

Internal Quality Control

- Suitable Quality Control material:
  - Stable, homogenous QC sample
    - Should be representative:
      - Matrix
      - Concentrations
      - Sufficient quantities available
      - Long term stability (if possible)
      - Homogeneous
  - Replicate analysis of routine test samples
  - Blanks
    - Standard blank
    - Process blank
  - Standard solutions / appropriate calibration material
  - Spiked samples
  - Blind samples

Internal Quality Control

- Quality Control Charts
  - Statistical process control (SPC) charts are simple graphical tools that enable process performance monitoring
    - Very powerful tool for internal quality control
    - Changes in quality can be detected quickly
  - Types:
    - Mean / X-control chart
    - Range Control chart

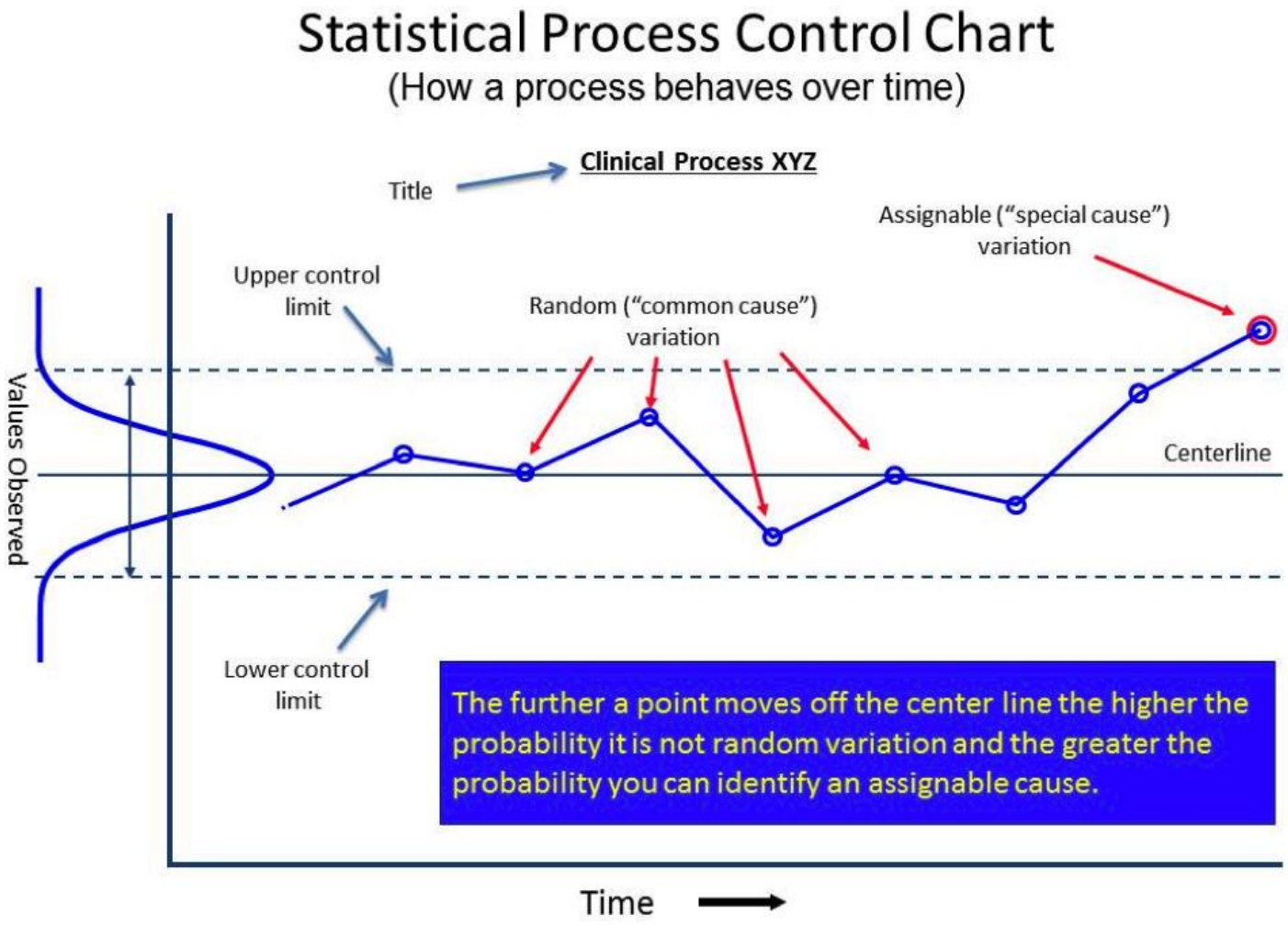
Control Charts

- General concepts:
  - Displays results vs. time
    - Target value
    - Limits
    - Based on routine analysis, i.e. typical intermediate precision
      - Repeatability: Too narrow limits
      - Reproducibility: Too wide limits

