Eurachem

A FOCUS FOR ANALYTICAL CHEMISTRY IN EUROPE

LGC

Demonstrating the fitness for purpose of analytical methods:

A practical guide for laboratories

Vicki Barwick

Chair, Eurachem Education and Training Working Group

Member, Eurachem Method Validation Working Group

BMTA/CSols Seminar

20 October 2015

Overview

- Introduction to Eurachem
- Eurachem activities
- Eurachem ‘method validation guide’
- Statistics & method validation
  - Precision
  - Bias
  - Capability of detection
  - Ruggedness

What is Eurachem?

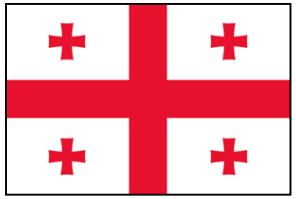
- A network of national and other organisations
- A focus on analytical quality
  - Method validation
  - Measurement uncertainty
  - Traceability of measurement results
- Providing authoritative guidance documents
- Organising workshops and training events
- Primary audience:
  - Laboratories for analytical measurement
  - Accreditation bodies and related organisations
- www.eurachem.org

LGC

Eurachem membership – 32 member countries

Members not shown on map

Cyprus



Georgia

Workshops organised since year 2000

- Validation, Traceability and Measurement Uncertainty (2000, 2012)
- Education & Training (2004)
- Proficiency Testing (2000, 2003, 2005, 2008, 2011, 2014)
- Measurement Uncertainty (2000, 2002, 2008, 2011)
- Reference Materials (2000)
- Sampling (2001, 2008)
- Metrology and Quality Assurance (2008, 2009, 2010, 2014)
- Decision making (2008, 2010)
- Internal Quality Control (2012)
- QA of measurements from Field to Laboratory (2013)

...as well as Quality Assurance (QA) and training events in conjunction with General Assembly meetings and in collaboration with other organisations

LGC

Guidance documents – www.eurachem.org

- Setting and using target uncertainty in chemical measurement (2015)
- The fitness for purpose of analytical methods: A laboratory guide to method validation and related topics, 2nd ed. (2014)
- Accreditation for microbiological laboratories, 2nd ed. (2013)

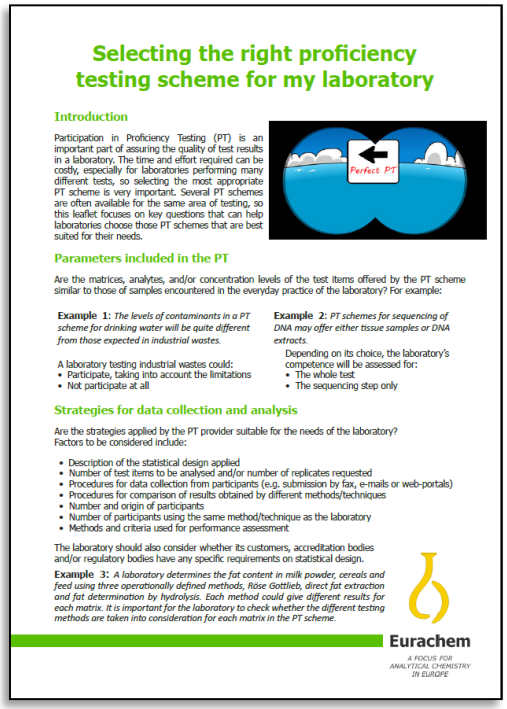
- Quantifying uncertainty in analytical measurement, 3rd ed. (2012)
- Selection, use and interpretation of proficiency testing (PT) schemes, 2nd ed. (2011)

Guidance documents – www.eurachem.org

- Terminology in analytical measurement – Introduction to VIM3 (2011)
- Measurement uncertainty arising from sampling – A guide to methods and approaches (2007)
- Use of uncertainty information in compliance assessment (2007)
- Traceability in Chemical Measurement – A guide to achieving comparable results in chemical measurement (2003)
- Guide to quality in analytical chemistry – An aid to accreditation (2002, *under revision)*
- Quality Assurance for Research and Development and Non-routine Analysis (1998, *under revision)*

Eurachem information leaflets

- Short briefing documents on specific topic, intended to inform a wide audience
  - Laboratory staff, managers, laboratory customers



Selecting the right proficiency testing scheme for my laboratory

Introduction

Participation in Proficiency Testing (PT) is an important part of assuring the quality of test results in a laboratory. The time and effort required can be costly, especially for laboratories performing many different tests, so selecting the most appropriate PT scheme is very important. Several PT schemes are often available for the same area of testing, so this leaflet focuses on key questions that can help laboratories choose those PT schemes that are best suited for their needs.

Parameters included in the PT

Are the matrices, analytes, and/or concentration levels of the test items offered by the PT scheme similar to those of samples encountered in the everyday practice of the laboratory? For example:

Example 1: *The levels of contaminants in a PT scheme for drinking water will be quite different from those expected in industrial wastes.*

A laboratory testing industrial wastes could:
- Participate, taking into account the limitations
- Not participate at all

Example 2: *PT schemes for sequencing of DNA may offer either tissue samples or DNA extracts.*

Depending on its choice, the laboratory's competence will be assessed for:
- The whole test
- The sequencing step only

Strategies for data collection and analysis

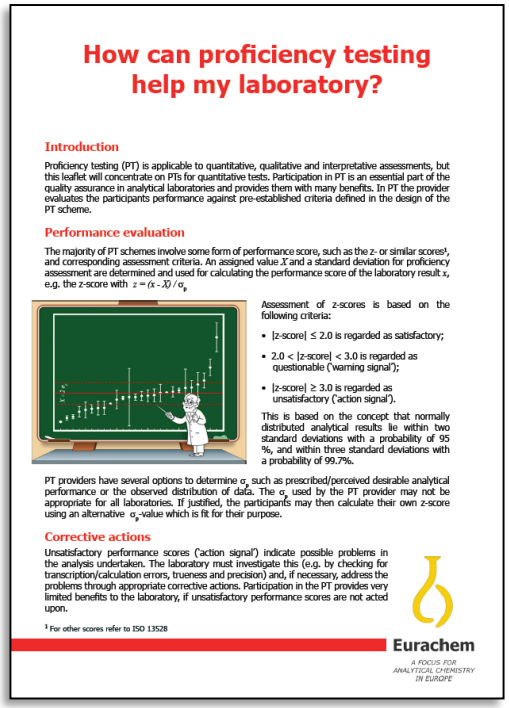
Are the strategies applied by the PT provider suitable for the needs of the laboratory? Factors to be considered include:

- Description of the statistical design applied
- Number of test items to be analysed and/or number of replicates requested
- Procedures for data collection from participants (e.g. submission by fax, e-mails or web-portals)
- Procedures for comparison of results obtained by different methods/techniques
- Number and origin of participants
- Number of participants using the same method/technique as the laboratory
- Methods and criteria used for performance assessment

The laboratory should also consider whether its customers, accreditation bodies and/or regulatory bodies have any specific requirements on statistical design.

Example 3: *A laboratory determines the fat content in milk powder, cereals and feed using three operationally defined methods, Röse Gottlieb, direct fat extraction and fat determination by hydrolysis. Each method could give different results for each matrix. It is important for the laboratory to check whether the different testing methods are taken into consideration for each matrix in the PT scheme.*

Eurachem
A FOCUS FOR ANALYTICAL CHEMISTRY IN EUROPE



How can proficiency testing help my laboratory?

Introduction

Proficiency testing (PT) is applicable to quantitative, qualitative and interpretative assessments, but this leaflet will concentrate on PTs for quantitative tests. Participation in PT is an essential part of the quality assurance in analytical laboratories and provides them with many benefits. In PT the provider evaluates the participants performance against pre-established criteria defined in the design of the PT scheme.

Performance evaluation

The majority of PT schemes involve some form of performance score, such as the z- or similar scores[1], and corresponding assessment criteria. An assigned value X and a standard deviation for proficiency assessment are determined and used for calculating the performance score of the laboratory result x, e.g. the z-score with $z = (x - X) / \sigma_p$

Assessment of z-scores is based on the following criteria:

- $|z\text{-score}| \leq 2.0$ is regarded as satisfactory;
- $2.0 < |z\text{-score}| < 3.0$ is regarded as questionable ('warning signal');
- $|z\text{-score}| \geq 3.0$ is regarded as unsatisfactory ('action signal').

This is based on the concept that normally distributed analytical results lie within two standard deviations with a probability of 95 %, and within three standard deviations with a probability of 99.7%.

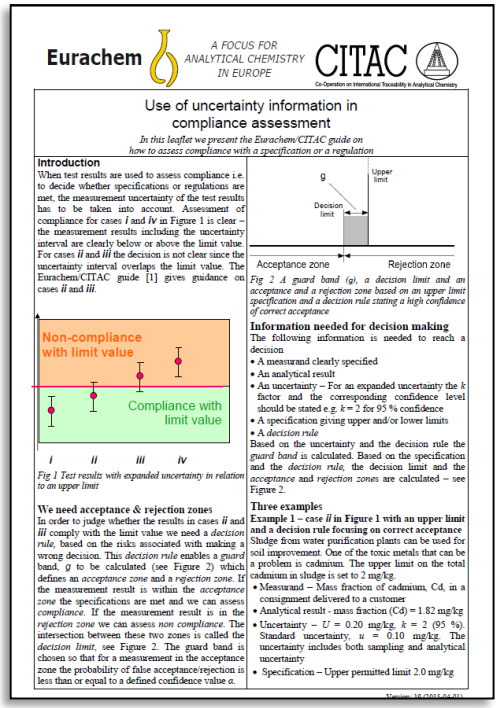
PT providers have several options to determine σ_p such as prescribed/perceived desirable analytical performance or the observed distribution of data. The σ_p used by the PT provider may not be appropriate for all laboratories. If justified, the participants may then calculate their own z-score using an alternative σ_p-value which is fit for their purpose.

Corrective actions

Unsatisfactory performance scores ('action signal') indicate possible problems in the analysis undertaken. The laboratory must investigate this (e.g. by checking for transcription/calculation errors, trueness and precision) and, if necessary, address the problems through appropriate corrective actions. Participation in the PT provides very limited benefits to the laboratory, if unsatisfactory performance scores are not acted upon.

[1] For other scores refer to ISO 13528

Eurachem
A FOCUS FOR ANALYTICAL CHEMISTRY IN EUROPE



Eurachem A FOCUS FOR ANALYTICAL CHEMISTRY IN EUROPE CITAC Co-Operation on International Traceability in Analytical Chemistry

Use of uncertainty information in compliance assessment

In this leaflet we present the Eurachem/CITAC guide on how to assess compliance with a specification or a regulation

Introduction

When test results are used to assess compliance i.e. to decide whether specifications or regulations are met, the measurement uncertainty of the test results has to be taken into account. Assessment of compliance for cases *i* and *iv* in Figure 1 is clear – the measurement results including the uncertainty interval are clearly below or above the limit value. For cases *ii* and *iii* the decision is not clear since the uncertainty interval overlaps the limit value. The Eurachem/CITAC guide [1] gives guidance on cases *ii* and *iii*.

