Use and treatment of measurement uncertainty in PT schemes

S L R Ellison
LGC Limited, Teddington, UK

Science
for a safer world

Introduction

- Provisions for participant data with measurement uncertainty in ISO 13528:2015
  - Uncertainties for consensus assigned values
  - Scoring methods using assigned value uncertainty
  - Scoring methods using participants' reported uncertainty
  - Examining participant uncertainties
- Consensus values using participants' reported uncertainty
  - Methods and challenges

Statistical procedures in 13528

- Study design
- Assigned value and uncertainty
- Standard deviation for PT
  - "target" standard deviation
- Performance statistics - scoring
- Combining scores
  - Within a PT round
  - Between rounds

Statistical procedures in 13528

- Study design
- Assigned value and uncertainty
- Standard deviation for PT
  - "target" standard deviation
- Performance statistics - scoring
- Combining scores
  - Within a PT round
  - Between rounds

Value assignment methods

Assigning values - uncertainty

- Certified value
- Reference value
  - Unusual in chemistry
- Consensus values (all)
- Formulation
- Read from certificate
- Calculated from CRM by difference
- From method of estimation
- Calculation, with GUM MU

A new model for assigned value uncertainty?

A new model for assigned value uncertainty?

$$u(x_{pt}) = \sqrt{u_{char}^2 + u_{hom}^2 + u_{trans}^2 + u_{stab}^2}$$

- Aligned with ISO Guide 35 (Reference materials)
- Includes explicit allowances for
  - 'Characterisation' – method of assigning value
  - Homogeneity – (possible) differences between test items
  - Transport effects
  - (in)Stability

A new model for assigned value uncertainty?

$$u(x_{pt}) = \sqrt{u_{char}^2 + u_{hom}^2 + u_{trans}^2 + u_{stab}^2}$$

Generally as in 13528:2005

Equivalent to 13528:2005 Annex B with z' score

Expected to be minimised 13528 7.2.2 Note 3*

* "Where this requirement is met, u_{stab} and u_{trans} may be set to zero"

Uncertainty for consensus assigned values

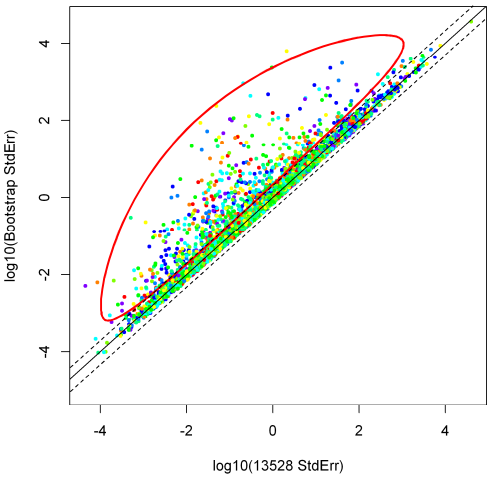
Assigned value uncertainty u_{char}

- Retained as in 2005:

$$u_{char} = 1.25\frac{s^*}{\sqrt{p}}$$

- Based on the distribution of the median for normally distributed data as $p \to \infty$
- *Slightly* conservative for efficient estimators applied to normally distributed data
- Use of estimator efficiency *permitted* with evidence
- Resampling methods ("bootstrapping") suggested for 'fully general approach'

Bootstrapping issues – Example Bootstrapping for H15 ('Algorithm A')



- Most bootstrap standard errors are similar to $1.253s^*/\sqrt{p}$
- Some are **much** larger
- Outlier proportions in bootstrap samples can be extreme

