Appropriate Statistical Techniques

considering degree of contamination of the PT results

Dan THOLEN
Piotr ROBOUCH

Eurachem PT 2023 – Windsor, 28th September 2023

Objective of Presentation

Actionable recommendations for PT providers and Accreditation Bodies

1. Improve traceability in PT schemes with limited participation;

2. Use statistical techniques that are appropriate for the experience of the scheme and its participants;

3 Evaluate the measurement uncertainties (MU) of PT results from participants in (some? all?) testing PT. This is essential in PT schemes that can be used to demonstrate competence in laboratories that make conformity assessment decisions.

1. Traceability in PT with limited participation

Ideal	- Assigned value is independent from participants & metrologically traceable to SI (or an appropriate reference) - SDPA not from consensus (fitness-for-purpose) - Consensus statistics are provided as additional information
Good	- Traceable reference within the dataset (bias check) - Set practical or technical limits on the SDPA (consensus can vary greatly)
Achievable	- Conduct a risk assessment on the impact of lack of traceability (e.g., regional bias, possible exclusion of traceable results) **ISO/IEC 17043:2023 clauses 7.2.3.1 and 7.2.3.3**

Statistical techniques for new schemes (1)

New scheme (open or closed, unknown ‘contamination’ level):

- ✓ Use traceable assigned values and fit-for-purpose limits where possible.
- ✓ Gross outlier removal and robust statistical techniques
- ❑ **When using consensus –** caution on evaluating if $u(x_{pt}) > 0.3\ \sigma_{pt}$ something is wrong
  - ❑ Place limits on sdpa and the acceptance range (e.g., avoid including 0 or crossing interpretation ranges)

Statistical techniques for experience of scheme (2)

Experienced scheme (open with low contamination):

- ✓ Use experience and technical knowledge to determine the sdpa
- ✓ Use fitness-for-purpose to limit acceptance ranges
  - ✓ LOOK at the acceptance ranges – would a customer take the same action on a product if the results from 2different labs were at the limits of the acceptance range?
  - ✓ IF the customer would NOT take the same decision, is the scheme fit for purpose?
- ✓ **Do not use** the "1.25" multiplier for the uncertainty of the robust mean in Algorithm A (use 1.0 or 1.05 at the highest)

Statistical techniques for experience of scheme (3)

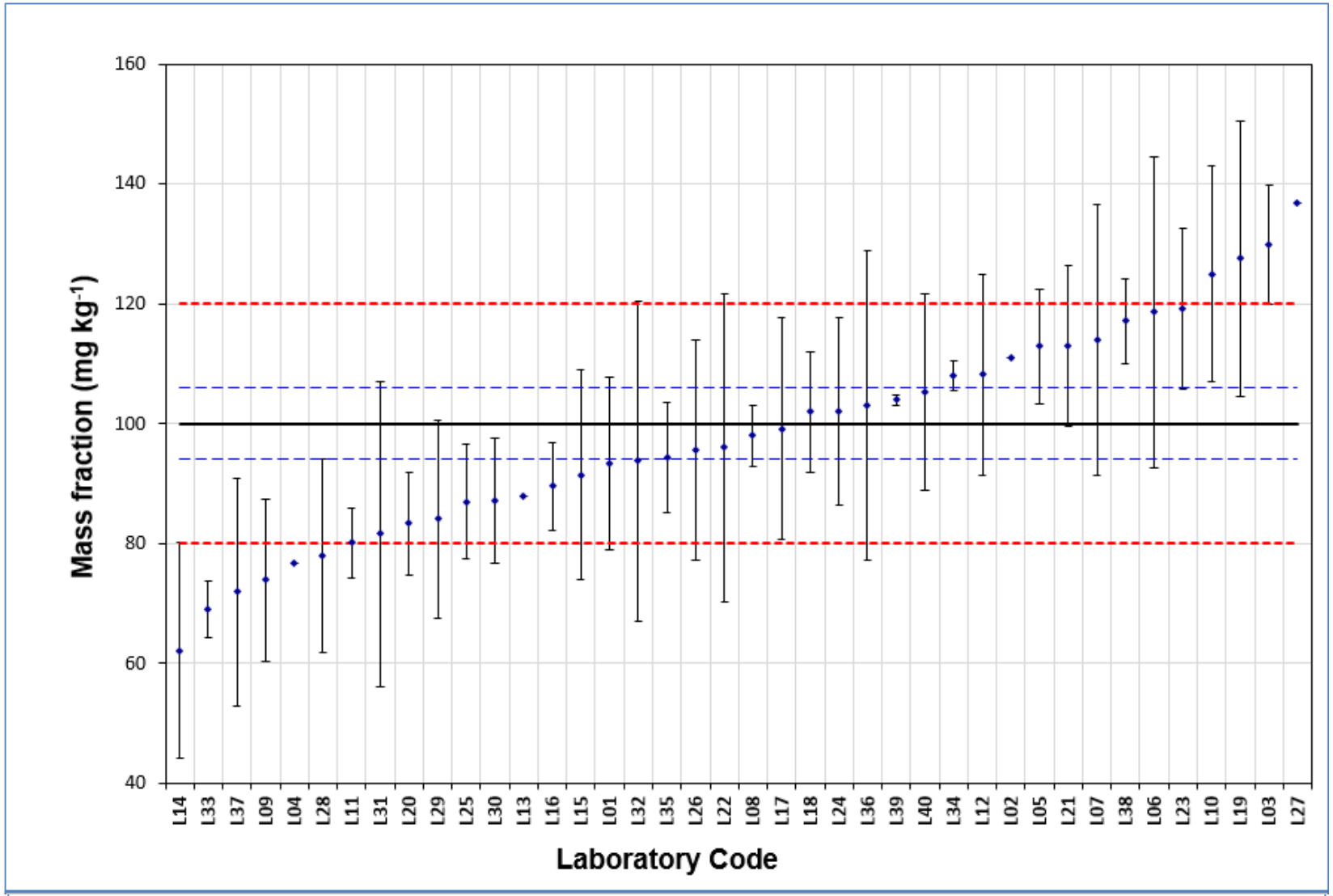
Closed scheme (rare contamination):

