INTERNATIONAL STANDARD

ISO 16140

First edition
2003-05-01

Microbiology of food and animal feeding stuffs — Protocol for the validation of alternative methods

Microbiologie des aliments — Protocole pour la validation des méthodes alternatives

Reference number
ISO 16140:2003(E)

ISO copyright office
Case postale 56 • CH-1211 Geneva 20
Tel. + 41 22 749 01 11
Fax + 41 22 749 09 47
E-mail copyright@iso.org
Web www.iso.org

Published in Switzerland

Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardization.

International Standards are drafted in accordance with the rules given in the ISO/IEC Directives, Part 2.

The main task of technical committees is to prepare International Standards. Draft International Standards adopted by the technical committees are circulated to the member bodies for voting. Publication as an International Standard requires approval by at least 75 % of the member bodies casting a vote.

Attention is drawn to the possibility that some of the elements of this International Standard may be the subject of patent rights. ISO shall not be held responsible for identifying any or all such patent rights.

ISO 16140 was prepared by the European Committee for Standardization (CEN) in collaboration with Technical Committee ISO/TC 34, *Food products*, Subcommittee SC 9, *Microbiology*, in accordance with the Agreement on technical cooperation between ISO and CEN (Vienna Agreement).

Throughout the text of this document, read "...this European Standard..." to mean "...this International Standard...".

Contents

Foreword

This document (EN ISO 16140:2003) has been prepared by Technical Committee CEN/TC 275 "Food analysis - Horizontal methods", the secretariat of which is held by DIN, in collaboration with Technical Committee ISO/TC 34 "Agricultural food products".

This European Standard shall be given the status of a national standard, either by publication of an identical text or by endorsement, at the latest by November 2003, and conflicting national standards shall be withdrawn at the latest by November 2003.

The annexes A, C to K and M to T are normative. The annexes B, L and U are informative.

This document contains also a Bibliography.

According to the CEN/CENELEC Internal Regulations, the national standards organizations of the following countries are bound to implement this European Standard: Austria, Belgium, Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Luxembourg, Malta, Netherlands, Norway, Portugal, Slovakia, Spain, Sweden, Switzerland and the United Kingdom.

Introduction

The need for the food industry to rapidly assess the microbiological quality of raw materials and finished products and the microbiological status of manufacturing procedures, has led to the development and refinement of alternative microbiological methods of analysis that are quicker and/or easier to perform than the corresponding reference method; some can also be automated.

Among these alternative methods, some can yield results that are equivalent to those provided by the reference method, while others can lead to results that differ appreciably.

The suppliers/producers of the alternative methods, the food and drink industry, the public health services and other authorities need a reliable common protocol for the validation of such alternative methods. The data generated can also be the basis for the certification of a method by an independent organisation.

Because of the extent of the methods comparative study described in this standard for use by the organising laboratory, the procedure is sometimes not appropriate for use as an "in house" method for the validation of an alternative method by an individual laboratory.

1 Scope

This document establishes the general principle and the technical protocol for the validation of alternative methods in the field of microbiological analysis of food, animal feeding stuff and environmental and veterinary samples (see 5.1.1.2.1) for:

— the validation of alternative methods which can be used in particular in the framework of the official control;

— the international acceptance of the results obtained by the alternative method.

It also establishes the general principles of certification of these alternative methods, based on the validation protocol defined in this document (see 4.2).

Where an alternative method is used on a routine basis for internal laboratory use without the requirement to meet (higher) external criteria of quality assurance, a less stringent comparative validation of the alternative method than that set in this standard may be appropriate.

2 Normative references

This European Standard incorporates by dated or undated reference, provisions from other publications. These normative references are cited at the appropriate places in the text, and the publications are listed hereafter. For dated references, subsequent amendments to or revisions of any of these publications apply to this European Standard only when incorporated in it by amendment or revision. For undated references the latest edition of the publication referred to applies (including amendments).

ISO 3534-1, *Statistics – Vocabulary and symbols – Part 1: Probability and general statistical terms.*

ISO 5725, *Accuracy (trueness and precision) of measurement methods and results.*

3 Terms and definitions

For the purposes of this European Standard, the following terms and definitions apply:

3.1
alternative method
method of analysis that demonstrates or estimates, for a given category of products, the same analyte (3.4) as is measured using the corresponding reference method (3.2).

NOTE 1 The method can be proprietary or non commercial, and does not need to cover an entire analysis procedure, that is from the preparation of samples to the test report.

NOTE 2 The alternative method exhibits attributes appropriate to the users' needs, for example:

— speed of analysis and/or response;

— ease of execution and/or automation;

— analytical properties (precision, accuracy, limit of detection, etc.);

— miniaturisation;

— reduction of cost.

NOTE 3 The term "alternative" is used to refer to the entire "test procedure and reaction system". This term includes all ingredients whether material or otherwise, required for implementing the method.

3.2
reference method
internationally recognised method and widely accepted.

NOTE For the purpose of this standard, these are International and European Standards and if not existing, certain national standards of equivalent standing.

3.3
validation of an alternative method
demonstration that adequate confidence is provided that the results obtained by the alternative method are comparable to those obtained using the reference method

NOTE The word "comparable" is defined in this EN ISO 16140 by a technical protocol adapted to each type of method (see clauses 5 and 6).

3.4
analyte
component measured by the method of analysis. It may be the microorganism

3.5
qualitative method
method of analysis whose response is either the presence or absence of the analyte (3.4) detected either directly or indirectly in a certain amount of sample

3.6
quantitative method
method of analysis whose response is the amount of the analyte (3.4) measured either directly (enumeration in a mass or a volume), or indirectly (colour absorbance, impedance, etc.) in a certain amount of sample

3.7
methods comparison study
study, performed by the organising laboratory of the alternative method against the reference method

3.8
inter-laboratory study
study of the method's performance using common samples in several laboratories and under the control of the organising laboratory

3.9
organising laboratory
laboratory having the qualified staff and skills to perform the method comparison study and organise the interlaboratory study.

NOTE The availability of an experienced statistician is essential for the analysis of the results.

4 General principles for the validation and the certification of alternative methods

4.1 Validation protocol

The validation protocol comprises two phases:

— a methods comparison study (3.7) of the alternative method (3.1) against the reference method (3.2) carried out in the organizing laboratory;

— an interlaboratory study (3.8) of each of the two methods.

If appropriate, the two phases may be undertaken in parallel.

The technical rules for performing the methods comparison study and the interlaboratory study are given in clauses 5 and 6, depending upon whether the alternative method is qualitative or quantitative in nature.

If the alternative method has already been validated and meets the requirements set by another organisation, specific rules are defined in annex A for accepting the results of this prior validation.

4.2 Principles of the certification

4.2.1 If a subsequent certification of the alternative method is required, the two following principles shall also be applied (in addition to 4.1):

Details on the organisation of the certification (management of the method comparison study and the interlaboratory study, all the different bodies involved including the expert laboratory – designated in this standard as the "organising laboratory"- the reviewers, the certification body, etc) are provided [8] by the certification body.

4.2.2 The manufacturer shall apply a **quality system** covering the production line of the product for which the certification is sought and based on the appropriate European Standard relative to quality systems or other equivalent international standard (for example EN ISO 9001).

In granting the certification, the certification organisation shall take into account the existence of any quality system certificate issued by a certification body accredited for quality systems.

4.2.3 A **regular verification** of the quality of the certified method shall be undertaken after the certification is granted. An audit is to be performed regularly to verify that the following are still met:

— the quality assurance requirements, (see 4.2.1);

— the product's production control requirements, (see 4.2.1).

In addition to the general requirements of the appropriate European Standard relative to the quality system, the manufacturer presents regularly to the certification organisation updated documentation that take into account any modification made to the product or production process which may affect the instructions for using the method and/or the method's performance. The certification organisation then decides whether these modifications affect the certification.

5 Qualitative methods - Technical protocol for their validation

5.1 Methods comparison study

5.1.1 Relative accuracy, relative specificity and relative sensitivity

5.1.1.1 Terms and definitions

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

5.1.1.1.1 relative accuracy (*AC*)

degree of correspondence between the response obtained by the reference method and the response obtained by the alternative method on identical samples[1)] (see 5.1.1.3.1.).

NOTE The term "relative accuracy" used here is complementary to the "accuracy" and "trueness" as defined in ISO 5725-1 and ISO 3534-1. These state that accuracy is "the closeness of agreement between a test result and the accepted reference value", and that the trueness is "the closeness of agreement between the average value obtained from a large series of test results and an accepted reference value". For the purpose of this standard, the accepted reference value is chosen as the value obtained by the reference method. Thus, the term "relative" implies that the reference method does not automatically provide the accepted reference value.

5.1.1.1.2 positive deviation (PD)

The alternative method becomes a false positive when it 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.

5.1.1.1.3 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.

---

1) Difficult to achieve if the pre-enrichment steps are different.

5.1.1.1.4 relative sensitivity (*SE*)

ability of the alternative method to detect the analyte when it is detected by the reference method (see 5.1.1.3.1.).

5.1.1.1.5 relative specificity (*SP*)

ability of the alternative method to not detect the analyte when it is not detected by the reference method (see 5.1.1.3.1.).

5.1.1.2 Measurement protocol

5.1.1.2.1 Food samples

It is of the highest priority to find food samples naturally contaminated with the analyte to be detected for the validation.

If it is sought to validate the method for all foods, study five categories of food. This number may be reduced to 1, 2, 3 or 4 categories if the validation of alternative method is restricted to these stated categories, at the producer's request. The recommended categories are listed in annex B.

Appropriate environmental samples may be included as one category. Veterinary samples may be treated as another category (see annex B).

It is desirable that food samples come from as wide a distribution as possible in order to reduce any bias from local food specialities and broaden the range of validation.

When analysing naturally contaminated samples, the range and distribution of contamination of the samples should be representative of the levels usually found in that product but with emphasis on smaller numbers.

If it is not possible to acquire a sufficient number of naturally contaminated foods for each of the categories, artificial contamination of food samples is permissible. The method and levels of contamination should result in samples behaving similarly to naturally contaminated ones. See methods of inoculation and restrictions in annex C.

5.1.1.2.2 Number of samples

The total number of test portions to be analysed is 60 for each food category chosen from the categories stated in annex B. Within each category, select representative food types and analyse 20 test portions of each food type by the proposed method and the reference method to produce at least 60 total results for each category by each method. For naturally contaminated food types prepare the sample as described in annex D. For artificially contaminated food type adjust the inoculation levels to achieve fractional positive recovery of the test portions analysed by at least one of the methods. Fractional recovery is achieved when some number, but not all, of the test portions are determined to be positive by one or both methods, alternative method or reference method.

It is desirable to produce approximately 50 % of the results that are positive and 50 % that are negative. This is, however, a recommendation, not an absolute percentage, provided that some number of the test portions are positive and some number are negative for the same food type.

5.1.1.2.3 Test sample preparation

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

Thus, if the first stage of the two methods is the same (for example the same pre-enrichment broth), perform the replication at the second step (case 1, annex D).

If this is not the situation, that is the first culture media, methodology or dilutions are different, prepare paired test portions for analysis. There are two primary methodologies for such preparations.

In the first instance, mix a double weight of sample with an equal weight/volume of sterile water or other suitable diluents and homogenize very thoroughly. Then divide into two portions taking particular care to increase the concentration of the primary enrichment by (approximately 10 %) to compensate for the dilution effect of the diluted, homogenised sample (case 2, annex D).

In the second instance, directly inoculate the food type with a starting inoculum sufficient to allow a fractional recovery of the micro-organisms in the test portions analysed by at least one of the methods after the microorganisms have equilibrated in the food type. Then weigh 25 g test portions and proceed as described in annex D. This may be preferred for liquid products but is acceptable for any food type provided that the food is properly homogenised.

5.1.1.3 Calculation and interpretation

5.1.1.3.1 Treatment of data

Tabulate the data of the paired results of the reference and alternative methods and calculate the following parameters for each food category (60 samples) according to the Table 1.

Table 1 - Paired results of the reference and alternative method

Responses	Reference method positive (R+)	Reference method negative (R-)
Alternative method positive (A+)	+/+ positive agreement (PA)	-/+ positive deviation (PD) (R-/A+)
Alternative method negative (A-)	+/- negative deviation (ND) (A-/R+)	-/- negative agreement (NA)

The calculations shall be performed on a number of negative results obtained by the reference method which for the results in Table 1 cannot exceed twice the number of positive results; the negative results being selected if necessary as immediately following a positive result, in the order of analysis of the samples.

Express the three criteria as follows:

— **Relative accuracy:** $AC = \frac{(PA + NA)}{N} \times 100\,\%$;

— **Relative specificity:** $SP = \frac{NA}{N_-} \times 100\,\%$;

— **Relative sensitivity:** $SE = \frac{PA}{N_+} \times 100\%$

where

N is the total number of samples (NA + PA + PD +ND);

N_- is the total number of negative results with the reference method (NA + PD);

N_+ is the total number of positive results with the reference method (PA + ND).

5.1.1.3.2 Confidence intervals

The calculation of confidence intervals associated with the number of samples tested is given in annex E.

5.1.1.3.3 Discordant results

Examine the discordant results as described in annex F (The McNemar test), by using the count of PD and ND (see 5.1.1.3.1).

When the values for PD and ND are high and almost equal, no statistical difference between the methods can be detected using the McNemar test. In this case, the organising laboratory shall pay further attention to explain the reasons for the high values of PD and ND. Moreover, it shows that the relative accuracy of a method shall never be interpreted by taking into account only the McNemar test.

