

NordVal International Protocol No. 1

Validation of microbiological alternative (proprietary) methods against a reference method

Table of contents

NORDVAL INTERNATIONAL PROTOCOL NO. 1	1
TABLE OF CONTENTS	1
1. FOREWORD	2
2. SCOPE	2
3. DEFINITIONS	3
4. QUALITATIVE METHODS	5
4.1 METHOD COMPARISON STUDY	5
4.1.1 Sensitivity study	5
4.1.1.3 Confirmation of results	7
4.1.2 Level of Detection and Relative level of detection study	11
4.1.3 Selectivity study	13
4.2 INTERLABORATORY STUDY, ILS	14
4.2.1 Number of laboratories	14
4.2.2 Number of samples	15
4.2.3 Summary of data and trueness calculation	15
4.2.4 Interpretation of data	18
5. QUANTITATIVE METHODS	20
5.1 METHOD COMPARISON STUDY	20
5.1.1 Relative trueness study	20
5.1.2 Accuracy profile study	24
5.1.3 Limit of quantification study	27
5.1.4 Inclusivity and exclusivity study	27
5.2 INTERLABORATORY STUDY, ILS	29
5.2.1 Number of laboratories and samples	29
5.2.2 Calculations, summary and interpretation of data	29
6. REFERENCES	34
7. APPENDIX OF RESOURCES	34
8. MAIN CHANGES IN THIS VERSION (29.06.2026)	34

1. Foreword

This Protocol follows the EN ISO 16140-2: Microbiology of the food chain — Method validation — Part 2: Protocol for the validation of alternative (proprietary) methods against a reference method.

Additional to the EN ISO 16140-2, the NordVal Protocol includes the requirement of and guide to estimate the Limit of Detection (LOD) (4.1.2).

In the version of November 2021, the requirement of using Kappa, a statistical entity for testing the agreement between two qualitative methods has been removed. The calculations for the sensitivity have been clarified.

Minor layout and graphical improvements have been made.

In the version of June 2026 (29.06.26), the additions given in ISO 16140-2 Adm. 1 relevant for the validation of microbiological alternative (proprietary) methods against a reference method are included. For the main revisions see [Section 8](#).

2. Scope

This NordVal International protocol describes the technical procedures for validation of alternative methods for microbiological analyses of food, water, animal faeces, feed, samples from equipment and processing areas and environmental samples.

The validation comprises two phases:

- A method comparison study of the alternative method against a reference method (performed by an organising laboratory)
- An interlaboratory study of the alternative method compared against the reference method (organised by an organising laboratory)

The protocol is valid for both qualitative and quantitative methods. The method performance characteristics to be determined are:

Qualitative methods	Quantitative methods
• Sensitivity • Specificity • Relative trueness • Level of detection (LOD) and relative level of detection (RLOD) • Selectivity (inclusivity/exclusivity)	• Relative trueness • Accuracy profile • Limit of quantification • Selectivity (inclusivity/exclusivity)

Some of the characteristics are evaluated against so-called acceptability limits (AL). These AL are empirically based (according to EN ISO 16140-2).

3. Definitions

Acceptability limit, AL

A numeric (minimum or maximum) value for required performance of the method.

Accuracy

The closeness of agreement between a measured quantity value and an assigned quantity value of an analyte.

Accuracy profile, AP

A quantitative comparison on the difference in the results obtained on equivalent samples by two different methods.

Alternative method

An alternative method is a method of analysis that demonstrates or estimates, for a given category of products, the same analyte as is measured using the corresponding reference method.

Analyte

The analyte is the component demonstrated or measured by the method of analysis. It may be the microorganism, its components or products.

Bias

Estimate of a systematic measurement error.

Organising laboratory

The organising laboratory is an independent expert laboratory having qualified staff and skills to perform the method comparison study and organise the collaborative study. The availability of an experienced statistician is essential for the analysis of the results. NordVal International requires that the organising laboratory:

- is economically independent of the applicant
- is accredited to perform the reference method
- has the competence and capacity to perform both the reference method as well as the alternative method
- has a competent person that is in charge of the validation

The organising laboratory should make a detailed validation plan and produce a validation report that must be approved by NordVal International.

Fractional recovery

Validation criterion that is satisfied when replicate samples of either the alternative or reference method yield 50% (in the interval from 25% to 75%) positive responses.

False positive test result, FP

Positive result by the tested method that is actually confirmed as a negative result

False positive ratio, FPR

Number of false positive test results related to the total number of samples tested.

Interlaboratory study, ILS

Study of the performance of the alternative method using identical samples in several laboratories and controlled by the organising laboratory.

Level of detection, LOD

Measured analyte concentration, obtained by a given measurement procedure, for which the probability of detection is x . In this protocol $x = 50$, i.e. LOD₅₀ is the level of detection for which 50 % of tests give a positive result.

Method comparison study

Study performed by the organising laboratory of the alternative method against the reference 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.

Relative accuracy, RA

The degree of correspondence between the response obtained by the alternative method and the reference method on artificially inoculated samples.

Relative level of detection, RLOD

Level of detection at $P=0.50$ (LOD_{50}) of the alternative proprietary method divided by the level of detection at $P = 0.50$ (LOD_{50}) of the reference method.

Relative trueness, TR

Closeness of agreement between the obtained average of several numbers of replicate measured by the alternative method and the reference method.

Seeding

Seeding is based on contamination of natural samples by a diluted culture and subsequent storage of the sample for an extended period in order for the microorganisms to adapt to the environmental conditions of the food.

Selectivity

The selectivity is a measure of a) the inclusivity: detection of the target microorganism from a wide range of strains, and b) the exclusivity: the lack of interference from a relevant range of non-target microorganisms.

Spiking

Spiking is inoculation of a diluted culture into the matrix after relevant stress conditions have been applied to the culture.

Sensitivity, SE

The sensitivity is the ability of the method correctly to detect the analyte.

Specificity; SP

The specificity is the ability of the method correctly not to detect the target microorganism when not present.

Reference method

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

Repeatability, r

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).

Reproducibility, R

The reproducibility is the closeness of agreement between single test results on identical test material using the same method and obtained by operators in different laboratories using different equipment.

Validation of an alternative method

The validation of an alternative method is the procedure to demonstrate if the alternative method provides equivalent results compared to the reference methods.

4. Qualitative methods

4.1 Method comparison study

The method comparison consists of three parts:

- a comparative study of the results of the reference method to the results of the alternative method in naturally and/or artificially contaminated samples (so-called sensitivity study)
- a comparative study to determine the level of detection (LOD) and the relative level of detection (RLOD) in artificially contaminated samples (so-called RLOD study)
- a selectivity (inclusivity/exclusivity) study of the alternative method

The results (tables and calculations) of the different parts and the interpretation of the results, including discrepant results, shall be given in a study report.

The organizing laboratory shall be competent to perform both the reference method as well as the alternative method.

NOTE: Competence can be demonstrated in different ways (e.g. for the reference method, a laboratory meeting the requirements of ISO/IEC 17025, and for the alternative method, a documented training)

The reference and the alternative methods shall be performed with, as far as possible, exactly the same sample (same test portion). In some cases, the same test portion can be used for both the reference and the alternative method, as both methods have exactly the same first step in the (enrichment) procedure. When using the same test portion for both the reference and the alternative method, it is called a **“paired study”**. In situations, where there is no shared initial (enrichment) step for both the reference and the alternative method, different test portions, coming from the same batch or lot of product, are used. Then this is named an **“unpaired study”**. If there is a common initial step in the (enrichment) procedures, a **paired study** design is mandatory.

When the reference and alternative methods are based on two different principles and are performed with the same test portion, but do not share a common enrichment procedure, an unpaired data study is performed. For example, when a qualitative alternative method is validated against a quantitative reference method at a limit of 100 cfu/g. In this case, a suspension of the (food) item can be used to inoculate both culture media for the reference method and the alternative method before any enrichment/multiplication of the microorganism.

4.1.1 Sensitivity study

The sensitivity study aims to determine the difference in sensitivity between the reference and the alternative method. This study is conducted using naturally and/or artificially contaminated samples. Different matrices shall be tested. For evaluating the results, Acceptability Limits (AL) have been defined for the maximum acceptable difference. The AL depend on the type of study (paired/unpaired) and the number of matrices tested.

4.1.1.1 Selection of matrices

The selection of categories and types used within the validation will depend on the type or group of microorganisms and the scope of the validation.

If the method is to be applied for a broad range of foods, at least 5 food categories shall be studied. If the method is to be applied for only 1 category, then only types within this category needs to be tested. In addition to food; feed samples, environmental samples and primary production stage samples can be included as additional categories.

For all selected categories, at least 3 different food types shall be included. Which matrices to be included in the study are given in separate documents for the specific microorganisms on the NordVal International website. Matrices for feed of other than plant origin, should include meat- and bone meal and fishmeal. Matrices for feces should include feces from poultry, swine and cattle unless a specific matrix is chosen. For other categories of matrices, a relevant number of matrices are selected.

If a method should be applicable for swabs, swabs must be included as one of the three different food types within a category.

4.1.1.2 Number of samples and contamination levels

For each category, analyse 20 samples for at least 3 types (3 types x 20 samples for each type = 60 samples). For each category, at least 30 samples shall have a positive result by the reference and/or the alternative method. Fractional positive results by either the reference or alternative method (i.e. samples should not be all positive or all negative) shall be obtained for each type tested. In the ideal situation 10 samples (50%) tested per type should be positive and 10 negative but could range between 25% and 75%.

Preferably, the samples for the study should be naturally contaminated. As this might be hard to obtain, artificial contaminated samples can be used. If artificial contamination is used the level of the inoculate should be close to the expected LOD.

Contamination of samples is often done by seeding or spiking. The choice of stress depends on the microorganism and the normal processing of the matrix (e.g. pasteurization). Stressing can be performed by physical or chemical treatment.

Example on artificial contamination / spiking:

Inoculate a tube of non-selective enrichment broth with the designated strain and incubate the broth at optimal conditions for the strain. After incubation, dilute the culture in a suitable diluent until the desired level(s). Stress the organisms by for instance heat treatment for 10 min. at 50 °C in a given temperature bath.

Evaluate the injury efficiency by enumerating the pure culture on both selective and non-selective agars. There should be more than 0.5 log₁₀ cfu/ml difference to indicate a sufficient stress application. Dilute the culture further if needed. More information on the spiking methodology is described in article by Jasson et.al (see References).

Inoculate by pipetting a known volume into individual samples. The volume of the inoculum should be as low as possible as it should not influence the water activity significantly. Generally, 0.25 ml per 25 gram of sample is used. After inoculation, mix the item thoroughly to insure homogeneity.

Select two strains relevant for the matrix.

