INTERNATIONAL STANDARD

ISO 17994

First edition
2004-05-15

Water quality — Criteria for establishing equivalence between microbiological methods

Qualité de l'eau — Critères pour établir l'équivalence entre les méthodes microbiologiques

Reference number
ISO 17994:2004(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

Contents

Page

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 document may be the subject of patent rights. ISO shall not be held responsible for identifying any or all such patent rights.

ISO 17994 was prepared by Technical Committee ISO/TC 147, *Water quality*, Subcommittee SC 4, *Microbiological methods*.

Introduction

This International Standard presents the criteria and procedures for assessing the average quantitative equivalence of the results obtained by two microbiological analytical methods one of which may but need not be a standard or reference method.

The methods considered are based on counts of colonies or of positive and negative liquid enrichment tubes (MPN and presence/absence methods).

NOTE It is possible that a method that is not quantitatively equivalent with a reference method would be accepted, especially if it appears “better” than the reference either quantitatively or otherwise.

Water quality — Criteria for establishing equivalence between microbiological methods

WARNING — Persons using this International Standard should be familiar with normal laboratory practice. This International Standard does not purport to address all of the safety problems, if any, associated with its use. It is the responsibility of the user to establish appropriate safety and health practices and to ensure compliance with any national regulatory conditions.

1 Scope

This International Standard defines an evaluation procedure for comparing two methods intended for the detection or quantification of the same target group or species of microorganisms.

This International Standard provides the mathematical basis for the evaluation of the average relative performance of two methods against chosen criteria of equivalence.

Any two enumeration methods based on counts (of colonies or positive tubes) or any two detection methods [presence/absence (P/A) methods] intended for the same purpose can be compared.

This International Standard provides no solution to directly compare a quantitative method (colony count or MPN) with a detection method (P/A).

2 Normative references

The following referenced documents are indispensable for the application of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO/TR 13843:2000, *Water quality — Guidance on validation of microbiological methods*

3 Terms, definitions and symbols

3.1 Terms and definitions

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

3.1.1 General terms

3.1.1.1
reference method
prescribed analytical method to analyse a given group or species of microorganisms

NOTE As a rule, the reference method is a standard or a commonly used method.

3.1.1.2
trial method
any method which is to be tested for equivalence with a reference method

3.1.1.3
equivalent method
method considered quantitatively equivalent with another method when the average relative difference of their confirmed counts is found “not different” when following the calculations specified in this International Standard

3.1.1.4
standard uncertainty
uncertainty of the result of a measurement expressed as a standard deviation

[GUM]

3.1.1.5
expanded uncertainty
quantity defining an interval about the result of a measurement that may be expected to encompass a large fraction of the distribution of values that could reasonably be attributed to the measurand

NOTE The fraction may be viewed as the coverage probability or level of confidence of the interval. To associate a specific level of confidence requires explicit or implicit assumptions regarding the probability distribution. The level of confidence may be attributed to this interval only to the extent to which such assumptions may be justified.

[GUM]

3.1.1.6
coverage factor
numerical factor used as a multiplier of the (combined) standard uncertainty in order to obtain an expanded uncertainty

NOTE The coverage factor, $k = 2$ is chosen for this International Standard because the distribution of the relative difference is unlikely to be normal; the expanded uncertainty corresponds only approximately to the 95 % confidence interval.

3.1.2 Specific terms

3.1.2.1
count
observed number of objects, e.g. colonies or cells of microorganisms, plaques of bacteriophages

NOTE In this International Standard, the result of an MPN estimation is also considered a count.

3.1.2.2
presumptive count
number of objects that according to their outward appearance should presumably be included in the count

3.1.2.3
confirmed count
count corrected for false positive results by further testing of the presumptive objects

3.1.2.4
relative difference
RD
difference between two results, a and b, measured on a relative (natural logarithmic) scale

NOTE The value of RD, x, expressed in percent, is given by

$$x = [\ln(a) - \ln(b)] \times 100\ \%$$

Essentially the same result is given by

$$x = \frac{2(a-b)}{(a+b)} \times 100\ \%$$

until the ratio between a and b is greater than about 3. This accounts for the usage of the term “relative difference” in both cases.

