



Nordic Committee on Food Analysis

National Veterinary Institute

PB 750 Sentrum

N-0106 Oslo

Norway

[www.nmkl.org](http://www.nmkl.org)

**NMKL Protocol No. 4, 2010:**

**MANUAL FOR NMKL PEER-VERIFICATION  
(INTERLABORATORY VERIFICATION)  
OF METHODS.**

Nordic Committee on Food Analysis

# **MANUAL FOR NMKL PEER-VERIFICATION (INTERLABORATORY VERIFICATION) OF METHODS**

## **CONTENTS**

1. INTRODUCTION .....	3
2. PROCESS OF NMKL PEER-VERIFICATION OF METHODS.....	3
3. PRACTICE FOR CHARACTERIZING METHOD PERFORMANCE.....	4
3.1. Minimum validation and performance data .....	4
3.2. Additional or Different Parameters for Qualitative Methods .....	7
3.3. Generic Descriptions.....	8
4. REQUIREMENTS FOR INDEPENDENT LABORATORIES.....	8
4.1. Identification and Selection of Independent Laboratory.....	8
4.2. Method Format.....	8
4.3. Sample Selection.....	9
4.4. Precision and Trueness.....	9
4.5. Evaluation .....	9
5. CHECKLIST FOR METHOD DEVELOPMENT, CHARACTERISATION AND RUGGEDNESS TESTING (see also [1] and [2]).....	9
5.1. Scope of Method .....	9
5.2. Ruggedness of Selected Method .....	9
5.3. Within-Laboratory Performance of Optimised Method.....	10
5.4. Characterisation/Specifications of Optimized Method .....	10
5.5. Quality Control of Method Performance .....	10
5.6. Detection records .....	10
6. CHECKLIST FOR INDEPENDENT LABORATORY PROTOCOL.....	10
7. INDEPENDENT LABORATORY REPORT FORM.....	12
8. TECHNICAL REVIEW FORM FOR NMKL PEER-VERIFICATION OF METHODS .....	13
APPENDIX A. RUGGEDNESS TESTING PROCEDURE [6] .....	14
REFERENCES.....	16

## **1. INTRODUCTION**

The purpose of the NMKL peer-verification programme is to provide a class of tested methods which have not been the subject of a full collaborative study.

The centralising of laboratory facilities and the use of more sophisticated, expensive and specialised analytical equipment have created a need for a procedure to verify analytical methods in situations where it is not possible to perform a full collaborative study due to lack of (participating) laboratories.

Verification of a method includes describing, validating and testing it for performance according to international recognised standards, as well as

- Establishing acceptable performance parameters within a laboratory
- Demonstrating acceptable performance in at least one more independent laboratory

The programme is intended only for chemical methods.

The protocol is elaborated by an NMKL working group chaired by Torben Leth, The National Food Institute, Technical University of Denmark. The other project members have been Flemming Hansen, Denmark, Lenna Saari, Finland, Guðjon Atli Auðunsson, Iceland, Per Lea, Norway, Lars Jorhem, Åsa Rosengren and Gunnar Forsgren, Sweden.

NMKL invites all readers and users of this manual to submit comments and views on its contents to the General Secretariat of NMKL.

© NMKL

## **2. PROCESS OF NMKL PEER-VERIFICATION OF METHODS**

Based on recommendations from the NMKL National Committees, the Secretary General appoints a method referee.

In collaboration with the Chair of the appropriate subcommittee, the Secretary General also appoints a technical referee.

The method referee appoints an independent laboratory [4].

The method referee

- Writes the method in the NMKL format according to ISO 78-2 [7].
- Conducts a full validation of the method in compliance with the NMKL procedure for validation of chemical methods [1] or the NMKL procedure for evaluation of results from qualitative methods [2]
- Designs a protocol for second laboratory testing of method performance according to the NMKL protocol for chemical analysis [3]
- If appropriate, selects a third laboratory [4]
- Prepares necessary collaborating samples and sends the method description, samples, protocol and report form to the second laboratory

- Analyses the data from the second laboratory and sends the method validation report and the independent laboratory report to the technical referee

The independent laboratory

- Conducts the method performance test according to the method and the protocol
- Submits the results to the method referee according to the report form

