Metrology in chemistry: some questions and answers

Ivo Leito [ID] * and Irja Helm [ID]

University of Tartu, Institute of Chemistry, Ravila 14a, Tartu, 50411, Estonia

Abstract: This article presents some typical questions from practitioners, trying to implement metrological concepts in their everyday chemical analysis work, and answers to them by the authors, in the context of the authors' pragmatic view on applying the metrological principles to chemical analysis. Several of the presented questions are staples at training seminars and during on-line courses. The answers to the questions reflect the authors' opinions and are not always fully in line with the generally accepted positions.

Keywords: Metrology in chemistry; reference values; repeatability; intermediate precision; constant improvement.

1. Introduction

In a recent paper [1] an attempt was made to present metrology in chemistry as a discipline in as simple and pragmatic terms as possible. In a somewhat simplistic fashion, three sets of key activities were outlined that minimally have to be performed in a laboratory in order to claim that the work is done according to metrological principles:

a) *Compare your values with reference values:* This is possibly the most important of all activities. If there is agreement then this first of all demonstrates acceptable accuracy and adequate measurement uncertainty. In addition, agreement with reference value indirectly confirms adequate selectivity of your analytical method (procedure) and absence of serious technical issues.
b) *Collect data over long time periods:* Analysis data of identical or similar samples collected over long time (formalized e.g. as a control chart [2]) are valuable in ensuring that the method is under control and stable. In addition, there is no way to evaluate trueness without replicate measurements, preferably over a long time period.
c) *"Do not stop there":* The data collected in the context of the two preceding items tend to (a) be insufficiently abundant and (b) become obsolete as time goes by. Therefore, it is important to never stop collecting data, as long as the method under question is in use.

These principles are expected to be universal – applicable to any type of measurement, not limited to chemical measurements. In addition, leaving aside some special cases, they are as a rule not difficult to apply in routine laboratory. Nevertheless, years of conducting training courses and six years of experience with the online measurement uncertainty course [3] have demonstrated that laboratories still need help with applying metrological principles to their everyday work.

The aim of this paper is to present some typical questions that practitioners have in this regard and answers to them. Several of the presented questions are staples at training seminars and during on-line courses and the answers are based on the experience and opinion of the authors, not necessarily always agreeing with generally accepted conventions.

---

* Corresponding author:E-Mail: ivo.leito@ut.ee

2. Questions and Answers

Question 1: *In determination of analyte A in matrix B there are no certified reference materials (CRMs) available. Therefore, there is no way to compare my results to any reference value, is it correct?*

Answer: CRMs are perhaps the best means of obtaining reference values,[4] but they are not the only means. There are other possibilities: spiking, preparing reference samples by formulation, analyzing your native samples with a different method (possibly in a different laboratory) and participating in an interlaboratory comparison, e.g. a proficiency test (PT). Situations when none of these is possible are rare. One often overlooked possibility is organizing bilateral interlaboratory comparisons. For this you only need to find somebody who does the same type of work. Then both you and the other laboratory can split a sample, send to each other and then compare your results.

It is true of course that when the aim of using a reference value for establishing traceability, rather than confirming adequate performance, then CRMs are more difficult to replace.

Question 2: *Comparing my results to such "reference values" as consensus value of a proficiency test or result of another lab of similar quality to mine will not yield any useful outcome, because the uncertainties of those values are of the same order as mine or even higher.*

Answer: Indeed, if a reference value that is used for comparison is not more accurate than the one determined with your own method then a single comparison, although with positive result, is insufficient to give much support for the adequacy of our method. However, if a number of such comparisons are available and every time your value agrees with the reference value then that can be considered quite serious evidence that your measurements are reliable. Moreover, although evidence that your method performs well is difficult to obtain, obtaining evidence that your method performs poorly is really easy: Just one comparison with strong disagreement is sufficient to reveal a problem.

