Beyond the dot-and-bar plot: Graphical methods for interlaboratory data analysis

S Ellison
LGC Teddington

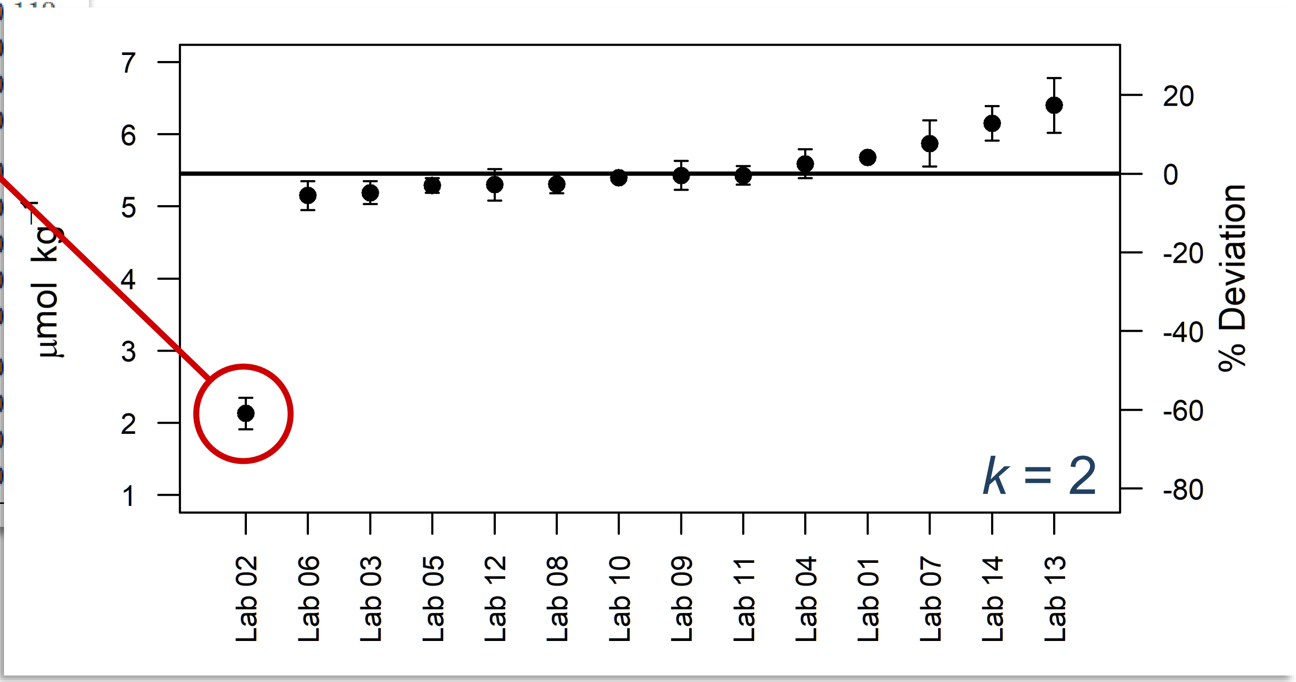
Why use graphical methods?

Why use graphical methods?

NML
UK NATIONAL MEASUREMENT LABORATORY
FOR CHEMICAL AND BIO-MEASUREMENT
HOSTED AT LGC

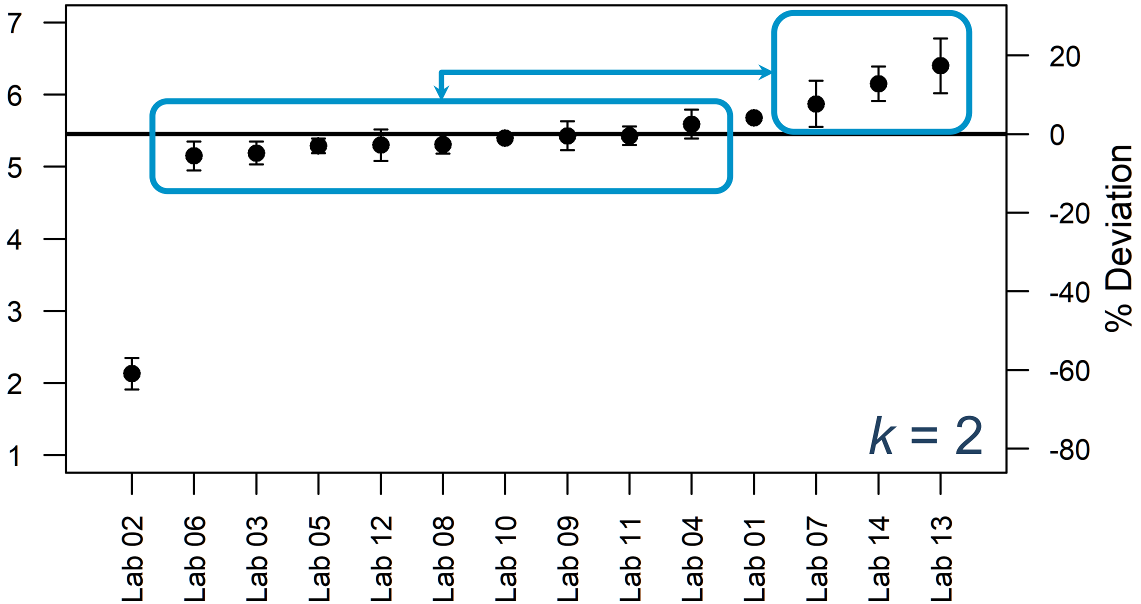
Institute	x / µmol kg^{-1}	u / µmol kg^{-1}
Lab01	5.681	0.029
Lab02	2.130	[illegible]
Lab03	5.190	[illegible]
Lab04	5.590	[illegible]
Lab05	5.290	[illegible]
Lab06	5.150	[illegible]
Lab07	5.870	[illegible]
Lab08	5.310	[illegible]
Lab09	5.430	[illegible]
Lab10	5.397	[illegible]
Lab11	5.430	[illegible]
Lab12	5.300	[illegible]
Lab13	6.400	[illegible]
Lab14	6.150	[illegible]

CCQM-K13
Cadmium in sediment



Why use graphical methods?

Institute	x / µmol kg^{-1}	u / µmol kg^{-1}
Lab01	5.681	0.029
Lab02	2.130	0.110
Lab03	5.190	0.080
Lab04	5.590	0.100
Lab05	5.290	0.050
Lab06	5.150	0.100
Lab07	5.870	0.160
Lab08	5.310	0.065
Lab09	5.430	0.100
Lab10	5.397	0.028
Lab11	5.430	0.065
Lab12	5.300	0.110
Lab13	6.400	0.190
Lab14	6.150	0.120



