

INTERLABORATORY COMPARISONS. PRACTICAL APPROACH FOR DATA EVALUATION

A.Chunovkina, N.Zviagin, N.Burmistrova

D I Mendeleyev Institute for Metrology, 19, Moskovsky pr. St Petersburg, Russia,
A.G.Chunovkina@vniim.ru

Abstract: This paper deals with evaluation of inconsistent data from interlaboratory studies in case when reference value is known a priori. Applications of different methods for evaluating inconsistent data [1-3] concerning this particular case are discussed. An explicit practical approach is proposed and illustrated by examples.

Keywords: data model, measurement model, inconsistent data, bias, reproducibility, uncertainty.

1. INTRODUCTION

Interlaboratory comparisons are widely used in metrology as an experimental tool for solving different tasks. Concerning their objections three main groups of interlaboratory comparisons are conventionally distinguished: proficiency testing, estimating measurement methods accuracy characteristics, assessment of the parameters of the objects under the investigation.

This paper deals with the first group of interlaboratory comparisons. Each of the above group requires corresponding data evaluation methods. Moreover within one group the methods can differ significantly depending on available information and the accuracy of the measurement data.

To some extent key comparisons of national measurement standards can be also regarded as proficiency studies with the aims to establish the degree of equivalence of measurement standard of the participating laboratories. But unlike the results of key comparison, where participants demonstrate the same level of accuracy, the measurement uncertainties of routine comparisons differ significantly from one laboratory to another. Quite often the interlaboratory comparison results are mutually inconsistent.

We would like to stress that consistency of the measurement data and the model applied is a crucial point for further data evaluation. Methods for data evaluation of interlaboratory comparisons are divided into two groups (see Fig 1).

The first group includes methods for evaluating consistent data. For key comparison of national standards consistency of measurement data means the equivalence of measurement results obtained by participating laboratories, i.e. the equivalence of measurement standards. Degree of equivalence of measurement standards can be established in agreement with CIPM MRA.

In case of key comparisons the detailed information of uncertainty budget is available. This allows providing bias estimates for the results of participating laboratories, based on measurement results of all participants and associated uncertainty budgets [2]. These estimates should not be regarded as degree of equivalence and no further corrections are implied. This additional information resulted from joint evaluation of all comparison data is a matter for within laboratory consideration and further improving the measurement procedure.

The second group comprises methods for evaluating inconsistent data. Mostly these methods are used for evaluating the data of pilot studies and routine interlaboratory comparisons. It should be stressed that all these methods are based on some assumptions which seem to be more or less reasonable for the intended use of applied methods [3]. In case of inconsistent data some quantitative measures of inconsistency (non equivalence) could be proposed.

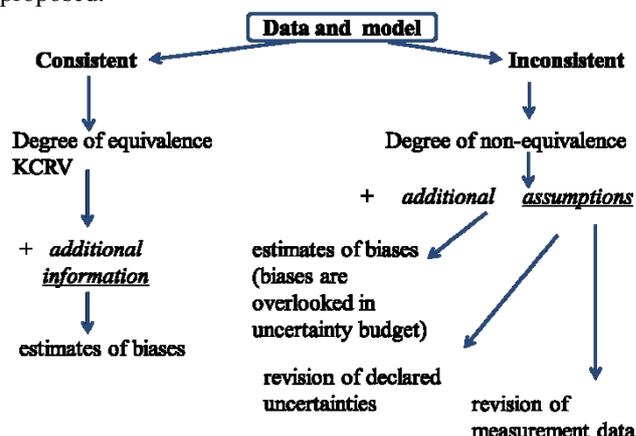


Fig. 1 Two approaches for evaluating measurement data of interlaboratory comparisons

The example of inconsistent data is presented in Fig.2. Generally the methods of inconsistent data evaluation are intended to solve the following tasks:

- to check declared accuracy of the results obtained by the participating laboratories,
- to reveal hidden errors, which weren't taken into account in uncertainty budget by some participating laboratories,
- to estimate corresponding systematic biases,

- to extend measurement uncertainties declared by the participants in case if they were initially underestimated.

There are several approaches and a lot of algorithms for evaluating inconsistent data. In this paper we address only few of them as having some compatibility with the approach discussed [4-7]. The method [4] offers measurement results and associated uncertainties modification to remove model and data non-conformity.