Question 3: *I am determining analyte A in chicken, but the only available CRM is with beef matrix. Analyzing that one is in my case useless, is it correct?*

Answer: In our opinion it is not useless. True, matrix match is very important and different "meat matrices" can lead to different recoveries or different interferences when the same method is used. Nevertheless, the same applies as with the previous question: if you have several comparisons with related but different matrices then the probability is quite high that you are not going completely wrong. And clearly, if with some related matrix – e.g. beef vs chicken – your result is in stark disagreement with the reference value then the chances are very high that there is something seriously wrong with your method. As a generalization – in the case of analyzing CRM with similar but not matching matrix, you cannot get from CRM analysis definitive confirmation that your method performs well. But if your method has a serious problem then that will most probably be revealed. Thus, analyzing a CRM with matrix that is related but different from the matrix of your sample is certainly better than not analyzing any CRM at all.

Question 4: *In my measurement field it is impossible to keep samples intact for a longer time than couple of days, thus there is no way to carry out determination of intermediate precision or run X control charts, is it correct?*

Answer: There are indeed cases when keeping samples for longer than a couple of days is a challenge. However, in the majority of cases the following approach works: (i) a large batch of sample is prepared and carefully homogenized; it is divided into small batches, each of the approximate size required for one determination; (ii) each batch is put in a different vessel/bag and carefully closed and (iii) all such samples are put in a freezer. This approach enables establishing control charts of many months' length with e.g. raw meat.

One of the measurement fields where keeping samples is next to impossible is determination of dissolved oxygen in water [5]. But with dissolved oxygen it is possible to generate in a water bath water

saturated with air [6] for which accurate dissolved oxygen concentration is known (as a function of temperature and air pressure). And, in this case control chart can be compiled using the difference between your measured value and calculated concentration [7].

Question 5: Reliably estimating measurement uncertainty of chemical analysis results continues to be a challenge. What possibilities are there for checking the validity of the obtained uncertainty estimates?

Answer: Probably the best check for the validity of your uncertainty estimate is to compare with an independent result obtained for the same sample. Very common is, e.g. analyzing a CRM and then comparing your result with the reference value of the CRM, using e.g. the zeta score approach [8,9]. If the absolute value of the zeta score is below 2 then the uncertainty estimate can be considered realistic. It is important to note that several such comparisons should be carried out, in order to reliably claim that the uncertainty estimate is adequate. If just one comparison is done then the agreement can be due to a chance. If the zeta score is above 3 then the uncertainty is underestimated or something has happened with the method that is not foreseen by the uncertainty estimate (which, in fact, also means that uncertainty is underestimated).

Also, if you participate in a PT then comparing your result with the PT consensus value is useful. In the case of a PT the consensus values usually do not have uncertainty estimates. Then a simple, although not fully rigorous, approach is to see, whether the PT consensus value is within the $k = 2$ uncertainty range of your result.

Question 6: Is "incomplete selectivity" really an uncertainty source? I have always known that we take care of demonstrating selectivity during validation and then when we apply the method within its scope we do not need to worry about selectivity any more…

Answer: In an ideal world, yes, this is correct. However, we live in a less than ideal world. In the context of this question non-ideality means that it is usually almost impossible to foresee and test all possible interferents that can occur in samples. An exception perhaps in production environment, where products of close to identical composition are produced day after day. Thus, in the majority of analyses it would be appropriate to have some estimate, how much the result could possibly change, should there be an interferent in the sample, which was not considered during validation. This kind of uncertainty contribution is nearly impossible to estimate rigorously and experience from similar situations has to be used. The good news is that the single-lab validation approach to uncertainty estimation (e.g. the formalization published by Nordtest) [10] enables addressing this uncertainty via the bias component, but only if the bias component is determined against different reference values (i.e. determined against different certified reference materials, different intercomparison reference values, etc).

Question 7: When I want to increase the reliability of my results, would it be a good approach to make as many replicate measurements as possible?

Answer: Making replicate measurements of our samples is almost always a good idea, if the extra cost can be tolerated. Two things should be considered, however. First of all, analysis results are influenced by random and by systematic effects. Making replicate measurements enables reducing the influence of the former, but not the latter. Since systematic effects – such as analyte losses during sample preparation or interferences from compounds similar to the analyte – often contribute significantly to the uncertainty of the result, making a large number of replicates is usually not very productive in reducing uncertainty. Secondly, it is very different situation if every replicate measurement is done on a separate subsample and involves sample preparation or if sample preparation is done only once and the obtained solution is analyzed in replicates. In the first case the spread of the results includes also random effects from sample preparation, while in the second case it does not. Thus, in the second case, essentially all uncertainty caused by sample preparation – usually the dominant uncertainty contribution in chemical analysis – comes in the form of a systematic effect and replicates do not help reducing it at all.