Scoring with assigned value uncertainty

Scoring for simple results

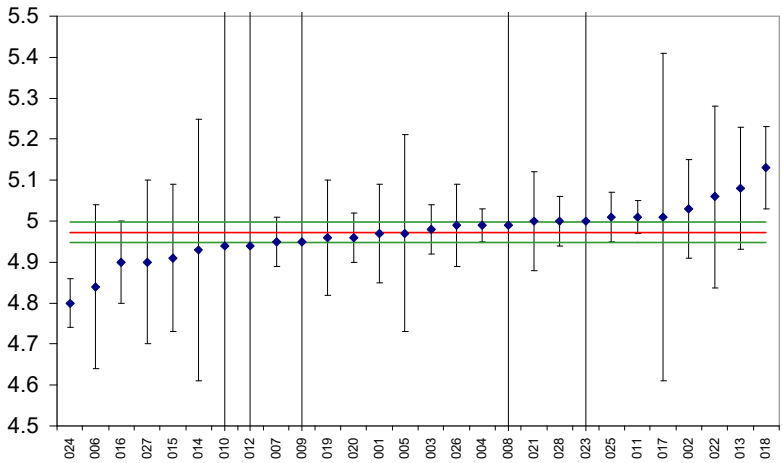
- Z-score
  - |z| < 2 acceptable

$$z = \frac{(x_i - \hat{x})}{\sigma_{\text{pt}}}$$

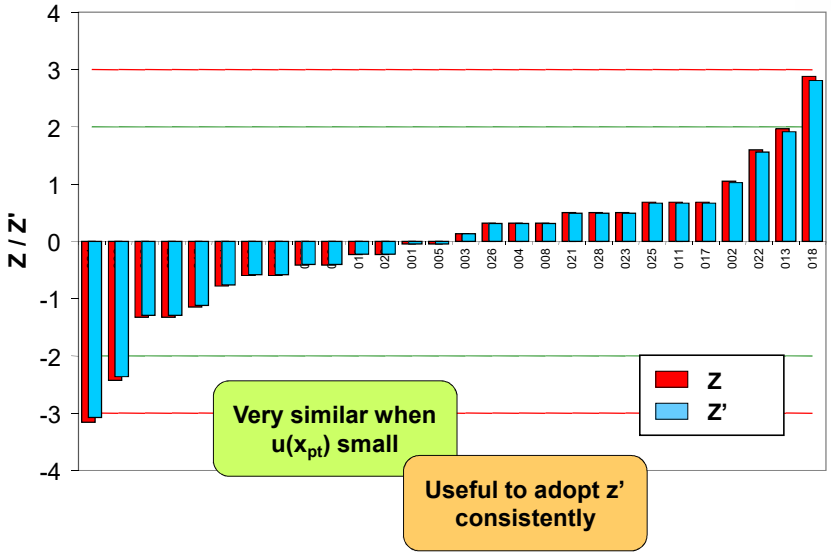
- Extended for MU in assigned value:

$$z' = \frac{(x_i - \hat{x})}{\sqrt{\sigma_{\text{pt}}^2 + u(\hat{x})^2}}$$

Example: Ethanol in beers



Z and Z'



Scoring with participant uncertainty

Scoring with participant uncertainty

'zeta' score

$$\zeta = \frac{(x_i - \hat{x})}{\sqrt{u(x_{pt})^2 + u(x_i)^2}}$$

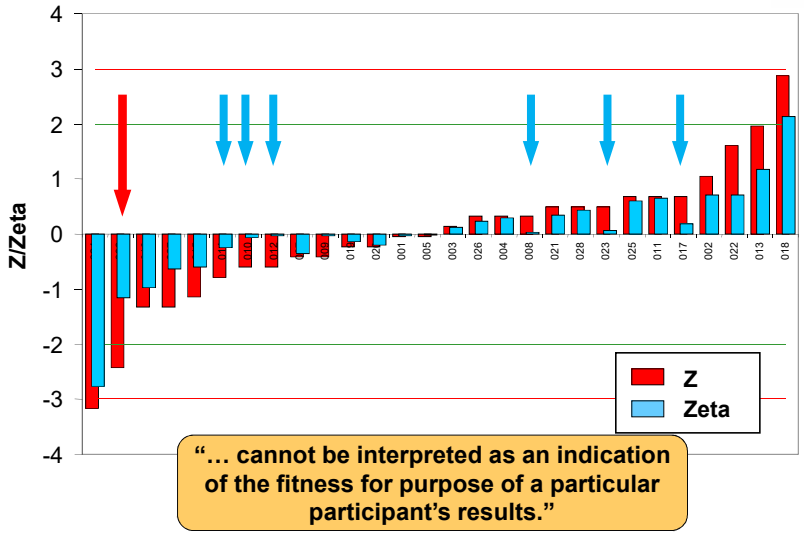
E_n number

$$E_n = \frac{(x_i - \hat{x})}{\sqrt{U(x_{\mathrm{pt}})^2 + U(x_i)^2}}$$

Note use of Expanded uncertainty:

Participant must evaluate *k*

Z and Zeta scores



"… cannot be interpreted as an indication of the fitness for purpose of a particular participant's results."