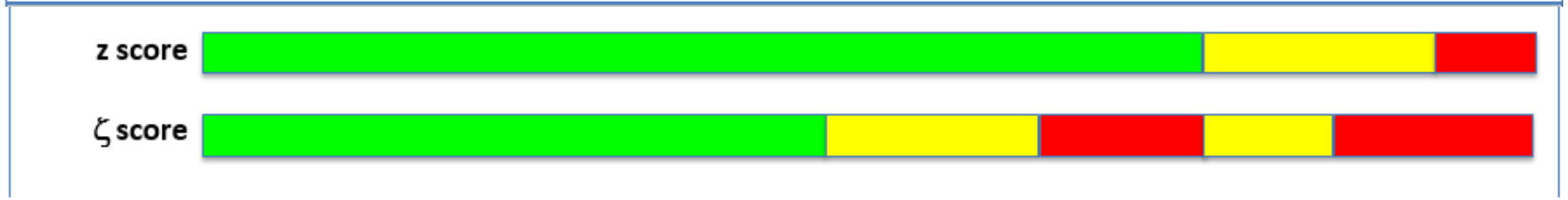
- ✓ Metrological traceability of assigned values is always recommended
- ✓ Can use conventional mean and SD, with review of statistical outliers.
  - ✓ Robust statistical methods may not be needed
- ✓ Gross outliers are removed and investigated.
- ✓ Evaluate measurement uncertainty (Zeta, Bias, graphical)
- ✓ Track performance metrics

Request participants to report MU in PTs

Good	- Reasonable MU: $u(x_{pt})_{rel,\%} < u(x_i)_{rel,\%} < \sigma_{pt,\ rel,\%}$ - Include a figure in the report showing $u(x_i)$ reported by participants
Better	- z , zeta (ζ) score for assessment + MU evaluation [a,b,c]
Best	- Naji2 plot (z, ζ, MU) (i) a guide to investigate poor performance in PT, or (ii) an overview of performance over time - https://doi.org/10.1007/s00769-022-01496-w (2022)

Evaluation of MU should be mandatory for any PT scheme where participants use the PT to support accreditation scopes that include making Conformity Assessment decisions





x_{pt} = 100

$u(x_{pt})$ = 3 (*k* = 1)

