

Stability Tests in the Process of Organization of Interlaboratory Comparison

Miloš Ujlaky^{1,2}, Teodor Tóth², Ing. Anton Frič^{1,3}, Daniel Kysler^{1,3}, Miroslav Dovica²

¹ Slovak Legal Metrology, Hviezdoslavova 31, Banská Bystrica, Slovakia

² Department of Biomedical Engineering and Measurement, Faculty of Mechanical Engineering, Technical University of Košice, Letná 1/9, 042 00 Košice, Slovakia

³ Institute of Automation, Measurement and Applied Informatics, Faculty of Mechanical Engineering, Slovak University of Technology in Bratislava, Nám. slobody 17, 812 31 Bratislava, Slovakia

Abstract: Success in programs of interlaboratory comparisons is based on two basic pillars. In addition to the inseparable share of the laboratory and the quality of the activities performed, the quality of the organized program of interlaboratory comparison also affects the final success. Organizing tests is a time consuming process. It includes several important stages that precede the laboratory's participation in the program itself. One of the most important stages is the selection of an object suitable for the purpose of comparative measurements. One of the criteria for selecting an interlaboratory comparison item is its stability. The article focuses on the basis of ILC, the causes of unsuccessful results, areas where a stability test is required and where not, and one of the possible approaches to checking the stability of a measuring instrument. The proposed approach is defined by the measurement model, described by methodological steps and was applied in practice, where the results were evaluated and analysed in the discussion. In conclusion, it can be stated that the proposed approach is particularly suitable for areas where it is possible to obtain a relatively large amount of data by automating the collection, as it is an application of statistical approaches. Even with this approach, cost-effectiveness must not be forgotten, with an impact on the final price for participation in the ILC. The situation therefore requires the search for optimal solutions in terms of the degree of risk and the cost of stability tests.

Keywords: Interlaboratory comparison, ILC provider, ILC scheme, criteria for participation, stability tests, repeatability, reproducibility, calibration and measurement capability

1. Introduction

The basic tool for proficiency testing is an interlaboratory comparison (ILC). Comparison, focused on a specific quantity (length, temperature, etc.), on the area of activity (calibration, measurement, testing) and the type of measuring instrument is called the ILC program. The focus of the program is based on the requirements of laboratories and the possibilities of the ILC provider. The requirements are usually focused on the purpose of the ILC, the type and range of the measuring instrument (ILC item), and the quality of the laboratory represented by its CMC (Calibration Measurement Capability). The task of the provider is to examine the possibilities of organizing the ILC program according to the above requirements and to set a criterion in the form of measurement uncertainty, which represents the quality level of the future ILC program. The quality of the program usually depends on the expected, potential participant with the best CMC. In another case, it is a requirement arising from a legislative requirement, where the maximum permissible measurement uncertainty can be determined. The value determined in this way can also be called the characteristic level of the ILC

* Corresponding author: Miloš Ujlaky, E-mail address: ujlaky@slm.sk

program. The criteria for the selection of the ILC items, the reference laboratory, the technical and organizational conditions of the measurement are derived from the stated level. This study in the field of secondary metrology focuses on the field of instrument selection, and is related to the examination of its properties such as accuracy, repeatability, and in an ILC, the most important ones as reproducibility and stability. The study proposes an approach determining under which conditions and how to perform stability tests of the ILC item and how to evaluate its stability. [1], [2], [4], [15], [17]

If in the past some programs failed, the reason was insufficient preparation and determination of technical and organizational conditions for participation, unplanned damage to the measuring instrument, etc., but the choice of the measuring instrument certainly played a significant role. If the basic property of the measuring instrument during ILC, which is stability, fails, some additional procedures must be applied. In these cases, not only the planned course is disrupted, the costs of the organization increase, but it can also have an impact on the quality level of the ILC. If a participant is unable to validate his CMC, simply because due to a change in the metrological characteristics of the object the relative uncertainty had to be increased and as a result exceeds the set quality level, participation in the program loses its justification. [1], [2], [3], [5], [6]