CCQM-K13
Cadmium in sediment

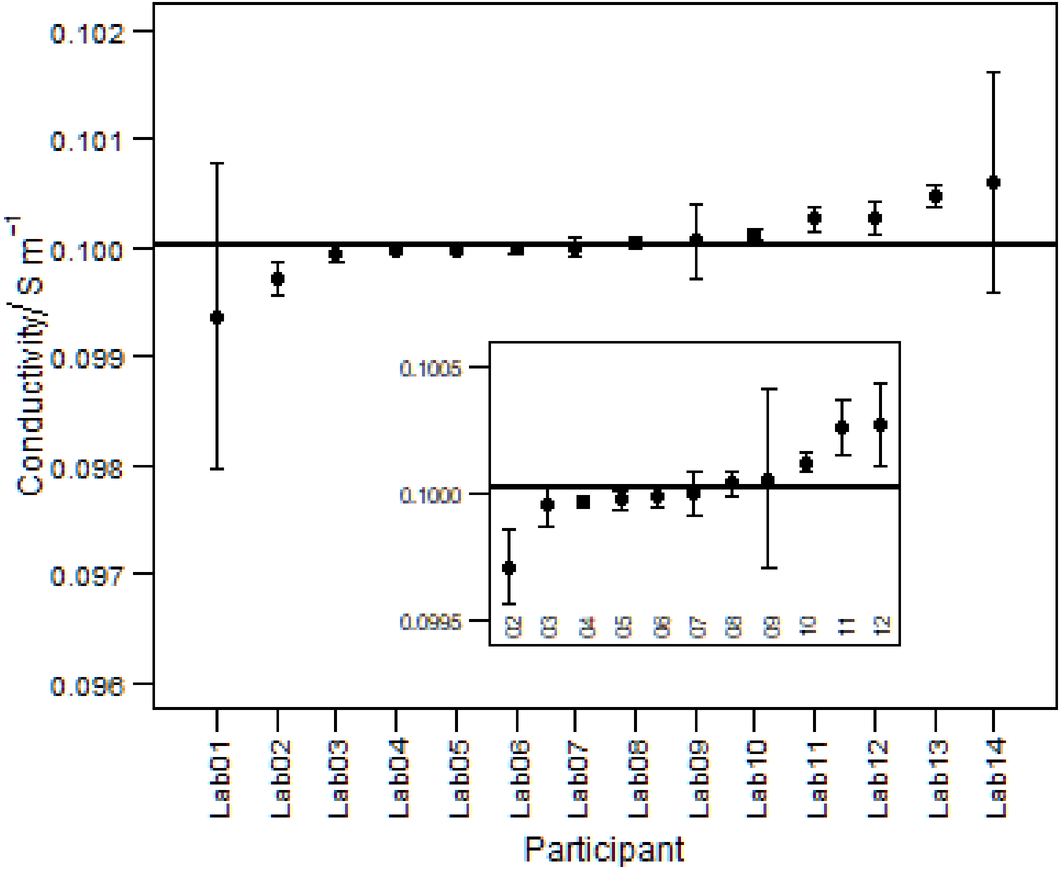
The purpose of visualisation is insight, not pictures

– Ben A Shneiderman

NML
UK NATIONAL MEASUREMENT LABORATORY
FOR CHEMICAL AND BIO-MEASUREMENT
HOSTED AT LGC

The dot-and-bar plot

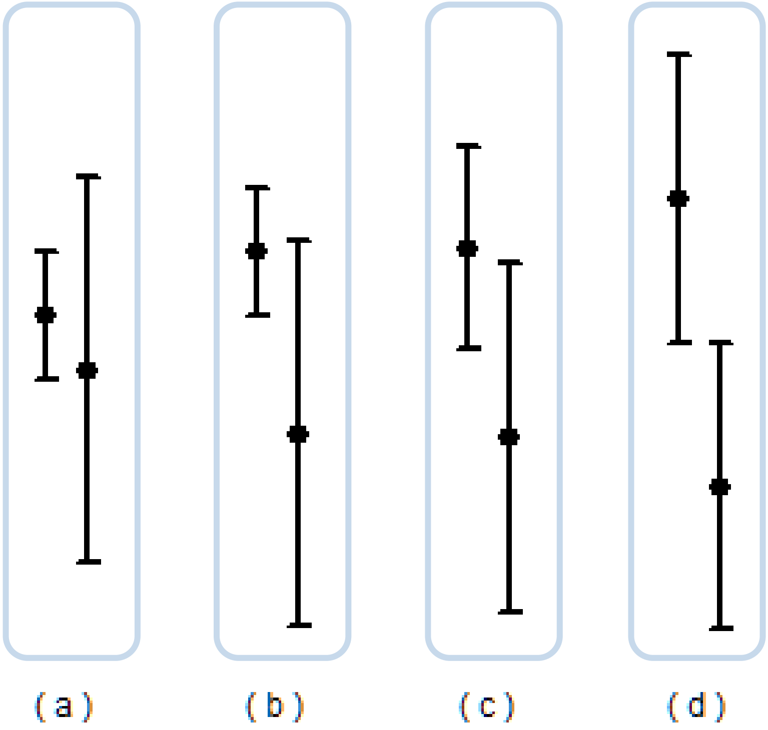
The dot-and-bar plot



Conductivity results from CCQM-P22, on conductivity of standard buffer solutions.

Error bars show expanded uncertainties at *k* = 2.

Inference from dot and bar plots
i) Interpreting overlap



All bars at 95% confidence

a) Both points within others' interval;

b) One point within another's interval:

Never significant at 95% level

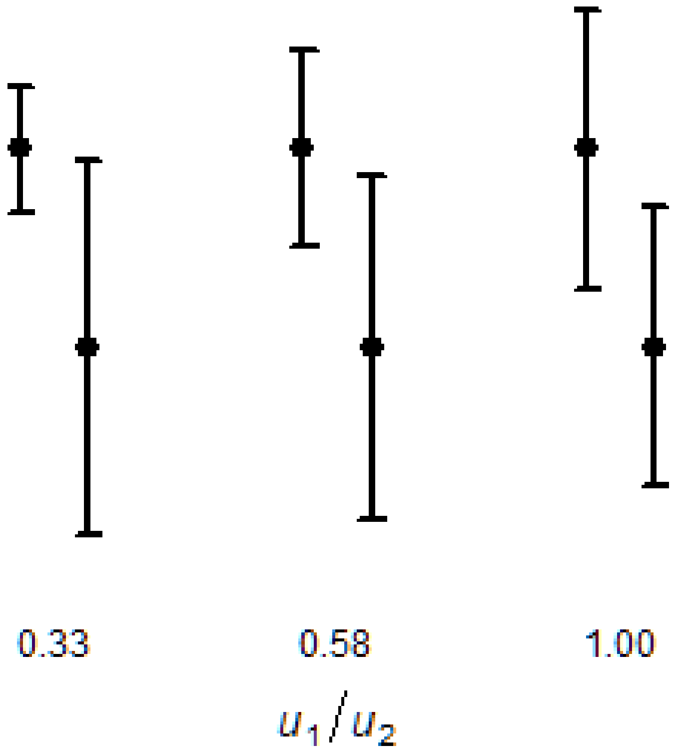
c) Partial overlap

No simple interpretation

d) 95% bars in contact

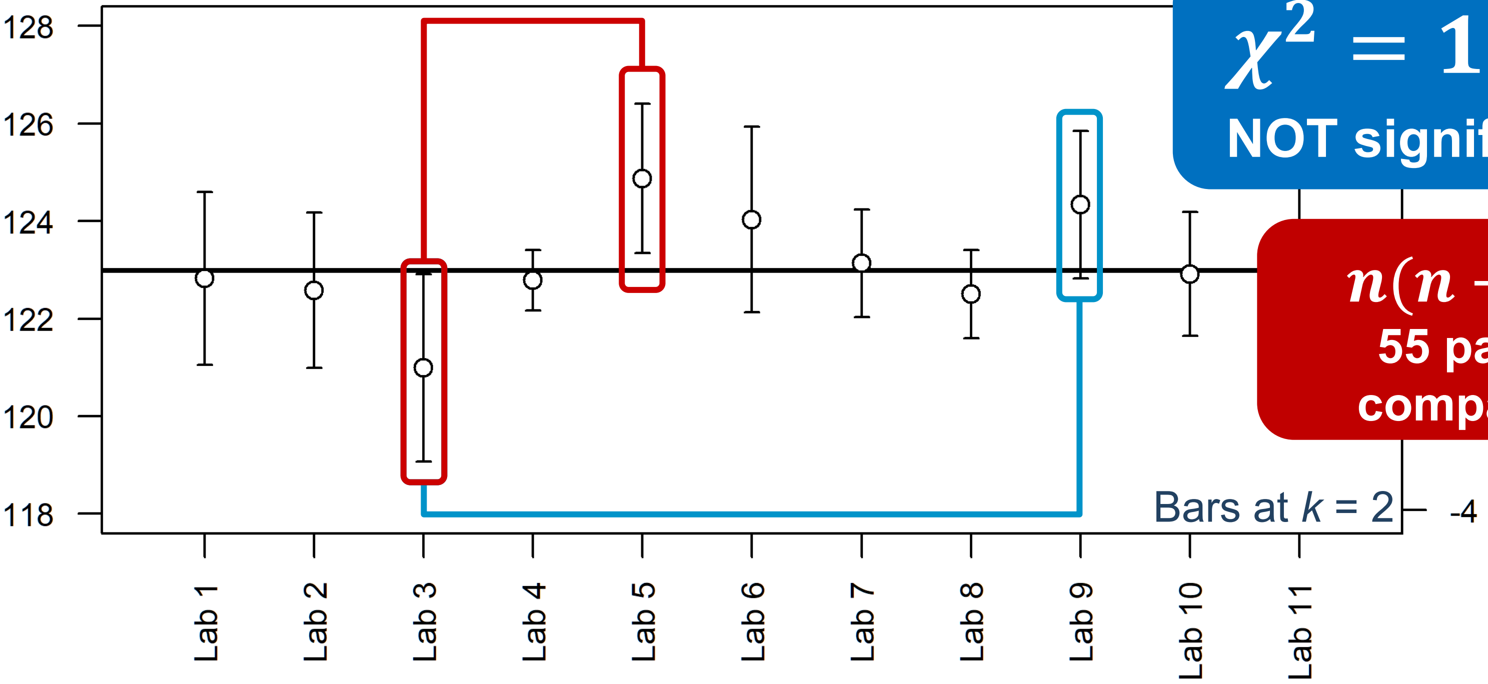
Significant at c. 99% level

Inference from dot and bar plots
ii) Overlap and significance



All pairs show significant difference at exactly 95% confidence

Inference from dot and bar plots
iii) Multiple comparisons



$$\chi^2 = 16.9$$

NOT significant

$$n(n-1)/2$$

55 pairwise comparisons

Dot-and-bar plots
summarise well
but are
easy to misinterpret

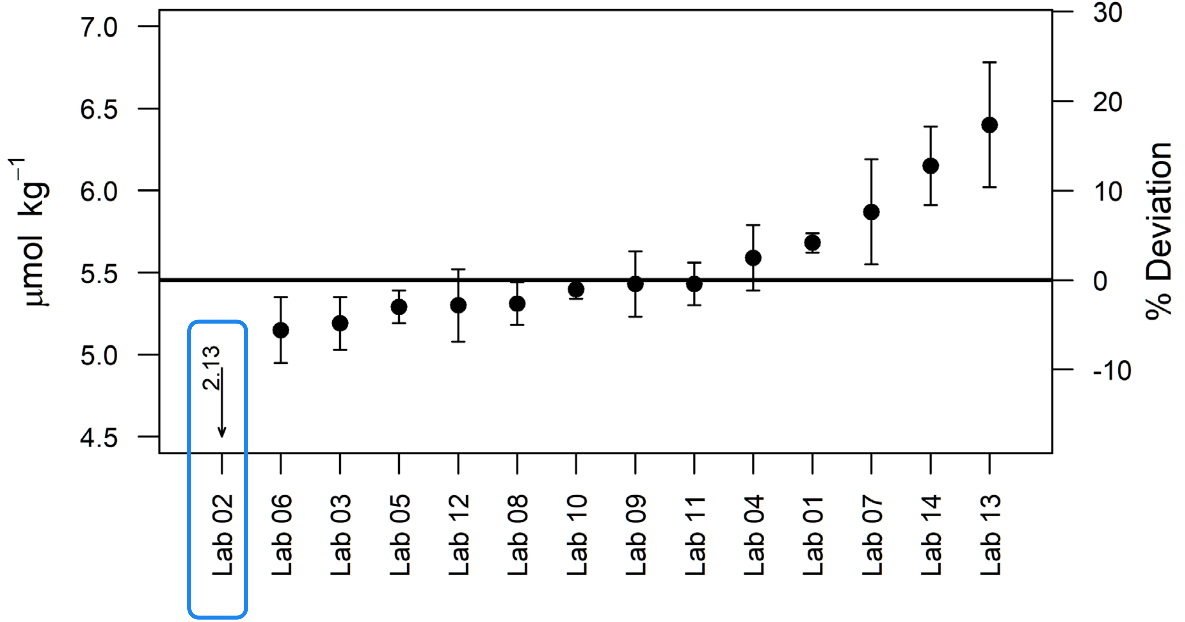
NML
UK NATIONAL MEASUREMENT LABORATORY
FOR CHEMICAL AND BIO-MEASUREMENT
HOSTED AT LGC

Other graphical tools

1. With a reliable reference value

Example data set: CCQM K13 Cadmium in a sediment

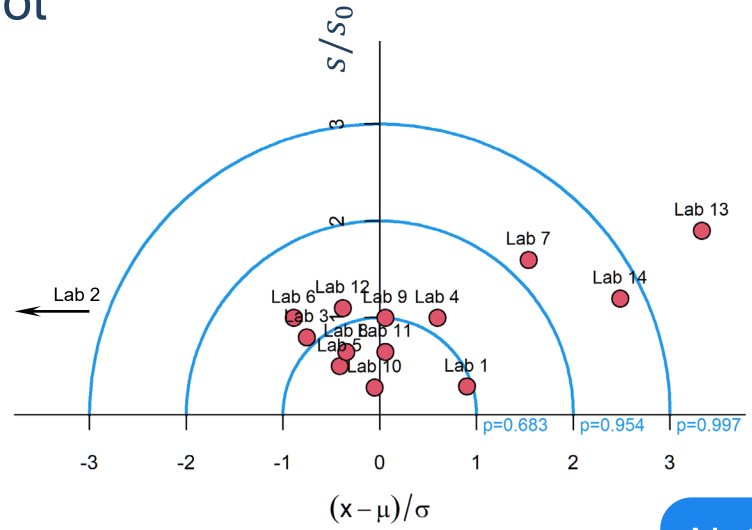
Institute	x / µmol kg^{-1}	u / µmol kg^{-1}
Lab01	5.681	0.029
Lab02	2.130	0.110
Lab03	5.190	0.080
Lab04	5.590	0.100
Lab05	5.290	0.050
Lab06	5.150	0.100
Lab07	5.870	0.160
Lab08	5.310	0.065
Lab09	5.430	0.100
Lab10	5.397	0.028
Lab11	5.430	0.065
Lab12	5.300	0.110
Lab13	6.400	0.190
Lab14	6.150	0.120



