

Validation of analysis methods

Application to water microbiology

**Validation protocol for an alternative commercial method
as compared with a reference method**

Revision 2 – Adopted by AFNOR Certification on **17 May 2013**
(after approval by the relevant Technical Board)

CONTENTS

Introduction	4
1 Purpose.....	4
2 Normative references.....	4
3 Terms, definitions and symbols.....	4
3.1 General terms.....	5
3.1.1 Reference method	5
3.1.2 Alternative method.....	5
3.1.3 Validation of an alternative method	5
3.1.4 Analyte	5
3.1.5 Quantitative method.....	5
3.1.6 Qualitative method.....	5
3.1.7 Organizing laboratory / expert laboratory	5
3.1.8 Collaborating laboratory	6
3.2 Validation criteria for qualitative methods.....	6
3.2.1 Positive deviation (PD)	6
3.2.2 Negative deviation (ND)	6
3.2.3 Relative sensitivity (SE).....	6
3.2.4 Relative accuracy (AC).....	6
3.2.5 Relative specificity (SP).....	7
3.2.6 Inclusivity and exclusivity (Selectivity)	7
3.2.7 Relative detection level or LOD ₅₀	7
3.2.8 Practicability.....	7
3.3 Validation criteria for quantitative methods	7
3.3.1 Linearity	7
3.3.2 Accuracy	7
3.3.3 Relative accuracy	7
3.3.4 Repeatability	7
3.3.5 Reproducibility	8
3.3.6 Limit of detection.....	8
3.3.7 Quantification limit (or determination limit)	8
3.3.8 Inclusivity and exclusivity (Selectivity)	8
3.3.9 Practicability (see 3.2.8)	8
3.3.10 Acceptability limit	8
4 Principle.....	8
5 Preliminary study	9

5.1	Qualitative methods	9
5.1.1	Protocol for the assessment of inclusivity and exclusivity (selectivity).....	9
5.1.2	Protocol to measure relative accuracy, sensitivity and specificity	10
5.1.3	Protocol to measure the relative detection level.....	12
5.1.4	Practicability.....	13
5.2	Quantitative methods	14
5.2.1	Principle and definitions.....	14
5.2.2	Protocol for the assessment of inclusivity and exclusivity (selectivity).....	14
5.2.3	Protocol to measure linearity and relative accuracy	15
5.2.4	Detection and quantification limits	16
5.2.5	Practicability of the alternative method.....	16
6	Interlaboratory study	17
6.1	Qualitative methods	17
6.1.1	Design.....	17
6.1.2	Calculations and interpretation	17
6.1.3	Interpretation.....	19
6.2	Quantitative methods	19
6.2.1	Accuracy profile: general principles	19
6.2.2	Design.....	22
6.2.3	Statistical calculations.....	23
6.2.4	Construction of the accuracy profile	26
6.3	Summary of the interlaboratory test	27
6.3.1	Interpretation of the results	27
6.3.2	Limit of quantification	27
6.4	Bibliography for section 6	28
Annex 1. Classification by categories of water		29
Annex 2. Stress to apply by type of water		30
Annex 3. Organisation of analyses carried out in parallel depending on the type of method		31
Annex 4. Qualitative methods. Test applied to the rejection of discordant results		33
Annex 5. Qualitative methods. Interlaboratory study, criteria of accordance, concordance and concordance odds ratio		34
Annex 6. Quantitative methods. Calculation of repeatability and reproducibility standard deviations		39
Annex 7. Example of an accuracy profile calculation		40
Annex 8. Guidelines for the organisation and conducting of interlaboratory studies		43

Introduction

This document sets out a procedure for the characterization and validation of a method of microbiological analysis (qualitative or quantitative) as compared with a standard reference method when it exists.

1 Purpose

This document provides guidelines that shall lead to the validation, for a specified scope (e.g. water intended for human consumption and/or treated water and/or untreated water and/or surface water and/or deep water and/or bathing water and/or swimming pool water and/or mineral water and/or bottled water, etc.), of a "rapid" or commercial method of microbiological analysis designed to find (qualitative method) or count (quantitative method) target microorganisms as compared with a standard reference method targeting the same microorganisms.

The procedure is organized in the following way:

1. A preliminary study carried out by an expert laboratory which supplies an initial characterization of the method in terms of classical validation criteria (repeatability, practicability ...).
2. An inter-laboratory study designed to estimate various criteria for qualitative methods (§ 6.1.3) and to construct the accuracy profile (§ 6.2.4), for quantitative methods.
3. Final interpretation of the results.

The methods to be validated may be direct counting methods, for example of colonies, or indirect counting, for example counting the most probable number MPN.

2 Normative references

The following normative documents contain provisions that, owing to the reference made in them, constitute valid provisions for this document. In the case of dated references, later amendments or revisions of these publications do not apply. However, stakeholders in agreements founded on this document are invited to research the option of applying the most recent editions of the normative documents indicated below. In the case of undated references, the latest edition of the normative document referred to applies.

ISO/TR 13843:2000, Water quality – Guidance on validation of microbiological methods.

EN ISO 16140:2003, Microbiology of food – Protocol for the validation of alternative methods.

3 Terms, definitions and symbols

For the purposes of this document, the following terms and definitions apply.

3.1 General terms

3.1.1 Reference method

A standardized method of analysis specified for the purpose of detecting or counting a species or a given group of microorganisms.

3.1.2 Alternative method

A commercial method used to detect or count, for a given type of water, the same microorganisms as those measured by a standardized method, and which has for example, one or more of the following criteria:

- Speed of analysis and/or response.
- Ease of execution and/or automation.
- Analytical characteristics (precision, detection limit, etc.).

The term "alternative method" designates the product, equipment and procedure as a whole. It includes all the ingredients, whether material or not, required to implement the alternative method.

3.1.3 Validation of an alternative method

According to VIM:2006 subclauses 2.44 and 2.45

Provision of tangible evidence that the alternative method meets the requirements specified, that is results comparable to the results obtained with the reference method, taking into account any measurement uncertainty and that these requirements specified are suitable for a pre-determined use.

3.1.4 Analyte

Component measured using the method of analysis (microorganisms).

3.1.5 Quantitative method

Method of analysis whose response is the amount of analyte measured either directly (counting in a mass or volume), or indirectly (colour absorbance, impedance, etc.) in a certain amount of sample.

3.1.6 Qualitative method

Method of analysis whose response is qualitative (presence/absence of the analyte) and measured either directly or indirectly (colour absorbance, impedance, etc.) in a certain amount of sample.

3.1.7 Organizing laboratory / expert laboratory

Laboratory having the qualified staff and skills required to perform the preliminary study and the inter-laboratory study. It is selected according to criteria defined by AFNOR Certification.

3.1.8 Collaborating laboratory

Laboratory participating in the inter-laboratory study, according to the instructions defined by the organizing /expert laboratory.

3.2 Validation criteria for qualitative methods

The different situations of agreement or deviation between the two methods are summarized in Table 1.

Table 1. Types of agreement and deviation of results between the reference and alternative methods.

Results	Reference method positive	Reference method negative	Total
Alternative method positive	<i>PA</i> Positive agreement +/+	<i>PD</i> Positive deviation -/+	<i>A+</i>
Alternative method negative	<i>ND</i> Negative deviation +/-	<i>NA</i> Negative agreement -/-	<i>A-</i>
Total	$N+ = PA+ND$	$N- = NA+PD$	<i>N</i>

3.2.1 Positive deviation (PD)

The alternative method presents a positive deviation if it gives a positive result when the reference method gives a negative result. A positive deviation becomes a false positive result when the true result can be proven as being negative. A positive deviation is considered as a true positive when the true result can be proven as being positive.

3.2.2 Negative deviation (ND)

The alternative method presents a negative deviation if it gives a negative result when the reference method gives a positive result. A negative deviation becomes a false negative result when the true result can be proved as being positive.

3.2.3 Relative sensitivity (SE)

Ability of the alternative method to detect the analyte when it is detected by the reference method.

$$SE = 100 \times \frac{PA}{N+}$$

3.2.4 Relative accuracy (AC)

Relative accuracy

$$AC = 100 \times \frac{PA+NA}{N}$$

3.2.5 Relative specificity (SP)

Ability of the alternative method to not detect the analyte when it is not detected by the reference method.

$$SP = 100 \times \frac{NA}{N -}$$

3.2.6 Inclusivity and exclusivity (Selectivity)

Inclusivity is the ability of the alternative method to detect the target analyte from a relevant range of strains.

Exclusivity is the ability of the alternative method not to detect non-target analytes from a relevant range of strains.

3.2.7 Relative detection level or LOD₅₀

The relative detection level is the smallest number of culturable microorganisms that can be detected in the sample, with a probability of 50 % by the alternative and reference methods. LOD₅₀ is estimated by a statistical model but never directly measured.

3.2.8 Practicability

Usability of a method, defined by a set of criteria given by the user.

3.3 Validation criteria for quantitative methods

3.3.1 Linearity

Ability of the method when used with a given matrix to give results that are in proportion to the amount of analyte present in the sample, that is an increase in analyte corresponds to a linear or proportional increase in results.

3.3.2 Accuracy

Closeness of agreement between a test result and the accepted reference value.

3.3.3 Relative accuracy

Degree of correspondence between the response obtained by the reference method and the response obtained by the alternative method on the same samples.

NOTE The term "relative accuracy" used here is complementary to 3.3.2. For the purpose of this document, the accepted reference value is chosen as the value obtained by the reference method. The term "relative" implies therefore that the reference method does not automatically provide the accepted reference value.

3.3.4 Repeatability

Closeness of agreement between successive measurements of the same quantity, carried out under the same conditions of measurement.

3.3.5 Reproducibility

Closeness of agreement between successive measurements of the same quantity, carried out under conditions of reproducibility.

3.3.6 Limit of detection

(VIM. 4.18) Measured value, obtained by a given procedure, for which the probability of falsely claiming the absence of a component in a material is β , given a probability α of falsely claiming its presence.

Or

(Draft standard XP T 90-465-1) The lowest degree of bacterial load for which the probability of a positive result (observation of at least one germ) is greater than or equal to 95 % or 99 %. Note: this definition involves a probability theory law of the Poisson or negative binomial type.

3.3.7 Quantification limit (or determination limit)

The lowest average concentration of an analyte that may be measured and quantified with known uncertainty, under the experimental conditions of the method considered.

3.3.8 Inclusivity and exclusivity (Selectivity)

Inclusivity is the ability of the alternative method to detect the target analyte from a wide range of strains.

Exclusivity is the lack of interference by a relevant range of non-target strains with the alternative method.

NOTE This criterion does not apply to counting total flora.

3.3.9 Practicability (see 3.2.8)

3.3.10 Acceptability limit

Maximum acceptable deviation between the value assigned by the reference method and a result obtained by the alternative method. It is recorded as λ . In the case of bacterial counting, it is expressed in log units.

4 Principle

The validation protocol comprises two phases:

1. A preliminary study carried out by a single, so-called "expert laboratory" to estimate the performance of the alternative method against the reference method, carried out on natural or artificial samples or on reference materials.
2. An interlaboratory study organised by the expert laboratory with a minimum set number of laboratories involving the use of artificially spiked samples analysed in parallel using the alternative method and the reference method.

Before starting the study, whether the scope of the alternative method is compatible with that of the reference method should be defined.

In particular, this includes:

- identification of the microorganism or microorganisms targeted.
- the categories of water concerned and the concentration range envisaged.

The different categories of water that can be covered by the study are summarised at Annex 1. Classification by categories of water.

5 Preliminary study

5.1 Qualitative methods

5.1.1 Protocol for the assessment of inclusivity and exclusivity (selectivity)

To study the selectivity of the alternative method, analyse the strains in the list in Table 2.

Table 2. Minimum number of strains for the selectivity study.