5.1.1.3.4 Summary of calculation

All the calculations shall be summarised in Table 2:

Table 2 - Calculation of the relative accuracy, the relative sensitivity and the relative specificity

Matrices	PA	NA	ND	PD	Sum	Relative Accuracy *AC* (%)	N_+	Relative sensitivity *SE* (%)	N_-	Relative specificity *SP* (%)
					N	$\frac{100\times(PA+NA)}{N}$	$PA+ND$	$\frac{100\times PA}{N+}$	$NA+PD$	$\frac{100\times NA}{N-}$
Food cat. 1										
Food cat. 2										
Food cat. 3										
Food cat. 4										
Food cat. 5										
TOTAL										

5.1.1.3.5 Interpretation

A table giving the raw results (that is **all** the positive and negative results, Table 1) shall be provided.

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 true positive results than the reference method is evaluated.

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

The procedure for the artificial contamination of test samples shall be described in the report of the study.

Data published elsewhere and meeting the conditions defined in annex A may be used for evaluating the relative accuracy.

5.1.2 Relative detection level

5.1.2.1 Definition

For the purpose of this standard, the relative detection level is the smallest number of culturable microorganisms (3.4) that can be detected in the sample in 50 % of occasions by the alternative and reference methods.

5.1.2.2 Measurement protocol

Test the following:

— use one food product within each food category chosen from 5.1.1.2.1, depending of the scope of the validation (see annex B);

— use five different target microorganisms (or less, depending on the scope of the validation) each one associated with one food category, if possible. (See annex G.1 for the definition of the target microorganism);

— preferably test five levels (but a minimum of three levels) of one target microorganism per food, including the negative control, etc. The first level shall be the negative control. The second level shall be the theoretical detection level. The third level shall be just above the theoretical detection threshold and any further levels shall be higher than the previous one. A factor of about three between each concentration in the upper levels could be applied;

— replicate each combination (food product, level of contamination) six times by both the alternative and reference methods. Perform the division at the level where the two methods differ as illustrated in annex D. Thus, if the 1st stage of each method is the same (for example the same pre-enrichment broth), perform the division at the 2nd step (case 1, annex D). If this is not the case, i.e. the first culture media, methodology or dilutions being different, mix a double weight of sample with an equal w/v of sterile water or other suitable diluent and then divide into two portions;

— apply the complete procedure of the alternative method and the reference method, including the preparation of the sample. Inoculation of each food sample may be prior to its addition to the culture medium or afterwards.

If necessary, for assuring a better precision of the lowest inoculum level, increase the amount of food sample or the number of replicate samples. For example, 75 g of food sample contaminated with three cells instead of 25 g contaminated with one cell.

The greater the number of inoculum levels used the more precise is the determination of the detection threshold.

5.1.2.3 Calculation

For each level L_i (i = 0 to 3) and each food/strain combination (j = 1 to 5), compare both methods as stated in Table 3:

Table 3 - Calculation of relative detection level

		Results		
		Negatives (-)	Positives (+)	Total
Method	Reference	a	$n - a$	n=6
	Alternative	b	$n - b$	n=6
	Total	$a+b$	$2n - (a + b)$	$2n$=12

For small 2 by 2 tables, perform exact Fisher tests [8].

Comparisons

Instead of only comparing both methods at each level and each food/strain, the same test to compare two food/strains at the same level can be used.

If food/strains seem to be comparable, the same test is available with $n > 6$ in pooling food/strains for each level L_i.

The levels can also be pooled to do checks, but using the ranking order: $L_0 + L_1$, $L_0 + L_1 + L_2$, $L_1 + L_2$, $L_0 + L_1 + L_2 + L_3$, $L_1 + L_2 + L_3$, $L_2 + L_3$… with or without pooling the food/strains.

Report all the significant differences between methods, food/strains and/or levels.

5.1.2.4 Interpretation

The interpretation shall be done by the organising laboratory in charge of the methods comparison study.

The relative detection level lies between the two contamination levels giving respectively less and more than 50 % detection level. The relative detection level is therefore expressed as a range.

5.1.3 Inclusivity and exclusivity

5.1.3.1 Definition

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

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

5.1.3.2 Measurement protocol

5.1.3.2.1 Selection of test strains

5.1.3.2.1.1 General

For microorganisms a range of strains is chosen to avoid any local bias.

Criteria for selecting test strains are given in annex G.

Each strain shall be characterised biochemically, serologically and if relevant genetically, in sufficient detail for its identity to be established and should be preferentially isolated from food. Also the food material from which it was originally isolated shall be known and recorded.

5.1.3.2.1.2 Target microorganisms

Select at least 50 pure cultures of microorganisms relevant to the alternative method and the food product being used (see G.3), except for *Salmonella*.

For *Salmonella* methods, select at least 30 pure cultures of microorganisms.

5.1.3.2.1.3 Non-target microorganisms

Select at least 30 pure cultures of microorganisms chosen from both the strains known to cause interference with the target microorganism and from strains naturally present in each food test material included in the validation (see G.4).

5.1.3.2.2 Inoculation

5.1.3.2.2.1 General

Each test is performed once. Inoculation of the growth medium is carried out using a dilution of a pure culture of each test strain. No food sample is added.

5.1.3.2.2.2 Target microorganisms

The inoculum level shall be 10 times to 100 times greater than the minimum relative detection level of the alternative method and the complete protocol of the alternative method shall be used, including pre-enrichment if stipulated. When false negative or doubtful results are obtained, the strain shall be tested once more together with the reference method.

5.1.3.2.2.3 Non-target microorganisms

The inoculum level of a strain shall be similar to the greatest level of contamination expected to occur in all the food categories being used.

The exclusivity shall be established. If the final nutrient medium of the culture medium is a selective broth this would be replaced by an appropriate non selective broth medium. When the alternative method gives positive or doubtful results with non-target microorganisms, the test shall be repeated using the complete protocol. The reference method shall be performed only once.

5.1.3.3 Expression of the results

Tabulate the results as in Table 4:

Table 4 - Presentation of the results for the selectivity

Microorganisms	Results			
	Reference method		Alternative method	
	Expected result	Actual result	Expected result	Actual result
Target strains				
1				
2				
etc.				
Non target strains				
1				
2				
etc.				

5.1.3.4 Interpretation

The interpretation shall be done by the laboratory in charge of the methods comparison study, as both quantitative and qualitative aspects (that is pathogenicity, prevalence, cultural aspects of test strains, for example motility, sensitivity to inimical agents etc.) shall be taken into account.

Other published data on the alternative method that meets the requirements of this EN ISO 16140, may also be used by the laboratory in charge of the methods comparison study, to provide further information on the above criteria. (See annex A that provides criteria for the acceptance of external results).

5.2 Interlaboratory study

NOTE The aim of the collaborative study is to determine the variability of the results obtained in different laboratories using identical samples and to compare these results with those obtained in the methods comparison study.

5.2.1 Measurement protocol

5.2.1.1 The interlaboratory study shall produce at least 10 collaborative laboratories having results without outliers.

Guidelines and requirements for the organisation, dispatching and conducting the interlaboratory studies are given in annex H.

It is necessary for the analyst in each collaborating laboratory to demonstrate his competence in the use of the alternative method and of the reference method prior to participating in the study proper.

5.2.1.2 The protocol is the following:

— one relevant food matrix (annex B) is used to prepare the test samples;

— the protocol for artificial contamination of the food sample shall be appropriate for the selected food substrate. Each sample shall be individually inoculated. Each blind replicate shall be prepared to ensure homogeneity between samples by the individual inoculation of each one. An appropriate number of samples shall be analysed to determine homogeneity (see annex H);

— at least three different levels of contamination shall be used: a negative control (L_0), one level slightly above the detection level of the alternative method (L_1), and one about 10 times greater than the detection level (L_2) (for example 0; 3; and 30 cells/25g);

— at least 8 blind replicates at each level of contamination are analysed by both reference and alternative methods by each collaborative laboratory;

— a suspension of the whole of each sample is prepared for analysis;

— the analysis of samples shall be performed in each laboratory at the stipulated date;

— the test samples are cultured according to annex J. Thus if the 1st culture step is common to both reference and alternative methods use this culture to inoculate each of the next stages (case 1, annex J). If the primary cultures of each method differ, then replicate by setting up each method individually (case 2, annex J);

— in either case, the combination "number of levels of contamination/number of replicates/number of non-outlier laboratories" shall be selected so that at least **480 results (240 by each method**) are generated for use in the calculations.

5.2.1.3 The organising laboratory using all recorded data (see H.3) shall determine which results are suitable and which are outliers for use in calculating the precision data. See annex K, that provides guidelines defining microbiological conditions for disregarding data.

WARNING Outliers: Do not exclude participating laboratory results if there is no clear explanation or gross error to explain them.

5.2.2 Calculation

5.2.2.1 For each level, put the positive results obtained with each method as in Tables 5 and 6:

Table 5 – Positive results 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 [a]	TP_1 [b]	TP_2 [c]

[a] False positive by the reference method
[b] True positive at level 1 by the reference method
[c] True positive at level 2 by the reference method

Table 6 – Positive results 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 [a]	TP_1 [b]	TP_2 [c]

[a] False positive by the alternative method
[b] True positive at level 1 by the alternative method
[c] True positive at level 2 by the alternative method

5.2.2.2 For level L_0 and for each method, calculate the percentage specificity *SP* as in equation (1):

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

where

N_- is the total number of all L_0 tests;

FP is the number of false positive.

5.2.2.3 For each contamination level and for each method, calculate the percentage of sensitivity SE as in equation (2):

$$SE = \frac{TP}{N_+} \times 100\ \% \qquad (2)$$

where

N_+ is the total number of all L_1 or L_2 tests respectively;

TP is the number of true positive.

5.2.2.4 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 method shall be reported as in Table 7:

Table 7 - Paired results of the alternative method and the reference method by the interlaboratory study

Alternative method	Reference method		Total
	+	-	
+	PA	PD	
-	ND	NA	
Total	N_+	N_-	N

Calculate the relative accuracy *AC* expressed in percentage, as in equation (3):

$$AC = \frac{(PA + NA)}{N} \times 100\ \% \qquad (3)$$

where

N is the number of tested samples (for the level L_i or all levels);

PA is the number of positive agreement;

NA is the number of negative agreement.

5.2.2.5 Calculate the confidence intervals for each proportion (see 5.1.1.3.2).

5.2.2.6 Examine the discordant results as described in annex F, by using the counts of PD and ND (see Table 8).

5.2.3 Interpretation

Compare *AC* (Table 7), *SE* and *SP* (Tables 5 and 6) with their relative counterparts obtained within the comparative study including the naturally contaminated samples and the relative detection level.

These criteria do not really address the variability within a laboratory and between laboratories of the method (notions of repeatability and reproducibility). Annex L provides further criteria (accordance, concordance and concordance odds ratio) which may help to address this variability (the criteria of repeatability and reproducibility have been defined for quantitative methods and cannot be used as such for qualitative methods).

6 Quantitative methods - Technical protocol for their validation

6.1 General

Colony counts from a given sample are the most common output from reference quantitative methods and all information should be recorded – sample weight, dilution series, inoculum volume and colony counts at each dilution. In microbiology, these *discrete* data are often truncated, for example by only counting plates with no more than 300 colonies and then applying an expansion factor inverse to the dilution factor to obtain the result. For this reason and other factors involving the growing process of the microorganisms, and also in order to obtain a symmetric distribution or a near normal one, the counting data are often transformed into their logarithms, or a square-root transformation. This transformation can be checked by using histograms of many data points (30 or more) all gathered under the same conditions.

In 6.2 and 6.3 for the validation of quantitative methods the gathering of ***continuous data*** (or *interval data*) is mainly described; however ***remarks about counts*** are included as a particular category.

6.2 Methods comparison study

6.2.1 Linearity and relative accuracy

6.2.1.1 Definitions

6.2.1.1.1 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.

NOTE 1 A **response curve** or a **signal function** is obtained when measuring the relationship between the **signal** or the method's **response** and the **analyte concentration** (doses) in different samples of **reference materials (RM)** having **known values**. In microbiology, where practically no stable reference material is available, these "known values" can be obtained after many replicated measurements using the **reference method**.

After data fitting, smoothing or another algorithm, the alternative method provider should establish a *monotonic* model over the whole application domain of the method in order to transform the measurements into the nearest values to the reference ones. The fitted model is a ***first (or original)* calibration curve** and in collecting all the **calibration factors** over the concentration domain; it is often not linear. But this is not included in the validation study.

A part of the validation is the **verification of the calibration**, which gives a ***final* calibration curve** setting the relationship between the transformed measurements and the corresponding "reference values" [2)], with the same unit system. With the same scales on the axes, this curve should be linear, having a null *intercept* (same lowest values), and a *slope* of 1 (same axes units) and with well estimated characteristics of spread. Extrapolation above and below the tested concentrations should not be performed, except if it is necessary to examine the behaviour near the concentration "zero".

NOTE 2 Counts by the following regression method the linearity is not correctly obtained for counts with low levels or with wide ranges. These counts have quasi-poissonian distributions with repeatability standard deviations proportional to the square root of the mean counts. It involves the same kind of estimation difficulty as mentioned under note in 6.2.3.1.

2) They are derived from the reference method with naturally contaminated samples, if RM's and «known» values are not available.

6.2.1.1.2 Accuracy

closeness of agreement between a test result and the accepted reference value [ISO 3534-1]

NOTE The term accuracy, when applied to a set of test results, involves a combination of random components and a common systematic error or bias component.

6.2.1.1.3 Bias

difference between the expectation of the test results and an accepted reference value [ISO 3534-1]

NOTE Bias is the total systematic error as contrasted to random error. There can be one or more systematic error components contributing to the bias. A larger systematic difference from the accepted reference value is reflected by a larger bias value.

