ISO 5725:1994

Accuracy (trueness and precision) of measurement methods and results.

An introduction

Franco Pavese

IMEKO TC21 “Mathematical Tools for Measurements”, Chairman

ISO TC69 “Applications of Statistical Methods”, Member

Italy,

frpavese@gmail.com

What ISO 5725:1994 **is not**

- A standard concerning measurement uncertainty in general
- A standard concerning proficiency tests: they are treated by ISO 13528:2005 and ISO 17034:2010
- A standard gathering new concepts' definitions, except a few: definitions basically come from ISO 3534:1993

What ISO 5725:1994 **is**

- A standard aimed at obtaining an evaluation of accuracy (precision and trueness) of a single specific measurement method.
- A standard concerning a **comparison** exercise between the results of a large number of participants
- The ISO 21748:2004 is a more recent guidance to ISO 5725

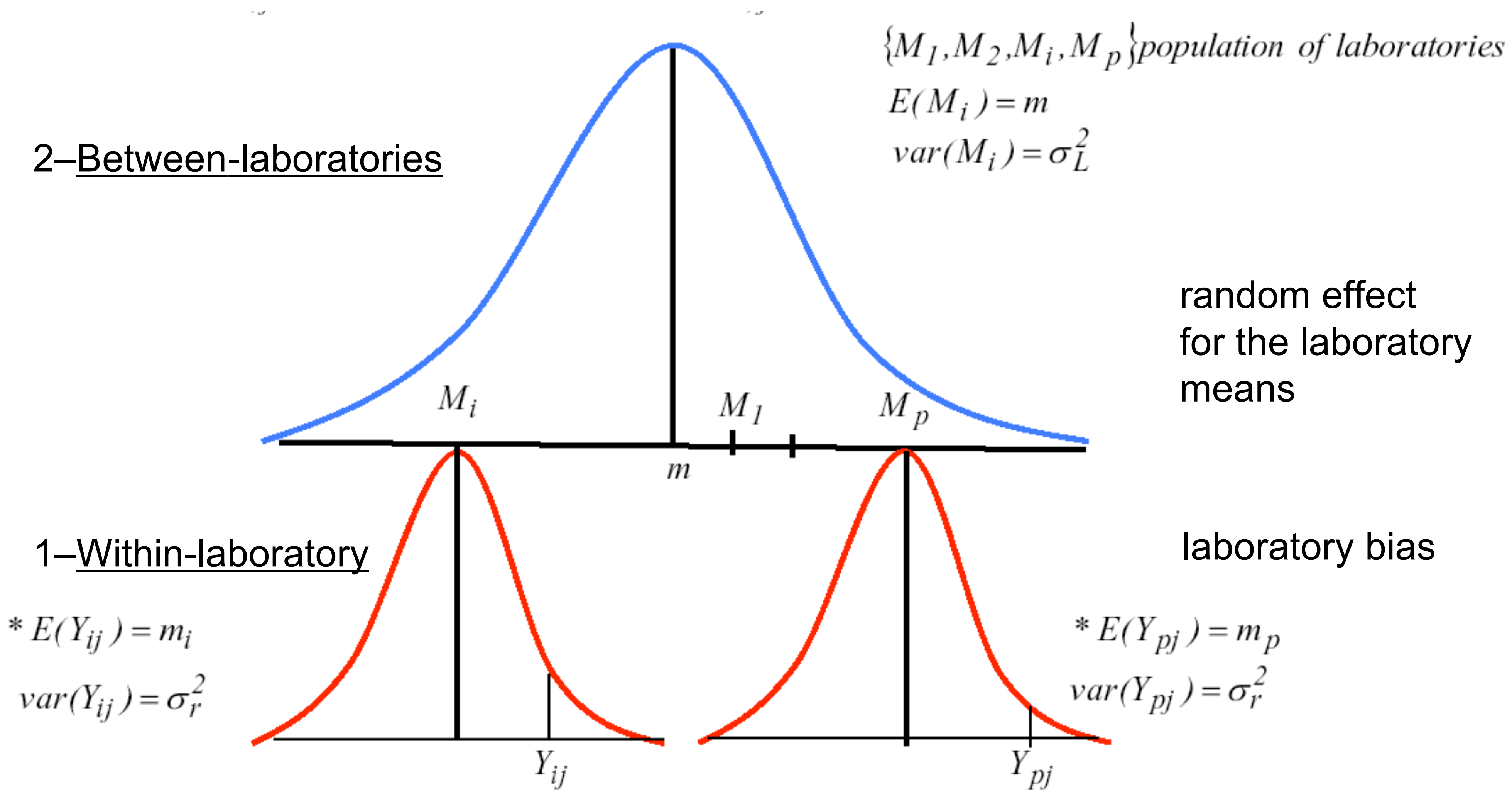
ISO 5725:1994 **vs** GUM:1995

- ISO 5725:1994 is born independent of the GUM:1995. However, the GUM measurement model is applicable to the ISO 5725 with the latter **limited to considering GUM Type A** factors of uncertainty.
- The GUM treats only within-laboratory measurements, while **ISO 5725 deals with *inter*-comparisons** (between-laboratories measurements)
- The GUM introduces and uses the “uncertainty approach” (Type A, Type B), while **ISO 5725 uses the “error approach”** (accuracy, reproducibility, repeatibility)