Check the contamination levels before using the samples in the validation study. If necessary, adjust the contamination levels to the desired contamination.

The alternative method shall be evaluated for a defined test portion size (e.g. 25 g, 200 g, 375 g) during the validation study. The method is considered to be validated for any test portion size up to the validated test portion size if the testing protocol (dilution ratio, incubation time and incubation temperature) is the same as that used during the validation study.

EXAMPLE. A reference method used in a validation study of this document was validated for a “broad range of foods” using a 25 g test portion and a 1:10 dilution ratio. The alternative method was validated for a “broad range of foods” using a 375 g test portion and a 1:5 dilution ratio at a determined incubation time. In practice, a user laboratory can use the alternative method for all food items (broad range of foods) using a test portion size up to 375 g test portion and a 1:5 dilution ratio at the validated incubation time (unless stated differently

by the organization involved in the method validation).

4.1.1.3 Confirmation of results

If the alternative method does not include confirmation step, then the (end) results of the alternative method will be the results obtained after enrichment and detection step. Only deviating results between the reference method and the alternative method need to be confirmed. This confirmation is needed to determine whether the result is a true-positive or false-positive result. The confirmation test or tests shall be able to recover and confirm the identity of the isolate as being the target of the method. These test(s) can be based on the confirmation procedure of the reference method, the confirmation step of the alternative method in case this procedure is able to isolate and confirm the identity of the target analyte, a combination of both, or by any other means that is able to isolate and confirm the identity of the target analyte. If isolation is not possible, then another appropriate method might be used.

4.1.1.4 Calculation and interpretation of results

The interpretation of the results (positive agreement, negative agreement, etc.) is based on a comparison of the reference method result (column 1 in **Tables 4.1** and **4.2**) and the alternative method result, including any confirmations as described in the alternative method protocol (column 2 in **Tables 4.1** and **4.2**). When positive or negative deviations are obtained, a footnote should be included at the end of each table to provide additional explanations for the interpretation of the deviations. The footnotes indicate if the result is due to a false positive or false negative result of the alternative method. The footnote is a comparison of the results of the alternative method (including any confirmations as described in the alternative-method protocol) (column 2 in **Tables 4.1** and **4.2**) and the confirmed alternative method (by any means) (column 3 in **Tables 4.1** and **4.2**).

Table 4.1: Comparison and interpretation of sample results for a paired study

Reference method	Result of each sample		
	Alternative method (including any confirmations as described in the alternative-method protocol)	Confirmed positive results (by any means) ^a	Interpretation based on confirmed alternative results
+	+	Not needed ^b	Positive Agreement (PA)
-	-	Not needed ^b	Negative Agreement (NA)
+	-	Not needed ^b	Negative deviation due to false negative result by the alternative method (ND _{FN(alt)})
-	+	+	Positive deviation (PD) The alternative method provides true positive result
-	+	-	Positive results for alternative method, turns out negative after confirmation (PD _{FP(alt)}) ^c

^a Confirmation of the alternative-method result is done according to 4.1.3.3.

^b No need for additional confirmation test(s). Confirmed alternative-method result is the same as the alternative-method result.

^c This false positive result (FP) shall also be used to calculate the false positive ratio.

Table 4.2 Comparison and interpretation of sample results between the reference and alternative methods for an unpaired study

Result of each sample			
Reference method	Alternative method (including any confirmations as described in the alternative-method protocol)	Confirmed positive results (by any means) ^{a,b}	Interpretation based on confirmed alternative results
+	+	+	Positive Agreement (PA)
+	+	-	Positive Agreement due to false positive alternative-method result (PA _{FP(alt)}) ^c
-	-	-	Negative Agreement (NA)
-	-	+	Negative Agreement due to false negative alternative-method result (NA _{FN(alt)})
+	-	-	Negative deviation. The alternative method provides false negative result (ND _{FN(alt)})
+	-	+	Negative Deviation due to false negative alternative-method result (ND _{FN(alt)})
-	+	+	Positive deviation (PD)
-	+	-	Positive results for alternative method, turns out negative after confirmation (PD _{FP(alt)}) ^c

^a Confirmation of the alternative-method result is done according to 4.1.3.3.

^b Confirmation by any means is only required when the results of the alternative method does not produce viable organisms. This is used as confirmed alternative method result in comparison to the reference method result.

^c These false-positive results (FP) shall also be used to calculate the false positive ratio.

Determine the Total Negative Deviation (TND) and Total Negative Agreement (TNA) for the validation study.

Paired evaluation: Total Negative Deviation: TND = ND_{FN(alt)}
 Total Negative Agreement: TNA = NA + PD_{FP(alt)}

Unpaired evaluation: Total Negative Deviation: TND = ND + ND_{FN(alt)} + PA_{FP(alt)}
 Total Negative Agreement: TNA = NA + NA_{FN(alt)} + PD_{FP(alt)}

Table 4.3 Summary of results obtained by the alternative and the reference methods of all samples for each category and for all categories.

	Reference method positive (R+)	Reference method negative (R-)
Alternative method positive (A+)	R+ / A+ Positive agreement (PA)	R- / A+ Positive Deviation (PD)
Alternative method negative (A-)	R+ / A- Negative Deviation (TND)	R- / A- Negative agreement (TNA)

How to estimate the sensitivity for both the alternative method and the reference method, relative trueness and false positive ratio for the alternative method are given in the Table 4.3.2. An Excel sheet is available for calculations and can be accessed here: [Statistics on qualitative results according to ISO 16140 2 amd 1.xlsx](#).

Table 4.3.1 Results and calculations for paired study

Matrices	PA	NA	ND _{FN(alt)}	PD	PD _{FP(alt)}	TNA	N	RT (%)	SE _{alt} (%)	SE _{ref} (%)	FPR (%)	FNR (%)
Equation			A			B	C	D	E	F	G	H
Food cat. 1												
Food cat. 2												
Food cat. 3												
Food cat. 4												
Food cat. 5												
TOTAL												

- A) Total negative deviation, TND = ND_{FN(alt)}
 B) Total negative agreement, TNA = NA + PD_{FP(alt)}
 C) Sum, N = PA + PD + TNA + TND
 D) Relative trueness, RT(%) = 100% x (PA+TNA)/N
 E) Sensitivity of the alternative method, SE_{alt}(%) = 100% x (PA+PD)/(PA+TND+PD)
 F) Sensitivity of the reference method, SE_{ref}(%) = 100% x (PA+TND)/(PA+TND+PD)
 G) False positive ratio, FPR(%) = 100% x PD_{FP(alt)}/TNA
 H) False negative ratio, FNR(%) = 100% x ND_{FN(alt)}/(PA+TND+PD)

PA, positive agreement; NA, negative agreement; ND_{FN(alt)}, negative deviation – the alternative method is false negative; PD, positive deviation; PD_{FP(alt)} false positive with alternative method;

Table 4.3.2 Results and calculations for unpaired study

Matrices	PA	NA	PA _{FP(alt)}	NA _{FN(alt)}	ND	ND _{FN(alt)}	PD	PD _{FP(alt)}	TND	TNA	(N)	RT (%)	SE _{alt} (%)	SE _{ref} (%)	FPR (%)	FNR (%)
Equation									A	B	C	D	E	F	G	H
Food cat. 1																
Food cat. 2																
Food cat. 3																
Food cat. 4																
Food cat. 5																
TOTAL																

- A) Total negative deviation, TND = ND + ND_{FN(alt)} + PA_{FP(alt)}
 B) Total negative agreement, TNA = NA + NA_{FN(alt)} + PD_{FP(alt)}
 C) Sum, N = PA + PD + TNA + TND
 D) Relative trueness, RT(%) = 100% x (PA+TNA)/N
 E) Sensitivity of the alternative method, SE_{alt}(%) = 100% x (PA+PD)/(PA+TND+PD)
 F) Sensitivity of the reference method, SE_{ref}(%) = 100% x (PA+TND)/(PA+TND+PD)
 G) False positive ratio, FPR(%) = 100% x (PA_{FP(alt)}+PD_{FP(alt)})/TNA
 H) False negative ratio, FNR(%) = 100% x (NA_{FN(alt)} + ND_{FN(alt)})/(PA+TND+PD)

PA, positive agreement; NA, negative agreement; PA_{FP(alt)} positive agreement due to false positive alternative-method result; ND_{FN(alt)}, negative deviation – the alternative method is false negative; PD, positive deviation; PD_{FP(alt)} false positive with alternative method;

4.1.1.5 Evaluation of the results - Acceptability requirements

Acceptability Limit (AL) for the sensitivity

The Acceptability Limits (ALs) are not based on data and consensus expert opinion. The interpretation of results shall be done per category and for all categories used in the validation study. The requirement is not met when the observed value is higher than the AL.

Calculate the difference between (TND - PD) for both paired and unpaired studies and the sum of (TND + PD) for paired studies. Check whether the difference and/or sum of PD and TND conform to the Acceptability Limit (AL) stated in Table 4.4 with respect to the type of study (**paired** or **unpaired**) and the number of categories used in the evaluation.

The interpretation of the results shall be done per category and for all categories used in the validation study. A sensitivity study can also exist of a partly paired and unpaired study. In that case the results for (TND + PD) shall be evaluated based on the number of positive samples obtained for the categories tested using the paired study design. The results for (TND - PD) shall be evaluated based on the number of positive samples obtained for the full study (so all categories belonging to both the paired and unpaired study design).

Table 4.4 Acceptability limit of the sensitivity for paired and unpaired study design in relation to the number of categories

Number of categories	Number of positive samples	Paired study		Unpaired study	Mixed study	
		(TND ^a - PD ^b)	(TND + PD)	(TND - PD)	(TND - PD)	(TND + PD)
1	30 to 59	3	6	3	3	6
2	60 to 89	4	8	4	4	8
3	90 to 119	5	10	5	5	10
4	120 to 149	5	12	5	5	12
5	150 to 179	5	14	5	5	14
6	180 to 209	6	16	6	6	16
7	210 to 239	6	18	7	7	18
8	240 to 269	6	20	7	7	20
9	270 to 299	7	22	8	8	22
10	300 to 329	7	24	8	8	24
11	330 to 359	7	26	9	9	26
12	360 to 389	8	28	9	9	28
13	390 to 419	8	30	10	10	30
14	420 to 449	8	32	10	10	32
15	450 to 479	9	34	11	11	34

If more categories see ISO 16140-2 Amd. 1

^a TND = total number of samples with Negative Deviation results.

^b PD = number of samples with Positive Deviation results.

^c Mixed study includes both paired and unpaired study design

Note:

- A negative value for (TND - PD) is acceptable as this indicates a better performance of the alternative method compared to the reference method.
- Information on differences observed between results of the alternative method before and after confirmation of the results according to the alternative-method protocol is commonly presented in the validation report as additional information but is not used in the overall assessment of the alternative-method performance.