Examining reported uncertainties

Examining reported uncertainties

- *zeta* and E_n do not directly assess reported uncertainty
  - They assess the **deviation** relative to reported uncertainty
- Overstating uncertainty gives good *zeta* and E_n scores

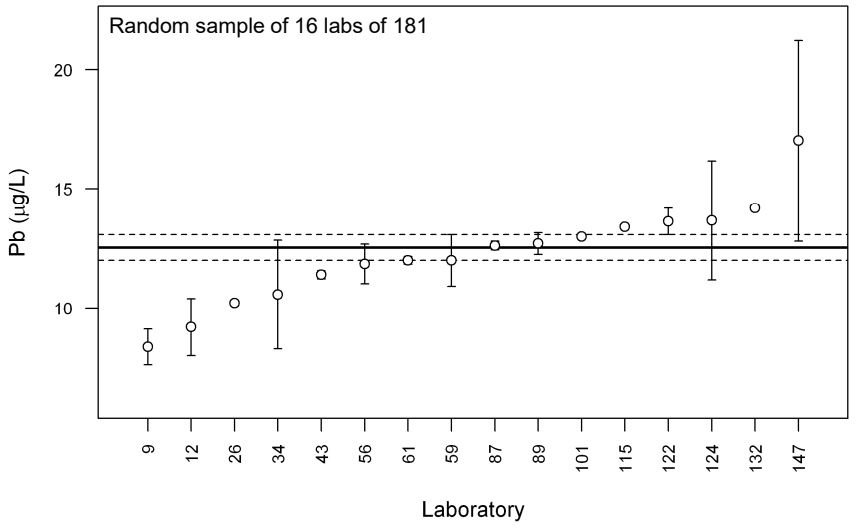
"… should be taken as an indicator of successful performance only if the uncertainties are valid and the deviation (x_i-x_{pt}) is smaller than needed by the participant's customers"

(ISO 13528, 9.7.2)

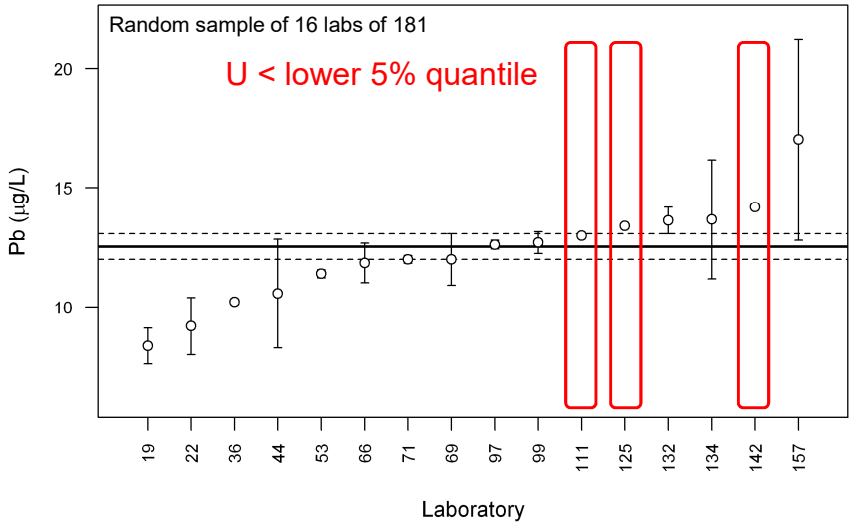
Examining reported uncertainties: New optional provisions

- Mark unrealistically low and high uncertainties
  - $u_i < u(x_{char})$ - unrealistically low
  - $u_i > 1.5\ s^*$ - probably high
- Mark reported uncertainties against quantiles
  - e.g. Lowest 5% and highest 5% of reported uncertainties
- Set limits based on an assumed distribution
  - e.g. Chi-squared, if degrees of freedom are known
- Set limits based on a required measurement uncertainty
  - e.g. some EU regulations set an upper limit for uncertainty

