VALIDATION OF ANALYTICAL METHODS AND LABORATORY PROCEDURES FOR CHEMICAL MEASUREMENTS

PIET VAN ZOONEN, HENK A. VAN 'T KLOOSTER, RONALD HOOGERBRUGGE, STEVEN M. GORT, AND HENK J. VAN DE WIEL

Analytical Chemical Laboratories Division, National Institute of Public Health and the Environment (RIVM), Bilthoven, The Netherlands

Method validation is a key element in the establishment of reference methods and in 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 the role of Quality Assurance/Quality Control (QA/QC). The paper puts validation in the context of the process generating chemical information, introduces basic performance parameters included in the validation processes, and evaluates current approaches to the problem.

Two cases are presented in more detail: the development of European standard for chlorophenols and its validation by a full scale collaborative trial and the intralaboratory validation of a method for ethylenethiourea by using alternative analytical techniques.

Key words:
method validation, quality assurance, proficiency testing, reference materials, standardisation

VALIDATION: ELEMENT OF LABORATORY QUALITY ASSURANCE

Quality is a relative notion; never "high" or "low" in an absolute sense. Rather, it is adequate or inadequate in terms of the extent to which a product, a process, or a 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

356

van Zoonen P. et al.: *METHODS AND PROCEDURES FOR CHEMICAL MEASUREMENTS*
Arh hig rada toksikol, Vol 49 (1998) No 4, pp. 355–370

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 the 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, a polymer chemist who wishes to verify the molecular structure of a product of synthesis, or a health researcher who wants to know whether the concentration of a certain toxic compound in certain food is above certain concentration level. It is not hard to imagine the consequences – in terms of costs, health risks, and so on – when, on closer examination or statistical evaluation of the measurement results, a "positive" finding turns out to be false, or the uncertainty margin of a measured concentration appears to be 100% and not the initially reported 10%. 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 on the infrastructure, equipment, operating procedures, personnel, and organization 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 International Organization for Standardization (ISO), Organisation for Economic Co-operation and Development (OECD), and European Committee for Standardization (CEN).

Validation of analytical methods is one, though essential step in the integral process of quality assurance and quality control of chemical measurements in material systems (1, 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). Figure 1 shows the cyclic nature of many chemical analytical processes.

The last step (interpretation and evaluation of results of analysis) should eventually provide an answer to the starting problem, generally stated by a client of the laboratory. 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 a development of a new method or (part of a) procedure in order, for example, to achieve better separation of certain components, or to attain a lower detection limit for specific compounds.

van Zoonen P. et al.: *METHODS AND PROCEDURES FOR CHEMICAL MEASUREMENTS*
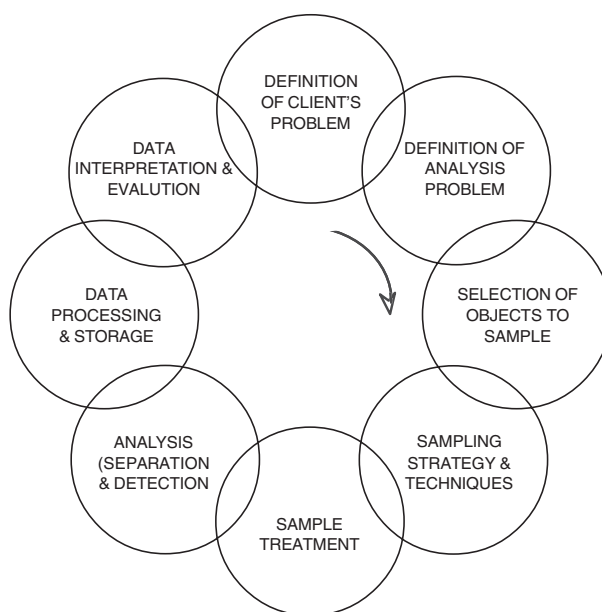Arh hig rada toksikol, Vol 49 (1998) No 4, pp. 355–370

357

Figure 1 *Chemical analysis as a cyclic process*

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

When the analytical laboratory is not responsible for the sampling, the quality management system often does not even take account of these weak links in the analytical process. Furthermore, if the preparation (extraction, clean-up, etc.) of the samples has not carefully been carried out, even the most advanced and quality controlled analytical instruments and sophisticated computer techniques cannot prevent that the results of the analysis become questionable. Finally, unless the interpretation and evaluation of results have a solid statistical base, it is not clear how significant these results are, which in turn greatly undermines their merit. We, therefore, believe that quality control and quality assurance should involve all the steps of chemical analysis as an integral process, of which the validation of the analytical methods is only one, though important, step.

In laboratory practice, quality criteria should concern 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.

358

van Zoonen P. et al.: *METHODS AND PROCEDURES FOR CHEMICAL MEASUREMENTS*
Arh hig rada toksikol, Vol 49 (1998) No 4, pp. 355–370

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 in 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.

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 and elucidation of molecular structure) is scarcely addressed. Yet, there are possibilities to resolve the problem, for example, through the use of 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 (6), it is possible to specify a quantitative threshold of uncertainty, comparable to that of confidence intervals used in quantitative analysis.

