

Scientific methodology in Analytical Chemistry

Jens E.T. Andersen, Botswana International University of Science and Technology
College of Science
Department of Chemical and Forensic Sciences
Palapye, Botswana
e-mail: andersenj@biust.ac.bw

Abstract— Ever since the work of Robert Boyle, scientists have used the same methodology to obtain new information about mechanisms of nature. Together with other scientists at the time, Robert Boyle suggested that all scientific discussions could be settled by means of experiments. Without proper tools of statistics, it was difficult to arrive at any conclusive decisions in many cases. However, later discoveries of statistical tools have limited dissemination of flawed discoveries of nature but to a great number of scientists, these results have not been understood or recognized. This poses a major issue in science and it can be shown that some publications do not meet requirements of credibility and reliability, owing to scientific misconduct. Journals thrive on scientists who wish to publish results where data have been seriously manipulated. At present time, it seems like scientific methodology has been forgotten, which may explain the lack of ground breaking development in science? The quest for quality assurance in science is greater than ever, and initiatives of the European Commission constitute the first steps towards an environment that might bring back scientific methodology to science. Unfortunately, guidelines published by the European Commission do not underpin scientific methodology, and it has become evident that science and industry use two different concepts of scientific methodology that cannot both be true. Therefore, it is suggested to improve the guidelines of the European Commission to a level of simplicity and transparency that allows scientists to perform reliable method validations and obtain correspondence between experimental uncertainty and predicted uncertainty, which is so fundamental to scientific methodology. It is proposed that new methods of quality assurance should be embraced by not only analytical chemistry but also by other fields of science.

Keywords— *Quality assurance, method validation, scientific misconduct*

Introduction

Imagine that scientist present only the best results they obtained during series of measurements aimed at solving a scientific issue. Clearly, all results obtained under such conditions relate to the statistical concept of precision whereas the real aim of science is to obtain information about trueness and accuracy [1,2]. If only the best results were shown in scientific publications, it would be a legitimate question to ask; what happened to all the other data? Such data may be denoted as dark data, and it would be interesting to know the magnitude of dark data that exist in the scientific production.

RESULTS AND DISCUSSION

Dark data

It is well known and almost an accepted fact that scientists are not able to reproduce results that are presented in coloured flyers with description of apparatuses, specifications and supreme performance. Under most circumstances, it is generally accepted that precision is slightly lower in the analytical laboratory as opposed to more controlled conditions that may be encountered only in manufacturer's laboratories. A slight decrease of precision is not really an issue and it rarely affects determination of figures of merits [3] that are obtained during the course of method validations [4]. The overall performance of most methods is acceptable, particularly with respect to qualitative analysis. Quantitative analytical chemistry is a completely different area of research where precision may reach unacceptably low levels [5]. However, there are instances where it becomes necessary to start thinking about how the technology and software are actually working together, and how it affects precision. Manufacturers of graphite-furnace atomic absorption spectrometry (GF-AAS) and electrothermal atomic absorption spectrometry (ET-AAS), prefer to allow for only certain data to be part of measurements. Data that deviate by more than 5 % (default) must be excluded from the data set. Suppose that the inherent precision of that type of apparatus result in deviations of e.g. 60 % [6]; this would make the likelihood of re-calibration very high. Further, it would make it almost impossible to construct a calibration curve with many data points because the risk would be high of an outlier lowering the precision to below 95 % that would trigger the re-calibration. Accordingly, scientists use only a few data to construct calibration curves. Therefore are expected excessive amounts of dark data with such technologies.

Now, in the description of these technologies was not mentioned the sample but it is evident that the risk is high of large deviations from expected values. If a calibration curve were obtained by repeated re-calibrations, the correspondence between expected value and the value obtained by the calibration curve, would be entirely random. However, when such repetitions are performed in great numbers, a value obtained by one of the calibration curves would fit the expected value quite neatly, and it might be used for publication or decision making. All the data not shown may then be relegated to the pool of data generated on days when the apparatus was not working well or the operator was not concentrating enough on the work. Thus, either the apparatus or the operator may be blamed for generation of poor data. Unfortunately, it is not an

easy task to decide if the apparatus or the operator was the cause of the problem and methods to distinguish between the two origins of bad data is currently investigated by science.