Extreme value off scale

Separating value and uncertainty - a common theme

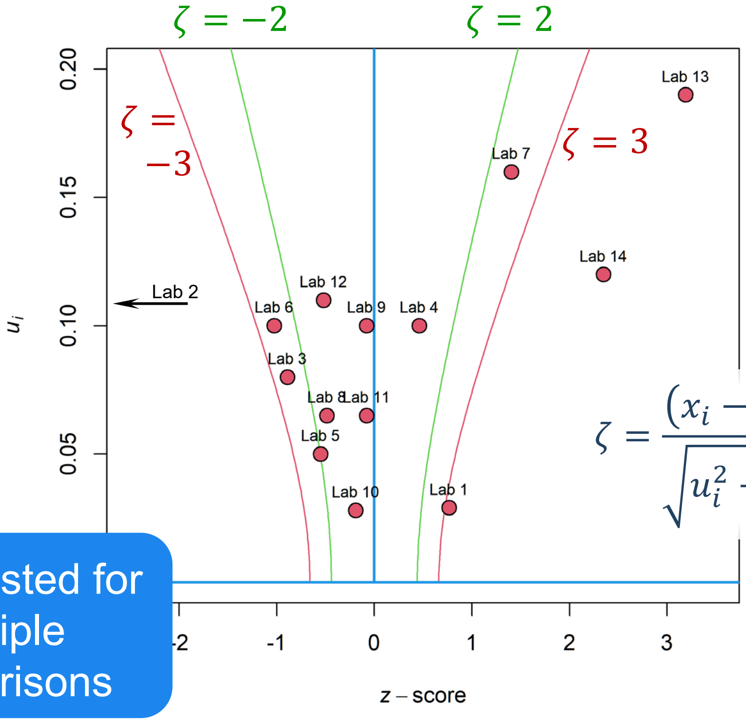
Duewer concordance/precision plot



Uncertainty

Value

Naji 2 plot



$$\zeta = \frac{(x_i - x_{ref})}{\sqrt{u_i^2 + u_{ref}^2}}$$

Not adjusted for multiple comparisons

Additional graphical tools

2. Without a reliable reference value

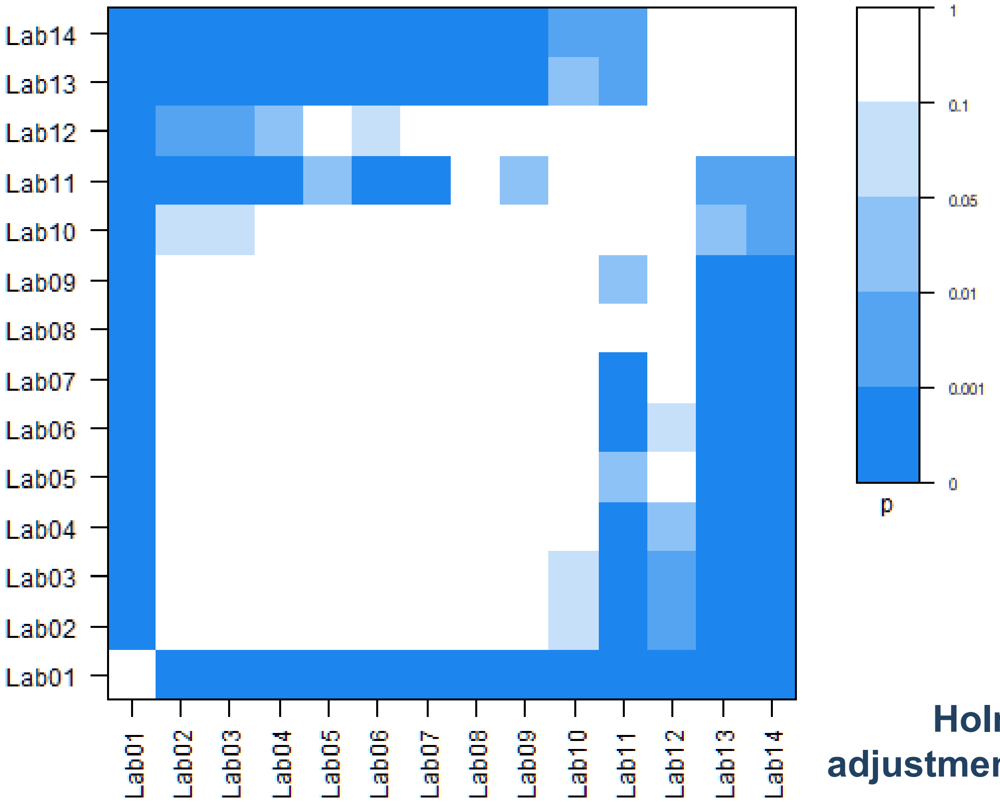
UK NATIONAL MEASUREMENT LABORATORY
FOR CHEMICAL AND BIO-MEASUREMENT
HOSTED AT LGC

Pairwise comparisons don’t need a reference value

UK NATIONAL MEASUREMENT LABORATORY
FOR CHEMICAL AND BIO-MEASUREMENT
HOSTED AT LGC

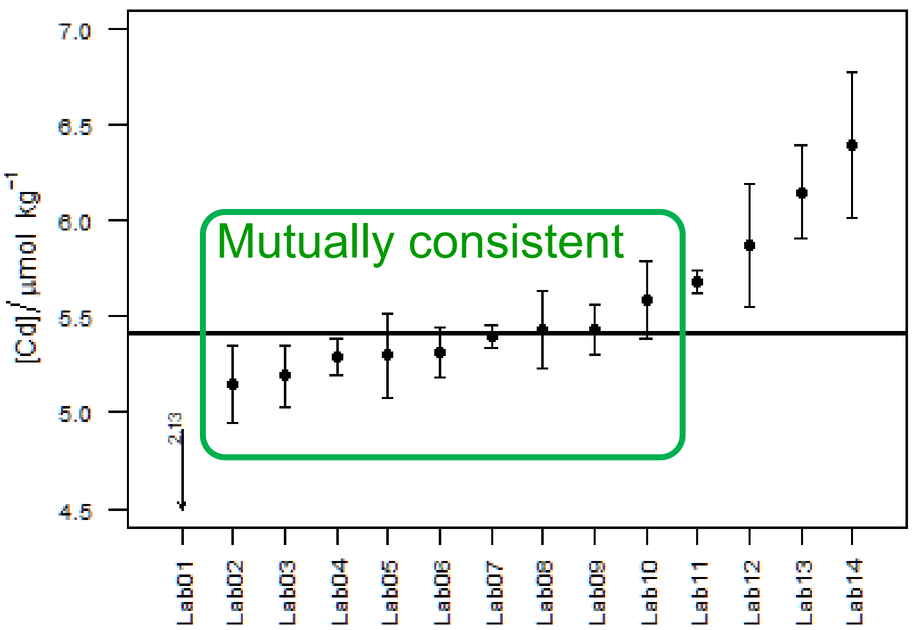
"Consistency plots" – a pairwise summary