3.2 Symbols and abbreviated terms

A	the (symbol for the idea of) trial method
a	a test result by Method A
a_i	the test result (confirmed count) of Method A in sample i
B	the (symbol for the idea of) reference method
b	a test result by Method B
b_i	the test result (confirmed count) of Method B in sample i
C	coefficient for computing the number of samples, given the value of the experimental variance
D	maximum acceptable deviation (value of confidence limit) in the case Methods A and B are “not different”
i	running index
k	coverage factor used for calculating the expanded uncertainty
L	smallest significant (i.e. maximum acceptable) relative difference between Methods A and B
MPN	most probable number quantitative method
m	number of parallel tubes per dilution in an MPN series
n	number of samples
n_A	number of samples where for the P/A Method, A is positive and B negative
n_B	number of samples where for the P/A Method, A is negative and B positive
P/A	presence/absence detection method
s	experimental standard deviation (standard uncertainty)
s^2	experimental variance
$s_{\bar{x}}$	standard deviation (standard uncertainty) of the mean
U	expanded uncertainty
x	relative difference
x_i	value of the relative difference between a_i and b_i in sample i
$\bar{x}$	arithmetic mean of x_i (i = 1,2,…,n)
x_L	value of the relative difference at the approximate lower 95 % confidence limit, derived by subtracting the value of the expanded uncertainty from the mean
x_H	value of the relative difference at the approximate upper 95 % confidence limit, derived by adding the value of the expanded uncertainty to the mean
X^2	experimental Poisson index of dispersion
y	conditional variable used in computing the number of samples for equivalence testing and/or verification

4 Principle

The basic data are pairs of confirmed counts (a_i, b_i) obtained from the examination of two equal portions taken from the same vessel of a carefully mixed test sample, one determination (count) per method. The complete design consists of a large number of similar determinations.

In this International Standard, two methods are considered quantitatively equivalent ("not different") if the mean relative difference of the paired confirmed counts does not differ significantly from zero and the expanded uncertainty does not extend beyond the level of the stipulated maximum acceptable deviation. The decision rules based on the above principle are detailed in 7.2 and 7.3 and a flow chart is given in Annex A.

NOTE 1 Fixing a value for the maximum acceptable deviation (D) implies indirectly that the smallest average difference (L) to be considered significant is one half of that value.

NOTE 2 It has been suggested that in international and inter-laboratory method performance tests a limit of D = 10 % for the "confidence interval" be the maximum acceptable deviation for drinking water[2].

NOTE 3 For chemical methods, mean and precision are used as criteria for equivalence. In microbiology, equal precision (equal variance) is not an equivalence criterion.

5 Basic requirements for an equivalence experiment

5.1 General

Both methods shall fulfil at least the minimum requirements of validity specified in ISO/TR 13843.

The most important basic requirement of equivalence trials is a wide range of samples. Participation by a number of laboratories is usually necessary to expand the sample range over large geographical areas. Also the credibility of a general conclusion is commonly believed to require the participation of several laboratories. The result of the comparison is generally valid only within the range of sample types studied. Collaborative trials are detailed in Annex B.

It is essential that all laboratories taking part in a collaborative study have recognized quality assurance systems in use and apply approved basic techniques of cultivation.

5.2 Types of samples

The requirements for method comparisons differ somewhat from the daily routine situation. It is useful and often necessary to pre-select or prepare special samples. Samples for method comparisons should contain enough bacteria that the likelihood of scoring a zero count is small.

Samples for method comparisons should represent types that are included in the scope of both methods. Natural samples are ideal. Appropriate samples may also be prepared by dilution, spiking, or mixing of different kinds of water to achieve the desired population in a suitable density. Spiking with pure cultures should be considered the last resort.