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

Validation characteristics/performance criteria

The key criteria for evaluation of an analytical method are: selectivity/specificity; accuracy/trueness; precision (repeatability, reproducibility); limit of detection; limit of quantification; sensitivity; working range and linearity; ruggedness/robustness; and recovery (7–9).

Selectivity/specificity

Specificity is a quantitative indication of the extent to which a method can distinguish between the analyte of interest and interfering substances on the basis of signals produced under actual experimental conditions. Random interferences should be determined using representative blank samples.

van Zoonen P. et al.: *METHODS AND PROCEDURES FOR CHEMICAL MEASUREMENTS*
Arh hig rada toksikol, Vol 49 (1998) No 4, pp. 355–370

359

Accuracy/trueness

Accuracy refers to closeness of agreement between the true value of the analyte concentration and the mean result obtained by applying experimental procedure to a large number of homogeneous samples. It is related to systematic error and analyte recovery. Systematic errors can be established by the use of appropriate certified reference materials (matrix-matched) or by applying alternative analytical techniques.

Precision (repeatability, reproducibility)

Both repeatability and the reproducibility are expressed in terms of standard deviation and are generally dependent on analyte concentration. It is therefore recommended that the repeatability and within-laboratory reproducibility are determined at different concentrations across the working range, by carrying out 10 repeated determinations at each concentration level. As stipulated by *Horwitz and Albert* (10), the variability among laboratories 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.

Limit of detection

The limit of detection is usually expressed as the analyte concentration corresponding to the sample blank plus three sample standard deviations, based on 10 independent analyses of sample blanks.

Limit of quantification

The limit of quantification is the lowest concentration of analyte that can be determined with an acceptable level of uncertainty or, alternatively, it is set by various conventions to be five, six, or ten standard deviations of the blank mean. It is also sometimes known as the limit of determination.

Sensitivity

Sensitivity is the measure of the change in instrument response which corresponds to a change in analyte concentration. Where the response has been established as linear with respect to concentration, sensitivity corresponds to the gradient of the response curve.

Working and linear ranges

For any quantitative method there is a range of analyte concentrations over which the method may be applied. At the lower end of the concentration range the limiting factor is the value of the limit of detection and/or limit of quantification. At the upper end of the concentration range limitations will be imposed by various effects depending on the detection mechanism.

Within this working range there may exist a linear range, within which the detection response will have a sufficiently linear relation to analyte concentration. The working and linear range may differ in different sample types according to the effect of

360

van Zoonen P. et al.: *METHODS AND PROCEDURES FOR CHEMICAL MEASUREMENTS*
Arh hig rada toksikol, Vol 49 (1998) No 4, pp. 355–370

interferences arising from the sample matrix. It is recommended that, in the first instance, the response relationship should be examined over the working range by carrying out a single assessment of the response levels to at least six concentration levels. To determine the response relationship within the linear range, it is recommended that three replicates are carried out at each of at least six concentration levels.

Ruggedness/robustness

This is a measure of how effectively the performance of the analytical method stands up to less than perfect implementation. In any method there will be certain parts which will severely affect the method performance, unless they are carried out with sufficient care. These aspects should be identified and, if possible, their influence on the method performance should be evaluated using the ruggedness tests, sometimes also called robustness tests. The ruggedness/robustness tests provide important information for the evaluation of the measurement uncertainty. The methodology for evaluating uncertainty given in the ISO Guide relies on identifying all parameters that may affect the result (that is, the potential sources of uncertainty) and on quantifying the uncertainty contribution from each source. This is very similar to procedures used in robustness tests which identify all the parameters likely to influence the result and determine the acceptability of their influence through control. If carried out with this in mind, the robustness tests can provide information on the contribution to the overall uncertainty from each of the parameters studied.

CURRENT APPROACHES TO VALIDATION

In a recent review, *van der Voet and co-workers* (11) discuss current approaches to validation of analytical chemical methods, identifying some shortcomings of existing validation schemes such as insufficient coverage of variability in space or time and mismatches between validation criteria and intended use of the method, giving an example of regulatory control. The authors make an attempt to link validation concepts used in different fields, such as measurement uncertainty, and the prediction error. They recommend general statistical modelling approach for combining different aspects of validation and illustrate it 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 inter-laboratory testing, and quality control. Validation is thus concerned with assuring that a measurement process produces valid measurements; this has also been called measurement assurance (12).

