Some practical examples of method validation in the analytical laboratory

Piet van Zoonen*, Ronald Hoogerbrugge, Steven M. Gort, Henk J. van de Wiel, Henk A. van ’t Klooster

National Institute of Public Health and the Environment (RIVM), P.O. Box 1, 3720 BA Bilthoven, The Netherlands

Method validation is a key element in both the elaboration of reference methods and the assessment of a laboratory’s competence in producing reliable analytical data. Hence, the scope of the term method validation is wide, especially if one bears in mind that there is or at least should be a close relation between validation, calibration and quality control QA/QC. Moreover, validation should include more than the instrumental step only since the whole cycle from sampling to the final analytical result is important in the assessment of the validity of an analytical result. In this article validation is put in the context of the process of producing chemical information. Two cases are presented in more detail: the development of a European standard for chlorophenols and its validation by a full scale collaborative trial, and the intralaboratory validation of a method for ethylene-thiourea using alternative analytical techniques. ©1999 Elsevier Science B.V. All rights reserved.

Keywords: Method validation; Quality control; Collaborative trial; Calibration

1. Introduction

Quality is a relative notion: never ‘high’ nor ‘low’ in an absolute sense, but rather adequate or inadequate in terms of the extent to which a product, process or service meets the requirements specified beforehand by an objective or a customer.
The principal product of an analytical chemical laboratory is information about the chemical composition of material systems, usually in terms of the identity and/or quantity of one or more relevant components in samples taken from these materials.

The quality of scientific information in general is evaluated by internationally accepted standards of objectivity, integrity, reproducibility and traceability, in any case prior to publication.

Essential criteria for the quality of produced chemical information are the utility and reliability, which are closely related to the margins of uncertainty in the measurement results, regarding both the identity and the concentration of the target components.

With respect to these correlated criteria, minimum requirements are generally set by the customer and usually deduced from a previously specified purpose. The quality of produced chemical information is therefore factually to be acknowledged by the customer as the end user of this information. For chemical measurements, this could be a clinical chemist who needs to know the identity of certain isolated compounds from a biological fluid, or a polymer chemist who wishes to verify the molecular structure of a synthesis product, or a health researcher who wants to know whether the concentration of a certain toxic compound in a certain food is above a certain concentration level. It is not hard to imagine the consequences — in terms of costs, health risks, etc. — when, on closer examination or statistical evaluation of the measurement results, a ‘positive’ turns out to be false, or the uncertainty margin of a measured concentration appears to be 100% and not the 10% initially reported. Evaluation and validation of analytical methods and laboratory procedures are therefore of paramount importance, prominent means being the use of adequate (preferably certified) reference materials and participation in interlaboratory proficiency tests.

Quality demands made upon the infrastructure, equipment, operating procedures, personnel and organisation of the laboratory are to be deduced from the quality requirements that the produced chemical information should meet. A formal recognition of this type of quality can be achieved through accreditation or certification, based on international quality standards and guidelines, as issued by ISO, OECD and CEN.

Validation of analytical methods is but one though an essential step in the integral process of quality assurance and quality control of chemical measurements in material systems [1,2].

2. Chemical analysis as an integral process

Chemical analysis of whatever material system can be described as a chain of decisions, actions and procedures [3]. The cyclic nature of many chemical analysis processes is schematically depicted in Fig. 1.

The last step (interpretation and evaluation of the results of the analysis) should eventually provide an answer to the starting problem, generally stated by e.g. a scientific problem or a legal context. If the answer is not satisfactory, the analysis cycle can be followed again, after a change or adaptation of one or more steps. Sometimes this leads to the development of a new method or (part of a) procedure, for example in order to achieve a better separation of certain components, or to attain a lower detection limit for specific compounds.

Like any chain, a chemical analysis chain is only as strong as its weakest link. In general, the weakest links in an analytical process are not the ones usually recognised as parts of a chemical analysis, such as chromatographic separation or spectrometric detection, but rather the preceding steps, often taking place outside the analytical laboratory: the selection of the object(s) to be sampled, design of the sampling plan, subsampling, sampling techniques and transportation and storage of samples.

When the analytical laboratory is not responsible for the sampling, the quality management system
often does not even account for these weak links in the analytical process!

Further, if elements such as preparation of the laboratory sample, extraction or clean-up of the samples have not carefully been carried out, even the most advanced and quality controlled analytical instruments and sophisticated computer techniques cannot prevent that the analytical results are questionable.