The technical referee:

- Selects at least 2 experts to provide technical reviews within 3 weeks, and informs the Secretary General of names and addresses
- Acts as follows:
  - If both reviews are negative, the technical referee notifies the method referee that the method is not acceptable.
  - If both reviews are positive, with minor comments, the technical referee notifies the method referee that the method is acceptable with minor revisions.
  - If the reviews are split, the technical referee can seek the advice of one or more additional experts to obtain other reviews.
  - If both reviews are positive, with moderate revisions, the technical referee requests the necessary revisions from the method referee before acceptance of the method. The NMKL chairman and subcommittee chairman carry out the final approval, and the secretariat publishes the method as an NMKL verified method.

### **3. PRACTICE FOR CHARACTERIZING METHOD PERFORMANCE**

#### ***3.1. Minimum validation and performance data***

##### **Trueness**

Trueness is determined by the use of certified reference materials, reference methods of known uncertainty or recovery from spiked samples.

Reference materials should be carried through the entire procedure with each batch. If fortified or spiked samples are used, the method of fortification should be described.

##### **Recovery**

For quantitative methods, recovery is measured for the following concentrations:

- Negative
- Low (max. 3×LOD)
- Medium (e.g. MRL if present)
- High

with at least 3 replicates for the analysis of both samples with the original amount and samples with added amount.

The recovery is calculated as follows:

$$\% \text{ Recovery} = \frac{\text{Amount found} - \text{Amount from original}}{\text{Amount added}}$$

Guidelines are available for expected recovery ranges in specific areas of analysis. The design of recovery experiments should be appropriate for the intended purpose of the method and should cover all relevant matrices for horizontal methods.

### **Standard Curve**

A sufficient number of standard solutions are needed to define the response in relation to the concentration. The number of standard solutions is a function of the range of concentration. In most cases, a minimum of 5 concentrations of standard solutions (not zero) is appropriate to prepare a standard curve. The curve should be statistically tested and expressed.

For nonlinear curves, more standard solutions are necessary.

### **Limit of Detection - LOD**

The LOD is calculated as follows:

$$\text{LOD} = \bar{X} + 3s$$

where  $\bar{X}$  is the mean value of the matrix blank readings (n=20) and s is the standard deviation of the mean, expressed in analyte concentration.

Without a signal for the blank sample, as in many chromatographic methods, the instrumental interference is magnified. Measure the width of the baseline at least 10 times with 1 min intervals and calculate the standard deviation.

For methods with less than 100% recovery, the limit of detection should be corrected for recovery.

### **Decision Limit - CC $\alpha$**

According to EU regulations, the decision limit CC $\alpha$  is required instead of the LOD for methods used to analyse for some residues in food [4].

For substances with no permitted limit, CC $\alpha$  can be established using either procedure 1) or 2) below.

1) By using the standard curve procedure according to ISO 11843 [5].

This procedure uses blank materials fortified at and above the minimum required performance level in equidistant steps. Analyse the samples. After identification, plot the signal against the added concentration. The corresponding concentration at the y-intercept plus 2.33 times the standard deviation of the within-laboratory reproducibility of the intercept, equals the decision limit. This is applicable to quantitative assays only ( $\alpha = 1\%$ ).

2) By analysing at least 20 blank materials per matrix to be able to calculate the signal-to-noise-ratio in the time window where the analyte is expected.

Three times the signal-to-noise-ratio can be used as the decision limit. This is applicable to quantitative and qualitative assays.

For substances with established permitted limit, CC $\alpha$  can be established using either procedure 3) or 4) below.

3) By using the standard curve procedure according to ISO 11843 [5].

As procedure 1), this procedure uses blank materials fortified around the permitted limit in equidistant steps. Analyse the samples. After identification, plot the signal against the added concentration. The corresponding concentration at the permitted limit plus 1.64 times the standard deviation of the within-laboratory reproducibility, equals the decision limit ( $\alpha = 5\%$ ).

4) By analysing at least 20 blank materials per matrix fortified with the analyte(s) at the permitted limit. The concentration at the permitted limit plus 1.64 times the corresponding standard deviation equals the decision limit ( $\alpha = 5\%$ ).

