XVII IMEKO World Congress
Metrology in the 3rd Millennium
June 22–27, 2003, Dubrovnik, Croatia

ANALYSIS OF MEASUREMENT OF INTER-LABORATORY COMPARISON THROUGH CONFIDENCE INTERVALS (No 46)

Adriana Horníková

Department for Interchangeable Manufacturing and Industrial Metrology,
University of Technology in Vienna, Vienna, Austria

Abstract – It is possible to compute uncertainty in the form of confidence intervals. In this article is exploited the confidence intervals determination through a simulation study that enables to evaluate uncertainty in the form of confidence intervals for real measurements of inter-laboratory comparison (ILC) even for small numbers of observations.

Here are listed estimation approaches (mathematical-statistical algorithms) for the determination of the consensus value (the true measured value); confidence interval for the true measured quantity in different laboratories; determination (estimate) of the inter-laboratory variance; confidence interval for the inter-laboratory variance; and within-laboratory variance in experiment of the inter-laboratory comparison with homoscedastic as well as heteroscedastic measurements. Different possibilities of evaluation of ILC when the model applied is a linear model with one random effect "laboratory" and estimation procedures are listed and discussed (also describing of the statistical features) in this contribution.

The merit of the simulation study is for a statistician (evaluator of ILC) in better approximation of needed confidence level to obtain the expected result precision in balanced and unbalanced experiment design for homoscedastic as well as heteroscedastic measurements when having small number of observations.

Keywords: Inter-laboratory Comparison, Uncertainty, Confidence Interval

1. INTRODUCTION

The Inter-laboratory Comparison (ILC) experiment is either the precision or trueness experiment. The aim of the ILC evaluation is (in the sense of statistical evaluation):

- determination of the consensus value and its uncertainty
- determination of the repeatability measures
- determination of the reproducibility measures (precision experiment)
- determination of the bias of the measurement method or of the laboratory bias (trueness experiment).

We will use the precision experiment. The precision experiment is an experiment for examination of laboratories, when each laboratory is using the same standard measurement method on identical material (homogenious) is examined. The aim of the precision experiment is to obtain the mean value estimate (consensus value), repeatability, reproducibility and between-laboratory variances (resp. their standard deviations).

2. MODEL OF INTER-LABORATORY COMPARISON (PRECISION EXPERIMENT)

The model of inter-laboratory comparison (ILC) used here coincides with the model of direct repeated measurement of one quantity with p representing participating laboratories. Each laboratory repeats its measurements on the sample in this way, in the i^{th} laboratory n_i times. The measured values $y_{11}, y_{12}, \ldots, y_{1n_1}$, $y_{21}, y_{22}, \ldots, y_{2n_2}, \ldots, y_{i1}, y_{i2}, \ldots, y_{in_i}, \ldots, y_{p1}, y_{p2}, \ldots, y_{pn_p}$ are considered to be realisations of random variables Y_{11}, Y_{12}, …, Y_{1n_1}, Y_{21}, Y_{22}, ..., Y_{2n_2}, …., Y_{i1}, Y_{i2}, …, Y_{in_i}, …, Y_{p1}, Y_{p2}, …, Y_{pn_p} ($i = 1,2,\ldots,p$). Y_{ij} represents measurement in individual laboratories.

If we assume in the mixed linear model only one factor $\boldsymbol{\alpha}$ („laboratory"), the model will be following:

$$\boldsymbol{Y} = \boldsymbol{i}\mu + \boldsymbol{Z}\boldsymbol{\alpha} + \boldsymbol{e} \qquad (1)$$

where $\boldsymbol{Y}$ is the vector of measurements in individual laboratories; μ is the true value of measured quantity; $\boldsymbol{i}$ is the vector of ones (dimension $N = \sum n_i$); $\boldsymbol{\alpha}$ is the vector of random or fixed effect "laboratory"; $\boldsymbol{Z}$ is matrix corresponding to vector $\boldsymbol{\alpha}$ and $\boldsymbol{e}$ is matrix of random errors.