6.2.1.1.4 Relative accuracy

See 5.1.1.1.1

6.2.1.2 Measurement protocol

6.2.1.2.1 Design

The verification of the calibration curve requires many different samples (if possible reference materials; see annex C); and the number of levels examined (doses or concentrations) is a function of the range of concentrations to be used in practice.

A minimum of five different **levels** of analyte in each food type is required to establish a calibration curve. The levels should **uniformly cover the whole range**[3] of interest, having a **minimum** (zero or otherwise), a **central**, a **maximum** , and two **intermediary** levels. The **choice of the scale** has to be made before the concentration levels are chosen and taking into account the contamination range and the expected detection limit (*LOD*): linear for small ranges (say < 3 x *LOD*, or 10 *LOD* to 100 *LOD*) including zero, logarithmic for wide ranges (say > 3 x *LOD*). This can influence the relationship between the standard deviations and the mean values of the response (see note in 6.2.1.3.1).

Priority is given to **naturally contaminated samples**. The word "level" is used here as an ***a priori*** guessed analyte concentration. In principle, an optimal design for regression analysis is obtained by choosing an appropriate and relevant range of analyte concentrations, or by not testing samples with similar counts or eliminating repetitive data with similar results. For liquids, the range could also be obtained by dilution; but many precautions concerning the homogeneity of both solid and liquid samples shall be taken into account.

The samples at one level shall be independently **duplicated** by preparing sub-samples; and the same number of sub-samples analysed at each level - at least two and ideally 5 to 10. When both the reference method and the alternative method use the same decimal dilutions of a sub-sample, duplicate the test at the Petri-dish level (annex M, case 1). When only the reference method needs decimal dilutions, perform the duplicate examination of samples (annex M, case 2).

Duplicate Petri-dishes are inoculated with each decimal dilution.

Therefore a minimum of 10 measurements (ideally 25 to 50 using naturally contaminated samples) by the alternative method shall be done for the verification of a calibration curve.

---

[3] Graphical examples of unacceptable and acceptable repartitions of measures are given in annex P.

If samples of reference materials are not available, it is not possible to estimate correctly the **accuracy** of the alternative method, but only a **relative accuracy** by measuring the same sub-samples with the reference method.

6.2.1.2.2 Food categories

The number of food categories to be studied is 5. This number may be reduced to 1, 2 or 3 categories if the alternative method is to be validated only for these stated categories (e.g. dairy products). The recommended categories are listed in annex B.

The samples shall be **naturally contaminated** with the analyte under study.

If appropriate, environmental samples may be used but only considered as one category (see annex B).

It is desirable that food samples come from as wide a distribution as possible in order to reduce bias from local food specialities and broaden the range of validation.

The range of contamination shall be well covered across the whole quantitative range under test.

If it is not possible to acquire a sufficient number of naturally contaminated samples at the required levels, it is permissible to **contaminate artificially** some samples; the number shall be kept to a minority of the total examined. The procedure of artificial contamination shall result in samples with analyte characteristics as similar to those in naturally contaminated samples (see annex C, 2nd and 3rd options).

Therefore, for each of the five food categories, at least five levels of target analyte are measured by both the reference and alternative methods, each samples being replicated the same number of times (2 to 10), which represents globally 10 to 50 measurements per method and food category. If the results obtained suggest that the matrix is not homogeneous, either in the composition of its components, or because of variations in the concentration of the target analyte throughout the matrix, further samples shall be examined.

After the **separate assessment by each food category**, the **global assessment** of the calibration curve across all the categories is a useful exercise to identify any discrepancy between the scope of application, the accuracy, and the repeatability spread within the whole analyte range.

6.2.1.3 Calculations

6.2.1.3.1 General

Before performing any calculations, plot a **graph** of the values as bi-dimensional points for the reference and alternative methods for each sample, **using the *y-axis* (vertical) for the alternative method** and **the *x-axis* (horizontal) for the reference method**. The points at each level should form a *discrete cluster*. In order to detect **outliers** and a **non-linearity,** visually check the graph for the presence of any **abnormal results**, that is those that are obviously outside each cluster. If any are present, retest the sample if possible and, if no explanation is provided or the result is still outside the cluster, discard temporarily that result and repeat the calculations below in order to estimate its effect in contrast to the calculations with all the data. If a dilution effect is involved, examine this point cautiously.

If all appears correct, use a linear regression programme[4)] which gives the probability of **lack-of-fit** or **non-linearity**, and, possibly also, to **weight unequally the regression** (see ISO 11095). If a less de-

[4)] Or a spreadsheet, like ***Excel***.

veloped calculation program or a smaller computer capacity is available, the calculations required are given in 6.2.1.3.2.

WARNING — When a simple linear regression gives a straight line fit *(y = a + bx*), its correlation coefficient *r* is not sufficient to give the required information. The statistical significance of *r*, or of the slope b, is not synonymous to the linearity test. Moreover, the two hypotheses (*a* = 0 and *b* = 1) shall be checked.

NOTE If the **repeatability error on *y*** (or x also) **depends on the *y* value** (which, unfortunately, is very frequently the case), e.g. a directly proportional increasing error, it is better to use a smoothed **weighting function** (for example: *weight [y] = constant/variance = constant/y^2*) to correctly fit the straight line $y = a + bx$ by the linear regression process (similar to a **WLS-method**, for weighted least squares). This brings the straight line nearer to the observed points which have less spread and more distant where the spread is greater: the estimates of slope *b* and intercept *a* are therefore very dependent on this weighting procedure. This is a difficult and technical subject, mainly for the statistician's use, and it is not therefore be developed in this EN ISO 16140. Statistical books describing General Linear Modelling GLIM/GLM, etc. can be referred to (see ISO 11095). Thus if after gathering the data and examination of the regression graphs they give questionable linearity, it is recommended that a statistician's expertise is used for this analysis.

6.2.1.3.2 Estimations using the regression method

6.2.1.3.2.1 Principle of the regression method

In the general case, the vertical ***y-axis*** **(dependent variable) is used for the alternative method** and the horizontal ***x-axis*** **(independent variable) for the reference method**. This independent variable *x* shall be very accurate, precise and having well known values.

If it is foreseen that a **repeatability error s_r (x) can happen on *x*, that is rather comparable or larger than s_r (y) on y**, the fitting functions y(x) or x(y) can lead to quite different straight lines. In the case where the repeatability standard deviations s_r (x) and s_r (y) are comparable (respectively on *x* and *y*), compute other estimates (see annex R.3). If the s_r (x)'s are much larger than s_r (y)'s, permute the axes *x* to *y*, and *y* to *x* to perform a regression y(x), or to use the regression x(y) without axes permutation (see annex R.4). For these choices (see annex R.2), use a cutpoint of 2 for the ratio of the repeatability errors on *x* and *y*, or their inverse.

For calculations: see annex R.

6.2.1.3.2.2 Further estimations

If the **repeatability errors s_r or the residual error $S_{y:x}$ depend strongly on *x* (or *y*)**, that is the residuals seems to increase or decrease clearly with *x* (or *y*; that is *s* is not constant), the fitted regression line could be incorrect and not pass through the more precise points. Moreover, the relationship of CL with *x* (or *y*) could not be reliable in the main part of the measurement range. In this case use a **weighted regression technique** (WLS, GLIM, etc.).

When it seems necessary, and in order to simplify this step, **monotone mathematical transformations of both axes *x* and *y*** can be used. For instance, $x' = \log x$ and $y' = \log y$, or $x' = x^m$ and $y' = y^m$ can be used for a good estimate of m (negative or positive, semi-integer or otherwise): the choice is made by visual inspection of residuals or of repeatability s_r 's for each *x*; constant errors correspond to the ideal transformation.

When using such a transformation, the intercept *a* shall not be very different from zero. Then, **when all estimates are provided**, like the estimated <y>'s and CL, **their inverse transformations** can finally be done.

6.2.1.4 Interpretation

6.2.1.4.1 The **relative accuracy relationship** between the alternative and reference method is assessed through the **linear model: *y = a + bx***.

There is **no systematic bias** (that is ideal **accuracy**) between the methods if this equation equals the theoretical equation ***y* = *x***, that applies if the two methods behave equivalently. The **intercept** is theoretically **null** within this ideal model. The estimated intercept *a* generated by the two methods is checked by $p\{a = 0\}$. If the alternative method has a systematic **bias** relatively to the reference method, the probability $p\{a = 0\}$ is less than $\alpha = 0{,}05$ (2-sided).

The theoretical **slope** of a line through the points giving a totally true equivalence is equal to **one**. The estimated slope *b*, generated by the two methods, shall be checked by $p\{b = 1\}$. If the alternative method does not give significantly the same values than those of the reference method, the probability $p\{b = 1\}$ is less than $\alpha = 0{,}05$ (2-sided). In this case, the alternative method has a **bias** relatively to the reference method, **depending on the concentration value (*x* or *y*)** and is maximal at the domain limits.

6.2.1.4.2 The **linearity** or the ***lack-of-fit*** can be visualised by a graphical representation (annex N). The best one is a **residuals plot** (see R.1). The $\{y_k\}$ are not plotted versus the $\{x_k\}$ values (k=1 to N, N=qn), but the estimated residuals $\{y_k - Y_k\}$ versus $\{x_k\}$. It gives the shape of the non-linearity if detected with p(F) in R.5.

6.2.1.4.3 The estimation of the **methods precision** by CL($<y_{U/L}>$) mainly comes from the residual standard deviation $s_{y:\,x}$ (see annex R). It gives limits within which the specified accuracy for the result is found with a probability of 95 %, when t is equal to about 2.

The corresponding CL($<x_{U/L}>$) (see annex R) is of better use because they are included in the reference system.

6.2.2 Detection and quantification limits

6.2.2.1 General

The following characteristics involve the reliability of the signal (or response) obtained with the alternative method to detect a non-null concentration obtained by the reference method. It gives the *precision limits,* including *bias,* at the lower end of the **concentration domain**.

The **background noise** or **blank** (instrumental or otherwise) is generally determined with at least six independent blank determinations. It also corrects the *general bias* of actual measurements by the use of a central value (average) and gives an important **estimate of spread, s_0**.

Often, when it is not possible to measure or detect something **less than a given low level**, samples are chosen with an analyte concentration near to the first estimation x_{LC} taken from the previous straight line for the verification of the calibration: $x_{LC} \approx 1{,}645\ s_a$ [3], [5], [1], where s_a is obtained in annex R. Around this concentration, a standard deviation is estimated which is expected near s_0, and used instead. Another procedure is to take the average and the standard deviation of all the measurements of the lowest doses detected after many replications of the same dilution scheme (dilution factor 1: 2 to ≤ 1: 10).

NOTE The estimation of the standard deviation s_0 includes many uncertainties, mainly when below the quantification level. Therefore, an initial s estimate (if possible with the robust method S_n given in annex Q) is often sufficient to provide the further estimates.

6.2.2.2 Definitions

6.2.2.2.1 Critical level (*LC*)

Smallest amount which can be detected (not null), but not quantified as an exact value. Below this value, it cannot be sure that the true value is not null.

NOTE At this level, the false negatives probability β is 50 % (β is the second type of statistical error (see 6.2.2.4).

6.2.2.2.2 Detection limit (*LOD*)

Higher than the critical level (6.2.2.2.1), because it involves a power, the probability 1 - β, which has to be well over 50 %, for example 95 %.

For instance: *LOD* = Average(blank) + $z\ s_0$(blank), where z = 2 x 1,645 ≈ 3,3, gaussian critical value with $\alpha = \beta$ = 0,05 (1-sided, see 6.2.2.4), for a rather high number n of blank samples.

6.2.2.2.3 Quantification limit (*LOQ*)

The smallest amount of analyte, (that is the lowest actual number of organisms), which can be measured and quantified with defined precision and accuracy under the experimental conditions by the method under validation.

NOTE AOAC defines, the quantification limit for quantitative methods as: *LOQ* = 10 s_0; it is equivalent of saying that a variation coefficient $CV_r = s_r/x$ is needed to be better than 10 %, that is lower [2].

6.2.2.3 Measurement protocol and samples

Perform at least six (preferably 10) *blank* sample determinations (negative, zero dose or near zero) to estimate the **baseline or threshold** spread s_0.

NOTE **For counts**, more than **five** samples at **a minimum non-zero level** are sufficient to ascertain that the contamination rate is ***p* <** 50 %, when **no *positives*** are detected (95 % confidence level). If **one *positive* out of five** is detected, it can be assessed that ***p* > 1 %**. Therefore, **starting from three analyte doses of reference material chosen** among 3, 10, 30, 100, 300 etc. cfu/unit, and **replicating each at least six times,** the **critical level** (50 % of detection) can be estimated: for example an ideal case could be: 0/6 at 10, 2/6 at 30, 5/6 at 100, giving an estimate around 30 cfu/unit. Although the approximate estimate provided by this procedure can be used (see annex P for a coherent estimation).

6.2.2.4 Calculations

In the general case, the critical level and the detection limit involve two types of statistical errors: α (to detect a non-existing difference (false-positive)) and β (to not detect a true difference (false-negative)). **The power 1 - β is the probability to detect significantly a value larger than *LC*.**

From the *blank* sample determinations x_{0i}, estimate their standard deviation s_0 (from S_n, annex R) and the **bias**: x_0 = median of the x_{0i}. It corresponds to a general lack of *specificity*, which is statistically significant if

$t = (x_0 \sqrt{n})/s_0 > t_\alpha > 1{,}645$ (see below).

