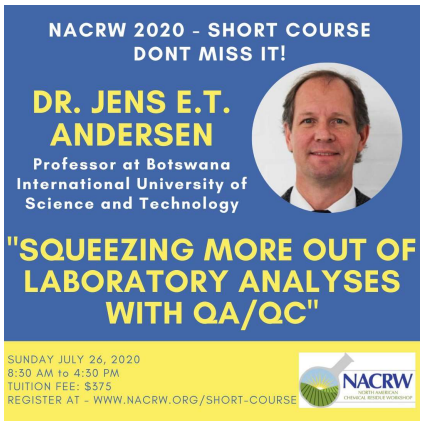
The Principle of Pooled Calibrations as Alternative to Conventional Practices and Procedures of QA/QC and Metrology

Jens E.T. Andersen

Botswana International University of Science and Technology, Faculty of Sciences, Department of Chemical and Forensic Sciences, Plot 10071, Boseja Ward, Private Bag 016, Palapye, Botswana; andersenj@biust.ac.bw

Analytical chemistry (AC) encompasses application of virtually any apparatus that is suitable for quantitation from weighing and pH measurements to liquid-chromatography-mass spectrometry and synchrotrons. All the technologies must be validated according to practices and procedures of QA/QC that have been developed by international organizations mainly for the industry over the past three decades. Introduction of QA/QC to academia is not as complete as it is to the industry, and some of the issues of these efforts to promote QA/QC in science are highlighted. Conventional methods of QA/QC and metrology have some built-in features of method validations that may provide contradictory results and disagreements even between professional laboratories. Recent results of Eurachem concur with the celebrated results of Horwitz, as they both claim that many of such discrepancies may be explained by incompetence or lack of training of laboratory staff. These findings have recently been contested, as evidence as to poor performance of apparatuses with respect to quantitation is more likely to be another source to the disputes. Introduction of the principle of pooled calibrations suggests that reproducibility can be assessed by a single laboratory. The uncertainty supplied by the manufacturer of the apparatus informs about precision, but it does not address the concept of accuracy. It is proposed that generally higher levels of uncertainty should be assigned to most apparatuses during the method validation of the laboratory. This will deliver long-lasting method validations that do not cause any stop in the production and it gives no disagreements between results of different laboratories.

Contribution of QA/QC to 57th NACRW



NACRW
NORTH AMERICAN CHEMICAL RESIDUE WORKSHOP

Dear Colleagues,

It is my pleasure and privilege to participate as presenter to your esteemed NACRW. Thank you very much to Steve, Jack, Katie, Mike, Teri and Paul for inviting me to give both a short course and a presentation on QA/QC. It also been great to meet so many members of your organisation at our on-line meetings during the lockdown. On the left-hand side, you see the advert to the short course on QA/QC that will be available to participants of the next edition of NACRW in 2021. I very much hope that the short course will be able to attract both young and senior scientists to discuss this important topic of QA/QC that influence the work of professionals, where to some extent, scientists of academia feel themselves excused or perhaps excluded from the development of QA/QC. The title is not the same as the title of today's presentation. Let us face it, nothing is more boring and uninteresting that QA/QC. Sorry colleagues, but this is really not interesting, as I can prove it because I have tried to communicate QA/QC to scientists of academia for decades. They simply do not want it to be part of scholarly work. Certainly, nobody can force QA/QC upon the work of scientists but experience from the courtroom tells me that it might be a good idea if they learn about some of the issues that people are dealing with in the laboratories of the industry. Believe me, there is much to learn!

BIUST
BOTSWANA INTERNATIONAL UNIVERSITY OF SCIENCE & TECHNOLOGY

Department of Chemical and Forensic Sciences

The Principle of Pooled Calibrations as Alternative to Conventional Practices and Procedures of QA/QC and Metrology

Presenter: Jens E.T. Andersen

Date: August 11th, 2020

Venue: Webinar 57th NACRW

www.biust.ac.bw

BIUST | Driving Change

QA/QC has developed over the years and it originates from physics, materials science, life sciences and metrology. For many years certified reference materials have successfully secured traceability and reproducibility. Companies have used the principles of QA/QC to develop LEAN technology that constitutes the backbone in the bodies who are supervising quality in companies worldwide. Since 1951, the ISO has delivered standards to companies but, in many cases, the standards were inaccessible to academia, owing to the high costs of the guidelines. However, BIPM and IUPAC developed in parallel guidelines that were available to a wider community and the IUPAC-BIPM constellation still publish guidelines that are available free of charge to everyone . In 1982 Horwitz published his celebrated paper on CV – value that attracted much attention because it documented that scientists are not always equally careful with delivering quality to their laboratory work. Hence, more education and training was declared to be the way forward. This notion was backed by the European Commission and International organisations comprising NIST, Eurachem and many others. Through the IMEP programme, the notion of the need for more education and training was confirmed by numerous publications and it led to development of a whole new range of standard reference materials and certified reference materials.

Botswana

Just a few words about Botswana that is know for the low density of population, the Okavango Delta, the Kalahari Desert, wildlife and diamonds. Diamonds constitute approx. 80 % of the country's GDP, and that's why the Government has decided to expand on the knowledge-based economy, in order to further diversify the economy. Hence the construction of BIUST that is situated in Palapye, a village 300 km north of the capital of Gaborone. In 2019, scientist found that the origin of the human race may be found in the northern region of the country between the Okavango Delta and the Chobe National Park. Just to induce curiosity to those of you who may be interested in finding your roots; well here they are, the outmost tips of your roots.

Agenda

- **Introduction and take-home messages**
- **Precision**

  - ISO 17025, 5725, IUPAC

  - BIPM, VIM3, Guide to the expression

of uncertainty in measurement (GUM), Quantifying Uncertainty in Analytical Measurement (QUAM)

All these organisations are struggling to make the best possible prediction of precision

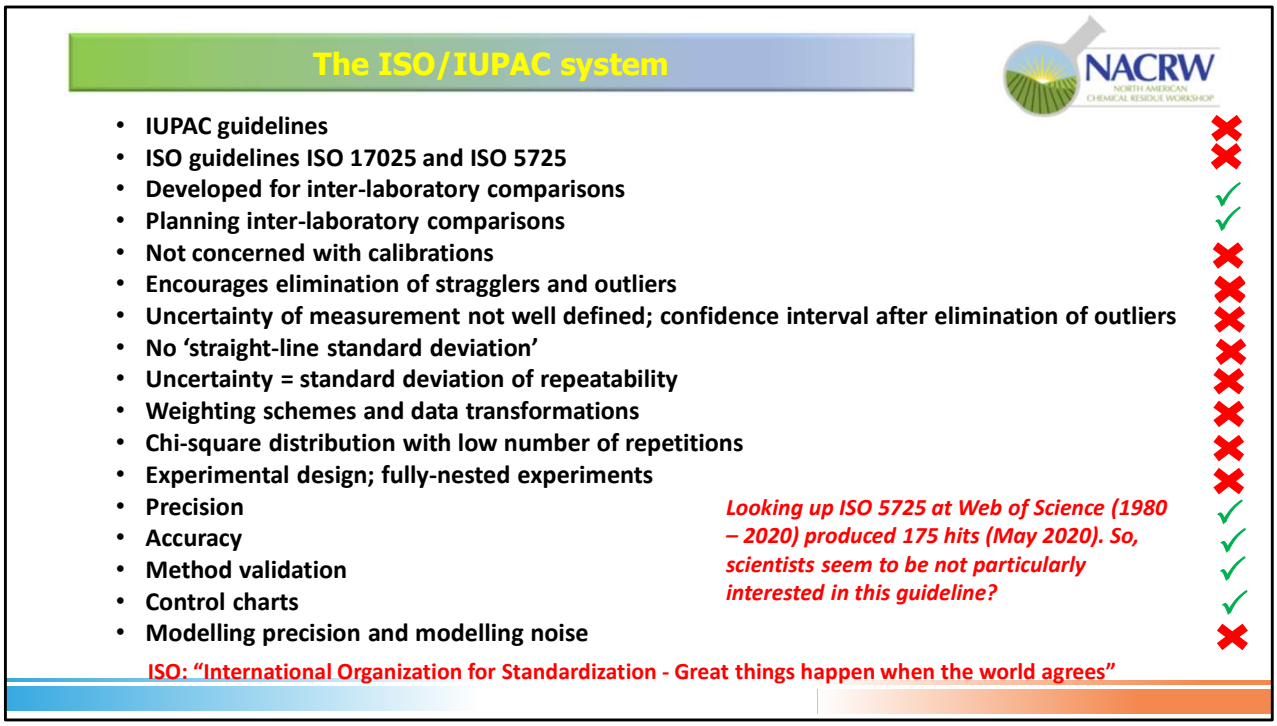
- **Accuracy**

  - The principle of pooled calibrations (PoPC)