Nobody knows the genuine content of a sample with unknown content, and apparatuses of analytical chemistry can only be used to give a qualified estimate of the content that is based on calibrations. To the people not familiar with analytical chemistry, it is a complete mystery and impossible to understand how two accredited laboratories are able to obtain two completely different values, when they used for their measurements the same type of apparatus and the same sample. To the scientists, however, it is relatively easy to understand this deviation because they are familiar with the concept of uncertainty of measurement. Uncertainty of measurement is the key to understanding results of any type. It is also easy to explain to laymen the principle of uncertainty of measurement, except when it is large. To laymen, it may seem incredible that a very expensive apparatus does not always provide data of high precision. Nuclear magnetic resonance spectrometry and mass spectrometry are notable examples of very expensive types of apparatuses that are not particularly well suited for quantitation. More research is needed, in order to elucidate the mechanisms associated with stability of detector responses, which constitute the sensing part of the apparatus. Some reports indicate that detectors, and not operators, are the main culprits with respect to generation of uncertainty.

Consensus value

The operational calibration is supposed to eliminate day-to-day variations of the apparatuses' parameters and the influence on results. However, in reality something else happens; the calibration may present excellent results, as can be recognized by a very straight line and a low standard deviation of the sample. Many calibration data would result in narrow confidence lines and narrow confidence limits of the sample [7]. Frequently, the value obtained for the sample differs significantly from known value, which should not occur, owing to the operational calibration. The known value may be obtained from measurements of a certified reference material, so in the case of any deviation, there is reason to expect that something was wrong with the measurements. Interestingly, in the majority of scientific publications, authors find complete correspondence between measured and certified values where both were associated with very low standard deviations. This is quite disturbing because many authors failed to notice that the certified value was obtained after elimination of numerous outliers. Therefore, it should not be possible to reproduce the certified value, unless the same number and magnitude of outliers were deleted from the laboratory trials, which obviously is very unlikely or impossible.

Unfortunately, the correspondence is less pronounced between laboratories analyzing samples of unknown content. However, since nobody knows the exact content of a sample of truly unknown content, everybody is right. Therefore, the correct way forward is to average values combined of all laboratories, and resultant average value is denoted as the consensus value [7]. Most frequently, the uncertainty of the consensus value exceeds by far the uncertainty for an average

value of a single laboratory. Contents of specific molecules of blood samples are truly unknown and the values of fundamental constants of nature are also not known but they may be estimated by various means of measurements.

International organisations

Issues of lack of correspondence between average values and uncertainties have been addressed by several international organizations, such as Bureau des Poids et Mesures (BIPM) [8], European Analytical Chemistry (Eurachem), Cooperation of International Traceability in Analytical Chemistry (CITAC), International Standardization Organization (ISO) and International Laboratory Accreditation Cooperation (ILAC). As stated on ISO website "Great things happen when the world agrees" but, paradoxically, ISO recommends rejection of outliers in the guide ISO 5725 [9], which is a route to disagreement (see above). During the past decades, the European Commission maintained a programme to deal with issues of analytical chemistry, and it delivered the organizations Eurachem and CITAC. The now 3rd edition of the Eurachem/CITAC guide [10] has obtained widespread recognition in the industry because it underpins traceability and it provides guidelines to calculation of uncertainty of measurement and uncertainty budgets. Although the Eurachem/CITAC guide is a step forward towards introducing to analytical chemistry the concept of uncertainty, it contains some serious flaws that, just like ISO 5725 [9], promote disagreement between results of different laboratories that analyze same samples. The main issue is concerned with the uncertainty budget where standard deviation of repetitions is included in the uncertainty budget. The uncertainty budget is designed to predict the uncertainty of the content of a sample. The sample is determined as an average value and a corresponding uncertainty that is supposed to match that of the uncertainty budget. Since the two are not separated in the Eurachem/CITAC guide, disagreement prevails again. Despite all the efforts and investments international guidelines do not promote scientific methodology. Scientific methodology is straightforward in the sense that it only requires that average values and standard deviations of measurements correspond to those of predictions and theory. In order to meet the requirements of scientific methodology, it is recommended that scientist use only certain sections of the guidelines and never reject any outliers whatsoever. It is always a good idea to produce an uncertainty budget for the method validation but the uncertainty of repeatability should not be included in the uncertainty budget because that is the parameter that is supposed to be predicted. However, with reference to guidelines of ISO, GUM and Eurachem/CITAC, scientists have the tools to treat data almost as they please, which continues to create confusion in analytical chemistry. Scientists of analytical chemistry may feel confused with statistical methods that are too complicated for everyday use, which excuses them from applying methods of QA to their investigations. Accordingly, further simplification is called for.