The critical level $LC \approx 1{,}65\ s_0$ (+x_0 for the alternative method), for α = 5 **%** (and 1-β = 50 **%**)

(More precisely, $LC = t_\alpha\ s_0$, with t_α being the Student value for a 1-side level α and n - 1 degrees of freedom; for α = 5 **%** and n = 6, t_α = 2,015 and then $L_c \cong 2{,}0\ s_0$; if $n \to \infty$ then $t_\alpha \to 1{,}645$).

The **limit of detection is *LOD* ≅ 3,3 s_0** (+x_0 for the alternative method), for **α = 5 %** and **1- β = 95 %**

(More precisely, $LD = (t_\alpha + t_\beta)\ s_0$; with t_β being the Student value for a 1-side level β and n - 1 degrees of freedom; e.g. for 1- β = 90 % and n = 6, t_β = 1,476 and then $LD \approx 3{,}5\ s_0$; for 1 - β = 80 %, t_β = 0,920).

The **limit of quantification** is $\boldsymbol{LOQ = 10\ s_0}$. ($+x_0$ for the alternative method).

6.2.3 Relative sensitivity and determination of unknown samples

6.2.3.1 General

An estimate of sensitivity is used in this EN ISO 16140 in order to ascertain that values given by the alternative method do not differ markedly from the reference method (less than 30 % in difference).

6.2.3.2 Definition

For the purposes of this EN ISO 16140, the relative sensitivity is defined as the ability of the alternative method to detect two different amounts of analyte measured by the reference method within a given matrix, at a specified average value, or over the whole measurement range; that is, it is the minimal quantity variation (increase of the analyte concentration x) which gives a significant variation of the measured signal (response y).

NOTE The sensitivity differs from the detection limit (6.2.2), because it is calculated for each value of the measurement range. It involves also the 2 types of statistical error α (2-sided) and b.

6.2.3.3 Calculations

The measurement protocol and the samples are described in 6.2.1.2.

Relative sensitivity: $\boldsymbol{\Delta C_S = 5{,}1\ s(\langle x(y)\rangle)}$, for **2-sided α = 5 %** and a **power 1 - β = 95 % ,** with $\boldsymbol{s(\langle x(y)\rangle) \cong s(\langle y\rangle)/b}$ from R.6.2.

(More precisely, $\Delta C_S = (t_\alpha + t_\beta)\ \sqrt{2}\ s(\langle x(y)\rangle)$. For α = 5 %, a power 1 - β = 90 % and n=5, t_α = 2,776 and t_β = 1,533, then $\Delta C_S = 6{,}1\ s(\langle x(y)\rangle)$

Identification and determination of unknown samples: $\langle x(y)\rangle = \langle x\rangle \pm t_\alpha\ s_{\langle x(y)\rangle}$.

Then **plot the** (im-)**precision profile** $s(\langle x(y)\rangle)$ or $CV(\langle x(y)\rangle)$ versus x(y).

6.2.4 Specificity, inclusivity and exclusivity

6.2.4.1 Definitions

6.2.4.1.1 Specificity

For the purposes of this EN ISO 16140, specificity is defined as the degree to which a method is affected (or not) by the other components present in a multi-component sample. That is the ability of a method to measure exactly a given analyte, or its amount, within the sample without interference from non-target components such as a matrix effect, or background *noise*.

6.2.4.1.2 Inclusivity and exclusivity

For the purposes of this EN ISO 16140, selectivity is defined as a measure of the degree of non-interference in the presence of non-target analytes. A method is selective if it can be used to detect

the analyte under examination, and that a guarantee can be provided that the detected signal can only be a product by that specific analyte.

This criterion is not applicable to a **total viable count**.

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

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

6.2.4.2 Measurement protocol

6.2.4.2.1 Samples

In microbiology (except in the case of total plate counts), the inclusivity and the exclusivity is established by the analysis of:

— at least 30 *positive* pure strains (of the analyte being studied);

— at least 20 *negative* pure strains (analytes other than that being studied), taken from strains known to commonly and consistently cause interference usually with the target analyte.

Select the *positive* and *negative* strains which regularly contaminate the food product(s) under examination. For strains not belonging to a collection, carry out a complete identification of the various strains tested. The real origin of a strain from food (not from the collection) shall be known.

Criteria for choosing the strains are given in annex G.

The pure strains from laboratory cultures constitute the test samples. The level of inoculation for each microorganism is fixed at **more than 100 times the limit of detection** of the target species.

Each sample containing a strain shall be analysed twice.

6.2.4.2.2 Measurement protocol

The specificity, the inclusivity and the exclusivity of the reference method need not be established if already known.

6.2.4.3 Expression of the results

This characteristic of selectivity is only given in a descriptive way (see Table 8) relative to the limit of detection of the target analyte; therefore, no calculations are needed. The lack of the inclusivity and exclusivity is a part of the overall lack of *specificity* (see also the *bias* in 6.2.1.4).

Table 8 — Presentation of the results for the inclusivity and the exclusivity

Strain tested	**Reference method** expected or demonstrated result Inoculum level etc.	**Alternative method** Results (give the detection limit, etc)
	(single test)	Duplicate
Target strains		
etc.		
Non-target strains		
etc.		

Data published and meeting the requirements of this EN ISO 16140 may also be used by the organising laboratory in charge of the methods comparison study, to provide further information to the above criteria (see annex A, providing criteria for the acceptance of external results).

6.2.4.4 Interpretation

It should be possible to evaluate whether the alternative method conforms to the manufacturer's specifications and to the specific requirements for its reliable use.

6.2.5 Additional characteristics of the alternative method

Concerning the alternative method, document all other pertinent characteristics affecting method performance: for example stability, reliability, robustness or ruggedness, etc.

6.3 Interlaboratory study

6.3.1 General

The collaborative study is aimed to **comparatively determine the performance characteristics** (accuracy and precision) of the alternative method against the reference method.

For guidelines to the organising laboratory regarding the collaborative study see annex H. See also the ISO 5725-2.

6.3.2 Terms and definitions

For the purposes of this EN ISO 16140, the following terms and definitions apply:

6.3.2.1 accuracy

closeness of agreement between a test result and the accepted reference value [ISO 3534-1]

NOTE It is approximately estimated as a bias, the bias being the systematic part of the error.

6.3.2.2 precision

closeness of agreement between independent test results obtained under stipulated conditions of repeatability and reproducibility (see 6.3.6 and 6.3.7) [ISO 3534-1].

6.3.2.3 outlier

extreme value which normally appears randomly in less than 1 % of tests, but more frequently, if abnormal situations occur. To quantify this probability, test procedures can be used

NOTE The "robust" approach of this EN ISO 16140 does not exclude an outlier, unless there is a clear indication of a sound microbiological reason for doing so. This approach can be used either as a test procedure for the outlier detection or for "robust" statistical estimations.

6.3.3 Measurement protocol and samples

6.3.3.1 The interlaboratory study shall produce at least eight collaborative laboratories having results without outliers.

The accuracy and precision estimates should be calculated from a large number of duplicate test results. This figure should be a minimum of 96 results for the one food matrix chosen.

Guidelines and requirements for organising, dispatching and conducting the interlaboratory studies are given in annex H.

The organising laboratory is also responsible for the preparation of the test protocol and a data sheet (see below) for recording of all experimental data and critical experimental conditions used by each laboratory (See H.3)

It is necessary for the analyst in each collaborating laboratory to demonstrate his competence in the use of the alternative method and of the reference method prior to participating in the study proper.

6.3.3.2 The protocol is the following:

— one relevant food matrix is used (see annex B);

— the analyte concentrations should be chosen to cover at least the lower, middle and upper levels of the entire range of the alternative method. The samples shall be shown to be homogeneous by the organising laboratory. A negative control should also be included;

— artificial contamination of a food sample with the target analyte may be used;

— to compare the alternative method with the reference method the same samples shall be used for each method. Four sub-samples from each level (or two aliquots, each measured by both methods) are prepared for each laboratory. These are blindly coded but labelled so that two are measured by the reference method and two measured by the alternative method;

— liquid samples (compared to solid samples) gives greater assurance of homogeneity if prepared and dispatched without change in microbiological content and used correctly. In specific cases, it could be necessary to subdivide the samples just before measurement with both methods;

— the analysis of samples shall be performed in each collaborative laboratory and the organising laboratory at a stipulated date using common batches of media and kits;

— the rounding of results shall be set by the organising laboratory. After pooling of each laboratory's results, the final data tabulation for each analyte level *j* shall be presented as shown in Table 9.

Table 9 — Presentation of the results of the interlaboratory study per each analyte level (j)

Laboratories (i)	Methods (k) and duplicate (r)			
	Reference method (coded)		Alternative method (coded)	
	Duplicate 1	Duplicate 2	Duplicate 1	Duplicate 2
1				
2				
etc.				
(n)				

6.3.4 Calculations

6.3.4.1 In microbiology, the data $\{y\}$ do not always show a normal statistical distribution, i.e. a Gaussian distribution. This distribution could be checked if enough values are available (many more than 30 results) all at the same level. Often, in order to get a more symmetric distribution, sparse counts are better transformed into logarithms. Other more complicated procedures can also be used.

It is often difficult to make reliable estimations (average, standard deviation, etc.) with a small bias and in presence of **outliers**. ISO 5725 includes **outlier tests** (Cochran, Dixon, Grubbs) in order to discard the badly influencing values and to obtain a better estimate; this however reduces the number of useful values for statistical analysis.

Therefore, in order to guard against these difficulties, **robust estimators** are used in this EN ISO 16140 as they are insensitive to any extreme values, and moreover do not depend too much on the distribution of the data, also they do **not exclude the laboratories** reporting extreme values unless exclusion is based on sound microbiological reasons. Estimate **three consensus values** for all participating laboratories:

— estimation of the global **centre** (for bias estimates): the median MED (instead of the mean *M*);

— estimation of the **spread *between* the duplicate means**, including the **reproducibility**: s_b based on the *recursive median* S_n (6), (see annex Q), instead of a classical standard deviation (*s*); and

— estimation of typical ***within* spread**, the **standard deviation of repeatability** derived from the median of the duplicate standard deviations: k_2 MED$\{s_i\}$, instead of a classical pooled standard deviation, coming from the variances mean.

6.3.4.2 For **each method k** (for example 1: reference, 2: alternative) **and level *j*, compute**:

Level j Method k		Replicates 1	 2	 M_i	 s_i
Laboratory (i):	1	y_{i1}	y_{i2}	$\frac{(y_{i1}+y_{i2})}{2}$	$\frac{\lvert y_{i1}-y_{i2}\rvert}{\sqrt{2}}$
	2	...	...	...	...
	...	...	...	...	...
	N	...	...	...	...
			MEDIAN:	MED $\{M_i\}$	$S_r = k_2$ MED $\{s_i\}$
Robust $s_R = S_b = s\{M_i\} = k_1 \times S_n$ $S_R = \sqrt{s_b^2 + \frac{S_r^2}{2}}$					

where the constants k_1=1,1926 and k_2= 1,4826; only this last constant is dependent on the fact that only duplicates (two laboratory values) are used and not more replicates [5].

S_n corresponds to the following procedure which does not use a centre estimate:

$S_n = \text{MED}_i \{\text{MED}_j |M_i - M_j|\}$. It consists in 2 phases of successive medianing:

— from each one of the *n* values, the median of its (*n*-1) **absolute differences** to the other values is estimated; and

— the spread estimate is the median of the *n* deviates medians, formerly calculated (for example, see annex P).

NOTE Only poor alternatives exist. The **Median Absolute Deviation *from the MEDian*** or **MAD** (Hampel, 1974): MAD = $\text{MED}_i\{|M_i - \text{MED}|\}$. This **robust** spread estimate is symmetrical around a centre estimate (MED). Gaussian case: $\sigma \cong 1{,}4826$ MAD. Another comparable alternative is based on the **interquantile range IQR** = Q3-Q1 = 3rd - 1st quartiles; for the gaussian case $\sigma \approx 1{,}4826$ IQR/2. This last one is easy to get with Excel [6].

6.3.5 Relative accuracy

6.3.5.1 Definition

For the purposes of this EN ISO 16140, the accepted reference values involved within a definition of trueness, are only given by the responses obtained by the reference method on identical samples.

---

[5] It is $\sqrt{\frac{\nu}{\chi^2}}$, with χ^2 = chi-square for p=0,5 (χ^2 = 0,455 here) and ν degrees of freedom within the replicates (ν = 2-1 here for duplicate).

[6] The ***Excel*** statistical functions to use here are: MEDIAN(range) equals to the 2nd quartile, and QUARTILE(range;k) with k =1, 2 or 3 for the 1st, 2nd or 3rd quartiles.

Therefore, the deviation of the measurements obtained by the alternative method better corresponds to a relative accuracy.

6.3.5.2 Calculations

For each analyte level j, compute $d_i = M_{i,alt} - M_{i,ref}$ for i = 1 to *n* laboratories,

where

$M_{i,alt}$ is the duplicate mean of laboratory i obtained by the **alternative** method;

$M_{i,ref}$ is the duplicate mean of laboratory i obtained by the **reference** method.

Thus, the median $MED\{d_i\}$= **bias D** and the robust standard deviation $SD\{d_i\} = k_1 S_n$ provide a t-like statistics:

***t*(d)**= $MED\{d_i\} \sqrt{n}/s\{d_i\}$, with *n* - 1 degrees of freedom (df).

In a Student t table is found the **probability that D = 0.** (For example for *n* = 12 (11 df), the critical value for 2-sided α = 0,05 is $t_{0,05;11}$= 2,201). If *t*(d) > $t_{\alpha;df}$ the bias D between the two compared methods is significant, that is **the alternative method lacks accuracy, relative to the reference method, for the level *j***.

The relationship between the relative accuracy and levels can be modelled (constant or proportional bias, etc), in connection with the linearity of the calibration curve.

