Development of methods for treatment of systematic effects in measurements of quantities realised by SI standards

F. Pavese, Torino, Italy

f.pavese@inrim.it

Aims of the talk

- This was a theme proposed within the JRP on maths at the 2007 EMRP Call
- This theme has been included among the activities of the SIGs of IMEKO TC21, following the last AMCTM Conference
- Open problems will be summarised for future collaboration work

Terms used in the talk

- *Within-laboratory knowledge* –

  Information, not exclusively statistical, available to each laboratory, allowing it to believe that its measurement results and the associated uncertainties obtained using a specific method are ‘correct’ for the intended purpose, i.e. accurately correspond to the intended measurand value.

  It is based on laboratory ‘*prior* information’, on which the uncertainty budget is based.

Terms used in the talk

- *Between-laboratory knowledge* –

  Additional information, not exclusively statistical, that is gained after the results of one laboratory are compared with others.

  The latter are typically the results of other laboratories, but may include also comparison with reference materials or with results obtained using a different method.

  This is typically '*posterior* information' with respect to the '*within*-laboratory knowledge'.

Terms used in the talk

- *Prior vs posterior information*
  - ✓ *Prior* information is the **input** information, including statistical information, to a process of data analysis aimed at providing a (set of) measurement result(s) and its associated uncertainty.
  - ✓ *Posterior* information is the **output** information obtained from a process of data analysis, or predicted on the basis of *prior* information (for stationary processes).

Systematic effects

There is consensus on their existence.

- In fact
  - ✓ corrections for them
  - ✓ comparison of results
  - ✓ consistency tests for results of comparisons

  are deemed necessary.
- They affect, to a different extent, all experimental results

Systematic effects

- The effects arise from the (known and unknown) influence quantities, and are estimated either by means of

  ✓ an (imperfect) model, with parameters (approximately) estimated
    – statistically
    – from first principles

  ✓ a reproducibility study, with results obtained using
    – an (incomplete) variational method
    – literature data or other *prior* sources of information

Problems arise in treatment

A well established distinction:

Repeatability conditions (**repeated** measurements)	Reproducibility conditions (**replicated** measurements)

But, *terms plagued by great confusion*. Two examples:

✓ GUM 3.1.5 "Variations in **repeated** observations are assumed to arise because influence quantities that can affect the measurement result are not held completely constant"

✓ VIM 2.17 "Component of measurement error that in **replicate** measurements remains constant or varies in a predictable manner"

Problems arise in treatment

a) Under repeatability conditions	b) Under reproducibility conditions
Each systematic effect has a fixed value	Each systematic effect shows a value variability

✓The VIM in 2.17, Note 3 "Systematic measurement error equals measurement error minus random measurement error": case (b) apparently includes bias variability in the random error

✓The GUM in 3.2.3 "It is assumed that, after correction, the expectation or expected value of the error arising from a systematic effect is zero" and in its Note "The **uncertainty** of a correction … is a measure of the uncertainty of the result due to incomplete knowledge of the required value of the correction": case (a) is not explicitly considered

Problems arise in treatment

a) Under repeatability conditions	b) Under reproducibility conditions
Each systematic effect has a fixed value: typical of *within*-laboratory knowledge	Each systematic effect shows value variability: typical of *between*-laboratory knowledge

Within-lab believe: zero expectation after correction	*Between*-lab evidence: lab-pair differences from comparison results

Problems arise in treatment

('true' measurand value x_M ; reference value x_R)

Input data
(from *within*-lab info)

Outcomes
(from *between*-lab info)

lab *i* data: x_i, U_i
within-lab: $E(X_1) = x_M$

$x_i - x_j = 0$

$E(B_i) = 0$

lab *j* data: x_j, U_j
within-lab: $E(X_j) = x_M$

Comparison process:

$$Y_{ij} = X_i + B_i - (X_j + B_j) + \varepsilon_{ij}$$

or, $Y_{iR} = X_i + B_i - x_R + \varepsilon_{ij}$

$x_i - x_j \neq 0$
$x_i - x_R \neq 0$

$B = \mathcal{D}(E(B_i))$

For comparison bias $B = \mathcal{D}(E(B_i))$, it may be $E(B) = 0$ (random effect), but for some laboratory biases $E(B_i) \neq 0$ (fixed effect)

*(valid only when **not** comparing different artefacts: $x_{Mi} \neq x_{Mj}$)*

What about expectation of the error arising from a systematic effect?