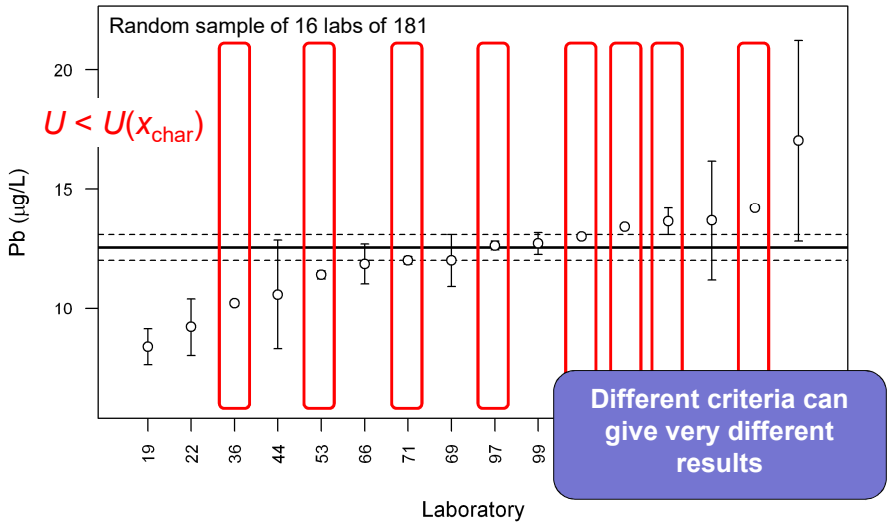
Examining reported uncertainties
Example (Pb in water: ISO 13528:2005)



Examining reported uncertainties
Example (Pb in water: ISO 13528:2005)



Examining reported uncertainties
Example (Pb in water: ISO 13528:2005)



Using reported uncertainties in consensus value estimation

Consensus values for data with reported uncertainties

- Methods used should:
  - include checks for validity of reported uncertainty estimates;
  - use a weighting procedure appropriate for the scale and reliability of the reported uncertainties;
  - allow for the possibility that reported uncertainties might not account fully for the observed dispersion;
  - allow for the possibility of unexpected outlying values for the reported result or the uncertainty;
  - have a sound theoretical basis;
  - **shall** have demonstrated performance sufficient for the purposes of the proficiency testing scheme.

Consensus values for data with reported uncertainties

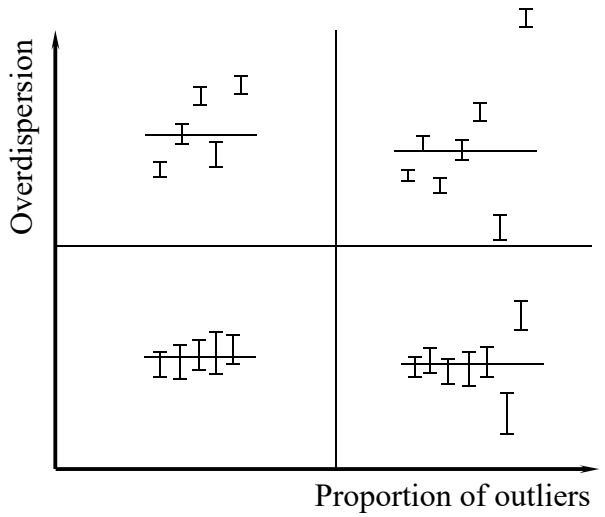
- Refers to a CCQM guideline for methods:

CCQM Guidance note: Estimation of a consensus KCRV and associated Degrees of Equivalence