The validation of an analytical method as a concept may be understood 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 ISO 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".

van Zoonen P. et al.: *METHODS AND PROCEDURES FOR CHEMICAL MEASUREMENTS*
Arh hig rada toksikol, Vol 49 (1998) No 4, pp. 355–370

361

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 statement: "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 International Union of Pure and Applied Chemistry (IUPAC), ISO, and Association of Official Analytical Chemists (AOAC International) (13). Results appear as a series of harmonised protocols. The second edition of the ISO 5725 standard (14) has much in common with the IUPAC/ISO/AOAC Protocol for design, conduct and interpretation of collaborative studies (15). Important contributions to some of the problems mentioned above were made in two other protocols on proficiency testing (PT)(16) and on internal quality control (IQC)(17). The harmonised international protocol for the proficiency testing of (chemical) analytical laboratories (16) 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 ref. 18). Although the primary purpose of proficiency testing is often the evaluation or improvement of laboratory performance, it is also reasonable to consider it as a method-performance validation in the wider sense of the definition (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. The prescribed repetition in proficiency testing schemes considers reasonable the frequencies of once every two weeks to once every four months. This solves the problem with the static nature of ISO 5725 validation, and, by varying the test materials, the problem of not assessing matrix variability. Despite all the advantages, proficiency testing according to the IUPAC guidelines (16) cannot be considered as 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, the current protocol has limited consideration of performance to laboratory bias, most often in the form of a z-score. This information alone may be insufficient to evaluate a method's fitness for the purpose.

It has been shown that an effective measurement assurance requires validation at different scales. Newly developed or implemented methods are usually first validated through 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 the method-evaluating schemes like ISO 5725. Currently, the three validation schemes, in-house validation, internal quality control, and proficiency testing, are not sufficiently linked. The model presented in this paper and the concepts of measurement of uncertainty and fitness-for-purpose provide a basis for the development of integrated validation approaches.

362

van Zoonen P. et al.: *METHODS AND PROCEDURES FOR CHEMICAL MEASUREMENTS*
Arh hig rada toksikol, Vol 49 (1998) No 4, pp. 355–370

A recent Joint FAO/IAEA Expert Consultation on validation of analytical methods for food control defined the "ideal validated method" as follows (8):

"*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 cost, lack of expert laboratories available and willing to participate in such studies, and overall time constraints.

Validation by using different analytical methods

Above all, the described validation strategies 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.

Hogendoorn and co-workers (19) discuss a number of practical examples such as screening and analysis of polar pesticides in environmental monitoring programmes by coupled-column liquid chromatography and gas chromatography-mass spectrometry (GC-MS). One example is a study on the levels of ethylenethiourea (ETU) in groundwater. In the validation of both methods the calibration procedure is very important and provides information for several of the criteria. The calibration is based on spiked samples which are comparable to samples of groundwater to be analysed, which, by comparison with the analysis of standards, in itself gives the recovery data. The calibration samples are ordered in time around and between the real samples. Possible influence of the conditions during the analysis (ruggedness) 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 Calwer© spreadsheet application (20). The result is shown in Figure 2. The application enables an extensive evaluation of the calibration curve with respect to the appropriate calibration model. The results of statistical tests are presented to the user for interpretation both through figures and through graphics. In the analysis of ETU by high performance liquid chromatography using an UV detector (HPLC-UV) the results show that the most simple calibration model ($y=bx$) is to be preferred since the more complex models ($y=a+bx$ or $y=a+bx+cx2$) show neither significant nor relevant improvements of the residual standard deviation. The Calwer© application also offers the opportunity to test the concentration dependence of the residues. Most calibration software uses ordinary least squares regression to calculate the calibration curve. This approach is based on the homoscedasticy of the measurements. The deviations at low concentrations should show deviations from a reasonable calibration model which are equal to deviations at the high limit of the working range. As for ETU, this assumption is clearly violated; a suitable model for the residues is, hence, used in combination with weighted least squares regression. In addition to the graphical presentation, the calculated log likelihood enables an objective selection