For these reasons, it is necessary to prevent these situations well in advance. Already in the preparation of the ILC itself, a detailed analysis and a strategy of the expected or assumed development of the ILC should take place. When to apply stability tests? In all cases where a change in metrological characteristics with respect to time can be expected. [5], [6]

2. Statistical approach to the evaluation of stability tests

The aim of the experimental part is to solve the most standard cases where the measuring instruments are usually assumed to drift only in the long run. [16] However, given the specific instrument, the number of participants and the time intervals between measurements, it is necessary to confirm the stability check of the assumption that there will be no significant changes in the metrological characteristics that could affect the

participants' results. This requirement is also based on International Standard 17043: 2010, where clause 4.4.3.1 states the requirement to have specified criteria for satisfactory stability. [10] The design of proficiency tests is based on one of the basic properties of the measuring instruments, which is repeatability and reproducibility. The number of measurement series and the time interval depend on the number of participants and the expected duration of the program. For the purpose of performing stability tests, ILC participants will be replaced by measurements in the laboratory performed at a specified time interval.

In the tests of stability will be used the same and standardized method, and the same measuring instrument, measuring procedure, laboratory and laboratory equipment will be used. It is possible to change the metrologist to perform stability tests to assess reproducibility. Subsequently we can look at individual measurement results statistically and evaluate whether it is a statistically controlled process. [7], [8], [11], [12]

We will build on a set of n measurements and a series of N repeated measurements at specified time intervals. One series of measurements corresponds to the result of one participant and we can determine it as a sample mean value, which is expressed as:

$$\bar{X}_j = \frac{1}{n} \sum_{i=1}^n X_{ij} \quad (1)$$

where: X_{ij} is the i -th measured value in the j -th series, $i = 1, \dots, n$, $j = 1, \dots, N$, n is the number of measurements per series.

The sample mean value for N series of measurements, in the graphs we denote as CL, we express as:

$$\bar{\bar{x}} = \frac{1}{N} \sum_{j=1}^N \bar{X}_j \quad (2)$$

where: N is the number of measurement series, $j = 1, \dots, N$.

We assume that the distribution of the quantity corresponds to the normal distribution (in the field of calibration activities it has its significant representation), meaning that we assume $e \sim N(0, \sigma^2)$. The variance of a random error from one series of measurements, also called the within-laboratory variance of the j -th series, and we estimate it as:

$$s_j^2 = \frac{1}{n-1} \sum_{i=1}^n (X_{ij} - \bar{X}_j)^2 \quad (3)$$

for $j = 1, \dots, N$.

As we take into account several series of measurements, we will talk about the repeatability variance, which we estimate as

$$s_r^2 = \bar{s}^2 = \frac{1}{N} \sum_{j=1}^N s_j^2 \quad (4)$$

In addition, we can determine the interlaboratory variance, which we estimate as

$$s_L^2 = \frac{1}{N-1} \sum_{j=1}^N (\bar{X}_j - \bar{\bar{X}})^2 \quad (5)$$

We can also determine the so-called reproducibility variance. Its estimate s_R^2 is determined as

$$s_R^2 = s_L^2 + s_r^2 = \frac{1}{(N-1)} \sum_{j=1}^N \sum_{i=1}^n (X_{ij} - \bar{\bar{X}})^2 \quad (6)$$

More often we use the square roots of these variances, which we call the standard deviation of repeatability (its estimate is denoted by s_r) and the standard deviation of reproducibility (its estimate is denoted by s_R).

Based on the performed N series of measurements after n measurements, we can proceed to the construction of a criterion for future stability monitoring. The control diagram method is proving to be a suitable method.

After obtaining the results from relations (3) and (7), we can proceed to establish criteria for stability monitoring. For this purpose, it is possible to proceed from the theory of control of measuring processes and use the so-called Shewhart control diagrams. [7], [8], [9]

The arithmetic mean diagram and the range diagram appear to be suitable.