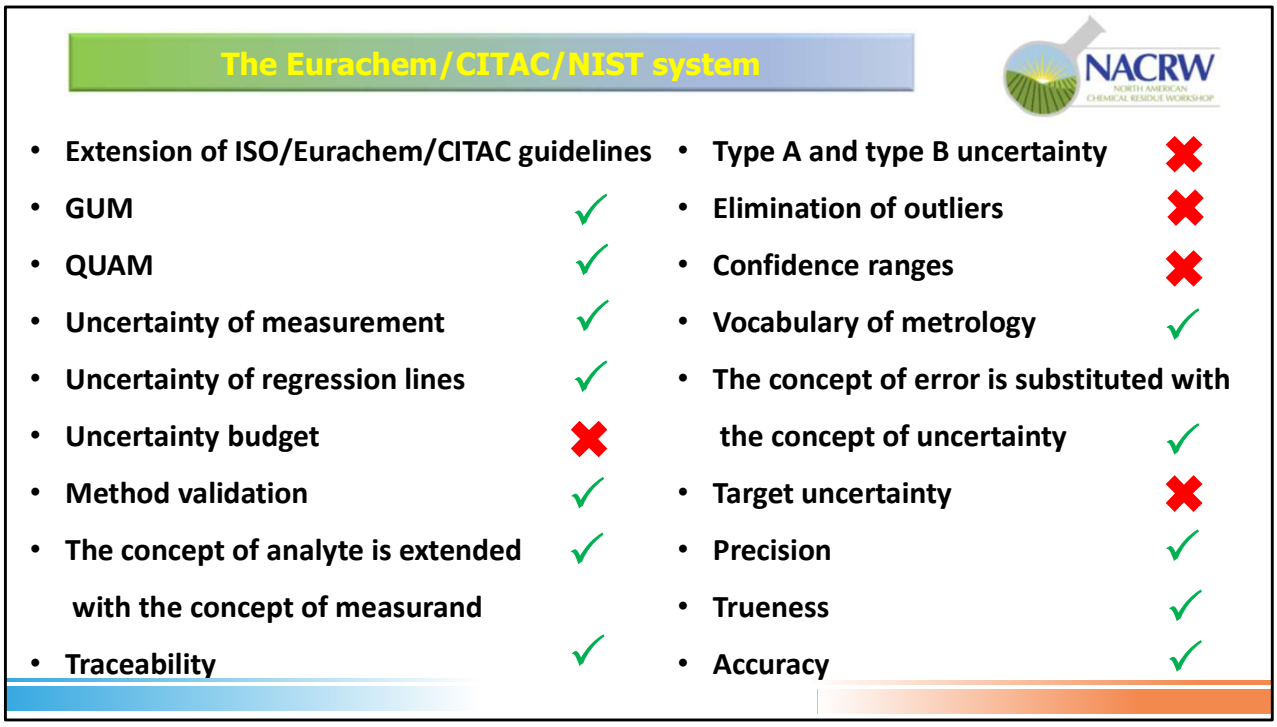
  - On the origin of the Horwitz equation

After this presentation, I would like you to increase your knowledge of uncertainty of measurement that is fundamental with respect to understanding the concepts of QA/QC. QA is maintained through method validations and QC is the laboratory manager establishes procedures of QC to supervise the performance of the laboratory and to secure the quality of the services. It has been recognised for centuries that QA/QC and certified reference materials are available to ensuring uniform levels of standards and to providing high levels of quality in trade and manufacturing. The guidelines of ISO have been generally available to QA/QC for approx. 70 years but they have never really been used within the are of science. The ISO guidelines were never really adopted by the scientific communities but recent attempts to improve the guidelines have increased the interest to consider applications of the methods of ISO in science. Despite the recent publication of the Eurachem/CITAC guide and its incorporation in the BIPM list of guidelines, there are still serious issues of quality that need be dealt with. One of the main issues is the lack of scientific methodology in the new guidelines. Scientific methodology is not mentioned and it seems to be not a prerequisite to understanding measurements of laboratories, which eventually leads to not only scientific disputes but also to disputes that can only be resolved by litigations. Since the majority of guidelines are concerned with delivery of precision and not accuracy, a high number of disputess are expected to be resolved by litigations in the industry. In 1982, Horwitz published his celebrated results of analytical chemistry, and he interpreted the lack of correspondence between results of the scientific

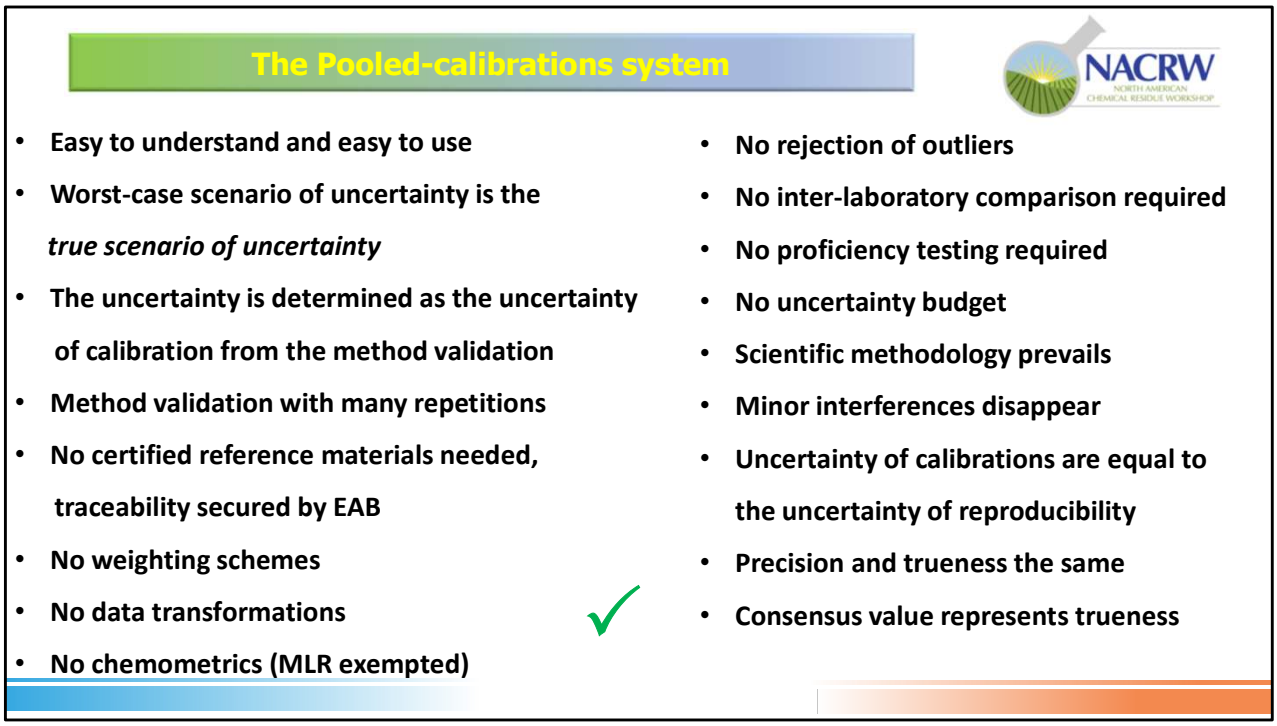
literature as a lack of training and lack of competences among laboratory staff. This notion was later confirmed by the IMEP programme where international organisations found huge discrepancies between results that were delivered by professional laboratories. However, there is something wrong with that interpretation, as it was found that uncertainties increased when the concentrations were low, which raises the question: Why would laboratory staff be more careless when handling low concentrations? This makes no sense, and there is an alternative explanation that seems to provide the true answer to that question. The uncertainty of measurement inherently depends on concentration and the uncertainty of measurement increases rapidly at low concentrations because of the lack of detectors' sensitivity when concentrations are low. Although detectors may provide linear responses, it was found that the slopes, intercepts and determination of concentrations of samples differed significantly between days of measurement. This observations clashes with the conventional understanding of calibration that is supposed to eliminate the influence of day-to-day variations of the analytical result. Therefore was developed the principle of pooled calibrations that satisfactorily provides the means to explain the day-to-day variations and supply the consensus values that everybody can agree about worldwide. The consequence of introducing the principle of pooled calibrations (PoPC) is a much higher level of uncertainty being assigned to the measurements, and in many cases it is in good agreement with the Horwitz equation. However, there are many cases where the PoPC can be used where the Horwitz equation cannot be used, as it will be demonstrated in this presentation.



Any change requires approval of at least 75 % of the member bodies. In this overview is indicated that I find them troublesome, inconvenient or simply not applicable to understand the comparison of data between laboratories. Without going in much detail here, as more details will be given at the short course, the handling of outliers may be highlighted as one of the major issues of QA/QC. Trueness is not in focus, as accuracy ~ trueness and precision. It should be noted that scientists rarely use QA/QC in the first place but when they do, they are not following the ISO guidelines, owing to the inherent self-contradictions of the system.



Major progress in QA/QC was made when international organisations devised a new scheme of QA/AC that is represented by the Eurachem/CITAC guide. The Eurachem/CITAC guide was adopted and endorsed by BIPM together with the preceding ISO system. Although it may be regarded as a step forward, the Eurachem/CITAC guide recommends a system that creates havoc in both industry and science. There is no mention of fulfilment of scientific methodology that is otherwise a fundamental concept of science. Hence, there is no requirement for correspondence between predicted and observed uncertainty. In addition, the outdated ISO system is maintained and it results in gigantic additional workloads to professional laboratories who are destined to end up in even more litigations over disputes that could have been resolved very easily by not following these guidelines.