	Types of strains	
Inclusivity	Target strains (species) (for example <i>E. coli</i> or <i>Legionella pneumophila</i>)	20
	Target strains belonging to a genus, a group or a family (for example coliforms, <i>Legionella</i> spp., or intestinal enterococci).	30
Exclusivity	Non-target strains known to cause interference with target strains (genus-species).	20
	Non-target strains known to cause interference with target strains belonging to a group or a family.	30

NOTE The number of strains to test may be increased according to the target (germ) of the alternative method to be validated (for example: 50 target strains on an inclusive basis when looking for salmonella).

Choose the target and non-target strains that regularly contaminate the class or classes, and/or categories of water selected. In addition, strains shall if possible come mainly from natural samples representative of the water's envisaged scope. The origin of each strain shall be known, recorded and traceable.

The level of inoculation of each microorganism is defined as follows:

- Target microorganisms (inclusivity): shall be around 10 to 100 times greater than the minimum relative detection level of the alternative method. The alternative method's complete protocol – without including the test portion – shall be implemented, but including pre-enrichment if stipulated. If the principle of the reference method is different from that of the alternative method, both methods are implemented in parallel.

- Non-target microorganisms (exclusivity): shall be around 1,000 to 100,000 times greater than the minimum relative detection level of the alternative method. If the enrichment medium is a selective broth, replace it with a non-selective broth. When the alternative method gives positive or doubtful results with non-target microorganisms, the test shall be repeated using the complete protocol. The laboratory may carry out the reference method in parallel if it so desires.

When there are discordant results between the alternative method and the reference method, the laboratory shall investigate to explain the results.

All these results shall be recorded in the final report and their interpretation shall be the subject of a comment by the expert laboratory.

5.1.2 Protocol to measure relative accuracy, sensitivity and specificity

5.1.2.1 Selection and preparation of samples

It is important to find samples of water that are naturally contaminated. In a general manner, all the categories/types for which validation is requested by the manufacturer must be studied. As a consequence, if validation is requested for all categories, they must all be studied. Please refer to Annex 1. Classification by categories of water.

When analysing naturally contaminated samples, the levels and distribution of contamination of the samples should be representative of the levels usually found in these classes of water.

If it is not possible to acquire a sufficient number of naturally contaminated samples for each of the categories, artificial contamination is permissible. The spiking method and levels targeted should as far as possible result in samples behaving similarly to naturally contaminated ones (for example, choose a low level of contamination for production or tap water). Thus, microorganisms used for spiking shall in particular have been stressed, a description of which is included in the final report.

Stress applied to microbial suspensions used for spiking artificial samples shall be relevant to the type of water concerned and the strain chosen. The types of stress applicable by water type are defined in Annex 1.

The types and modalities of stress, as well as the intensity of stress applied (a minimum of 0,5 log) shall be included in the expert laboratory's preliminary study report.

The total number of samples to be analysed is 60 per category of water. It is desirable to obtain approximately 50 % positive results and 50 % negative results. For each category of water, all the types of water listed in Annex 1. Classification by categories of water shall be tested, with equitable distribution of the samples between the types representative of the scope of the method being validated.

The reference and alternative methods shall always be implemented from two test portions from the same sample. Finally, each test portion is subject to the whole of the method's procedure. However, the first phases of both methods are often common; in this case, the two test portions are produced at the end of the final common phase (see Annex 3).

Comment With artificial contamination, directly inoculate the sample then homogenize it before taking the test portions.

5.1.2.2 Interpretation of the results

Collect the pairs of results obtained with the alternative and reference methods in the form of a table, like Table 3, using the conventions described in Table 1.

Table 3. Calculation of the relative accuracy, relative sensitivity and the relative specificity

Category	PA	NA	ND	PD	N	Relative accuracy AC	N+	Relative sensitivity SE	N-	Relative specificity SP
Water 1										
Water 2										
...										
Total										

For each category of water, calculate the following parameters:

$$AC = 100 \times \frac{PA + NA}{N} \text{ Relative accuracy}$$

$$SE = 100 \times \frac{PA}{N+} \text{ Relative sensitivity}$$

$$SP = 100 \times \frac{NA}{N-} \text{ Relative specificity}$$

Discordant results can be examined by the PD and ND count, using the McNemar test for example, as described in Annex F of the EN ISO 16140 standard.

When the values for PD and ND are high and almost equal, no statistical difference between the two methods can be detected using the McNemar test. In this case, the expert laboratory shall pay particular attention to the reasons for these high values of PD and ND.

5.1.2.3 Interpretation

Taking into account the number of positive deviations and the number of negative deviations, the capability of the alternative method to give more or fewer positive results than the reference method is evaluated.

The study report shall distinguish the results obtained with naturally contaminated and artificially contaminated samples.

By way of illustration of the values of the validation criteria, Figure 1 presents 10 extreme cases in order to understand how to interpret the results better. The left-hand scale gives the counts for 100 samples, and the right-hand scale gives the values of the criteria in %.

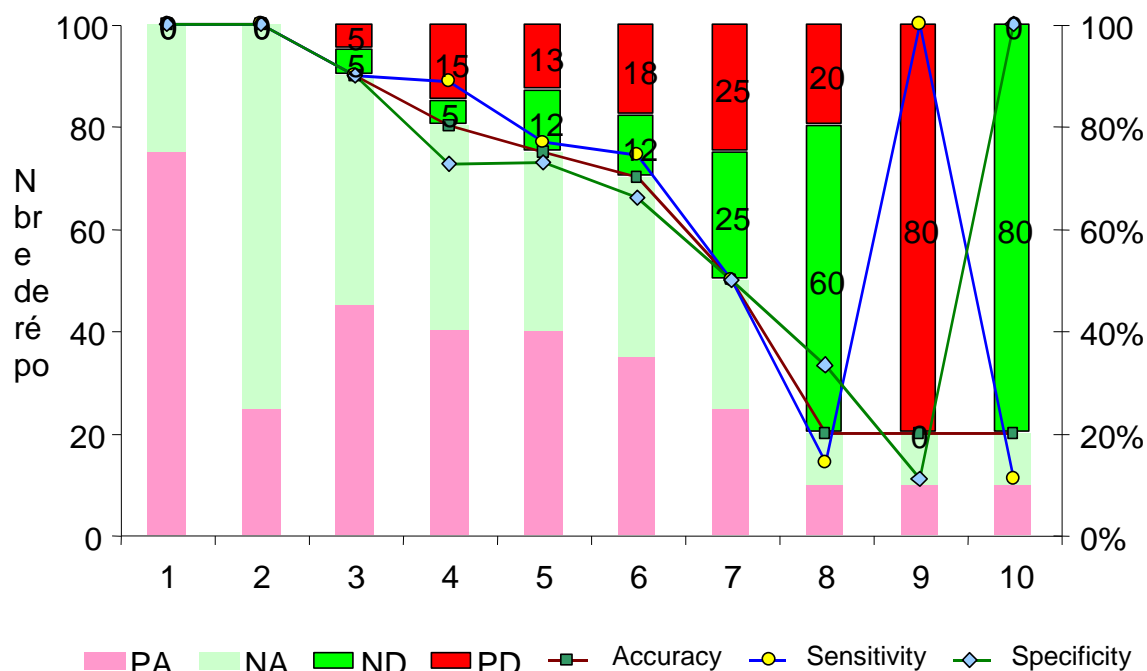


Figure 1. Example of the calculation of validation criteria for different situations.

5.1.3 Protocol to measure the relative detection level

5.1.3.1 Design

The relative detection level is defined in 3.2.7. To measure it, test in parallel with the reference method and alternative method according to the following design. For each category of water, choose a strain that is normally present. However, ensure that the strains are diversified, so that two categories of water are not associated with the same strain.

Initially determine the level of flora contamination associated with each category of water chosen.

For each "strain/category" combination, independently prepare 6 spiked samples for a minimum of 4 levels of contamination, including a negative control; that is, a total number of at least 24 spiked samples per combination. The choice of levels of contamination follows the following recommendation:

1. Level 1 corresponds to the absence of contamination.
2. Level 2 corresponds to contamination that gives less than 50 % positive results ($\leq 3/6$) with one of the two methods.
3. Level 3 shall give over 50 % positive results.
4. Level 4 shall be such that 100 % of results are positive.

Comment The expert laboratory may produce an inoculum level of less than 1 cell per test portion. Low contaminations are controlled by precise calibration and experience acquired with the strains tested.

Depending on the methods, arrange the measurements according to the recommendations in Annex .

5.1.3.2 Calculation and interpretation

For each level of contamination and each "category/strain" combination, estimate the relative detection level setting the limit point to 50 %, called LOD₅₀. This calculation is carried out according to the Spearman-Kärber method which also supplies an associated confidence interval. LOD₅₀ can also be interpreted as the level of contamination that theoretically corresponds to a recovery rate of 50 %.

The method's overall relative detection level is expressed by taking the minimum and maximum of all the confidence intervals of the various LOD₅₀.

5.1.4 Practicability

The following 13 criteria are to be completed by the expert laboratory.

Table 4. List of practicability criteria identified to characterise the alternative method

	Criteria to check	Communication concerning the criteria	Method of checking the criteria
1	Procedure for conditioning the elements of the method	Packaging or instructions	Verification by the expert laboratory
2	Reagent volume	Packaging or instructions	Verification by the expert laboratory
3	Conditions of storage of the elements (expiry date for unopened products)	Packaging or instructions	Verification by the expert laboratory that the conditions exist
4	Modalities of use after the first use (expiry dates for use)	Packaging or instructions	Verification by the expert laboratory that the modalities exist
5	Specific equipment or premises required	Instructions	Verification by the expert laboratory of the veracity of the written information
6	Reagents ready-to-use or to be reconstituted (in this case, is there a procedure)	Packaging or instructions	Verification by the expert laboratory of the veracity of the written information
7	Period required to train an operator not initiated into the method	Report	Measured by the expert laboratory (option of using periods implemented by collaborating laboratories) and allocated into one of the following 3 categories: less than 1 day, between 1 day and a week, more than a week.
8	Real-time handling and flexibility of the method depending on the number of samples to be analysed.	Report	Handling time measured in comparison with the reference method: less than, equal to or greater than the reference method handling time.

	Criteria to check	Communication concerning the criteria	Method of checking the criteria
9	Time required for obtaining the results	Report and certification	2 cycles are established describing each phase of the method only in terms of time: 1st cycle: negative samples- 2nd cycle: positive samples
10	Operator qualification type	Report	Specified by the expert laboratory as compared with the level required for the reference method: level identical to, or different from, that required for the reference method (the expert laboratory may use data from the collaborating laboratories)
11	Phases shared with the reference method	Report	Verification by the expert laboratory
12	Means or traceability of the analysis results for the user	Instructions	Verification by the expert laboratory
13	Obligation to maintain specific apparatus for the user	Report	Period and frequency

5.2 Quantitative methods

5.2.1 Principle and definitions

In the case of quantitative methods, the purpose of the preliminary study is to measure the performance of the alternative method, and if need be, the reference method for the following parameters:

- selectivity, specificity (inclusivity/exclusivity)
- linearity, relative accuracy
- detection limit, quantification limit
- practicability

5.2.2 Protocol for the assessment of inclusivity and exclusivity (selectivity)

To study the selectivity of the alternative method, analyse the strains in the list in Table 5.

This protocol does not apply to non-selective methods (for example, counting revivifiable aerobic germs).

Table 5. Minimum number of strains for the selectivity study

	Types of strains	
Inclusivity	Target strains (species) (for example <i>E. coli</i> or <i>Legionella pneumophila</i>)	20
	Target strains belonging to a genus, a group or a family (for example coliforms, <i>Legionella</i> spp., or intestinal enterococci).	30
Exclusivity	Non-target strains known to cause interference with target strains (genus-species).	20
	Non-target strains known to cause interference with target strains belonging to a group or a family.	30

Choose the target and non-target strains that regularly contaminate the class or classes, and/or categories of water selected. In addition, strains shall if possible come mainly from natural samples representative of the water's envisaged scope. The origin of each strain shall be known, recorded and traceable.