Pooled calibrations and scientific methodology

The notion of consensus value is somewhat disturbing because it is never obtained during e.g. blood testing of patients and optimization of chemical syntheses. It is imperative to know the true level of uncertainty for optimization purposes. Blood testing is an important part of the general practitioners tools of diagnosis [11] and too much trust given to results of clinical analysis may jeopardize the validity of diagnosis. This entails risk of dosing medicine to somebody who is not sick or, even worse, not dosing medicine to somebody who is actually sick. Similarly, optimization cannot be performed with any degree of reliability if it were based upon underestimated standard deviations (short-term precision).

Usually, the uncertainty of repeated measurements of samples is larger than the uncertainty that was predicted upon the basis of calibration lines. This disagreement has been puzzling to scientist for a long time, and the guidelines of ISO and Eurachem/CITAC are published with the intent to better understand and quantify unexpected deviations. Recently, it was suggested to consider the worst-possible performance of a given apparatus, in order to model the lowest uncertainty of repeated measurements of the sample. The idea is to perform a series of completely independent measurements where the apparatus is shut down, turned on and calibrated between each series of measurements, and then calibration lines are pooled with the purpose of modelling the uncertainty [12]. Pooled calibrations are thus reserved for calculation of uncertainty of calibrations, $u(\text{cal})$. Results of individual determinations of the contents of samples are also pooled and the corresponding standard deviation of repetitions, $u(\text{rep})$ is calculated [13–15]. When $u(\text{cal})$ corresponds to $u(\text{rep})$, it is said that the system is in statistical control. For a number of different technologies, it seems possible to obtain this correspondence, at least within an order of magnitude. Although this procedure yields elevated uncertainties, it may be the correct uncertainty overall that also corresponds to the uncertainty of inter-laboratory testing and the uncertainty predicted by the Horwitz equation [16].

Certified reference materials

As a substitute for inter-laboratory testing, it has been generally accepted to validate analytical methods by using samples of certified reference materials, where the content of analytes was determined by many independent laboratories using different technologies. The average value of all measurements minus outliers is denoted as the certified value that can be used for comparison with results of other laboratories [12]. It seems that it becomes difficult to compare measured values to certified references, when the principle of pooled calibrations was applied to the measurements. This indicates that the issue of outliers needs to be considered in more detail. Occasionally, the uncertainty of determination of samples of unknown content is so large that it becomes easily comparable to certified references but this may be regarded as a token of poor method performance or relegated to influence of interferences.