Non-compliance with limit value
Compliance with limit value
i ii iii iv

Fig 1 Test results with expanded uncertainty in relation to an upper limit

We need acceptance & rejection zones

In order to judge whether the results in cases *ii* and *iii* comply with the limit value we need a *decision rule*, based on the risks associated with making a wrong decision. This *decision rule* enables a *guard* band, *g* to be calculated (see Figure 2) which defines an *acceptance zone* and a *rejection zone*. If the measurement result is within the *acceptance zone* the specifications are met and we can assess *compliance*. If the measurement result is in the *rejection zone* we can assess *non compliance*. The intersection between these two zones is called the *decision limit*, see Figure 2. The guard band is chosen so that for a measurement in the acceptance zone the probability of false acceptance/rejection is less than or equal to a defined confidence value α.

g
Upper limit
Decision limit
Acceptance zone
Rejection zone

Fig 2 A guard band (g), a decision limit and an acceptance and a rejection zone based on an upper limit specification and a decision rule stating a high confidence of correct acceptance

Information needed for decision making

The following information is needed to reach a decision
- A measurand clearly specified
- An analytical result
- An uncertainty – For an expanded uncertainty the *k* factor and the corresponding confidence level should be stated e.g. $k = 2$ for 95 % confidence
- A specification giving upper and/or lower limits
- A *decision rule*

Based on the uncertainty and the decision rule the *guard band* is calculated. Based on the specification and the *decision rule*, the decision limit and the *acceptance* and *rejection zones* are calculated – see Figure 2.

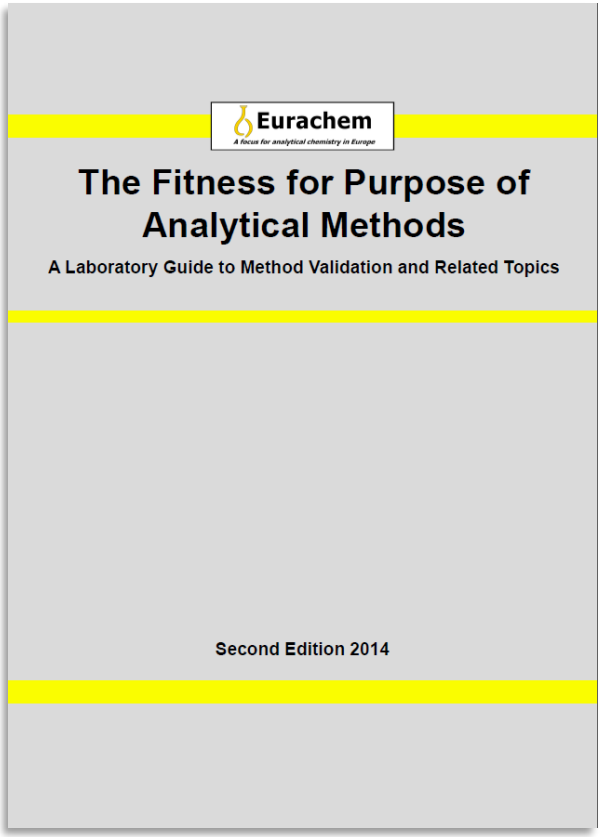
Three examples

Example 1 – case *ii* in Figure 1 with an upper limit and a decision rule focusing on correct acceptance

Sludge from water purification plants can be used for soil improvement. One of the toxic metals that can be a problem is cadmium. The upper limit on the total cadmium in sludge is set to 2 mg/kg.
- Measurand – Mass fraction of cadmium, Cd, in a consignment delivered to a customer
- Analytical result - mass fraction (Cd) = 1.82 mg/kg
- Uncertainty – $U = 0.20$ mg/kg, $k = 2$ (95 %). Standard uncertainty, $u = 0.10$ mg/kg. The uncertainty includes both sampling and analytical uncertainty
- Specification – Upper permitted limit 2.0 mg/kg

Version 10 (2015-04-01)

The fitness for purpose of analytical methods: A laboratory guide to method validation and related topics



- What is method validation?
- Why is method validation necessary?
- When should methods be validated or verified?
- How should methods be validated
- Method performance characteristics (selectivity, precision, trueness, etc.)
- Using validated methods
- Using validation data to design quality control
- Documentation
- Implications of validation data for calculating and reporting results

LGC

What is validation?

'The **confirmation** by examination and the provision of **objective evidence** that the particular requirements for a **specific intended use** are fulfilled' *

- specific intended use = analytical requirement
- objective evidence = experimental data
  (method performance parameters)
- Confirmation = comparison between requirement and (evidence) data

Can the method deliver results that are fit for a particular purpose?

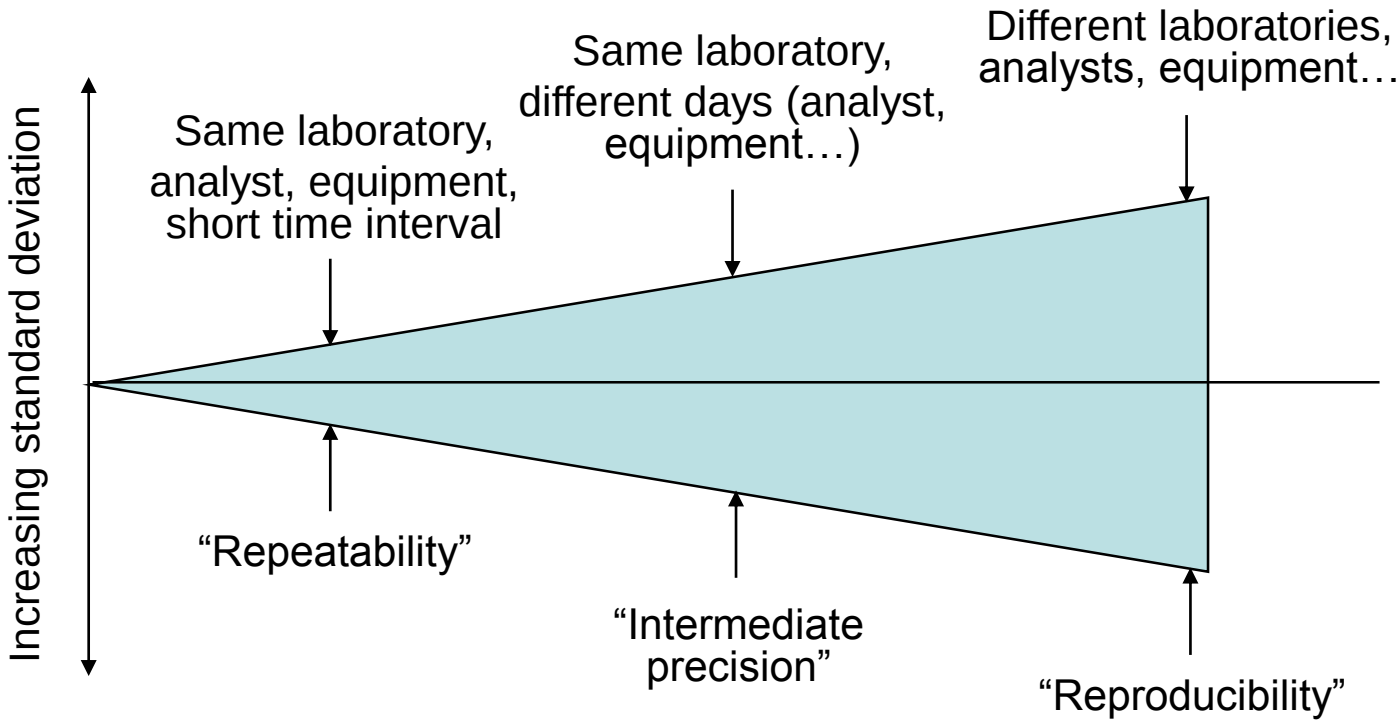
* [ISO/IEC 17025 definition]