The degree of inoculation of each microorganism is defined as follows:

- Target microorganisms (inclusivity): count a suspension of the target germ by the alternative method and in a non-selective medium. If the principle of the reference method is different from that of the alternative method, both methods are implemented in parallel.
- Non-target microorganisms (exclusivity): use non-target germ concentrations around 10^4 to 10^5 CFU / volume analysed.

When there are discordant results between the alternative method and the reference method, the laboratory shall investigate to explain the results.

All these results shall be recorded in the final report and their interpretation shall be the subject of a comment by the expert laboratory.

5.2.3 Protocol to measure linearity and relative accuracy

NOTE Validation applies to the whole range of the alternative method. The levels of contamination chosen shall include the regulatory limits, if such exist, and the maximum range. If the calculation of the limit of quantification is high, a level once the assumed LOQ and another twice the assumed LOQ should be set.

5.2.3.1 Linearity

To check the linearity, for each category of water studied as described in Annex 1, prepare J samples of known concentration, such as reference materials or spiked natural samples. The number J is called the number of levels and it depends on the scope envisaged by the alternative method and shall be ≥ 3 .

At each level and for each category of water, carry out k complete and independent replications with each method ($k \geq 2$). Note y_{jk} the result obtained with the alternative method and x_{jk} the result obtained with the reference method. These two measurements are paired.

The linearity can generally be studied graphically by transferring the pairs of values obtained with the 2 methods (reference and alternative) onto a graph. If need be, the straight line $y_{jk} = a_0 + a_1 x_{jk}$ that passes between the points of the graph can be calculated with a linear regression method. Its slope a_1 shall be close to 1 and its y-intercept a_0 close to 0. However, it is not advantageous to carry out statistical compliance tests.

The raw values obtained for the samples analysed are presented in the form of two-dimensional graphs by categories of water and possibly by type of water if necessary. The counting results per category are presented on two-dimensional graphs as log N and as CFU.

5.2.3.2 Relative accuracy

For each category of water tested, a minimum of 20 samples are analysed in duplicate with both methods, that is, 80 analyses in all per category, with homogenous distribution for each type of water and a minimum of 6 to 7 samples per type.

The stress applied to microbial suspensions used for spiking artificial samples shall be relevant to the type of water concerned and the strain chosen. The types of stress applicable by water type are defined in Annex 2.

The types and modalities of stress, as well as the intensity of stress applied (a minimum of 0.5 log) shall be included in the expert laboratory's preliminary study report.

The relative accuracy between the alternative method and the reference method is evaluated by calculating the differences $d_{jk} = y_{jk} - x_{jk}$ for each pair of values. For each level j , group the values of d_{jk} and calculate a relative accuracy for the level via the median:

$$D_i = \text{mediane}(d_{jk})$$

The relationship between the relative accuracies D_j and the values of the levels (x_j) may be modelled. A null value ($D = 0$) is expected for each level. If the values D are too large for all or part of the scope, it may be concluded that the alternative method differs too much from the reference method and does not merit further validation. In addition, the values of D can be used to start to define a value for the limit of acceptability (λ) that will be used to interpret the accuracy profile (section 6.2.1). However, the repeatability and purpose of the method must also be borne in mind when fixing the value λ .

The raw values obtained for the samples analysed are presented in the form of two-dimensional graphs by categories of water and possibly by type of water if necessary. The counting results per category are presented on two-dimensional graphs as log N and as CFU.

When there are discordant results between the alternative method and the reference method, the laboratory shall investigate to explain the results.

5.2.4 Detection and quantification limits

NOTE The calculation shall take the type of matrix and regulatory threshold values into consideration.

The detection and quantification limits are checked, according to the manufacturer's recommendations and the reference method, in accordance with 6.2.2.4 of EN ISO 16140.

The detection and quantification limits are determined by analysing a pure culture of the target strain. As a minimum, 6 replicates for each level are tested in sterile water. As a minimum, 3 to 4 levels shall be tested (in addition to level 0), of which, at least one level with less than 50 % positive responses and one level with more than 50 % positive responses.

5.2.5 Practicability of the alternative method

The 13 criteria to be completed by the expert laboratory are listed in Table 4.

6 Interlaboratory study

The interlaboratory study shall be used to evaluate the performance of the method in several laboratories under real conditions that represent its routine application. It is organized under the control of the expert laboratory and generally follows the preliminary study. Its interpretation is "method" oriented and not "laboratory" oriented.

6.1 Qualitative methods

6.1.1 Design

The interlaboratory study applicable to qualitative methods must involve at least 8 collaborating laboratories having results without outliers.

The collaborating laboratory must demonstrate its competence in the use of the alternative method and of the reference method prior to participating in the study.

- one relevant water matrix is used to prepare the test samples.
- each sample shall be individually inoculated.
- at least three different levels of contamination shall be used: a negative control (L_0), one level slightly above the relative detection level of the alternative method (L_1), and one level about 10 times greater than the detection level (L_2) (for example 0, 3, and 30 cells/25 mL).
- the stability of the microorganism in the water matrix used shall be demonstrated.
- blind replicates at each level of contamination are analysed with the reference method and with the alternative method in each of the collaborating laboratories, that is $X=24$ samples analysed per laboratory.
- if the alternative method and the reference method do not have a common phase, the expert laboratory prepares and sends the 24 samples to each collaborating laboratory in duplicate (2X).
- the samples shall be analysed in each laboratory on the date stipulated.
- the temperature of the control sample shall not exceed 10 °C during transport, until the samples are received by the collaborating laboratory. A temperature recorder must be used to record changes in temperature during shipment (for example to create heat profiles).

The expert laboratory using all recorded data shall determine which results are suitable and which are outliers for use in calculating the precision data.

6.1.2 Calculations and interpretation

For each level, put the positive results obtained with each method as illustrated in

Table 6 and Table 7.

Table 6. Positive results obtained by the reference method

Laboratories	Contamination level		
	L_0	L_1	L_2
Laboratory 1	/8	/8	/8
Laboratory 2	/8	/8	/8
Laboratory 3	/8	/8	/8
Etc.	/8	/8	/8
Total	FP	$TP1r$	$TP2r$

FP: false positive by the reference method

TP1r: true positive at level 1 by the reference method

TP2r: true positive at level 2 by the reference method

Table 7. Positive results obtained by the alternative method

Laboratories	Contamination level		
	L_0	L_1	L_2
Laboratory 1	/8	/8	/8
Laboratory 2	/8	/8	/8
Laboratory 3	/8	/8	/8
Etc.	/8	/8	/8
Total	FP	$TP1r$	$TP2r$

FP: false positive by the alternative method

TP1r: true positive at level 1 by the alternative method

TP2r: true positive at level 2 by the alternative method

For level L_0 and for each method, calculate the percentage specificity SP as in the following equation:

$$SP = 100 \times \left(1 - \frac{FP}{N-} \right)$$

Where $N-$ is the total number of all tests L_0 .

FP is the number of false positives.

For each positive contamination level (L_1 and L_2), calculate the percentage sensitivity SE using the following equation:

$$SE = 100 \times \left(\frac{TP}{N+} \right)$$

Where $N+$ is the total number of L_1 or L_2 tests respectively.

TP is the number of true positives.

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 and to examine the discordant results.

Each pair of results from a sample measured by the alternative and the reference methods shall be reported as in Table 8.

Table 8. Paired results of the alternative method and the reference method in the context of the collaborative study.

Alternative method	Reference method		Total
	+	-	
+	<i>PA</i>	<i>PD</i>	
-	<i>ND</i>	<i>NA</i>	
Total	<i>N+</i>	<i>N-</i>	<i>N</i>

Calculate the relative accuracy AC expressed in percentage, as in the following equation:

$$AC = 100 \times \left[\frac{PA + NA}{N} \right]$$

Calculate the confidence intervals for each proportion (see 5.1.2.2.2 of EN ISO 16140).

Examine the discordant results (see 5.1.2.2.3 of EN ISO 16140).

6.1.3 Interpretation

Compare AC, SE and SP with their corresponding values, obtained in the context of the preliminary study. These criteria do not really address the variability of the method within a laboratory and between laboratories.

The degree of agreement, concordance and odds ratio criteria are used to help address this notion of variability (repeatability and reproducibility) (cf. Annex 5 of this document).

6.2 Quantitative methods

6.2.1 Accuracy profile: general principles

The accuracy profile allows a statistical approach to validation to be used. It is based on the direct application of the principles described in the standards in the ISO 5725 series, parts 1, 2 and 3. These standards describe a statistical method for estimating the accuracy (trueness and precision) of a method or results. In this case, the model applied to explain a measurement z of a measurand Z from a laboratory is of the type:

$$z = m + B + e$$

where m is the general mean of the homogenous sample sent to the laboratories, B the laboratory's bias component under conditions of repeatability and e the random error occurring in each measurement under conditions of repeatability. The bias component comes from the laboratory, in the case of an interlaboratory study, but it can also come from another source of uncertainty, like the day, the operator, the instrument, etc. in the case of an intralaboratory study. The best known application of the ISO 5725 standard is the calculation of precision criteria in an interlaboratory study that results in standard deviations for repeatability and reproducibility. Apart from methods for the calculation of accuracy criteria, these standards also specify the organization of collection in detail and the precautions to be observed.

Based on the reproducibility standard deviation calculated according to the recommendations of the ISO 5725 standard, a so-called tolerance (or prediction) interval is constructed which contains a proportion β of future results. **It is important to note that the tolerance interval shall not be confused with the confidence interval.** A confidence interval describes the dispersion of results around a parameter, such as the mean, based on past experiments. Thus the confidence interval for the mean is expressed as follows, with k_{IC} its coverage factor

and $s_{\bar{x}}$ the standard deviation from the mean:

$$\bar{x} \pm k_{IC} \times s_{\bar{x}}$$

A tolerance interval describes the dispersion of **future** results. It takes account of past experiments and uncertainty concerning the mean. It is expressed with k_{IT} another coverage factor calculated in a completely different manner from k_{IC} and s_R , the reproducibility standard deviation:

$$\bar{x} \pm k_{IT} \times s_R$$

This calculation is repeated for each level of concentration j to obtain a series of tolerance intervals which, together, cover the whole of the method's scope and form what is called the accuracy profile.

Figure 2 gives an example of an accuracy profile constructed with 3 levels of concentration, between 2.2 and 4.2 log units, which define the validation range or scope of the method to validate.

The horizontal axis represents the theoretical concentration of the levels and the vertical axis the difference between the theoretical concentration and the concentration found expressed as a log, that is, the absolute bias. The tolerance interval limits define a range where a proportion β of future results is located (in this case 80 %).

Finally, the accuracy profile may be compared to an interval of acceptability, defined according to the method's purpose. The limits of the intervals of acceptability, recorded as $\pm\lambda$, are also expressed by a log difference. The interpretation strategy is described in detail in Annex 7. Example of an accuracy profile calculation. It is defined as informative since it calls on a limit of acceptability λ , which depends on the method's context of use and the β proportion chosen.

In the range marked by the broken vertical line the method is capable of producing a proportion β of results between the limits of acceptability. The method is then said to be valid in this range indicated by the double arrow. The scope represents the range initially chosen in which to conduct validation.