### **Limit of Quantification - LOQ**

The LOQ is calculated as follows:

$$LOD = \bar{X} + 10s$$

where  $\bar{X}$  is the mean value of the matrix blank readings ( $n=20$ ) plus 10 standard deviations of the mean, expressed in analyte concentration.

For methods with less than 100% recovery, the limit of quantification should be corrected for recovery.

### **Detection Capability - CC $\beta$**

In the EU, the detection capability CC $\beta$  is required instead of the LOQ for methods used to analyse for some residues in food [4].

For substances for which no permitted limit has been established, CC $\beta$  can be established using either procedure 1), 2) or 3) below.

1) By using the standard curve procedure according to ISO 11843 [5].

This procedure uses representative blank materials fortified at and below the minimum required performance level in equidistant steps. Analyse the samples. After identification, plot the signal against the added concentration. The corresponding concentration at the decision limit plus 1.64 times the standard deviation of the within-laboratory reproducibility of the mean measured content at the decision limit, equals the detection capability ( $\beta = 5\%$ ).

2) By analysing at least 20 blank materials per matrix fortified with the analyte(s) at the decision limit. Analyse the samples and identify the analytes. The value of the decision limit plus 1.64 times the standard deviation of the within-laboratory reproducibility of the measured content, equals the detection capability ( $\beta = 5\%$ ).

3) Where no quantitative results are available, the detection capability can be determined by the investigation of fortified blank materials at and above the decision limit. In this case, the concentration level at which only  $\leq 5\%$  false compliant results remain, equals the detection capability of the method. Therefore, at least 20 investigations for at least one concentration level have to be carried out in order to ensure a reliable basis for this determination.

In the case of substances for which a permitted limit has been established, CC $\beta$  can be established using either procedure 4) or 5) below.

4) By using the standard curve procedure according to ISO 11843 [5].

This procedure uses representative blank materials fortified around the permitted limit in equidistant steps. Analyse the samples and identify the analyte(s). Calculate the standard deviation of the mean measured content at the decision limit. The corresponding concentration at the value of the decision limit plus 1.64 times the standard deviation of the within-laboratory reproducibility, equals the detection capability ( $\beta = 5\%$ ).

5) By analysing at least 20 blank materials per matrix fortified with the analyte(s) at the decision limit. The value of the decision limit plus 1.64 times the corresponding standard deviation, equals the detection capability ( $\beta = 5\%$ ).

### **Precision**

The relative standard deviation of individual results should be based on results from different days, different analysts, different standard curves, different batches of reagents, and different matrices. Precision is usually defined in terms of reproducibility and repeatability:

#### **Reproducibility**

Ideally, the precision should be measured for performance between laboratories. Obtain values for different samples, different concentrations ( $\frac{1}{2}x$ ,  $1x$  and  $1\frac{1}{2}x$  or  $2x$ ), and separate days. Reproducibility data is generated from the referee's own lab plus one or two independent labs.

#### **Repeatability**

This component of reproducibility can be determined within a laboratory.

### **Specificity**

Reagent blanks and field blanks should be run to ensure that no interfering compounds are present. To verify the specificity of the method for the analyte(s) of interest, results should be tested under different experimental conditions, e.g. two different analytical principles or two different detection techniques. The method should be able to distinguish the analyte from known interfering materials; and the behaviour of the analyte during analysis should be indistinguishable from that of the corresponding standard material in the appropriate matrix.

### **Comparison to Existing Methods**

When applicable, comparison to an existing method is strongly recommended (preferably a validated or collaboratively studied method). Depending on topic, a suitable number of comparative analyses, if possible at levels tested in a collaborative study of the "old" method, are generally sufficient to show equivalent performance.

## **3.2. Additional or Different Parameters for Qualitative Methods**

### **Cross-Reactivity**

Samples should be tested for cross-reactivity with other known substances.

### **Specificity**

Usually determined as cross-reactivity. All samples should be tested for cross-reactivity with other known substances.