Uses of statistics in method validation

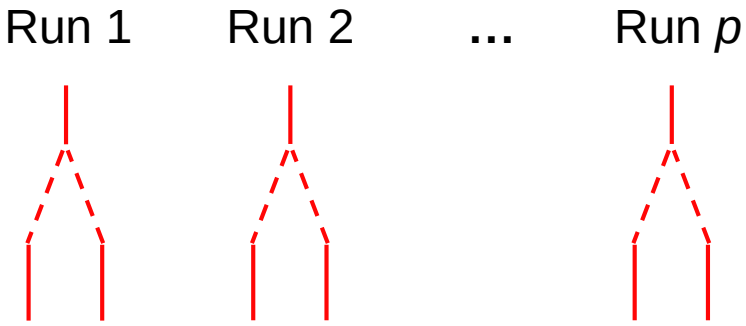
- Summarising data
- Planning efficient studies
- Checking for significant differences
  - Is there a significant bias in my results?
  - Are these two methods equally precise?
  - Is there a significant between run effect?
  - Is my method rugged/robust?
- Assessing capability of detection
- Include data analysis as part of the validation planning process

Planning efficient precision studies

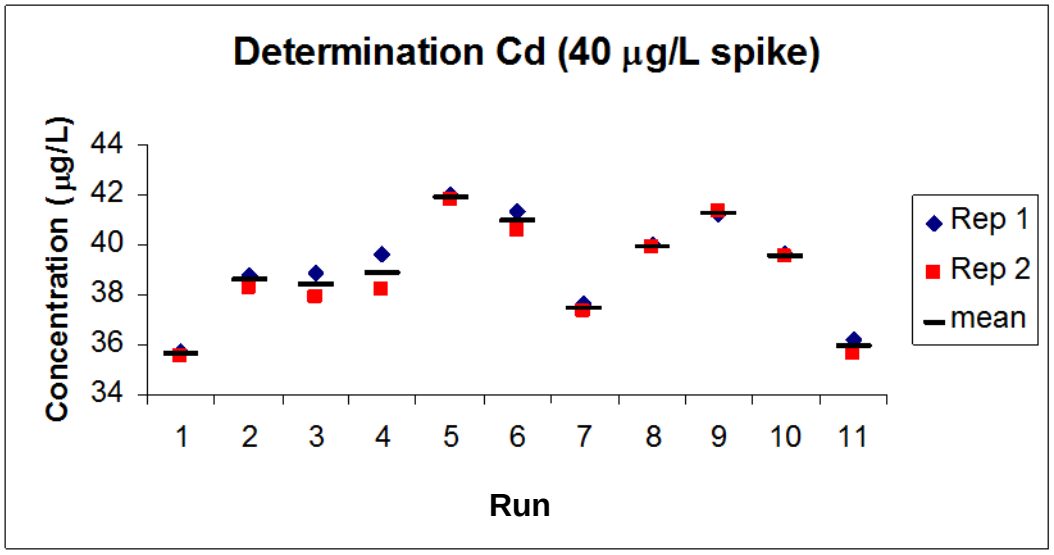
- Precision – Closeness of agreement between independent test/measurement results obtained under stipulated conditions



Nested designs – an efficient study



Sample analysed *n* times in each of *p* runs
Vary parameters between runs – day, analyst, equipment…



"11x2 design"
11 runs containing duplicate measurements (repeatability conditions) of each sample

Nested designs – advantages

- Saves effort where several sets of conditions are to be studied
  - Repeatability and intermediate precision
  - Small groups allow different samples to be analysed in a run (different matrices, concentrations…)
- Practical solution to gaining enough data
  - E.g. when the measurement time is long
  - Several small sets of data are combined to give sufficient data (degrees of freedom)
  - Evaluated using analysis of variance (ANOVA)

Nested designs - analysis

ANOVA						
Source of Variation	*SS*	*df*	***MS***	*F*	*P-value*	*F crit*
Between Groups	85.18455	10	**8.518455**	43.992	2.04×10^{-7}	2.854
Within Groups	2.13	11	**0.193636**			
Total	87.31455	21				