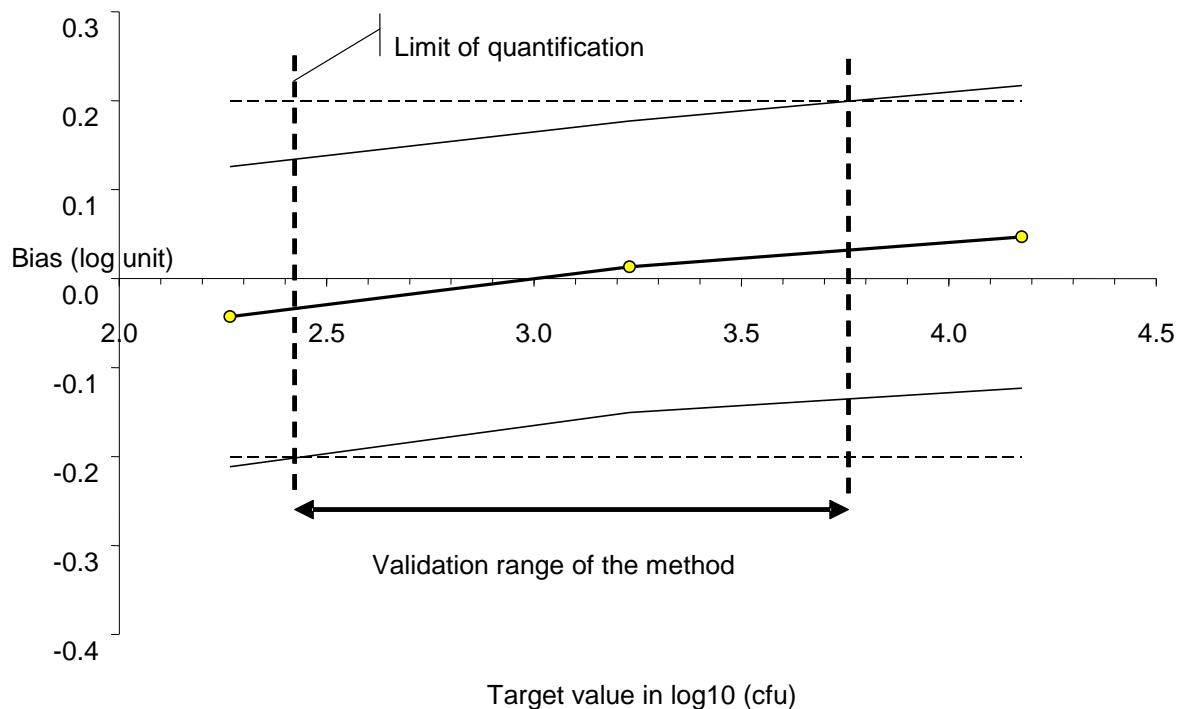


Figure 2. Accuracy profile calculated for 3 levels of concentration. Each grey circle represents the average concentration found for the level and quantifies the trueness. The dotted lines mark the interval of acceptability and the continuous lines the tolerance interval calculated from the intermediate precision standard deviations for each level. The vertical lines mark the range of validity, in which the method is capable of producing a known proportion (80 %) of acceptable results.

The operations leading to the construction of the accuracy profile may be summarised as follows:

1. Organise the collection of data according to the design described in Section 6.2.2.
 2. Transform the counts into log₁₀.
- IMPORTANT NOTE** Present the accuracy profiles in log form and as a number of microorganisms, except in specific cases validated by the Technical Board.
3. For each level (low, medium and high) calculate the reproducibility standard deviation with Z replicates of the alternative method, according to ISO 5725.
 4. For each level, calculate the median of X replicates of the reference method; these medians are the **target values** that shall be achieved by the alternative method, taking the acceptability limits into account.
 5. For each level, calculate the mean of Z replicates of the alternative method.
 6. For each level, calculate the tolerance interval limits according to the publication's formulae (Hubert, 2006). Remember that the tolerance interval defines a range where microbiologists can guarantee they can produce a known percentage (80 % in the example) of future measurements.
 7. Summarise all the calculations in a table that combines data from Table 10 and from Table 11 and a figure designed as in Figure 2.

6.2.2 Design

6.2.2.1 Organisation of the interlaboratory study

The interlaboratory study is used to evaluate the alternative method in relation to the reference method. It shall be organized by the expert laboratory which selects a sufficient number of laboratories (preferably at least 10) to obtain a minimum of 8 laboratories with no measurement rejected as an outlier. It is based on sending samples which will be analysed in duplicate and in parallel by the alternative method and the reference method. The purpose of these tests is to calculate the performance criteria for the alternative method which will be used to construct the accuracy profile and make a decision concerning its validity. Guidelines and requirements for organising, dispatching and conducting interlaboratory studies are given in Annex 8.

The expert laboratory is responsible for the preparation of the test protocol and the measurement data sheet for recording all measurements and critical experimental conditions used by each laboratory. The analyst in each participating laboratory shall demonstrate his/her competence in the use of the alternative method and of the reference method prior to participating in the study.

The protocol is the following:

- Choose a relevant matrix generally defined after the preliminary study (see 6.2.2.3).
- Choose the levels of concentration for artificially or naturally contaminated samples and which cover at least the lower, middle and upper levels of the expected scope of the alternative method. A negative control may also be included but the measurements carried out are not included in the calculations.
- The expert laboratory shall demonstrate the homogenous nature and stability of the three concentration levels tested.
- Use the same samples with each method.
- The organising laboratory prepares 4 aliquots from each level (that is, two aliquots each measured by both methods) for each laboratory. These aliquots are coded *blind*, such that two are measured by the reference method and two measured by the alternative method. In some special cases, subdivide the aliquots, just before measurement with the two methods (see Annex 3. Organisation of analyses carried out in parallel depending on the type of method).
- Each participating laboratory shall analyse the aliquots on a stipulated date using common batches of media, reagents and/or kits.
- The expert laboratory shall define the rules for rounding results.
- The expert laboratory pools the laboratories' results and sorts them by level j and by laboratory i as shown in Table 9.

Table 9. Presentation of the results of the interlaboratory study

Laboratory <i>i</i>	Level <i>j</i>	Reference method X		Alternative method Z	
		Replicate 1	Replicate 2	Replicate 1	Replicate 2
Lab 1	1	x_{111}	x_{112}	z_{111}	z_{112}
Lab 2	1				
Labo <i>i</i>	1				
Labo <i>l</i>	1				
Labo <i>i</i>	<i>J</i>				
Labo <i>l</i>	<i>J</i>				

6.2.2.2 Choice of the numbers of levels, participating laboratories and replicates

The minimum conditions consist of choosing a number of levels $J = 3$ and a number of replicates $K = 2$. It is essential that $J \geq 3$ because the linearity between the reference values and the values found must be checked. Three levels are needed for this verification. For the value of the number of laboratories *l*, reference can be made to part 1 of ISO 5725. To compensate for low values of *J* and *K*, it was decided to take $l \geq 8$ laboratories. Preferably, these laboratories shall be accredited to EN ISO 17025 but above all have good experience of microbiological methods and finally accept the participation and quality assurance (traceability) rules described by the organising laboratory.

Finally, the number of measurements collected shall therefore be at least equal to $8 \times 2 \times 2 \times 3 = 96$.

6.2.2.3 Type of samples

The interlaboratory study only covers a single matrix. The choice of category to which it belongs shall be defined following the preliminary study. Concentration levels may be prepared by spiking in order to achieve the values required.

6.2.3 Statistical calculations

NOTE A laboratory's results shall only be excluded as an exception and this shall be systematically documented (enquiry carried out by the organising laboratory with the collaborating laboratory).

Two Excel files are provided in a controlled manner by AFNOR Certification and shall be used to construct the accuracy profile, one for the interpretation of the CFU data and the other for the Log data.

6.2.3.1 Verification of linearity

Before any computer processing, it is always a good idea to make a prior illustration in graph form and carry out a visual examination of the data in order to detect any obvious errors, such as incorrectly recorded data. These graphs may be placed in the appendix of the assessment file. The presence of abnormal data leads to the suspicion of incorrect development of the method and may invalidate the validation procedure or a particular laboratory.

In microbiology, the measurements X and Z are not always distributed according to normal distribution, and in particular a high level of asymmetry of distribution can be observed. Often in order to get a more symmetric distribution, counts are better transformed into logarithms. Other transformations may also be used for the small numbers found in water microbiology.

6.2.3.2 Calculation of target reference values per concentration level

Use the data X from Table 9 to calculate the target reference value per level. The simplest solution consists of using the median:

$$\tilde{x}_j = \text{mediane}(x_{i,k}) \quad (1)$$

These medians are used to calculate the trueness of the method and the accuracy of the measurements. It can be seen that it may not be an advantage to carry out as many replicates with the reference method, in as much as they are not always used as such.

6.2.3.3 Calculation of precision standard deviations per concentration level

Use the data Z from Table 9 to calculate the repeatability s_r , inter-series s_B and reproducibility s_R standard deviations. This calculation is carried out independently for each concentration level j (with $1 \leq j \leq 3$) according to the principle of ISO 5725-2 described in Annex 6. Collect the data in a table in the form of Table 10.

Table 10. Precision and trueness criteria per level

Level		1	...	J
Number of laboratories	I			
Number of replicates	K			
Average theoretical target concentration	\tilde{x}_j			
Average concentration found	$\bar{\bar{z}}_j$			
Repeatability standard deviation	$s(j)_r$			
Inter-series standard deviation	$s(j)_B$			
Intermediate precision standard deviation	$s(j)_R = \sqrt{s(j)_r^2 + s(j)_B^2}$			
Absolute mean bias	$\bar{\bar{z}}_j - \tilde{x}_j$			

Obtaining reliable estimates of the mean or the standard deviation can be problematic in the presence of outliers. ISO 5725 proposes tests for discarding outlier means or variances (Grubbs and Cochran tests respectively) which can be used to exclude values with a strong influence and so obtain a better estimate. This nevertheless reduces the number of values that can be used for statistical analysis and is not recommended for this type of study.

6.2.3.4 Calculation of the tolerance interval

Reminder. The tolerance interval defines an interval in which there is an expectation of finding on average a proportion β of future results obtained by applying the method routinely, that is under the conditions of precision applied (in this case reproducibility). For this reason, this is also called the prediction interval.

The method proposed by Mee [Mee, 1984] is the one chosen for the protocol. It has also been the subject of several publications [Hubert, 2003, Hubert 2006a, Hubert, 2006b]. The calculation is carried out using data Z from Table 10 for each level of concentration j (with $3 \leq j \leq J$). In order to simplify the formulas, the index j has been omitted from the following formulae.

The tolerance interval is symmetrical around the average concentration found \bar{z} and is expressed by the formula:

$$\bar{z} \pm k_{tol} \times s_R \quad (2)$$

where k_{tol} is a coverage factor equal to:

$$k_{tol} = t_{v, \frac{1+\beta}{2}} \sqrt{1 + \frac{1}{KIB^2}} \quad (3)$$

with:

$$B = \sqrt{\frac{R+1}{IR+1}} \quad (4)$$

$$R = \frac{s_B^2}{s_r^2} \quad (5)$$

In addition, $t_{v, \frac{1+\beta}{2}}$ is the quantile of the Student t distribution for v degrees of freedom, calculated according to an approximation method proposed by Satterthwaite [Mee, 1984].

$$v = \frac{(R+1)^2}{\frac{\left(R + \frac{1}{I}\right)^2}{K-1} + \frac{1}{KI}} \quad (6)$$

For each level j , calculate the coverage factor k_{tol} and the lower and upper limits of the tolerance interval. The value chosen for β shall be at least 80 %. All the calculations are collected in a table in the form of Table 11.

Table 11. Tolerance interval limits per level

Level		1	...	J
Average theoretical concentration	\tilde{x}			
Lower tolerance limit	$\bar{z} - k_{tol} \times s_{FI}$			
Upper tolerance limit	$\bar{z} + k_{tol} \times s_{FI}$			
Differential lower tolerance limit	$(\bar{z} - k_{tol} \times s_{FI}) - \tilde{x}$			
Differential upper tolerance limit	$(\bar{z} + k_{tol} \times s_{FI}) - \tilde{x}$			

NOTE The number v is rarely an integer and Student tables which accept non-integer degrees of freedom should be used. However, the quantile value can be approximated via linear interpolation between the 2 integer numbers of degrees of freedom that surround v .