So, which guidelines are we then supposed to follow, if the purpose is to obtain agreement of measurements worldwide? Well, one way of reaching this goal would be to consider the work of Horwitz and accepting that the elevated levels of uncertainties he observed in his investigation were not a result of lack of skills or lack of competence among members of laboratory staff, but it is simply related to the way in which the apparatuses' detectors are working. Within a short time frame, they will give excellent linear responses, but over long periods of time, the linear responses are significantly different, which has not been recognised by neither Horwitz, ISO, or any other international organisation. The reason for this is easy to understand, as recognition of the lack of long-term stability of detectors imposes serious issues with determination of uncertainty of measurement. Therefore, the worst=case scenario was introduced to investigate into the significance of accepting all data and not rejecting any outliers, as long as the outliers did not occur as a result of errors of handling chemicals or apparatuses. The general idea is, that outliers cannot be rejected on statistical grounds alone; there must be some other reason for erasing data from data sets. This approach is denoted as the principle of pooled calibrations (PoPC), that reveals serious issues with decision making within all areas of science. After the introduction of the PoPC, scientists can no longer disregard the application of QA/QC in the search to understanding fundamental mechanisms of nature and delivery of accurate results to the public. If the PoPC was adopted by science, it would mean a radical change of scientific methodology as we know it. Numerous technologies of analytical chemistry were tested according to PoPC,

and, so far, the PoPC was found to never fail.

Why do we perform a method validation?

To make sure that the method is fit for purpose

- **To make sure that the uncertainty of the sample is not influenced by the uncertainty of the apparatus**
- **To make sure the uncertainty that is generated by the apparatus does not influence the decision making with respect to determination of the contents of samples**
- **To make sure that all measurements are clear of the noise range**
- **To determine the uncertainty as a function of concentration for all the chemical species**
- **To make sure that the predicted uncertainty corresponds to the observed uncertainty**
- **To deliver results that are universally correct to the customers**

The generally accepted reason for introducing method validations is that it should be fit for purpose. To me, this indicates that as long as the laboratory and the customers are happy with the results, everything is fine, even if the results are completely wrong. In order to improve quality, as to strive for high quality, it may be useful to consider some additional goals of the method validation. No doubt, it is important to perform a full method validation, but it should be aimed at delivering tools of decision making that makes it possible to provide universally correct results to the customers. Therefore, the method validations should provide the means for the laboratory to predict all future uncertainties of analysis and make sure that the results are outside the noise range of the methods. It should not be possible for a customer to go to another laboratory and get another result that is significantly different from the result that was delivered by your laboratory.

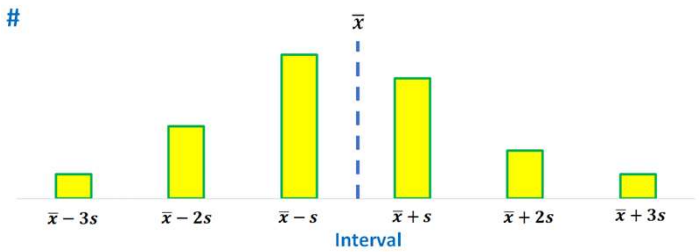
Core ingredients of QA/QC

$$\bar{x} = \frac{\sum x_i}{N} \pm 2 \cdot s$$

Average value

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{N - 1}}$$

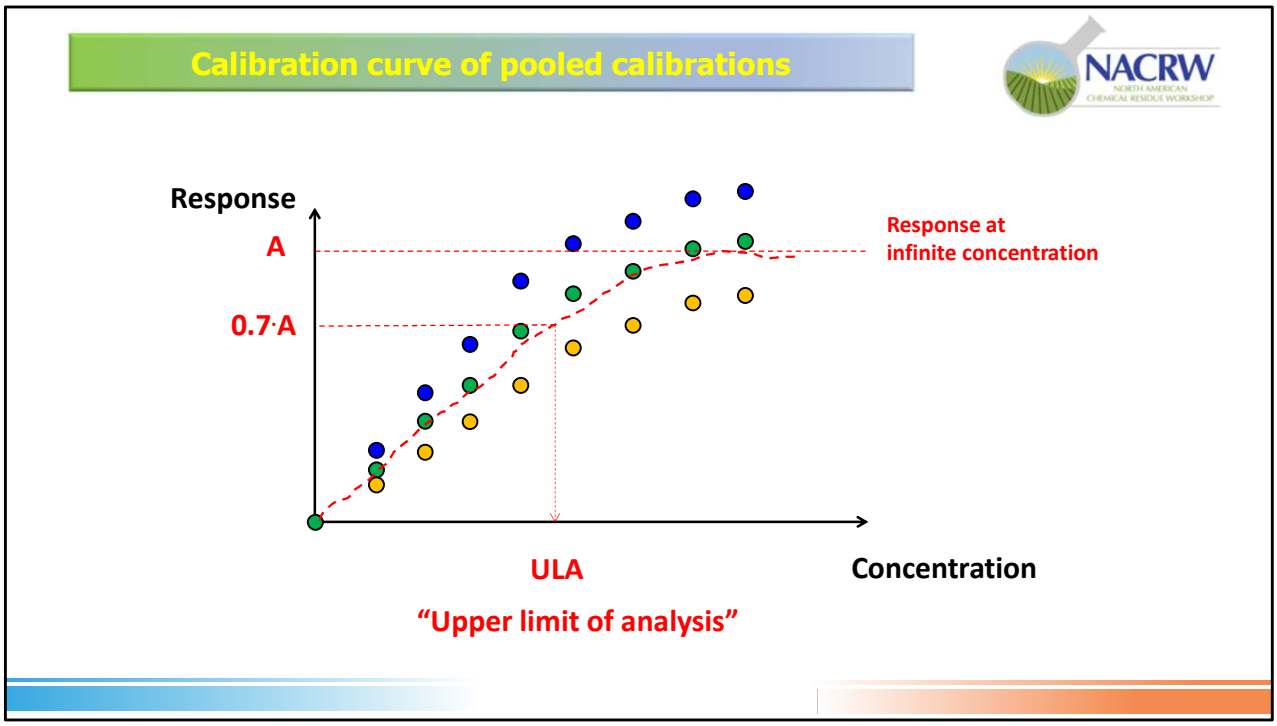
Variance or standard deviation



Canonical distribution

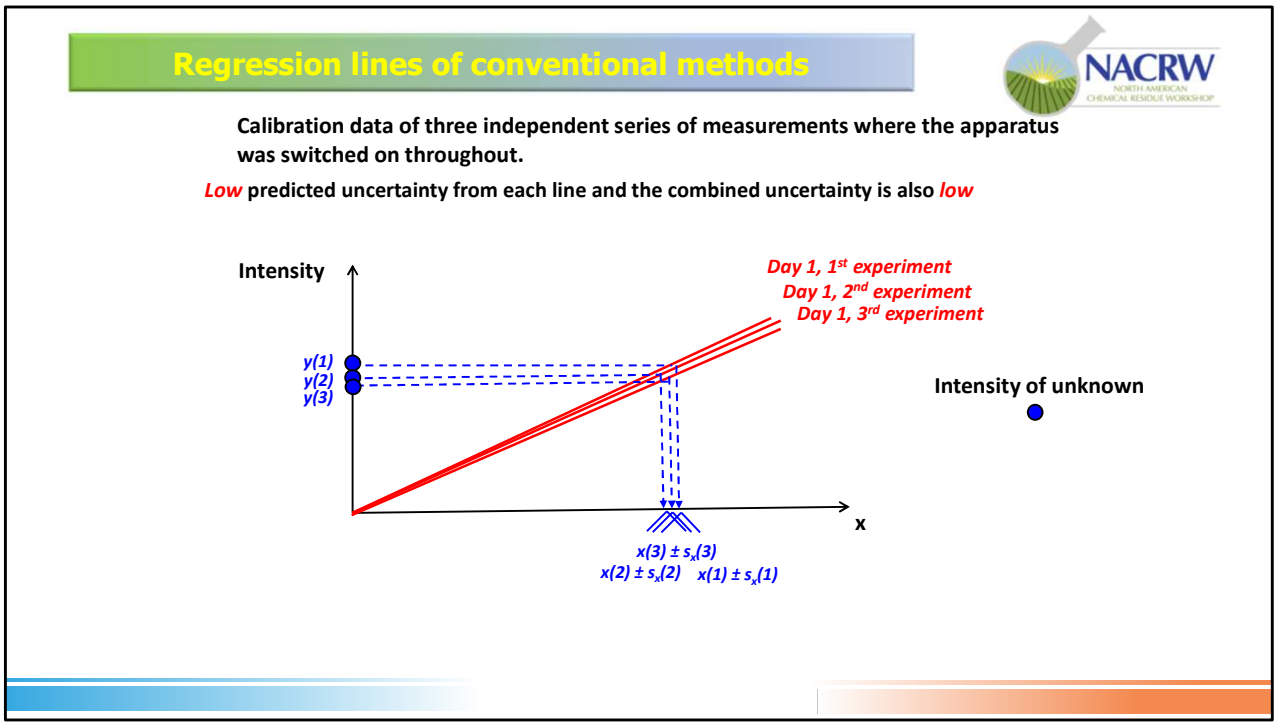
I promised Steve to not make this presentation into a series of mathematical equations, but it may be useful to make a recap on the understanding of some of the fundamental concepts of analytical chemistry, just to make sure that we are on the same page from the beginning of the presentation. In many instances, there is a clash between the understanding of the methodologies by mathematicians and other scientists. We agree about the way a results should be reported; an average value +/- the standard deviation where the latter includes division with N-1 and not N. If we have a large number of repetitions, that we must have, then it makes no difference to us to subtract 1 or not subtract 1 from the number of repetitions. Only when precision is considered as a result of a low number of repetitions, it becomes relevant to subtract 1. However, under such circumstances, the standard deviation becomes highly unreliable and it makes no sense to perform testing, such as F-testing and t-testing, which effectively hampers comparison and decision making when the number of repetitions is low. Accordingly, it should be evident that all results of analytical chemistry should be reported only after a high number of repetitions. The PoPC anticipates that the standards respond to detectors the same way as that of samples. Thus, the samples may be regarded as part of the population of standards, unless serious interferences influence the value of the results. Therefore, it is of paramount importance to perform a comprehensive method validation with 100 – 200 repetitions where the apparatuses are switched off and on in between many series of experiments. Otherwise the uncertainty of standard and samples are not modelled correctly, and