The factor $\boldsymbol{\alpha}$ can be understood as a random effect (random effect model) or fixed effect. Fact, that the factor is assumed to be fixed or random in the model, requires to use different estimation methods and then one obtains different estimates (having also different statistical features).

If we assume $\boldsymbol{\alpha}$ to be random effect and the measurements Y_{ij} are supposed to be normally distributed as:

$$Y_{ij} \sim N\left(\mu, \sigma_{\mathrm{L}}^2 + \sigma_{ei}^2\right), \qquad (2)$$

with mean μ (the true measured value) and dispersion $\sigma_{\mathrm{L}}^2+\sigma_{ei}^2$, where σ_{L}^2 is the inter-laboratory variance (between-laboratory variance) and σ_{ei}^2 is the within-laboratory variance of i[th] laboratory. These variances are also called variance components.

In the first step of ILC evaluation the requirement of normal probability distribution of measurements eq.(2) is tested, using Shapiro-Wilk's, d´Agostin's or Pearson's χ^2-goodness of fit tests. If the requirement of normal probability distribution of measurements in laboratories is not met, non-parametric versions of evaluation (robust) must be used or the procedure for outlier determination and their elimination from the overall evaluation is used [1].

2.1 Homoscedastic and heteroscedastic data

The second requirement is the homoscedasticity of measurements (achievement of comparable repeatability of measurements in the individual laboratories):

$$\sigma_{e1}^2=\sigma_{e2}^2=\cdots=\sigma_{ep}^2, \tag{3}$$

where σ_{ei}^2 is the within-laboratory variance of i[th] laboratory and can be tested using Cochran's, Bartlett's or Hartley's tests. In a case of measurements do not meet this condition, they are heteroscedastic [2].

2.2 Evaluation of the ILC

The aim of the ILC evaluation is to determine the consensus value $\hat{\mu}$ and variance components $\sigma_{ei}^2, \sigma_{\mathrm{L}}^2$ (and their confidence intervals). The consensus value is obtained in unbalanced experiment (balanced as well) with homoscedastic measurements using the generalised least squares method as:

$$\hat{\mu}=\left(\boldsymbol{i}_N^T\,\mathrm{var}(\boldsymbol{Y})^{-1}\boldsymbol{i}_N\right)^{-1}\boldsymbol{i}_N^T\,\mathrm{var}(\boldsymbol{Y})^{-1}\boldsymbol{Y}=\frac{\sum_{i=1}^{p}\frac{n_i\overline{Y}_i}{\sigma_e^2+\sigma_{\mathrm{L}}^2 n_i}}{\sum_{i=1}^{p}\frac{n_i}{\sigma_e^2+\sigma_{\mathrm{L}}^2 n_i}} \tag{4}$$

where $\overline{Y}_i=\frac{1}{n_i}\sum_{l=1}^{n_i}Y_{ij}$ and $\mathrm{var}(\boldsymbol{Y})$ is the covariance matrix of observations. In a case that the variance components are not known, one of the following methods must be used to estimate them: MINQUE, Henderson's method, maximum likelihood, restricted maximum likelihood or GLSE [3-6]. Following the similarity we can say, that most of these estimates of the consensus value differ in the weight that's in variance components [7].

If $\sigma_{\mathrm{L}}^2=0$, the random effect „laboratory" does not have any influence on results. This statement is tested using a statistical test [8], and this can be done also using a confidence interval. In this contribution is preferred the procedure of testing using the confidence intervals as it is more concise and efficient, because the uncertainty as a measure is also defined in the sense of a confidence interval.

The estimation of the confidence interval for the inter-laboratory variance was proposed by several authors: Tukey-Williams, Thomas-Hultquist, Burdick-Eickman, Hartung-Knapp, Wald which gave confidence intervals suitable only for homoscedastic measurements [9,10].

