



Guide in Validation of Alternative Proprietary Chemical Methods

Contents

DEFINITIONS.....	2
PART 1: VALIDATION AND EVALUATION OF QUALITATIVE PROPRIETARY METHODS	5
A. METHOD COMPARISON STUDY	5
B. INTERMEDIATE STUDY	9
C. INTERPRETATION.....	9
PART 2: VALIDATION AND EVALUATION OF QUANTITATIVE PROPRIETARY METHODS	10
A. METHOD COMPARISON STUDY	10
B. INTERMEDIATE STUDY	14
REFERENCES:	15

SCOPE

This NordVal Protocol describes validation of proprietary chemical methods (test kits). The objective is to guide expert laboratories, NordVal technical committees and the NordVal steering group in the validation, evaluation and certification of test kits. This protocol consists of the following two parts, describing

- √ validation and evaluation of qualitative proprietary methods
- √ validation and evaluation of quantitative proprietary methods

Preferably, the alternative method should be validated against a reference method. However, when no such method is available the validation can be carried out using certified reference materials, control materials and/or spiked samples at different levels in various matrixes.

DEFINITIONS

Accuracy, relative

The relative accuracy is the degree of correspondence between the response obtained by the proprietary method and the reference method on artificially contaminated samples, or “the expected/true” results of the spiked samples.

Analyte

The analyte is the component demonstrated or measured by the method of analysis.

Intermediate study

A study of the proprietary method’s performance by at least one additional independent laboratory.

Cut-off Level

The Cut-Off Level is the response or signal from a screening test which indicates that a sample contains a substance at or above the Screening Target Concentration. If the Cut-Off Level is exceeded a subsequent confirmatory test should be carried out. During the initial validation process (i.e. the comparison test if not already determined in the manufacturer’s R&D programme), the Cut-Off Level may be established through analysis of matrix blank samples and replicates of those same samples spiked (fortified) at the Screening Target Concentration.

Detection capability $CC\beta$

Detection capability ($CC\beta$) is the smallest content of the substance that may be detected, identified and/or quantified in a sample with an error probability of β . The β error is the probability that the tested sample is truly non-compliant even though a compliant measurement has been obtained. For screening tests the β error (i.e. false compliant rate) should be $< 5\%$.

False negatives and false positives

The false negative rate is the probability that the test is negative for samples that contain the analyte. The false positive rate is the probability that the test is positive for samples that do not contain the analyte at the screening level.

Inclusivity

Inclusivity is the ability to detect the relevant members of a target analyte group or the target analyte from a wide range of sources, e.g. detecting the soy or peanut from Asia as well as those coming from USA.

Limit of Detection (LOD)

The lowest amount or concentration of the analyte in a sample, which can be reliably detected (but not necessarily quantified).

Limit of quantification (LOQ)

The lowest amount of an analyte which can be determined quantitatively with a closely defined confidence.

Method comparison study

Study performed by the expert laboratory of the proprietary method against the reference method/spiked samples.

Negative deviation

The proprietary method presents a negative deviation if it gives a negative result when the reference method gives a positive result.

Positive deviation

The proprietary method presents a positive deviation if it gives a positive result when the reference method gives a negative result.

Precision

The degree of agreement between independent analysis results obtained under specific circumstances.

Proprietary method

A proprietary method means that a party, or proprietor, exercises private ownership of the method.

Qualitative method

A qualitative method is a method of analysis whose response is either the presence or absence of the analyte in a certain amount of sample.

Quantitative method

A quantitative method is a method of analysis whose response is the amount of the analyte measured either directly or indirectly in a certain amount of sample.

Reference method

A reference method is a method which is internationally recognised and accepted (e.g. NMKL, ISO, CEN and AOAC International methods, methods given in EU/national legislations and certain national standards of equivalent standing).

Repeatability

The repeatability is the closeness of agreement between successive and independent results obtained by the same method on identical test material under the same conditions (apparatus, operator, laboratory and short intervals of time).