6.3.5.3 Interpretation

A null bias (D = 0) for each analyte level is expected. However, if either the estimated D magnitude is too big for the methods purpose, or statistical significance is reached, the alternative method is different from the reference one.

6.3.6 Repeatability

6.3.6.1 Terms and definitions

For the purposes of this EN ISO 16140, the following terms and definitions apply:

6.3.6.1.1 repeatability

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; that is *repeatability conditions*)

6.3.6.1.2 repeatability limit (*r*)

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

Use: if the difference between two results exceeds *r*, the results should be considered as suspect.

6.3.6.2 Calculations

For each method k and analyte level j, compute:

— **repeatability limit $r = 2{,}8\ s_r$**, with [7] the repeatability standard deviation s_r;

— **relative standard deviation of repeatability $RSD_r = 100\ \% \ s_r/MED\{M_i\}$.**

The repeatability of the **alternative and reference methods are compared** with a F-distribution:

$F = (s_{r;alt}/s_{r;ref})^2$ with n and n degrees of freedom. A F-table gives the **probability p(F) that $s_{r;alt} = s_{r;ref}$**; for example for $n = 12$, the critical value for 2-sided $\alpha = 0{,}05$ is $F_{0,05;12;12} = 2{,}69$.

If F (or 1/F) > $F_{\alpha;n;n}$ then the compared **methods have different repeatability, for the level j.**

The relationship between repeatability and levels can be modelled, as well as for the relative standard deviation of repeatability. For this purpose, see ISO 5725.

6.3.6.3 Interpretation

A comparable repeatability for each analyte level is expected. However, if the statistical signification p(F) < α (2-sided) is reached, the alternative method is different from the reference one.

At present no criteria are available for the rejection of an alternative method.

6.3.7 Reproducibility

6.3.7.1 Terms and definitions

For the purposes of this EN ISO 16140, the following terms and definitions apply:

6.3.7.1.1 reproducibility

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 (that is *reproducibility conditions*)

6.3.7.1.2 reproducibility limit (*R*)

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

Use: if the difference between two results from different laboratories exceeds *R*, the results shall be considered suspect.

6.3.7.2 Calculations

The value of R is calculated from the repeatability standard deviation s_r and the between laboratories standard deviation s_b, for each method k and each analyte level j.

— **Reproducibility limit $R = 2{,}8\ s_R$**

---

[7] The constant $2{,}8 \cong z_\alpha \sqrt{2}$ with $z_\alpha = 1{,}960$ and 2-sided $\alpha = 0{,}05$ or 95 % confidence level.

with the **reproducibility standard deviation** $s_R = \sqrt{s_L^2 + s_r^2} = \sqrt{s_b^2 + \frac{s_r^2}{2}}$

and $s_L^2 = s_b^2 - \frac{s_r^2}{2}$ = laboratories variance, with $s_b = k_1 S_n$

relative standard deviation of reproducibility $RSD_R = 100\ \% \ s_R / MED\{M_i\}$

The reproducibilities of the **alternative and reference methods are compared** with a F-distribution:

$F = (s_{R;alt} / s_{R;ref})^2$ with (n - 1) and (n - 1) degrees of freedom. A F-table gives the **probability p(F) that $s_{R;alt} = s_{R;ref}$**; for example for n = 12 (11 df), the critical value for 2-sided α = 0,05 is $F_{0,05;11;11} = 2,82$. If F (or 1/F) > $F_{\alpha;n-1;n-1}$ then the compared **methods have different reproducibility, for the level j.**

The relationship between reproducibility and levels can be modelled, as well as for the relative reproducibility RSD_R. For this purpose, see ISO 5725.

At this point, estimate the **probability p(F) of having homogeneity** for the **"between laboratory spread** that is the laboratory differences are less than their own typical determination spread: **$F = 2 (s_b / s_r)^2$** with df(NUM)= n - 1 and df(DENOM)= n degrees of freedom. A F-table gives the probability p(F) indicating the importance of the between labs variation s_L.

6.3.7.3 Interpretation

A comparable reproducibility for each analyte level is expected. However, if the statistical significance p(F) < α (2-sided) is reached, the alternative method is different from the reference one (worse or better, along R).

At present no criteria are available for the rejection of an alternative method.

$R \leq 2\ r$ is generally correct (see Horwitz criterion). However it depends on method responses and the context of use. The probability to have non-homogeneity with a large spread between laboratories generally indicates that the standardisation of the method has to be improved, that is the instructions for their operation are not well understood (wording, explanations, etc.), analyses are not well done (skilfulness), or there are abnormal time changes for reactions, cultures, etc. as well as for all external conditions in the laboratory.

Annex A
(normative)

Specific rules for the acceptance of external results already obtained in a prior validation scheme

(see references: 4.1; 5.1.1.3.5; 5.1.3.4; 6.2.4.3)

A.1 General

This considers data generated by International or European Standards and other reference methods.

A.2 Alternative methods which have changed or were previously validated to an International or European Standard

When assessing whether existing external data can be accepted in conformance with this EN ISO 16140, establish if the alternative method has been changed in any manner after the data were produced.

If this is the case, and the changes are considered to be major, the results cannot be accepted.

If there are no changes or if the changes are minor, then the quality system of the laboratory carrying out the validation shall be addressed.

If the organising laboratory was not accredited according to EN ISO/IEC 17025, an audit shall be performed to check calibration, media, temperature, equipment and training records from the time when the validation was carried out. If the results of the audit are unsatisfactory, the results cannot be accepted. The audit shall be documented in writing.

If, however, the organising laboratory is accepted on the basis of the audit, or the laboratory had an EN ISO/IEC 17025 accreditation, the validation criteria are judged to be of satisfactory reliability.

Also information on the collaborative laboratories (their capability and QA practises) shall be assessed.

A.3 Alternative methods which have not been validated against a reference method in an International or European Standard

Prior studies are be acceptable under this standard provided:

— that they have been conducted according to validation protocols approved by a recognised panel of technical reviewers and the results of such studies had been fully accepted by them;

— that the technical reviewers were operating under the sponsorship of internationally recognised organisations performing method validations (for example AFNOR, NORVAL, AOAC International, AOAC Research Institute);

— that the validations include studies that conform to at least the total sample number and food matrix requirements of this standard.

When the alternative method has been compared with an internationally recognised reference method (such as AOAC International) that differs in minor aspects from the reference method in the International or European Standard and if the protocol is similar to this standard, then the results can be accepted.

If the method is substantially different from the reference method in the International or European Standard, an assessment is made to determine if these differences would have a minor or major impact on method performance. An assessment is made regarding supplementary data, if any, required to resolve procedural and/or reference method differences (for example different primary enrichment broth). Decisions that reference methods contain major difference shall be substantiated by the organising laboratory, with documented data. The data required to resolve the perceived differences shall be stipulated.

Annex B
(informative)

Classification of sample types for validation studies

(see references: 5.1.1.2.1; 5.1.2.2; 5.2.1.2; 6.2.1.2.2 and annex H)

B.1 Food and animal feeding stuffs samples

Table B.1 — Food categories relevant to major foodborne pathogen

Type of product	*Yersinia* spp.	*Clostridium perfringens*	*Listeria monocy-togenes*	*E. coli* O157 & VTEC	*Staphylococcus aureus*	*S.aureus enterotoxins*	*Campy-lobacter* spp.	*Salmonella* spp.	*Bacillus cereus*
Meat products									
Raw	x		x	x			x	x	x
Heat processed			x		x	x			
Frozen			x	x					
Fermented			x	x					
Cured		x	x		x	x			
Other		Dishes/ gravy	Pâté						
Poultry									
Raw	x						x	x	
Heat Processed									
Frozen									
Others		Dishes/ gravy							
Fish and Seafood products									
Raw	x		x				x	x	
Heat processed									
Frozen									
Smoked			x		x				
Others									
Fruits and vegetable based products									
Raw	x		x	x			x	x	
Heat processed									
Frozen									
Dry									x
Fermented									
Cured/Salted									
Juices/Concentrates				x				x	
Low moisture/IMF									
Others						Processed Mushrooms			
Dairy products									
Raw	x		x	x	x	x	x	x	x
Heat processed									x
Frozen			x	x	x	x		x	x
Fermented			x	x	x	x		x	
Dry					x	x		x	x
Other									
Chocolate/Bakery products									
Low moisture/IMF								x	
Dry								x	
Others					Pastry				Custards
Other products									
Beer									
Dressings									
Spices		x						x	x
Mayonnaise						x		x	
Pasta						x			
Egg and derivatives								x	
Cereals/ rice									x
Animal feeds									
Miscellaneous								x	

Table B.2 - Food Categories relevant to major non-pathogenic microorganisms

Type of product	Yeasts & Moulds	Lactic acid bacteria	Total viable counts	Coliforms	*Escherichia coli*
Meat Products					
Raw		x	x	x	x
Heat processed		x	x	x	
Frozen			x	x	
Fermented	x	x			
Cured		x			
Others	x				x
Poultry products					
Raw		x	x	x	x
Heat processed		x	x	x	
Frozen			x	x	
Others					
Fish and seafood products					
Raw		x	x	x	x
Heat processed		x	x	x	
Frozen			x	x	
Smoked		x	x	x	
Others					
Fruit and vegetable based products					
Raw		x	x	x	x
Heat processed			x	x	
Frozen			x	x	
Dry	x		x	x	
Fermented	x				
Cured/Salted	x				
Juices/Concentrates	x	x			
Low moisture/IMF	x				
Others					
Dairy products					
Raw			x	x	x
Heat processed			x	x	
Frozen			x	x	x
Fermented	x				x
Dry			x	x	
Others					
Chocolate and Bakery products					
Low moisture/IMF	x		x	x	
Dry			x	x	
Others					
Other products					
Beer	x				
Dressings	x	x			
Spices					
Mayonnaise	x	x			
Pasta					
Egg derivatives					
Cereals/rice					
Animal feed					
Miscellaneous	x			x	

B.2 Veterinary samples

For the screening of the following farmed animals the cited samples are examined for *Salmonella*:

Table B.3 – Veterinary samples for *Salmonella*

Animal, product or its environment	Sample for analysis
Degenerated eggs	Composite faecal samples
Laying hens	Composite faecal samples
One day-old chicks	Bedding straw from boxes
Pullets of breeding stock	Pooled faecal samples
Carcasses of breeding stocks	Pooled samples of meconium
Fresh poultry meat	Pieces of neck skin or pieces of tissue
Poultry	Pooled samples of faeces
Fresh beef, veal and pork	Surface swabs from carcasses and pieces of tissue and drip water.

Annex C
(normative)

Use of naturally contaminated samples and preparation of artificially contaminated samples in validation studies

(see references: 5.1.1.2.1; 5.1.1.2.3; 6.2.1.2.1; 6.2.1.2.2; 6.3.3.2; annex H)

Selection of a food type within a category in order to increase the proportion of naturally contaminated samples should be considered. If only a limited number of naturally contaminated samples are available then the second option below can be taken to increase sample numbers. The third option shall only be used in well justified cases. Reference materials containing stressed microorganisms, preferably with well defined levels of target analyte (microorganisms) and quantified shelf life may be used where stated.

1st option: **Naturally contaminated samples**

Use samples collected from products analysed on a routine basis by either the organising laboratory or other laboratories.

Storage should minimise both microbiological change and minimise microbial stress. Contamination (qualitative or quantitative as the case may be) should be confirmed immediately prior to the validation study.

2nd option: **Contamination by mixture**

Inoculate liquid and semi-solid products to be tested by dilution with a naturally contaminated sample of similar type.

The background microflora in the "diluted product" shall be of a similar type and in a similar state of stress to those occurring in naturally contaminated products.

3rd option: **Spiked samples**

If available, reference material[8)] should be used for performing the spiking. If not available, the strains used should have been isolated from the same type of product.

A protocol for stressing the target microorganism is to be defined and the stress demonstrated at the time of inoculation.

The level of microflora should be representative of the contamination which occurs in that product naturally.

8) 10 replicates at each of 3 or 4 levels of contamination are examined.

4th option

Reference materials

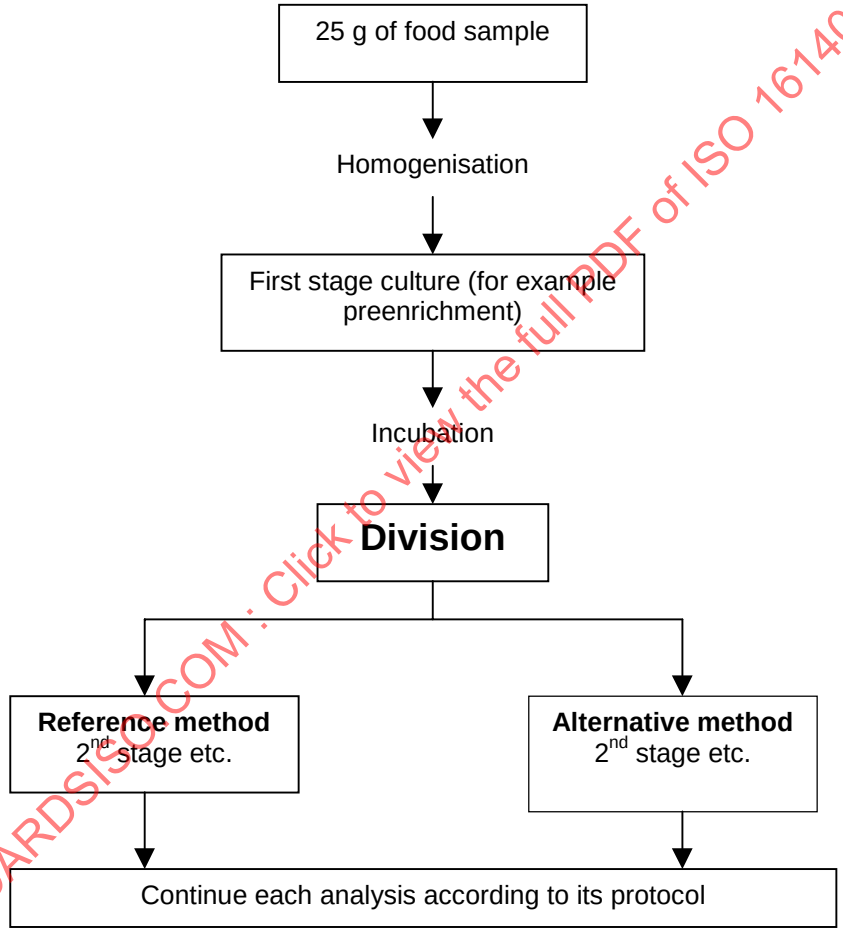
Reference materials, such as certified reference material, containing appropriate but well defined levels of target analyte (microorganisms) in a stable but stressed state, may be used to spike samples for analysis by both qualitative and quantitative methods. For qualitative studies their use should be limited when only a few strains or serotypes of food origin of the target analyte are available as reference materials.