Finally, when the interpretation and evaluation of the analytical results have no statistical basis, it is not clear how significant these are, which greatly determines their utility.

Therefore, in our opinion, quality control and quality assurance should involve all the steps of the chemical analysis as an integral process, of which validation of the analytical methods used is only one, though important, step.

In laboratory practice, quality criteria should be concerned with the rationale of the sampling plan, the validation of methods, instruments and laboratory procedures, the reliability of identifications, the accuracy and precision of measured concentrations and the comparability of laboratory results with relevant information produced earlier or elsewhere.

3. Quality, information and uncertainty

According to Shannon [4], gaining information is reducing the amount of uncertainty, as provided by the results of measurements.

The following uncertainty sources in quantitative analysis have been identified by EURACHEM [5]:

- Incomplete definition of the measurand
- Sampling
- Incomplete extraction/preconcentration
- Matrix effects and interferences
- Contamination during sampling or sample preparation
- Personal bias in reading analogue instruments
- Lack of awareness/imperfect measurement of effects of environmental conditions on the measurement procedure
- Uncertainty of weights and volumetric equipment
- Instrument resolution or discrimination threshold
- Values assigned to measurement standards and reference materials
- Values of constants and other parameters obtained from external sources, used in the data reduction algorithm
- Approximations and assumptions incorporated in the measurement procedure
- Random variation.

However, one must realise that for a particular problem one should try to assess the relative influence of the alternative sources on the final outcome, e.g. if the dominating source of uncertainty of the procedure has a RSD of 25% it is not very efficient to demand a < 0.1% accuracy in measuring or weighing the initial sample volume.

As stipulated by Horwitz and Albert [6], the among-laboratories variability is the dominating error component in the world of practical ultratrace analysis. They conclude that a single laboratory cannot determine its own error structure, except in the context of certified reference materials or consensus results from other laboratories.

Again, the importance of the use of reference materials is underlined, since these provide information on the combined effect of many of the potential sources of uncertainty.

In the literature as well as in laboratory practice, quantification of uncertainty in qualitative analysis (identification, molecular structure elucidation) is scarcely addressed. Yet, there are possibilities, for example when using computer-aided library search of molecular spectra for identification of organic compounds. If an adequate similarity index is being used, such as Cleij’s reproducibility-based system for different kinds of molecular spectra [7], it is possible to specify a quantitative threshold of uncertainty, like confidence intervals are used in quantitative analysis.

In a next version of the EURACHEM syllabus, quantification of uncertainty will also refer to qualitative analysis.

4. Validation characteristics/performance criteria

The key criteria for evaluation of an analytical method are the following [8–10]:

- selectivity/specificity
- limit of detection
- limit of quantification
- recovery
- working range and linearity
- accuracy/trueness
- precision (repeatability, reproducibility)
- ruggedness/robustness.
Whether or not these criteria apply to validation of a particular method depends on the purpose and the nature of the method. In its guidelines the Dutch Board for accreditation distinguishes between ‘new’, ‘standard’, or ‘modified’ methods in order to characterise the nature of a method; a standard method in this definition is a fully validated (inter)national standard, while on the other hand methods are classified according to their purpose, e.g. qualitative/quantitative or trace-analytical vs major components. The validation requirements for such methods are summarised in Fig. 2 [11].

5. Current approaches to validation

In a recent review [12] current approaches to validation of analytical chemical methods are discussed. Some shortcomings of existing validation schemes are identified, such as insufficient coverage of variability in space or time, and mismatches between validation criteria and intended use of the method, for example the use in regulatory control. An attempt is made to link validation concepts used in different fields, such as measurement uncertainty, and prediction error. A general statistical modelling approach for combining different aspects of validation is recommended, and illustrated with an example. This type of modelling should be the basis for the development of new statistically underpinned validation schemes which integrate current validation and quality assurance activities.

It is stated that validation includes the initial assessment of performance characteristics, several types of interlaboratory testing, and quality control. Validation is thus concerned with assuring that a measurement process produces valid measurements; this has also been called measurement assurance [13].