2.2 The Confidence Interval for the Between-laboratory Variance after Burdick and Eickman

This confidence interval for the between-laboratory variance is derived on the principle of the confidence interval after Tukey and Williams and is [9,10]:

$$\left\langle\left(\frac{\tilde{n}\cdot L}{1+\tilde{n}\cdot L}\right)\cdot\frac{(p-1)MS3}{\chi^2_{1-\alpha/2}(p-1)};\left(\frac{\tilde{n}\cdot U}{1+\tilde{n}\cdot U}\right)\cdot\frac{(p-1)MS3}{\chi^2_{\alpha/2}(p-1)}\right\rangle \tag{5}$$

where the lower and upper boundaries are

$$L=\max\left\{0,\frac{MS3}{MS_E}\cdot\frac{1}{F_{1-\alpha/2}(p-1;N-p)}-\frac{1}{n_{\min}}\right\}$$

$$U=\min\left\{0,\frac{MS3}{MS_E}\cdot\frac{1}{F_{\alpha/2}(p-1;N-p)}-\frac{1}{n_{\max}}\right\}$$

and mean squares

$$MS3=\frac{\sum_{i=1}^{p}\left(\overline{Y}_i-\frac{1}{p}\sum_{i=1}^{p}\overline{Y}_i\right)^2}{p-1}$$

$$MS_E=\frac{\sum_{i=1}^{p}\sum_{j=1}^{n_i}\left(Y_{ij}-\overline{Y}_i\right)^2}{N-p}$$

and

$N=\sum n_i$, $\tilde{n}=\frac{p}{\sum_{i=1}^{p}\frac{1}{n_i}}$ is the harmonic mean of repeatings and $n_{\min}$ is the minimum number of repeating, $n_{\max}$ is the maximum number of repeatings, $\chi^2_{\alpha/2}(p-1), \chi^2_{1-\alpha/2}(p-1)$ are the critical values of the chi-square distribution and the critical values of the Fisher distribution are $F_{1-\alpha/2}(p-1;N-p), F_{\alpha/2}(p-1;N-p)$.

3. SIMULATION ANALYSIS

A simulation analysis was done regarding the estimate differences for heteroscedastic measurements using the following methods: maximum likelihood, Mandel-Paule and modified Mandel-Paule's method to estimate the consensus value and variance components [2-6].

3.1 Simulation Experiment Configuration

The set-up of experiments was combined as:

- the number of laboratories is one of three values: 5 (few), 10 (middle), 20 (a lots of laboratories);
- measured quantity (μ): 10 (small) or 100 (great);
- the number of replicates of measurement in individual laboratory is either 5 or 15 in balanced model, or in unbalanced model (2,4,6,8,10 or 2,4,6,8,10,12,14,16,18,20 resp. 2,2,3,4,5,6,7,8,9,10,11, 12,13,14,15,16,17,18,19,20), marked as „unbalan“;
- within-laboratory variance (σ_{ei}^2): constant 0,1 or 1 (for the mean value 10) and 1 or 10 (for the mean value 100), or in increasing order (1, 4, 7, 10, 15 or 1, 2, 3, 5, 7, 9, 11, 14, 17, 20 resp. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20), marked „differ“;
- between-laboratory variance (σ_L^2): 0,1-times within-laboratory variance (eventually its average value) or 10-times within-laboratory variance. If the within-laboratory variance in individual laboratories is different (marked as „differ“) it is needed to calculate with the average value from all within-laboratory variances, that is $\overline{\sigma}_{ei}^2$.

Table I Simulation analysis in experiment with μ =100; σ_{ei}^2=[2 4 6 8 …] (differ); σ_L^2=0,1·$\overline{\sigma}_{ei}^2$.

Experiment	Maximum Likelihood estimate	Estimate after Mandel and Paule	Modified estimate after Mandel and Paule
Lab.x repeats	Consensus value $\hat{\mu}$ (empirical variance for $\hat{\mu}$ in 10 000 simulations)		
5 x 5	100,2026 (0,3908)	100,1593 (0,3718)	100,1738 (0,3715)
5 x 15	100,1710 (0,2034)	100,1508 (0,2015)	100,1632 (0,2001)
10 x 5	100,1196 (0,2446)	100,0880 (0,2330)	100,0931 (0,2339)
10 x 15	100,0698 (0,1284)	100,0644 (0,1275)	100,0666 (0,1272)
20 x 5	100,0694 (0,1546)	100,0559 (0,1503)	100,0578 (0,1510)
20 x 15	100,0450 (0,0833)	100,0427 (0,0831)	100,0436 (0,0830)

Table II Simulation analysis in experiment with μ =10; σ_{ei}^2=[2 4 6 8 …] (differ); σ_L^2=0,1·$\overline{\sigma}_{ei}^2$.