Annex D
(normative)

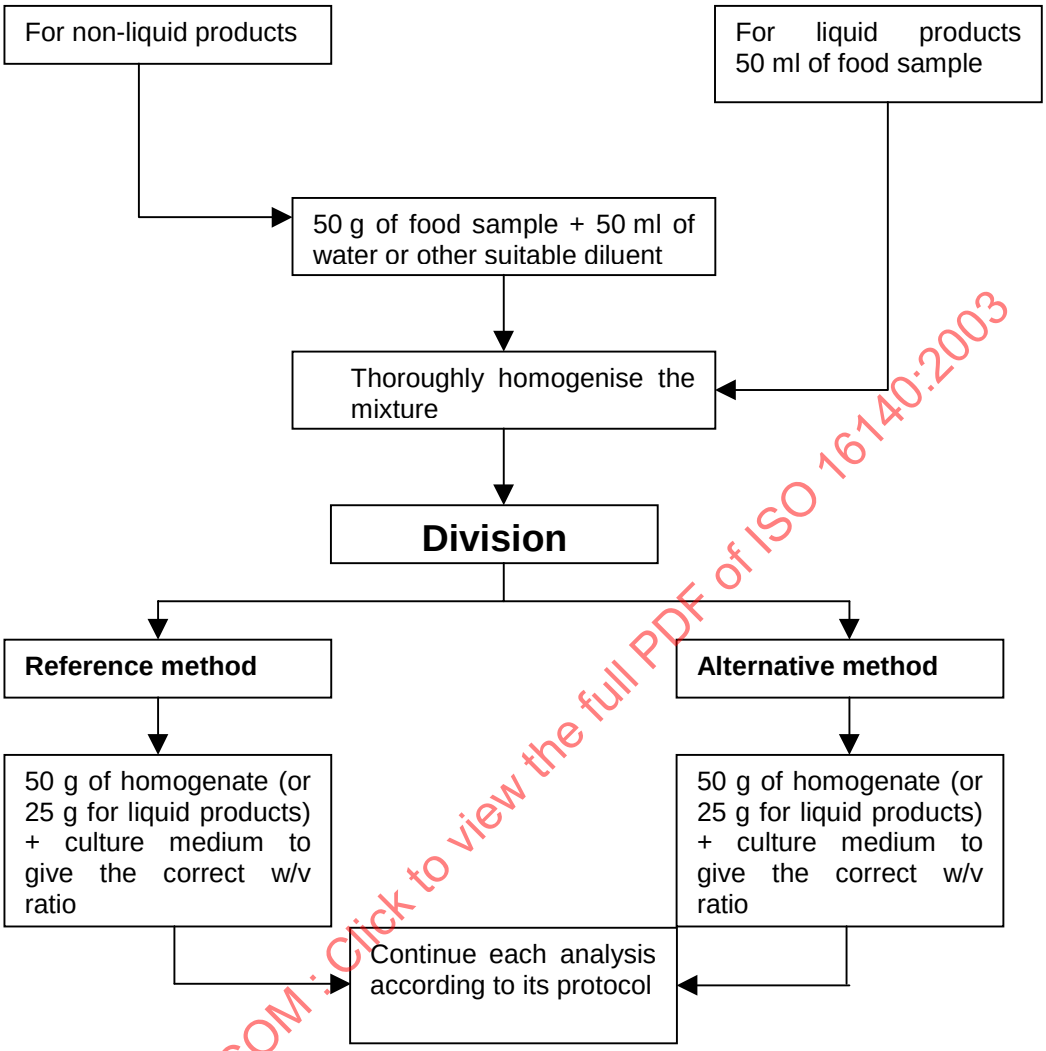
Duplication of samples for the determination of relative accuracy and of relative detection level for qualitative methods

(see references: 5.1.1.2.3; 5.1.2.2; annex H)

D.1 Case 1 - Where all the conditions of the first stage of culture of both methods are identical.



D.2 Case 2 - Where the first stage of culture of both methods differ.



NOTE It should be noted that in case 2, due to the dilution of the media by the sample + diluent, ready-to use media cannot be used and the concentration of ingredients in laboratory prepared broth should be increased by about 10 %.

Annex E
(normative)

Calculation of the confidence intervals associated with the number of samples tested

(see reference: 5.1.1.3.2)

For each *AC*[9], *SE*, and *SP* percentage (p) (see 5.1.1.3.1) calculate the confidence intervals (CI):

— if 10 % < p < 90 %, calculate the approximate two sides confidence intervals at 95 %:

$$\text{CI (at 95 \%)} \approx p \pm 2\sqrt{\frac{p(1-p)}{n}}$$ with $n = N, N_+, N_-$ respectively for p (in %) = *AC*, *SE*, *SP*;

— for $p \geq 90$ %, calculate the lower confidence limit at 95 % (1-sided), with $n \approx N, N_+, N_-$ respectively for *p* (in %) = *AC*, *SE*, *SP*.

For this case, it is preferable to use the binomial table for $n = 10, 20, 30, 40, 50, 60$. See Table E.1:

Table E.1 - Lower confidence limit at 95 % for *p* = 90 % and greater

	n =	10	20	30	40	50	60
p =	0,90	0,75	0,83	0,82	0,84	0,83	0,84
	0,92	0,85	0,83	0,85	0,86	0,87	0,88
	0,94	0,85	0,88	0,88	0,89	0,89	0,89
	0,96	0,85	0,93	0,92	0,91	0,93	0,93
	0,98	0,95	0,93	0,95	0,96	0,95	0,96
	0,99	0,95	0,98	0,98	0,96	0,97	0,98

EXAMPLE $n = 20$, $p = 94$ % then LCL (p; at 95 %) = 88 %.

[9] See annex W for the explanations of these abbreviations.

Annex F
(normative)

Test applied to the examination of discordant results

(see references: 5.1.1.3.3; 5.2.2.6)

Count the total number of discordant results Y as follow:

$Y = PD + ND$ (for example $PD = 2$, $ND = 10$, then $Y = 12$)

Check if the two methods could be different for the balance of sensitivity versus specificity:

— for $\boldsymbol{Y < 6}$, (less than six disagreements): no test is available;

— for $\boldsymbol{6 \leq Y \leq 22}$, (i.e. between 6 and 22 disagreements), 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 according to the following Table F.1:

if $\boldsymbol{m \leq M}$ for a given Y, the two methods are **different** at $\alpha < 0{,}05$ (2-sided).

Table F.1 - M values for Y disagreements ($6 \leq Y \leq 22$)

Disagreements Y = PD + ND	6 to 8	9 to 11	12 to 14	15 to 16	17 to 19	20 to 22
M = Max(m) for $\alpha < 0{,}05$	0	1	2	3	4	5

For example, for $Y = 12$ disagreements and $m = 2$, $M = 2$ and $m \leq M$: thus, the two methods are different with $p < 0{,}05$.

— for $\boldsymbol{Y > 22}$, (more than 22 disagreements), use the McNemar test with the chi-square distribution for 1 degree of freedom:

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

The two methods are **different** at $\alpha < 0{,}05$ (2-sided) **if** $\boldsymbol{\chi^2 > 3{,}841}$.

This chi-square test corresponds to the minimal d for each Y of the following Table F.2 for $\alpha < 0{,}05$.

(That is for a given Y, d shall be equal or superior to the value given in Table F.2 for concluding that the two methods are different.)

Table F.2 - d values for Y disagreements ($Y > 22$)

Disagreements Y = PD + ND	22 to 26	27 to 31	32 to 37	38 to 44	45 to 51	52 to 58
d = $\lvert$PD - ND$\rvert$ ≥	10	11	12	13	14	15

Annex G
(normative)

Points to be considered when selecting strains for testing selectivity

(see references: 5.1.2.2; 5.1.3.2.1; 5.1.3.2.1.2; 6.2.4.2.1; annex H)

G.1 General

This annex outlines the minimum test requirements for general use. In the selection of test strains the majority should originate from the range of food materials used in the study and cover the recognised range of the target analyte with respect of the following – geographical distribution, incidence, diversity in identification characteristics e.g. biochemical, serotype, phage type, etc. and any claims made by the producers of the alternative method.

G.2 Target group categories

a) Undefined group for example total count, coliform, yeast, lactic acid bacteria;

b) family for example Enterobacteriaceae;

c) genus for example *Salmonella, Pseudomonas, Listeria*;

d) species for example *Listeria monocytogenes, Staphylococcus aureus, Escherichia coli*;

e) strain for example *Salmonella* enteritidis phage type 4.

G.3 According to the target group specified in G.1 a range of positive microorganisms can be chosen

f) For undefined groups for which the target group is defined by the reference method, the strains used shall be selected from those capable of typical growth in the reference method;

g) For families: use strains from a range of genera in that family and if possible include a representative member of all genera in the family;

h) For genera: use a range of species from that genus and if possible test all species in the genus;

i) For species: a range of strains from that species. The definition of strain should be taken into account when making this judgement. At present pathogens such as *Salmonella* and *Listeria* are serotyped, however *Listeria* spp are also phage typed, in the future other genetic typing methods such as RAPD typing, PFGE typing and ribotyping will be used. In defining the positive strains to use in tests, organising laboratories should use available up to date information to ensure that strains are relevant at that time to the target food categories;

j) for strain: a range of sources of that strain.

G.4 Non target groups used in selectivity study

a) The non-target groups (that is those expected to be negative and being used for cross reactivity tests) should be specified according to the target group;

b) When the target group is a family: non-target strains shall include families;

c) When the target group is a genus: non-target strains shall include other genera considered to be similar to the target genus;

d) When the target group is a species: non-target strains shall include other species within the target genus;

e) When the target group is a strain: non-target strains shall include other strains within the same species.

Annex H
(normative)

Guidelines for the organisation and conducting collaborative studies

(see references: 5.2.1.1; 5.2.1.3; 6.3.1; 6.3.3.1; annex J)

H.1 Preparation of food samples

Negative (control) samples.

In the general case only artificially contamination is to be used and the target analyte shall not be demonstrated in uninoculated food. Take a suitable numbers of samples in accordance with annex J and analyse them to confirm this. Negative controls shall also be used where stipulated during the trial.

Positive samples containing the target analyte.

For some quantitatives studies, e.g. total bacteria counts, yeasts and moulds, coliforms, etc, naturally contaminated samples, containing the target analyte and/or representative microflora should be used for collaborative studies. In the absence of suitable numbers of naturally contaminated samples, or where the level of contamination is too narrow or inappropriate, artificially contaminated samples may be used. Samples shall be contaminated by either the second or third option outlined in annex C. Where reference materials (annex C) 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.

In the case of both qualitative and quantitative methods, the food and target analyte/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.

As outlined in annex C, the food sample should contain a 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 food type, chosen from the food categories previously defined in annex B. The quality and suitability of the food shall be determined prior to the study.

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

H.2 Transport of the samples

The organising laboratory shall define the food type to be used and check that transit packaging is suitable for the intended journey.

The individual samples shall be double sealed to ensure that leakage from any one sample does not affect the integrity of all 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 meet international and national postal regulations.

The organising laboratory shall test posting and distribution methods by sending a set of samples to a collaborating laboratory requesting their immediate return, in order to ascertain the effects of transit conditions (safety and microbiological stability) prior to the study. Any detrimental effects shall be assessed and measures taken to minimise/prevent their further occurrence.

In order to maintain the target analyte or naturally occurring microflora unchanged during distribution it may be necessary to use refrigerated or frozen conditions during transport to the collaborating laboratories. The packaging conditions and best method of transport shall be determined by the organising laboratory. Ideally a suitable means of monitoring the temperature of the samples during transit is desirable.

For each laboratory, one additional pack, identical to the test samples, for measuring their temperature upon receipt, shall be provided.

H.3 Organisation of the collaborative study (see ENV ISO/TR 11133-1)

— Operating protocols

  Standard operating procedures shall be distributed to the collaborators for comment and familiarisation prior to commencement of the trial. 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 count of a designated sample. A standard method for enumeration shall be submitted to each collaborator by the organising laboratory.

— Alternative method and reference method

  Operatives conducting the analysis in each collaborative laboratory shall be skilled to an appropriate standard (and trained if necessary) and fully conversant with the alternative method and the reference method prior to the commencement of the trial.

— Reagents and operating conditions

  Additional factors, for example quality and composition of culture media and reagents, control of incubation temperatures etc. can have a profound effect on the outcome of the test. For this reason, such variability shall be minimised (for example by dispatching media/reagents to all collaborators) or be taken into consideration in the interpretation of data.