The value of v depends on the relationship $R = \frac{s_B^2}{s_r^2}$ between the inter-series and repeatability variances.

The relationship expresses the relative importance of the series effect. For example, when the day effect is the series effect, if the method is very stable from one day to the next and can supply very close results for the same sample, the correction introduced by R will be low. By contrast, it will increase with this relationship. In fact, the lower the number of degrees of freedom, the higher the Student t distribution quantile and the more the tolerance interval is extended.

The following table illustrates this mechanism for a design V with $K = 3$ and $l' = 3$ and s_{Fl} constant and equal to 1. When the variance ratio R increases from 0 to 9, the coverage factor k_{tol} and the scope of the tolerance interval increase from 1.97 to 3.07.

Table 12. Influence of the variance ratio on the scope of the tolerance interval ($\rho = 0.90$)

K	l'	s_r	s_B	s_{Fl}	R	v	$t_{v; \frac{1+\beta}{2}}$	k_{tol}
3	3	1.00	0.00	1.0	0.0	7.714	1.869	1.970
3	3	0.71	0.71	1.0	1.0	4.154	2.109	2.332
3	3	0.58	0.82	1.0	2.0	3.219	2.290	2.569
3	3	0.50	0.87	1.0	3.0	2.842	2.408	2.722
3	3	0.45	0.89	1.0	4.0	2.642	2.489	2.826
3	3	0.41	0.91	1.0	5.0	2.518	2.549	2.902
3	3	0.38	0.93	1.0	6.0	2.434	2.594	2.959
3	3	0.35	0.94	1.0	7.0	2.374	2.629	3.004
3	3	0.33	0.94	1.0	8.0	2.328	2.658	3.041
3	3	0.32	0.95	1.0	9.0	2.292	2.681	3.070

6.2.4 Construction of the accuracy profile

The accuracy profile may be constructed in different ways, depending on the type of data being processed. The most traditional method, when relative concentrations are involved, is the one shown in Figure 2 where the results are expressed by the coverage rate or relative bias. But if the data is a matter of counts expressed in logarithms, it is better to use the absolute bias in as much as a difference between 2 quantities expressed by their logarithms is equivalent to their ratio.

To construct the accuracy profile select the following 5 lines from Table 10 and Table 11:

1. Average theoretical concentration (horizontal axis)
2. Average concentration found
3. Relative lower tolerance limit
4. Relative higher tolerance limit
5. Average coverage rate

Transfer the data to a graph using the average theoretical concentration to draw the x-axis.

Construct at least two profiles with $\beta=80\%$ and $\beta=90\%$, for interpretation purposes.

6.3 Summary of the interlaboratory test

6.3.1 Interpretation of the results

Interpret the graph using the following rules:

- Suggest a value for the acceptability criterion λ . Justify this choice according to the problem being processed and prior experience. The final choice is the Technical Board's responsibility.
- Depending on the value of λ , the alternative method is said to be valid in the whole range where the tolerance interval is between the acceptability limits.
- Finally, define the quantification limit as the point where the tolerance interval intersects one of the two acceptability limits. This is the limit beyond which the microbiologist can no longer guarantee that a percentage β of the results obtained by the alternative method are acceptable.

6.3.2 Limit of quantification

The quantification limit is calculated by simple linear interpolation between the points where the tolerance interval intersects the acceptability limit. Figure 3 illustrates this calculation.

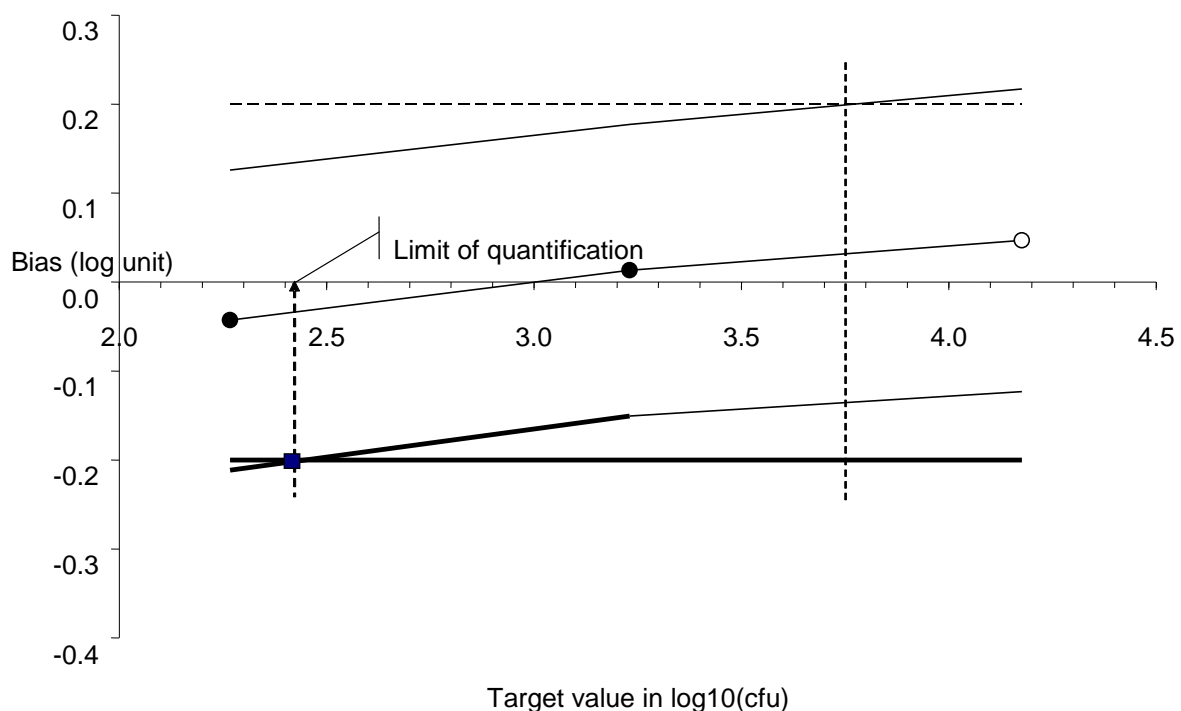


Figure 3. Calculation of the quantification limit via interpolation

The straight line formed by the lower limit of the tolerance interval (segment shown in bold) will be considered. Assuming that its equation is formed using the values of the target levels on the x-axis and the value of the lower limit of the tolerance interval on the y-axis, we have,

$$y = c_0 + c_1X$$

Since the two pairs of values $\{x_1, y_1\}$ and $\{x_2, y_2\}$ are known, the slope and the y-intercept can easily be calculated by rules of three and the value of the x-axis for the value $-\lambda$ can be predicted. When the intersection occurs with the upper limit, $-\lambda$ is replaced by $+\lambda$:

$$c_1 = \frac{y_2 - y_1}{x_2 - x_1}$$

$$c_0 = y_1 - c_1 x_1$$

$$LOQ = \frac{-\lambda - c_0}{c_1} \tag{7}$$

In the example shown in Figure 3, the values found are as follows:

Points	1	2
Value of the x-axis (target concentration)	2.267	3.230
Value of the y-axis (IT lower limit)	-0.211	-0.150
Slope	0.0633	
Origin	-0.3546	
Limit of acceptability (lambda)	0.2	
LOQ	2.44	

The LOQ, depending on the choices made for acceptability λ and the proportion β , is 2.4 log units.

If there is no intersection, as shown in Erreur ! Référence non valide pour un signet..

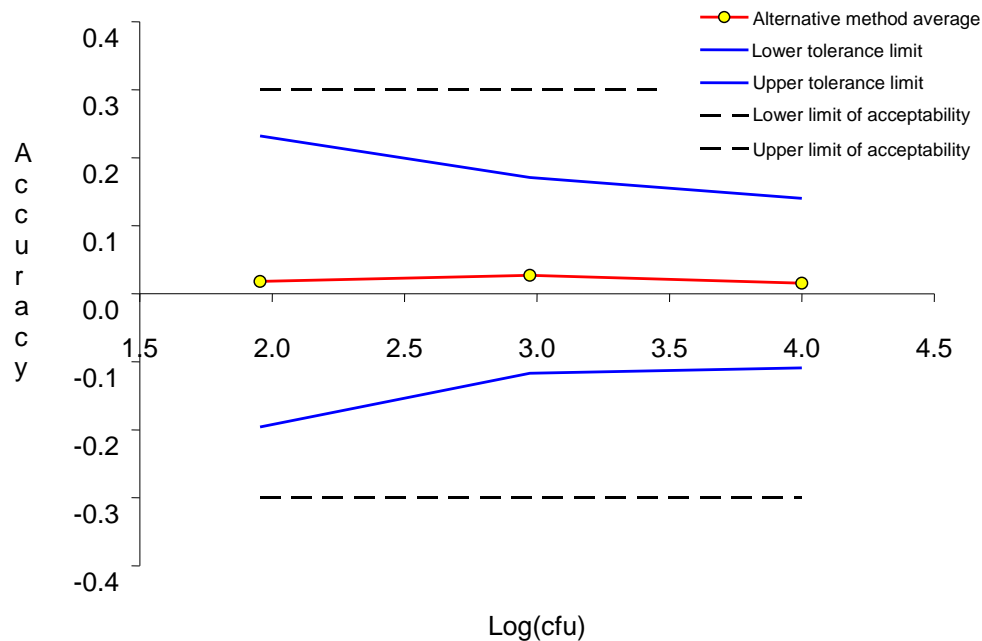


Figure 4, LOQ corresponds to the lowest value of the validation range.

6.4 Bibliography for section 6

Hubert P., Nguyen-Huu J.J., Boulanger B., Chapuzet E., Chiap P., Cohen N., Compagnon

P.A., Dewe W., Feinberg M., Lallier M., Laurentie M., Mercier N., Muzard G., Nivet C., Valat L. (2003) Validation of quantitative analytical procedure, Harmonization of approaches. Part I. STP Pharma Pratiques 13(3) 101-138.

Hubert P., Nguyen-Huu J.J., Boulanger B., Chapuzet E., Chiap P., Cohen N., Compagnon P.A., Dewe W., Feinberg M., Lallier M., Laurentie M., Mercier N., Muzard G., Nivet C., Valat L. (2006a) Quantative analytical procedures: Harmonization of the approaches. Part II – Statistics STP Pharma Pratiques 16(1) 28-58.

Hubert P., Nguyen-Huu J.J., Boulanger B., Chapuzet E., Chiap P., Cohen N., Compagnon P.A., Dewe W., Feinberg M., Lallier M., Laurentie M., Mercier N., Muzard G., Nivet C., Valat L. (2006b) Quantitative analytical procedures validation: harmonization of the approaches. Part III. Examples of application. STP Pharma Pratiques 16(2) 87-121.