The concept validation of a method of analysis may be used in (at least) three senses: in the narrow and traditional sense the term denotes validation of a chemical method as described in a standard operating procedure (SOP). In a wider sense validation may be concerned with a method of analysis (e.g. in an (inter)national standard) which explicitly leaves freedom to adapt the procedure to the infrastructure in a specific situation. In this case there are more SOPs, all in conformity with the ‘master method’. Finally, in a still wider and perhaps unconventional sense, validation of analytical methods may be considered from the perspective of those who use analytical results for other purposes. The method of analysis for end users of analytical results amounts to the specification of an analytical result (e.g. ‘ clenbuterol in liver’), with the implied ‘analysed by any reasonable method’. Accordingly, a specific method of analysis in the analytical chemical sense can be considered as just one realisation of the class of all methods currently applied to measure component X in matrix Y.

In principle, each modification of the protocol invalidates an existing validation according to ISO 5725.

Much work on validation has been performed in joint efforts of IUPAC, ISO and AOAC International (see [14] for its early history). Results appear as a series of harmonised protocols. The second edition of standard ISO 5725 [15] has much in common with the IUPAC/ISO/AOAC Protocol for design, conduct and interpretation of collaborative studies [16]. Important contributions to some of the problems mentioned above were made in two other protocols [17,18], on proficiency testing (PT) and on internal quality control (IQC), respectively. The international harmonised protocol for the proficiency testing of (chemical) analytical laboratories [17] considers
laboratory-performance studies, in which each laboratory uses its own analytical method as opposed to the method-performance studies of ISO 5725 (nomenclature according to [19]). Although the purpose of proficiency testing is often primarily the evaluation or improvement of laboratory performance, it is reasonable to consider it also as a method-performance validation with our wider definition of method (e.g. from the perspective of a customer interested in ‘clenbuterol in liver’ measurements). This would solve problems of different laboratories having different SOPs, and of SOPs changing every now and then in each laboratory. There is a prescribed repetition in proficiency testing schemes with frequencies of once every 2 weeks to once every 4 months being considered reasonable. This solves the problem with the static nature of ISO 5725 validation, and, by varying the test materials, also the problem of not assessing matrix variability.

Despite all the advantages, proficiency testing according to the IUPAC guidelines [17] cannot be considered a complete validation methodology on its own. First of all, it does not provide for SOP-specific validation. More importantly, the scheme requires repeated interlaboratory studies, which severely restricts the amount and variety of samples that can be analysed. Therefore, proficiency testing is an extensive validation methodology. Finally, in the current protocol the only performance characteristic considered is laboratory bias, most often in the form of a z-score. This information alone may be insufficient to judge fitness for purpose.

It has been shown that an effective measurement assurance requires validation at different scales. Newly developed or implemented methods are usually first validated using in-house validation. This type of validation should be supplemented by ongoing internal quality control validation in each laboratory, and by participation of the laboratory in interlaboratory schemes. Considering the complex nature of many modern methods of analysis, proficiency testing schemes, allowing laboratory-specific SOPs, are more to the point than method-evaluating schemes like ISO 5725. Currently, there is insufficient linkage between the three validation schemes in-house validation, internal quality control and proficiency testing. The model presented in this article, together with the concepts of measurement uncertainty and fitness-for-purpose, provides a basis for the development of integrated validation approaches.

During a recent Joint FAO/IAEA Expert Consultation on validation of analytical methods for food control the ‘ideal validated method’ was defined as follows [9]: “The ideal validated method is one that has progressed fully through a collaborative study in accordance with internationally harmonised protocols for the design, conduct and interpretation of method performance studies. This usually requires a study design involving a minimum of 5 test materials, the participation of 8 laboratories reporting valid data, and most often includes blind replicates or split levels to assess within-laboratory repeatability parameters.”

Limiting factors for completing ideal multi-laboratory validation studies include high costs, lack of sufficient expert laboratories available and willing to participate in such studies, and overall time constraints.

6. Validation by using alternative analytical methods

The validation strategies described above all assume that methods are applied on a routine basis in various laboratories. In a research environment a rather unique method might be developed and validated for the use in only one or a few studies. Then of course one would like to establish the same amount of information on the validity of the method. However, some of the usual tools, like participation in proficiency schemes and the use of reference materials are probably not possible.

A number of practical examples, i.e. the screening and analysis of polar pesticides in environmental monitoring programmes by coupled-column liquid chromatography and gas chromatography–mass spectrometry, are discussed by Hogendoorn et al. [20]. One of the examples is a study on the levels of ethylene-thiourea (ETU) in groundwater. Below the main aspects of the validation are described, further details can be found in [20].

