# CONSENSUS SCIENCE AND THE IMPACT ON ANALYTICAL CHEMISTRY

Jens E.T. ANDERSEN

Botswana International University of Science and Technology, Department of Chemistry and Forensic Sciences,
Private Bag 16, Palapye, Botswana
E-mail: andersenj@biust.ac.bw

The origin of uncertainty is discussed vigorously in analytical science and, at present stage, it is unclear as to what extent scientists themselves impose uncertainty on the measurements or the apparatuses possess large levels of uncertainty of measurement.[1–3] Introduction of certified-reference materials (CRM's) and standard-reference materials (SRM's) in analytical chemistry has revealed that scientists and professional laboratories disagree about value of results and quality of measurement. Reproducibility is poor in many cases and it is impossible to understand why some laboratories are able to produce excellent results, whereas other laboratories produce poor results despite the fact that both analyzed the same compound with the same type of apparatus. Science is divided about this issue where one party believes that uncertainty originates from incompetence of laboratory staff and the other party believes that uncertainty relates entirely to the quality of the apparatus.

The alleged lack of compliance[4] will be demonstrated by examples and case studies with determination of calcium by flame-atomic absorption spectrometry (FAAS). It is suggested that discrepancies can be explained entirely with inherent properties of the system, and human error is negligible. It is shown that universal compliance can be obtained by accepting the real uncertainty of measurement without rejection of outliers. These results strongly indicate that scientific methodology should be altered by introduction of consensus science. The consensus value is the most likely result that can be obtained under a given set of circumstances but the real value is genuinely unknown. Consensus science has some immediate and fundamental consequences for analytical chemistry, which are discussed in detail.

## INTRODUCTION

It is the aim of the present investigation to demonstrate that there is a demand for altering scientific methodology in analytical chemistry. Further is it shown, that a novel view on scientific methodology has an impact on other fields of science as well. Recent results of metrology and quality assurance suggest that research must be associated with uncertainty estimates, in order to secure correct decision making. Mistaking short-term precision for long-term precision constitutes a major pitfall of decision making[5] because method validation and statistical testing based on short-term precision promotes decisions that are different from those of long-term decisions. In brief, decision making based on short-term precision is not valid; decisions in science must be based on long-term precision.[6] The reason is that precision of an apparatus does not guarantee a high level of accuracy. The real goal of science is, however, true and accurate values that are important for the understanding of mechanisms of nature; the goal is not the precision of an apparatus. Uncertainty of measurement is the key parameter of scientific methodology, and much effort has recently been devoted to establish practices and procedures that are suitable for evaluation of results in science and industry. The uncertainty budget gives an overview of all contributions to the overall uncertainty that is supposed to explain observed uncertainties. Large deviations between results obtained from analysis of certified reference materials (CRM's) at professional laboratories may be explained by incompetence[7] but there is also another possible explanation that absolves staff at laboratories: Uncertainty may originate from apparatuses. Trueness is assessed by analysis of CRM's that are available with a plethora of chemical species in various matrices and they serve as 'fundamental constants' for analytical chemistry.[8] Metrology of

analytical chemistry has developed towards investigations of CRM's. Methods for analysis of these chemical species are validated on CRM's and the average result of many independent laboratories constitutes the consensus value that may be different from certified value. The idea of a genuinely unknown concentration of a chemical species in a given sample has led to re-definitions of fundamental concepts of analytical chemistry.[8] Accuracy was substituted with trueness in cases of reproducing certified values, and trueness is evaluated by the proximity to the certified value. Accuracy is evaluated in terms of proximity to a consensus value that may or may not be true. Thus cannot accuracy be assigned a number, as opposed to trueness that in many respects resembles the old definition of accuracy? This set of new definitions may seem confusing or obsolete but they make good sense; it becomes possible to investigate in more detail the degree of uncertainty and compare it to a tentative bias. Manufacturers rightfully promise very high precision but it is the obligation of scientists to investigate accuracy. Analysis of chemical species in e.g. samples of blood[9] is often associated with a high degree of uncertainty that jeopardizes accuracy and decision making with respect to medical treatment. In a data set of a limited number of experiments is it of no value to identify outliers and subsequently reject outliers.[10,11] Even when the certified value is known with high accuracy is it unacceptable to reject outliers because uncertainty of measurement thus decreases to a level that prevents data to be reproduced.[12] A high degree of accuracy may be endorsed by successful treatment based on analytical results, since nobody knows the genuinely true number of measurand species in the sample. The problem is discussed of not knowing the genuinely true value thus has significance to science of analytical chemistry but it also has an impact to other fields of science.