Experiment	Maximum Likelihood estimate	Estimate after Mandel and Paule	Modified estimate after Mandel and Paule
Lab.x repeats	Consensus value $\hat{\mu}$ (empirical variance for $\hat{\mu}$ in 10 000 simulations)		
5 x 5	10,0636 (0,0364)	10,0498 (0,0356)	10,0542 (0,0353)
5 x 15	10,0548 (0,0219)	10,0484 (0,0213)	10,0524 (0,0213)
10 x 5	10,0389 (0,0247)	10,0296 (0,0233)	10,0312 (0,0234)
10 x 15	10,0301 (0,0136)	10,0275 (0,0134)	10,0286 (0,0134)
20 x 5	10,0179 (0,0152)	10,0142 (0,0151)	10,0148 (0,0152)
20 x 15	10,0138 (0,0083)	10,0130 (0,0083)	10,0133 (0,0083)

Table III Simulation analysis in experiment with μ =100; σ_{ei}^2=[2 4 6 8 …] (differ); σ_L^2=10·$\overline{\sigma}_{ei}^2$.

Experiment	Maximum Likelihood estimate	Estimate after Mandel and Paule	Modified estimate after Mandel and Paule
Lab.x repeats	Consensus value $\hat{\mu}$ (empirical variance for $\hat{\mu}$ in 10 000 simulations)		
5 x 5	101,1317 (13,8)	101,1258 (13,8)	101,1312 (13,8)
5 x 15	101,0028 (13,2)	101,0008 (13,2)	101,0027 (13,2)
10 x 5	100,5125 (8,5)	100,5100 (8,5)	100,5113 (8,5)
10 x 15	100,5293 (8,4)	100,5288 (8,4)	100,5292 (8,4)
20 x 5	100,2394 (5,3)	100,2389 (5,3)	100,2392 (5,3)
20 x 15	100,2852 (5,1)	100,2851 (5,1)	100,2851 (5,1)

Table VI Simulation analysis in experiment with μ =100; σ_{ei}^2=10; σ_L^2=0,1·$\overline{\sigma}_{ei}^2$.

Experiment	Maximum Likelihood estimate	Estimate after Mandel and Paule	Modified estimate after Mandel and Paule
Lab.x repeats	Consensus value $\hat{\mu}$ (empirical variance for $\hat{\mu}$ in 10 000 simulations)		
5 x unbalan	100,0579 (0,7078)	100,0722 (0,6519)	100,0715 (0,6719)
10 x unbalan	100,0135 (0,2536)	100,0250 (0,2523)	100,0246 (0,2551)
20 x unbalan	100,0101 (0,1185)	100,0170 (0,1230)	100,0170 (0,1236)

Table V Simulation analysis in experiment with μ =100; σ_{ei}^2=[2 4 6 8 …] (differ); $\sigma_L^2 = 0{,}1 \cdot \overline{\sigma}_{ei}^2$.

Experiment	Maximum Likelihood estimate	Estimate after Mandel and Paule	Modified estimate after Mandel and Paule
Lab.x repeats	Consensus value $\hat{\mu}$ (empirical variance for $\hat{\mu}$ in 10 000 simulations)		
5 x unbalan	100,1489 (0,4253)	100,1366 (0,3655)	100,1450 (0,3778)
10 x unbalan	100,0631 (0,1764)	100,0647 (0,1647)	100,0666 (0,1666)
20 x unbalan	100,0372 (0,1122)	100,0369 (0,1077)	100,0376 (0,1085)

Table VI Simulation analysis in experiment with μ =100; σ_{ei}^2=[2 4 6 8 …] (differ); $\sigma_L^2 = 10 \cdot \overline{\sigma}_{ei}^2$.