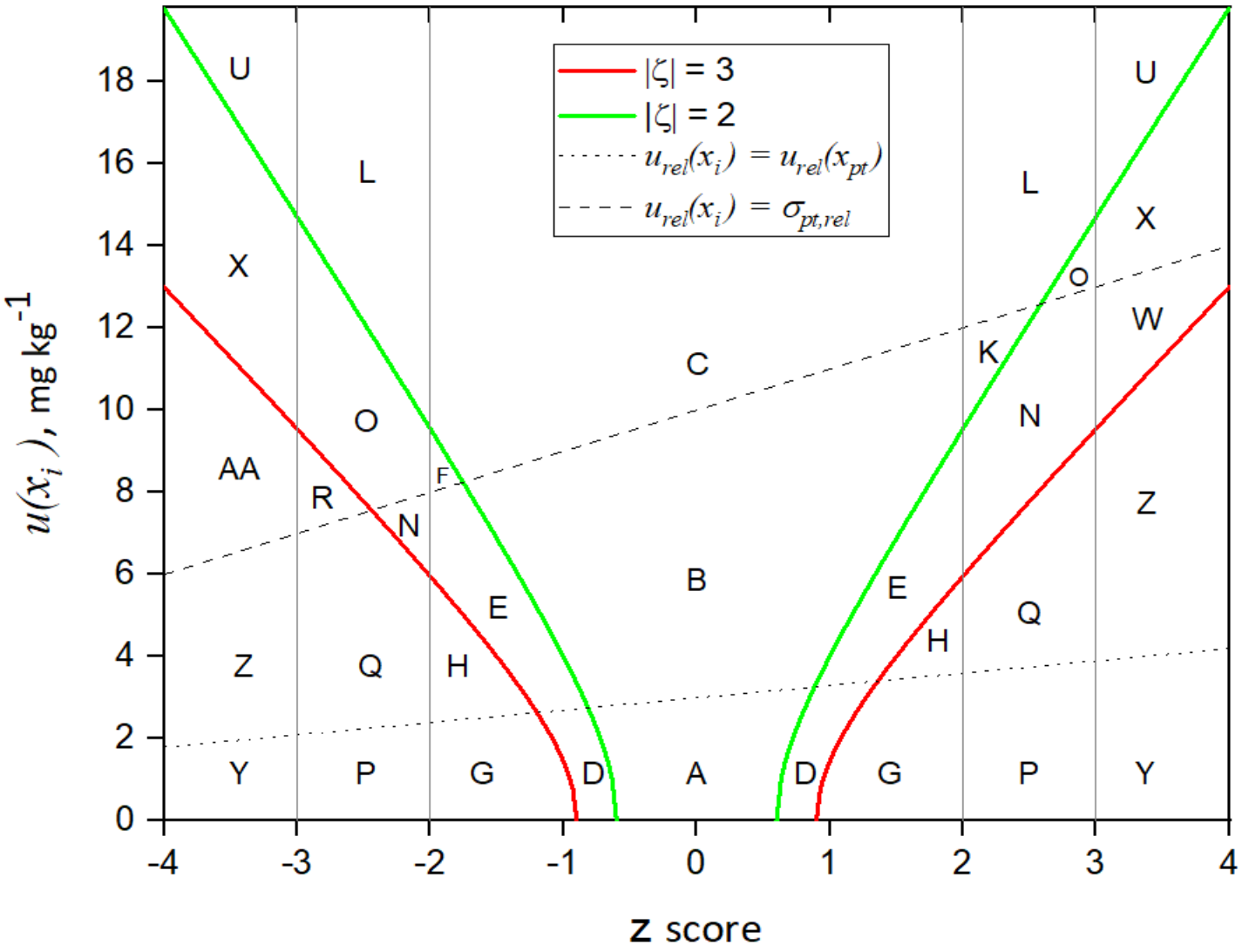
σ_{pt} = 10 (10 %)