## EXPERIMENTAL

Experiments with determination of calcium with flame-atomic absorption spectrometry (FAAS) were conducted by using manufactures procedures of two different types of apparatuses (Perkin-Elmer Analyst 100 and Perkin-Elmer Analyst 4100).

## RESULTS AND DISCUSSION

A certified value for speed of light in vacuum[13] was obtained as an average of two laboratories that provided values that were equal within statistical uncertainty (Table 1). Speed of light in vacuum is a fundamental constant of nature, according to theory of relativity, and a quantity genuinely unknown to mankind; it can only be estimated by measurement and a universally correct value is not known. However, it has been argued that speed of light may have changed, in fact decreased over the past one and a half century, as indicated by the depiction of measurements since 1862 in Fig. 1a.

Without question, relative uncertainty of measurement did decrease from 0.17% in 1862 to $2.4 \cdot 10^{-7}$% in 1977 and some early average values were determined at levels apparently above today's certified value. Most certainly, it cannot be decided whether or not speed of light decreased or remained constant unless a detailed knowledge of number of repetitions, handling of outliers and statistical testings' were included in the evaluation of data. At the $17^{th}$ CGPM conference in 1983, was it decided that the average value of Table 1 was adopted as the exact value of zero uncertainty ignoring all previous values.[14]

However, if no errors were committed in earlier estimates would it be natural, according to principles of consensus science, to adopt an average of all values since 1862? This would give us the alternative value of 299807 ± 54 km/s with an RSD(rep) of merely 0.012%. Such a low RSD is by far below the reach of analytical methods but much greater than contemporary value that is practically close to naught. The difference between the exact value and the consensus value is that the consensus value can be reproduced whereas the exact value is irreproducible in dissension with scientific methodology. The corresponding RSD, that was calculated by combined uncertainties (Fig. 1b) of all reported uncertainties, was determined as an order of magnitude higher than RSD(rep) thus indicating lack of statistical control. The issue with discrepancy between RSD(cal) and RSD(rep) is of fundamental significance in science. Presumably was some of the old methods imprecise but if the scientific community were to agree about a consensus value would it be reasonable to suggest that more than two laboratories deliver a result; more than one hundred independent measurements would be appropriate in order to meet requirements of statistics and consensus science.
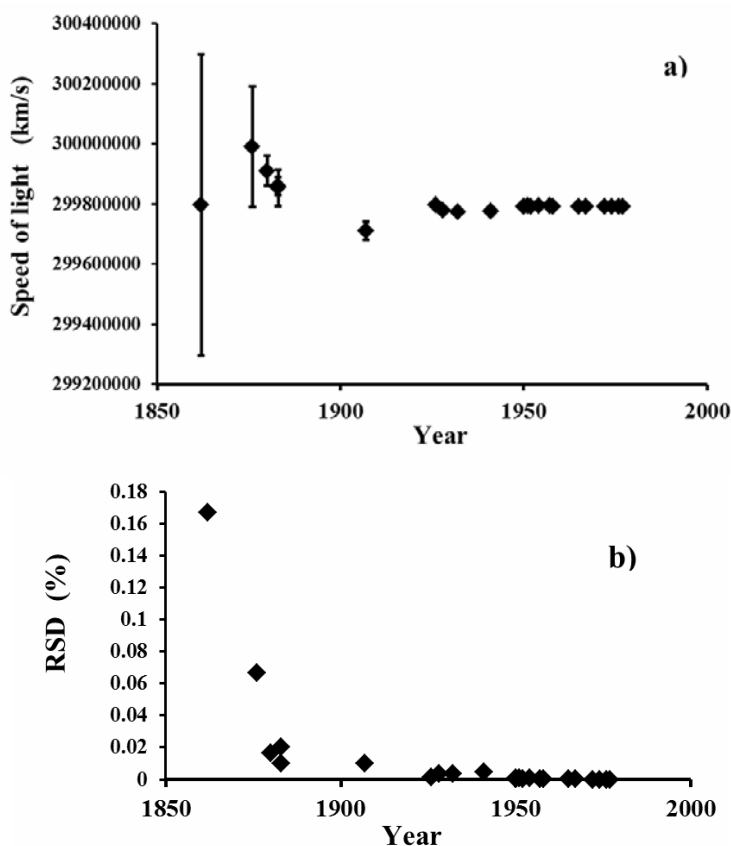
Fig. 1 – Average values and associated uncertainties of determinations of speed of light (a) within the time period 1862 – 1977 (b) Relative-standard deviations (RSD's) of results (1862 – 1977).