CONCLUSION

In the research area of analytical chemistry, it has been recognized earlier that it is difficult to obtain correspondence between uncertainty of calibration and uncertainty of measured values. However, recent findings indicate that full correspondence between predicted and observed uncertainty is the key to understanding research results of analytical chemistry. By using the principle of pooled calibrations, it becomes possible to obtain statistical control in analytical chemistry but at the expense of relatively poor precision. In many instances, the uncertainty of measurement is underestimated, owing to rejection of outliers, which hampers comparison with contents of certified reference materials. Certified reference has become one of the prime requirements for method validations and for research to be published in scientific literature. Since comparison is hampered, it becomes difficult to maintain the good reputation of analytical chemistry. Issues with ISO, GUM and Eurachem/CITAC need to be resolved together with the practice of eliminating outliers related to CRMs before analytical chemistry can be gaining more in credibility.

REFERENCES

- [1] J.C.F.G.I.M. JCGM, JCGM 200 : 2008 International vocabulary of metrology — Basic and general concepts and associated terms (VIM) Vocabulaire international de métrologie — Concepts fondamentaux et généraux et termes associés (VIM), Int. Organ. Stand. Geneva ISBN. 3 (2008) 104. doi:10.1016/0263-2241(85)90006-5.
- [2] J.E.T. Andersen, S.-M. Glasdam, D.B. Larsen, N. Molenaar, New Concepts of Quality Assurance in Analytical Chemistry: Will They Influence the Way We Conduct Science in General?, Chem. Eng. Commun. 203 (2016). doi:10.1080/00986445.2016.1180290.
- [3] J. Uhrovčík, Strategy for determination of LOD and LOQ values - Some basic aspects, Talanta. 119 (2014) 178–180. doi:10.1016/j.talanta.2013.10.061.
- [4] E. Working Group, The Fitness for Purpose of Analytical Methods A Laboratory Guide to Method Validation and Related Topics, 1998.
- [5] R. Chomchoei, E.H. Hansen, J. Shiowatana, Utilizing a sequential injection system furnished with an extraction microcolumn as a novel approach for executing sequential extractions of metal species in solid samples, Anal. Chim. Acta. 526 (2004) 177–184. doi:10.1016/j.aca.2004.09.043.
- [6] J.E.T. Andersen, Determination of iron by Z-GFAAS and the influence of short-term precision and long-term precision, Cent. Eur. J. Chem. 12 (2014). doi:10.2478/s11532-013-0362-9.
- [7] J.E.T. Andersen, Consensus science and the impact on analytical chemistry, Rev. Roum. Chim. 60 (2015) 397–402.
- [8] JCGM, BIPM, IEC, IFCC, ILAC, ISO, et al., Evaluation of measurement data — Guide to the expression of

uncertainty in measurement, Int. Organ. Stand. Geneva ISO. 1 (2008) 1–120. doi:10.1373/clinchem.2003.030528.

[9] ISO 5725, Accuracy (trueness and precision) of measurement methods and results, 1994.

[10] CITAC, Eurachem, Quantifying Uncertainty in Analytical Measurement, BIPM. 3rd ed. (2012) 1–133. doi:0948926 15 5.

[11] C. Lam, C. Law, K. Sze, K.K. To, Quantitative metabolomics of urine for rapid etiological diagnosis of urinary tract infection: Evaluation of a microbial–mammalian co-metabolite as a diagnostic biomarker, Clin. Chim. Acta. 438 (2014) 24–28. doi:10.1016/j.cca.2014.07.038.

[12] J.E.T. Andersen, On the development of quality assurance, TrAC - Trends Anal. Chem. 60 (2014). doi:10.1016/j.trac.2014.04.016.

[13] J. Cohen, Statistical Power Analysis for the Behavioral Sciences, 2nd ed., Hillsdale: Lawrence Erlbaum, 1988.

[14] O.I. Hedges L. V., Statistical methods for meta-analysis, Academic Press, San Diego, 1985.

[15] J. Cohen, How to Calculate the Pooled Standard Deviation, (n.d.). <http://www.statisticshowto.com/pooled-standard-deviation/>.

[16] W. Horwitz, Evaluation of analytical methods used for regulation of foods and drugs, Anal. Chem. 54 (1982) 67–76. doi:10.1021/ac00238a002.