Control Charts

- General concepts:
  - Limits
    - Warning limit
      - Exceeding once allowed
      - Represents 5% limit, i.e. 5% of correct results can be expected to exceed this limit
    - Control limit
      - Stop immediately if result exceeds this limit
      - Represents 0.3% limit, i.e. only 0.3% of correct results can be expected to exceed this limit – very unlikely

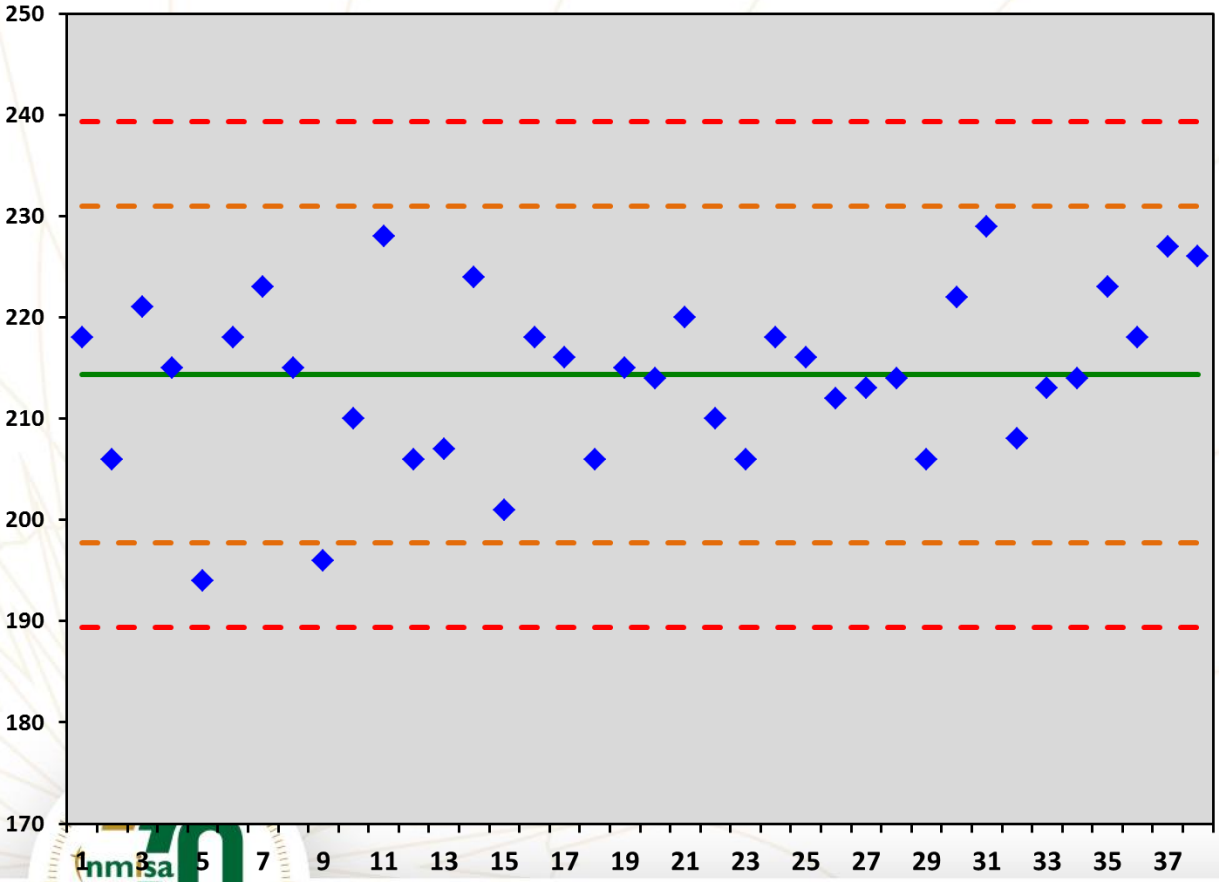
Control Charts



Control Charts

- Mean / X-control chart
  - QC sample
    - Intermediate precision
    - Changes in systematic error
    - Trueness (if CRM is used)
  - Blank
    - Reagents
    - Potential environmental contamination
  - % Recovery
    - Changes in systematic error