Repeatability limit (r)

The repeatability limit is the value less than or equal to which the absolute difference between two tests results obtained under repeatability conditions is expected to be with a probability of 95%.

Note: If the difference between two results exceeds r, the results should be considered as suspect.

Replicates

In this document, the term replicates is used about duplicates, i.e. samples containing the same matrix with the same concentration level, handled as separate samples.

Reproducibility as internal reproducibility

The internal reproducibility is the closeness of agreement between single test results on identical test material using the same method obtained at the same laboratory at different days.

Internal reproducibility limit (R)

The reproducibility limit is the value less than or equal to which the absolute difference between two test results obtained under internal reproducibility conditions is expected to be with a probability of 95%.

Ruggedness

The sensitivity of an analytical method to minor deviations in the experimental conditions of the method. A method which is not influenced by such minor deviations, is said to be rugged when it comes to these experimental conditions.

Semi-Quantitative method

A semi-quantitative method gives an approximate indication of the concentration of the analyte. Whilst the numerical result may not be regarded as reportable, this may be useful to the analyst in deciding the calibration range for the subsequent (quantitative) confirmation test. (e.g. ELISA, which include a calibration curve).

Sensitivity

For specified test conditions, the sensitivity represents the proportion of test samples that contain the analyte and respond positively to the test.

The relative sensitivity is the ability of the proprietary method to detect the analyte compared to the reference method.

Specificity

Specificity is here defined as the ability of an analytical method to distinguish the analyte to be determined from other substances present in the sample.

For specified test conditions, the specificity is the proportion of test samples that do not contain the analyte and respond negatively to the test.

Spiked sample

A spiked sample is a sample to which known concentrations of specific analytes have been added in such a manner as to minimize the change in the matrix of the original sample. Every spiked sample analysed should have an associated reference to the spike solution and the volume added.

Trueness: The degree of agreement between a sample's true content of a specific analyte and the result of the analysis.

Validation of a proprietary method

The validation of a proprietary method is the procedure to demonstrate if the proprietary method provides adequate results and is in compliance with its claims.

PART 1: VALIDATION AND EVALUATION OF QUALITATIVE PROPRIETARY METHODS

A. METHOD COMPARISON STUDY

RUGGEDNESS

The nature of the analytical method in question will determine which parameters, which need to be tested. The most frequently tested ruggedness parameters, which may be critical to an analytical method, are:

- √ the composition of the samples
- √ pH
- √ timing of individual (assay) steps
- √ temperature
- √ presence of potentially interfering substances (e.g., tannins or other complexing agents, varying levels of lipids, endogenous enzymes)

Blank samples can be used for ruggedness testing as they will reveal effects caused by the matrix or the chemical batch. Information from the ruggedness test can be used to specify the conditions under which a method should be used.

BATCH-TO-BATCH VARIATION

The expert laboratory review the manufacturer's descriptions and documentation on how they have ensured that there is no significant batch-to-batch variation of their products.

METHOD PERFORMANCE CHARACTERISTICS

The method comparison study is carried out by the expert laboratory only. The proprietary method is tested against the reference method. When no reference method is available, the proprietary method can be compared against "expected results" of certified reference materials (CRM), control materials and/or various spiked samples. The following performance characteristics should be estimated for qualitative methods:

- √ field of application, concentration range
- √ limit of detection
- √ sensitivity
- √ accuracy
- √ inclusivity
- √ specificity
- √ false positives
- √ false negatives
- √ the agreement between methods

The method comparison should be carried out on real incurred samples if possible to obtain in the levels of interest. Otherwise, it should be carried out on artificially contaminated samples. If the validation is requested for all food matrices, at least 5 relevant food matrices are selected. For other categories of matrices, select a relevant number of matrices. The concentration levels tested should be low, medium and high in addition to a blind sample, i.e. a sample matrix free of the analyte of interest. The number of replicates should be about 10.