*Table 1*

Two values of speed of light, as determined by two independent laboratories. The two values were equal within limits of statistical uncertainty, and the average value has been promoted as certified value of speed of light in vacuum

| Organisation and average | Speed of light in vacuum   (km/sec) |
|---|---|
| **US National Bureau of Standards** | 299792.4574 ± 0.0011 |
| **The British National Physical Laboratory** | 299792.4590 ± 0.0008 |
| **Average** | 299792.4582 ± 0.0014 |

Determinations of speed of light in vacuum transfers directly to analytical chemistry and the example illustrates the idea of a value that is known to nobody; the value of a given parameter can only be determined as consensus value from numerous contributions comprising of measurements that were performed by different laboratories using different technologies.[15] Similar results has been published earlier with the celebrated work of Taylor and de Bièvre[15] in focus. Although less apparatuses and fewer analyses included in the production of results with in-house experiments was found a clear resemblance with the spread of results observed in ILC's of Taylor and de Bièvre where 181 results of different laboratories as included in the evaluation. Taylor and de Bièvre[15] indicated that deviations from certified value was a results of staff's incompetence; an idea that later was supported by other scientists[7,16] as well. However, if uncertainty of measurement were underestimated as a result of elimination of outliers would it be impossible to test for statistical significance?

These comments also apply to the results of Taylor and de Bièvre[15] but results were evaluated in terms of z-scores. However, application of z-scores for decision making cannot be trusted when the number of repetitions is low because this provides an unreliable estimate of standard deviation. The average value of all results of Taylor and de Bièvre may be designated as the consensus value and the corresponding uncertainty would be high (not calculated, however). Such an evaluation of data is missing but it would be worthwhile to perform because it might show that all laboratories agreed about the content but some laboratories underestimated uncertainty simply owing to recommendations of IUPAC and ISO.[10,17]

Contemporary articles in scientific literature are reluctant to mediate information about outliers[18] and unexpected results that are supposed to occur on a regular basis according to common principles of statistics. Almost perfect straight lines are displayed in conjunction with operational calibrations in analytical chemistry, despite the fact that the responses of the majority of apparatuses are non-linear. In a series of experiments with determination of calcium by flame-atomic absorption spectrometry was found that an average of three repetitions yielded a value of $2.5554 \pm 0.0094$ mg/L that was significantly different from expected 2.66 mg/L. Procedures of IUPAC and ISO[10,17] provided an RSD of 3.3 % whereas the RSD of repetitions RSD(rep) was determined as 0.36 % or almost an order of magnitude lower that the RSD of calibration (RSD(cal)).[12] Further investigations with pooled data (consensus science) showed a 20 % relative uncertainty that immediately explained the observed difference between results, and it explained the relative uncertainty observed for multiple repetitions of the same sample. The discrepancy between RSD(cal) and RSD(rep) is at the heart of the problem. According to scientific methodology these two quantities must be equal but, owing to the limited number of repetitions, they deviate to some extent. It may be stated that they should comply within less than an order of magnitude, which was also found in a series of experiments with different measurands and different technologies.[12]

The average value between ratios of RSD(cal) to RSD(rep) was found to be indifferent from square root of two, which has been explained as earlier as the constant that corrects for correlations between parameters of regression.[19] The correspondence between RSD(cal) and RSD(rep) may be further illustrated by comparing measurements of calcium by FAAS performed during several days of independent series of measurement observing retainment of outliers.[20] Experiments were performed with (Table 2) and without lanthanum modifier (Table 3), in order to investigate the influence of modifier on samples that contained phosphate interferences. In experiments with lanthanum modifier was found an excellent correspondence between SD(cal) and SD(rep) and the SD predicted by Horwitz.[21] The results thus strongly indicate that the empirical formula of Horwitz in fact provides the correct SD and the true SD is much larger than expected from manufacturers' indications. An overall RSD of approx. 20 % is the correct level of relative uncertainty for samples in pure samples. Accordingly are uncertainties higher than 20 % expected for real samples of more complex matrices, such as samples of blood.