It may be appropriate to stress the microbial population of some samples by controlled application of disinfectants[2] or by refrigerated storage in order to simulate situations encountered in routine laboratory practice.

5.3 Number of samples and participating laboratories

5.3.1 General

It is not possible to determine beforehand the exact number of samples required for a valid comparison. The number depends on the actual difference observed, on the experimental standard deviation and on the difference considered significant. This International Standard includes an adequacy clause based on a

stipulated “maximum acceptable deviation” and the expanded uncertainty. If the data are found inadequate for deciding that the methods are “not different”, more samples are to be collected and examined.

If the methods happen to differ markedly, a small number of samples might suffice to determine this fact. It is therefore advisable to proceed in stages. The first stage should be planned to detect large differences between the methods. If large differences are not found (result inconclusive), more samples are taken until the system is able to detect the average difference that corresponds with the maximum acceptable deviation chosen at the beginning of the trial. Tables are given in 5.3.3 and 5.3.6 to provide help for planning.

5.3.2 The number of laboratories

There are no previous standards or rules about the number of laboratories in inter-laboratory equivalence trials. Six is tentatively suggested as minimum number.

5.3.3 Number of samples, two colony methods

The total number of samples, n, sufficient for the detection of a given average relative difference at about 95 % confidence depends on the experimental variance according to the equation:

$$n = Cs^2$$

where

s^2 is the variance;

C is a coefficient that depends on the chosen least significant difference.

The value of C is derived from the relationship: $C = 4/L^2$. The relationship between D (the maximum acceptable deviation) and L (the least significant difference) is: $L = D/2$ (see Table 1).

Table 1 — Coefficients for determining the number of samples required for the detection of a given relative difference (L)

D[a] %	L %	C
60	30	0,004 4
40	20	0,010 0
30	15	0,017 8
20	10	0,040 0
10	5	0,160 0

[a] The corresponding maximum acceptable deviation (D) is shown for comparison.

EXAMPLE A rather frequently observed value for the experimental standard deviation of the relative difference is approximately s = 80. Inserting this value in the equation gives n = 6 400C. In order to detect an average relative difference of 10 % units (L = 10 %), n = 6 400 × 0,040 0 = 256 samples should be sufficient.

5.3.4 Number of samples, two MPN methods

With MPN methods the number (n) of samples depends on the number (m) of parallel tubes according to the equation:

$$n = 1\,700/m$$

With five parallel tubes per dilution, 1 700/5 = 340 samples should suffice for the detection of a 10 % relative difference.

5.3.5 Number of samples, mixed comparisons

When one of the methods is an MPN Method and the other a colony method the number of samples likely to be required is halfway between that in (5.3.3) and (5.3.4).

NOTE With some recent methods based on chromogenic substrates, it is possible to estimate two bacterial groups simultaneously. One of the groups may be ten or more times as numerous as the other. The number of samples sufficient for making a final decision of equivalence with the more numerous type may not be sufficient for the organism present in low numbers.

5.3.6 Number of samples, two P/A methods

Testing the equivalence of P/A methods is wasteful of samples. Samples where both methods give the same result (+ +) or (– –) do not contribute to the statistical evaluation of equivalence. The number of samples expected to be sufficient for detecting an average relative difference of L at 95 % confidence is indicated in Table 2. The number given is the total number of samples with unequal results (+ –) and (– +). The experiment should not be done with fewer samples to avoid the false verdict "not different" due to insufficient data.

Table 2 — The number of samples required in order to detect whether the average relative recovery of two P/A methods is greater than L

L %	n
40	100
30	170
20	380
15	680
10	1 540
5	6 140

5.3.7 Number of samples when comparing a counting method with a P/A method

Counts and qualitative test results are not commensurable. The comparison of a counting method with a P/A method requires either "down-grading" the count results or "up-grading" the P/A method.