In [4] the validity of the model is assumed. Similar assumption is made in the given paper where the reference value is provided by the reference laboratory with significantly less uncertainty. In the context of this paper the approaches mentioned above are advocated for obtaining meaningful values of measurement uncertainties (if they were originally underestimated by participants). As a result a heuristic practical approach for interlaboratory comparisons inconsistent data evaluating is proposed.

2. FORMALIZATION AND MODELING

Further we will consider evaluation of inconsistent data of interlaboratory studies. The value of the measurand is provided by reference laboratory. Stability of the travelling standard is assumed and is confirmed by series of measurements obtained by reference laboratory. So there could be two reasons for data inconsistency: underestimating measurement uncertainty by some participants or significant systematic hidden bias in some results.

Actually both reasons can be regarded as a result of incorrect uncertainty evaluation. In order to separate these

two reasons we need additional information about meaningful value of measurement uncertainty. In the proposed approach this information is inferred from the data provided by participants. In this sense this approach has some similarity with [4,5]. But unlike [4, 5] the given approach provides estimates on the basis of results from reference group of participating laboratories and doesn't imply moving measuring values themselves.

So the idea is to validate some meaningful value for measurement uncertainty and to consider the measuring values that stand far apart from reference values comparing with the uncertainty as having significant hidden systematic bias.

Further, the following measurement model is considered.

Each laboratory measures the same measurand X :

$$X = X_i \quad (1)$$

and provides measurement result x_i and associated standard uncertainty u_i , $i = 1, \dots, n$:

$$x_i = x + b_i + e_i \quad (2),$$

where x – measurand value, b_i – systematic bias, e_i – random error.

Some laboratories take into account the biases within the uncertainty budget and provide combined standard uncertainty, so data model (2) for their results is reduced to:

$$x_i = x + e_i \quad (2a)$$

The others declare underestimated uncertainties. The problem is that initially these two groups of laboratories can't be separated. The similar models are discussed in [7].

The aim of comparison is just to reveal hidden errors in some data. We assume that reference value of the measurand is provided by reference laboratory: x_{ref}, u_{ref} ; $u_{ref} \ll u_i$.

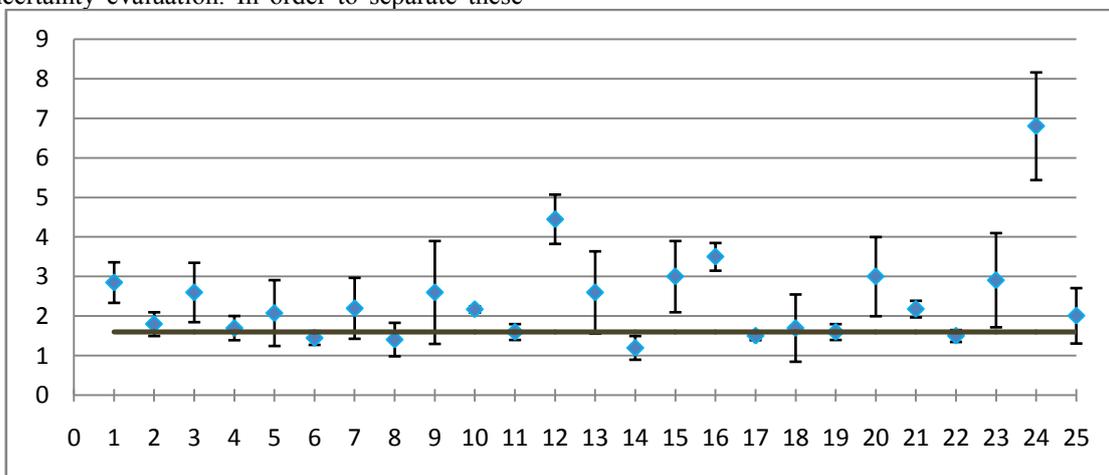


Fig.2 Example of inconsistent data. The error bars show the expanded uncertainties ($k=2$). The line indicates reference value provided by reference laboratory.

3. EVALUATION PROCEDURE

The following evaluation procedure is proposed.

Step 1: Checking the declared uncertainties

To check the declared uncertainty the criterion E_n is used:

$$E_n = \frac{|x_i - x_{ref}|}{2\sqrt{u_i^2 + u_{ref}^2}}$$

If $E_n > 1$, the correspondent measurement result can be regarded as an outlier.