*Table 2*

Determination of calcium by FAAS using a relatively low number of repetitions (N(rep) = 27). The number of data of pooled data for the regression line was N(cal) = 180.[19] Lanthanum modifier was added in order to correct for phosphate interference in sample no. 2. LLA = 0.15 mg/L

| Phosphate | Sample | x | Certified | $s_x$(rep) | $s_x$(cal) | Horwitz eq. 1 |
|-----------|--------|------|-----------|-------|-------|---------------|
| No | 1 | 2.49 | 2.66 | 0.50 | 0.45 | 0.37 |
| Yes | 2 | 2.46 | 2.66 | 0.50 | 0.45 | 0.37 |
| No | 3 | 1.43 | 1.50 | 0.31 | 0.26 | 0.23 |

Similar experiments were performed with standards and samples without modifier (Table 1). In these series of experiments was found a less protruding correspondence between SD(cal) and SD(rep) but the lack of correspondence might be a result using a low number of repetitions (N(rep) = 30). This indicates that construction of the regression line corrects for day-to-day variations of the apparatus but recent results[12] show that the deviation may be accidental because large biases may be expected for most types of apparatuses despite using higher numbers of repetitions.

The present investigations indicate that consensus science provide results of elevated uncertainties and uncertainties to good approximation follow predictions by Horwitz.[21] The empirical formula of Horwitz may thus be explained in terms of the uncertainty predicted by pooled calibrations of regression lines. Relative-standard deviations (RSD's) or coefficients of variations (CV's) for all types of measurands and all types of apparatuses may be estimated by the empirical formula (eq. 1) of Horwitz:[22]

$$CV = 2^{1-0.5 \cdot \log(x)} \cong 2 \cdot x^{-0.1505} \tag{1}$$

where 'x' is concentration measured in gram of solute per gram of solvent (w/w). The CV may also be calculated by utilizing the equation for determination of CV of a regression line:[20]

$$CV = \frac{s_x}{x} \cdot 100\% = \frac{100}{m} \cdot \sqrt{\frac{s_b^2}{2 \cdot x^2} + \frac{a^2 + s_m^2}{2}} \tag{2}$$

where parameters have been defined earlier.[20] By considering the extremes, the relation to the empirical equation (2) may be approximated as follows:

$$CV \xrightarrow{x \to 0} \frac{100}{\sqrt{2}} \cdot \frac{s_b}{m} \cdot x^{-1} \% \tag{3a}$$

$$CV \xrightarrow{x \to \infty} \frac{100}{\sqrt{2}} \cdot \frac{1}{m} \cdot \sqrt{a^2 + s_m^2} \% \tag{3b}$$

*Table 3*

Determination of calcium by FAAS using a relatively low number of repetitions, N(rep) = 30. The number of data of pooled data for the regression line was, N(cal) = 162.[19] No modifier added, in order to study effect on result of sample no. 2.
LLA = 0.17 mg/L

| Phosphate | Sample | x | Certified | $s_x$(rep) | $s_x$(cal) | Horwitz eq. 1 |
|-----------|--------|-------|-----------|------------|------------|---------------|
| No | 1 | 2.64 | 2.66 | 0.17 | 0.57 | 0.37 |
| Yes | 2 | 2.590 | 2.66 | 0.075 | 0.54 | 0.36 |
| No | 3 | 1.467 | 1.50 | 0.067 | 0.35 | 0.22 |

The power function of eq. 1 is then identified in eq. 3a but the empirical relation (eq. 1) does not approach a constant value at high concentrations as predicted in eq. 3b. All equations 2, 3a and 3b were derived upon the basis of pooled data, and this principle has proved useful also for the assessment of in-house experiments[12] were a startling resemblance with results of ILC's was identified.[11,15]

Consensus science may provide results that lead to incomplete information about natural levels of chemical species in blood samples. It is believed that the natural level of e.g. vitamin D is approx. 30 microgram/L in healthy adults.[23] However, additional measurements show that results deviate significantly, as dependent on the type of technology used for the analysis.[24] Since testing for the content of vitamin D is a common clinical analysis performed by the general practitioner is it important to focus on the concept of accuracy rather than focusing on precision. Contents are typically associated with large standard deviations and average values are very different, which leaves the consensus value virtually undetermined.[23–26] These results show that consensus science may reveal a genuine problem of analytical chemistry where more basic research is needed, in order to assess reliable values that are reproducible.

**CONCLUSIONS**

Determination of physical constants and concentrations of unknowns such as CRM's have common features related to level of uncertainty. The concept of consensus science means that results should be collected from many different laboratories, in order to provide information of accuracy that is the genuine interest of science, as opposed to precision. It was shown that good correspondence between predicted and observed uncertainty was obtained by using the concept of pooled calibration to estimate predicted uncertainty. Consensus science means that a generally higher level of uncertainty should accompany measurements of analytical chemistry thus promoting agreement between results obtained by several independent laboratories. Results of consensus science resemble results of ILC's thus revealing that outliers

most likely have been removed in earlier investigations of analytical chemistry, in order to reach relative uncertainties of only a few percent. Implementation of consensus science inevitably leads to rejection of practices and procedures of IUPAC and ISO.