Mee R.W. (1984) β -Expectation and β -Content Tolerance Limits for Balanced One-Way ANOVA Random Model. *Technometrics* 26:3, 251-254

Annex 1. Classification by categories of water

Water for human consumption:

Water from wells, springs and boreholes

Bottled water: spring water, mineral and other water

Mains and drinking fountain water

Bathing water (untreated):

Sea water

Fresh water

Swimming pool water

Industrial water:

Cooling tower water

Sanitary hot water

Waste water, raw and processed waste water

Others (Specific requests from manufacturers)

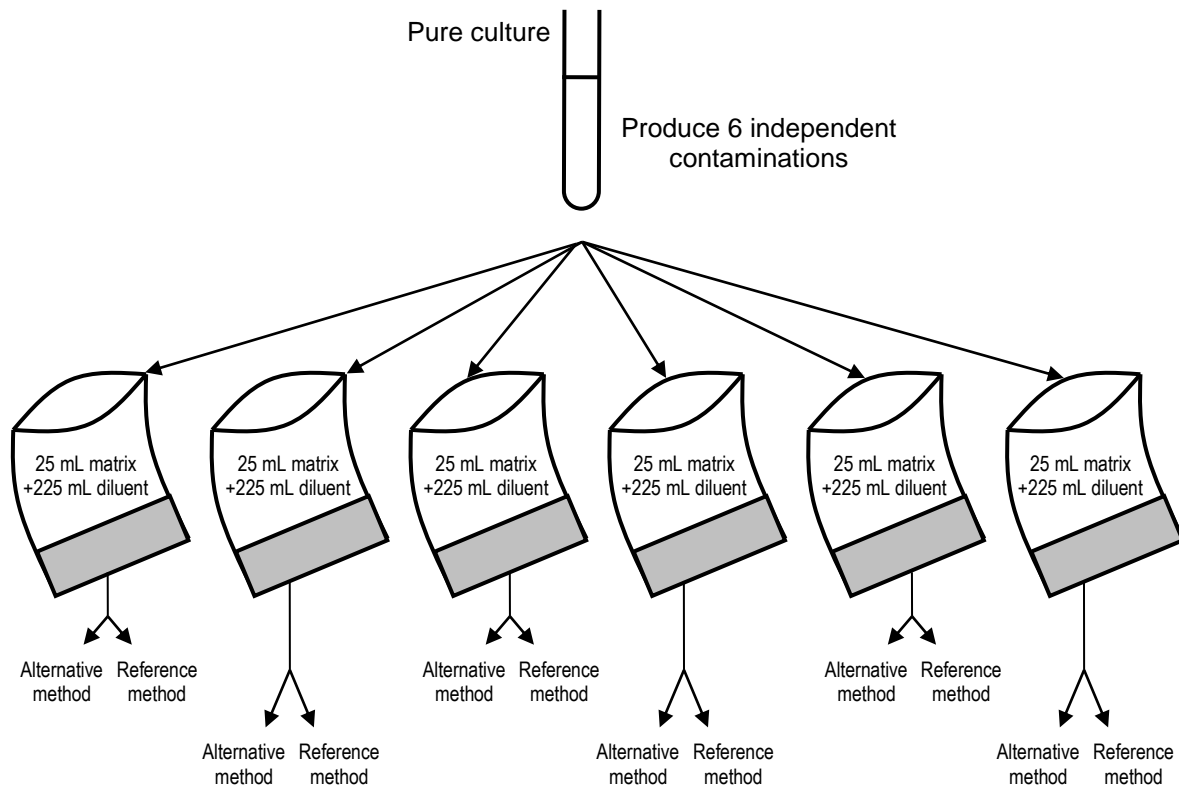
For example: thermal waters (spa water, ...)

Annex 2. Stress to apply by type of water

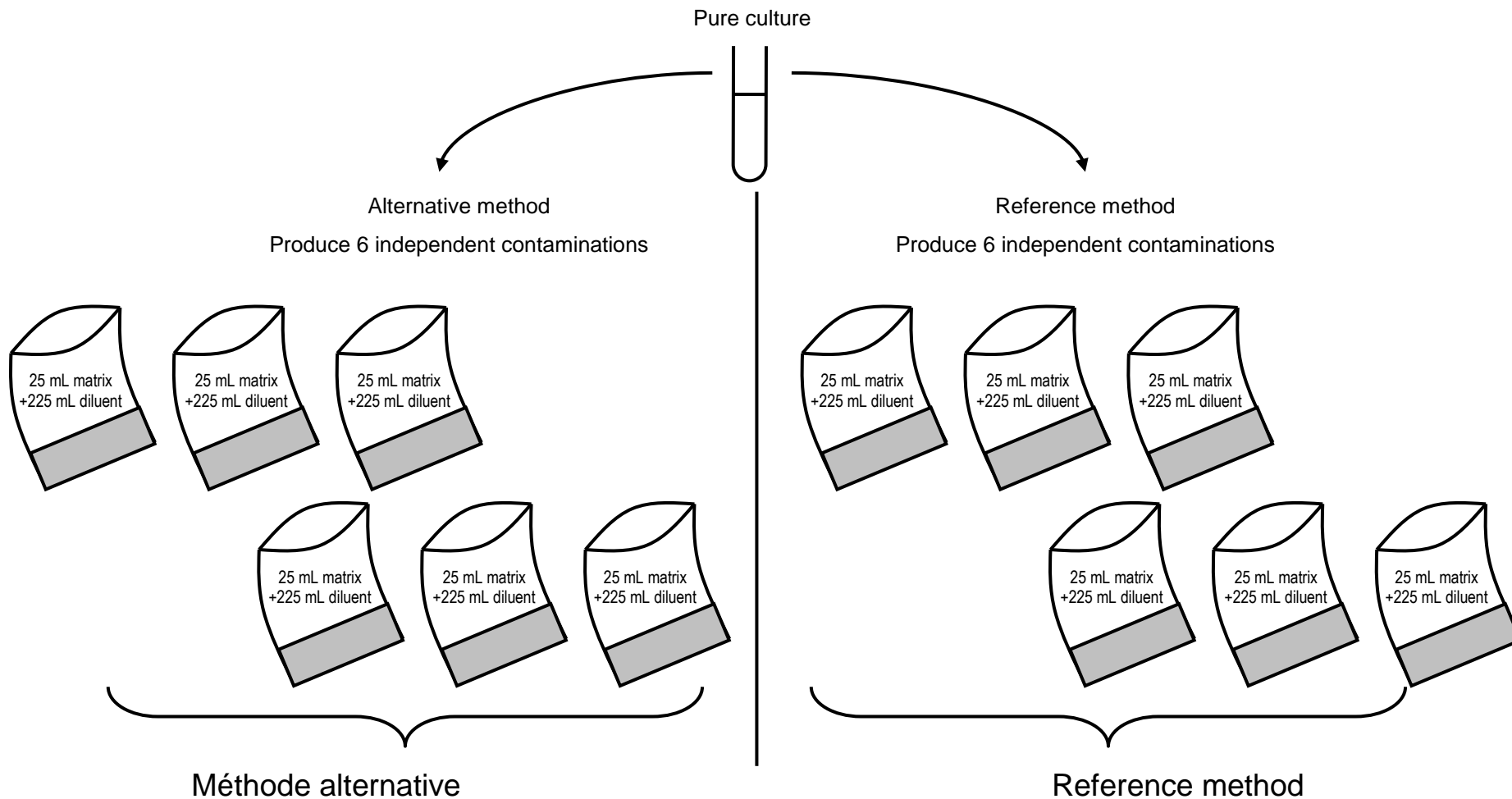
Category	Type	Stress
Water for human consumption	Water from wells, springs and boreholes	none
	Bottled water: spring water, mineral and other water	thermal (hot and/or positive cold)
	Mains and drinking fountain water	chlorination, thermal (hot and/or positive cold)
Bathing water (untreated)	Sea water	thermal (positive cold), osmotic
	Fresh water	thermal (positive cold)
Swimming pool water		chlorination, thermal (hot)
Industrial water	Cooling tower water	biocide
	Sanitary hot water	chlorination, thermal (hot)
	Waste water, raw and processed waste water	chemical (acid and/or basic pH, etc.)
Others (Specific requests from manufacturers)	for example: Thermal waters (spa water, ...)	chlorination, thermal (hot)

Annex 3. Organisation of analyses carried out in parallel depending on the type of method

Methods with a common phase



Methods without common phases



Annex 4. Qualitative methods. Test applied to the rejection of discordant results

Procedure

Count the total number of discordant results RD :

$$RD = PD + ND$$

For example, if $PD = 2$ and $ND = 10$, then $RD = 12$)

IF $RD < 6$ (less than 6 discordant results), no test is available.

IF $6 \leq RD \leq 22$ (between 6 and 22 discordant results), determine m as the smallest of the two values of PD and ND (for example $m = PD = 2$, because $PD < ND$) and use the binomial law in accordance with Table 13.

If $m \leq M$ (M defined in Table 13) for a given Y , the two methods are different at $\alpha < 0.05$ (two-sided).

Table 13. Values of M for Y discordant results ($6 \leq Y \leq 22$)

Discordant results $RD = PD + ND$	6-8	9-11	12-14	15-16	17-19	20-22
$M = \text{Max}(m)$ with $\alpha < 0.05$	0	1	2	3	4	5

For example, with $Y = 12$ discordant results and $m = 2$, $M = 2$ and $m \leq M$: therefore the two methods are different with $\alpha < 0.05$.

if $Y > 22$ (more than 22 discordant results), use the McNemar test with chi-square distribution for 1 degree of freedom:

$$\chi^2 = d^2/Y, \text{ with } d = |PD - ND| \text{ and } Y = PD + ND$$

The two methods are different for $\alpha < 0.05$ (two-sided) if $\chi^2 > 3.841$

This χ^2 test corresponds to the minimal d for each Y of the following Table 14 with $\alpha < 0.05$ (that is, for a given Y , d shall be greater than or equal to the value given in Table 14 for concluding that the two methods are different).

Table 14. Values of d for Y discordant results ($Y > 22$)

Discordant results $Y = PD + ND$	22-26	27-31	32-37	38-44	45-51	52-58
$d = PD - ND \geq$	10	11	12	13	14	15

Annex 5. Qualitative methods. Interlaboratory study, criteria of accordance, concordance and concordance odds ratio

Taken from Annex L of EN ISO 16140

General

The criteria of accuracy, sensitivity and specificity (see 5.2.2) do not really address the variability of the method within and between laboratories (precision of the method).

This annex provides further criteria (accordance, concordance and *odds ratio*) which can help to approach this variability.

Repeatability and reproducibility criteria measure the likely difference between two samples sent to either the same or to different laboratories. Since the difference for data that are not quantitative cannot be used, statistics for qualitative methods are instead based on the probability (expressed as a percentage) that two identical samples both produce the same result.

These criteria have been developed by the European SMT4 CT 96 2098 project funded by the European Commission/ DG XII to validate the six main standardised methods used in food microbiology (co-ordinator: Dr C. Lahellec, AFSSA, France).

The calculations are illustrated by means of the following data (see Table 15).

Table 15. Example of numerical values

Lab	1	2	3	4	5	Number of positives (out of 5)
1	+	+	+	+	+	5
2	+	+	+	+	+	5
3	+	+	+	+	+	5
4	+	+	+	+	+	5
5	-	-	+	+	+	3
6	+	+	+	+	+	5
7	-	-	+	+	+	3
8	+	+	+	+	+	5
9	+	+	+	+	+	5
10	+	+	+	+	+	5

The data shown in this table are just for one level of one food type. In practice, an interlaboratory test would be larger, but a smaller dataset makes it easier to explain the calculations.

Accordance

Definition

Accordance is the percentage chance of finding the same result (i.e. both negative or both positive) from two identical samples analysed in the same laboratory, under repeatability conditions (i.e. one operator using the same apparatus and the same reagents within the shortest feasible time interval).

Accordance is therefore the equivalent of repeatability for quantitative methods.

Calculations

To derive the accordance from the results of an interlaboratory study, the probability that two identical samples give the same result is calculated for each participating laboratory in turn, and this probability is then averaged over all the laboratories.

For laboratories (such as laboratory 1) where all the replicates were found to be positive, the best estimate of the probability of getting the same result is clearly 1.00 or 100 %.

For the other laboratories (in the example, laboratories 5 and 7) the probability that one replicate will be positive is $3/5=0.60$. Square this probability to obtain the probability of a pair of replicates being positive ($0.6^2 = 0.36$). Do the same for the probability of a pair of replicates both being negative ($0.4^2 = 0.16$). Then add these two figures together to get the overall probability that two replicates will give the same result ($0.36 + 0.16 = 0.52$). Do this for all the laboratories (see Table 16).