Experiment	Maximum Likelihood estimate	Estimate after Mandel and Paule	Modified estimate after Mandel and Paule
Lab.x repeats	Consensus value $\hat{\mu}$ (empirical variance for $\hat{\mu}$ in 10 000 simulations)		
5 x unbalan	100,9714 (13,7)	100,9728 (13,6)	100,9754 (13,6)
10 x unbalan	100,3317 (8,8)	100,3316 (8,8)	100,3317 (8,8)
20 x unbalan	100,2305 (5,0)	100,2303 (5,0)	100,2303 (5,0)

Table VII Simulation analysis in experiment with μ =10; σ_{ei}^2=[2 4 6 8 …] (differ); $\sigma_L^2 = 0{,}1 \cdot \overline{\sigma}_{ei}^2$.

Experiment	Maximum Likelihood estimate	Estimate after Mandel and Paule	Modified estimate after Mandel and Paule
Lab.x repeats	Consensus value $\hat{\mu}$ (empirical variance for $\hat{\mu}$ in 10 000 simulations)		
5 x unbalan	10,0421 (41,4)	10,0390 (35,6)	10,0414 (36,7)
10 x unbalan	10,0147 (17,1)	10,0147 (15,9)	10,0153 (16,1)
20 x unbalan	10,0106 (10,6)	10,0104 (10,3)	10,0106 (10,4)

Table VIII Simulation analysis in experiment with μ =10; σ_{ei}^2=[2 4 6 8 …] (differ); $\sigma_L^2 = 10 \cdot \overline{\sigma}_{ei}^2$.

Experiment	Maximum Likelihood estimate	Estimate after Mandel and Paule	Modified estimate after Mandel and Paule
Lab.x repeats	Consensus value $\hat{\mu}$ (empirical variance for $\hat{\mu}$ in 10 000 simulations)		
5 x unbalan	10,3811 (1,3864)	10,3819 (1,3856)	10,3827 (1,3851)
10 x unbalan	10,1540 (0,8505)	10,1540 (0,8503)	10,1541 (0,8503)
20 x unbalan	10,0817 (0,4964)	10,0819 (0,4964)	10,0819 (0,4964)

Table IX Simulation analysis in experiment with μ =10; σ_{ei}^2=1; $\sigma_L^2 = 0{,}1 \cdot \overline{\sigma}_{ei}^2$.

Experiment	Maximum Likelihood estimate	Estimate after Mandel and Paule	Modified estimate after Mandel and Paule
Lab.x repeats	α needed for covering the μ confidence interval in 9500 incident		
5 x 5	0,0001	0,00005	0,00005
5 x 15	0,001	0,001	0,001
10 x 5	0,0005	0,005	0,005
10 x 15	0,01	0,01	0,01
20 x 5	0,02	0,02	0,02
20 x 15	0,02	0,03	0,03

Table X Simulation analysis in experiment with μ =100; σ_{ei}^2=10; $\sigma_L^2 = 0{,}1 \cdot \overline{\sigma}_{ei}^2$.

Experiment	Maximum Likelihood estimate	Estimate after Mandel and Paule	Modified estimate after Mandel and Paule
Lab.x repeats	α needed for covering the μ confidence interval in 9500 incident		
5 x unbalan	0.00005	0.00005	0.00005
10 x unbalan	0.001	0.001	0.001
20 x unbalan	0.01	0.01	0.01

Table XI Simulation analysis in experiment with μ =100; σ_{ei}^2=10; σ_L^2=10·$\overline{\sigma}_{ei}^2$.

Experiment	Maximum Likelihood estimate	Estimate after Mandel and Paule	Modified estimate after Mandel and Paule
Lab.x repeats	α needed for covering the μ confidence interval in 9500 incident		
5 x unbalan	0.00005	0.00005	0.00005
10 x unbalan	0.01	0.01	0.01
20 x unbalan	0.02	0.02	0.02

3.1 Confidence Interval for the Consensus Value

In the tables I to VIII are determined the point estimators of the true measured value obtained as average values from 10 000 simulations of the consensus value and its variance using three different estimators, that are maximum likelihood estimator, after Mandel and Paule, modified after Mandel and Paule are listed for different experiments and different numbers of laboratories and repeatings.