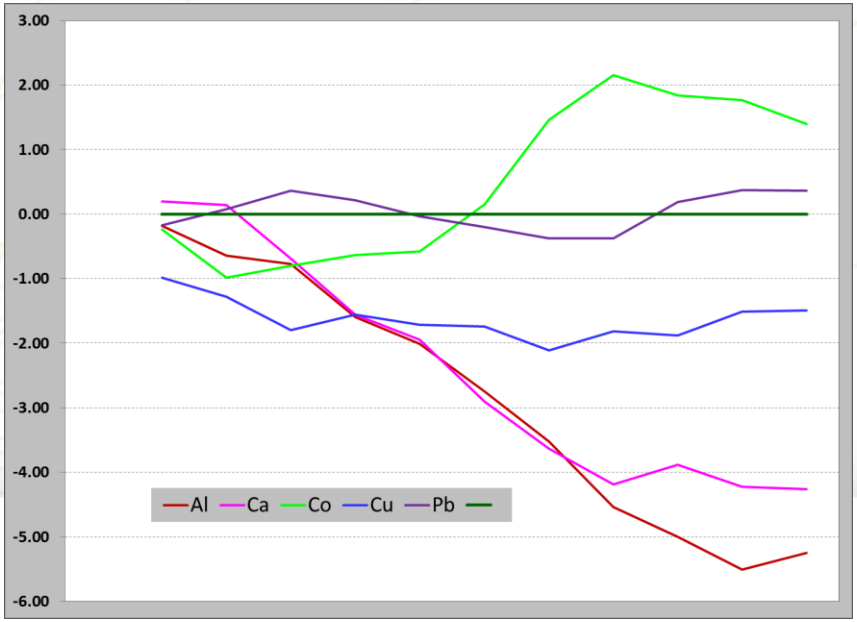
Quality Control:
Mean Control Chart



- **Action limit:**
  Mean ± 3SD
- **Warning limit:**
  Mean ± 2SD

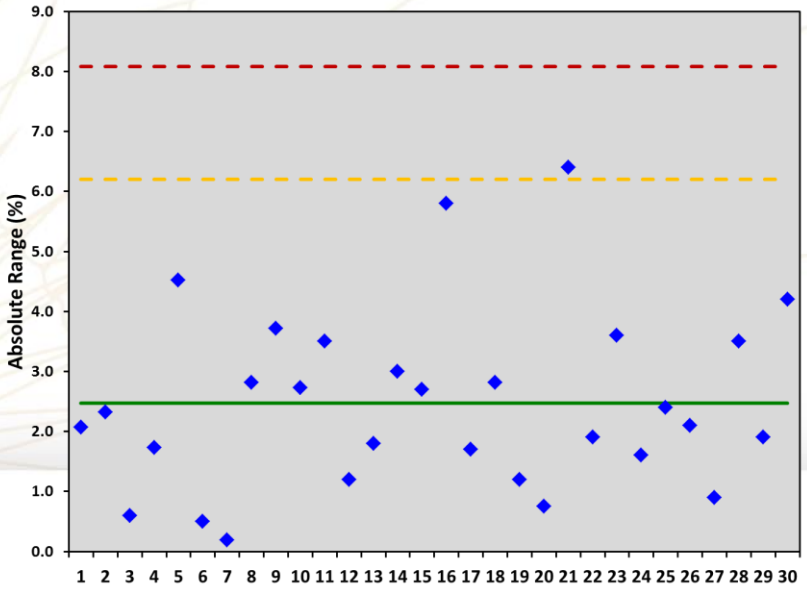
Control Charts

- CUSUM chart
  - Cumulative sum of all errors from one target value
    - Faster detection of change in process
    - Can identify point at which process went out of control