Repeatability (s_r)

$$s_r = \sqrt{\text{within group MS}}$$

$$s_r = \sqrt{0.194} = 0.440\ \mu\text{g/L}$$

Intermediate precision (s_I)

$$s_b = \sqrt{\frac{\text{between group MS - within group MS}}{n}}$$

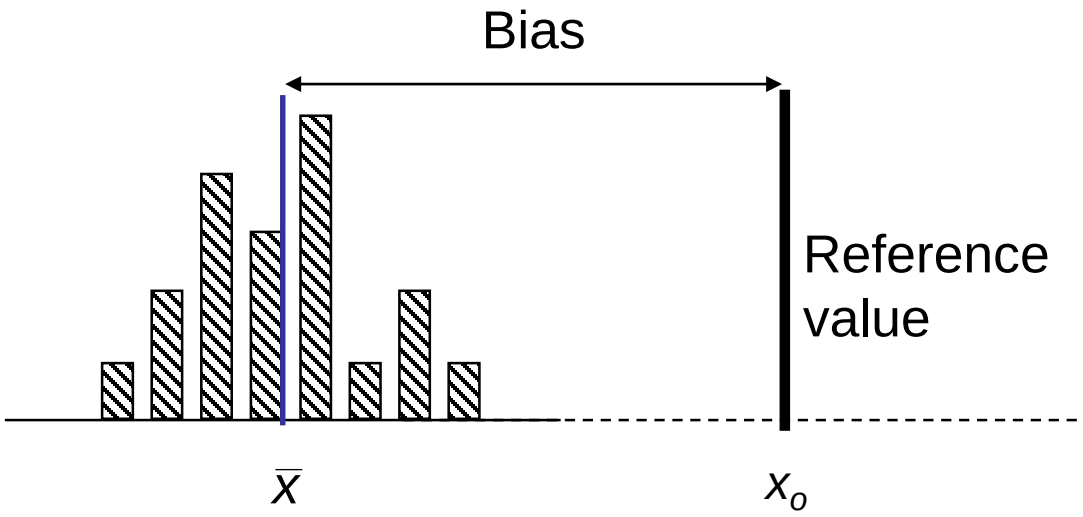
$$s_I = \sqrt{s_r^2 + s_b^2}$$

$$s_b = \sqrt{\frac{8.518 - 0.194}{2}} = 2.04\ \mu\text{g/L}$$

$$s_I = \sqrt{0.440^2 + 2.04^2} = 2.09\ \mu\text{g/L}$$

Checking for significant differences

- Difference between mean of observations and a reference value (bias assessment)



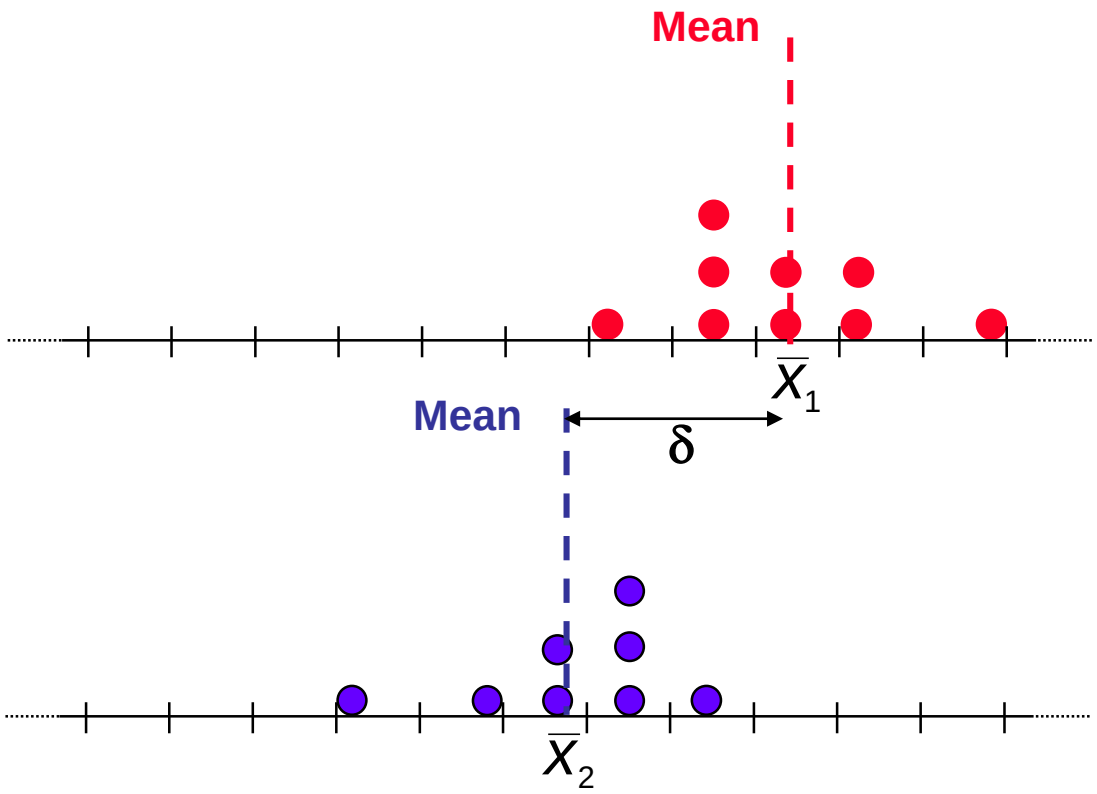
One-sample *t*-test

$$t = \frac{\bar{x} - x_o}{s/\sqrt{n}}$$

$t > t_{crit}$?