There are two reasons why E_n -test could fail. The first one is the hidden error in measurement result and the second – underestimated uncertainty. The question is how to recognise what case has occurred and what uncertainty value can be regarded reasonable.

We assume that submitted data is characterised by some joint reproducibility. In some cases it can be interpreted as a method reproducibility and can be estimated by experimental standard deviation S . It's proposed to estimate S using not all the submitted results, but only the results from so-called reference group.

Step 2: Determining the reference group

The reference group comprises the measurement results which are consistent with the reference value and are mutually consistent. The first type of consistency means that laboratories correctly calculate the measurement uncertainties. While the second type of consistency means that measuring data form homogeneous set of data that can be characterised by joint reproducibility. For checking the both types consistency presence it's proposed to use two criteria E_n, z_n jointly:

$$z_n = \frac{|x_i - \bar{x}|}{2S}, \quad \bar{x} = \frac{1}{m} \sum_{i=1}^m x_i, \quad S^2 = \frac{1}{m-1} \sum_{i=1}^m (x_i - \bar{x})^2$$

The results with $E_n, z_n < 1$ are included in the reference group, m – number of results in the reference group.

So the reference group includes the measurement results which are mutually consistent (they passed z_n check) and are consistent with the reference value (they passed E_n check).

Step 3: Estimating measurement uncertainty level

The joint estimate of overall measurement uncertainty is based on the deviations of reference group measurement results from the reference value :

$$\widehat{u}^2 = \frac{1}{m} \sum_{i=1}^m (x_i - x_{ref})^2 \quad (3)$$

Step 4: Extending the measurement uncertainties

The uncertainties of the results, that do not pass the E_n criterion initially, are proposed to be extended to the recommended level \tilde{u} . After this the checking procedure is repeated. If the E_n test (with modified uncertainties) fails again, it means that the corresponding result is shifted significantly from the reference value. Consequently the systematic bias can be estimated (step 5) and the corresponding laboratory needs to analyse the measurement procedure carefully.

Step 5: Evaluating the systematic biases

Systematic bias is estimated as deviation from the reference value:

$$b_i = x_i - x_{ref}, \quad u(b_i) = \sqrt{u_i^2 + u(x_{ref})^2} \quad (4)$$

In case when systematic bias is considered as significant the corresponding laboratory should analyze the measurement procedure and reveal reasons for this bias. It can result in recalculating the combined uncertainty u_i^2 .

4. EXAMPLES

The proposed analysis is illustrated by its application to the example data shown in figure 2. The initial data are inconsistent with the reference value (for 11 of 25 participating laboratories $E_n > 1$). The measurement uncertainties of 3 from 11 laboratories were enlarged according to proposed analysis so that the revised data demonstrate the consistency with the reference value. For the last 8 laboratories the systematic biases are recognised significant and the measurement results require further analysis within the laboratory. Fig.3 illustrates the application of the proposed methods. Green dots indicate initial measurement data and red dots indicate revised data with enlarged associated uncertainties.

The given approach is compared with two other approaches for evaluating inconsistent data [4, 5]. In [4] the authors propose to remove inconsistency of data and model by modifying the data and uncertainties so that the modified data provide χ^2 value equals to its expectation (the number of participating laboratories). The modified data are mutually consistent with the reference value due to their strong correlation. This correlation is caused by the analysis applied. Fig.4 illustrates the application of method [4] to the example data. To achieve the consistency the initial measuring data are slightly moved within the associated uncertainties and uncertainties are enlarged. Comparable analysis with the proposed analysis shows similarity in enlarging uncertainties. Actually the same results remain inconsistent with the reference value in sense of E_n checking.

In [5] the author considers procedure for estimating joint reproducibility which is close to one proposed in discussed method. The main difference is that in [5] the estimate of variance is based on all data while in discussed method only the results of the reference group are used. Fig.5 illustrates the application of method [5] to the example data. The results significantly differ from presented in Fig.3 and Fig.4. Fig.5 shows that for the example data the initial measurement uncertainties actually don't influence their revised values. The measurement uncertainties are changed significantly in order to take into account the systematic biases of measurement results. This approach isn't initially intended for revealing systematic biases. The revised data demonstrate good consistence with reference value in sense of E_n checking.