ISO 5725:1994 **vs** GUM:1995

- The GUM terminology is based on VIM2:1993, while the **ISO 5725 terminology is based on ISO 3534:1993**
- This fact must be carefully taken in consideration when comparing the two documents today, when the terminology references are **VIM3:2008** and **ISO 3534:2006**, respectively, many definitions of which have changed (in both).
- Inconsistencies may arise today due to the latter fact, indicating that **both documents are today obsolete in many respects**.

Basic scheme of an ISO 5725 exercise



ISO 5725:1994 **vs** MRA:1999

- Dealing with *inter*-comparisons, ISO 5725 can also be related to the Key Comparisons of the MRA:1999 and to the proficiency tests standards ISO 13528:2005 and ISO 17034:2010.

- Some of the main features of the *inter*-comparison exercise are different in the three cases.

ISO 5725	MRA KC	ISO 13528
The number of participants is generally sufficiently high and a random choice of them is assumed. Non-hierarchical	The number of participants is generally small and a random choice of them cannot be assumed, Non-hierarchical	Reference materials can be used. Hierarchical
All participants strictly use the method under test	Measurements may be performed with many different methods	Normally, the same method has to be used by all participants
The estimate of the laboratory bias is one of the aims. This bias is assumed to be a *random effect*	NMIs with different levels of uncertainty (up to > 10:1) and using different methods can take part in each exercise	When there are a few participants in a scheme (e.g., fewer than 30), the recommendations of the IUPAC/ CITAC TR should be followed
The estimate of the method bias is *not* one of the aims	Implicitly assumed that each local sample represents the local population	Homogeneity and stability of samples must be checked
The results of the exercise are the general mean of the test and an estimate of the *between*-laboratory variance in addition to the *within*-laboratory variance	A key KCRV is normally—but not always— defined for each exercise, based on the results of the participants, and usually with an associated uncertainty estimate	An assigned value is used, which can be done in several ways (reference or consensus)
The precision of the model is established	The difference of each NMI result and the KCRV is called “degree of equivalence”, which may be a biased estimator	One of several scoring methods can be used to evaluate the results
Outlying results can be defined and rejected	No result can be considered an outlier, nor be discarded	Outlying results can be defined and rejected
From F. Pavese, Mathematical and statistical tools in metrological measurement, 2013, Chapter in Encyclopedia EOLSS, UNESCO. *http://www.eolss.net*	The difference between pair of NMIs is called bilateral degree of equivalence, which is an unbiased estimator and does not require a KCRV to be established	One of the aims is the estimate of the method accuracy and precision and of the laboratory biases

ISO 5725:1994
Needed adjustments

- As to terminology, now the references are ISO 3534:2006 and VIM3:2012
- As to the treatment of uncertainty, now also GUM:1995 and its Supplements should be considered, when applicable
- Some structure and notation problems should be fixed
- From 2009 and 2012 a revision was going on within ISO TC69, but failed to reach a conclusion
- From 2014 it is recognised, also on input from other ISO Committees, the need to restart a revision procedure
- An inquiry about the interest for this revision, and the availability to have part on it, has been launched in June 2014 by the ISO TC69

ISO 5725:1994
an example of notation problem: the model

<table>
<tr><th>Present (1994)</th><th>Tentative (2012)</th></tr>
<tr><td>

$$y = m + B + e$$

“m is the general mean (expectation);
B is the laboratory component of bias under repeatability conditions;
e is the random error occurring in every measurement under repeatability conditions.”
<u>Problems</u>:
–not a standard notation
–does not reflect the two-step procedure (first within-laboratory, then between-laboratories): see previous figure.

</td><td>

$$Y_{i,j} = M_i + E_{i,j}$$

M_i random variable associated with the ith laboratory;
$Y_{i,j}$ random variable associated with the jth observation of the ith laboratory;
$E_{i,j}$ random variable on each observation performed under repeatibility conditions.
Realizations of the variable $Y_{i,j}$:

$$y_{i,j} = m_i + e_{i,j}$$

With $L_i = M_i - m$, one can write

$$Y_{i,j} = m + L_i + E_{i,j}$$

m general mean
L_i random variable associated with the ith laboratory

</td></tr>
</table>