Inclusivity

Where relevant, check the inclusivity, by analysing samples from different regions/countries in order to learn if the method is able to detect the analyte not only in different foodstuffs but also in foodstuffs from different

regions/ countries. This is relevant for instance when analysing soy or peanut from different countries. The performance of a group specific test (e.g., betalactam or sulphonamide antibiotics) should be checked with all the relevant members of the compound group.

The concentration range and the limit of detection (LOD)

Start by examining the applicable concentration range of the method. This is done by analysing series of samples comprising a blank sample and samples which contain different concentrations of the analyte. It is recommended to carry out at least 10 parallels on each concentration level. Draw up a response curve by plotting the portion of positive results on the y-axis and the concentration on the x-axis. It is then possible to read from the curve the threshold concentration at which the method starts to become unreliable.

In the example below, the reliability of the detection method becomes less than 100% at concentrations below 100 µg/g.

Table 1: Results at different levels – for the determination of the concentration range and LOD

Concentration (µg/g)	N	Positive	Negative	Positive/negative (%)
25	10	0	10	0
50	10	1	9	11
75	10	5	5	50
100	10	10	0	∞
200	10	10	0	∞

Detection capability (CCβ)

The detection capability (CCβ) is to be tested against the method's claimed screening/threshold limit. Analyse at least 20 blank materials and 20 blanks per matrix fortified with the analyte(s) at the screening limit. Analyse the samples and identify the analyte(s). No more than one sample obtained can be false negative (β = 5%, 95% confidence).

The agreement between methods or spiked samples - sensitivity, accuracy, specificity, false positives, false negatives, the agreement between methods

For estimating all the method performance characteristics, except for the limit of detection, a cross table as illustrated below is a helpful tool.

Table 2: Table for comparing two qualitative methods or obtained versus expected results:

		Proprietary method (method 1) Obtained result		Total
		Positive/ Detected	Negative/ Not Detected	
Reference method or expected results (method 2)	Positive/ Detected	N_{11}	N_{12}	$N_{1_}$
	Negative/ Not Detected	N_{21}	N_{22}	$N_{2_}$
Total		$N_{_1}$	$N_{_2}$	$N = N_{1_} + N_{2_}$ or $N = N_{_1} + N_{_2}$

where:

N_{11} = the number of samples which were positive with both methods

N_{12} = the number of samples which were negative with method 1 but positive with method 2

N_{21} = the number of samples which were positive with method 1 but negative with method 2

N_{22} = the number of samples which were negative with both methods

In many cases the reference method would be a quantitative method rather than a qualitative one, and would most probable be more sensitive than the proprietary method. Thus, for the reference method, negative samples might be defined as the levels obtained at and below the limit of detection (LOD) of the proprietary method. Further, positive samples could be defined as the samples with concentrations above the LOD of the proprietary method.

Sensitivity (SE)

The sensitivity, i.e. the number of obtained positive results, that are expected to be positive, divided by the total number of expected positive results, is calculated according to the following:

$$SE = \frac{N_{11}}{N_{1_}}, \text{ Relative sensitivity, SE (\%)} = \frac{N_{11} \cdot 100}{N_{1_}}$$

The sensitivity is partly dependent of the concentration level. At the detection level the sensitivity should be about 50%. In validation of alternative methods, the methods are generally considered acceptable if the overall sensitivity is 95% or higher.

Relative Accuracy (RA)

The relative accuracy is the degree of correspondence between the response obtained by two different methods:

$$RA = \frac{N_{11} + N_{22}}{N}, \text{ Relative accuracy, RA (\%)} = \frac{(N_{11} + N_{22}) \cdot 100}{N}$$

False negative rate (FN)

Number of obtained negative results, that are expected to be positive, divided by the total number of expected positive results:

$$FN = \frac{N_{12}}{N_{1-}} = 1 - SE$$

As $FN = 1 - SE$, FN does not provide additional information if SE is given.

Specificity rate (SP)

Number of obtained negative results, that are expected to be negative, divided by the total number of expected negative results:

$$SP = \frac{N_{22}}{N_{2-}}, \text{ Relative specificity SP (\%)} = \frac{N_{22} \cdot 100}{N_{2-}}$$

Specificity is tricky in particular for immunoassays as spiked samples would not contain precursors or metabolites that might cross-react and affect the response. Thus, one should aim for analysing incurred real samples if feasible.