Control Charts

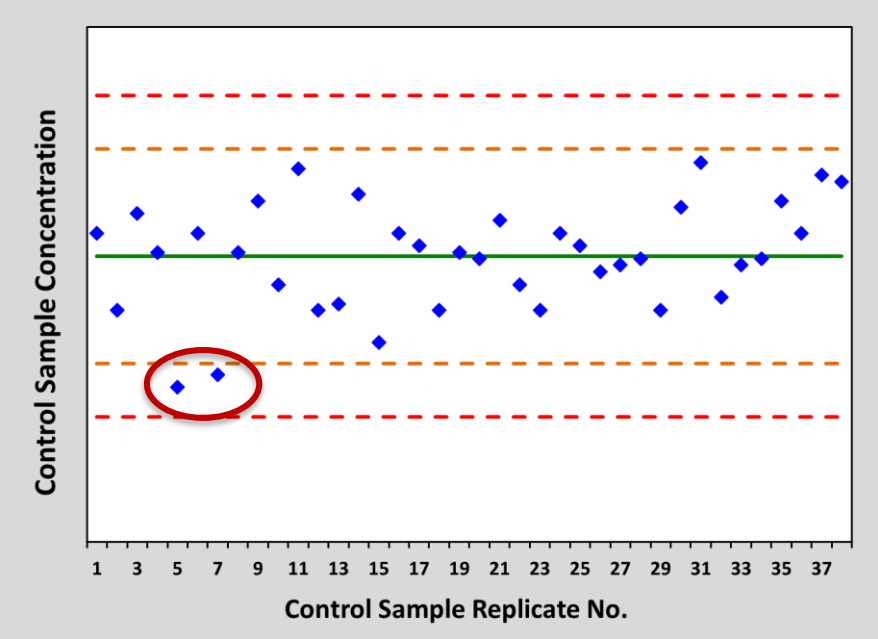
- Range control chart
  - Replicate analysis of routine test samples
    - Difference between highest and lowest value
    - Only upper control limits
    - Check repeatability