### **Incident of False Positives/False Negatives**

For qualitative methods, the precision may be expressed as true and false positive (or negative) rates. These rates should be determined at several concentrations including the specification level. Data from a confirmatory method comparison study should be provided if the method(s) is(are) applicable to the same matrix/cases and concentration range(s). In the absence of method comparison, populations of negative and positive fortified samples should be analysed. Formulas used in the calculation must be defined and included in the method. Response rates can also be expressed in terms of specificity, sensitivity, and predictive rates. Exact definitions of the used specificity, sensitivity and predictive rates should be described. Statistical calculations and evaluation of the results must be carried out according to NMKL Procedure No. 20 [2].

### **3.3. Generic Descriptions**

#### **Performance-Based Descriptions of Reagents and Apparatus**

Apparatus and reagents specified in methods should be specified by performance or generic characteristics rather than brand name. To avoid referring to specific brand products, critical parameters should be identified and defined. System suitability standards and tests should be established and clearly presented so that the product is defined generically, and equivalency can be readily determined. Vendor or brand names may be included and stated as "examples". A disclaimer should be included in a footnote stating that "NMKL does not endorse the brand names used in this method and these are only included in the method as examples".

#### **System Suitability Tests**

Where appropriate, methods should contain system suitability tests for determining the acceptability of reagents and apparatus. A system suitability test is a clear, concise, specific technical statement of one or more definitive tests, which indicate that all controllable variables attributable to a particular product used in a method are within acceptable limits. This test is performed to determine whether a product used in the method has any measurable effect on the method results.

## **4. REQUIREMENTS FOR INDEPENDENT LABORATORIES**

### **4.1. Identification and Selection of Independent Laboratory**

The method referee selects the independent laboratory, which has the assignment of verifying and evaluating the method. It is not expected to duplicate all the data generated by the referee, but to verify the performance characteristics. The independent laboratory can be a part of the method referee's organisation as long as the personnel used to conduct the independent laboratory validation

- do not report to the person involved in developing or submitting the method,
- do not use the same equipment or supplies.

The laboratory should be one where the type of assay and matrices in question are used on a routine basis.

### **4.2. Method Format**

The independent laboratory should receive the method in the step-by-step format specified by ISO 78-2 [7], which has been submitted to NMKL for publication. This is very important because if the laboratory has any problems, the exact step can be identified and clarified.

### **4.3. Sample Selection**

The type of samples encompassing all relevant matrices should be selected by the method originator 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 method originator, 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. Trial or practice samples should also be made available to the independent laboratory for analyst familiarisation.

### **4.4. Precision and Trueness**

The independent laboratory should verify the precision and trueness of the method. Precision should be determined in at least some of the representative matrices of the target samples. Trueness should be determined through the use of reference materials, other known samples, or standard additions. If available, all three methods should be used.

### **4.5. Evaluation**

The independent laboratory should prepare a written evaluation of the method, including suggestions for alternative steps or procedures. Matters that should also be addressed, are trueness and precision as experienced by the second laboratory, ease-of-use of the method, special glassware/instrumentation/etc. requirements, and general comments. The independent laboratory should maintain a record of all communications with the method referee pertaining to the method and interpretation of results. This record should be submitted along with the evaluation. The independent laboratory should retain detection records (chromatograms, spectra, etc.) for all samples analysed, including blanks and standards, and submit these along with the method evaluation. In the event of non-agreement by the second laboratory, the method should not be published until all issues are resolved. For quantitative methods, a HorRat value  $\leq 2$  is required for a method to be accepted.

## **5. CHECKLIST FOR METHOD DEVELOPMENT, CHARACTERISATION AND RUGGEDNESS TESTING (see also [1] and [2])**

### **5.1. Scope of Method**

Characterise the scope of the method as follows:

- Need/purpose of method (e.g. surveillance, monitoring, quality control)
- Type of method (e.g. reference, screening, confirmatory)
- Applicability of method (e.g. concentration range, analyte(s), matrix/ces))

### **5.2. Ruggedness of Selected Method**

Follow procedures for ruggedness testing as outlined, for example, in one or both of the following:

Youden & Steiner (1975) Statistical Manual of the AOAC, AOAC International, Arlington, VA. (See APPENDIX A), Dols & Armbrrecht (1976) J. Assoc. Off. Anal. Chem. **59**, 1204-1207 or in annex A of this document.