User data

Name	Ellen Dijkman
Project	715801/95/01
Analyte	ETU
Internal standard	
Summary file	
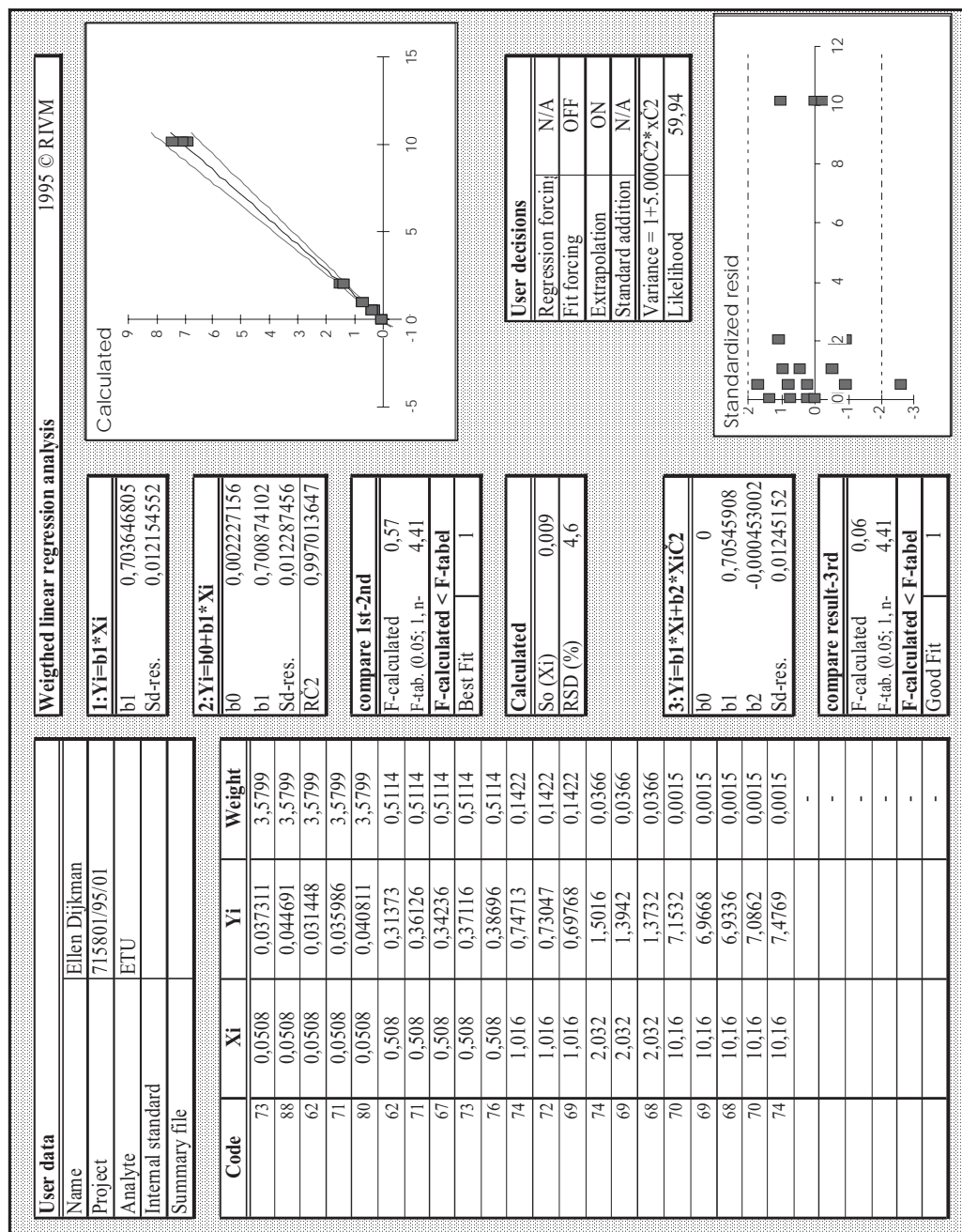
Code	Xi	Yi	Weight
73	0,0508	0,037311	3,5799
88	0,0508	0,044691	3,5799
62	0,0508	0,031448	3,5799
71	0,0508	0,035986	3,5799
80	0,0508	0,040811	3,5799
62	0,508	0,31373	0,5114
71	0,508	0,36126	0,5114
67	0,508	0,34236	0,5114
73	0,508	0,37116	0,5114
76	0,508	0,38696	0,5114
74	1,016	0,74713	0,1422
72	1,016	0,73047	0,1422
69	1,016	0,69768	0,1422
74	2,032	1,5016	0,0366
69	2,032	1,3942	0,0366
68	2,032	1,3732	0,0366
70	10,16	7,1532	0,0015
69	10,16	6,9668	0,0015
68	10,16	6,9336	0,0015
70	10,16	7,0862	0,0015
74	10,16	7,4769	0,0015
			-
			-
			-
			-
			-

Weigthed linear regression analysis — 1995 © RIVM

Calculated

1:Yi=b1*Xi

b1	0,703646805
Sd-res.	0,012154552

2:Yi=b0+b1* Xi

b0	0,002227156
b1	0,700874102
Sd-res.	0,012287456
RC2	0,997013647

compare 1st-2nd

F-calculated	0,57
F-tab. (0,05; 1, n-	4,41
F-calculated < F-tabel	
Best Fit	1

Calculated

So (Xi)	0,009
RSD (%)	4,6

3:Yi=b1*Xi+b2*XiC2

b0	0
b1	0,70545908
b2	-0,000453002
Sd-res.	0,01245152

compare result-3rd

F-calculated	0,06
F-tab. (0,05; 1, n-	4,41
F-calculated < F-tabel	
Good Fit	1

User decisions

Regression forcing	N/A
Fit forcing	OFF
Extrapolation	ON
Standard addition	N/A
Variance = $1+5.000 \dot{C}2*xC2$	
Likelihood	59,94

Standardized resid

Figure 2   *Calibration curve for ethylenethiourea (ETU) analysed by HPLC-UV. Several calibration models are statistically tested and a model for the variance is applied.*

criterion between several variance models. The weighted regression has been applied for about four years in the RIVM Division of Analytical Chemistry, 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 (21).

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
☐ standard solutions are checked by IR spectroscopy for impurities
☐ a parallel study in another laboratory (familiar with this analysis) for comparison of results
☐ all 60 groundwater samples are analysed by two independent methods, HPLC-UV and GC-MS.