In tables IX to XI are considered the interval estimators. Results of the simulation analysis are evaluated and listed in tables in such a way that it is obvious which confidence level is necessary to obtain 95 % covering of the theoretical mean value by the confidence interval for different numbers of laboratories and repeating (balanced and unlabanced experiment).

4. RESULTS OF THE SIMULATION ANALYSIS

The simulation analysis shows that the differences in estimates of consensus value and the length of the confidence interval for the true value for three methods of estimation (maximum likelihood, Mandel-Paule and modified Mandel-Paule's method) are minimal, independent of the structure of the experiment (balanced and unbalanced) and for different types of measurements (homoscedastic or heteroscedastic).

The simulation analysis offers proof that the consensus value (its variance) is independent of the true value (see and compare tables I and II) and only minimally influenced by the inter-laboratory variance (compare tables I II), but considerably affected by the within-laboratory variance as well as whether measurements are homoscedastic or not.

The tables IX-XI obtained in the simulation analysis makes it possible to determine the α value in a way that the confidence interval for the true value reaches the confidence level of 95% in both an unbalanced experiment with heteroscedastic measurement. It makes it possible to determine if the laboratories are significantly "biased" by each other, which is one of very important problems in evaluation of ILC.

For the evaluation of confidence intervals of inter-laboratory variance the simulation analysis proves that the Burdick-Eickman interval is suitable even for heteroscedastic measurements (balanced and unbalanced experiment). It is possible to determine α in such a way that the obtained confidence interval covers at confidence level 95 % the true value of the inter-laboratory variance.

REFERENCES

[1] G. Wimmer, R. Palenčár, V. Witkovský: Use and evaluation of measurements. Spracovanie a vyhodnotenie meraní. STU, Bratislava, 2001

[2] G. Wimmer: Bio-statistics (Model with random efects) Bioštatistika (Model s náhodnými efektmi). MATFYZ Bratislava, 1999

[3] J. Mandel, R. C. Paule: Interlaboratory Evaluation of a material with Unequal Numbers of Replicates, Anal. Chem. no. 42, vol. 11, pp. 1194-1197, Nov. 1970.

[4] R. C. Paule, J. Mandel: Consensus Values and Weighting Factors. Journal of Research of the National Bureau of Standards no. 87, vol. 5, pp. 377-385, 1982.

[5] A. L. Rukhin, M. G. Vangel: Estimation of a Common Mean and Weighted Means. Journal of the American Statistical Association, no. 441, vol. 93, pp. 303-308, March 1998

[6] A.L.Rukhin, B.J.Biggerstaff, M.G. Vangel: Restricted maximum likelihood estimation of a common mean and the Mandel-Paule algorithm, J. of Statistical Planning and Inference vol. 83, pp. 319-330, 2000.

[7] V. Witkovský, G. Wimmer: On Statistical Models for Consensus Values. Proceedings of the 3rd Measurement Conference, 14-17 May 2001 Smolenice SR.

[8] G. Wimmer: Between Groups Variance Component Interval Estimation in Inter-laboratory Comparison. 4th International Conference on Mathematical Statistics ProbaStat 2002, 4-8 February 2002 Smolenice SR.

[9] J. Hartung, K.H. Makambi, D. Argacs: An extended ANOVA F-test with Applications to the Heterogeneity Problem in Meta-Analysis, Biometrical journal no. 43, vol. 2, pp. 135-146, Feb. 2001.

[10] J. Hartung, G. Knapp: Confidence intervals for the between group variance in the unbalanced one-way random effects model of analysis of variance. J Statist Comput simul vol. 65, pp. 311-323, 2000.

Author: Ing., Mgr., Adriana Horníková, PhD, Department. for Interchangeable Manufacturing and Industrial Metrology, TU Vienna, Karlsplatz 13, A-1040 Wien Austria, Tel.: 0043 1 58801 31143, Tel.: 00421 2 54411 258 (home), Fax: 0043 1 58801 31196, E-mail: ahornik@pobox.sk.