  The degree of tolerance permitted, in all aspects of the analysis, e.g. 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.

— Guidance

  The organising laboratory shall be available to provide advice or guidance to the collaborators during the entire period of the validation.

— Data collection

The organising laboratory shall develop a questionnaire that generates information regarding 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 content of the questionnaire shall be agreed, by both the organising laboratory and the alternative method manufacturer, in order to comply with the critical steps of the assay or test procedure of the alternative method.

— Confirmation of quality of test samples

Representative aliquots of the samples should be analysed by the organising laboratory on the day the analysis is scheduled to begin to confirm the presence and homogeneity of the target analyte. Samples, previously held under ideal conditions, should also be analysed by the organising laboratory on the day of commencement of analysis by the collaborating laboratories. These analyses shall be used only to confirm the inoculation procedure and stability of the analyte. The results should not be used in the statistical evaluation of the method under test.

Annex I
(normative)

Determination that negative controls are free of target analyte

(see reference: 5.1.2.2.)

In order to check the absence of the target microorganism in the product for use as either negative controls or before artificial inoculation, analysis of the test materials is essential. The larger the number of sub-samples examined the greater is the security for a reliable determination. Negative control samples are also confirmed during the study.

There is a probability that very low levels of analyte (numbers of microorganisms) are not detected by the amount of sample tested, or if present may not be homogeneously distributed through the sample.

On statistical grounds, if eight samples are tested, the probability of detecting the presence of 1, 2, 3, 4 or 5 cells per 600 g of product are 0,33; 0,57; 0,72; 0,83 and 0,90 respectively.

Divide the number of samples analysed into two categories:

— at least six samples are analysed before launching the experiment, in order to check the absence of the microorganisms, the number being determined by the test protocol;

— further negative samples are examined during the experiment, the number being determined by the test protocol.

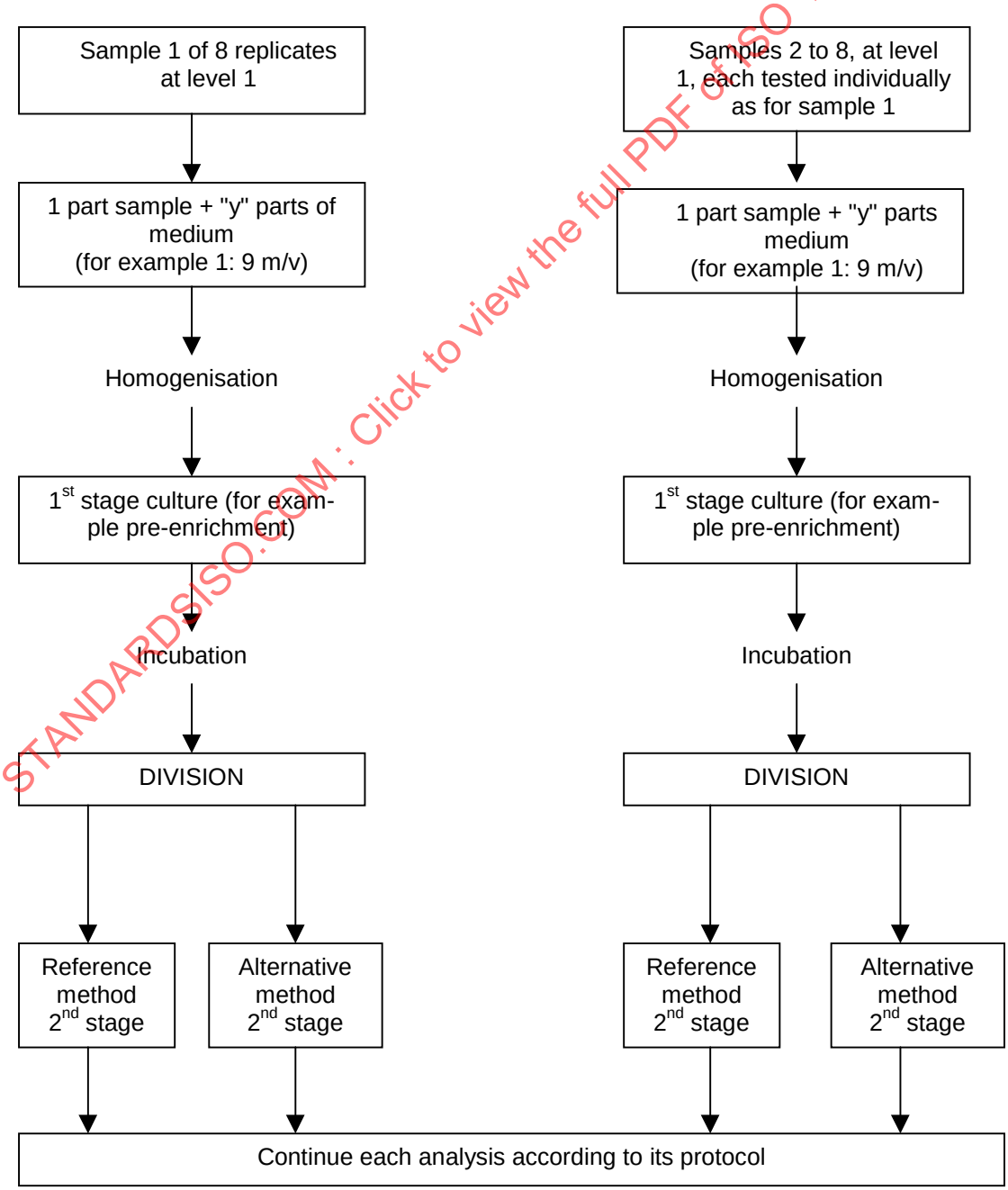
Annex J
(normative)

Replication of samples for interlaboratory studies of qualitative methods

(see reference: 5.2.1.2)

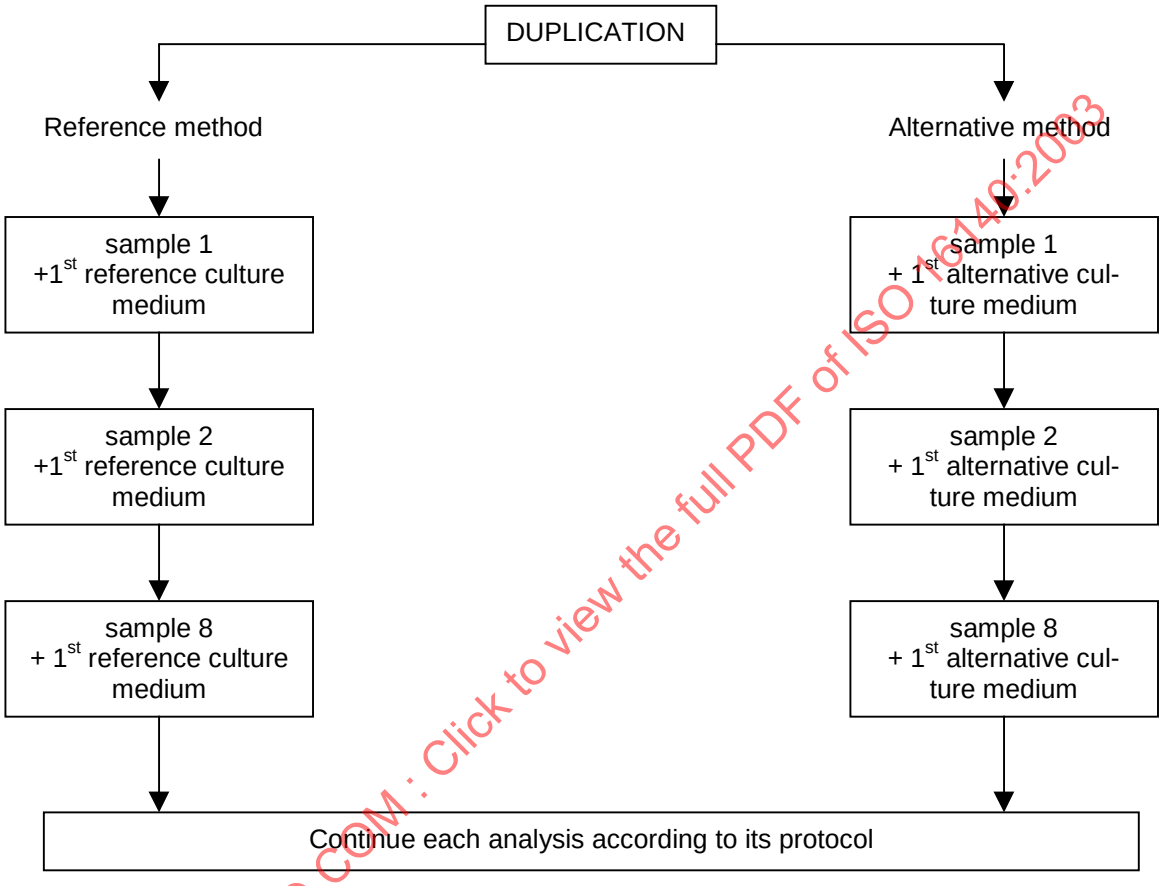
J.1 Case 1 - Where the first culture step of both reference and alternative methods are identical.

Eight replicate samples at each of three or four levels of contamination are tested by each method.



J.2 Case. 2 - Replication is by duplication when the first culture step of each method differs.

Eight replicate samples at each of three or four levels of contamination are tested in each method.



Annex K
(normative)

Consideration of data

(see reference: 5.2.1.3)

When the interlaboratory study is completed, all the information on data sheets and the results shall be submitted to the organising laboratory and examined as follows:

— check that samples/test kits, etc. were not damaged during transit;

— give the trademark and batch number of the media used. The expert laboratory should check if the formula is in accordance with the reference method;

— where sample damage occurs the data shall be disregarded. Cross contamination of other samples in the same packaging should also be considered;

— disregard data if transit conditions and times fall outside the specified acceptable tolerances;

— disregard data if enumeration values fall outside the specified acceptable tolerances;

— disregard data if the questionnaire suggests that the laboratory has deviated from either the standard protocol or the critical operating conditions.

Annex L
(informative)

Interlaboratory study of qualitative methods: criteria of accordance, concordance and concordance odds ratio

(see reference: 5.2.3)

L.1 General

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

This annex provides further criteria (accordance, concordance and concordance 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 different laboratories. Since the difference for data that is not quantitative cannot be used, statistics for qualitative methods are instead based on the probability (expressed as a percentage) that two samples both produce the same result.

These criteria have been developed by the European project SMT CT 96 2098 funded by European Commission/ DG XII to validate the six main standardised methods used in food microbiology (co-ordinator: Dr C. Lahellec, AFSSA, France) [9].

The calculations are illustrated by means of the following data (see Table L.1).

Table L.1 — Example of numerical values

	Replicate number					
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 a collaborative trial would be larger, but a smaller dataset makes it easier to explain the calculations.

L.2 Accordance

L.2.1 Definition

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

The accordance is therefore the equivalent of repeatability for quantitative methods.

L.2.2 Calculation

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

For examples, see Tables L.1 and L.2.

For those laboratories (such as laboratory 1) where all samples were found positive, the best estimate of the probability of getting the same result is clearly 1,00 or 100 %.

For the others (in the example, laboratories 5 and 7) the probability that one replicate will be positive is 3/5=0,60 and 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 negatives ($0,4^2$ = 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 laboratories (see Table L.2).

Table L.2 — Calculation of accordance

Lab	Number of positives	Proba of positive	Proba of pair of positives	Proba of negative	Proba of pair of negatives	Proba of pair of same 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 (mean) 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.

L.3 Concordance

L.3.1 Definition

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

The concordance is therefore the equivalent of reproducibility for quantitative methods.

L.3.2 Calculation

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

The concordance is the percentage of all pairings giving the same results on all the possible pairings of data.

For example, see Tables L.1 and L.3.

Taking each replicate in each laboratory in turn, start with the first replicate of laboratory 1 which is positive. This can be paired with any of the 45 replicates from other laboratories, and all but 4 of these pairings (those with two replicates of laboratories 5 and 7) match (i.e. give a pair with both positive), 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) between-laboratory pairings of replicates involving laboratory 1, of which 205 (5 x 41) give the same result.

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

For laboratory 5, with 3 replicates out of 5 positives, the 2 negative replicates each match with just 2 other negative replicates of 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 (2x2 + 3x43).

Table L.3 — Calculation of concordance

Lab	Number of positives	Between-lab pairings with the same results	Total between-lab 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		**1 906**	**2 250**

The concordance is the percentage of all pairings of duplicates giving the same result; in this example this is 84,7 % (1 906/2 250 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.

L.4 Concordance odds ratio

L.4.1 General 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 ones, suggesting that there can be 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 strongly dependent on the level of accuracy, making it difficult to assess easily the degree of between-laboratory variation.

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

$$\mathrm{COR} = \frac{\text{accordance} \times (100 - \text{concordance})}{\text{concordance} \times (100 - \text{accordance})}$$

L.4.2 Significance tests

A value for the odds ratio of 1,00 would be expected if accordance and concordance were equal, and the larger the odds ratio is, the more inter-laboratory 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 evidence for extra variation between laboratories is convincing. The “exact test” is the best recommended test for this[10]. The philosophy behind such tests is that the probabilities of occurrence are calculated for all sets of replicate results that could have produced the overall numbers of positives and negatives.

For example, see Table L.1.

With in total 46 positives and 4 negatives, the possible arrangements are given as columns in Table L.4.

Table L.4 — Possible arrangements of positive replicates (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

[10] This test can be performed using statistical packages such as SAS®.