The AL is not met when the observed value is higher than the AL. When the AL is not met, if the number of positive samples is higher than expected according to the number of categories (e.g. having 60 or more positive samples for one single category), it is possible to use the second column of Table 4.4 and switch to higher AL. When the AL is not met, investigations should be made (e.g. root cause analysis) in order to provide an explanation of the observed results. Based on the AL and the additional information, it is decided whether the alternative method is regarded as not fit for purpose for the category or categories involved. The reasons for acceptance of the alternative method in case the AL is not met shall be stated in the study report.

4.1.2 Level of Detection and Relative level of detection study

A comparative study is conducted to estimate the level of detection (LOD) of the alternative and to evaluate the level against the LOD of the reference method.

4.1.2.1 Selection of categories, number of samples and replicates tested

For the RLOD study:

A minimum of three levels per type shall be prepared consisting of at least a negative control level, a low level and a higher level. Ideally, the low level shall be the theoretical detection level (i.e. 0,7 cfu per test portion) and the higher level just above the theoretical detection level (e.g. 1 cfu to 1,5 cfu per test portion). For fastidious bacteria, the low level can significantly exceed the theoretical detection level; thus, the low and high level should be adjusted appropriately. A fixed ratio (e.g. 1:2) between the low- and high-level contamination should be used to aid in determining the final contamination levels. At least the low level should provide fractional recovery by either the reference method or the alternative method (fractional recovery at the low level should be between 25 % and 75 % of the number of samples tested). An evaluation shall be performed to ensure the relationship and consistency of the number of positives of the intermediate and high level. In cases where the alternative method produces fractional recovery at the low level and the reference method produces all positive results, the results of the RLOD study are not valid and a root cause analysis shall be performed.

The level of contamination of the sample used (except for the negative control) shall be determined by performing a most probable number (MPN) analysis on the (stabilized) inoculated samples (preferably) and/or through the enumeration of the inoculum at the time of inoculation.

At the negative control level, at least five replicate samples should be tested by both methods. For the second (low) level (theoretical detection level), at least 20, and for the third (higher) level, at least 5 replicate samples should be tested by both methods. The negative control level shall not produce positive (by isolation of the target organism) results. When positive results are obtained, the experiments have to be repeated for all levels.

For the LOD₅₀ study:

LOD₅₀ shall be calculated for each category for the alternative method and optionally for the reference method.

For each category evaluated, determine the contamination of the low level by performing a 3-level MPN for the particular (food) item tested using the reference method (preferably) at the time of the RLOD experiment and/or by the enumeration of the inoculum at the time of inoculation using a non-selective medium. When enumeration is performed, the inoculum for the low and high contamination levels should be determined. The MPN and 95 % confidence interval shall be determined for the fractional level only. For the low (or fractional) contamination level, analyse 20 test portions plus 5 test portions at approximately 2 times the test portion size and 5 test portions at approximately half of the test portion size evaluated in the validation study (e.g. if the reference method test portions were 25 g, evaluate 5 test portions at 50 g and 5 test portions at 10 g). To each test portion, add a proportionate amount of enrichment broth as described in the reference method to maintain the enrichment volume to mass ratio (e.g. a reference method with a 1:10 enrichment ratio, add 450 ml to the 50 g test portions and 90 ml to the 10 g test portions).

Analyse the test portions following the reference method from enrichment to confirmation.

Enumerate the inoculum at the time of inoculation by plating onto non-selective agar (see ISO 7218 for guidance on plating). Agar plates should be incubated under conditions to allow for optimal growth of the target microorganism.

Use the number of positive results per test portion size and the MPN value or the results of the enumeration of the inoculum to calculate the LOD50 and 95 % confidence interval.

NOTE The 20 test portions from the low level are the same as the 20 test portions used in the RLOD study. Therefore, only 5 test portions at 2 times and 5 at approximately half of the test portion size are analysed in addition.

4.1.2.2 Calculation and interpretation of LOD and the RLOD

An Excel®-based program for calculating LOD50 values is freely available for download at <https://www.statistics-wilrich.info/> (download the file PODLOD version 12). The LOD50 value is calculated per category tested in the RLOD study and shall be expressed as cfu/test portion.

The LOD is defined as LOD50 and is the level of detection for which 50 % of tests give a positive result. Transform the results into $\log_{10}$ and use the excel spreadsheet PODLOD available at the link below. It is only required to estimate the LOD50 for the alternative method. Tabulate the LOD50 Detection Limit and the Lower and Upper confidence limits for each matrix and for the combined results as illustrated in Table 4.5.

Table 4.5 Presentation of the LOD50 after confirmation of the alternative method results

Matrix	LOD50 log cfu / test portion	Lower confidence limit log cfu/ test portion	Upper confidence limit Log cfu/test portion
1			
2			
3			
4			
5			
Combined results			

The RLOD is defined as the ratio of the LODs of the alternative method and the reference method:

$$RLOD = LOD_{alt} / LOD_{ref}$$

The contamination levels are not needed for the calculations of the RLOD, since they are included in the model, resulting in curves in a graph of probability of detection versus log dose (contamination level). However, the known contamination levels are needed for the estimation of LOD50. Prof. Dr. Wilrich, Freie Universität Berlin has elaborated Excel spreadsheet for calculating LOD and RLOD.

Tabulate the results as indicated in **Table 4.6**.

Table 4.6 Presentation of RLOD before and after confirmation of the alternative-method results

	RLOD using the alternative- method results	RLOD using the confirmed alternative-method results
Item (category) (<i>i</i>)	$RLOD_i$	$RLOD_i$
1		
2		
...		
<i>K</i>		
Combined		

An acceptability limit (AL) for the RLOD, based on the confirmed alternative-method results, specifies the maximum increase in LOD of the alternative versus the reference method that would not be considered as relevant in consideration of the fitness for purpose of the method. Consequently, AL will be a value >1 . The interpretation should be made for each item.

For **paired** study data, the AL of RLOD is set at 1.5, meaning that the LOD for the alternative method may not be higher than 1.5 times the LOD of the reference method.

For **unpaired** study data, the AL of RLOD is set at 2.5, meaning that the LOD for the alternative method may not be higher than 2.5 times the LOD of the reference method.

A LOD value for the alternative method smaller than the LOD value for the reference method is always accepted, as this means that the alternative method is likely to detect lower levels of contamination than the reference method.

The AL is not met when the observed value is higher than the AL. When the AL is not met, investigations should be made (e.g. root cause analysis) in order to provide an explanation of the observed results. Based on the AL and the additional information, it is decided whether the alternative method is regarded as not fit for purpose for the item or category involved. The reasons for acceptance of the alternative method in case the AL is not met shall be stated in the study report.

4.1.3 Selectivity study

The selectivity is a measure of

- a) inclusivity: detection of the target microorganism from a wide range of strains, and
- b) exclusivity: the lack of interference from a relevant range of non-target microorganisms.

A range of strains shall be used. Each strain used shall be characterized biochemically and/or serologically and /or genetically in sufficient detail for its identity to be known. The original origin of all isolates should be known, and they should be held in a local (e.g. organising) national or international culture collection to be able to be used in future testing, if required.

The selectivity study is carried out on the alternative method (including a confirmation step if prescribed in the alternative-method protocol). When doubtful results are obtained, the test should be repeated and with the reference method included, checking that the strain could be detected with the appropriate reference method. Inoculation of a suitable growth medium is carried out with a dilution of a pure culture of each test strain. No sample is initially added.

For some microorganisms it will be difficult to obtain the required number of strains. The strains tested have to be approved by NordVal International.

4.1.3.1 Inclusivity

Select 50 pure cultures of target microorganisms relevant to the alternative method and the food matrices. For *Salmonella* methods, select at least 100 pure cultures of different serotypes of *Salmonella*.

From each test strain a growth medium is inoculated with a level of 10 to 100 times greater than the detection level. The analyses are carried out according to the respective method descriptions.

4.1.3.2 Exclusivity

Select 30 pure cultures of non-target microorganisms chosen from both the strains known to cause interference with the target microorganism and from strains naturally present in the food matrices.

From each test strain a growth medium is inoculated with a level similar to the greatest level of contamination expected to occur in the food matrices.

Table 4.7 Results of the selectivity study

Microorganisms	Alternative method	
	Test result	Confirmed* result
Inclusivity (target strains)		
1		
2		
Etc.		
Exclusivity (non-target strains)		
1		
2		
Etc.		

* Confirmed by reference method in case of doubtful result

4.2 Interlaboratory study, ILS

The aim of the ILS is to determine the difference in sensitivity between the reference and the alternative method when tested by different laboratories using identical samples (reproducibility conditions).

The interlaboratory study shall be conducted with collaborators belonging to more than one country. The collaborators shall be competent to perform both the reference method as well as the alternative method.

NOTE Competence can be demonstrated in different ways (e.g. for the reference method, a laboratory meeting the requirements of ISO/IEC 17025, and for the alternative method, a documented training).

4.2.1 Number of laboratories

At least 10 valid data sets from at least 10 collaborators (analysts) shall be included. The collaborators shall come from a minimum of 5 different organisations, but preferably 10 organisations, excluding the organising laboratory. (Laboratories in different locations, but belonging to one company or institute, are accepted as different organisations.) A maximum of 3 data sets can be produced by one organisation. Technicians, involved in the preparation of the samples used in the ILS, shall not take part in the testing of those samples.

4.2.2 Number of samples

One relevant food matrix is selected to prepare the test samples. Use at least 3 levels, a negative control (L_0), and two levels (L_1 and L_2). At least one of these should produce fractional positive results. The level of contamination needed to obtain fractional recovery shall be based on the RLOD study data of the reference method in the method comparison study. Theoretically, an average level of contamination of 1 cfu/sample is adequate to obtain fractional recovery. The level of contamination shall be determined. This allows calculation of the LOD_{50} of the alternative method. For instance the following levels could be used L_0 : negative control, L_1 : 1-5 cells per 25 g (according to the RLOD study) and L_2 : 5-50 cells per 25 g.

At least eight blind replicates at each level of contamination are analysed by each collaborator by both methods, so in total a minimum of 48 results (8 replicates x 3 levels x 2 methods) per collaborator.

If the alternative method comprises more than one method procedure a relevant food matrix for each procedure should be selected for the study.

The organising laboratory shall determine which data from the participating laboratories are suitable for the calculation of the precision data. Data from collaborators have to be discarded if the method procedures are not strictly followed, the sample/test kits have been damaged during transportation etc.

For paired study, if the result obtained by the reference method is negative but positive with the alternative method, additional confirmation is required for the alternative method as described in the protocol for the validation study.

For unpaired study, all results need confirmation for the alternative method as described in the protocol for the validation study.