typically, the uncertainties are may be vastly underestimated, which his one of the main reason for ending up in court when the customers cannot accept the results. Suppose that we subtract/add one standard deviation of the mean, two standard deviations from the mean and prepare a distribution of the data. Then, it is important to realise that the frequency of data exhibits a peak. Otherwise, in case you observe a square distribution, that would be a sign of random data being measured by the method. This is likely to occur when many outliers are rejected when using a method that is essentially not working properly. This situation can be illustrated by generating random numbers between 0 and 1 in an excel spreadsheet and subsequently depicting the distribution by means of the histogram function. That provides a square distribution that should never be observed in analytical chemistry, as it indicates that the method is nothing but a generator of random numbers. This observation suggests that enough data should be available to check the corresponding distribution, just to make sure that the average value has a genuine meaning according to an analytical result.

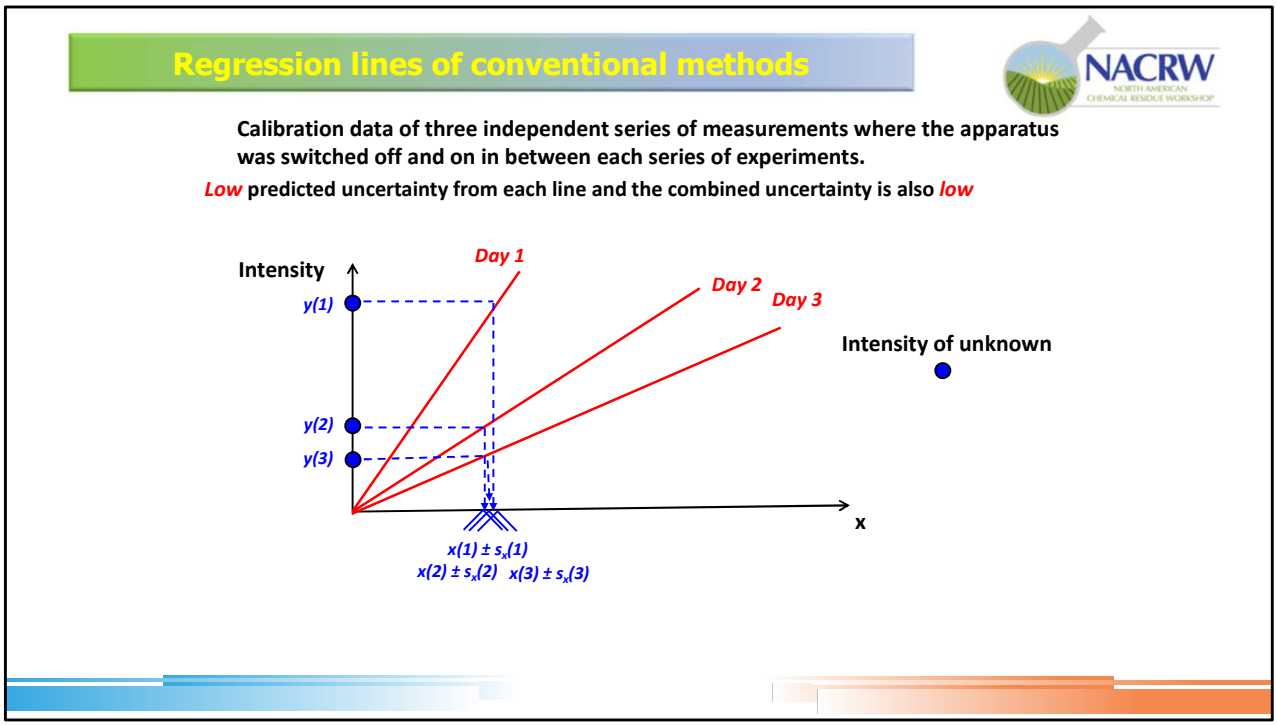


Today, it is possible to perform calibrations by means of non-linear calibration and many manufacturers are using non-linear calibrations to their technologies. However, it is not easy to calculate the corresponding uncertainties, which poses a problem to reporting results of analytical chemistry. It is possible to linearize some equations and transform a non-linear problem into a linear problem, but this approach is outdated by the development of computers. With an excel spreadsheet at hand, there is no need to linearize any equations, as it is straightforward to perform curve fitting and calculate corresponding coefficients of regression. The world has gone digital, which means there is no longer a need to linearize equations such as the Lineweaver-Burk equation or the Arrhenius equation. All calibration curves are inherently non-linear, but we only utilize the linear range, in order to maximize sensitivity and to facilitate uncertainty calculations. Emission technologies are generally recognised for their extremely high sensitivities, which often make manufacturers cut off the signals of the detectors after a certain magnitude of responses. That makes the response versus concentration linear up to the limit of response that is defined by the cut off value. Although detectors have different characteristics, the majority of response curves exhibits the features shown on the graph where the sensitivity is at its maximum within the linear range of responses and then it decreases to become constant at large concentrations. It is possible to use non-linear response curves to expand the calibration interval but that complicates the calculations of uncertainty of measurement to a degree that makes it difficult to handle under normal operating

conditions. Hence, we revert to the straight line of regression that provides reasonable equations to calculate the corresponding uncertainties as a function of concentration. Any analytical chemist is familiar with the concept of the linear range, but the method of defining the upper limit of analysis (ULA) or the maximum concentration to the calibration line is not explained in any of the guidelines of ISO, BIPM, Eurachem or others. It seems like a regression coefficient close to one is the only criteria that is available to decide on the interval of linear response. Within the framework of the PoPC, the linear range is found where the mathematical expansion to first order of the response curve is no longer valid, which occurs at approx. 70 % of the maximum responses of very high concentrations. Therefore, as part of the method validation, the maximum value of every single species should be measured, as it defines the ULA in a straightforward manner.



Suppose that we have determined a series of regression lines of calibrations and we want to determine the concentration of unknowns by means of those calibrations. When the apparatus is heated up to normal operating temperatures and we decide upon the stability of the signals, three consecutive series of measurements may provide calibration lines of almost the same slopes and intercepts and the responses of a single sample. The ensuing uncertainties are relatively low and it does not make much difference to calculate the uncertainty upon the basis of a single calibration line or use all three calibration lines in combination. This may be considered as an ideal situation where uncertainties always come out low to the satisfaction of the customers. However, sometimes the result may be completely wrong, as compared to the value that was expected by analysis of CRMs/SRMs. This latter experience may be devastating to the operators who struggled to make everything work perfectly well with delivery of almost perfect calibration lines. The discrepancy between expected and observed result may attain extraordinary dimensions. Note that these calibration lines all pass through the origin and that is most frequently the case after conducting the operational calibration. However, in the vast majority of publications, you will observe negative or positive values to the intercepts, which indicates that it was not checked by statistical methods if the intercept were significantly different from zero. If the intercept was not significantly different from zero, it should simply be reported as ‘zero’ and whatever the value of the intercept, it should not be used for calculation of concentrations.

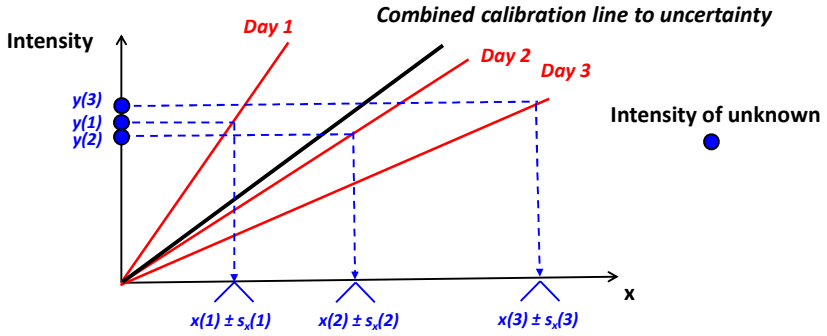


This figure illustrates the expected outcome of the operational calibrations that are performed on a routine basis to eliminate the influence of variations of the apparatuses on the results. Day-to-day variations and non-stability of the apparatus are expected to be compensated by the value of slopes and intercepts of the calibration lines in such a manner that large differences between slopes and intercepts always provide the same result. That corresponds to large differences in signals of the same sample being converted into the same concentration from day to day, irrespective of the uncertainty that was added by the apparatus. However, this is almost never observed in practice, which then provides a genuine paradox of analytical chemistry because it contradicts the very reason for preparing the calibration lines in the first place. In this event, the concentrations of unknowns would be virtually independent of the slopes.