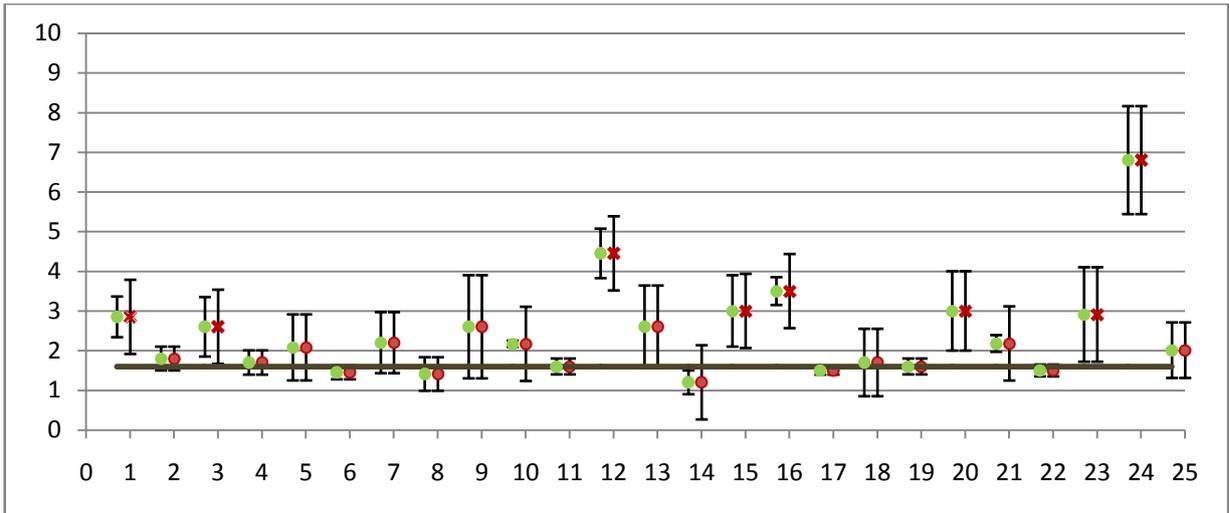


Fig.3 Evaluation of the example data by proposed method

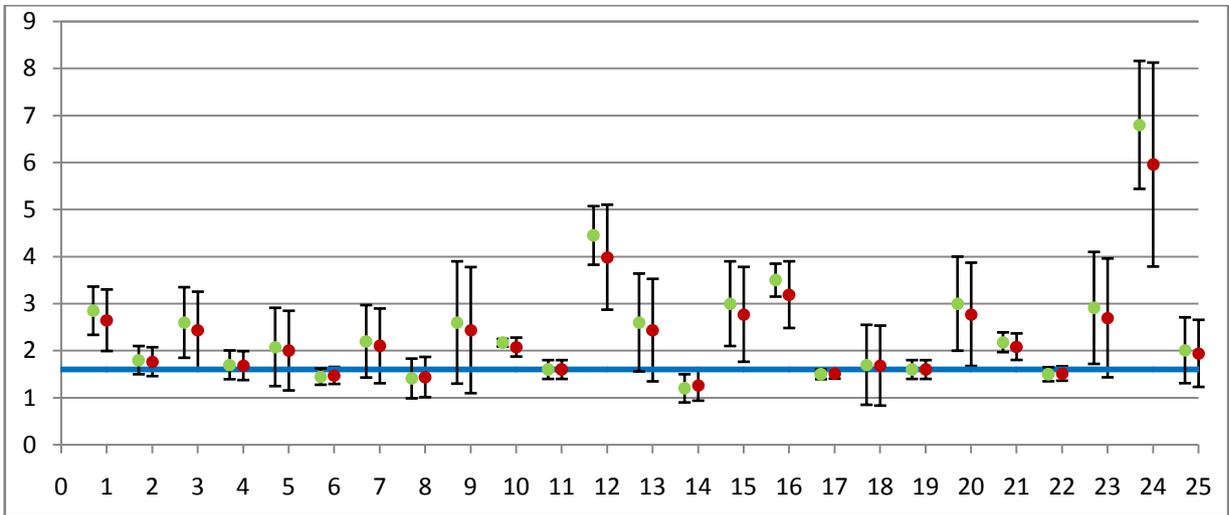


Fig.4 Evaluation of the example data by method [4].