Down-grading is effected by ignoring the numerical value of each count and replacing it with a (+) meaning a positive detection of the presence of an organism. Most of the information contained in the counts is thereby lost. In this case, the number of samples required for comparing two P/A methods applies (5.3.6).

NOTE 1 Because the conversion of the count to a mere (+) takes place after the count has been confirmed, no work is saved in confirmation.

Subdividing a P/A sample after inoculation into several small volumes "up-grades" the P/A method into a single dilution MPN method. For instance, a 100 ml volume can be subdivided into 10 portions of 10 ml or, preferably 20 portions of 5 ml providing an MPN estimate that can be compared with the count of the other method. In this case, the number of samples is the same as required for the comparison of two MPN methods (5.3.4) or for mixed comparisons (5.3.5).

NOTE 2 Upgrading a P/A method into an MPN may seem like a radical change. All essential properties of microbiological methods except the precision are, however, determined by the chemical formulation of the nutrient medium and the incubation conditions. They remain unchanged. Precision is not an equivalence criterion in microbiology. Upgrading merely improves the precision thus making the comparison more efficient.

5.3.8 The minimum requirements for verification

It may be sometimes necessary to verify the result of a comparison under ecological conditions or in a geographical area not represented in a collaborative trial. When a laboratory only needs to verify a method already tested and officially accepted, it can take full advantage of the previous test results.

The laboratory should have access to the report of the collaborative comparison. Accordingly, it should have at its disposal estimates of the mean and standard deviation of the relative difference. In this case, it can estimate the recommended number of samples from the formula given below. The same formula can be applied to estimate the number of additional samples when the initial number has been found inadequate (see 7.2.4 and 7.3.5). However, whatever the result of the calculation the number of samples studied should not be less than thirty.

$$n = 4\left(\frac{s}{y}\right)^2$$

where

n is the number of samples required;

s is the standard deviation of the relative difference;

y is the larger of the two quantities:

$$y_1 = \bar{x}$$
$$y_2 = |\bar{x}| - |D|$$

D is the maximum permitted deviance from 0 of the expanded uncertainty in the case that the methods are “not different” in % units;

$\bar{x}$ is the arithmetic mean of the relative difference in % units.

NOTE In the case of a one-sided evaluation (7.3), y_2 is calculated as $y_2 = \bar{x} - D$, with the algebraic signs of D and $\bar{x}$ taken into account.

If the only prior information is the knowledge that the method has been officially approved, the recommended numbers of samples for verification are as follows.

The minimum requirement for MPN and colony methods is twice the number required of each laboratory in a collaborative trial.

In the case of two P/A methods, enough samples should be examined to give a total of at least 400 samples with (+ −) and (− +) results.

5.4 Counting and confirming

5.4.1 Counts

Counts shall be recorded to the last digit observed. Rounding of the counts to two figures, as often specified for reporting routine results, shall not be practiced. MPN results shall be rounded to the nearest whole number when results per 100 ml are given.

5.4.2 Confirmation

Most of the methods are of a type that requires confirmation of the presumptive observations. The basic rule of confirmation in method comparison trials is to confirm every count by testing every presumptive positive observation (colony or tube).

Detailed advice concerning confirmation shall be given in the instructions from the panel of experts (see Annex B).

To confirm a colony count reliably, the presumptive colonies should be well separated from each other and from background colonies. For practical and statistical reasons, the ideal range is from 10 to 30 target colonies per plate. Dilution of the samples should be chosen accordingly.

It is recommended to use dilutions of 1:3, 1:2 or even less. This should allow the selection of the plate(s) that fulfil the basic quality criteria of reliable confirmed counts. If the plan fails to produce any plates with well separated colonies, the sample shall be discarded. Another sample from the same source (if considered indispensable) should be taken and diluted such that the objective of a reliable count can be achieved.

Both the presumptive and confirmed counts shall be recorded and reported. This provides data for comparison of the confirmation rates. Methods with the possibility of *in situ* confirmation (by transplantation of the membrane filter, for instance) give the confirmed count directly. In such cases, the presumptive results need not be reported.