False positive rate (FP)

Number of obtained positive results, that are expected to be negative, divided by the total number of expected negative results:

$$FP = N_{21} / N_{2-} = 1 - SP$$

As $FP = 1 - SP$, FP does not provide additional information if SP is given.

Checking the degree of agreement by kappa

The degree of agreement might be quantified by kappa. If all the observed values lie on the diagonal of the cross table, there is perfect agreement between the two methods. On the other hand, more values placed outside the table diagonal, indicate less agreement between the methods. To get an estimate of the agreement between the methods, Cohen's kappa may be calculated as follows:

The observed proportion of agreement, the accuracy, is: $RA = p_o = \frac{N_{11} + N_{22}}{N}$

The expected frequency of agreement, the expected accuracy, or repeatability by chance is:

$$p_e = \frac{(N_{1-} \cdot N_{-1}) + (N_{2-} \cdot N_{-2})}{N^2}$$

To measure the agreement between the methods, Cohen's κ (kappa) can be applied:

$$\kappa = \frac{p_o - p_e}{1 - p_e}$$

In general, the following κ values are used in the interpretation of kappa:

- | |
|---|
| <ul style="list-style-type: none"> • $\kappa \leq 0.20 \rightarrow$ Poor agreement • $\kappa \in \{0.21 - 0.40\} \rightarrow$ Fair agreement • $\kappa \in \{0.41 - 0.60\} \rightarrow$ Moderate agreement • $\kappa \in \{0.61 - 0.80\} \rightarrow$ Good agreement • $\kappa > 0.80 \rightarrow$ Very good agreement |
|---|

For method validation, considering the overall agreement (number of test results are relatively high), "very good agreement" is often required, i.e. $\kappa > 0.80$

B. INTERMEDIATE STUDY

The aim of the intermediate study is to confirm the obtained results on at least one additional laboratory.

At least three relevant food materials, artificially contaminated at 3 levels (low, medium and high) and a negative control should be used. The lowest level should be about the detection/screening level. If the intermediate study is conducted at only one additional laboratory the number of replicates for each matrix has to be at least five. The levels of the samples should be unknown to the laboratory.

If the proprietary method comprises more than one method procedure a relevant food matrix for each procedure should be selected for the study. Carry out the calculation in the same way as for the comparison study (see table 2).

C. INTERPRETATION.

Compare the accuracy, sensitivity, specificity and kappa with the obtained results with results from the comparative study.

PART 2: VALIDATION AND EVALUATION OF QUANTITATIVE PROPRIETARY METHODS

A. METHOD COMPARISON STUDY

RUGGEDNESS

The nature of the analytical method in question will determine which parameters need to be tested. The most frequently tested ruggedness parameters, which may be critical to an analytical method, are:

- √ the composition of the samples
- √ pH
- √ timing of individual (assay) steps
- √ temperature
- √ presence of potentially interfering substances (e.g., tannins or other complexing agents, varying levels of lipids, endogenous enzymes)

Blank samples can be used for ruggedness testing as they will reveal effects caused by the matrix or the chemical batch. Information from the ruggedness test can be used to specify the conditions under which a method should be used.

BATCH-TO-BATCH VARIATION

The expert laboratory review the manufacturer's descriptions and documentation on how they have ensured that there is no significant batch-to-batch variation of their products.

METHOD PERFORMANCE CHARACTERISTICS

The method comparison study is carried out by the expert laboratory only. The proprietary method is tested against the reference method. When no reference method is available, the proprietary method can be compared against "expected results/true results" of certified reference materials (CRM), control materials and/or various spiked samples. In addition to testing the ruggedness of the method, the following performance characteristics should be estimated for quantitative methods:

- √ field of application, concentration range
- √ limit of quantification
- √ specificity
- √ inclusivity
- √ trueness
- √ repeatability
- √ recovery

The method comparison should be carried out on real incurred samples if possible to obtain in the levels of interest. Otherwise, it should be carried out on artificially contaminated samples.