Regression lines of conventional methods

Calibration data of three independent series of measurements where the apparatus was switched off and on in between each series of experiments.

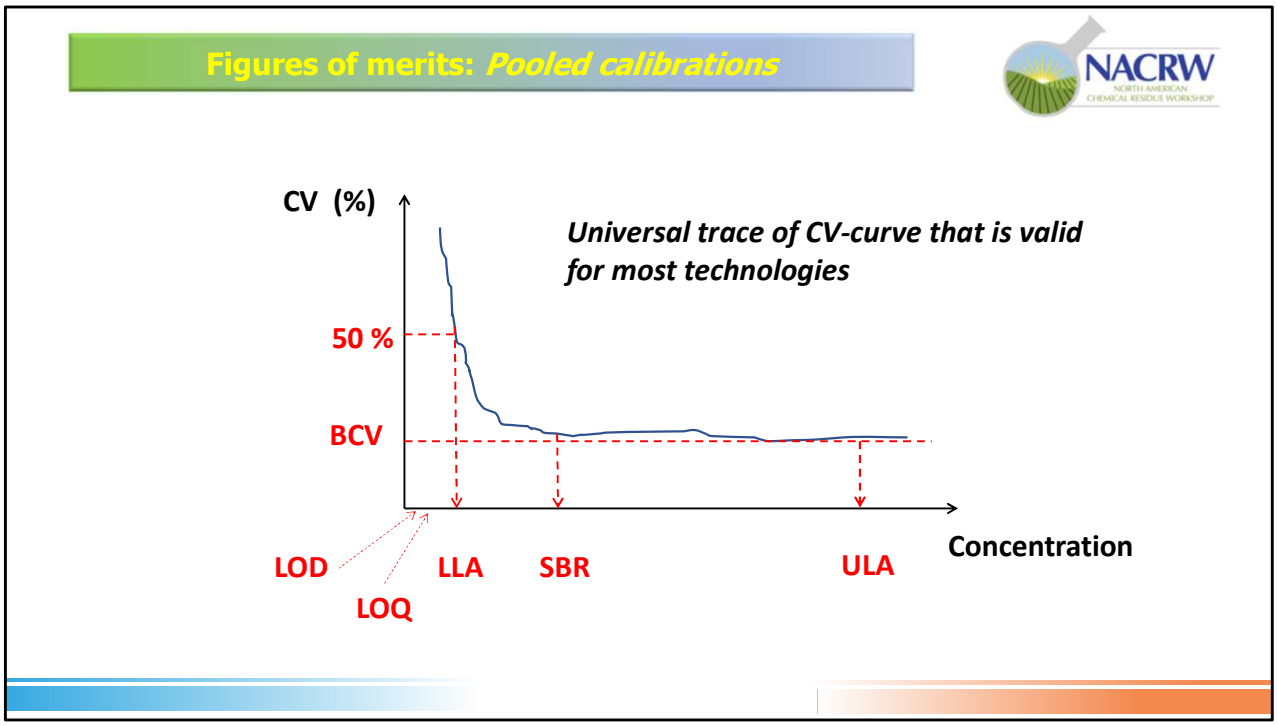
Low **predicted uncertainty from each line but the predicted uncertainty of the combined line is** ***high***



Note: The calibration line is used for calculation of uncertainty only. Concentrations are calculated by means of every single calibration line; three in this case

This scenario is similar to the one that was presented in the previous slide, except that in the present slide, the results between days differ significantly, and they are virtually incompatible. The response values may be almost similar or far apart, but the fundamental requirement to calibrations is not met in this scenario. The operational calibration does not compensate for day-to-day variations, which contradicts the theory. However, the main discovery related to PoPC is that this worst-case scenario of uncertainty actually represents the truth. That is, numerous investigations of the performance of analytical technologies have disclosed this feature of QA/QC that was previously associated with lack of skills, lack of training or lack of competences. The large uncertainties of the PoPC originates from inherent properties of the detectors that cannot be altered or compensated for by statistical calculations. A pH-meter utilises a glass electrode to sense the concentration of protons, and it does that with a universal level of uncertainty that is almost the same for all pH-meters worldwide. The same is true for other types of detectors, but the characteristics of each type of detector is determined by means of the method validation. In terms of uncertainty, it is important to recognize that the average regression line of calibration (full line in black) should not be used for determination of concentrations of unknowns before the true level of uncertainty has been confirmed by the method validation. Hence, it is postulated that once the method validation has been completed, the uncertainty to all future results are determined by the uncertainty of the PoPC. The uncertainty never changes, and it is not necessary to re-validate, unless the detector deteriorates to a

unacceptable low level of sensitivity. Therefore, the method validation should be very thorough and it may consume vast amounts of resources but it will be rewarded during long-term operation of the apparatus because there will be no stop in production of results and the risk of providing inaccurate results will be diminished.



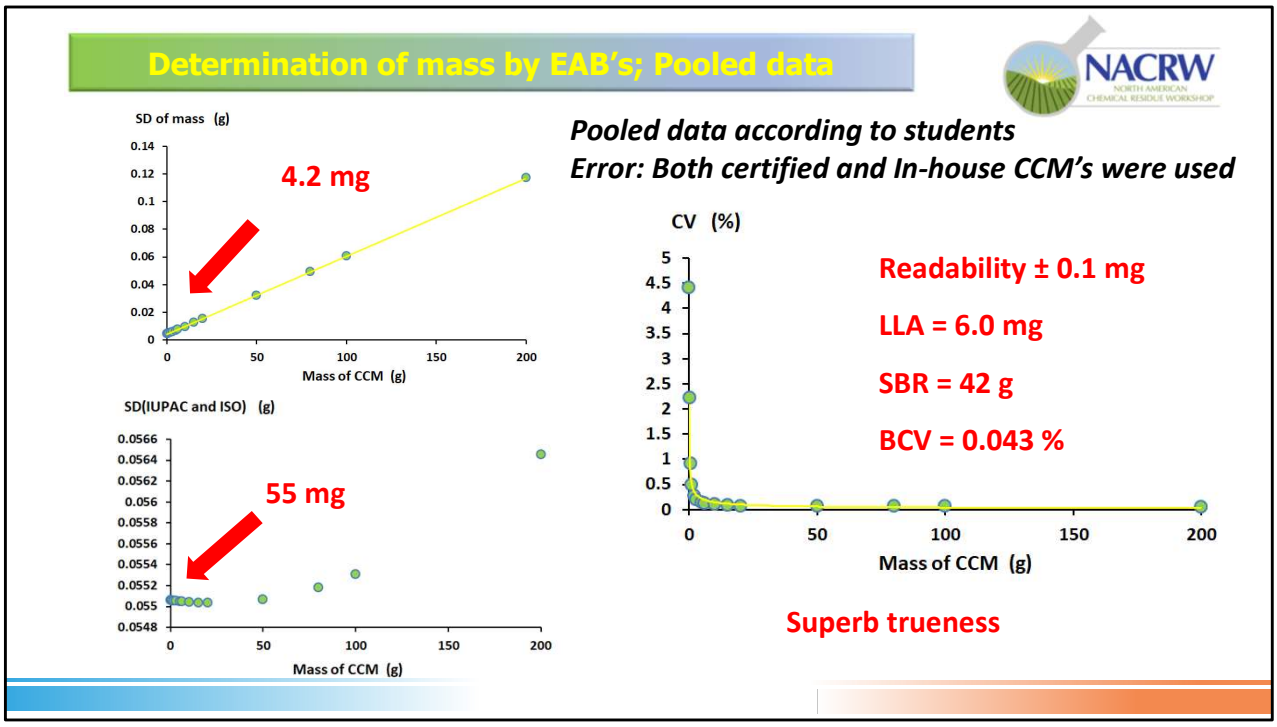
This is a graph that is key to the understanding of the concept of quality itself and to the understanding of the PoPC. The blue curve represents the CV-value as a function of concentration, and it will have this shape for all technologies. However, if only a few repetitions was applied to the analysis or to the regression line, it might look very differently and it may attain any shape because then the standard deviations would be unreliable and comparable to random numbers. As of today, almost any publication of analytical chemistry names the LOD as an important parameter and figure of merits that underpins the notion of reliable analysis at very low concentrations. However, according to the PoPC, the LOD is a completely redundant parameter, as it only tells about the minimum signal of the analyte/measurand that can be registered by the detector. There is no mention about the uncertainty of measurement at the concentration of the LOD and in many cases, it would be construct a calibration line with all concentrations below that of the LOD. After a full method validation by the PoPC, it becomes possible to calculate the uncertainty of the LOD and LOQ, which then indicates if it makes sense to deliver results of concentrations close to the LOQ. The lower-limit of analysis (LLA) is defined as the concentration where the CV-value exceeds 50 %, which translates into an uncertainty of 100 %, and it should be determined by means of the PoPC. For many technologies, the LLA attain values much higher than those of the corresponding LOQs, which is rather disturbing. Especially in mass spectrometry, the difference between LOQ and LLA can be as high as three orders of magnitudes. That is, one thousand times higher or one hundred

thousand percent higher. This is very bad news to science but it also shows the way to go forward. Three additional figures of merits, the SBR, the ULA and the BCV, should be considered for the validation of the method. Above a certain concentration, that is given by the start of best range (SBR), the CV-value and thus the uncertainty remains constant all the way up to the upper-limit of analysis (ULA). Preferably, all solutions should be diluted only to the ULA because the uncertainty is at its minimum at that concentration. However, the uncertainty is virtually the same in the full interval that is determined by the SBR and the ULA. Concentrations that are determined between the LLA and the SBR, may be associated with rather large uncertainties of up to 50 % but that might be acceptable if it were not possible to e.g. pre-concentrate the samples. In summary, it is proposed that this graph that be constructed by means of the PoPC constitute the fundamental prerequisite to define the quality of the performance of an apparatus.