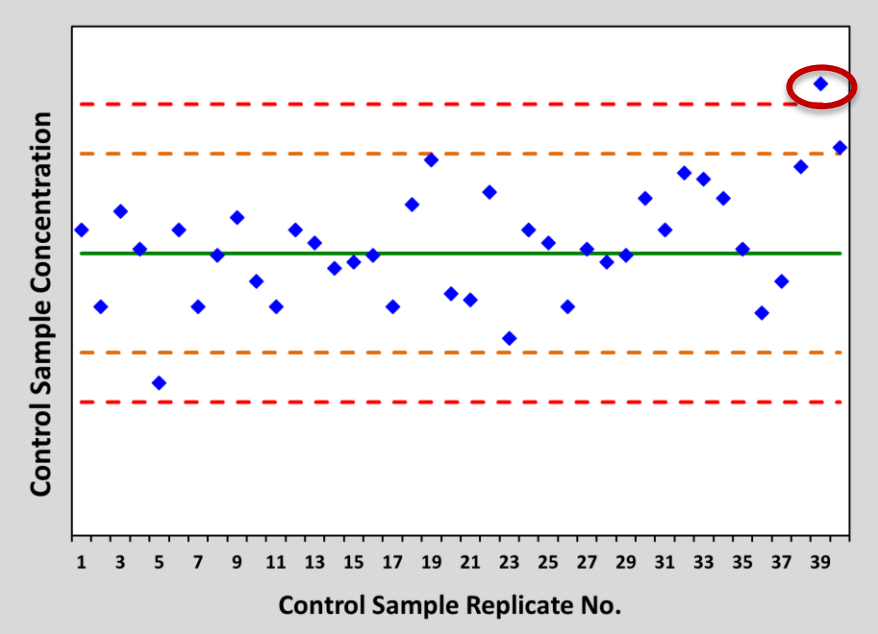


- **Action limit:**
  Mean ± 3.69SD
- **Warning limit:**
  Mean ± 2.83SD

Interpretation of Control Charts

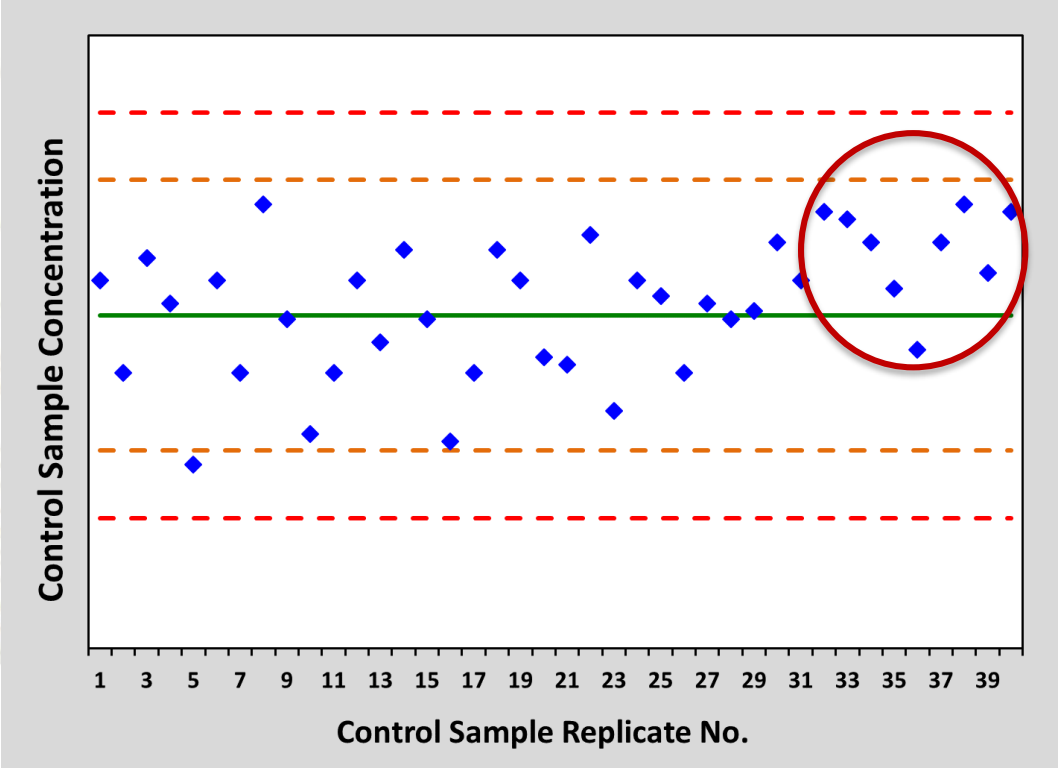
- Indication of “out-of-control” analytical procedure
  - Control limits
    - Warning: 2 out of 3 consecutive values outside limits
    - Action limit: 1 value outside limits