4.2.3 Summary of data and trueness calculation

Tabulate the data obtained with each method as follows:

Table 4.8: Positive results by the reference method

Laboratories	Contamination level		
	L_0	L_1	L_2
Laboratory 1	$/8^a$	$/8^b$	$/8^c$
Laboratory 2	$/8$	$/8$	$/8$
Laboratory 3	$/8$	$/8$	$/8$
Etc. Etc	$/8$	$/8$	$/8$
Total	P_0	P_1	P_2
L_0 = negative control L_1 = 1-5 cells per 25 g L_2 = 5-50 cells per 25 g a Number of positive reference method results at level 0 b Number of positive reference method results at level 1 c Number of positive reference method results at level 2			

Table 4.9: Positive results (before and after confirmation) by the alternative method

Laboratories	Contamination level					
	L ₀		L ₁		L ₂	
	Screening	Confirmed	Screening	Confirmed	Screening	Confirmed
Laboratory 1	/8 ^a	/8 ^b	/8 ^c	/8 ^d	/8 ^e	/8 ^f
Laboratory 2	/8	/8	/8	/8	/8	/8
Laboratory 3	/8	/8	/8	/8	/8	/8
Etc. Etc	/8	/8	/8	/8	/8	/8
Total	P ₀	CP ₀	P ₁	CP ₁	P ₂	CP ₂

L₀ = negative control
 L₁ = 1-5 cells per 25 g
 L₂ = 5-50 cells per 25 g
 a Number of positive alternative method results at level 0
 b Number of confirmed alternative method results at level 0
 c Number of positive alternative method results at level 1
 d Number of confirmed alternative method results at level 1
 e Number of positive alternative method results at level 2
 f Number of confirmed alternative method results at level 2

Calculate the percentage of the specificity (%) of both the reference method and the alternative method, using the data after confirmation, based on the level L₀ as follows:

$$\text{Specificity for the reference method } SP_{\text{ref}} = \left[1 - \frac{P_0}{N_-} \right] \cdot 100\%$$

$$\text{Specificity for the alternative method } SP_{\text{alt}} = \left[1 - \frac{CP_0}{N_-} \right] \cdot 100\%$$

Where:

N₋ is the total number of all L₀ tests

P₀ is the number of false positive obtained by the reference method

CP₀ is the number of false positive obtained by the alternative method

For each positive contamination level L₁ + L₂, the results for all collaborators are combined as shown in Table 4.10 (1 and 2) and summed up as illustrated in Table 4.11.

Table 4.10.1 Summarised results for all collaborators for each sample, paired study

Reference method	Alternative method ^a	Confirmed alternative method ^b	Interpretation (based on the confirmed alternative-method result)
+	+	Not needed ^c	Positive agreement (PA)
-	-	Not needed ^c	Negative agreement (NA)
+	-	Not needed ^c	Negative Deviation (ND) due to false negative alternative-method result (ND _{FN(alt)})
-	+	+	Positive Deviation (PD)
-	+	-	Positive Deviation due to false positive alternative-method result (PD _{FP(alt)}) ^d

^a The alternative-method results includes any confirmations as described in the alternative-method protocol.

^b The confirmed alternative-method result is the result after additional confirmation as described in the protocol for the validation study.

^c No need for additional confirmation test(s). Confirmed alternative-method result is the same as the alternative-method result.

^d This false positive result (FP) shall also be used to calculate the false positive ratio.

Table 4.10.2 Summarised results for all collaborators for each sample, unpaired study

Result of each sample			
Reference method	Alternative method ^a	Confirmed Alternative method ^b	Interpretation (based on the confirmed alternative-method result)
+	+	+	Positive Agreement (PA)
+	+	-	Positive Agreement due to false positive alternative-method result ($PA_{FP(alt)}$) ^c
-	-	-	Negative Agreement (NA)
-	-	+	Negative Agreement due to false negative alternative-method result ($NA_{FN(alt)}$)
+	-	-	Negative deviation (ND)
+	-	+	Negative Deviation due to false negative alternative-method result ($ND_{FN(alt)}$)
-	+	+	Positive deviation (PD)
-	+	-	Positive results for alternative method, turns out negative after confirmation ($PD_{FP(alt)}$) ^c

^a The alternative-method result includes any confirmations as described in the alternative-method protocol.

^b The confirmed alternative-method result is the result after additional confirmation as described in the protocol for the validation study

Determine the Total Negative Deviation (TND) and Total Negative Agreement (TNA) for the validation study.

Paired study: Total Negative Deviation: $TND = ND$
Total Negative Agreement: $TNA = NA + PD_{FP(alt)}$

Unpaired study: Total Negative Deviation: $TND = ND + ND_{FN(alt)} + PA_{FP(alt)}$
Total Negative Agreement: $TNA = NA + NA_{FN(alt)} + PD_{FP(alt)}$

For each level of contamination and the totality of the results, compare the alternative method and the reference method in order to calculate the relative accuracy in the study carried out by the organising laboratory. Each pair of results from a sample measured by the alternative and the reference method shall be reported in a cross table as table 4.11.

Table 4.11: Summary of results for all collaborators obtained with the reference and alternative methods for Level L1 or L2.

Alternative method	Reference method	
	R+	R-
A +	R+ / A+ (PA)	R- / A+ (PD)
A -	R+ / A- (TND)	R- / A- (TNA)

Based on data summarized in Table 4.11, calculate the sensitivity of the two methods as well as the relative trueness and false positive ratio and the false negative method for the alternative method as follows

$$\text{Sensitivity of the alternative method: } SE_{alt} = \frac{(PA+PD)}{(PA+TND+PD)} \cdot 100\%$$

$$\text{Sensitivity of the reference method: } SE_{ref} = \frac{(PA+TND)}{(PA+TND+PD)} \cdot 100\%$$

$$\text{Relative trueness: } RT = \frac{(PA+TNA)}{N} \cdot 100\%$$

False positive ratio (FPR) and false negative ratio (FNR) for the alternative method:

$$\text{For paired study, false positive ratio: } FPR = \frac{PD_{FP(alt)}}{TNA} \cdot 100\%$$

$$\text{For paired study, false negative ratio: } FNR = \frac{ND_{FN(alt)}}{PA+TND+PD} \cdot 100\%$$

$$\text{For unpaired study, false positive ratio: } FPR = \frac{(PA_{FP(alt)} + PD_{FP(alt)})}{TNA} \cdot 100\%$$

$$\text{For unpaired study, false negative ratio: } FNR = \frac{(NA_{FN(alt)} + ND_{FN(alt)})}{PA+TND+PD} \cdot 100\%$$

Where:

N is the number of tested samples (TNA + PA + PD + TND),

FP is false positive results explained and

FN is the false negative results

For explanation of the abbreviated terms see Table 4.10.1 and Table 4.10.2.

The confirmed alternative-method results shall be used to determine whether the alternative method produces comparable results to the reference method.

An Excel sheet is available for calculations and can be accessed here: [Statistics on qualitative results according to ISO 16140 2 amd 1.xlsx](#).

4.2.4 Interpretation of data

4.2.4.1 Paired study

For a paired study, calculate the difference between (TND – PD) and the sum of (TND + PD) for the level(s) where fractional recovery was obtained (so L1 and possibly L2). The values found for (TND – PD) and (TND + PD) shall not be higher than the Acceptability Limits (ALs) given in Table 12 with respect to the number of participating laboratories (N_{lab})

Table 4.12 Acceptability limits for a paired study design in relation to the number of collaborating laboratories

Number of laboratories	TND-PD	TND+PD
10	3	4
11	4	4
12-13	4	5
14-16	4	6
17	4	7
18	5	7
19-20	5	8

The AL is not met when the observed value is higher than the AL. When the AL is not met, investigations should be made (e.g. root cause analysis) in order to provide an explanation of the observed results. Based on the AL and the additional information, it is decided whether the alternative method is regarded as not fit for purpose. The reasons for acceptance of the alternative method in case the AL is not met shall be stated in the study report.

4.2.4.2 Unpaired study

For an unpaired study, calculate the difference between (TND-PD) for the level(s) where fractional recovery was obtained (i.e. L1 and possibly L2). The observed value found for (TND-PD) shall not be higher than AL. The AL is defined as $(\text{TND-PD})_{\max}$ and calculated per level, where fractional recovery was obtained, as follow:

$$(\text{TND-PD})_{\max} \leq \sqrt{3N_{x(\text{ref})} \cdot ((p+)_{\text{ref}} + (p+)_{\text{alt}} - 2 \cdot ((p+)_{\text{ref}} \cdot (p+)_{\text{alt}}))}$$

where

$$(p+)_{\text{ref}} = \frac{P_{x(\text{ref})}}{N_{x(\text{ref})}} \quad \text{and} \quad (p+)_{\text{alt}} = \frac{CP_{x(\text{alt})}}{N_{x(\text{alt})}}$$

$P_{x(\text{ref})}$ is number of samples with a positive result obtained with the reference method at level x (L1 or L2) for all laboratories;

$N_{x(\text{ref})}$ is number of samples tested at level x (L1 or L2) with the reference method by all laboratories.

$CP_{x(\text{alt})}$ is the number of confirmed positive results obtained with the alternative method for the specific level by all laboratories

$N_{x(\text{alt})}$ is the number of samples for the specific levels analysed with the alternative method by all laboratories

The AL is not met when the observed value is higher than the AL. Based on the AL and the additional information, it is decided whether the alternative method is regarded as not fit for purpose for the category or categories involved.

4.2.4.3 Calculation of the relative level of detection and LOD

For both a paired and unpaired study, the variation of the relative levels of detection (RLOD) between laboratories shall be assessed. Conduct the calculations in accordance with [PODLOD-interlab statistical programs by Wilrich & Wilrich](#). The ALs for the RLOD of paired and unpaired studies are found in 4.1.4.2.

In addition, the data can be evaluated using the probability of detection (POD) model described in J. AOAC Int. 2011, 94 pp. 335–337. The evaluation using the POD model can give additional information on the equivalence of the methods.