Table 16. Calculation of accordance

Lab	Number of positives	Probability of positives	Probability of a pair of Positives	Probability of a negative	Probability of a pair of negatives	Probability of a pair of identical results
1	5	1.00	1.00	0.00	0.00	1.00
2	5	1.00	1.00	0.00	0.00	1.00
3	5	1.00	1.00	0.00	0.00	1.00
4	5	1.00	1.00	0.00	0.00	1.00
5	3	0.60	0.36	0.40	0.16	0.52
6	5	1.00	1.00	0.00	0.00	1.00
7	3	0.60	0.36	0.40	0.16	0.52
8	5	1.00	1.00	0.00	0.00	1.00
9	5	1.00	1.00	0.00	0.00	1.00
10	5	1.00	1.00	0.00	0.00	1.00
					Average:	0.904 =90.4%

The accordance is the average of the probabilities that two replicates give the same result for each laboratory: 90.4 % in this case.

NOTE Since repeatability measures difference whereas accordance measures similarity, high values of accordance indicate a reliable method, in contrast to quantitative methods where low values of repeatability are desirable.

Concordance

Definition

Concordance is the percentage chance of finding the same result for two identical samples analysed in two different laboratories.

Concordance is therefore the equivalent of reproducibility for quantitative methods.

Calculations

To calculate concordance from the results of an interlaboratory study, take in turn each replicate in each participating laboratory, and pair it with identical results from all the other laboratories.

Concordance is the percentage of all pairings giving the same results for all the possible pairings of data.

FOR EXAMPLE (see Tables L.1 and L.3 of EN ISO 16140).

Take each replicate in each laboratory in turn, starting with the first replicate from laboratory 1 which is positive. This can be paired with any of the 45 replicates from the other laboratories, and all but 4 of these pairings (those with two replicates from laboratories 5 and 7) match (i.e. a pair with two positives), 41 pairs thus give the same result.

The same applies to the other four replicates from laboratory 1, so there are a total of 225 (5 x 45) inter-laboratory pairings of replicates involving laboratory 1, of which 205 (5 x 41) give the same result.

The same applies to all the other laboratories with all replicates found positive.

For laboratory 5, with 3 positive replicates out of 5, the 2 negative replicates each match with just 2 other negative replicates from laboratory 7, whilst the 3 positive replicates each match with 43 positive replicates. Thus the total number of pairs with the same result is 133 (2 x 2 + 3 x 43).

Table 17. Calculation of concordance

Lab	Number of positives	Interlaboratory pairings with the same result	Total number of interlaboratory pairings
1	5	205	225
2	5	205	225
3	5	205	225
4	5	205	225
5	3	133	225
6	5	205	225
7	3	133	225
8	5	205	225
9	5	205	225
10	5	205	225
Total		1906	2250

Concordance is the percentage of all pairings of replicates giving the same result; in this example this is 84.7 % (1906/2250 x 100).

NOTE Since reproducibility measures difference whereas concordance measures similarity, high values of concordance indicate a reliable method, in contrast to quantitative methods where low values of reproducibility are desirable.

Concordance odds ratio

General remarks and definition

If the concordance is smaller than the accordance, it indicates that two identical samples are more likely to give the same result if they are analysed by the same laboratory than if they are analysed by different laboratories, suggesting that there can be a predominant variability in performance between laboratories. This is the same situation as when reproducibility is greater than repeatability for a quantitative method.

Unfortunately, the magnitude of the concordance and accordance is highly dependent on the level of accuracy, making it difficult to assess the degree of interlaboratory variation easily.

It is therefore helpful to calculate the *concordance odds ratio* (COR) defined as follows:

$$COR = \frac{\text{deg ré d' accord} \times (100 - \text{concordance})}{\text{concordance} \times (100 - \text{deg ré d' accord})} (\text{accordance/concordance})$$

Significance tests

A value for the odds ratio of 1,00 would be expected if accordance and concordance were equal, and the higher the odds ratio, the more interlaboratory variation is predominant.

Nevertheless, values above 1,00 can occur by chance variation, and so a statistical significance test should be used to confirm whether the variation between laboratories is actually predominant. The “exact test” is the one that is most recommended for this purpose¹. The philosophy behind such tests is that the probabilities of occurrence are calculated for all combinations of replicate results that could have been obtained with the overall numbers of positives and negatives.

FOR EXAMPLE (see Table L.1 of EN ISO 16140).

With a total of 46 positives and 4 negatives, the possible arrangements are given by columns in Table L.4 of EN ISO 16140.

Table 18. Possible arrangements of positive replicates between laboratories (columns) to give a total of 46 positives.

4	3	3	2	1
4	4	3	4	5
4	4	5	5	5
4	5	5	5	5
5	5	5	5	5
5	5	5	5	5
5	5	5	5	5
5	5	5	5	5
5	5	5	5	5
5	5	5	5	5

1) This test can be performed using statistical software such as SAS®.

The actual arrangement is shown in bold (third column) in Table M.4. The test adds up the probabilities P of all the possible arrangements that show at least as much evidence for interlaboratory variation as the real arrangement – here they correspond to the permutations in the two columns on the right. If this sum of probabilities is less than the conventional value of 0,05 or 5 %, it is unlikely that this degree of interlaboratory variation could have occurred by chance and thus it is concluded that there is a significant variation in performance between laboratories.

In the example of Table M.4, P = 0,039 indicates that the variation between laboratories is significant at a level of 5 %.

Where “exact test” software is not available, an ordinary χ^2 analysis for the contingency tables can be used as an alternative. The results of this test are less reliable than the “exact test” with the number of replicates usually used in collaborative studies, but simulations suggest that the results provide a reasonable guide to the significance of interlaboratory differences.

With either test, it shall be remembered that the ability to detect interlaboratory differences is dependent on the number of laboratories and the number of replicates analysed at each laboratory. A non-significant test result should not be taken to mean that performance does not vary between laboratories, but rather that such differences have not been proved; this is particularly true where the P-value is only just above 0,05. The ideal solution would be to quote the odds ratio with the standard deviation or confidence limits, but the distribution of the odds ratio is highly skewed making it very difficult to estimate reliable limits.

Annex 6. Quantitative methods. Calculation of repeatability and reproducibility standard deviations

This annex is based on clause 7 of ISO 5725-2. The calculations are conducted according to the principle described in ISO 5725. They are applied at each level j to data z . As a result, the number of replicates per level is considered to be constant and equal to K .

Calculate for each level j :

- the repeatability variance, recorded as $s(j)^2_r$.
- the inter-series variance, recorded as $s(j)^2_B$.
- the intermediate precision variance $s(j)^2_{PI} = s(j)^2_r + s(j)^2_B$.

In order to simplify scoring, the index j for the level is deleted from the formulae that follow.

The starting point consists of writing the following equality which breaks down the sum of the deviations from the general average into two sums of squared deviations. n_k represents...

$$\sum_{k=1}^K \sum_{i=1}^I (z_{ik} - \bar{z})^2 = \sum_{k=1}^K (\bar{z}_k - \bar{z})^2 + \sum_{k=1}^K \sum_{i=1}^I (z_{ik} - \bar{z}_k)^2 \quad (8)$$

This equation is traditionally written in an abbreviated form, involving three sums of squared deviations (SSD). General equation for variance analysis

$$SSD_t = SSD_B + SSD_r$$

Where each of the sums is defined as follows to facilitate interpretation:

SSD_t Sum of total deviations from the level's general mean

SSD_B Sum of the inter-series deviations

SSD_r Sum of the intra-series deviations

As explained in ISO 5725, the sums of squares are developed to make calculation easier; SSD_B is always calculated by difference. This method poses problems if the result is negative: the value of SCE_B is then forced to 0.

$$SSD_B = SSD_t - SSD_r$$

if $SSD_B > 0$
if $SSD_B \leq 0$ then $SSD_B = 0$

NOTE An alternative to this method of calculation is to use an algorithm to estimate s^2_B using the restricted maximum likelihood method (REML) available in most statistical software.

Calculate the repeatability variance (intra-series) for the level:

$$s_r^2 = \frac{1}{(I-1)K} \sum_{i=1}^I \sum_{k=1}^K (z_{ik} - \bar{z}_k)^2 = \frac{1}{(I-1)K} SCE_r \quad (9)$$

Calculate the inter-series variance in the following manner:

$$s_B^2 = \frac{(K-1) \left(\frac{SCE_B}{K-1} - s_r^2 \right)}{I} \quad (10)$$

Finally the intermediate precision variance for the level is equal to: $s_R^2 = s_r^2 + s_B^2$ (11)

Annex 7. Example of an accuracy profile calculation

Lab	Series	Level	Alternative (cfu)	Reference (cfu)	Alter log.	Ref log.
A	1	1	60	120	1.778	2.079
A	1	1	40	75	1.602	1.875
B	2	1	70	95	1.845	1.978
B	2	1	90	85	1.954	1.929
C	3	1	80	90	1.903	1.954
C	3	1	50	80	1.699	1.903
D	4	1	90	90	1.954	1.954
D	4	1	170	120	2.230	2.079
E	5	1	40	90	1.602	1.954
E	5	1	70	130	1.845	2.114
F	6	1	50	130	1.699	2.114
F	6	1	120	65	2.079	1.813
G	7	1	90	95	1.954	1.978
G	7	1	130	110	2.114	2.041
H	8	1	90	100	1.954	2.000
H	8	1	110	110	2.041	2.041
I	9	1	140	70	2.146	1.845
I	9	1	90	80	1.954	1.903
J	10	1	90	120	1.954	2.079
J	10	1	110	110	2.041	2.041
L	11	1	90	140	1.954	2.146
L	11	1	120	110	2.079	2.041
M	12	1	110	95	2.041	1.978
M	12	1	130	130	2.114	2.114
N	13	1	90	95	1.954	1.978
N	13	1	150	80	2.176	1.903
O	14	1	110	65	2.041	1.813
O	14	1	50	95	1.699	1.978
A	1	2	810	1300	2.908	3.114
A	1	2	520	1200	2.716	3.079
B	2	2	940	1000	2.973	3.000
B	2	2	950	1100	2.978	3.041
C	3	2	1000	1000	3.000	3.000
C	3	2	1100	1000	3.041	3.000
D	4	2	1000	1000	3.000	3.000
D	4	2	900	940	2.954	2.973
E	5	2	940	930	2.973	2.968
E	5	2	920	1100	2.964	3.041
F	6	2	810	1000	2.908	3.000
F	6	2	1000	1100	3.000	3.041
G	7	2	910	910	2.959	2.959
G	7	2	1000	900	3.000	2.954
H	8	2	1200	1100	3.079	3.041
H	8	2	1100	1000	3.041	3.000
I	9	2	1100	1000	3.041	3.000
I	9	2	980	1100	2.991	3.041
J	10	2	1200	950	3.079	2.978