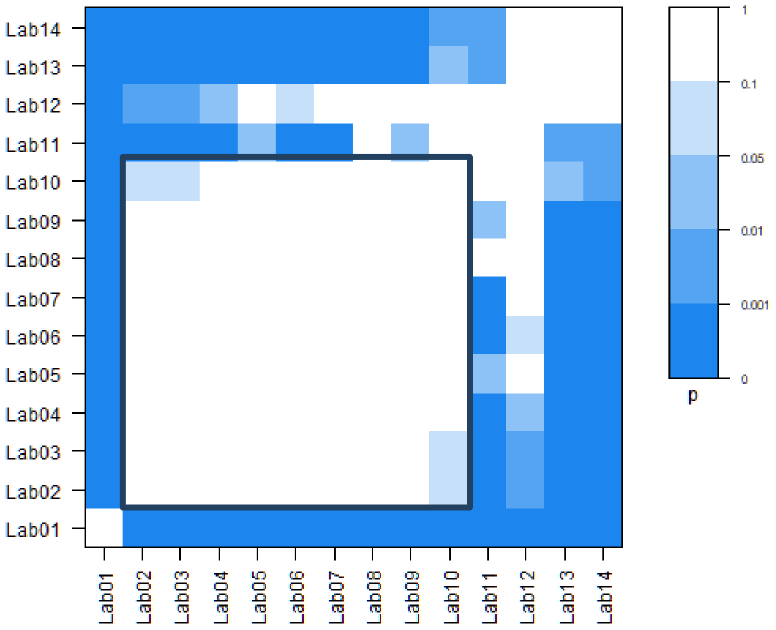
1. Calculate z_{ij} for pairwise differences
$$z_{ij} = \frac{|x_i - x_j|}{u(x_i - x_j)}$$
2. Determine probability (*z*- or *t*-test)
3. Adjust for multiple comparisons
4. Plot as pairwise display