4.2.4.4 Calculation of the LOD50

LOD50 shall be calculated for each category used in the interlaboratory study for the alternative method and optionally for the reference method. The MPN and 95 % confidence interval shall be determined for the fractional level only. Determine for the L1 level the contamination level by performing a 3-level MPN for the particular (food) item tested in the interlaboratory study using the reference method. For the low (or fractional) contamination level, analyse 20 test portions plus 5 test portions at approximately 2 times the test portion size and 5 test portions at approximately half of the test portion size evaluated in the validation study (e.g. if the reference method test portions were 25 g, evaluate 5 test portions at 50 g and 5 test portions at 10 g). To each test portion, add a proportionate amount of enrichment broth as described in the reference method to maintain the enrichment volume to mass ratio (e.g. a reference method with a 1:10 enrichment ratio, add 450 ml to the 50 g test portions and 90 ml to the 10 g test portions). Analyse the test portions following the reference method from enrichment to confirmation. Use the number of positive results per test portion size to calculate the MPN value. An Excel®- based program for calculating MPN values is freely available for download at <https://standards.iso.org/iso/7218> (download the file “MPN calculation Excel program”)

Use the number of positive results per test portion size from the alternative method and the MPN value

obtained using the reference method to calculate the LOD₅₀ of the alternative method and 95 % confidence interval. An Excel®-based program for calculating LOD₅₀ values is freely available for download at <https://standards.iso.org/iso/16140/-2/ed-1/en/amd/1/> and at [Statistical programs • Wilrich](#) (download-PODLOD interlab). The LOD₅₀ value shall be expressed as cfu/test portion.

5. Quantitative methods

5.1 Method comparison study

The method comparison study is performed in the organising laboratory and consists of four parts.

- A comparative study of the results of the reference method to the results of the alternative method, in a variety of different items (naturally and/or artificially) contaminated samples (so-called relative trueness study).
- A comparative study of the results of the reference method to the results of the alternative method in artificially contaminated samples, using replicates of a single item per category. The data are analysed using the Accuracy Profile, AP (so-called AP study).
- A limit of quantification, LOQ, study of the results of the alternative method in artificially contaminated samples using replicates of a single item per category. The data are used to calculate the LOQ of the alternative method. This study is only done for instrumental-like methods (i.e. methods that are not based on the counting of individual colonies).
- A selectivity (inclusivity/exclusivity) study of the alternative method.

The organizing laboratory shall be competent to perform both the reference method as well as the alternative method.

NOTE: Competence can be demonstrated in different ways (e.g. for the reference method, a laboratory meeting the requirements of ISO/IEC 17025, and for the alternative method, a documented training).

5.1.1 Relative trueness study

The relative trueness study is a comparative study between the results obtained by the reference method and the results of the alternative method. This study is conducted using naturally and/or artificially contaminated samples of different categories, types and items.

5.1.1.1 Selection of matrices and number of samples

The selection of categories and types depend on the type or group of microorganism and the scope of the validation. If the method is to be applied for a broad range of foods, choose at least 5 categories of food. If the method is to be validated for a restricted number of food categories, e.g. 'ready-to-eat, ready-to-reheat meat products', and 'heat-processed milk and dairy products', then only these categories should be studied. In addition to food categories, feed samples, environmental samples and primary production stage samples can be included as additional categories.

For all selected categories (food and others), include at least 3 different types per category. For each type analyse at least 5 samples, this results in a minimum of 15 samples per category. Relevant types and categories per type of microorganism that might be relevant for the validation is given in separate documents at the NordVal webpage as guidance, not as mandatory.

Use naturally contaminated samples. If it is not possible to acquire a sufficient number of naturally contaminated samples, use artificial contamination of samples. Select different types with both high and low (natural) background microflora, different types of stresses due to processing and raw (unprocessed) items. For stressing see chapter 4.1.1.2.

The samples should be contaminated at a level that is representative for the natural variation in level of contamination. All samples combined should cover the range of concentration normally observed for the type of microorganism used.

Some naturally contaminated samples may contain high numbers of target analyte, and this can result in difficulty in achieving the required range of contamination. In such cases, the naturally contaminated sample can be 'diluted' with uncontaminated material of the same item.

The reference and alternative methods shall be performed with, as far as possible, exactly the same sample.

5.1.1.2 Calculation and interpretation of relative trueness study

The results (in $\log_{10}$) obtained by the reference method and the alternative method can be illustrated in a scatter plot. The results obtained by the reference method are plotted on the x-axis against the results of the alternative on the y-axis. Plot the data for each sample per category and for each sample in all categories and draw the line of identity on which all points would lie if the two methods gave identical results for each sample analysed.

Figure 5.1 presents an example for a single category. The plot for a single category should show the results of each type tested with a distinct symbol. The plot for all categories should show the results of each category tested with a distinct symbol. This provides a rapid visual assessment of the extent to which the two methods (do not) agree. If any result (either reference or alternative method) is below the quantification limit, the data should be plotted using a substituted value of 1 $\log_{10}$ unit less than the observed value in case of a lower than value (e.g. 6 $\log_{10}$ units will be amended to 7 $\log_{10}$ units). For example, a reference method result of 2.54 $\log_{10}$ cfu/g and an alternative-method result < 2 $\log_{10}$ cfu/g will be plotted in the graph as 2.54 for the reference method and 1.0 for the alternative method. Information (based on investigation, e.g. root cause analysis) can be given in the study report in order to provide an explanation of the findings when one of the methods gave a result below the quantification limit.

These amended values shall not be used for the calculation of the data presented in Table 5.1 but shall be included in the scatter plot (see **Figure 5.1**).

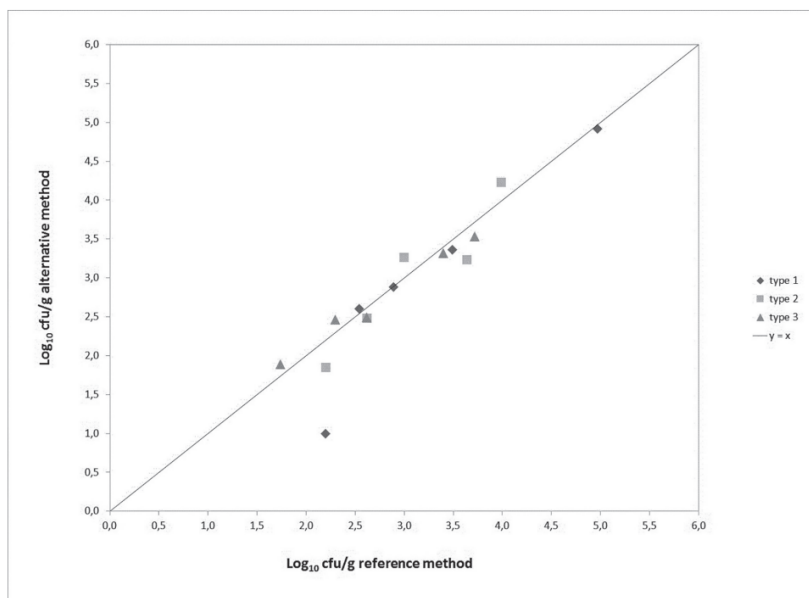


Figure 5.1 Scatter plot of reference method versus alternative method results for a single category

Determine the average of each pair of data values and the difference between the values as in 5.1 and plot these differences against the corresponding averages per category and for all categories to illustrate the degree of agreement between the reference method and the alternative method. **Figure 5.2** shows the line of identity (zero difference), the line of bias (average difference) as well as the upper and lower 95 % prediction limits of the individual sample-specific bias values. These limits of the prediction range are called “limits of agreement”.

Table 5.1 Summarised results for all categories

Category	Type	Sample	Log10 cfu		Mean	Difference (bias)
			Reference-method result	Alternative-method result		
1	1	1	R1	A1	$(R1 + A1)/2$	$D1 = A1 - R1$
		2	R2	A2	$(R2 + A2)/2$	$D2 = A2 - R2$
		3	R3	A3	$(R3 + A3)/2$	$D3 = A3 - R3$
		4	R4	A4	$(R4 + A4)/2$	$D4 = A4 - R4$
		5	R5	A5	$(R5 + A5)/2$	$D5 = A5 - R5$
1	2	6	R6	A6	$(R6 + A6)/2$	$D6 = A6 - R6$
		7	R7	A7	$(R7 + A7)/2$	$D7 = A7 - R7$
		8	R8	A8	$(R8 + A8)/2$	$D8 = A8 - R8$
		9	R9	A9	$(R9 + A9)/2$	$D9 = A9 - R9$
		10	R10	A10	$(R10 + A10)/2$	$D10 = A10 - R10$
1	3	...	...	...	...	...
Average category 1				D_1		
Standard deviation category 1				s_{D1}		
...	...	...	...	...	...	...
x4			Rx	Ax	$(Rx + Ax)/2$	$Dx = Ax - Rx$
Average category x				D_x		
Standard deviation category x				s_{Dx}		
Average all categories				D_{all}		
Standard deviation all categories				s_{Dall}		

Compute per category and for all categories the average difference D , the standard deviation of differences sD and the limits of agreement using Formula $D \pm 1.96 \cdot sD$

Plot as in Figure 5.2 (Example 1) the individual sample differences against the mean values on a graph that shows the line of identity (zero difference), the line of bias (average difference) as well as and the upper and lower 95 % prediction limits (limits of agreement) of the individual sample-specific bias values (both separately per category and across all categories). This illustrates the degree of bias and the (lack of) agreement of the data.

Example 1: Estimation of relative trueness – Bland Altman Plot

For a broad range of foods the number of test portion analysed by each method should be: 5 categories x 3 types per category x 5 replicates = 75 samples. In this example 2 categories are chosen, and the number of samples for each method would be: 2 categories x 3 types per category x 5 replicates each type = 30 samples.

Table 5.2: Results obtained in a relative trueness study, including 2 categories. The results are given in log₁₀ cfu/g

Category	Type	Replicates	Sample No.	Ref method	Alt method	Mean	Difference
1	1	1	1	2.00	2.38	2.19	0.38
		2	2	3.18	3.04	3.11	-0.13
		3	3	3.04	3.15	3.09	0.10
		4	4	3.66	3.73	3.70	0.07
		5	5	3.58	3.48	3.53	-0.10
	2	1	6	3.76	3.84	3.80	0.08
		2	7	5.10	5.17	5.13	0.07
		3	8	4.68	4.78	4.73	0.10
		4	9	4.74	4.83	4.79	0.09
		5	10	4.88	4.94	4.91	0.06
	3	1	11	3.48	3.26	3.37	-0.22
		2	12	2.90	3.11	3.01	0.21
		3	13	4.23	4.20	4.22	-0.03
		4	14	3.83	3.90	3.87	0.07
		5	15	4.38	4.54	4.46	0.16
Average of difference for category 1							0.06
Standard deviation of the difference for category 1							0.14
Category	Type	Replicates	Sample No.	Ref method	Alt method	Mean	Difference
2	1	1	1	3.04	3.09	3.07	0.05
		2	2	3.88	3.96	3.92	0.08
		3	3	3.96	4.29	4.12	0.33
		4	4	4.16	4.26	4.21	0.10
		5	5	4.38	4.39	4.38	0.01
	2	1	6	4.20	4.17	4.19	-0.03
		2	7	4.18	4.26	4.22	0.08
		3	8	5.62	5.71	5.66	0.09
		4	9	3.70	3.30	3.50	-0.40
		5	10	5.25	5.22	5.23	-0.02
	3	1	11	4.11	4.26	4.18	0.14
		2	12	3.26	3.46	3.36	0.21
		3	13	4.20	4.24	4.22	0.04
		4	14	2.83	2.78	2.81	-0.05
		5	15	3.38	3.46	3.42	0.08
Average of difference for category 2							0.05
Standard deviation of the difference for category 2							0.16
Average of the difference for all categories, D							0.05
Standard deviation of the difference for all categories, SD							0.15
Upper limit, 95% confidence = $D + 1.96 \cdot SD$							0.34
Lower limit, 95% confidence = $D - 1.96 \cdot SD$							-0.24