The analyses of all samples using the two methods gives the opportunity to compare the variance between real samples and the expected variance based on the validation results of both methods (Figure 3). For mediate and relatively high concentrations of ETU in groundwater (>1 μg/L) the variance between both methods corresponds to the expected. However, for very low concentrations (<1 μg/L) the variance between the results of both methods is much larger than expected. This deviation might indicate that the natural spread of interfering materials in the groundwater is
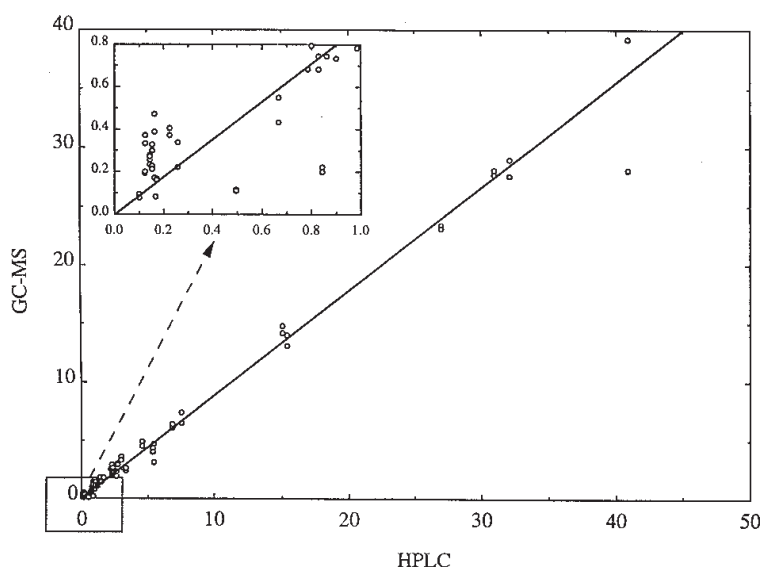

Figure 3 *Comparison of ethylenethiourea (ETU) levels found using two analytical methods. At the very low end of the concentration range deviations in results between the two methods seriously exceed the expectation based on the two validations.*

van Zoonen P. et al.: *METHODS AND PROCEDURES FOR CHEMICAL MEASUREMENTS*
Arh hig rada toksikol, Vol 49 (1998) No 4, pp. 355–370

365

not completely covered by the samples used for the calibration/validation experiments. These interferences, apparently, substantially affect uncertainty measurements only at low levels (<1 μg/L) of the analyte.

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, and interlaboratory confirmations), and an abstract of the validation results are summarised in a validation sheet.

Interlaboratory method validation

Another approach to validation is by interlaboratory tests. An example of the validation of GC determination of chlorophenols in water is summarised below. Details can be found in reports of *Hoogerbrugge and co-workers* (22, 23).

An interlaboratory study was organised to validate the (preliminary) CEN method Provisional European Norm (PrEN) 12673 "Water quality – Gas Chromatographic determination of some selected chlorophenols in water" (24). The intercomparison study comprised a total of nine 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 when compared to variation between the results of the participants.

Figure 4 gives an example of the results of pentachlorophenol measurement in a waste water sample with the high addition of spikes. Results were received from 24 laboratories from eight European countries. In this data set about 7% of the data were detected as statistical outliers and were subsequently rejected. The number of eliminated outliers is comparable with the "experience in the United States with environmental analytes indicating that about 9% of the data represents out-of-control performance" (*W. Horwitz*, personal communication). For the remaining data set, the relative standard deviations for repeatability varied between 5% and 25% and the reproducibility between 26% and 56% (Figure 5). Both these ranges complied with the general variation in interlaboratory studies as found by *Horwitz* (25).

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

The data set was also evaluated for differences in results due to degrees of freedom in the standard method. The largest difference was found between participants who perform internal and those who perform 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 needed evaluation was the use of a mass spectrometer (MS) instead of an electron capture detector (ECD). The MS results varied, among the components, between minus 25% and plus 20% compared to the ECD results.