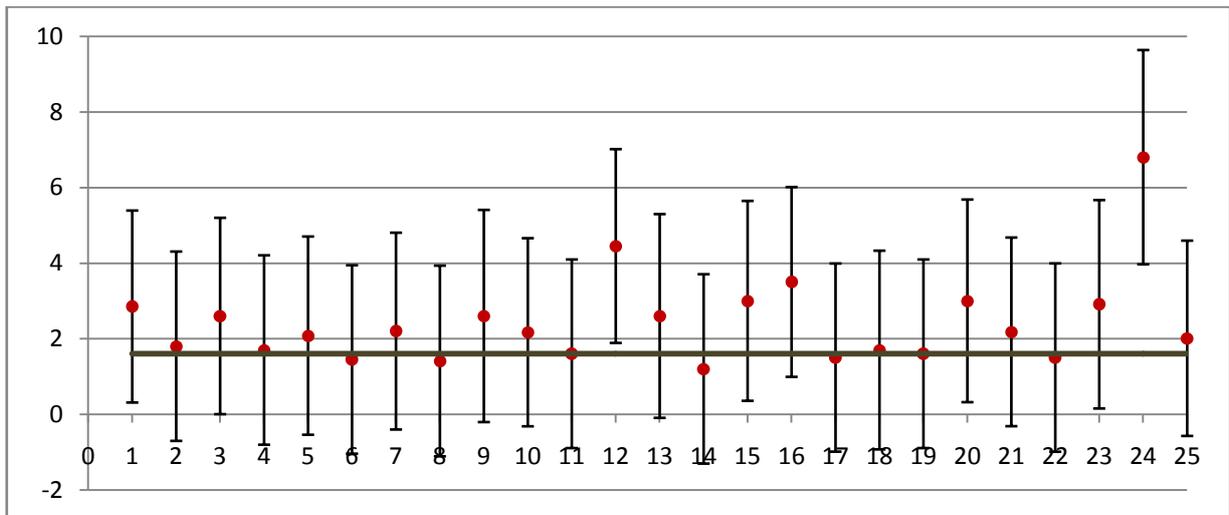


Fig.4 Evaluation of the example data by method [5].

5. CONCLUSION

The paper addresses the problem of data evaluation of comparisons organized with the aim to estimate the quality of measurements provided by participating laboratories. The consistency and inconsistency of the measurement data are regarded as a crucial point for choosing the method for data evaluation.

The paper discusses the evaluation of inconsistent data of interlaboratory comparison. It addresses the following items: checking the declared uncertainties, extending underestimating uncertainties and estimating significant systematic biases of discrepant results.

Extending the measurement uncertainties on the results of comparisons is always a questionable point and requires sound foundation. Meaningful arguments for such extending are discussed. The proposed procedure implies that the reference value of the measurand is provided by the reference laboratory and the dispersion of the results included into the reference group can be characterised by the overall reproducibility. The criteria for construction the reference group are proposed and discussed.

The comparison of proposed procedure with other approaches for evaluating inconsistent data is discussed and illustrated by the examples.

6. REFERENCES

- [1] CIPM MRA. Bureau International des Poids et Mesures 1999
- [2] Chunovkina A. G., Elster C., Lira I., Woeger W. "Analysis of key comparison data and laboratory biases", 2008, *Metrologia*45, pp. 211-6
- [3] Lira I. "Combining inconsistent data from interlaboratory comparisons", 2007, *Metrologia*44, pp. 415-21.
- [4] Weise K. and Wöger W. "Removing model and data non-conformity in measurement evaluation", 2000, *Meas. Sci. Technol.* 11, pp. 1649-58.
- [5] J. Willink R. "Statistical determination of a comparison reference value using hidden errors", 2002, *Metrologia*39, pp. 343-54.
- [6] [4] Cox M. G. "The evaluation of key comparison data: determining the largest consistent subset", 2007, *Metrologia*44, pp. 187-200
- [7] Elster C., Toman B. "Analysis of key comparisons: estimating laboratories' biases by a fixed effects model using Bayesian model averaging", 2010, *Metrologia*47, pp. 113-119