# REFERENCES

1.   J. E. T. Andersen, W. Buchberger and P. Worsfold, *Anal. Bioanal. Chem.*, **2013**, *405*, 5361-5364.
2.   H. Emons, *Trends Anal. Chem.*, **2013**, *48*, xv-xvi.
3.   J. E. T. Andersen, *Anal. Bioanal. Chem.*, **2013**, *405*, 7177-7179.
4.   A. Williams, *Accredit. Qual. Assur.*, **2008**, *13*, 633-638.
5.   J. E. T. Andersen, *Anal. Bioanal. Chem.*, **2014**, *406*, 6081-6082.
6.   J. E. T. Andersen, *Cent. Eur. J. Chem.*, **2013**, *12*, 194-205.
7.   I. Kuselman, F. Pennecchi, A. Fajgelj and Y. Karpov, *Accredit. Qual. Assur.*, **2012**, *18*, 3-9.
8.   BIPM; IEC; IFCC; ILAC; ISO; IUPAC; IUPAP; OIML. International vocabulary of metrology - Basic and general concepts and associated terms (VIM3) *http://www.bipm.org/utils/common/documents/jcgm/JCGM_200_2012.pdf* (accessed Jun 7, 2013).
9.   L.-L. Løbger, H.W. Petersen and J. E. T. Andersen, *Anal. Lett.*, **2008**, *41*, 2564-2586.
10.  ISO. ISO 5725 Accuracy (trueness and precision) of measurement methods and results *http://www.iso.org/iso/home.html* (accessed Apr 30, 2013).
11.  M. Bednarova, Y. Aregbe, C. Harper and P.D.P. Taylor, *Accredit. Qual. Assur.*, **2006**, *10*, 617-626.
12.  J. E. T. Andersen, *TrAC Trends Anal. Chem.*, **2014**, *60*, 16-24.
13.  R. Price and J. Zizka, *Report of Arizona State University https://www.phys.ksu.edu/personal/rprice/SpeedofLight.pdf* **2002**, 1-9 (accessed May 22, 2015).
14.  P. Giacomo, *Metrologia*, **1984**, *20*, 25-30.
15.  P. De Bièvre and P.D.P. Taylor, *Fresenius. J. Anal. Chem.*, **2000**, *368*, 567-573.
16.  L. Van Nevel, P.D.P. Taylor U. Örnemark and P. De Bièvre, *Accredit. Qual. Assur.*, **1998**, *3*, 444-446.
17.  K. Danzer and Ll. A. Currie, *Pure Appl. Chem.*, **1998**, *70*, 993-1014.
18.  I. Taverniers, M. De Loose and E. Van Bockstaele, *TrAC Trends Anal. Chem.*, **2004**, *23*, 535-552.
19.  J.E.T. Andersen, *J. Chem. Educ.*, **2009**, *86*, 733.
20.  J.E.T. Andersen and H.S.H Alfaloje, *Rom. J. Chem.*, **2013**, *58*, 691-703.
21.  W. Horwitz, *Anal. Chem.*, **1982**, *54*, 67A-76A.
22.  W. Horwitz and R. Albert, *Fresenius J. Anal. Chem.*, **1995**, *351*, 507-513.
23.  M.F. Holick, R.M. Biancuzzo, T.C. Chen, E.K. Klein, A. Young, D. Bibuld, R. Reitz, W. Salameh, A. Ameri and A.D. Tannenbaum, *J. Clin. Endocrinol. Metab.*, **2008**, *93*, 677-681.
24.  G.D. Carter, J.L. Berry, E. Gunter, G. Jones, J.C. Jones, H.L.J. Makin, S. Sufi and M.J. Wheeler, *J. Steroid Biochem. Mol. Biol.*, **2010**, *121*, 176-179.
25.  L. Carlucci, G. Favero, C. Tortolini, M. Di Fusco, E. Romagnoli, S. Minisola and F. Mazzei, *Biosens. Bioelectron.*, **2013**, *40*, 350-355.
26.  S.-Y. Kong, C.-S. Ki, H.-J. Kim, K.-O. Lee, J.-C. Bae, S.-H. Kim and J.-W. Kim, *Clin. Chem.*, **2004**, *50*, 2193-2195.