Checking for significant differences

- Difference between the means of two data sets



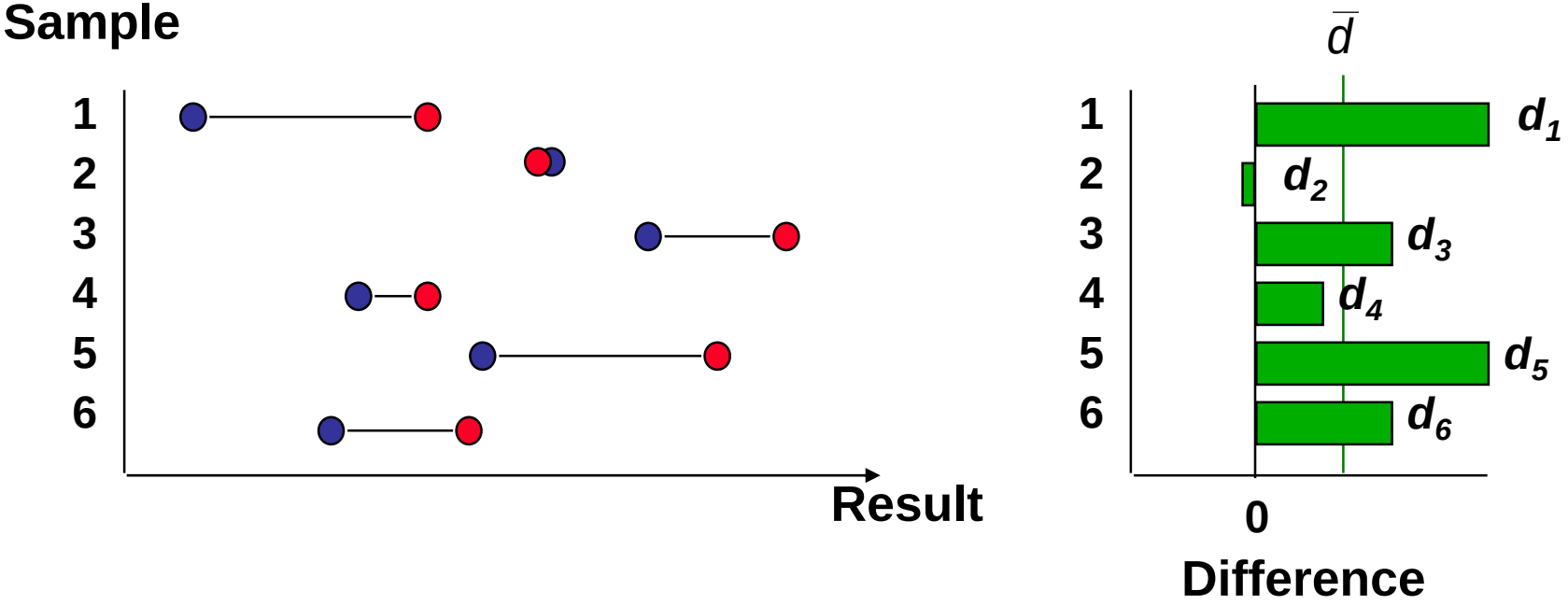
Two-sample *t*-test

$$t = \frac{\bar{x}_2 - \bar{x}_1}{s_{pool}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

***t* > t_{crit}?**

Checking for significant differences

- Difference between pairs of data



Paired-sample *t*-test

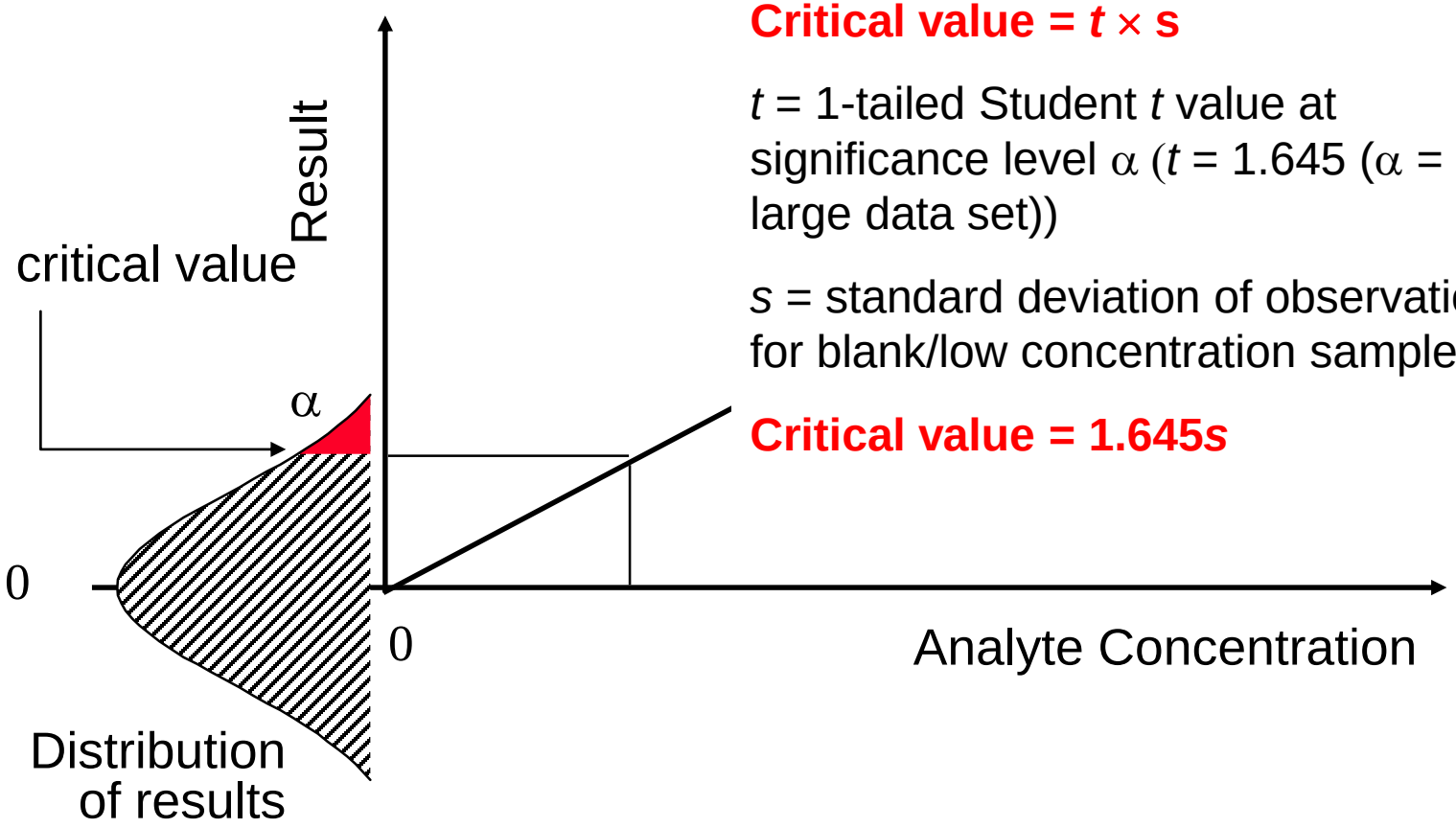
$$t = \frac{\bar{d}}{s(d)/\sqrt{n}}$$

$t > t_{crit}$?