Completed method validations

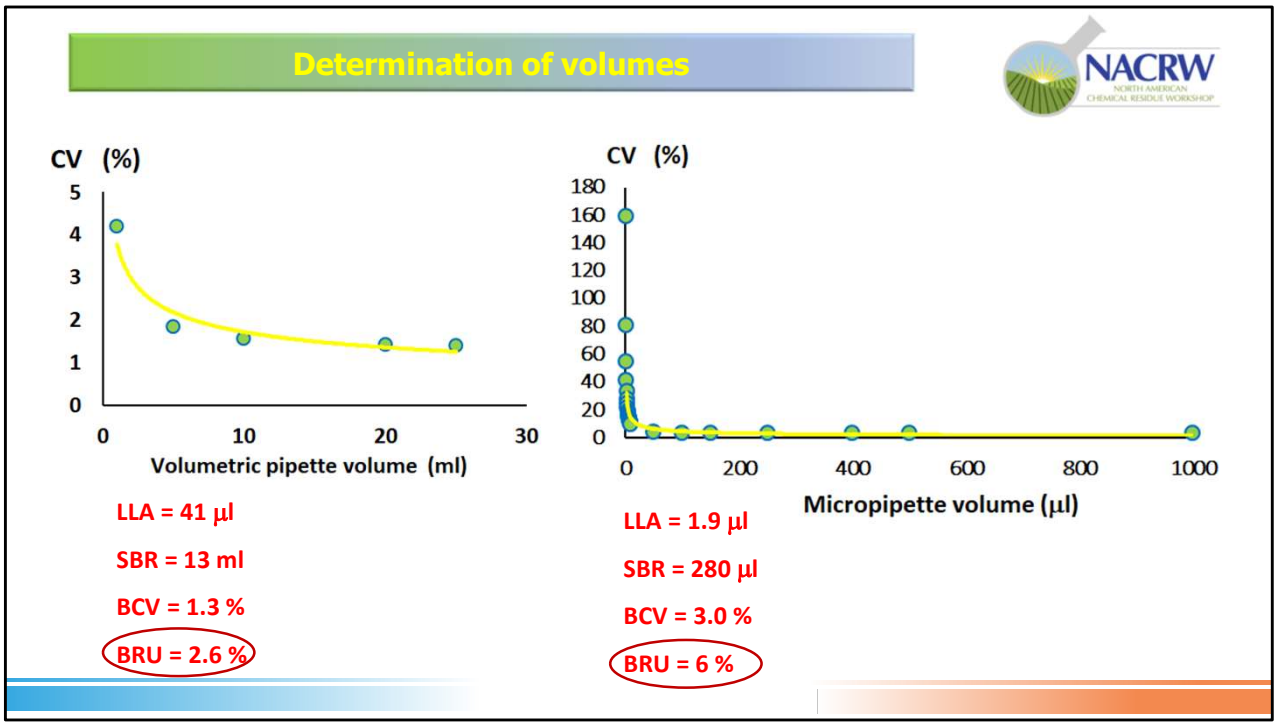
- **Electronic analytical balances (EABs)**
- **Volumetric pipettes**
- **Micropipettes**
- **Ion chromatography (IC)**
- **pH-measurements**
- **Flame-atomic absorption spectrometry (F-AAS)**
- **Graphite-furnace atomic absorption spectrometry (GF-AAS)**
- **Inductively-coupled plasma mass spectrometry (ICP-MS)**
- **Inductively-coupled plasma optical emission spectrometry (ICP-OES)**
- **Liquid-chromatography triple-quadrupole mass spectrometry (LC-QQQ-MS)**

So far, several, but far from all analytical technologies have been validated by the PoPC that provided increased understanding of the performance and limitations of that apparatus when it was used for quantitation. Only for GF-AAS was observed homoscedasticity with respect to uncertainty of measurement but that has only been confirmed in this single investigation that was the first of its kind. Note that the PoPC also works for common laboratory utilities such as glassware and EABs, as we will now demonstrate by a few examples.

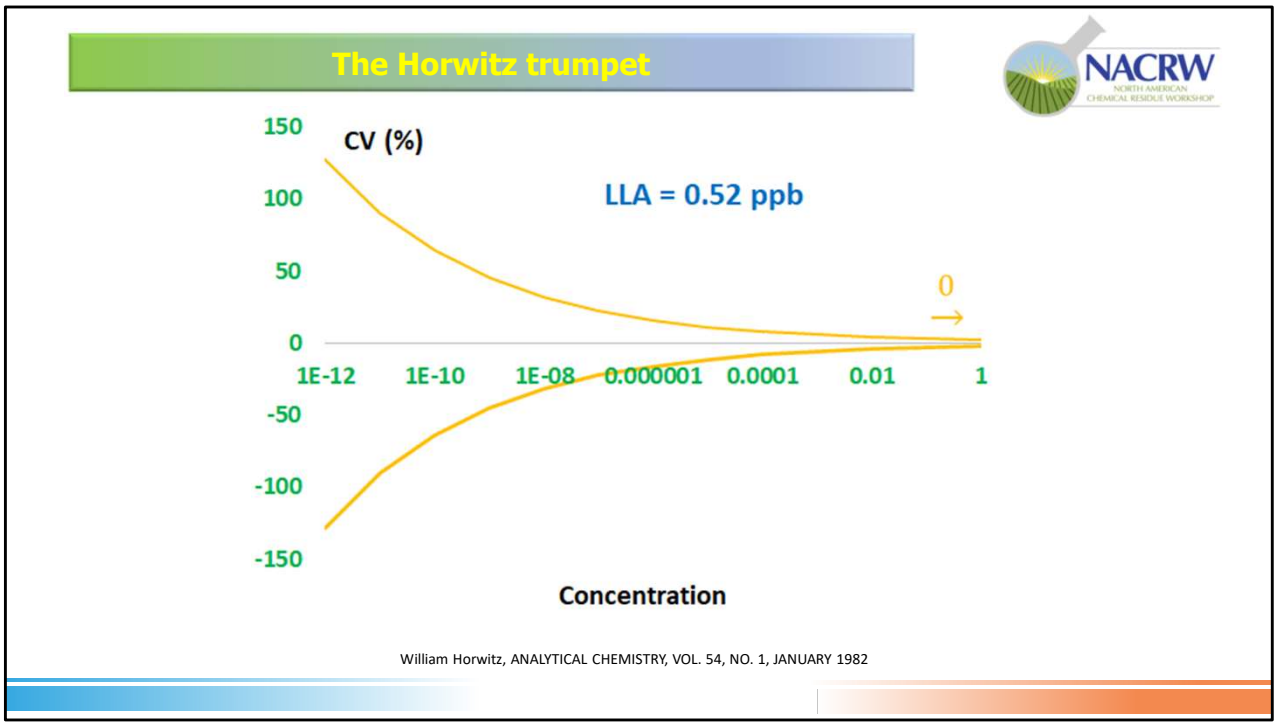


In the following few examples is shown some results from labs of students who followed a module on statistics and chemometrics. Based on experience from previous editions of the module, it was found necessary to carefully instruct the students in handling the laboratory tools before they carried out the exercises. Otherwise, outliers related to incorrect handling of the tools and apparatuses were likely to occur. The careful instructions before each exercise provided results where no major deviations or unexpected large deviations were found among the measurements of the students who obviously understood well the instructions before the measurements were carried out. Statistical outliers were identified but none of them were eliminated, as it is completely unnecessary to reject outliers when there is a large number of repetitions to the measurements, up to 700 in these examples. We committed a slight error, as to which I am also responsible: We used two types of CCM where it was known that small systematic deviations from the certified values were assigned to the in-house CCMs. However, it does not matter for the illustration of the difference between using the PoPC that is based on the law-of-propagation of uncertainty and the IUPAC-ISO method that is based on the variance-covariance matrix. The many measurements of the CCMs showed a non-zero intercept (4.2 mg) to the SD of CCMs. It is not common practice to show the corresponding SD versus mass for the IUPAC-ISO method but for pooled data, it provides a vast overestimated value of 55 mg for the same data. Clearly, the IUPAC-ISO method results in overestimated SDs if pooled data were accepted, which is a commonly recognised issue of QA. At the same time, the SDs of the IUPAC-ISO

method also shows underestimated values to the SDs when the mass, or any other parameter for that matter, is large. The SD of IUPAC-ISO remains almost constant within a relative large interval, which is also a characteristic feature of this methodology. It led to the concept of homoscedasticity where it may be anticipated that SDs should be independent of the parameter of the x-axis. This clearly contradicts the prediction of the PoPC where the SD depends linearly on the x-value. Since it is always found that SDs of results of different laboratories depend linearly on the x-value, it may be concluded that the PoPC present the more suitable description of reality. The important parameter of interest with respect to estimating quality is the CV-value of the RU that is equal to the CV-value multiplied by a factor of two. On the majority of EABs is printed the readability as 0.1 mg that in many cases is mistaken for uncertainty of measurement. This becomes evident when the CV-value versus mass, in this case, is calculated where it attained values of more than 50 % when the mass was lighter than 6,0 mg. Above the start of best range (SBR), the best CV-value (BCV) was 0.043 %, which is extremely low, as compared with other parameters that contribute to the uncertainty budget. However, the SBR was rather large (42 g), so it is important tp use the curve to calculate uncertainties when the masses are weighed at masses between LLA and SBR.