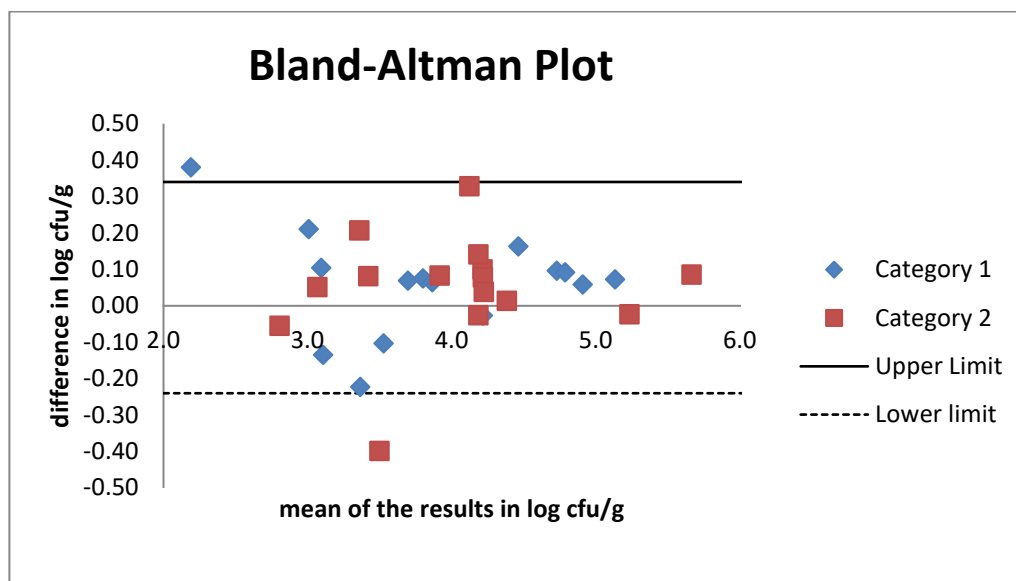


Figure 5.2 Bland-Altman difference plot for all categories

The results of the difference and scatter plot shall be interpreted based on a visual observation of the overall bias, the spread of the individual sample-specific bias values, and any bias values lying outside the limits of agreement. If the individual bias values are normally distributed, it can be expected that 1 in 20 values will lie outside the limits of agreement. Discrepancies can be seen as an indication that the individual bias values do not follow a normal distribution, e.g. due to the presence of outliers. Any deviations from the expectations should be documented.

5.1.2 Accuracy profile study

The accuracy profile study is a comparative study between the results obtained by the reference method and the results of the alternative method. This study is conducted using artificially contaminated samples. One type per category will be tested.

5.1.2.1 Selection of matrices and number of samples

The selection of categories and types depend on the type or group of microorganisms and the scope of the validation. If the method is to be applied for a broad range of foods, choose at least 5 categories of food.

For each category being examined, test at least 1 type, using 6 samples per type. Of the 6 samples, there should be 2 at a low level, 2 at an intermediate level and 2 at a high level of contamination. These levels should cover the whole range of contamination of the selected type. Of each sample, use 5 replicates, representing 5 different test portions from the same sample, shall be used.

NOTE: The 6 samples might be different belonging to the same type, but not necessarily the same item. For example, one sample might be full fat milk powder, another infant formula; belonging to the same food type (dried milk products) but not the same food item.

5.1.2.2 Calculation and interpretation of accuracy profile study

For each type of food category, calculate the median and the standard deviation, SD, for the 5 replicates for all the 6 levels obtained by both the reference method and the alternative method. Further calculate the combined standard deviation, the bias and the upper and lower limits as illustrated in the Example 2 below. The excel spreadsheet is available in the end of chapter 5.

Example 2: Accuracy profile

In this example, 1 specific type of a food category is chosen. Then 6 levels, each with 5 replicates are analysed by both the reference and the alternative method. The results obtained by the reference method are given in Table 5.3, the results obtained by the alternative method are given Table 5.4 and a summary of the results and the values needed for the accuracy profile (Figure 5.2) is given in Table 5.5.

Table 5.3 Results of the replicates for the different levels obtained by the reference method.

Levels	Replicates Reference Method, levels in $\log_{10}/g$					Median	$SD=\sqrt{1/(n-1) \cdot \sum(x-\text{mean})^2}$
	1	2	3	4	5		
1 low	2.00	2.04	2.48	2.34	<2	2.18	0.587
2 low	2.42	2.36	2.56	2.41	2.59	2.42	0.099
3 intermediate	3.54	3.70	3.63	3.69	3.69	3.69	0.066
4 intermediate	3.65	3.43	3.69	3.61	3.71	3.65	0.111
5 high	5.40	5.45	5.23	5.41	5.24	5.40	0.103
6 high	5.41	5.32	5.59	5.40	5.52	5.41	0.107
Combined standard deviation, $SD_{REF} = \sqrt{1/6 \sum SD_i^2}$							0.256

Table 5.4 Results of the replicates for the different levels obtained by the alternative method.

Levels	Replicates Alternative Method, levels in $\log_{10}/g$					Median	$SD=\sqrt{1/(n-1) \cdot \sum(x-\text{mean})^2}$
	1	2	3	4	5		
1 low	2.26	2.18	2.00	2.46	2.54	2.26	0.219
2 low	2.30	2.40	2.34	2.28	2.65	2.34	0.151
3 intermediate	3.71	3.62	3.75	3.65	3.76	3.71	0.060
4 intermediate	3.61	3.26	3.61	3.32	3.43	3.43	0.164
5 high	5.50	5.53	5.46	5.48	5.48	5.48	0.026
6 high	5.28	5.30	5.46	5.36	5.40	5.36	0.073
Combined standard deviation, $s_{alt} SD_{ALT} = \sqrt{1/6 \sum SD_i^2}$							0.134

Calculate the bias (the difference between the medians of the replicates obtained by the alternative method and the reference method). Further, compute the interval where the expected portion of future results will fall, upper level, U, and lower level, L, (in 16140-2 named upper and lower β -ETI). This interval is calculated as

$$\text{Bias} \pm T \cdot s_{alt} \sqrt{1+1/n},$$

where

- bias is the difference between the obtained medians
- T is a factor from the student t- distribution for a chosen probability (80%) and degrees of freedom (here 24), $T = T(1/2(1-0.8), q(n-1))$; where q=number of levels and n=number of replicates, $T = T(0.1; 6(5-1)) = T(0.1; 24)$ so for the validations $T = 1.71$ (formula in the excel spreadsheet is: = TINV (0.1;24))
- s_{alt} is the combined standard deviation of results obtained by the alternative method
- n is the number of replicates = 5

The expected level will then be: Bias (for each level) $\pm s_{alt} \cdot 1.87$

In this example the $s_{alt} = 0.134$, hence the expected levels would be Bias (for each level) ± 0.25

Note: As the calculations are carried out in excel spreadsheet there might be some rounding issues.

The obtained results are satisfactory if the upper and lower levels do not cross the upper and lower acceptability limits, AL, of ± 0.5 log units.

For each category, summarise the results as shown in Table 5.5 and illustrated graphically in Figure 5.2

Table 5.5 Accuracy profile for one category

Levels	Median Ref	SD	Median Alt	SD	Bias	Upper level, U	Lower level, L	Upper AL	Lower AL
1 low	2.18	0.587	2.26	0.219	0.08	0.33	-0.18	0.5	-0.5
2 low	2.42	0.099	2.34	0.151	-0.08	0.17	-0.33	0.5	-0.5
3 intermed	3.69	0.066	3.71	0.060	0.02	0.27	-0.23	0.5	-0.5
4 intermed	3.65	0.111	3.43	0.164	-0.22	0.03	-0.47	0.5	-0.5
5 high	5.40	0.103	5.48	0.026	0.08	0.33	-0.17	0.5	-0.5
6 high	5.41	0.107	5.36	0.073	-0.05	0.20	-0.30	0.5	-0.5
Combined SD		0.256		0.134					

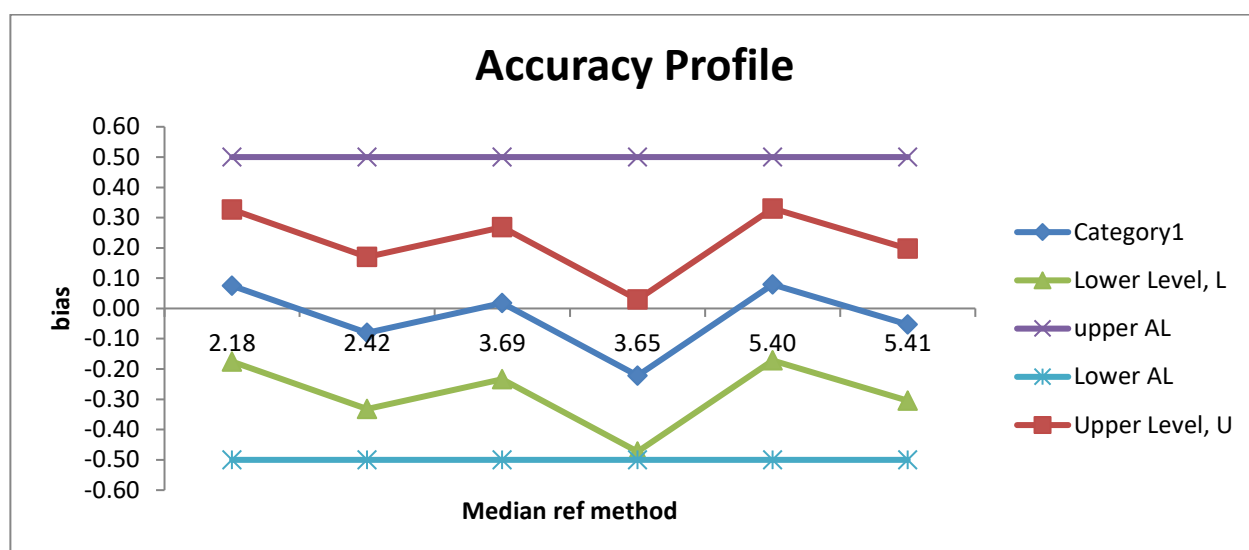


Figure 5.2 Accuracy profile for one category

Whenever no biases exist, these values for the category are located on the horizontal reference line. In the figure above, the acceptability limits (AL) are represented by the purple and the light blue lines. The levels where the results might be expected to vary between (upper and lower levels) are given as red and green lines. The bias (the difference obtained by the results obtained by the alternative method and the reference method) is given as the blue line.