Every presumptive positive observation in MPN and P/A methods shall be confirmed.

6 Calculations

6.1 Preliminary editing of the raw data

Samples shall be excluded from calculations when both methods give a confirmed count of zero (0,0) or either method gives a result other than a count (e.g. TNTC, "larger than", etc.). The primary observations, except the presumptive result (0,0), should be reported nevertheless.

6.2 The basic relative differences

6.2.1 Regular count data

An estimate of the relative difference is calculated for every pair of non-zero confirmed counts (a_i, b_i) as follows:

$$x_i = \left[\ln(a_i) - \ln(b_i)\right] \times 100\ \%$$

6.2.2 Results with zeroes

Some results of the types (a_i, 0) and (0, b_i) are almost inevitable. To avoid omitting these samples, the relative differences are calculated using the formulae below.

When the result is of the type (a_i, 0), the relative difference is obtained from

$$x_i = \ln(a_i + 1) \times 100\ \%$$

When the result is of the type (0, b_i), the relative difference is obtained from

$$x_i = -\ln(b_i + 1) \times 100\ \%$$

NOTE 1 The equations are the result of adding the constant one (1) to each "count" when one of them is a zero. For instance, in the case of (a_i, 0), the values entered in the general formula are (a_i +1, 0+1).

NOTE 2 Adding a constant reduces the mathematical problems caused by zeroes but does not completely erase them. It is therefore advisable to try to minimize the number of samples with zero results. At least 75 % of the samples should contain regular count data.

6.3 The mean relative difference

The average relative performance is estimated in % units according to

$$\bar{x} = \frac{\sum x_i}{n}$$

where

n is the number of samples;

x_i is the relative difference in sample i.

6.4 The expanded uncertainty

The standard uncertainty (standard deviation) of x is obtained from the conventional formula for experimental standard deviation

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$$

The standard uncertainty of the mean (formerly "standard error") is computed from

$$s_{\bar{x}} = \frac{s}{\sqrt{n}}$$

The expanded uncertainty is derived from the standard uncertainty of the mean by using the coverage factor $k = 2$

$$U = k s_{\bar{x}} = \frac{2s}{\sqrt{n}}$$

To evaluate the result of the comparison the "confidence interval" of the expanded uncertainty around the mean is calculated by computing the limits

Lower limit: $x_L = \bar{x} - U$

Upper limit: $x_H = \bar{x} + U$

6.5 P/A methods

For statistical evaluation of P/A results two sums are needed:

n_A is the number of samples where Method A is positive and Method B negative;

n_B is the number of samples where Method A is negative and Method B positive.

The evaluation of P/A methods is based on the Poisson index of dispersion, X^2, (sign test and McNemar's test are other suitable alternatives). The value of the index is given by

$$X^2 = \frac{(n_A - n_B)^2}{(n_A + n_B)}$$

7 Evaluation

7.1 Preliminary evaluations

7.1.1 Examination by groups

The evaluation of equivalence should begin with different groupings of the data to detect possible differences between laboratories or categories of samples. The analysis of variance or its non-parametric equivalents are likely to be suitable methods. Large differences attributable to the origin of samples might lead to a recommendation to exclude certain sample types from the scope of the method.

If great differences between laboratories seem to be observed, they should be critically examined. In extreme cases, the statistical analyst may recommend exclusion of the results of a laboratory from a collaborative trial. Deviation from the agreed protocol or demonstrable technical problems are valid reasons for exclusion.

In addition to the grouping by laboratory, the data should be grouped by the type or origin of the samples for similar analyses of heterogeneity.

NOTE In microbiology, the method giving the highest confirmed results is always closest to the truth. Reference methods occupy no protected position. They can be deemed unsuitable for the types of samples studied when a trial method yields significantly higher results than the reference method.