### **5.3. Within-Laboratory Performance of Optimised Method**

Provide data on the following:

- Trueness or Bias (systematic error)
- Recovery
- Response factor /Standard curve function (response vs concentration in solution)
- Analytical function (response vs concentration in matrix, including blanks)
- Limit of detection, limit of quantification
- Repeatability
- Mathematical transformations needed, if any

### **5.4. Characterisation/Specifications of Optimized Method**

Provide information on the following:

- Interferences (specificity): impurities, contaminants, additives, matrices, metabolic products
- Performance specifications and acceptability criteria for reagents, instruments
- Suitability tests for systems
- Critical steps or parameters
- Comparison with other methods, if available

### **5.5. Quality Control of Method Performance**

Provide information/specifications/requirements for the following:

- Acceptance limits
- Standard curve checks / Regression
- Reference materials
- Blank checks
- Standards
- Fortified samples

### **5.6 Detection records**

Copies of chromatograms, spectra, absorbance readings, etc. for samples, standards and controls, clearly labelled.

## **6. CHECKLIST FOR INDEPENDENT LABORATORY PROTOCOL**

### 1. Method

- NMKL Peer-Verified Method format according to ISO 78-2 [7]

### 2. Samples and Standards

- Clearly labelled w/ identification, quantity, potency (for standards) and storage conditions

### 3. Laboratory Evaluation

- Directions for preparation of standards and standard curve
- Directions for familiarisation (at least 2 control samples and 2 fortified samples)

4. Substantiation of Within-Laboratory Performance

- Trueness and/or recovery
- Repeatability

5. Between-Laboratory Performance

- Reproducibility

6. Report Form

- Specify number of significant figures as needed according to appropriate topic

## 7. INDEPENDENT LABORATORY REPORT FORM

Author(s):

Method:

Independent Laboratory:

Independent Laboratory Contact (name, address, phone, e-mail):

<b>Technical Evaluation Criteria:</b>	<b>Yes</b>	<b>No</b>
1. Is the method technically sound?		
2. Do the method description and peer evaluation cover all pertinent items on the Checklist for Method Development, Characterisation and Ruggedness Testing?		
3. Are the conclusions and method applicability statements, as proposed by the author, valid based on the data?		
4. Are sufficient data points evaluated?		
5. Is the method practical, including safety and environmental considerations?		
6. Are critical steps identified?		
7. Are quality control measures identified/adequate?		
<b>Editorial Evaluation Criteria:</b>		
8. Is the method description clear and complete?		
9. Are all tables, figures, equations and terms sufficiently explained?		
10. Have reagents and apparatus been described in performance terms with system suitability tests where necessary?		

Submit with Report Form: All data; chromatograms, absorbance curves, etc. for blanks, standards, control samples and samples.

Comments:

## 8. TECHNICAL REVIEW FORM FOR NMKL PEER-VERIFICATION OF METHODS

Author(s):

Method:

Technical Recommendation:

\_\_\_\_\_ The data indicate that the method is acceptable.

\_\_\_\_\_ The method cannot be properly evaluated until additional information is supplied as indicated below or specified on a separate sheet.

\_\_\_\_\_ The data indicate that the method is not acceptable.

<b>Technical Evaluation Criteria:</b>	<b>Yes</b>	<b>No</b>
1. Is the method technically sound?		
2. Do the method description and peer evaluation cover all pertinent items on the Checklist for Method Development, Characterisation and Ruggedness Testing?		
3. Are the second laboratory's comments addressed?		
4. Are the conclusions and method applicability statements valid based on data presented?		
5. Are sufficient data points evaluated?		
<b>Editorial Recommendation:</b>		
The report presentation is acceptable.		
The report presentation requires revisions as indicated below or on a separate sheet.		
<b>Editorial Evaluation Criteria:</b>		
6. Is the method presented in the correct format?		
7. Is the method description clear and unambiguous?		
8. Are all tables, figures, equations and terms explained?		
9. Have reagents and apparatus been described in performance terms with system suitability tests where necessary?		