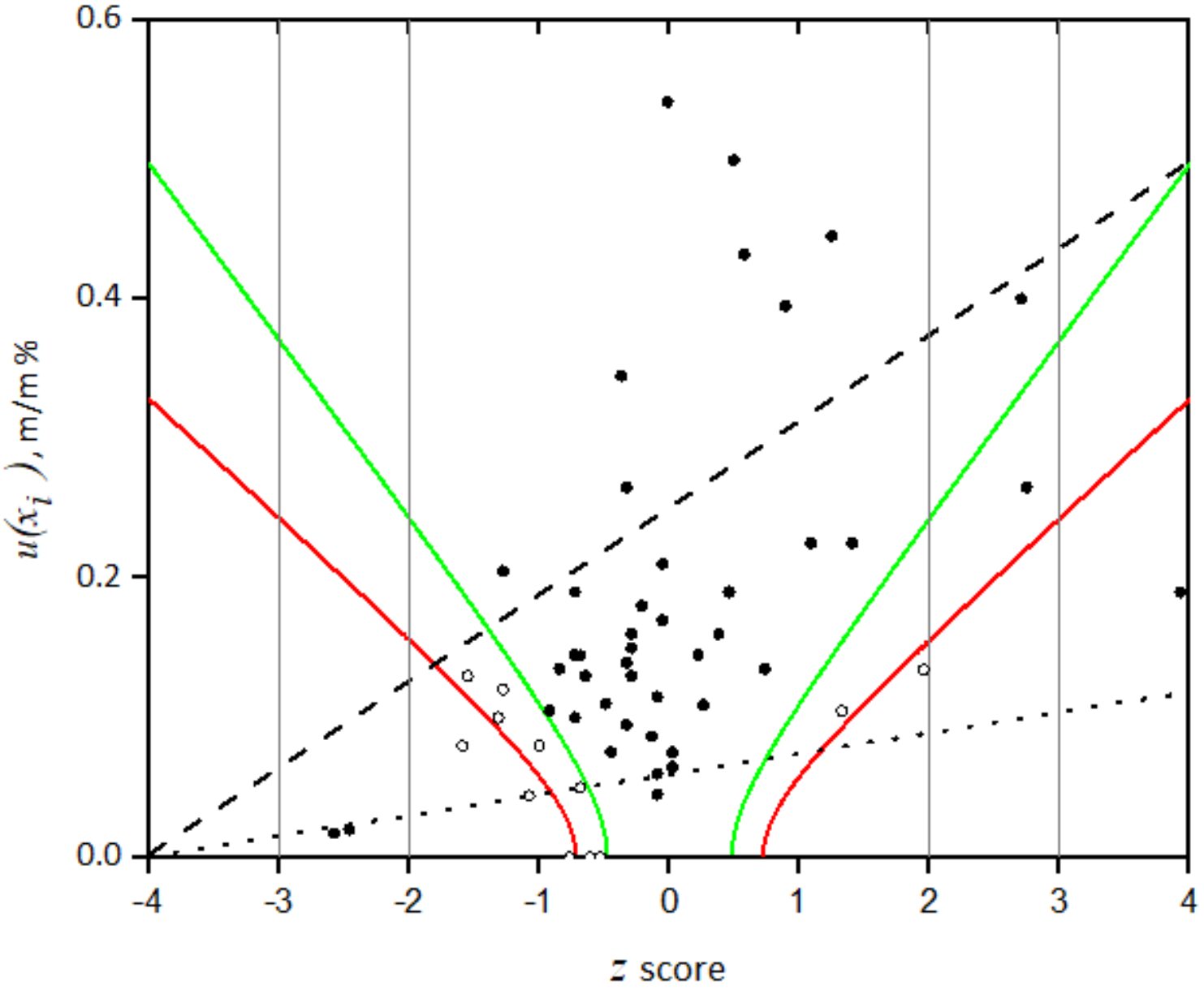
in mg/kg

European Commission

Naji2 Plot explanation

z score	ζ score	MU	Plot areas
Satisfactory	Satisfactory	Under-estimated	A
		Realistic	B
		Over-estimated	C
	Questionable	Under-estimated	D
		Realistic	E
		Over-estimated	F
	Unsatisfactory	Under-estimated	G
		Realistic	H
		Over-estimated	
Questionable	Satisfactory	Under-estimated	
		Realistic	K
		Over-estimated	L
	Questionable	Under-estimated	
		Realistic	N
		Over-estimated	O
	Unsatisfactory	Under-estimated	P
		Realistic	Q
		Over-estimated	R
Unsatisfactory	Satisfactory	Under-estimated	
		Realistic	
		Over-estimated	U
	Questionable	Under-estimated	
		Realistic	W
		Over-estimated	X
	Unsatisfactory	Under-estimated	Y
		Realistic	Z
		Over-estimated	AA





GMFF-19-02:
GM Soybean 40-3-2 in pig feed

$x_{pt} = 1.014$

$u(x_{pt}) = 0.061$ $(k = 1)$

$\sigma_{pt} = 0.254$ (25 %)

in m/m %

European Commission

Oversight Body Responsibility

- These recommendations will not be implemented by PT Providers **unless** requested by customers or by specifiers, e.g.
  Regulatory Authorities, Accreditation Bodies
  & Large PT Customers/communities (e.g. Food, Environment, Geology)

Essential for the credibility of accreditation:

- ✓ **ISO/IEC 17025** requires MU estimation, since 2005
- ✓ **ILAC MRA** ➔ international equivalence of competence, & compatibility of results, since 2010

Let's start collecting the evidence to support claims for compatibility in measurements and accreditation

Thank you

tholen.dan@gmail.com

piotr.robouch@ec.europa.eu

European Commission