Field of application, concentration range

If validation is requested for all food matrices, at least 5 relevant food matrices are selected. For other categories of matrices, select a relevant number of matrices. The levels tested should be low, medium and high in addition to a blind sample, i.e. a sample matrix free of the analyte of interest. Some methods might have a rather narrow working range, and thus it is important that the applicable concentration range is defined.

Inclusivity

Where relevant, check the inclusivity, by analysing samples from different regions/countries in order to learn if the method is able to detect the analyte not only in different foodstuffs but in foodstuffs from different regions/ countries. This is relevant for instance when analysing soy or peanut from different countries. The performance of a group specific test (e.g., betalactam or sulphonamide antibiotics) should be checked with all the relevant members of the compound group.

Limit of quantification, LOQ

Analyse a number of blank samples. The quantification limit is 10 times the standard deviation for the average of the blank sample.

Alternatively, the limit of quantification can also be determined by means of the standard deviation with added amounts of the analyte equal or close to the declared limit of quantification of the test kit.

Instead of blank samples, use, if possible, real samples that do not contain the analyte of interest. That would be the best choice.

For totally banned compounds, LOD should be estimated as 3 times the standard deviation for the average of the blank sample.

Detection capability (CC β)

The detection capability (CC β) is to be tested against the method's claimed screening/threshold limit.

Analyse at least 20 blank materials and 20 blanks per matrix fortified with the analyte(s) at the screening limit. Analyse the samples and identify the analyte(s). No more than one sample obtained can be false negative ($\beta = 5\%$, 95% confidence).

Specificity

A blank sample and one or more samples, to which a known amount of the analyte has been added, are analysed to check that there are no interferences with the analyte from other compounds in the sample, from degradation products, metabolites or know additives. In some cases, for example in the analysis of pesticides, a more concentrated extract of the blank may be analysed in order to demonstrate that no signals occur.

The specificity of a method should preferably be checked by comparing it with other methods based on other principles of analysis. Specificity can also be examined by carrying out determinations in the presence of substances suspected of interfering with the analysis. However, the analyst must be aware of the fact that the analyte may be present in the sample in more than one chemical form.

Internal Reproducibility

For examining the reliability of the method, a minimum of three different levels of analyte in each food type is required. For horizontal methods, 5 food types are required. The levels should cover the range of interest (low, intermediate and maximum levels). The number of analysis per level for the proprietary and the reference method, respectively, are five. In order to calculate the internal reproducibility, the analyses should be carried out on different days.

Calculate the mean, the standard deviation and internal reproducibility for each level and matrix as given in table 3.

Table 3: Calculations to be made on each level and matrix:

Day	Replicates for level A					calculations	Mean
1	x ₁₁	x ₁₂	x ₁₃	x ₁₄	x ₁₅	$s_1 = \sqrt{\frac{\sum_{i=1}^5 (x_i - \bar{x})^2}{4}}$	$\bar{y}_1 = \sum_{i=1}^{n=5} \frac{x_i}{5}$
2	x ₂₁	x ₂₂	x ₂₃	x ₂₄	x ₂₅	$s_2 = \sqrt{\frac{\sum_{i=1}^5 (x_i - \bar{x})^2}{4}}$	$\bar{y}_2 = \sum_{i=1}^{n=5} \frac{x_i}{5}$
.						.	.
.						.	.
.						.	.
N	x _{n1}	x _{n2}	x _{n3}	x _{n4}	x _{n5}	$s_n = \sqrt{\frac{\sum_{i=1}^5 (x_i - \bar{x})^2}{4}}$	$\bar{y}_n = \sum_{i=1}^{n=5} \frac{x_i}{5}$
Calculate the sum of the s ²						s ² =s ₁ ² +s ₂ ² +..+ s _n ²	
Calculate the repeatability, s _r ²						s ² /5n	
Calculate the mean							m = Σy/n
Calculate the standard deviation of y						S _y	
Square the standard deviation of y						S _y ²	
Calculate the between series variance						S _L ² = S _y ² - (s _r ² /2)	
Calculate the variance of the internal reproducibility						S _R ² = S _L ² + s _r ²	
Calculate the standard deviation of the repeatability, s _r , and the standard deviation of the internal reproducibility, s _R						s _r = √s _r ² s _R = √s _R ²	
Calculate the relative standard deviation of the repeatability, RSD _r , and the standard deviation of the internal reproducibility, RSD _R						RSD _r = (s _r /m) · 100 RSD _R = (s _R /m) · 100	
Calculate the limit of the repeatability, r, and the limit of the internal reproducibility, R						r = 2.8 · s _r R = 2.8 · s _R	
Calculate the theoretical value for the RSD _R , where C is the concentration ratio (e.g. mg/kg = 10 ⁻⁷). This is applicable for C ≥ 10 ⁻⁷ , for C < 10 ⁻⁷ the (RSD _R) _{theoretical} = 22%						(RSD _T) = 2C ^{-0,1505}	
Calculate the HorRat value (applicable for C ≥ 10 ⁻⁷)						(RSD _R) _{obtained} / (RSD _T)	