The actual arrangement is shown in bold (third column) in Table L.4. The test adds up the probabilities P of all those possible arrangements that show at least as much evidence for between-laboratory variation as the real arrangement - here that means all permutations of the three columns on the right. If this probability is less than the conventional value of 0,05 or 5 **%**, it is unlikely that this degree of between-laboratory variation could have occurred by chance and hence, and it is concluded that there is significant variation in performance between laboratories.

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

Where software for the “exact test” is not available, an ordinary chi-squared analysis for contingency tables can be used as an alternative. The results of this test is 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 between-laboratory differences.

With either test, it shall be remembered that the ability to detect between-laboratory differences is dependent on the number of laboratories and the number of replicate samples 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 a standard error or confidence limits, but the distribution of the odds ratio is highly skewed making it very difficult to produce reliable limits.

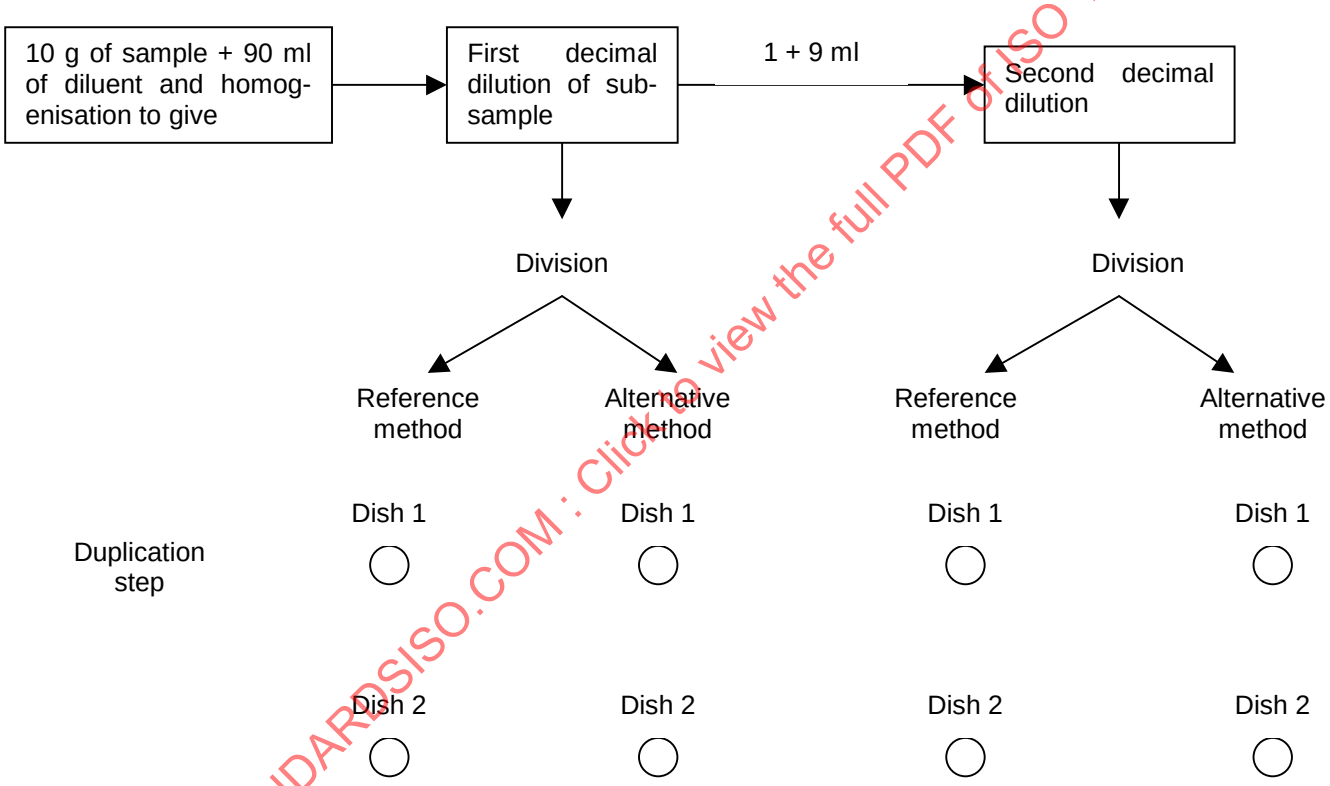
Annex M
(normative)

Replication of samples for the determination of relative accuracy of quantitative methods

(see reference: 6.2.1.2.1)

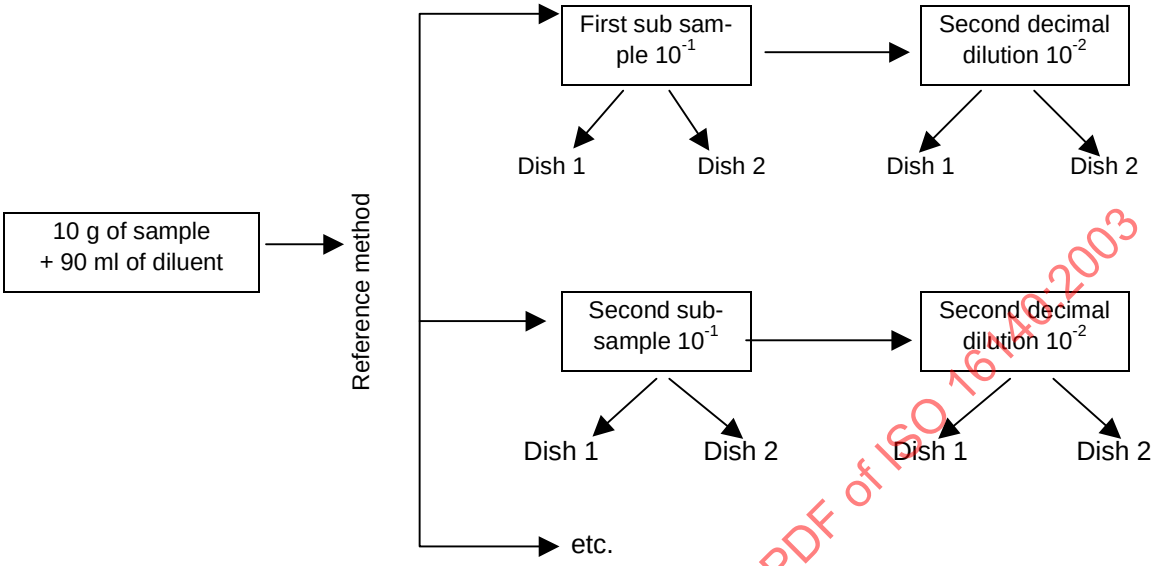
M.1 Case 1 - Where both the reference and alternative methods use the same decimal dilutions of each sub-sample

The same number of sub-samples (at least 2 and preferably 5 to 10) are cultured by both methods.



M.2 Case 2 - Where only the reference method requires decimal dilutions

The same number of sub-samples (at least 2 and preferably 5 to 10) are cultured by both methods.



Annex N
(normative)

Examples of acceptable and unacceptable situations and range of measurements for the estimation of the regression line for quantitative methods

(see reference: 6.2.1.1.1)

N.1 Unacceptable situations

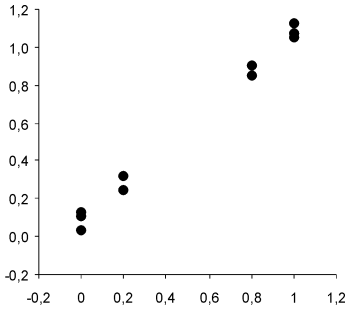


Figure N.1 — Lack of centre points

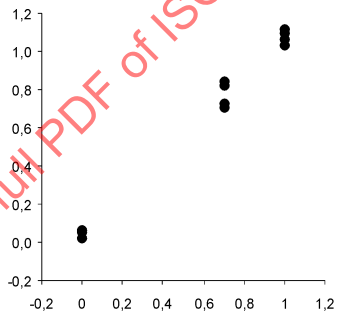


Figure N.2 — Bad repartition

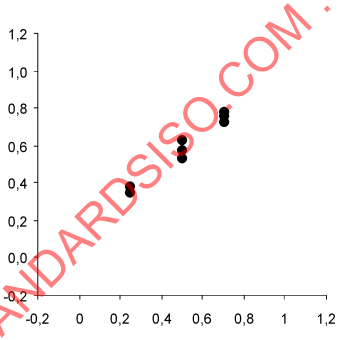


Figure N.3 — Range too narrow to cover 0 to 1

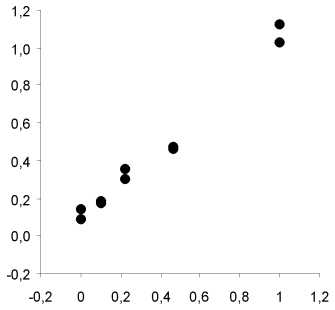


Figure N.4 — For use on a long-scale, only if SD increases with y

N.2 Acceptable situation

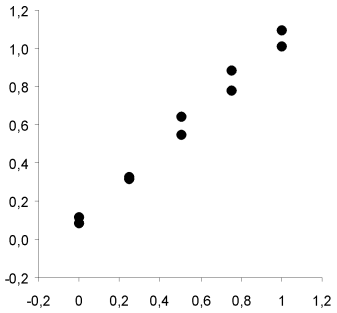


Figure N.5 – Acceptable situation

Annex O
(normative)

Assessment of the linearity of quantitative methods by graphical representation

(See reference: 6.2.1.4.2)

O.1 Cases of non-linearity

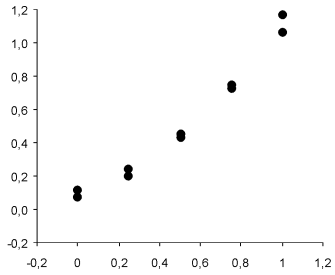


Figure O.1

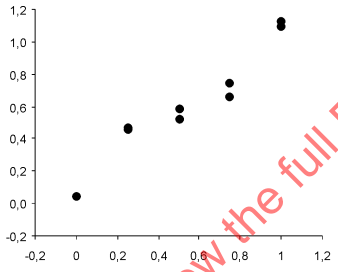


Figure O.2

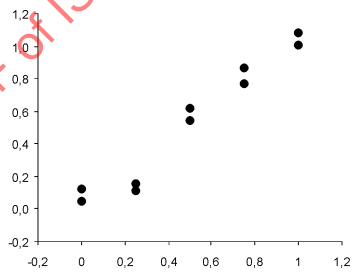


Figure O.3

O.2 Cases of linearity

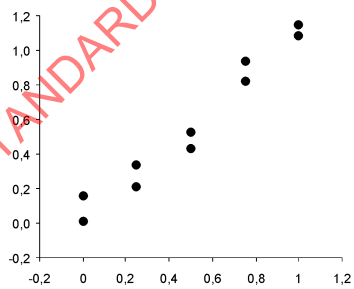


Figure O.4

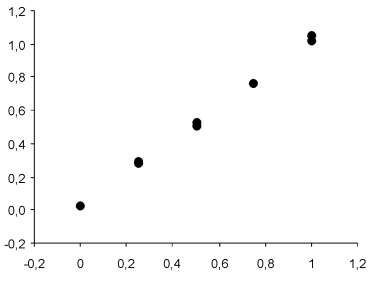


Figure O.5

Annex P
(normative)

Detection and quantification limits for counts

(see reference 6.2.2.3)

P.1 Detection limit

Normally, in the presence of method uncertainty, related for example to the sample used, the observation of only **one colony forming unit (cfu)** on the medium is not sufficient to presume the *presence* of analyte. At this low level, a dichotomisation of the phenomenon into *presence* (1 or more cfu) and *absence* (no cfu) is necessary, as for **qualitative** methods. Therefore, the process follows either a poissonian law (low or high rates) or a binomial one (intermediate rates) along the *presence* frequency *np*.

Considering that a theoretical negative control gives no result of *presence* of analyte, a detection limit for a low contamination (poissonian) is reached by at least *np presences* corresponding to the **decision levels *LC*=3** (at 1 - α = 95 %) or **5** (at 99 %). **Detection limits** are also obtained including a type-2 statistical error β for instance ***LOD* = 7** at 1-sided α = 5 % and 1 - β > 90 %.

The poissonian law gives the **minimum number *np* of *positive*** results among *n* (proportion *p*) to obtain, in order to reach a significantly *positive* statement at the 1 - α confidence level (see Table P.1):

Table P.1 — Minimum number *np* of *positive* results among *n*

n	***np* for 1 - α = 0,95**	***np* for 1 - α = 0,99**
1	1	1
2	2	2
3	2	3
4	3	3
5 to 15	3	4
16 and over	3	5

However, the other type of error is not avoided: the **lack of power** if the sampling size *n* is too small. For instance, having *no positive with n = 2 samples* only implies that the probability of contamination is less than 78 % (with 95 % of confidence). *1 positive over 2* implies that the probability of contamination is 2,5 % or more, with the same confidence! It can be seen that, **to reach a significantly *positive* statement at the 95 % or 99 % confidence levels**, the minimal probability *p* of contamination decreases when *n* increases: ***p* ≥ 3/*n*** and respectively **5/*n*** (more precisely 2,996/*n* and 4,605/*n*). It implies mainly that, if it is aimed to **detect a probability of contamination *p*** = 10 % (at 95 %), more than 30 samples (***n* ≥ 3/*p***) have be to taken. The minimum level *p* is therefore indicative.

P.2 Determination limit

The application of the AOAC's rule ***to counts*** could give ***LOQ* ≥ 100…** An *a priori* poissonian distribution is used in this EN ISO 16140 with *s*(n) = $\sqrt{n}$, and CV(n) = *s*(n)/n = 1/$\sqrt{n}$ ≤ 10 %, which gives $n \geq 100$.