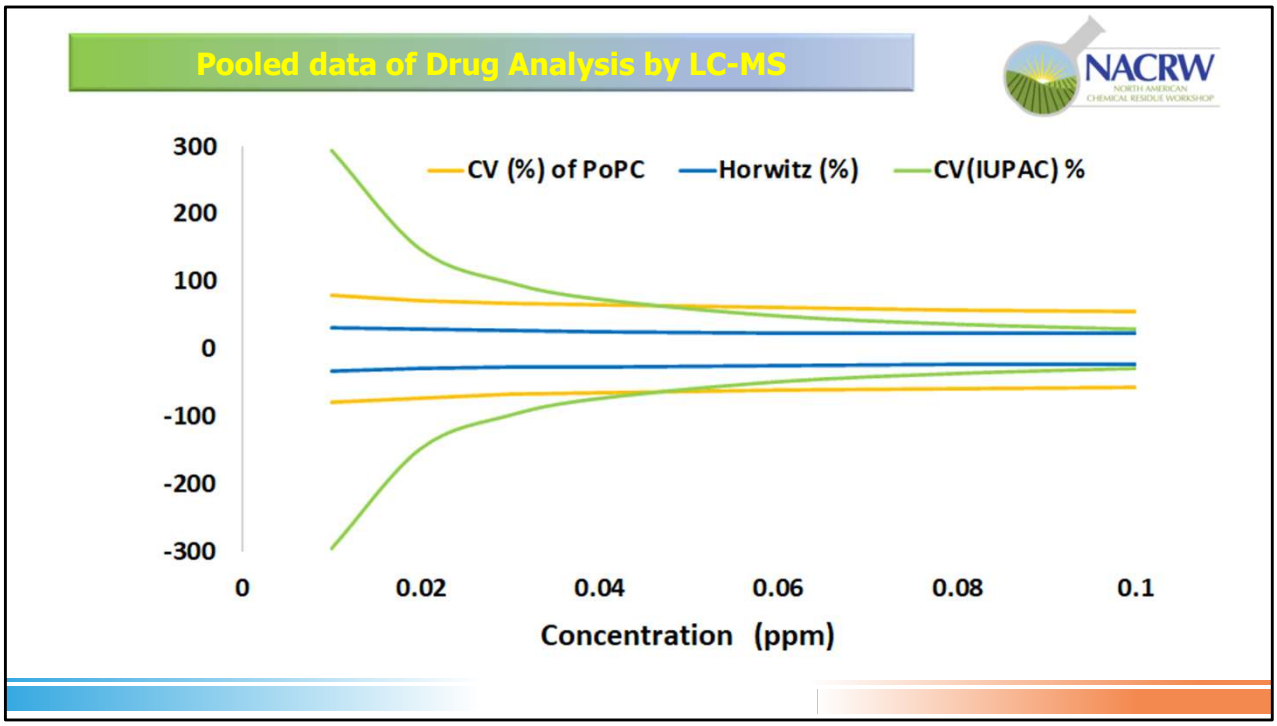


Since the CV-value or the RU is now identified as an important parameter of quality, it should be applied to all the laboratory tools of analytical chemistry where volumetric pipettes and micropipettes are very important. The classical exercise of measuring the masses of the volumes and the corresponding temperatures can be used to check the performance of these tools. After many hundreds of repetitions, it was found that it is safe to use volumetric pipettes with volumes larger than 41 microliters, which a volume far below volumes of commonly used pipettes. However, it should be noted that the SBR starts at 13 mL, which means that the CV-value is above 1.3 % and the RU is above 2.6 between the LLA and the SBR. The best possible RU that can be obtained by using volumetric pipettes is some 2.6 % that is a considerable contribution to the uncertainty budget. A similar investigation carried out with micropipettes showed that between 1.9 microliters and 280 microliters, the RU must be larger than 6 %. Compare this value to uncertainties reported in scientific publications where applications of micropipettes, complicated procedures of sample pre-treatments and subsequent measurements resulted in reporting the results with uncertainties of less than 5 %. Of course, results are frequently reported with a single SD, but that indicates anyway that the detector added virtually zero to the overall uncertainty, which cannot be correct.



In the recent edition of the SOP to United States Department of Agriculture Agricultural Marketing Service, Science & Technology Pesticide Data Program, the Horwitz CV-value was referred to as a useful rule of thumb with respect to estimating the uncertainty of the uncertainty budget. This is a very god rule of thumb, but then it has also to be recognized that the uncertainty budget in many cases underestimates the uncertainty of measurement. The Horwitz formula does not provide data for all types of uncertainties, as hown in the previous slides, but it is a great tool for estimating the uncertainty of measurement, even when the method validation was not performed. In the present diagram is shown the Horwitz trumpet depiceted from low concentrations towards high concentration where the opposite order was used in the original publication of Horwitz. It is interesting that the LLA of the Horwitz equation is approx. half a microgram per kg. This contradicts most uncertainties of contemporary determinations at the ultra-trace level of concentrations, but of course, technologies have improved much since 1982 after the publication of Horwitz. At the time, the curve was interpreted as a lack of training or lack of skills of laboratory staff who were allegedly careless when handling solutions of low concentrations. Obviously, this explanation makes no sense because there is not reason to believe that continued dilutions of solutions would result in such large increments of the CV-values. However, this is still the prevailing explanation to unexpected high levels of uncertainties that was also found by the investigations of the IMEP. The Horwitz equation is strictly empirical and it is based on neither the law-of-propagation of uncertainty nor the

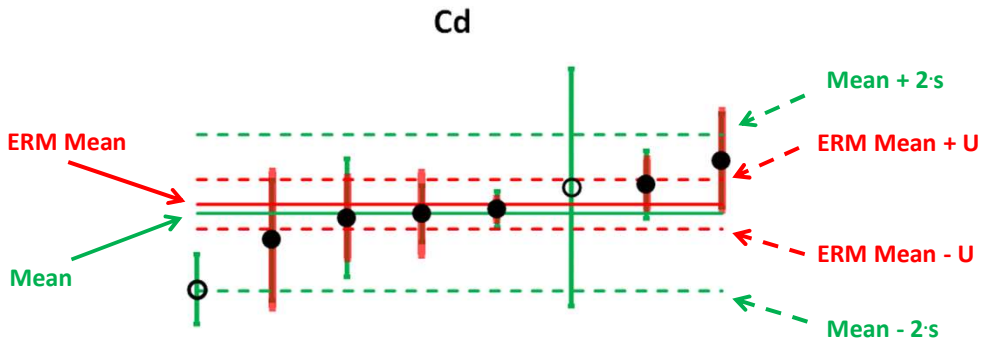
variance-covariance matrix. It correctly predicts infinite CV-values at zero concentration but at high concentrations, it predicts zero CV-values, which is not observed in practice. If the explanation of lack of skills and training is not the explanation to the shape of the Horwitz trumpet, what is the alternative explanation? This is where the PoPC comes in handy.



These graphs represent results of analysis of synthetic cannabinoids with LC-MS where the CV-value were compared for the three methods, PoPC, Horwitz and IUPAC-ISO. It may be expected that the Horwitz curve would give rise to the trumpet-like shape of the curve, but it is in fact more apparent for the IUPAC-ISO formula. In the present interval of concentrations, the curve of the PoPC and Horwitz are comparable, but at high concentrations, the Horwitz curve decreases towards zero, whereas the curve of the PoPC approaches a constant value that is better aligned with what is observed in practice. Again, the IUPAC-ISO formula overestimates the uncertainty at low concentrations and it tends to underestimate the uncertainty at high concentrations. Further results from the present analysis of synthetic cannabinoids showed that the Horwitz formula was not able to predict that some of the compounds, 2 out of 10, could not be analysed at these levels of concentrations because the uncertainties exceeded 100 % according to the PoPC.