Draw a diagram for each matrix. Use spread sheet program. Plot the results obtained by the reference and the proprietary method; the levels make the x-axis and the means are illustrated on the y-axis. Include the confidence interval (±2s) of the reference method for each level in the graph. If the mean results obtained by the proprietary method fall within the relevant confidence intervals, there are no significant difference between the methods. That is, if the standard deviation of the reference method and the proprietary method are satisfactory.

When no reference method is available, the “expected /true results” of the spiked samples should be found within the confidence levels obtained by the proprietary method.

Acceptance criteria for the precision

The HorRat value should be no more than 2. As guidance the acceptance criteria for the precision at different levels are given in table 4. For some analyses, using advanced techniques, better precision might very well be obtained.

Table 4. Precision requirement at different concentrations based on the Horwitz/Thompson equation.

	Thompson	Horwitz equation ($2C^{-0.1505}$)							
Concentration ratio (C)	$< 10^{-7}$	10^{-7}	10^{-6}	10^{-5}	10^{-4}	10^{-3}	10^{-2}	10^{-1}	1
Concentration unit	< 0.1 mg/kg	0.1 mg/kg	1 mg/kg	10 mg/kg	0.1 g/kg	1 g/kg	10 g/kg	100 g/kg	1000 g/kg
RSD _T (%)	= 22	22	16	11	8	6	4	3	2
RSD _R (%)	≤ 44	≤ 44	≤ 32	≤ 22	≤ 16	≤ 12	≤ 8	≤ 6	≤ 4

Recovery

Recovery is usually estimated from spiked samples. However, there are some limitations in recovery information obtained from spiking. The most important is the fact that there is often no guarantee for the same behaviour of the added analyte and the native analyte. The native analyte can be physically or chemically bound to the sample matrix, while this is less likely in the case of the added analyte. Therefore the recovery can appear to be 100% in the formula below even if the yield of the native analyte present in the sample is lower. This can lead to a negative bias in a corrected analytical result.

However, recovery information is an important tool for testing the analytical method for bias (systematic error, trueness) and for the fate of unstable analytes throughout the entire analytical process.

If there is no CRM available, the recovery has to be determined by experiments using fortified blank matrix using, for example, the following scheme:

Select 18 aliquots of a blank material and fortify six aliquots at each of 1, 1.5 and 2 times the required performance limit, and then calculate the recovery (represented by R) as follows:

$$R = \frac{Q_{\text{Found}}}{Q_{\text{Original}}}$$

where Q_{Found} is the amount of the analyte recovered after processing the sample, and Q_{Original} is the known, original amount. If standard addition or spiking is used for calculating the recovery, the recovery is calculated according to the following formula:

$$R = \frac{Q_{\text{Found}} - Q_{\text{Originalsample}}}{Q_{\text{Spiked}}}$$

where Q_{Found} is the amount of analyte measured (which contains the original amount of analyte plus the added amount), $Q_{\text{Originalsample}}$ is the amount of analyte measured in the original sample, and Q_{Spiked} is the added amount of analyte. For expected recovery ranges, see table 5.