Limit of detection calculations

- “3 times standard deviation of the blank”
- Where does the factor of 3 come from?
- Concepts
  - Critical value – method response taken to indicate analyte is present
  - Detection limit – lowest concentration of analyte that can be detected at a specified level of confidence

Statistical basis of limits



α false positive rate (typically 5%)

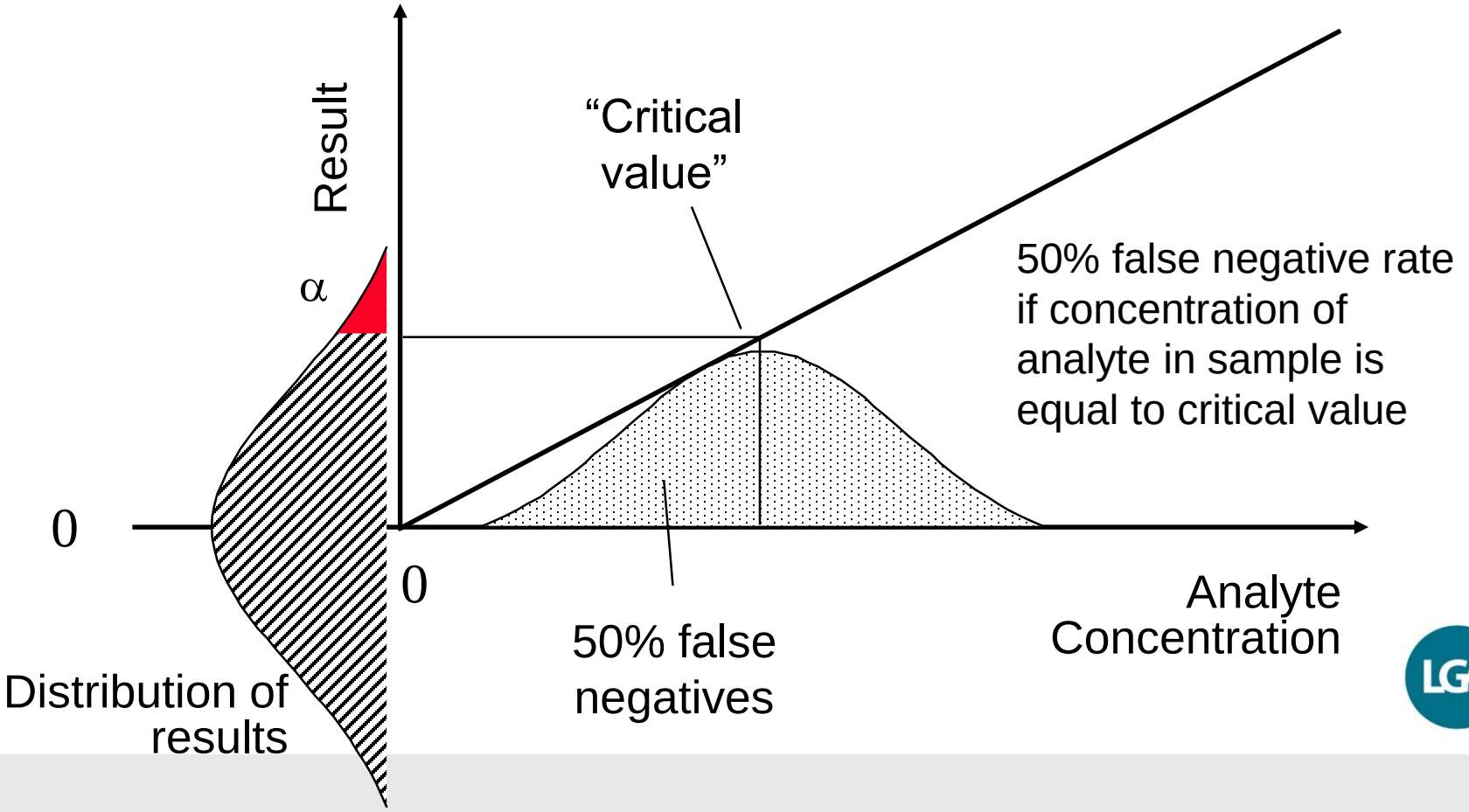
Critical value = $t \times s$

t = 1-tailed Student t value at significance level α (t = 1.645 (α = 0.05, large data set))