In the validation of both methods the calibration procedure is very important and provides the information for several of the criteria. The calibration is based on spiked samples which are comparable to the samples of groundwater to be analysed which gives in itself, by comparison with the analysis of standards, the recovery data. The calibration samples are ordered in time around and between the real samples. Therefore, the possible influence of the conditions during the analysis (robustness) of the samples is automatically captured in the calibration sequence.

To determine the working range of both methods, the calibration data are evaluated by the spreadsheet application Calwer [21]. The result is shown in Fig. 3. The application enables an extensive evaluation of the
calibration curve for example with respect to the appropriate calibration model. The results are presented to the user both by number, as results of statistical tests, and by graphs, to enable a graphical interpretation. These results show for the ETU/HPLC-UV calibration that the most simple calibration model \( y = bx \) is to be preferred since the more complex models \( y = a + bx \) or \( y = a + bx + cx^2 \) show neither significant nor relevant improvements of the residual standard deviation. Calwer also offers the opportunity to test the concentration dependence of the residuals. In most calibration software ordinary least squares regression is used to calculate the calibration curve. This approach is based on the homoscedasticity of the measurements. Then the deviations at low concentrations should show deviations from a reasonable calibration model which are equal to the deviations at the high limit of the working range. For ETU this assumption is clearly violated; therefore a suitable model for the residuals is used in combination with weighted least squares regression. In addition to the graphical presentation, also the log likelihood is calculated, which enables an objective selection criterion between several variance models. The weighted regression has been applied in our laboratory for about 5 years showing that the assumption of equal variance over the working range is nearly always severely violated. Application of a reasonable variance model implicitly gives the standard deviation at low concentrations and therefore estimates the limit of detection.

The strategy of variance models, weighted regression in combination with the maximum likelihood criterion is formalised for method comparison calibration/validation in ISO 13752 [22].

The selectivity of the method is checked by comparing the shape and the retention of real samples and calibration samples with the chromatogram of standard solutions.

For the analysis of ETU in ground water neither reference materials nor proficiency testing schemes
are available. A reasonable accuracy is assured by a number of precautions:

- standard solutions are made by two independent routes
- solid reference material is checked by IR spectroscopy for impurities
- a comparison study with one other laboratory (familiar with this analysis)
- all 60 groundwater samples are analysed by two independent methods, HPLC-UV and GC-MS.

The results are shown in Fig. 4.

The analyses of all samples by the two analyses methods offers the opportunity to compare the variance between real samples with the expected variance based on the validation results of both methods with the same type of regression. For intermediate and relatively high concentrations of ETU in groundwater (>1 µg/l) the variance between both methods complies with its expectation. However, for very low concentrations (<1 µg/l) the variance between the results of both methods is much larger than expected. This deviation might be an indication that the natural spread in interfering materials in the groundwater is not completely covered by the samples used for the calibration/validation experiments. These interferences apparently only have a substantial effect on the measurement uncertainty at low levels (<1 µg/l) of the analyte.

This additional source of uncertainty might disturb the assumption that the method with the best validation characteristics should be preferred as the reference method (the x in the regression equation \( y = a + bx \)). Without precautions the additional source of variance is treated by usual regression schemes as due to the \( y \)-method yielding estimates for the slope of the calibration relation (\( b \)) which are too low. In method comparison this implies that the ratio of the \( y \)-method results with respect to the \( x \)-method results is underestimated. An alternative approach to compare analytical methods explores the 'known' variances of both methods [23,24]. However these results will also be disturbed by the additional matrix variance because it will be shared according to the ration of these 'known' variances. Recently Carrol et al. [25] described the statistical implications of additional sources of variance in regression analysis, suggesting a rather simple correction (method of moments) to obtain a reasonable estimate of the ratio between the two methods.
6.1. Validation sheet

In our laboratory, analytical procedures are always validated. Key elements of the method, references to relevant documentation (study plan, SOPs, files of raw data, reference materials, (interlaboratory) confirmations) and an abstract of the validation results (see above) are summarised in a validation sheet, that again gives access to the original raw data.

7. Interlaboratory method validation

Another approach to validation is by interlaboratory tests. As an example, the validation of a GC determination of chlorophenols in water is summarised below; details can be found in [26,27].

In order to validate the (preliminary) CEN method PrEN 12673 ‘Water quality – Gas chromatographic determination of some selected chlorophenols in water’ [28] an interlaboratory study was organised. The intercomparison study comprised a total of 9 samples; a high-level, a low-level and a blank sample of three types of water: drinking water, surface water and waste water. The levels of the spikes were based on the water quality objectives. The variation in the samples due to homogeneity and stability of the components was tested extensively and appeared to be negligible in comparison with variation between the results of the participants.