The results of the interlaboratory study were compiled into a large data set for multivariate evaluation. Since only alternative selections of chlorophenols were added to each of the samples, this data set contained a lot of missing values ($>50\%$). The usual calculation techniques for principal component analysis (PCA) can not deal with
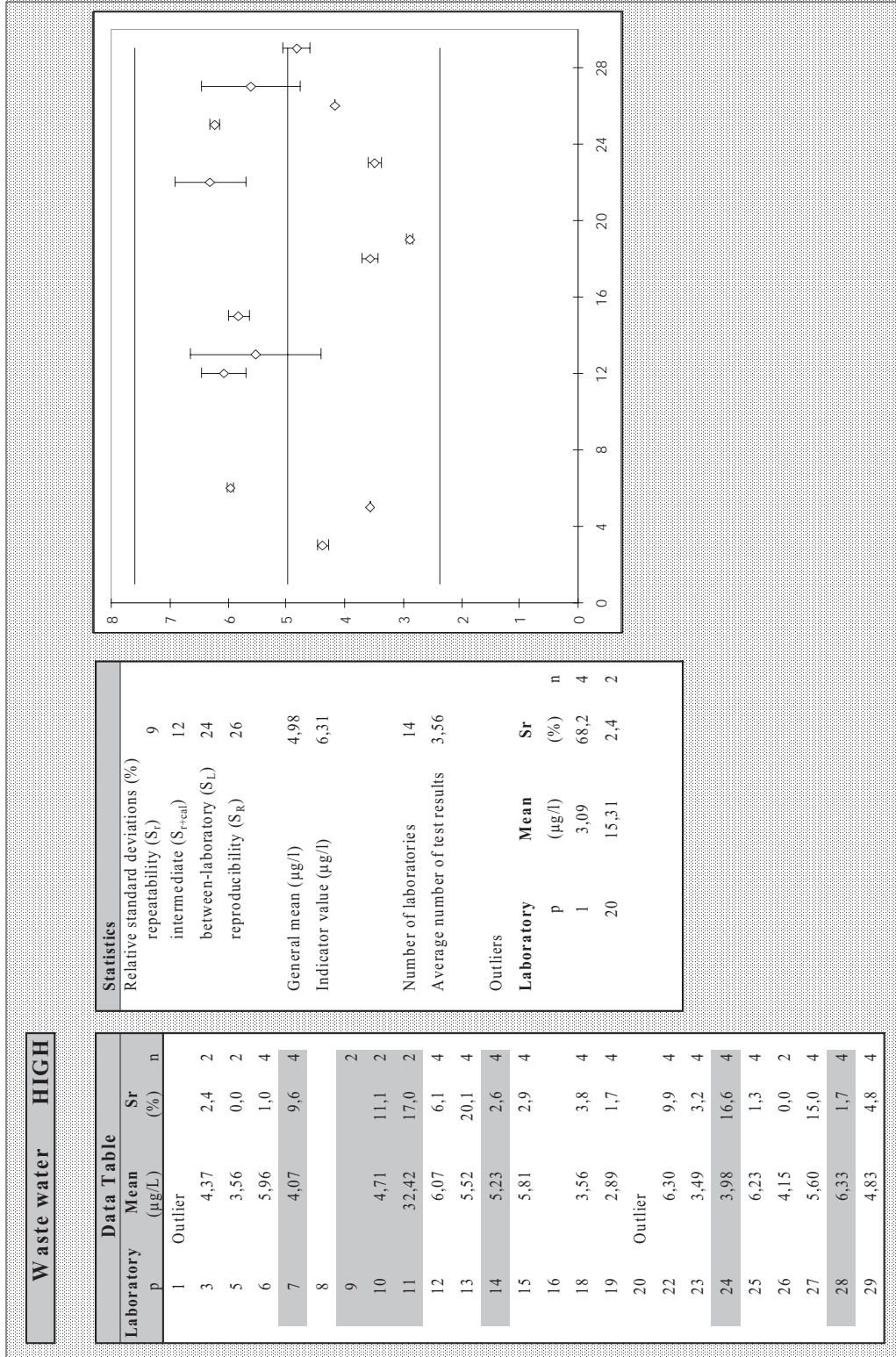
Waste water **HIGH**

Data Table

Laboratory p	Mean (μg/L)	Sr (%)	n
1	Outlier		
3	4,37	2,4	2
5	3,56	0,0	2
6	5,96	1,0	4
7	4,07	9,6	4
8			2
9	4,71	11,1	2
10	32,42	17,0	2
11	6,07	6,1	4
12	5,52	20,1	4
13	5,23	2,6	4
14	5,81	2,9	4
15			
16	3,56	3,8	4
18	2,89	1,7	4
19	Outlier		
20	6,30	9,9	4
22	3,49	3,2	4
23	3,98	16,6	4
24	6,23	1,3	4
25	4,15	0,0	2
26	5,60	15,0	4
27	6,33	1,7	4
28	4,83	4,8	4
29			

Statistics

Relative standard deviations (%)	
repeatability (S_r)	9
intermediate (S_{r+cal})	12
between-laboratory (S_L)	24
reproducibility (S_R)	26
General mean (μg/l)	4,98
Indicator value (μg/l)	6,31
Number of laboratories	14
Average number of test results	3,56

Outliers

Laboratory p	Mean (μg/l)	Sr (%)	n
1	3,09	68,2	4
20	15,31	2,4	2

Figure 4 *The participants' measurement results for pentachlorophenol in the waste water sample with the high addition. The tabulated results are shown in combination with the detected outliers. The statistical results are based on data set after the removal of the outliers.*
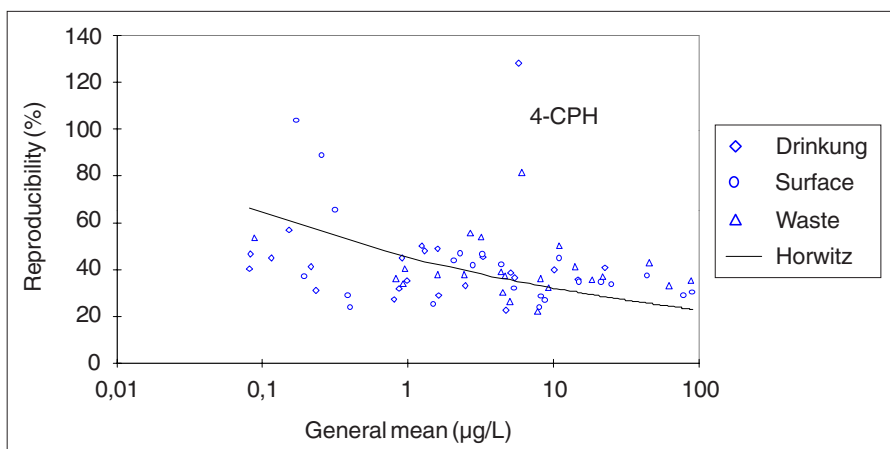
van Zoonen P. et al.: *METHODS AND PROCEDURES FOR CHEMICAL MEASUREMENTS*
Arh hig rada toksikol, Vol 49 (1998) No 4, pp. 355–370