J	10	2	1100	1100	3.041	3.041
L	11	2	830	1000	2.919	3.000
L	11	2	1300	1000	3.114	3.000
M	12	2	1100	1000	3.041	3.000
M	12	2	830	1200	2.919	3.079
N	13	2	660	940	2.820	2.973
N	13	2	810	980	2.908	2.991
O	14	2	640	820	2.806	2.914
O	14	2	940	1200	2.973	3.079
A	1	3	9300	9700	3.968	3.987
A	1	3	9900	9700	3.996	3.987
B	2	3	7800	9400	3.892	3.973
B	2	3	9700	9000	3.987	3.954
C	3	3	11000	10000	4.041	4.000
C	3	3	11000	12000	4.041	4.079
D	4	3	9500	11000	3.978	4.041
D	4	3	10000	11000	4.000	4.041
E	5	3	11000	9600	4.041	3.982
E	5	3	9400	10000	3.973	4.000
F	6	3	8900	11000	3.949	4.041
F	6	3	9500	10000	3.978	4.000
G	7	3	12000	13000	4.079	4.114
G	7	3	13000	12000	4.114	4.079
H	8	3	9400	12000	3.973	4.079
H	8	3	9500	9400	3.978	3.973
I	9	3	11000	12000	4.041	4.079
I	9	3	12000	12000	4.079	4.079
J	10	3	12000	12000	4.079	4.079
J	10	3	11000	9500	4.041	3.978
N	13	3	9500	12000	3.978	4.079
N	13	3	10000	13000	4.000	4.114
O	14	3	9000	7600	3.954	3.881
O	14	3	9500	8500	3.978	3.929

Table 19. Results of the interlaboratory study

Criteria	Level 1	Level 2	Level 3
Target value	1.95	2.97	4.00
Alternative method average	0.018	0.027	0.016
Lower tolerance limit	-0.196	-0.117	-0.109
Upper tolerance limit	0.232	0.171	0.140
Lower limit of acceptability	-0.3	-0.3	-0.3
Upper limit of acceptability	0.3	0.3	0.3

Construct the accuracy profile like the one illustrated in **Erreur ! Référence non valide pour un signet..**

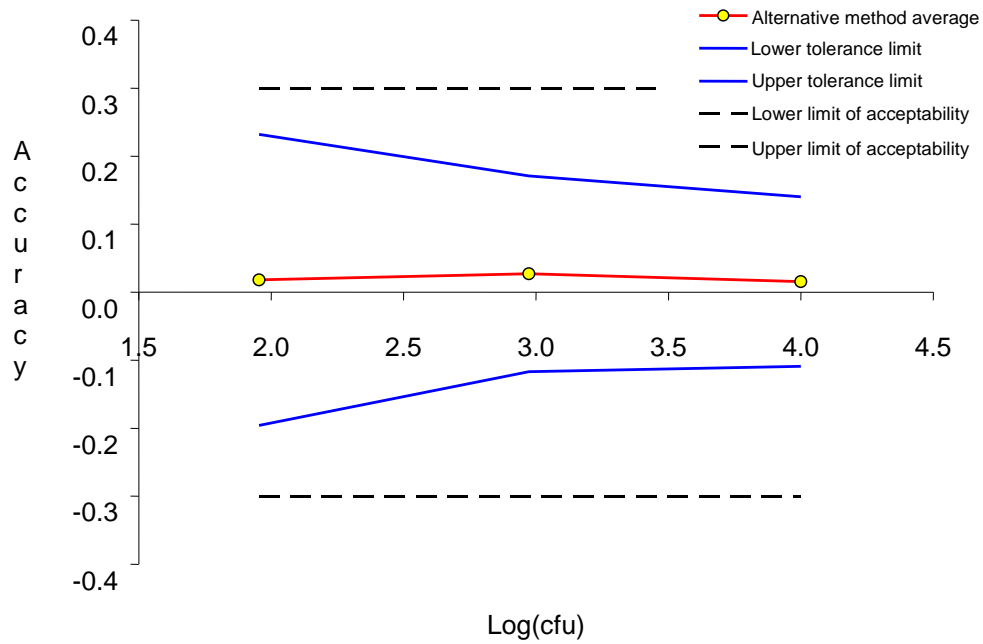


Figure 4. Accuracy profile The limits of acceptability have been chosen at ± 0.3 log

To show the importance of the acceptability criterion,

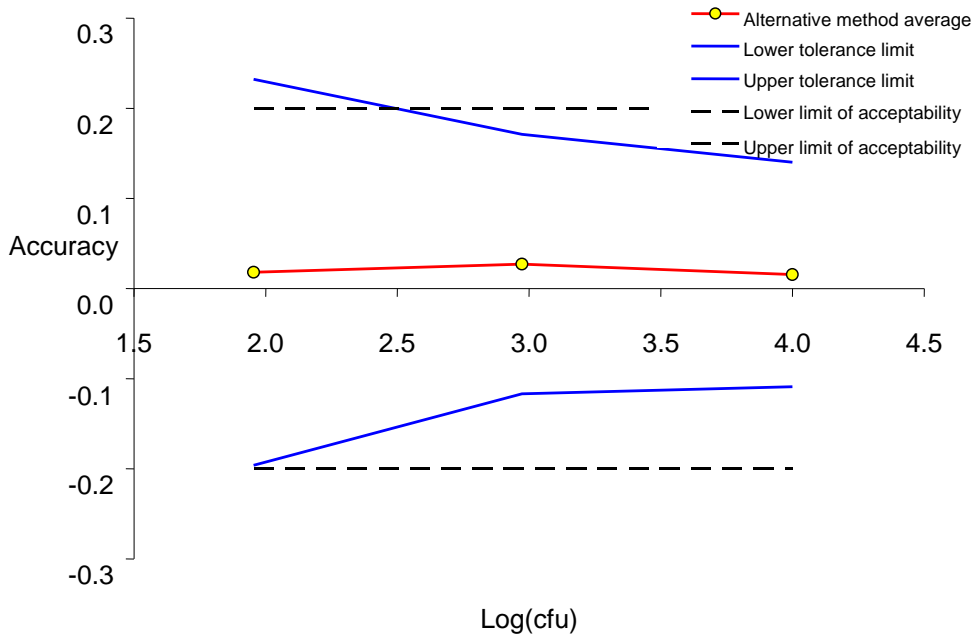


Figure 5 illustrates the situation where the acceptability limits were chosen at ± 0.2 log. The LOQ is then 2.51 log units.

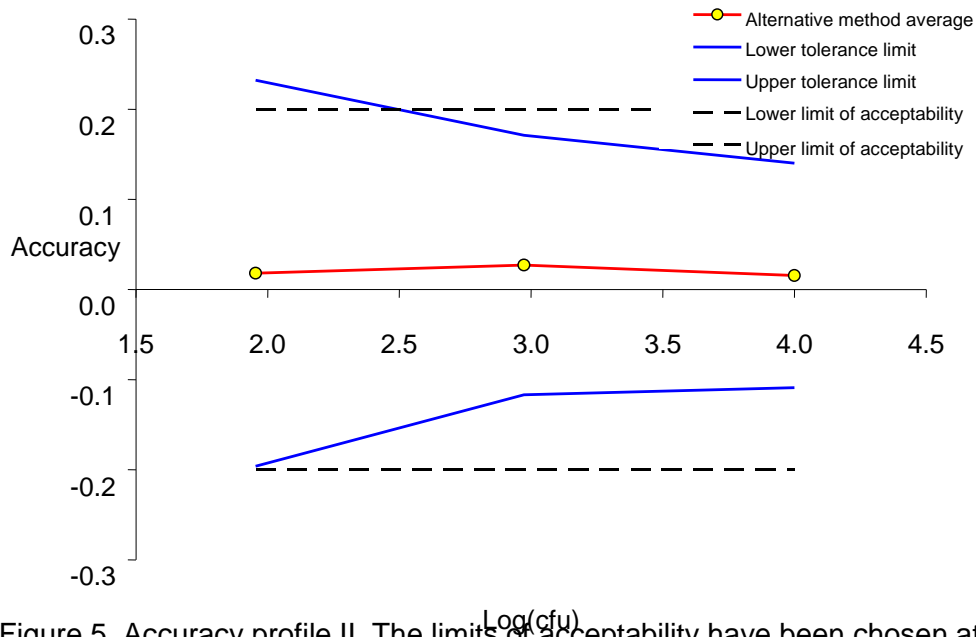


Figure 5. Accuracy profile II. The limits of acceptability have been chosen at ± 0.2 log

Annex 8. Guidelines for the organisation and conducting of interlaboratory studies

This annex is based on Annex H (normative) of EN ISO 16140.

Preparation of water samples

Negative (control) samples

The target germ shall not be present in the water matrix. Choose a suitable number of samples in accordance with 6.1 or 6.2 in this document and analyse them to confirm this. Negative controls shall also be used during tests where stipulated.

Positive samples containing the target germ.

In the case of **qualitative** and **quantitative** studies, samples shall be contaminated artificially as shown in options 2 and 3 of Annex C of EN ISO 16140. Where reference materials are used, it is the responsibility of the organising laboratory to send both the reference materials and details of the inoculation procedure to each collaborating laboratory.

The sample and microorganisms shall be capable of being homogenised and should remain stable both in transit and for the duration of the analyses. Homogeneity and stability studies shall be conducted prior to despatching the test material from the organising laboratory.

The water sample should contain representative background microflora or interfering components that shall also remain stable in transit and for the duration of the analyses.

An alternative method shall be tested using a minimum of one water type, chosen from the categories defined. The quality and suitability of the water samples shall be determined prior to the study.

Based on the comparative study, the characteristics of the chosen microbial strain being used to validate a method shall be representative of the genus/species being sought, for example growth rate, antigenic characteristics and sensitivity to inimical agents etc.

Transport of samples

The organising laboratory defines the type of water to be used and checks that the transit packaging is suitable for the intended journey.

Individual samples shall be double sealed to ensure that leakage from any one sample does not affect the integrity of all the other samples.

Each collaborating laboratory is responsible for supplying the organising laboratory with relevant details that ensure that the distribution of samples for the study meets international and national postal regulations.

Concerning refrigeration, the following conditions apply: the temperature of samples during transport shall be lower than or equal to 10 °C, and, on arrival at the laboratory, between 0 °C and 10 °C.

The packaging conditions and best method of transport shall be determined by the organising laboratory. An appropriate means of monitoring the samples during transit is desirable. An additional pack, identical to the test samples, shall be provided in order to measure their temperature upon receipt.

Organisation of the interlaboratory study

Operating protocols: operating procedures shall be distributed to collaborators for comment familiarisation prior to the commencement of the test. A final protocol shall be issued in advance of the evaluation.

Confirmation of sample quality: each collaborating laboratory should be instructed to undertake enumeration of the total bacterial flora in a specific additional sample prepared by the expert laboratory. The organising laboratory shall provide each collaborator with a standard enumeration method.

Alternative method and reference method: the collaborator responsible for carrying out the analysis in each participating laboratory shall be trained in advance (if necessary) and shall be fully conversant with the alternative method and the reference method prior to the commencement of the test.

Reagents and operating conditions: certain additional factors such as the quality and composition of the culture media and reagents, control of incubation temperatures, etc. can have a profound effect on the outcome of the test. This variability shall be minimised (for example, by dispatching media/reagents to all collaborators) or be taken into consideration when interpreting data.

The degree of tolerance permitted for the different aspects of the germ (for example, times, temperatures, masses, total plate count and day of analysis), shall be stated in the protocol. A clear warning shall be issued if no tolerances are permitted.

Guidelines: the organising laboratory shall be available to provide advice or guidance to collaborators during the whole period of the study.

Data collection: the organising laboratory shall prepare a questionnaire that can be used to collect information concerning the critical points of the procedure. All collaborators should be asked to record details such as media pH, times of incubation, incubator start and finish temperatures, quality control data, state of samples on arrival, time of arrival, temperature on arrival, storage condition/times, etc. The questionnaire shall be the subject of agreement by both the organising laboratory and the manufacturer, in order to guarantee its suitability for the critical phases of the test or test procedure of the alternative method.

Confirmation of the quality of test samples: representative aliquots of the samples should be analysed by the organising laboratory on the day of the distribution in order to confirm the presence and homogeneity of the target germ. Samples, previously held under ideal conditions, should also be analysed by the organising laboratory on the day of the start of analysis by the collaborating laboratories. These analyses shall be used only to confirm the inoculation procedure and stability of the germ. The results should not be used in the statistical evaluation of the method under test.

The fact of using branded media and different batches introduces additional variability.