Table 5: Illustration of method criteria for levels of interest at increasing orders of magnitude:

concentration unit	0.001 mg/kg	0.01 mg/kg	0.1 mg/kg	1 mg/kg	10 mg/kg	100 mg/kg	1 g/kg	10 g/kg
Concentration ratio	10 ⁻⁹	10 ⁻⁸	10 ⁻⁷	10 ⁻⁶	10 ⁻⁵	10 ⁻⁴	10 ⁻³	10 ⁻²
Recovery (%) *	40 - 120	60 - 115	80 - 110	80-110	80 - 110	90 - 107	95 - 105	97 - 103

* Other guidelines are available for expected recovery ranges in specific areas of analysis. In cases where recoveries have been shown to be a function of the matrix other specified requirements may be applied.

Trueness

In some cases there are no (certified) reference materials available. The trueness of the proprietary method might be determined by comparing it against the reference method or by use of a (certified) reference material (could well characterized material be an option where CRM's are not available?). A z-score is often used as an estimate for the trueness and expressed as follows:

$$z\text{-score} = \frac{(\bar{X}_{\text{found}} - X_{\text{certified}})}{\left(\frac{X_{\text{certified}}}{100} \times 2(X_{\text{certified}} \times C)^{-0.1505} \right)}$$

where C = concentration ratio, e.g. % = 0.01, mg/kg = ppm = 1×10^{-6} , $\mu\text{g}/\text{kg}$ = ppb = 1×10^{-9} .

The z-score using the HorRat value is normally expected to be below 2. This gives a first estimate of the method bias. The importance of the bias is considered on a case to case basis.

B. INTERMEDIATE STUDY

An independent laboratory should verify the results obtained in the comparison study. The type of samples encompassing all relevant matrices should be selected by the organiser and should address the requirements of the second laboratory to substantiate the range and within-laboratory precision and trueness. The independent laboratory should, when possible, receive at least some of the same samples assayed by the organising laboratory, so that between-laboratory precision data can be obtained. If this is not possible, then, depending on the number of samples involved, at least half of the samples should be known (reference materials) or previously assayed by other methods.

At least three relevant food materials, artificially contaminated at 3 levels (low, medium and high) and a negative control should be used. The lowest level should be about the detection/screening level. If the intermediate study is conducted at only one additional laboratory the number of replicates for each matrix has to be at least five. The levels of the samples should be unknown to the laboratory.

Trial or practice samples should also be made available to the independent laboratory for analyst familiarisation.

REFERENCES:

AOAC – 1999: Methods Committee Guidelines for Validation of Qualitative and Quantitative Microbiological Methods.

NMKL Procedure No 20, 2007: Evaluation of results from qualitative methods.

NMKL Procedure No 4, 3rd. Ed. 2009: Validation of chemical analytical methods.

NMKL Procedure No 9, 2nd. Ed. 2007: Evaluation of method bias using certified reference materials.

NMKL-protokoll nr. 1, 2005 - Referentvejledning for det kemiske område - Udarbejdelse af analysemetoder indenfor NMKL. (only available in Danish)

NMKL Protocol No. 4, 2010: Manual for NMKL peer-verification (interlaboratory verification) of methods.

NordVal Protocol for the validation of alternative microbiological method.

CX/MAS 09/30/7 Guidelines for establishing methods criteria for the identification of relevant analytical methods.

ISO 11843: 1997 Capability of detection – Part 1: Terms and definitions, Part 2: Methodology in the linear calibration case Part 3: Methodology in the linear calibration case.

2002/657/EC: Commission Decision of 12 August 2002 implementing Council Directive 96/23/EC concerning the performance of analytical methods and the interpretation of results (Text with EEA relevance) (notified under document number C(2002) 3044)