367

Figure 5 *Relative standard deviations of reproducibility as a function of the concentration for all sample/component combinations. The Horwitz curve is plotted for comparison.*
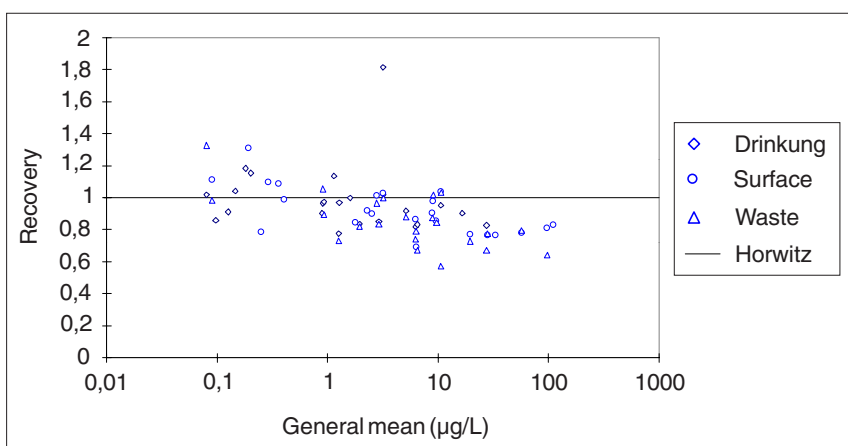

Figure 6 *Recovery of the additions, calculated as the ratio between the indicator value and the general mean as function of the concentration of the indicator value.*

such 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 principle 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 in a number of laboratories, implying 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 suggest that the application of standard multivariate statistical tools easily obtains essential information from inter-laboratory data sets which might otherwise be overlooked.

OTHER USEFUL ELEMENTS IN IN-HOUSE VALIDATION PROCEDURE

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 laboratory application one should, however, realise that additional control experiments are often 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.

Proficiency testing schemes

The data from proficiency testing schemes are helpful in assessing the performance of a method in a 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 data produced by the other laboratories in the ring test.

Control charts

Validation studies often demonstrate the performance of an analytical method before its routine application. The validity of the assessed performance for the routine measurements can be controlled by 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.

Double blind replicates

With respect to control, the proficiency reference materials and control samples have limited value, as they are expensive (reference materials) and liable to bias through recognition by operators. The possible measurement errors in known concentrations are, of course, not fully representative of the possible measurement errors in un-known concentrations. To avoid this limitation, the laboratory can ask a sampling organisation to provide a number of unknown replicates (with independent sample codes). The codes of the replicates are revealed when the measurements are com-pleted, which renders the data quite representative of the precision in other samples.

CONCLUSION

The development of alternatives for the conventional collaborative trial sets important trends in method validation. The limitations are associated with 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,

van Zoonen P. et al.: *METHODS AND PROCEDURES FOR CHEMICAL MEASUREMENTS*
Arh hig rada toksikol, Vol 49 (1998) No 4, pp. 355–370

369

which in turn may prove its merits in the method evaluation. Furthermore, the more intensive use of QA/QC schemes in recent years has been a valuable source of performance data.