IMEP-40: Cd of ERM CA-403 Seawater

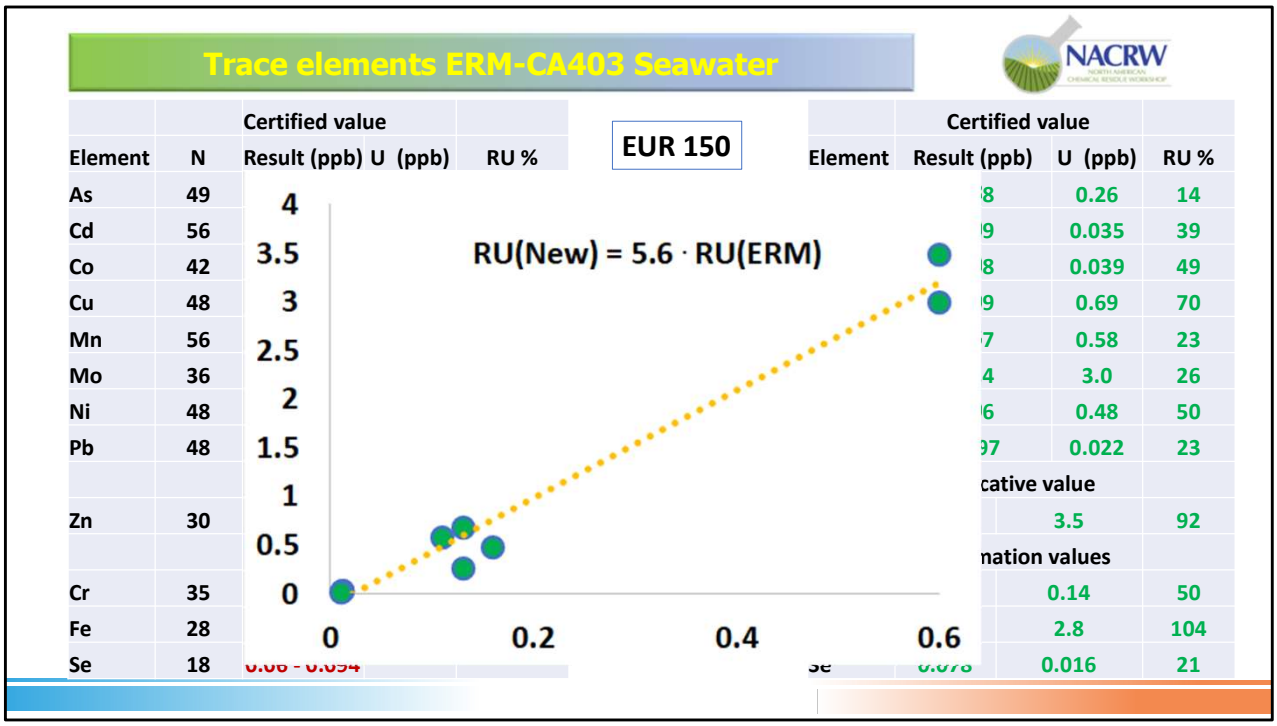
"The tables in this annex also contain the data sets that were discarded for technical reasons."



Data from certification report: https://crm.jrc.ec.europa.eu/p/40455/40463/By-material-matrix/Water/ERM-CA403-SEAWATER-element-content/ERM-CA403

Certified reference materials and standard reference materials have been introduced to analytical chemistry as the 'fundamental constants of analytical chemistry' that are used to check traceability of analytical measurements. Ever since the inception of CRMs/SRMs, it has been difficult to understand why one laboratory was able to produce results of much higher quality than any other professional laboratory in the world. SRMs of NIST are notoriously known as certified reference materials of high quality but it may be difficult to reproduce the results because they are reported with extremely low uncertainties. It is impossible to extract information about the raw data to the SRMs but all that information is readily available to the CRMs where the full report of the investigation can be downloaded from the website of ERM from where the CRMs can be purchased. However, the ERM also eliminate outliers form the data sets with application of more or less accepted methods of outlier detection. The analysis of inorganic species is considered as one of the most precise and accurate methods of analysis throughout the spectrum of analytical methodologies, so it may be slightly difficult to understand why it is necessary to reject results of professional laboratories that are accredited to perform this type of analyses on a routine basis. It is also difficult to grasp why the ERM is incapable of understanding that it becomes impossible for the laboratories to reproduce the results of CRMs when the corresponding uncertainties are grossly underestimated. As it has been pointed out before, it is not necessary to purchase CRMs when there are EABs in the laboratory. EABs are accurate, and it is easy to prepare solutions of reliable concentrations

when the solutions are prepared w/w. Therefore, there is no reason other than increasing business to apply CRMs and SRMs to analytical chemistry and to accuse any laboratory staff to have lack of training and lack of skills and eliminating their results from the pool of data is simply unacceptable and indecent. In the present example is show a result of the IMEP-40 program with analysis of Cd in seawater. The Red line indicates the average value of the ERM and the green line represents that average value that was determined upon the basis of all contributing laboratories without eliminating any outliers. It seems that we cannot agree about the value of the standard deviations, even when we are using the same data as those printed in the report to ERM CA-403. The red broken lines indicate the uncertainties that were determined by ERM after removal of outliers, which is in line with the practices and procedures of the BIPM, Eurachem and ISO. The black dots show the data that were accepted for calculation of certified values according to ERM. The open black circles represent the data that were rejected by the ERM. However, if all data of that investigation were accepted and no outliers were rejected, then some completely different results would be obtained, as indicated in green. Although the average value did not change very much, in accordance with the always small standard deviation of the mean, for this element was the overall SD more than tripled. It is natural to expect that the data point to the far left was rejected by the outliers test, as it deviates very much from the limits that are defined by the ERM uncertainty (red broken lines). However, it is slightly surprising that the data point third from the right hand side was rejected, as its average value corresponded very well with the certified value (solid red line). The standard deviation of this rejected data point was determined as too large, as compared with the standard deviations of the other data point, which may have been the reason for rejecting it. It is clear though, that the correspondence between the standard deviation of this point (third from right-hand side), is in excellent agreement with the uncertainty that was obtained by rejecting no data points from the data set. Accordingly, the best determination of Cd in the seawater (third from right-hand side) was rejected by ERM, which is unacceptable, regardless which method was applied to rejection of outliers. The consequence of manipulating the data in this unfavourable manner is, that it becomes difficult or impossible for laboratories to reproduce the results of this investigation, unless they reject exactly the same outliers that were supplied by the professional laboratories to ERM. Of course that would be impossible, which then disqualifies the treatment of data for this CRM.



The results of the analysis of other elements of the ERM seawater CRM is shown in this table where the majority of them were designated as 'certified', Zn was relegated to 'indicative value' and Cr, Fe and Se were assigned to the group of 'information' values that are not to be trusted as accurate. It is known that iron occurs at extremely low concentrations of 0.5 ng/L or less in ocean water/seawater where the pH value is too high for that element to stay in solution. After eliminating outliers, an interval was given within which it is likely that buyers of the CRM will determine the concentration of iron. An interval starting at 2.7 microg/L and ending at 4.2 microg/L is highly unexpected and contrary to expectation from results of the literature (e.g. Paul Worsfold). With respect to the other elements, the relative uncertainties were determined within the interval of 4.5 % and 15 %, which are extraordinary good values for analysis of contents of seawater. However, when the same data are used to calculation of uncertainties without rejecting any outliers, the picture changes completely, and RUs are now found within the range 14 % to 104 % by also considering the information values. The corrected results show that the concentration of Se could be moved to the certified category, whereas concentrations of Zn and Fe should be denoted as 'undetermined', which would be aligned with the expected chemical behaviour of these compounds in saline water and pH values close to 7. Some of the certified elements Cd, Cu, and Ni should be moved to the category of 'indicative values', which leaves only 4 elements to be certified by this analysis, in stark contrast to the postulated 8 elements by the ERM. The RU of the data is 5.6 times higher than the RU of

data that were subjected to elimination of outliers, so the conclusions and decisions ended up being completely different. ERM has ensured that it is not possible for anybody to agree about anything.

Conclusion

- **In order to improve results of analytical chemistry, researchers and developers need to focus on the performance of detectors, not the performance of people**
- **The method validation is performed to make sure that you are not working in the noise range**
- **The values and uncertainties of fundamental constants may have to change into *consensus values***

$$G = (6.67 \pm 0.67) \cdot 10^{11}\ m^3 \cdot kg^{-1} \cdot s^{-2}$$
$$c = (300 \pm 30) \cdot 10^5\ km/s$$
$$h = (6.63 \pm 0.66) \cdot 10^{-34}\ J \cdot s$$
$$N_A = (6.02 \pm 0.60) \cdot 10^{23}\ mol^{-1}$$
$$e = (1.60 \pm 0.16) \cdot 10^{-19}\ C$$

?

- **Scientific methodology and correspondence between predicted and observed uncertainty can be fulfilled by means of the PRINCIPLE OF POOLED CALIBRATIONS and consensus values**

It is concluded that consensus science is the way forward because it increases understanding of scientific results and it allows the correct decisions to be made. However, consensus science and the principle of pooled calibrations come at a price, which is highly elevated uncertainties. The staff is doing good work and they are not to be blamed for the adverse effects associated with the uncertainty of the detectors. The detector is the culprit, not humans, when it comes to the observation of high uncertainties. Horwitz was right with respect to the dependence of uncertainty as a function of concentration but the explanation to the dependence should be found in the performance of detectors. The principle of pooled calibrations is universal and it applies to all technologies, also those of physics. Therefore, it is proposed that the value of the fundamental constants may have to change as to reduce the number of significant digits and decimal points. Perhaps an RU of 10 % is exaggerated but this value is proposed to highlight the extremely high number of significant digits that are associated with fundamental constants that were determined by only a few so-called expert laboratories. Fundamental constants should be determined as consensus values where values are accepted from professional laboratories and universities around the world. You may have to reconsider your method validations?

Contact details

For more information contact:

Prof. Jens E.T. Andersen
Faculty of Sciences
Department of Chemical and Forensic Sciences
Tel: (+267) 4931537
Cell: (+267) 77174215
E-mail: andersenj@biust.ac.bw

Please submit e-mail to this address, if you wish to receive a copy of the notes.

Thank you very much for your attention!

Etosha National Park, Namibia, 2018

Thank you for your attention. Much appreciated.