2.1. Diagram of arithmetic means

These diagrams consist of a plot of the horizontal central line CL, represented by the mean value of a series of measurements and control limits, otherwise the criteria for the acceptability of the stability results. The control limits are set to $3s_{\bar{X}}$ (99.7%), or $2s_{\bar{X}}$ (95%) on both sides. In common metrological practice, measurement results with assigned uncertainty are reported with a coverage interval of approximately 95%, so we set the acceptance criterion at $2s_{\bar{X}}$. [13]

The limits of the eligibility criteria can be calculated as:

$$UCL = \bar{\bar{X}} + 2 s_R / \sqrt{m} \quad \text{and} \quad LCL = \bar{\bar{X}} - 2 s_R / \sqrt{m} \quad (7)$$

where: m is the number of measurements during the control in the process of the stability monitoring, which we will subsequently do, UCL - upper control limit, LCL - lower control limit.

It is advisable to record the results graphically.

2.2. Range diagram

For a broader analysis, it is possible to analyse the results in terms of the range of measured values, but the measurements in the stability control do not differ so much that we would have to normally monitor also the range diagram. The calculation procedure is available in the relevant literature [7], [8], [9].

2.3. Application in practice

For the purpose of verifying the suitability, two areas of possible ILC programs were selected. As the items that could potentially be used for the purpose of ILC, resistance decade and electricity meter were selected. These are areas where significant drift of the instrument is not expected in the short term. [16] In order to confirm this assumption, the procedure described in point 5 was applied to selected ILC items. In this case, three repeated measurements were performed, with a certain time period, corresponding to the expected duration of the ILC program.

In standard metrological practice, for the purpose of stability, 2 to 3 repeated measurements are usually performed. It is based mainly on: - the quality of the selected ILC items, - the quality of the program, - the complexity of the metrological performance, - wider knowledge about the measuring instrument (knowledge from long-term monitoring of the characteristics of the measuring instruments), etc. The fact is that a higher number of repeated measurements will reassure us on the one hand about the suitability of the measuring instruments, but we must not ignore the fact that this will be reflected in the final fees for participation in the ILC program. Therefore, the approach to stability tests is based on efforts to optimize the scope of testing, which also applies to the selection of measuring points, e.g. depending on the expected characteristics of the measuring instrument. Figure 1 shows some selected stability results. The results were also supplemented by assessing the degree of agreement of the two outermost measurements using the standard deviation criterion (agreement is considered satisfactory if $En < 1$). [4], [9], [10], [11], [12] In the second case, stability tests were



Figure 1 Example of evaluation of stability of resistance decade at selected measuring points

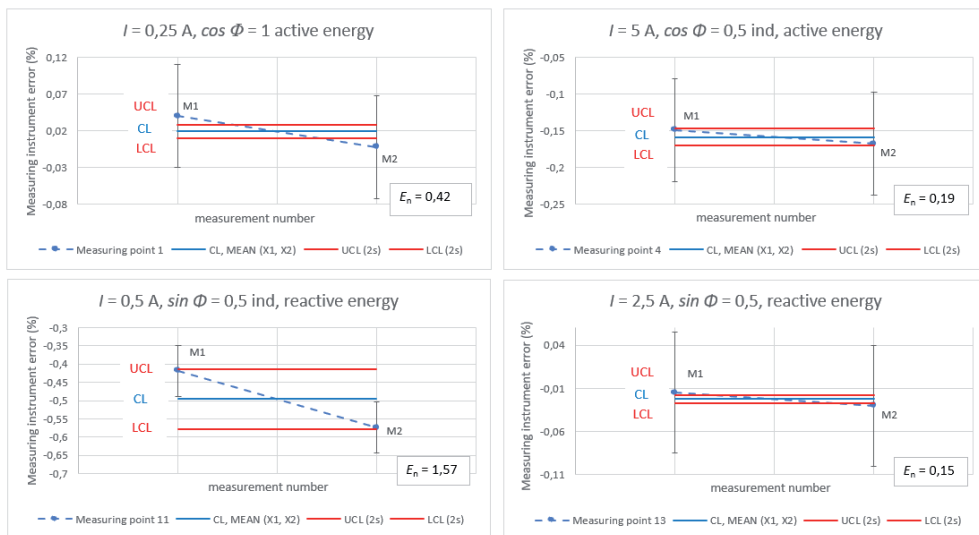


Figure 2 Examples of evaluation of electricity meter stability at selected measuring points