REFERENCES

1. *Christensen JM, Kristiansen J, Hansen AM, Nielsen JL*. Method validation: An essential tool in total quality management. In: Parkany M, ed. Quality Assurance and Total Quality Management for Analytical Laboratories. Cambridge: The Royal Society for Chemistry, 1995:46–54.
2. *Juniper IR*. Method validation: An essential element in quality assurance. In: Parkany M, ed. Quality Assurance and Total Quality Management for Analytical Laboratories. Cambridge: The Royal Society for Chemistry, 1995:114–20.
3. *van 't Klooster HA*. Chemical analysis as a cyclic process (A chain is only as strong as its weakest link). In: Angeletti G, Bjørseth A, eds. Organic Micropollutants in the Aquatic Environment. Dordrecht: Kluwer Academic Publishers, 1991:163–71.
4. *Shannon CE*. A mathematical theory of communication. J Bell Syst Tec 1948;27.
5. *Williams A, Wegscheider W, eds*. Quantifying uncertainty in analytical measurement. EURACHEM report. Teddington: LGC, 1995.
6. *van 't Klooster HA, Cleij P, Luinge HJ, Kleywegt GJ*. Computer-aided spectroscopic structure analysis of organic molecules using library search and artificial intelligence. In: Warr WA, ed. Chemical Structures: the International Language of Chemistry. Berlin: Springer Verlag, 1988: 219–33.
7. *Holcombe H, ed*. The fitness for purpose of analytical methods. In: A Laboratory Guide to Method Validation and Related Topics. EURACHEM Guide. Teddington: LGC, 1998.
8. *Food and Agriculture Organization (FAO) and International Atomic Energy Agency (IAEA)*. Food and Nutrition Paper 68: Validation of Analytical Methods for Food Control. Rome: FAO/UN, 1998.
9. *Hill ARC, Reynolds SL*. Guidelines for validation of analytical methods for monitoring trace organic components in food stuffs and similar materials (unpublished).
10. *Horwitz W, Albert R*. Reliability of the determinations of polychlorinated contaminants (Biphenyls, Dioxins, Furans). J AOAC Int 1996;79:589–621.
11. *van der Voet H, van Rhijn JA, van de Wiel HJ*. Interlaboratory and time aspects of effective validation (unpublished).
12. *Carey MB*. Measurement assurance: role of statistics and support from international statistical standards. Int Stat Rev 1993;61:27–40.
13. *Horwitz W*. History of the IUPAC/ISO/AOAC harmonization program. J AOAC Int 1992;75:368–71.
14. *International Organization for Standardization (ISO)*. ISO 5725: Accuracy (trueness and precision) of measurement methods and results. Parts 1–4, 6. Geneva: ISO, 1994.
15. *International Union of Pure and Applied Chemistry (IUPAC)*. Protocol for the design, conduct and interpretation of collaborative studies. Pure Appl Chem 1988;60:855–64.
16. *International Union of Pure and Applied Chemistry (IUPAC)*. The international harmonized protocol for the proficiency testing of (chemical) analytical laboratories. Pure Appl Chem 1993;65:2123–44.
17. *International Union of Pure and Applied Chemistry (IUPAC)*. Harmonized guidelines for internal quality control in analytical chemistry laboratories. Pure Appl Chem 1995;67:649–66.
18. *International Union of Pure and Applied Chemistry (IUPAC)*. Nomenclature of interlaboratory analytical studies. Pure Appl Chem 1994;66:1903–11.

19. *Hogendoorn EA, Hoogerbrugge R, Baumann RA, Meiring HD, de Jong APJM, van Zoonen P.* Screening and analysis of polar pesticides in environmental monitoring programmes by coupled-column liquid chromatography and gas chromatography-mass spectrometry. J Chromatogr A 1996;754:49–60.

20. *Gort SM, Hoogerbrugge R.* A user-friendly spreadsheet program for calibration using weighted regression. Chemometr Intel Lab Syst 1995;28:193–9.

21. *International Organization for Standardization (ISO).* ISO 13752: Air Quality – Assesment of uncertainty of a measurement method under field conditions using a second method as reference. Geneve: ISO, 1998.

22. *Hoogerbrugge R, Ramlal MR, Stil GH, et al.* Interlaboratory validation of PrEN 12673: Water Quality – Gas Chromatographic determination of some selected chlorophenols in water. RIVM-report no. 219101006. Bilthoven: RIVM, 1997.

23. *Hoogerbrugge R, Gort SM, van der Velde EG, van Zoonen P.* Multi- and univariate interpretation of the interlaboratory validation of PrEN 12673; GC determination of chlorophenols in water. Anal Chim Acta (accepted).

24. *European Committee for Standardization (CEN).* PrEN 12673: GC determination of chlorophenols in water. Brussels: CEN, 1998.

25. *Horwitz W.* Evaluation of analytical methods used for regulation of foods and drugs. Anal Chem 1982;54:67A–76A.

Sažetak

VREDNOVANJE ANALITIČKIH METODA I LABORATORIJSKIH POSTUPAKA U KEMIJSKIM MJERENJIMA

Vrednovanje metoda ključni je postupak u utvrđivanju referentnih metoda i procjeni kompetencije laboratorija da proizvodi pouzdane rezultate analiza. Stoga je značenje širine okvira ovog termina posebice važno jer valja imati na umu i ulogu osiguranja kakvoće i kontrole kakvoće.

Autori stavljaju postupak vrednovanja metoda u kontekst proizvodnje kemijskoanalitičkih podataka, upoznaju čitatelja s osnovnim parametrima pri ocjeni uspješnosti te ocjenjuju trenutne pristupe ovom problemu.

U dva je primjera posebna pozornost posvećena razvoju europske norme za klorofenole i njezinoj ocjeni opsežnim pokusom te unutarlaboratorijskom potvrdom metode za određivanje etilentiouree rabeći alternativne analitičke tehnike.

Ključne riječi:
osiguranje kakvoće, referentni materijal, standardizacija, testiranje vrsnosti

Requests for reprints:

Piet van Zoonen, Ph.D.
Analytical Chemical Laboratories Division,
National Institute of Public Health and the Environment (RIVM)
P.O. Box 1, 3720 BA Bilthoven, The Netherlands
E-mail: *piet.van.zoonen@rivm.nl*