Holm adjustment

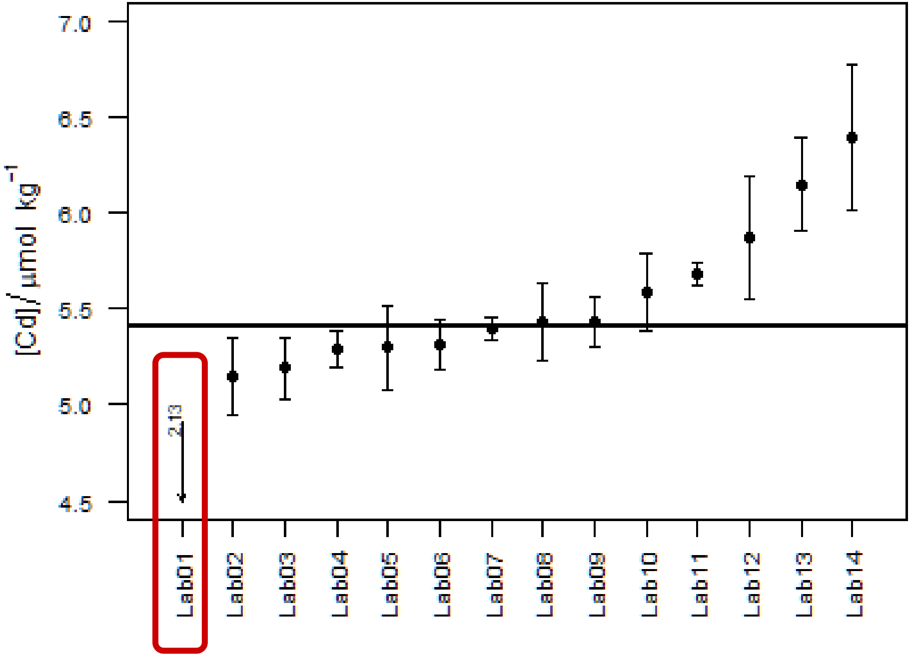
“Consistency plots” – interpretation

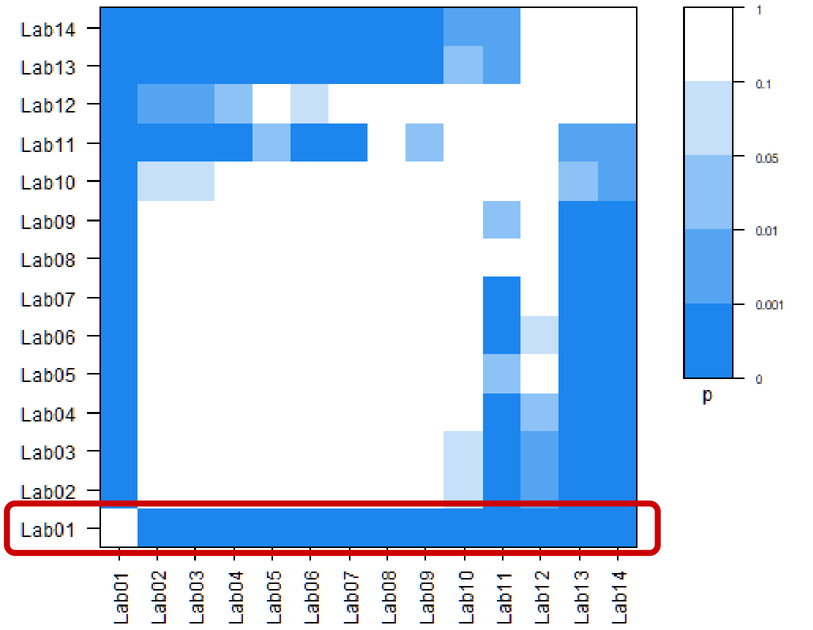




“Consistency plots” – interpretation

NML
UK NATIONAL MEASUREMENT LABORATORY
FOR CHEMICAL AND BIO-MEASUREMENT
HOSTED AT LGC

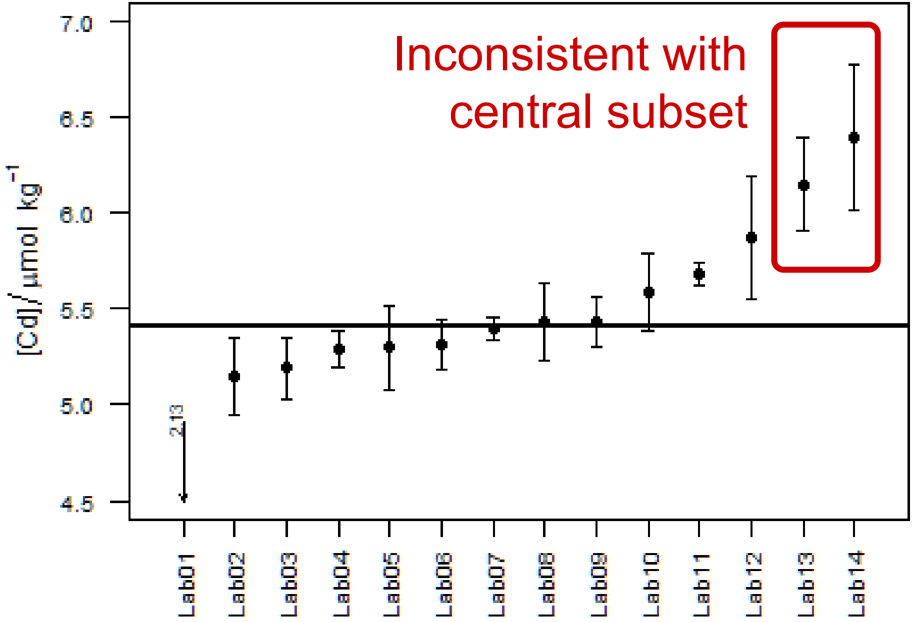


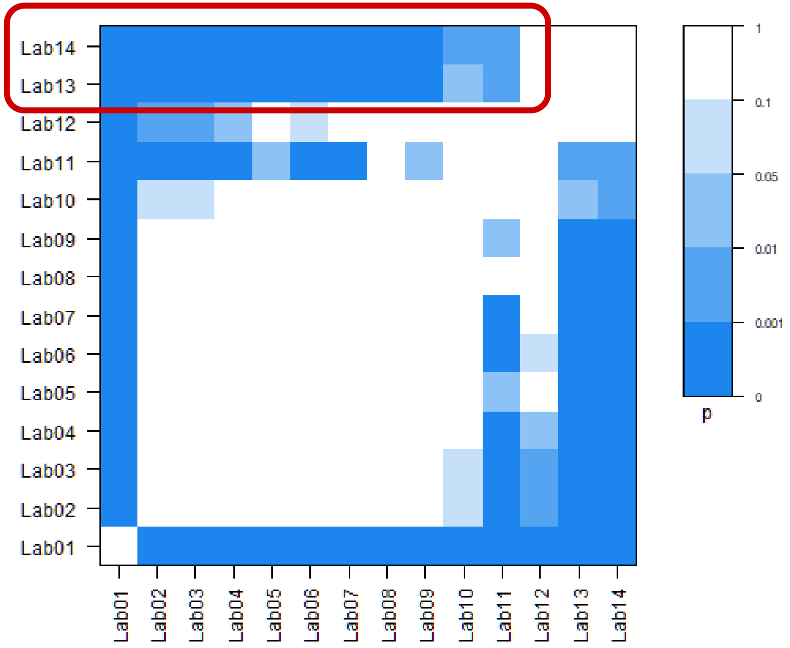


Extreme outlier

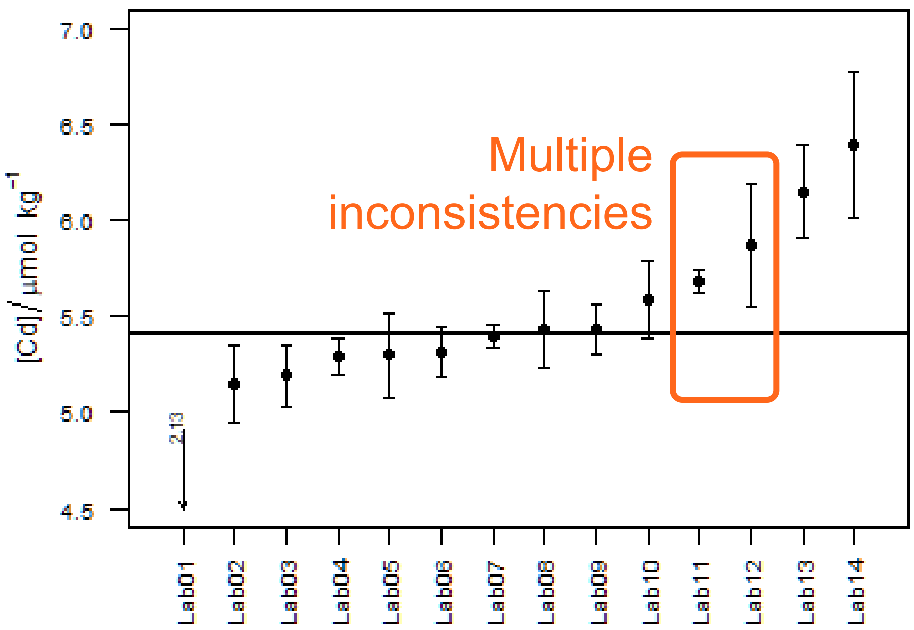
LGC

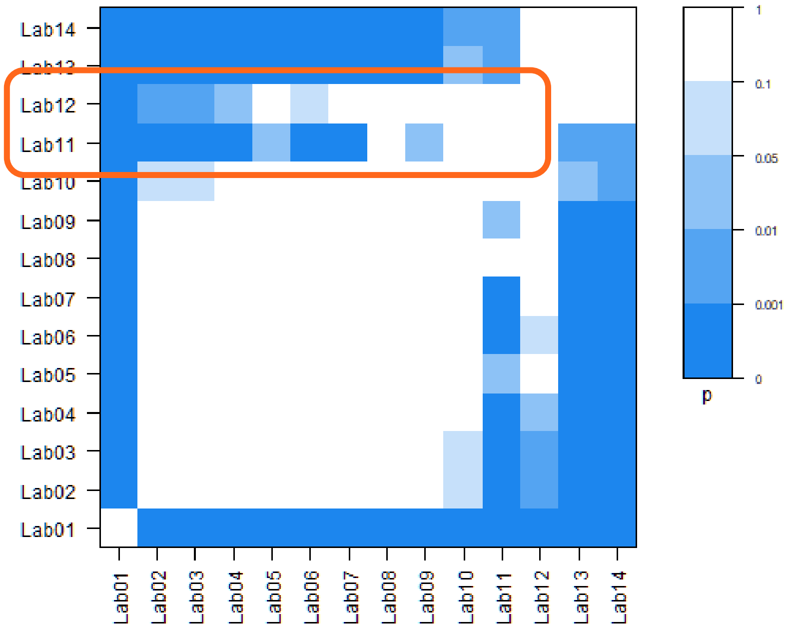
"Consistency plots" – interpretation





"Consistency plots" – interpretation

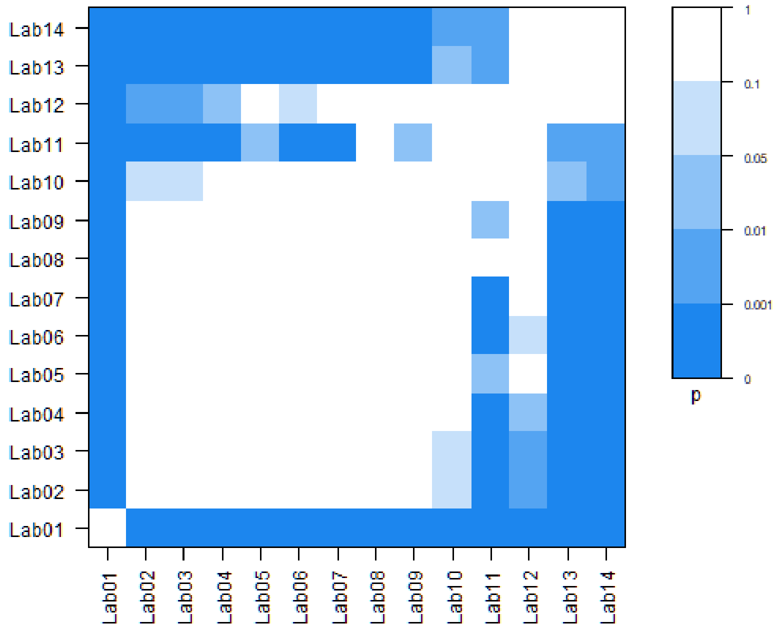




"Consistency plots" – Pros and cons

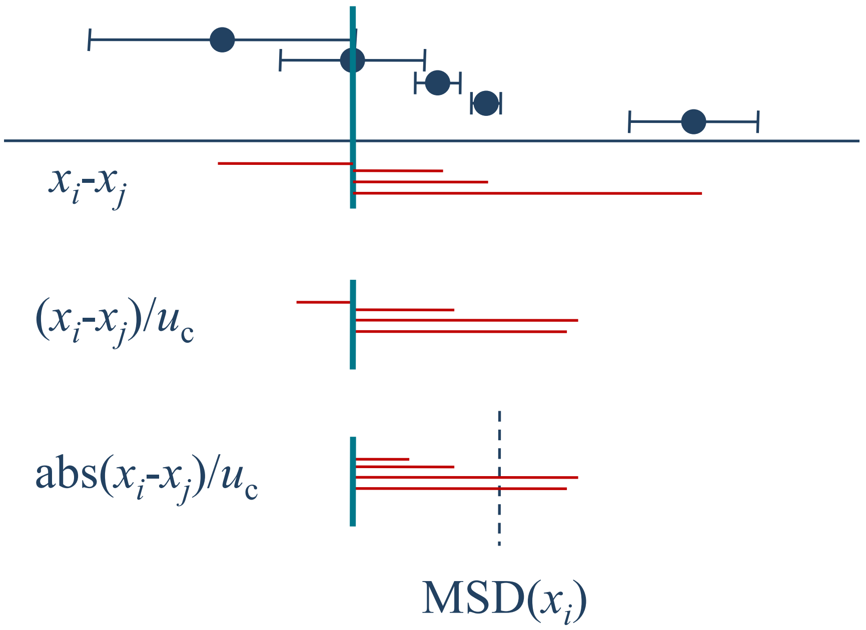
- **Pro**
  - Not reliant on a particular estimator
  - High information content
  - Unambiguous interpretation