- Decide in which cases and for which purposes a laboratory fixed effect applies

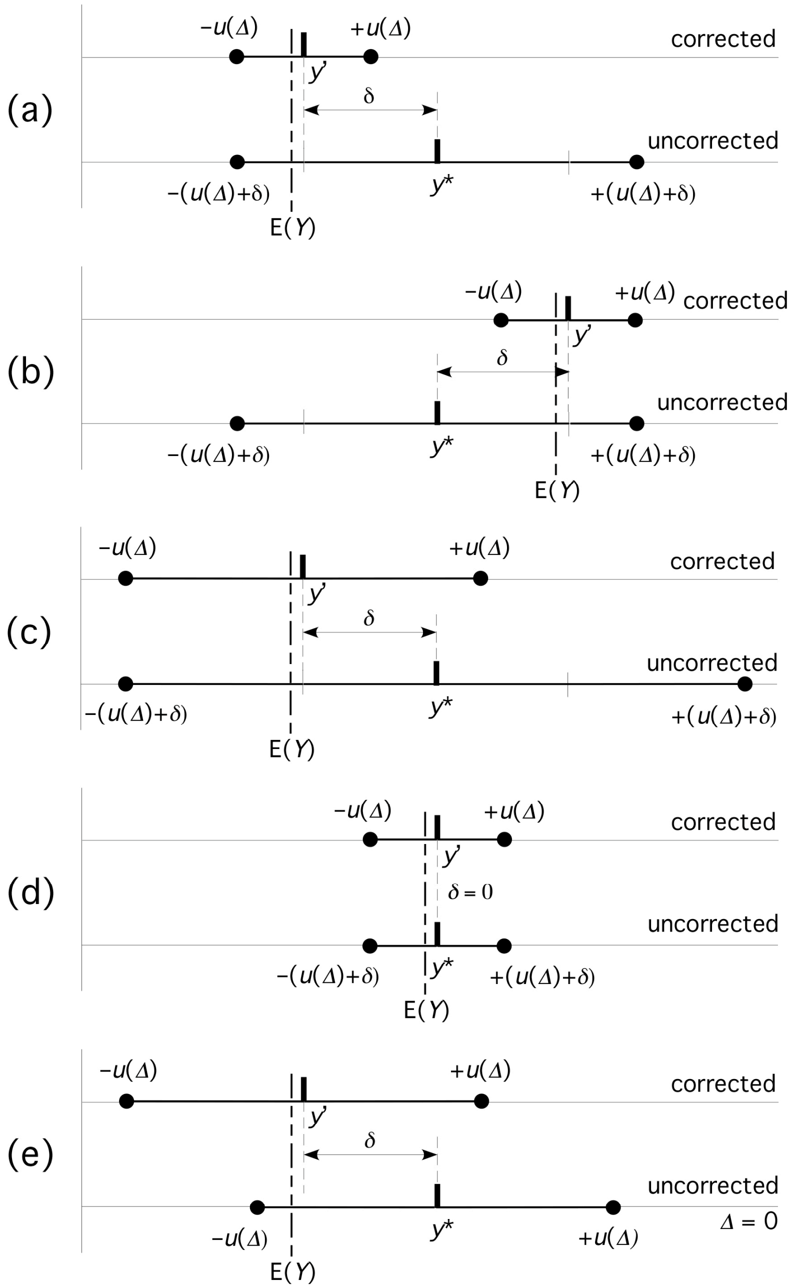
- Decide in which cases and for which purposes a laboratory random effect applies

- Decide in which cases and for which purposes a laboratory mixed effect applies

Depends on the type of comparison and on the framework. Related concept: Outliers.

What to do when a correction is **not** applied

- Allowed in GUM Note to 6.3.1, in exceptional cases for *known* corrections
- The correction can instead be critically not reliable
- The estimated (*but not applied*) bias value comes into the uncertainty component due to bias
  - Several proposed expressions
  - Unreliability of bias estimate is less critical for uncertainty estimate than for ‘correcting’ the measured value
- **This case is different from the case when the estimated bias value is set to <u>zero</u>**



here $\delta = -\Delta$ or $-\Delta/k$

(a) positive bias δ with small correction Δ uncertainty;

(b) negative bias δ with small correction Δ uncertainty;

(c) positive bias δ with large correction Δ uncertainty;

(d) bias δ value *estimated* be equal to zero;

(e) correction value Δ *chosen* be null

E(Y) intended, y^* measured, y' estimated

Aims for possible joint activity

- Develop approaches for providing a *prior* characterisation of uncertainty components in an uncertainty budget so that the likely observable contribution in adjustment exercises, inter-laboratory comparisons, etc., in different time frames can be estimated

  (*Although uncertainty components should be treated equally in providing an overall uncertainty statement, how they manifest themselves in inter-laboratory comparisons and adjustment exercises will be different*).

- Develop appropriate statistical analysis tools to provide a *posterior* state of knowledge about the component effects taking into account the aggregation of knowledge and all relevant information available in the *prior* analyses and uncertainty budgets. These tools (frequentist, Bayesian, fiducial, etc) should be applicable namely to key comparisons, other inter-laboratory comparisons and other metrological studies.