Results were received from 24 laboratories from 8 European countries. In the total data set about 7% of the data were detected as statistical outliers and were subsequently rejected. The number of eliminated outliers is comparable with “experience in the United States with environmental analytes indicating that about 9% of the data represents out-of-control performance” [29].

For the remaining data set the relative standard deviations for repeatability varied between 5% and 25% and reproducibility between 26% and 56%. Both these ranges complied with the general variation in interlaboratory studies as found by Horwitz [30]. The comparison of the reproducibility with the Horwitz relation is shown in Fig. 5.

The recovery of the spikes found by the participants generally varied between 60 and 140% (Fig. 6).

The data set was also evaluated for differences in results due to the degrees of freedom in the standard method. The largest difference was found between participants which perform internal versus external calibration. The latter method obtained results that were on average up to 50% lower than the results of the participants applying internal standards. Supported by this result the use of an internal standard will be mandatory in the EN 12673.

Another degree of freedom in the standard method that should be evaluated was the use of a mass spectrometer (MS) instead of an electron capture detector.
The results of the interlaboratory study were compiled into one large data set to enable a multivariate evaluation. Since only alternative selections of the chlorophenols were added to each of the samples, this data set contained a lot of missing values ( > 50%). The usual calculation techniques for PCA can not deal with such a data set, therefore a least squares procedure was implemented to estimate the PCA-like results iteratively. The scores were calculated as the projection of the part of the data that was actually measured. The result showed a prominent first principal component that describes about 40% of total variance with all loadings on the same side. This indicated that a major multicomponent systematic source of variation was present at a number of laboratories. This implies that the elimination of that source of variance, which might be relatively simple, will have a major influence on the comparability of EN 12673 measurements.

Both the study of the systematic differences (like the differences between GC-ECD and GC-MS results) and the result of the PCA that the application of standard multivariate statistical tools easily obtains essential information from interlaboratory data sets which might otherwise be overlooked.

8. Other useful elements for an in-house validation procedure

8.1. Reference materials

If relevant reference materials are available they are a powerful tool in the assessment and control of the accuracy of the performance of the applied analytical method. In practical application in the laboratory one should, however, realise that additional control experiments often are necessary since reference materials usually do not reflect the complete range of application of the method with respect to concentration, matrix effects and possible inhomogeneity of samples.

8.2. Proficiency testing schemes

The data from proficiency testing schemes are helpful in assessing the performance of a method in the laboratory. In most instances it has the same practical limitations as the use of reference materials and, additionally, the usefulness of the results strongly depends on the quality of the data produced by the other laboratories in the ring test.

8.3. Control charts

Validation studies often demonstrate the performance of an analytical method before its routine application. The validity to the routine measurements of the assessed performance can be controlled by the repeated analyses of control samples. The results are monitored in a control chart with warning and action limits. Application of a stable control sample also provides necessary information for the interpretation of long-term trend studies.

8.4. Double-blind replicates

With respect to proficiency control, reference materials and control samples have limited value due to their high price (only for reference materials) and since the operators might recognise them. The possible measurement errors in known concentrations are of course not fully representative to the possible measurement errors in unknown concentrations. To avoid this limitation the laboratory can ask the sampling organisation to provide a number of unknown (with independent sample codes) replicates. After finalising the measurements the codes of the replicates are revealed and these data are quite representative for the precision in the other samples.

9. Conclusion

Important trends in method validation are the development of alternatives for the conventional collaborative trial, which has its limitations due to the large amount of work to be done to establish a standard method. The proliferation of laboratory accreditation has prompted the need for practical in-house validation procedures, which in turn may prove their value in method evaluation. Also the more intensive use of QA/QC schemes in recent years is a valuable source of performance data.

References


C.E. Shannon, A mathematical theory of communication, J. Bell Syst. Tech. 27 (1948).

W. Horwitz, personal communication.

W. Horwitz, Evaluation of analytical methods used for monitoring trace organic compounds (biphenyls, dioxins, furans), J. AOAC Int. 79 (1996) 589–621.


A. Madansky, The fitting of straight lines when both methods are subject to error, J. Am. Stat. Assoc. 54 (1959) 173.


W. Horwitz, personal communication.