Question 8*: Is the concept of true value a valid concept or not?*

Answer: In the opinion of the authors, the concept of true value has its merits [11]. Yes, the true value will always remain unknown to us. But at the same time, true value is useful as "the ultimate aim" of a measurement. It can even be argued that without true value, it is not clear what a measurement performer should be aiming at.

Question 9*: For the calculation of repeatability (s_r) and intermediate precision (s_{Rw}), can I use the routine samples (customer samples) that have been received in the lab or do I have to use a control sample prepared for that purpose and analyze it every time when I analyze the customer samples?*

Answer: You can use the real samples received in the laboratory. For s_r several measurements with one sample have to be done and the repeatability data of different samples can be combined using pooled standard deviation. When determining the s_{Rw} you have to be certain, that the sample is stable in the timeframe (typically at least couple months, preferably one year) that you use this as the one control sample.

Question 10*: Which is more important to determine, repeatability or interim precision?*

Answer: In general, it is a good idea to determine both, because they show different things. However, if the question is "which of these two is more useful in evaluation the reliability of my method?" then the answer is: intermediate precision is more useful. The reason is that the intermediate precision standard deviation s_{Rw} includes more effects that influence the measurement result. All effects that are systematic within day but become random in the long term are accounted for by s_{Rw} but not by repeatability standard deviation s_r. A good example is calibration graph: if the results of all measurements on the same day are calculated using the same calibration graph, then any deviation of the graph from an "ideal" one will be a systematic effect within day and will not be accounted for by s_r but will be accounted for by s_{Rw}.

3. Conclusions

The questions and answers presented here illustrate that it is nearly always possible to apply metrological approach to chemical measurements. Situations, such as "there is no CRM available with the exact analyte-matrix combination and concentration range", are of course disturbing but do not destroy the possibility of confirming the adequacy of results: useful confirmatory evidence can be obtained in different ways.

ORCID

Ivo leito: 0000-0002-3000-4964
Irja Helm: 0000-0002-5964-2361

References

[1]   I. Leito (2015). A personal view on metrology in chemical analysis, *Accred. Qual. Assur.* **20**, 229–231.
[2]   H. Hovind, B. Magnusson, M. Krysell, U. Lund and I. Mäkinen (2012). Internal quality control–Handbook for chemical laboratories, Nordtest TR 569, 4th Ed., Nordtest, Oslo.
[3]   I. Leito, I. Helm and L. Jalukse (2015). Using MOOCs for teaching analytical chemistry: experience at University of Tartu, *Anal. Bioanal. Chem.* **407**, 1277–1281.
[4]   F. Ulberth (2006). Certified reference materials for inorganic and organic contaminants in environmental matrices, *Anal. Bioanal. Chem.* **386**, 1121–1136.
[5]   T. Näykki, L. Jalukse, I. Helm and I. Leito (2013). Dissolved oxygen concentration interlaboratory comparison: what can we learn?, *Water* **5**, 420–442.
[6]   I. Helm, L. Jalukse and I. Leito (2012). A highly accurate method for determination of dissolved oxygen: gravimetric Winkler method, *Anal. Chim. Acta* **741**, 21–31.
[7]   I. Helm, G. Karina, L. Jalukse, T. Pagano and I. Leito (2018). Comparative validation of amperometric and optical analyzers of dissolved oxygen: a case study, *Environ. Monit. Assess.* **190**, 313 (18 pages).

[8]   ISO 13528:2015, Statistical Methods for Use in Proficiency Testing by Interlaboratory Comparison, ISO, Geneva, **2015**.

[9]   T. Linsinger. Application Note 1. Comparison of a measurement result with the certified value. European Communities, **2005**.

[10]   B. Magnusson, T. Näykki, H. Hovind and M. Krysell (2012). Handbook for calculation of measurement uncertainty in environmental laboratories (NT TR 537 - Edition 3.1) - Nordtest.Info, Nordtest, Oslo.

[11]   G. Meinrath (2002). The merits of 'true values', *Accred. Qual. Assur.* **7**, 169–170.

A C G
publications