    - Adjusted p-values correct for implicit multiple comparisons
  - Relatively easy to explain

- **Con**
  - Adjustment method is a choice
    - Strong adjustment reduces power
  - Not a 'summary'

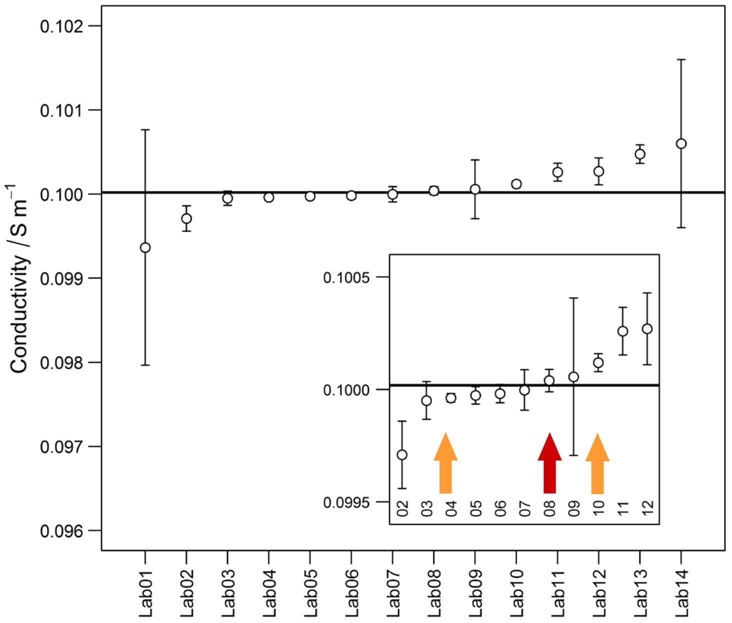


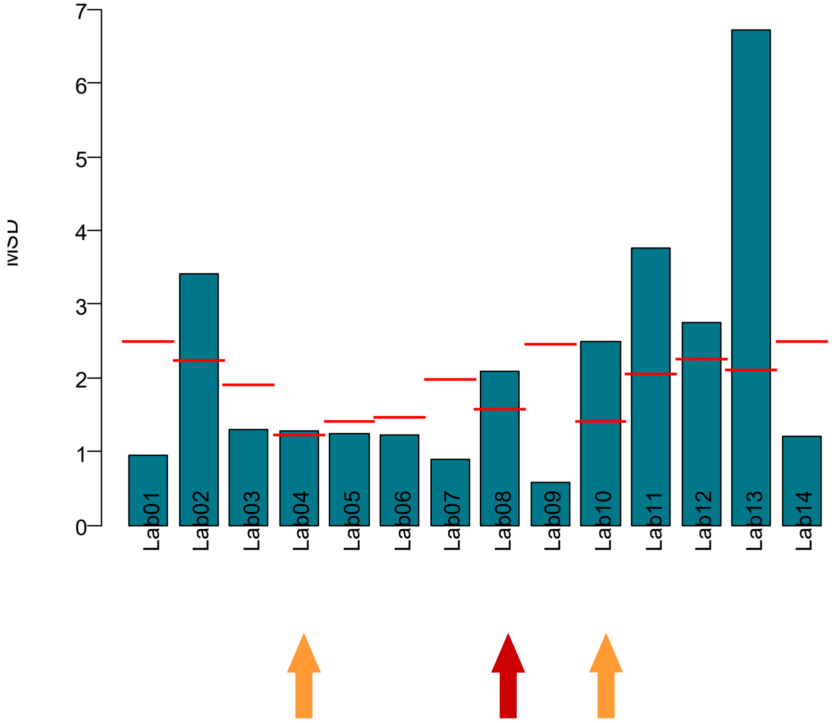
A pairwise summary indicator: Median scaled difference

$$\mathrm{MSD}_i = \operatorname*{med}_{j,i\neq j}\left(\frac{\left|x_i - x_j\right|}{u\left(x_i - x_j\right)}\right)$$



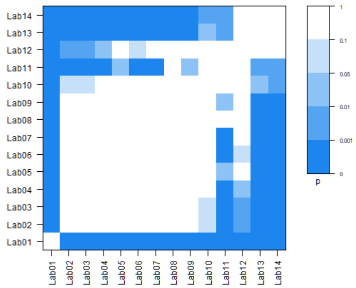
MSD example:
CCQM-P22 - Non-IID bootstrap quantiles

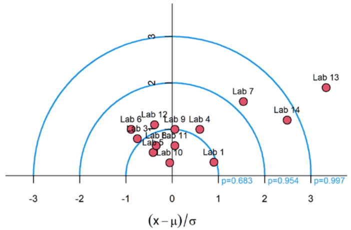


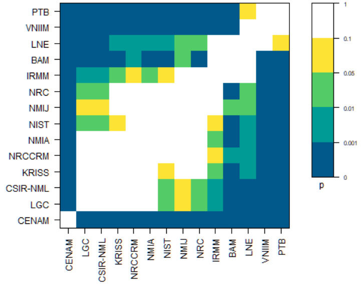


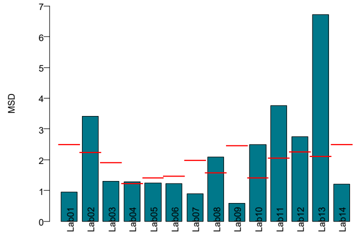
SUMMARY

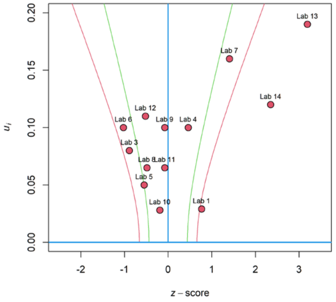
There's more to interlab graphics than dot-and-bar plots