If the upper level (red line) is below the upper AL (purple line), and the lower level (green line) is above the lower AL (light blue line) the alternative method is accepted as being equivalent to the reference method.

Accuracy profile is conducted for all categories.

If any of the upper or lower levels exceeds the Acceptability Limits, AL, and the standard deviation of the reference method, $s_{ref} > 0.125$ but ≤ 0.25 , then calculate new Acceptability Limits as $AL_s = 4 \cdot s_{ref}$. In cases of $s_{ref} > 0.25$, no recalculation is allowed. In this case, investigations should be made (e.g. root cause analysis) to provide an explanation of the high s_{ref} , correct possible issues and, as an option, repeat the experiment.

If the $U \leq AL_s$ and $L \geq AL_s$ for all replicates, the alternative method is accepted as being equivalent to reference method.

For the Accuracy Profile calculations, an Excel® spreadsheet is available for download here:

<https://standards.iso.org/iso/16140/-2/ed-1/en/>

5.1.3 Limit of quantification study

5.1.3.1 General considerations

For some alternative methods, it is of interest to determine the limit of quantification (LOQ). The LOQ is only relevant when the measurement principle of the alternative method is not based on counting visible colonies of the target microorganism and shall therefore be determined in these cases. Examples of methods for which the LOQ needs to be determined, are the instrumental measurement of conductivity or fluorescence, which is related to the growth of the microorganism.

5.1.3.2 Selection of matrices and number of samples

Select the same categories and types as used for the accuracy profile study.

If the LOQ needs to be determined, blank samples are tested per type/category used. These blank samples are used to verify the limit of quantification of the alternative method. A minimum of 10 test portions from the same sample shall be used. Examine the test portions with the alternative method.

5.1.3.3 Calculation and interpretation of limit of quantification study

The 10 results per type/category are used to estimate the baseline or threshold standard deviation s_0 . Calculate the standard deviation s_0 of the n results as follows:

$$s_0 = \frac{1}{n-1} \sum_{j=1}^n (y_j - \bar{y})^2$$

where

n is the total number of test portions used;

y_j is the $\log_{10}$ transformed result of test portion j ;

$\bar{y}$ is the average $\log_{10}$ transformed result of all test portions.

The limit of quantification is calculated as $LOQ = 10 s_0$.

5.1.4 Inclusivity and exclusivity study

Inclusivity and exclusivity testing is not required for general enumeration methods such as total plate count (TPC) and yeast & mould (Y&M) methods. It is required for enumeration methods designed for specific microorganisms (e.g. *Listeria*).

5.1.4.1 Selection and number of test strains

A range of strains should be used. The strains used should take into account the measurement principle of the alternative method (e.g. culture-based, immunoassay-based, and molecular). Different measurement principles may require the use of different test panels of strains.

Each strain used should be characterized biochemically and/or serologically and/or genetically in sufficient detail for its identity to be known. Strains used should preferentially have been isolated from foods, feeds or the food-processing environment, or primary production taking into account the scope of the validation. However, clinical, environmental and culture collection strains can also be used. The original source of all isolates should be known and they should be held in a local (e.g. an organising laboratory), national or international culture collection to enable them to be used in future testing if required.

For inclusivity testing, at least 50 pure cultures of (target) microorganisms should be tested.

For exclusivity testing, at least 30 pure cultures of (non-target) microorganisms should be tested.

NOTE 1: Some microorganisms will be difficult or impossible to culture, like viruses or protozoan parasites. Where the target organism cannot be cultured, pure suspensions of the test strains should be used for spiking at the earliest appropriate step of the method.

NOTE 2: For some microorganisms it will be difficult to obtain the required number of strains for inclusivity and exclusivity. In these cases an agreed set of test strains should be selected by the parties involved in the validation study.

5.1.4.2 Target microorganisms (inclusivity)

Each test is performed once and with the alternative method, the reference method and a non-selective agar. The inoculum level should be at least 100 times greater than the minimum level for quantification of the alternative method being validated. When using a plate method as the alternative method, the inoculum level shall obtain a countable number on the plate. If results are negative, consideration could be given to repeat the test with the addition of a food item.

5.1.4.3 Non-target microorganisms (exclusivity)

Each test is performed once and with the alternative and the reference method. The inoculum level should be similar to the greatest level of contamination expected to occur in any of the categories being used. No sample is added. The pure culture should be grown in a suitable non-selective broth under optimal conditions of growth for at least 24 h and diluted to an appropriate level before testing is begun.

NOTE: If the organism cannot be cultured a stock suspension should be diluted to an appropriate level before use.

5.1.4.4 Expression and interpretation of results

Tabulate the results as in Table 5.5 for the inclusivity tests and Table 5.6 for the exclusivity tests. The interpretation shall be done by the laboratory in charge of the method comparison study. The report should state any anomalies from the expected results.

Table 5.5 Presentation of results for inclusivity

Microorganisms	Reference method	Alternative method	Non-selective agar
1			
2			
Etc.			

NOTE: The interpretation of the inclusivity data for an alternative method using a plating medium is made on a qualitative basis. However, quantitative data should facilitate the interpretation of the data.

Table 5.6 Presentation of results for exclusivity

Microorganisms	Reference method	Alternative method
1		
2		
Etc.		

5.2 Interlaboratory study, ILS

5.2.1 Number of laboratories and samples

The aim of the ILS is to compare the performance of the alternative method to the reference method by different collaborators, using identical samples examined under reproducibility conditions. Further, to compare these results with pre-set criteria for the acceptable difference between the reference method and the alternative method. Wherever possible the study conditions should reflect the normal variation between laboratories. The ILS is organised by the organising laboratory.

The ILS shall produce at least 8 valid data sets from at least 8 collaborators. The collaborators shall come from a minimum of 4 different organizations, but preferably 8 organizations, excluding the organizing laboratory. (Laboratories in different locations, but belonging to one company or institute, are accepted as different organizations.) A maximum of 3 data sets can be produced by one organization. Technicians, involved in the preparation of the samples used in the ILS, shall not take part in the testing of those samples within the ILS.

The ILS shall be conducted with collaborators belonging to more than one country. The collaborators shall be competent to perform both the reference method as well as the alternative method.

NOTE: Competence can be demonstrated in different ways (e.g. for the reference method, a laboratory meeting the requirements of ISO /IEC 17025, and for the alternative method, a documented training).

The accuracy and precision estimates should be calculated from a large number of duplicate test results. Minimum 8 collaborators, 3 levels of contamination, 2 methods of enumeration (reference and alternative) and duplicate measurements, i.e. $8 \times 3 \times 2 \times 2 = 96$ results.

The organizer is responsible for the preparation of the test protocol and a data sheet for the recording of all experimental data and critical experimental conditions used by each laboratory. Prior to the study, it is a good idea to send a test sample to each collaborator for checking their competence in the methods.

The samples in the ILS should be

- relevant matrix containing a natural background microflora
- inoculated with the target organism. The protocol for inoculation of the samples shall be appropriate for the selected item. Samples shall be prepared and tested by the organising laboratory ensuring homogeneity
- 3 different levels of contamination lower, middle and upper levels of the entire concentration range of the alternative method + a negative control level
- 2 replicates each level; all samples should be blind coded to ensure that the analysts are not aware of their level of contamination
- analysed by each laboratory at the stipulated date.

The organising laboratory is reviewing the recorded data. No outlier tests are performed on the selected data.

5.2.2 Calculations, summary and interpretation of data

Use $\log_{10}$ transformed results obtained from the different collaborators for both the reference and alternative method. For each level, calculate the mean, the repeatability (s_r), the between series variation (S_L^2), the reproducibility (S_R), the bias and the expected levels as shown in the Example 3.

Example 3: Results from ILS

8 laboratories analysed 2 replicates (A and B) on 3 different levels on one type of sample by both the alternative method and the reference method. The results are collected in 3 different tables; one for each level. Some of the calculations are also shown in the tables 5.7- 5.9.

Table 5.7 Results ILS, low level

Collaborators	Low level log ₁₀ cfu/g									
	Reference method					Alternative method				
	Results Replicates		Calculation			Results Replicates		Calculation		
	A	B	(A-B) ²	Mean X	X ²	A	B	(A-B) ²	Mean .y	y ²
1	2.48	2.41	0.0049	2.45	5.9780	2.08	2.34	0.0676	2.21	4.884
2	2.40	2.32	0.0064	2.36	5.5696	2.41	2	0.1681	2.21	4.862
3	2.26	2.20	0.0036	2.23	4.9729	2.04	2.18	0.0196	2.11	4.452
4	2.18	2.15	0.0009	2.17	4.6872	2.34	2.32	0.0004	2.33	5.429
5	2.28	2.21	0.0049	2.25	5.0400	2.18	2.11	0.0049	2.15	4.601
6	2.15	2.28	0.0169	2.22	4.9062	2.2	2.15	0.0025	2.18	4.731
7	2.26	2.38	0.0144	2.32	5.3824	2.2	2.2	0	2.20	4.840
8	2.11	2.18	0.0049	2.15	4.6010	2.15	2.34	0.0361	2.25	5.040
Sum			0.0569	18.13	41.14			0.2992	17.62	38.87
Mean	2.27					2.20				
S _r	0.060	=√(1/16·0,0569)				0.137	=√(1/16·0.2992)			
s _L ²	0.009	= [(n·Σ(X) ² - (ΣX) ²)/n·(n-1)] - s _r ² /2 = [(8·41.14 - 18.13 ²)/(8·7)] - 0.060 ² /2				0	= [(8·38.87 - 17.62 ²)/(8·7)] - 0.299 ² /2 = -0.005 If the value is < 0, the s _L ² is sat to 0			
S _R	0.110	=√(s _r ² + s _L ²) = √(0.06 ² + 0.009)				0.137	=√(0.137 ² + 0 ²)			
Bias						-0.07				

The results should be expressed as $y \pm k \cdot s_R$, where y is the mean value of the alternative method and k is a constant (= 1.41 when using $T(0.2; n-1)$ n is number of (here the degree of freedom is $8 - 1 = 7$) and s_R is the standard deviation of reproducibility. [In ISO 16140-2 the degree of freedom is calculated more exactly, however, the results will be similar when rounding the figures into relevant figures¹. The example given in ISO 16140-2 is repeated in NordVal Excel Spreadsheet with the “short cut” in calculation, providing similar result].

$$y \pm k \cdot s_R = 2.20 \pm 1.41 \cdot 0.137 = 2.20 \pm 0.19$$

$$\text{Upper level, } U = 2.20 + 0.19 = 2.39$$

$$\text{Lower level, } L = 2.20 - 0.19 = 2.01$$

$$U - X = \text{Upper level} - \text{mean of the reference value: } 2.39 - 2.27 = 0.13$$

$$L - X = \text{Lower level} - \text{mean of the reference value: } 2.01 - 2.27 = -0.26$$

¹ For the Interlaboratory study, ILS, NordVal International has a simpler approach for the degree of freedom of the constant, k in $y \pm ks$. However, the results will be exactly the

same. In ISO 16140-2, the degree of freedom is calculated as

$$v = \frac{(H+1)^2}{\frac{(H+1)^2}{n} + \frac{1 - \frac{1}{n}}{p-1}}$$

however as $H = \frac{s_L^2}{s_r^2}$ and $s_L^2 \ll s_r^2$, the $H \approx 0$ and the equation is then equal to $v = p-1$,

where p = the number of laboratories.