performed on the electricity meter, where 17 calibration points were measured. At least at one point after the evaluation, a problem was revealed, either by the measuring instrument itself or by the calibration laboratory. Selected results are shown in Figure 2.

3. Evaluation and Discussion

Individual experiments were analysed and subsequently evaluated.

– Figure 1. After evaluating the graphical representation of the

results, it can be stated that the resistance decade stability test showed a statistically controlled measurement process and a satisfactory agreement of repeated measurement results was demonstrated.

– Figure 2, graph - measuring point 4. The example, which represents another 14 calibration points, is a statistically controlled process, as well as the agreement of two repeated measurement results.

– Figure 2, graph for measuring point 1. Statistically uncontrolled measurement process. It is clear from the results that there was a drift of the ILC item. Nevertheless, comparison and evaluation of the standard deviation criterion showed a good agreement of repeated measurement results.

– Figure 2, graph for measuring point 11. Statistically controlled measurement process. Nevertheless, the two repeated measurements do not match. This measuring point is not suitable for ILC purposes. Such a case requires further detailed analysis of the causes also in connection with the decision of the suitability of the measuring instruments for the expected purpose.

– Figure 2, graph for measuring point 13. Statistically controlled measurement process, but at the same time good agreement of two measurements. It is worth noting that the variance of the two measurements is an order of magnitude smaller than the expanded measurement uncertainties themselves. [13], [14]

4. Discussion of the results

The application of the new approach in the process of stability tests showed cases where the set criteria were exceeded. In this context, we must realize that we cannot expect the measuring instruments to have ideal stability. Especially in the field of secondary metrology, where standard to routinely calibrated types of measuring instrument should be used for the purpose of ILC. The use of high-quality measuring instruments with top stability would result not only in increased demands on the security of the measuring instruments, in transport, but especially in the problems of the participants with the calibration of such instruments, as they are not part of normal metrological practice. Therefore, when evaluating stability, it is appropriate to focus on the contribution of the change to the measurement uncertainty that characterizes the quality level of the ILC program, to consider its significance. If the change is not significant, then in this case it is assumed that, as in the stability test, a successful agreement of the first and last measurement should be demonstrated in the ILC itself. If the change is significant, it is necessary to monitor this change during the program and include it competently in the evaluation of the ILC results. Alternatively, if the exceedance relates to a measuring point at the level of the measuring range minimum, it may be considered to exclude the measuring point from the ILC range, as there is an increased risk of extreme measuring instrument errors at the extreme points of the range.

5. Conclusions

This approach to evaluating the stability of the ILC items is only one of the possible approaches applicable in this area. It is especially suitable for areas where it is possible to obtain a relatively large

amount of data by automating the collection, as it is an application of statistical procedures. A wide range of testing will be difficult to apply in areas where measurements are time consuming. From another point of view, even in this area, there is a constant pressure on cost-effectiveness, with an impact on the final price for participation in the program. The situation therefore requires the search for optimal solutions in terms of the degree of risk and the cost of stability tests.