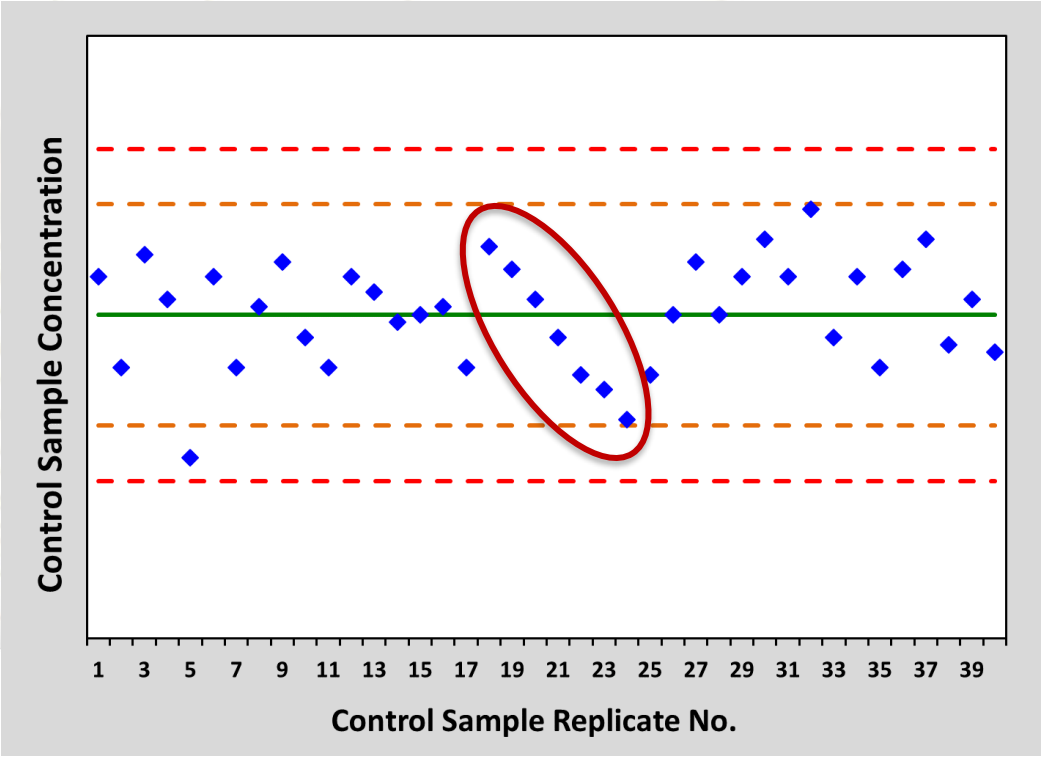
Interpretation of Control Charts

- Indication of “out-of-control” analytical procedure
  - Shift
    - 10 out of 11 consecutive values above or below mean



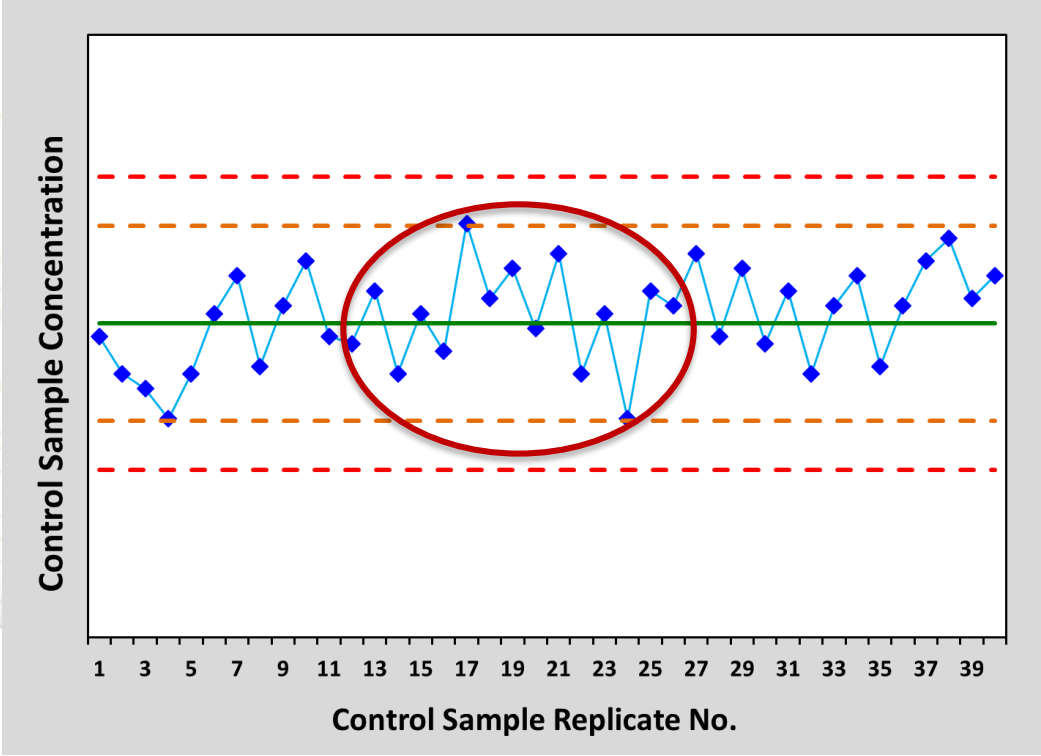
Interpretation of Control Charts

- Indication of “out-of-control” analytical procedure
  - Trend
    - 7 consecutive values either increasing or decreasing