Version: 10
Date: 2013-04-12
Status: Released for reference

http://www.bipm.org/cc/CCQM/Allowed/19/CCQM13-22_Consensus_KCRV_v10.pdf

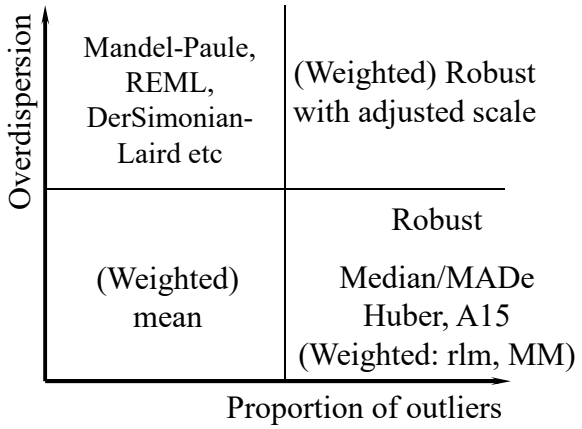
Problems in interlaboratory consensus



Other factors:

- Uncertainty estimate reliability
- Differences in uncertainties
- Number of data points
- Correlation
- Homogeneity
- ....

Recommended Estimators



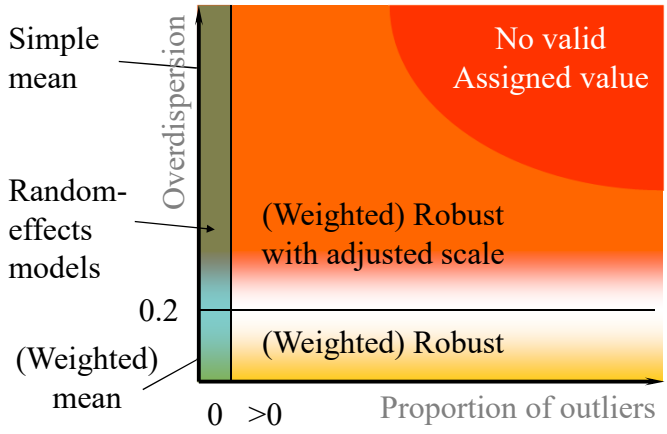
"Most efficient" estimator preferred

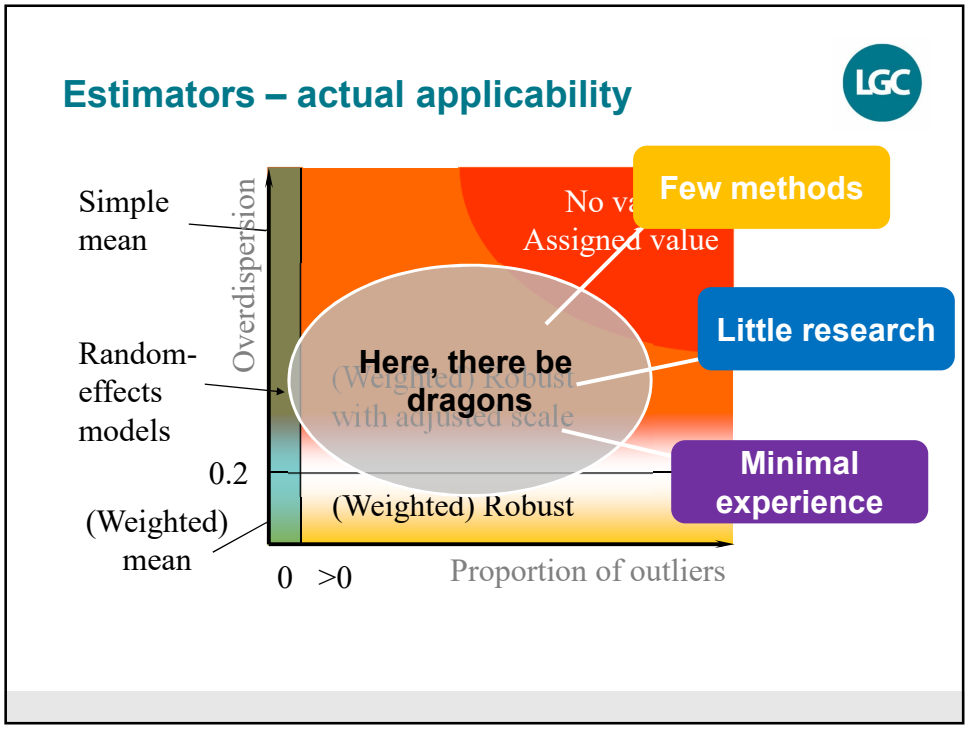
NB: "Weighted" here indicates variance-weighting