Acknowledgments

This work has been produced with support under the projects KEGA 1/0168/21, KEGA 040TUKE-4/2019 and VEGA 1/0168/21. APVV-18-0066, APVV-19-0032

References and Notes

1. VOICULESCU, R.M. – OLTEANU, M. C. – NISTOR, V. M.: Design and operation of an interlaboratory comparison scheme. In: Nuclear, (2013)
2. BRIGGS Philip: Proficiency testing for calibration laboratories, XX IMEKO World Congress, (2012)
3. Jeff C. Gust: A Discussion of Stability and Homogeneity Issues in Proficiency Testing for Calibration Laboratories, In: NCSLI Measure The Journal of Measurement Science, Volume 4, (2009)
4. WIMMER, Gejza – PALEŇČÁR, Rudolf – WITKOVSKÝ, Viktor – ĎURIŠ, Stanislav: Vyhodnotenie kalibrácie meradiel: Štatistické metódy pre analýzu neistôt v metrológii. Bratislava: STU, 2015, ISBN 978-80-227-4374
5. PALEŇČÁR, R. – RUIZ, J. M. – JANIGA, I. – HORNÍKOVÁ, A.: Štatistické metódy v metrologických skúšobných laboratóriách, 2001. ISBN 80-968449-3-8
6. UJLAKY, Miloš – TÓTH, Theodor – ŠMIGURA, Dušan: Kvalitatívna úroveň programov MLPM ako kritérium pre účasť v programe medzilaboratórnych porovnávacích meraní. In: Metrológia a skúšobníctvo. Roč. XXIV, č. 2, (2019), s. 18-21, ISSN 1335-2768
7. Ujlaky, Miloš – TÓTH, Theodor – ŽIVČÁK, Jozef: Medzilaboratórne porovnávacie merania ako súčasť procesu zabezpečenia kvality výsledkov meraní, Trendy v biomedicínskom inžinierstve 2019: 13. konferencia slovenských a českých pracovísk biomedicínskeho inžinierstva. – Žilinská univerzita v Žilina, (2019), ISBN 978-80-554-1587-113
8. Ujlaky, Miloš – TÓTH, Theodor – FRIČ, Anton – KYSLER, Daniel: Inovatívne prístupy v procese organizácie programov MLPM. In: Novus Scientia 2021: 18. Medzinárodná vedecká konferencia doktorandov strojných fakúlt technických univerzít a vysokých škôl; (2021), s. 278-282, ISBN 978-80-

553-3798-2

9. UJLAKY, Miloš: Analýza kľúčových porovnávacích meraní v kontexte optimalizácie skúšok spôsobilosti v oblasti sekundárnej metrológie. In: Zborník z 49. Fórum metrológov, (2019), s. 241-260, ISBN 978-80-553-3329-8
10. UJLAKY, Miloš – TÓTH, Theodor – ŠMIGURA, Dušan: Analýza ovplyvňujúcich faktorov v procese účasti v porovnávacích meraniach. In: Metrológia a skúšobníctvo. Roč. XXIV, č. 1, (2019), s. 17- 0, ISSN 1335-2768
11. Cox, M.G.: The evaluation of key comparison data. In: Metrologia, (2002), Vol. 39, 589-595
12. BUCHER Jay L.: The Metrology Handbook, Second Edition, (2012)
13. Standard EN ISO/IEC 17043: 2010 Conformity assessment - General requirements for proficiency testing, International Organization for Standardization, Geneva, Switzerland
14. Standard ISO 13528: 2016 Statistical methods for use in proficiency testing by interlaboratory comparison, International Organization for Standardization, Geneva, Switzerland
15. Standard ISO 5725: 2000 Accuracy (trueness and precision) of measurement methods and results. Part 2: Basic method for the determination of repeatability and reproducibility of a standard method, Slovak Office of Standards, Metrology and Testing, Bratislava, Slovakia
16. Guide EA-4/02 M:2013 Guide to the Expression of Uncertainty in Measurement
17. Guide EA-4/21 INF: 2018 Guidelines for the assessment of the appropriateness of small interlaboratory comparisons within the process of laboratory accreditation, European accreditation
18. MLPM: EURAMET Guide on Comparisons, Version 1.0 (05/2016)
19. JCGM 200:2012 International vocabulary of metrology – Basic and general concepts and associated terms, BIPM, 2012
20. PTB for PT: Basic Rules PTB Ex Proficiency Testing Scheme, Physikalisch-Technische Bundesanstalt, version 1, 19. 03. 2015