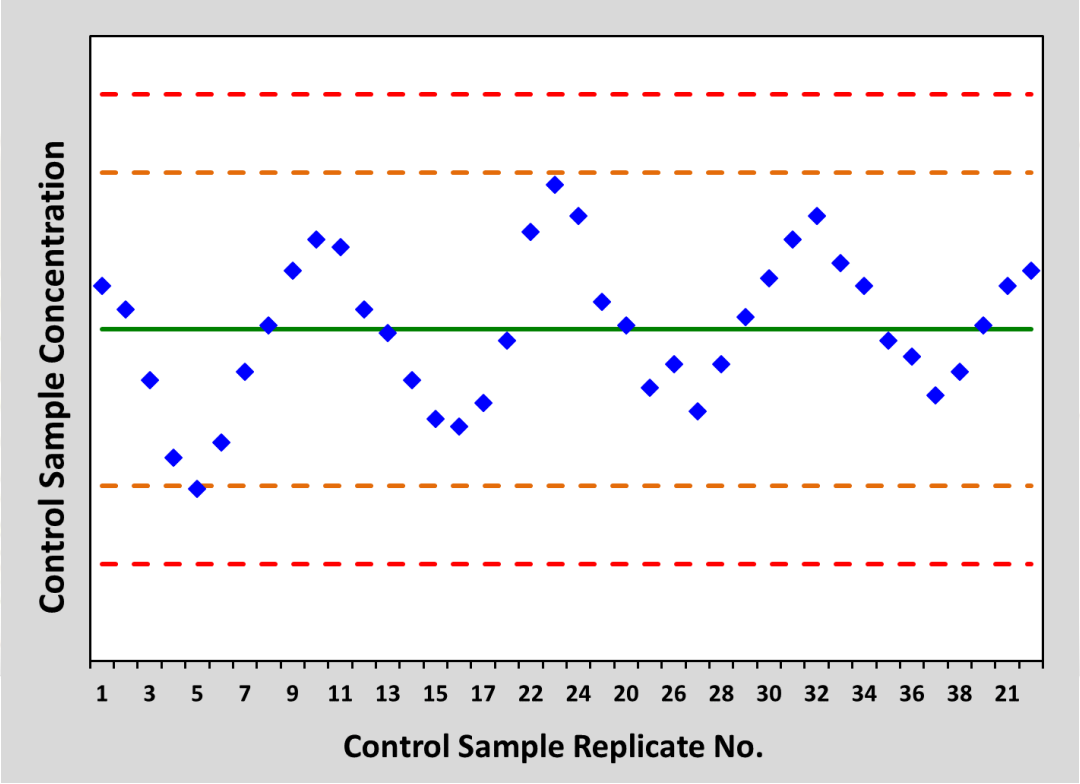
Interpretation of Control Charts

- Indication of “out-of-control” analytical procedure
  - Zig-Zag
    - 14 or more consecutive values increasing and decreasing alternatively



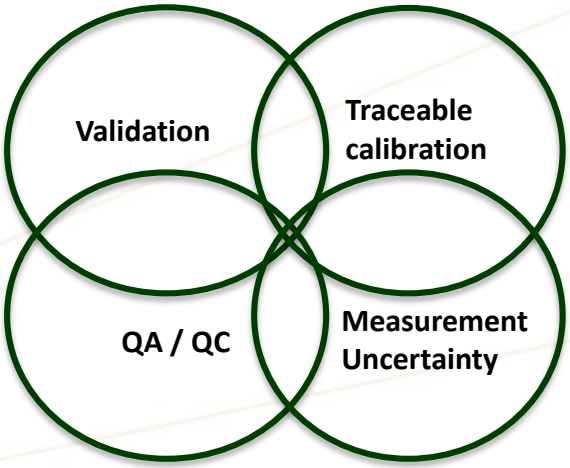
Interpretation of Control Charts

- Indication of “out-of-control” analytical procedure
  - Cyclical pattern
    - Pattern observed over time



Conclusions

- Method Validation
  - One part of ensuring valid, traceable results
  - Fit-for-purpose
    - Performance characteristics
    - Acceptance criteria
- Quality control
  - Internal processes
  - External processes



References

- Eurachem / CITAC Guides:
  - The Fitness for Purpose of Analytical Methods, 2007
  - Traceability in Chemical Measurement, 2003
  - Quantifying Uncertainty in Analytical Measurement
  - The Selection and Use of Reference Materials
- Statistics and Chemometrics for Analytical Chemistry, JN Miller, JC Miller
- Validation of Analytical Methods, Ludwig Huber, Agilent
- GUM: Guide to the Expression of Uncertainty in Measurement
- Nordtest: Handbook for calculation of measurement uncertainty in environmental laboratories (NT TR 537)