Table 5.8 Results ILS for the medium level

	Medium Level log cfu/g									
Lab	Reference method					Alternative method				
	A	B	(A-B) ²	Mean, X	X ²	A	B	(A-B) ²	Mean, y	y ²
1	3.26	3.28	0.0004	3.27	10.69	3.30	3.46	0.0256	3.38	11.42
2	3.08	3.11	0.0009	3.10	9.58	3.32	3.11	0.0441	3.22	10.34
3	3.23	3.18	0.0025	3.21	10.27	3.32	3.08	0.0576	3.20	10.24
4	3.15	3.11	0.0016	3.13	9.797	3.18	3.38	0.0400	3.28	10.76
5	3.23	3.26	0.0009	3.25	10.53	3.11	3.20	0.0081	3.16	9.954
6	3.28	3.23	0.0025	3.26	10.60	3.32	3.26	0.0036	3.29	10.82
7	3.36	3.41	0.0025	3.39	11.46	3.32	3.30	0.0004	3.31	10.96
8	3.08	3.11	0.0009	3.10	9.579	3.34	3.15	0.0361	3.25	10.53
Sum			0.0122	25.68	82.50			0.2155	26.08	85.02
Mean	3.21					3.26		if < 0, then omitted		
S _r	0.028		=√(1/16·0.0122)			0.116				
s _L ²	0.010		= [8·82.50–25.68 ²]/(8·7)] – 0.028 ² /2			-0.002				
S _R	0.102		=√(0.028 ² + 0.010)			0.116				
Bias						0.05				
U=y + 1.4 s _R = 3.26 + 1.41· 0.116 =						3.42				
L=y - 1.4 s _R = 3.26 - 1.41· 0.116 =						3.10				
U-X	=3.42 - 3.26 =					0.21				
L-X	= 3.10 - 3.26 =					-0.11				

Table 5.9 Results ILS for the high level

Lab	High Level log cfu/g									
	Reference method					Alternative method				
	A	B	(A-B) ²	Mean, X	X ²	A	B	(A-B) ²	Mean, y	y ²
1	4.32	4.18	0.0196	4.25	18.06	4.23	4.43	0.04	4.33	18.75
2	4.38	4.20	0.0324	4.29	18.40	4.28	4.36	0.0064	4.32	18.66
3	4.18	4.15	0.0009	4.17	17.35	4.32	4.04	0.0784	4.18	17.47
4	4.11	4.11	0	4.11	16.89	4.15	4.15	0	4.15	17.22
5	4.18	4.34	0.0256	4.26	18.15	4.15	4.23	0.0064	4.19	17.56
6	4.36	4.30	0.0036	4.33	18.75	4.32	4.28	0.0016	4.30	18.49
7	4.08	4.18	0.01	4.13	17.05	4.11	4.08	0.0009	4.10	16.77
8	4.11	4.08	0.0009	4.10	16.77	4.26	4.30	0.0016	4.28	18.32
sum			0.093	33.63	141.43			0.1353	33.845	143.24
Mean	4.20					4.23				
S _r	0.076		$= \sqrt{(1/16 - 0.093)}$				0.092			
S _L ²	0.005		$= [8 \cdot 141.43 - 33.63^2] / (8 \cdot 7) - 0.076^2 / 2$				0.004			
S _R	0.105		$= \sqrt{(0.076^2 + 0.005)}$				0.109			
Bias						0.03				
U = y + 1,4 S _R	= 4,23 + 1,41 · 0,109 =					4.38				
L = y - 1,4 S _R	= 4.23 - 1.41 · 0.109 =					4.08				
U-X	4.38-4.20=					0.18				
L-X	4.08-4.20=					-0.13				

Summary of the results

The results of the ILS on the 3 levels are summed up in Table 5.10. These results are used for the graphical illustration of the accuracy profile

Table 5.10 Summary of the results of the ILS

Levels	Ref method	Alt. Method	Bias	U-x	L-x	AL 0.5	AL -0.5
1	2.27	2.20	-0.06	0.13	-0.25	0.5	-0.5
2	3.21	3.26	0.05	0.21	-0.11	0.5	-0.5
3	4.2	4.23	0.03	0.18	-0.13	0.5	-0.5

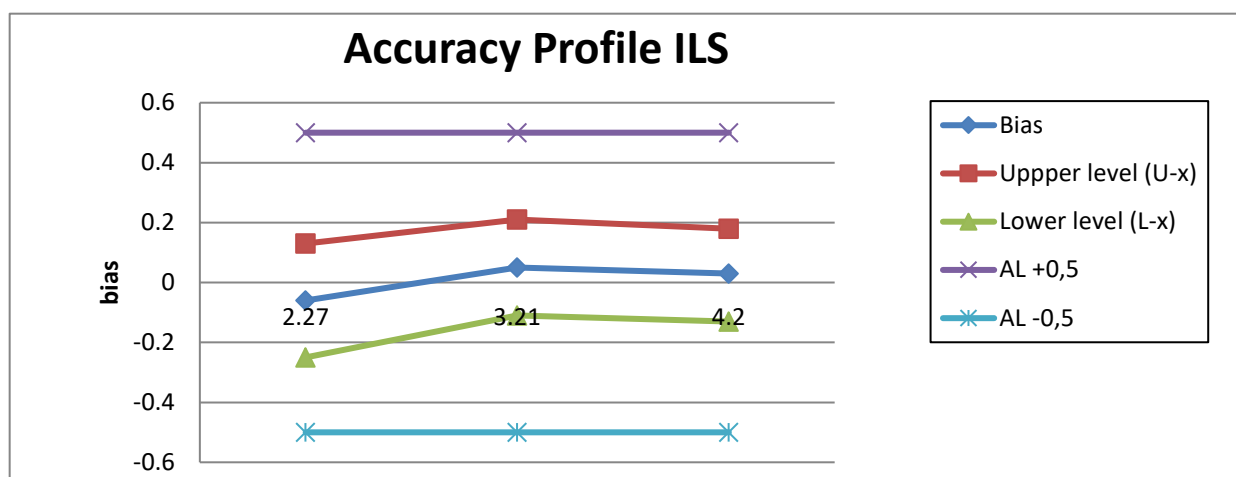


Figure 5.3 Accuracy profile for Interlaboratory Study

As the upper and lower level fall within the Acceptance Levels, the results are satisfactory.

The examples for the quantitative analyses are given in the Excel file available for download here: [NordVal-Prot-1 A3 Quantitative-methods-plots\(new\).xlsx](#)

For the Accuracy Profile calculations an Excel® spreadsheet is available for download here: [A5 AP calculation tool ILS.xlsx](#)

The results (tables and calculations) of the different parts and the interpretation of the results, including discrepant results, shall be given in a study report.

6. References

ISO 16140-2 Microbiology of food chain – Method validation- Part 2: Protocol for the validation of alternative (proprietary) methods against a reference method.

ISO 16140-2 Amendment 1: Revision of qualitative method comparison study data evaluation, relative level of detection calculations in the interlaboratory study, calculation and interpretation of the relative trueness study, and inclusion of a commercial sterility testing protocol for specific products

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

ISO 5725-2, 1994 Accuracy (trueness and precision) of measurement methods and results - Part 2: Basic method for the determination of repeatability and reproducibility of a stand measurement method.

NMKL Procedure No 8, 4. Ed. 2008: Measurement of uncertainty in quantitative microbiological examination of foods.

NMKL Procedure No 23, 2008: Guide on quality assurance in microbiological laboratories.

Bland, J.M, Altman D.G. Statistical methods for assessing agreement between two methods for clinical measurement. Lancet, 1986, i: 307-310.

Jasson. V., Uyttendaele, M., Rajkovic, A., Debevere, J. Establishment of procedures provoking sub-lethal injury of *Listeria monocytogenes*, *Campylobacter jejuni* and *Escherichia coli* O157 to serve method performance testing. Int. J. Food Microbiol. 2007, **118** (3) pp. 241–249.

Wehling. P. et al. Probability of Detection (POD) as a Statistical Model for the Validation of Qualitative Methods. J. AOAC Int. 2011, 94 pp. 335–337

7. Appendix of resources

- Appendix 1 Sensitivity study: [Statistics on qualitative results according to ISO 16140 2 amd 1\(new\).xlsx](#)
- Appendix 2 Qualitative analysis: [Statistics on qualitative results according to ISO 16140 2 amd 1\(new\).xlsx](#)
- Appendix 3 Quantitative methods: [NordVal-Prot-1_A3_Quantitative-methods-plots\(new\).xlsx](#)
- Appendix 4 Accuracy profile multiple comparison study: [A4_AP calculation tool MCS.xlsx](#)
- Appendix 5 Accuracy profile interlaboratory study: [A5_AP calculation tool ILS.xlsx](#)

8. Main changes in this version (29.06.2026)

- The protocol has been aligned with ISO 16140-2 Amd 1, except for the recognised errors in the amendment, and LOD has been included as described.
- For qualitative studies, new formulas have been included for both MCS and ILS. These are combined in a single Excel file, with separate worksheets: [Statistics on qualitative results according to ISO 16140 2 amd 1\(new\).xlsx](#)
- For the ILS, calculations are included for paired studies only (not unpaired studies).

- For quantitative studies, the Excel file includes Bland–Altman plot and Accuracy Profiles for both MCS and ILS, again organized in separate worksheets. For the Bland–Altman analysis, there is a minor modification to the confidence factor, using 1.96 instead of 2, i.e. no big change for the plots: [NordVal-Protocol A3 Quantitative-methods-plots\(new\).xlsx](#)
- For the Accuracy Profile, the ISO spreadsheets (free to download) are recommended: <https://standards.iso.org/iso/16140/-2/ed-1/en/>.