Comments:

## APPENDIX A. RUGGEDNESS TESTING PROCEDURE [6]

When the initiating laboratory develops and standardises the procedures of a method, data may be collected for a set of operations and equipment that is never varied. This process does not reveal what will happen during a collaborative study in a number of laboratories, each of which has its own set of reagents, equipment and routines. Preparation of standards, time and temperature variations, instrument calibration and performance, and analyst technique all contribute to minor variations even when a procedure is followed "exactly". The only way to forecast the performance of a method under different laboratory conditions, is to deliberately introduce reasonable variations and observe what happens. If the procedure is "rugged", it should be transparent to minor variations and the results should not be affected.

A ruggedness test is a scheme of attack that will conserve labour but be sensitive enough to pick up small effects of variation. The scheme does not study one alteration at a time, but introduces several changes at once in such a manner that the effects of individual changes can be ascertained.

### EXAMPLE:

Let A, B, C, D, E, F and G denote the nominal values for seven different factors that might influence the results if their values were varied slightly. Let their alternative values be denoted by a, b, c, d, e, f and g.

The conditions for running a determination can be completely specified by the seven letters, each letter being either capital or lower case. There are  $2^7$  or 128 different combinations, and this makes it impractical to test. However, it is possible to choose a subset of eight combinations that adequately balances upper and lower case conditions.

Eight combinations of seven factors used to test the ruggedness of an analytical method								
Combination or Determination No.								
Factor Value	1	2	3	4	5	6	7	8
A or a	A	A	A	A	a	a	A	a
B or b	B	B	b	b	B	B	B	b
C or c	C	c	C	c	C	c	C	c
D or d	D	D	d	d	d	d	D	D
E or e	E	e	E	e	e	E	E	E
F or f	F	f	f	F	F	f	f	F
G or g	G	g	g	G	g	G	G	g
Observed result	s	t	u	v	w	x	y	z

The table specifies the value for seven factors to be used while running eight determinations. The results are designated by s through z. How can one isolate the separate effects of factor changes, when 4 factors are always altered from the initial procedure conditions of all capital letters?

To find out whether changing factor A had an effect, compare the average  $(s + t + u + v)/4$  with the average  $(w + x + y + z)/4$ . Determinations 1, 2, 3 and 4 were run with factor level A, and determinations 5, 6, 7, and 8 with factor level a. This gives two groups of 4 determinations, and each group includes the other six factors twice at the capital level and twice at the lower case level. The effects of these factors, if present, cancel out, leaving only the effect of changing A to a.

Whenever the eight combinations are split into two groups of four on the basis of one of the letters, all other factors cancel out. Every one of the combinations is tested by only eight determinations. The effect of altering G to g, for example, is examined by comparing the average  $(s + v + x + y)/4$  with  $(t + u + w + z)/4$ .

Collect the seven average differences for A - a, B - b, etc., and list them in order of size. If one or two factors are having an effect, their differences will be substantially larger than the group of differences associated with the other factors. This ranking is a direct guide to the method's sensitivity to modest alterations. A useful, i.e. rugged, method should not be affected by changes that will most certainly be encountered between laboratories.

If there is no outstanding difference, the most realistic measure of the analytical error is given by the seven differences obtained from the averages for capitals minus the average for corresponding lower case letters. Denote these seven differences by  $D_a, D_b, \dots, D_g$ . Estimate the standard deviation by squaring the differences and take the square root of  $2/7 * \text{the sum of their squares}$ .

To check the calculation, compute the standard deviation obtained from the eight results, s through z; obtain the mean of the eight results. Square the eight differences from the mean, sum the squares, divide by  $8 - 1$ , and take the square root.

## REFERENCES

- 1] NMKL Procedure No. 4, 2009, 3rd. Ed., Validation of chemical analytical methods
- 2] NMKL Procedure No. 20, 2007, Evaluation of results from qualitative methods
- 3] NMKL-protokoll nr. 1, 2005 Referentvejledning for det kemiske område (Guide for the referees within NMKL, available in Danish only)
- 4] Commission Decision of 12 August 2002 implementing council Directive 96/23/EC concerning the performance of analytical methods and the interpretation of results
- 5] ISO 11843: 1997 Capability of detection – Part 1: Terms and definitions, Part 2: Methodology in the linear calibration case
- 6] Youden & Steiner (1975) Statistical Manual of the AOAC, AOAC International, Arlington, VA
- 7] ISO 78-2:1999. Chemistry – Layouts for standards – Part 2: Methods of chemical analysis