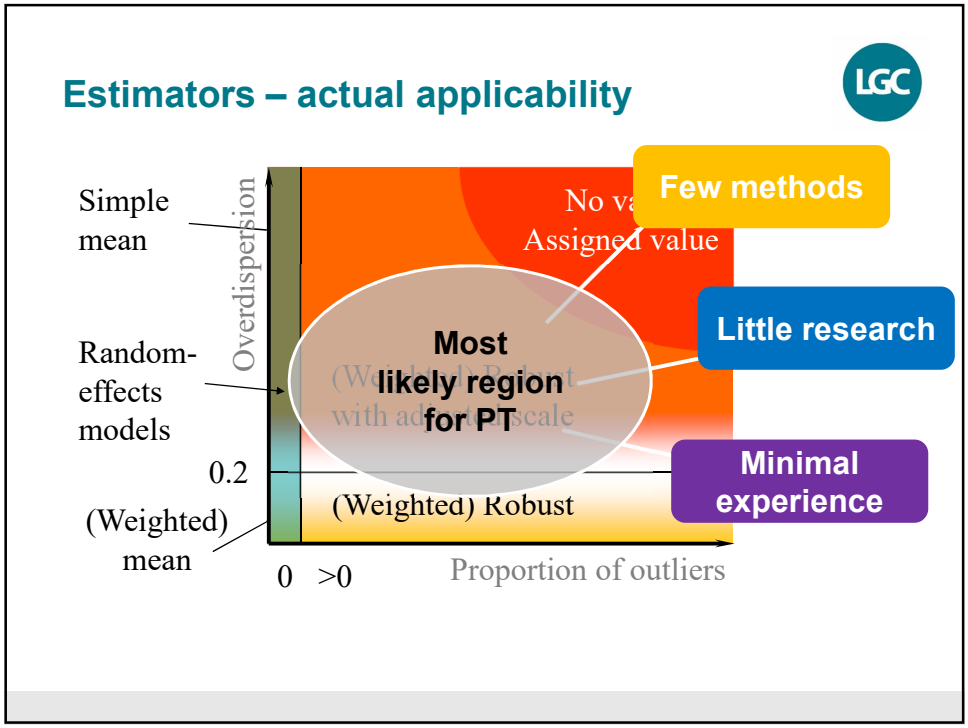
What the different estimator families do

Estimator	Uncertainty*	Properties
Mean	$s/\sqrt{n}$	Ignores lab uncertainties; Limit of added-variance estimators as dispersion increases
Median	$MAD_e\sqrt{\frac{\pi}{2n}}$	Ignores lab uncertainties; very resistant to outliers.
Weighted mean	$\sqrt{\left(\sum \frac{1}{u(x_i)^2}\right)^{-1}}$	Uses lab uncertainties. Limit of added-variance estimators as excess dispersion decreases. **Not recommended**
REML, DSL, M-P, V-R	$\sqrt{\left(\sum \frac{1}{u(x_i)^2+\tau^2}\right)^{-1}}$	Uses lab uncertainties Adds a (constant) variance to model excess variation. **Recommended** when outliers absent
Huber, MM-etc	Based on robust SD or individual weights	Some variants use lab uncertainties Resistant to outliers Excess dispersion accommodated by multiplicative correction

Estimators – actual applicability




Estimators – actual applicability
LGC
Simple mean
Random-effects models
(Weighted) mean
Overdispersion
0.2
0
>0
Proportion of outliers
No va
Assigned value
Few methods
Little research
Minimal experience
Here, there be dragons
(Weighted) Robust
with adjusted scale
(Weighted) Robust


Estimators – actual applicability
LGC
Simple mean
Random-effects models
(Weighted) mean
Overdispersion
0.2
0
>0
Proportion of outliers
No va
Assigned value
Few methods
Little research
Minimal experience
Most likely region for PT
(Weighted) Robust
with adjusted scale
(Weighted) Robust

Conclusions

- More attention to measurement uncertainty in ISO 13528:2015
- New assigned value uncertainty ‘model’ is largely equivalent to existing practice
- Additional provisions for reporting on participant uncertainties
  - No scoring mechanism yet
- Provision for using participant uncertainties in consensus value assignment
  - IF they are known to be reliable