s = standard deviation of observations for blank/low concentration sample

Critical value = 1.645s

Statistical basis of limits



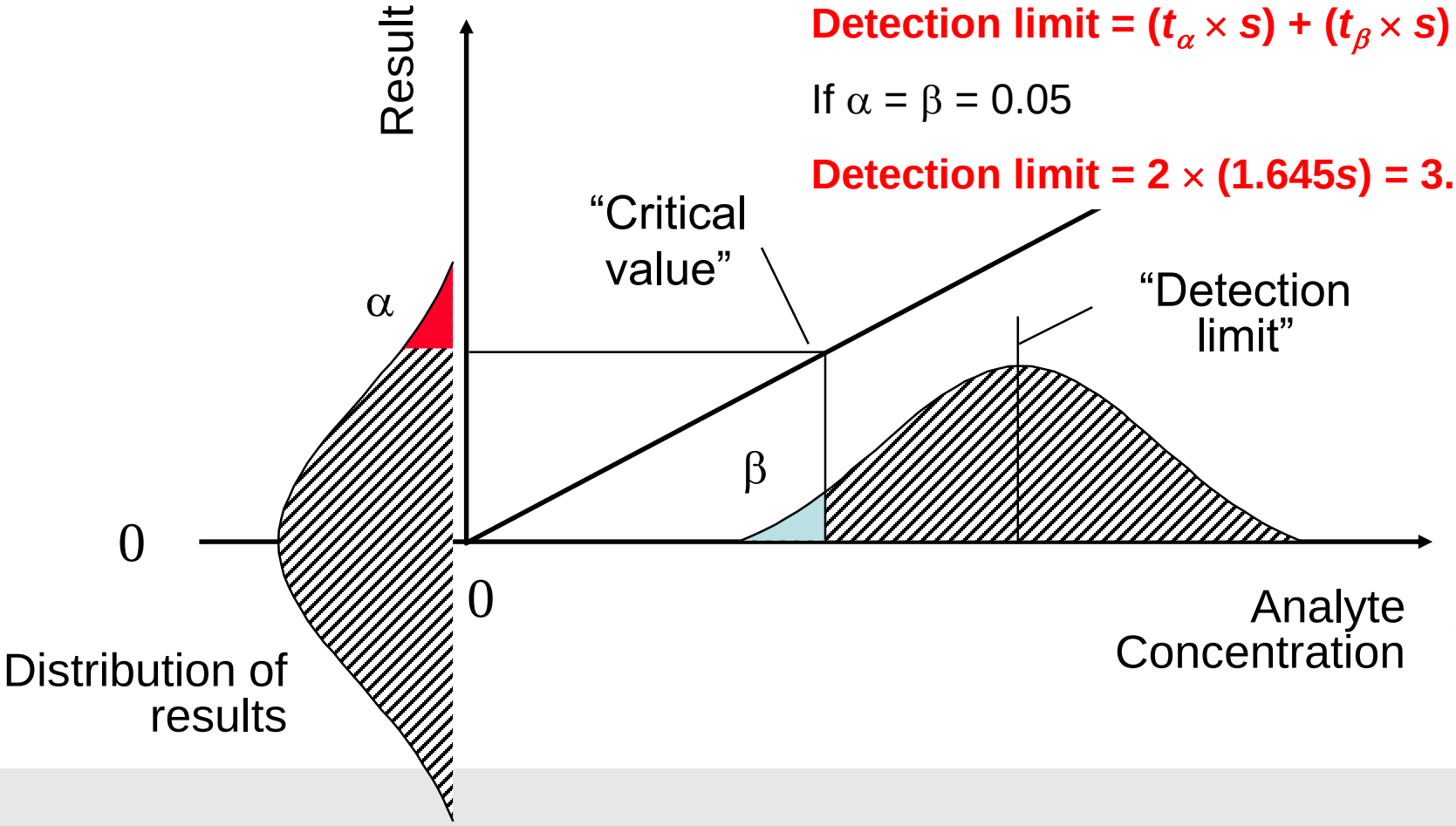
Statistical basis of limits

β false negative rate (typically 5%)

Detection limit = ($t_\alpha \times s$) + ($t_\beta \times s$)

If α = β = 0.05

Detection limit = 2 × (1.645*s*) = 3.3*s*



Ruggedness

- A measure of a method's capacity to remain unaffected by small, but deliberate variations in method parameters
  - Ruggedness provides an indication of the method's reliability during normal usage
- Ruggedness study – deliberately change method operating parameters
  - Determine if there is a significant effect on the measurement result

Typical method parameters

- Concentration of reagents
- Volumes of reagents
- pH
- Extraction time
- Extraction temperature
- Flow rates through chromatographic systems
- Age of chromatographic column

Plackett-Burman experimental design

Experimental parameter	Experiment number							
	1	2	3	4	5	6	7	8
A or a	A	A	A	A	a	a	a	a
B or b	B	B	b	b	B	B	b	b
C or c	C	c	C	c	C	c	C	c
D or d	D	D	d	d	d	d	D	D
E or e	E	e	E	e	e	E	e	E
F or f	F	f	f	F	F	f	f	F
G or g	G	g	g	G	g	G	G	g
Observed result	s	t	u	v	w	x	y	z

- 7 parameters at 2 levels
- 8 experiments
- Representative test material
- Effect of each parameter can be isolated from effect of changing the others

LGC

Plackett-Burman experimental design – data evaluation

- Calculate differences for each parameter

$$D_A = \frac{(s + t + u + v)}{4} - \frac{(w + x + y + z)}{4}$$

Experimental parameter	Experiment number							
	1	2	3	4	5	6	7	8
A or a	A	A	A	A	a	a	a	a
B or b	B	B	b	b	B	B	b	b
C or c	C	c	C	c	C	c	C	c
Observed result	s	t	u	v	w	x	y	z

Magnitudes of difference indicate relative significance of each parameter

Can also apply significance test – is a difference *D* significantly different from zero?

Summary

- Eurachem develops guidance and organises workshops on key quality assurance issues
  - Visit www.eurachem.org
- Statistics are essential for planning and evaluating efficient validation studies
- Plan data analysis from the outset
  - Statistics should not be a salvage operation!

Method validation – Current practices and future challenges

- 9-10 May 2016, Gent, Belgium
- www.belab-eurachem2016.com

Current practices

- International guidance
- Setting method performance requirements
- Extent of validation/verification studies
- Planning effective validation studies
- Analysis of validation data
- Examples of best practices in different fields

Future challenges

- Future developments – Accreditation Body viewpoint
- Validation of microbiological methods
- Validation of multiparameter methods
- Implementing principles of Quality by Design (QbD)

LGC

Acknowledgements

LGC contribution to Eurachem activities is supported by the UK National Measurement System.

National Measurement System

The National Measurement System delivers world-class measurement for science and technology through these organisations

National Measurement & Regulation Office

NPL National Physical Laboratory

LGC

nel technology for life