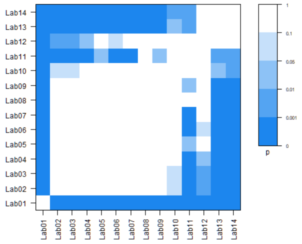




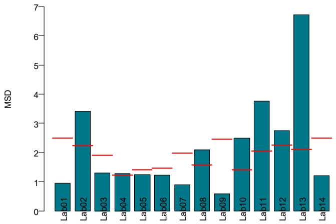


NML
UK NATIONAL MEASUREMENT LABORATORY FOR CHEMICAL AND BIO-MEASUREMENT HOSTED AT LGC

References



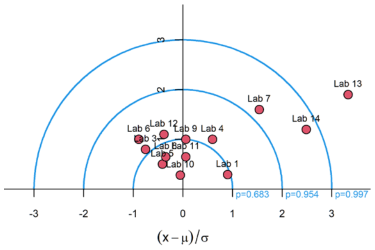
Ellison, S.L.R. Consistency plots: a simple graphical tool for investigating agreement in key comparisons. *Accred Qual Assur* 27, 341–348 (2022). https://doi.org/10.1007/s00769-022-01520-z



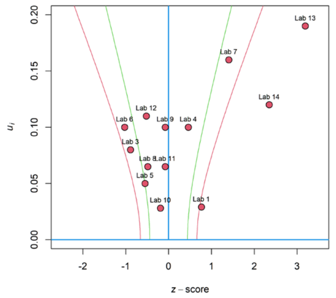
Ellison, S.L.R. An outlier-resistant indicator of anomalies among inter-laboratory comparison data with associated uncertainty *Metrologia* 55 840 (2018)

https://orcid.org/0000-0002-3008-6656

References



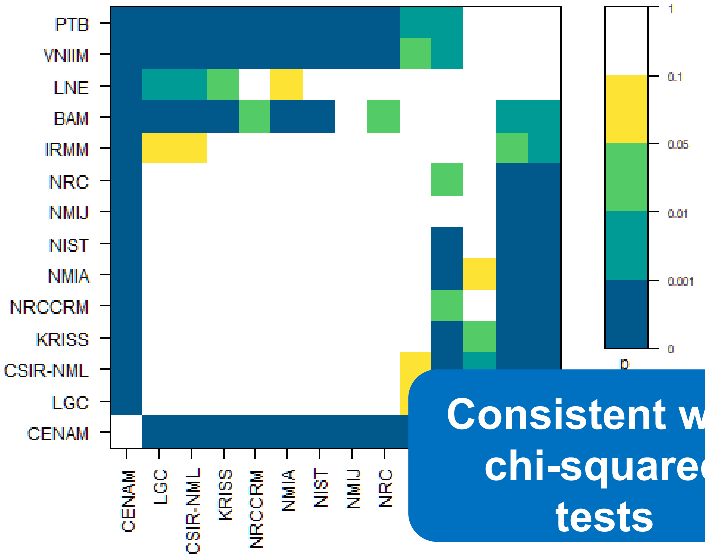
Duewer D. L. , Kline M.C., Sharpless K.E., Brown Thomas J., and Gary K. T. Micronutrients Measurement Quality Assurance Program: Helping participants use interlaboratory comparison exercise results to improve their long-term measurement performance.
Anal. Chem. 71, 1870-1878 (1999)
https://doi.org/10.1021/ac981074k



Cordeiro F., Emons H. and Robouch P. Is the z score sufficient to assess participants' performance in proficiency testing? The hidden corrective action.
Accred Qual Assur 27, 145–153 (2022).
https://doi.org/10.1007/s00769-022-01496-w

Comparing adjustment methods

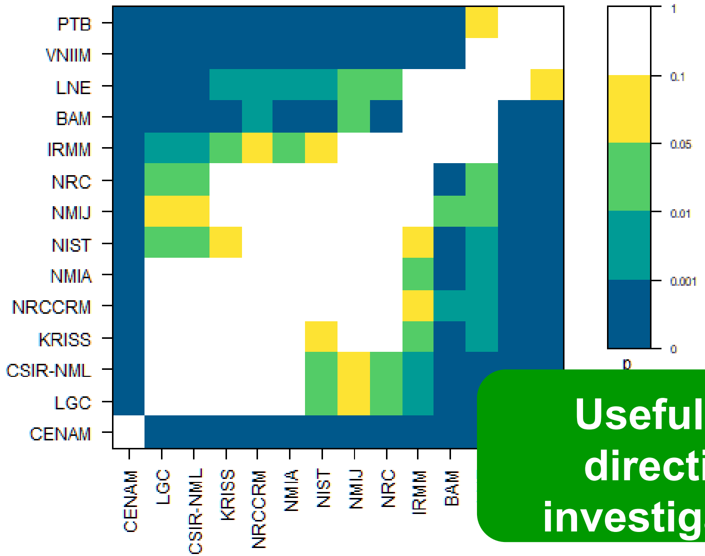
Holm



Consistent with chi-squared tests

Preserves family-wise error rate α

Benjamin-Hochberg

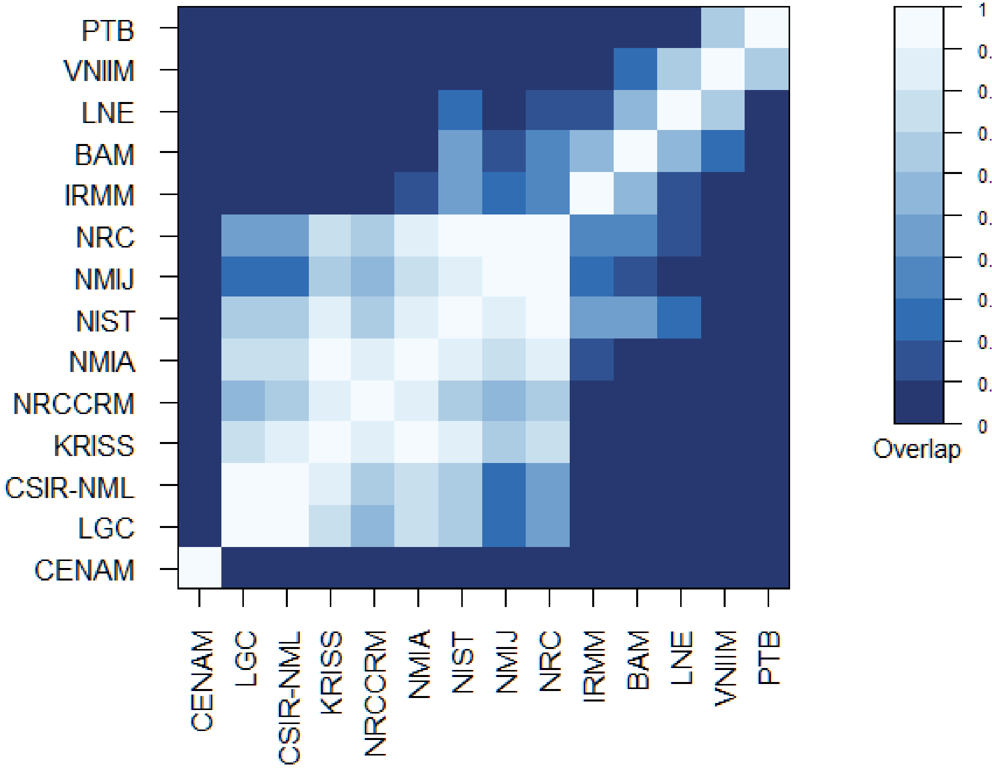


Useful for directing investigation

Controls false discovery rate

Pairwise overlap

- **Plot of mutual distribution overlap**
  (Following Cofino/Quasimeme*)



*See, for example:
Burrell S, Crum S, Foley B, Turner AD (2016), Proficienc
testing of laboratories for paralytic shellfish poisoning toxins in shellfish by QUASIMEME: A review,
Trends in Analytical Chemistry, 75, 10-23,
https://doi.org/10.1016/j.trac.2015.09.004.