7.1.2 Outlier detection

This International Standard provides no formal tests for the detection of outlier laboratories, unsuitable sample types, or individual outlying test results. It is left for the statistical analyst to decide upon the use of outlier tests. Individual outlier values can usually be detected graphically by plotting $\ln(a_i)$ against $\ln(b_i)$ points. Omission of results may be advised by the statistical expert. The great random variation associated with the low counts characteristic of method comparisons is a major difficulty. Particular attention should be paid to the higher counts where differences between methods become pronounced and the linearity of the association may fail.

If no problems with outliers, or significantly different behaviour between laboratories or sample categories are observed, the data from all laboratories and all samples can be merged into one analysis to provide a powerful general test of equivalence. Otherwise the evaluation should proceed by groups and the conclusions should be formulated accordingly.

7.2 Two-sided evaluation

7.2.1 General

Whether one method can be named the reference or not, equivalence is normally understood to mean that neither method gives significantly higher or significantly lower results than the other. For special practical purposes, asymmetric maximum acceptable deviation limits might be agreed, i.e. different values for the lower ($-D$) and higher ($+D$) side.

7.2.2 Methods “not different”

The methods are “not different” when

$$-D \leqslant x_L \leqslant 0 \text{ and } 0 \leqslant x_H \leqslant +D$$

7.2.3 Methods “different”

The methods are “different” when

$x_L > 0$ or $x_H < 0$

7.2.4 Inconclusive

The data are insufficient for a decision when

$x_L < -D$ and $x_H > 0$ or

$x_L < 0$ and $x_H > +D$

More samples should be examined. The number of additional samples required can be estimated as shown in 5.3.8 (see also Annex C).

7.2.5 Indifferent

The methods are statistically different but the difference is too small to be of practical significance (microbiologically) when

$x_L > -D$ and $x_H < 0$ or

$x_L > 0$ and $x_H < +D$

It is an arbitrary choice which argument is considered decisive (statistical or practical). From a practical point of view, the methods are considered to be equivalent.

7.3 One-sided evaluation

7.3.1 General

It is possible that the expert panel or a regulatory agency decides to accept an alternative method whenever its average performance is either quantitatively equivalent or higher than the reference method[2]. In such cases, only the lower value of the maximum acceptable deviation ($-D$) is of concern in the evaluation. The result categories differ somewhat from those in the two-sided evaluation.

7.3.2 Methods “not different”

Methods are “not different” when

$-D \leqslant x_L \leqslant 0$ and $x_H > 0$

7.3.3 Trial method: higher recovery

The trial method has a (significantly) higher recovery than the reference method when

$x_L > 0$

In a one-sided evaluation, the alternative method is considered acceptable even though the methods are not mathematically equivalent.

7.3.4 Trial method: lower recovery

The trial method has a (significantly) lower recovery than the reference method when

$$x_H < 0$$

7.3.5 Inconclusive

The data are insufficient for a decision when

$$x_L < -D \text{ and } x_H > 0$$

More samples should be examined. The number required can be estimated as shown in 5.3.8 (see also the Note in 5.3.8).

7.3.6 Indifferent

The trial method gives a significantly lower recovery (statistically) but the average relative difference is probably of no practical significance (microbiologically) when

$$x_L > -D \text{ and } x_H < 0$$

7.4 Two P/A methods

When the value of the Poisson index of dispersion $X^2 \geqslant 4$, the methods are considered to be “different”, when $X^2 < 4$ the methods are considered to be “not different”.

8 Test report

The test report of a collaborative evaluation trial should include the following information:

a) a reference to this International Standard (ISO 17994:2004);

b) an unambiguous exposition of, or reference to, the methods;

c) relevant descriptive details of the experiment (numbers of samples, participants, the maximum acceptable deviation);

d) evaluation in words (“not different”, “different”, etc.);

e) mean relative difference;

f) standard deviation of the relative difference;

g) an annex of the raw data.

Annex A
(informative)

Flowchart

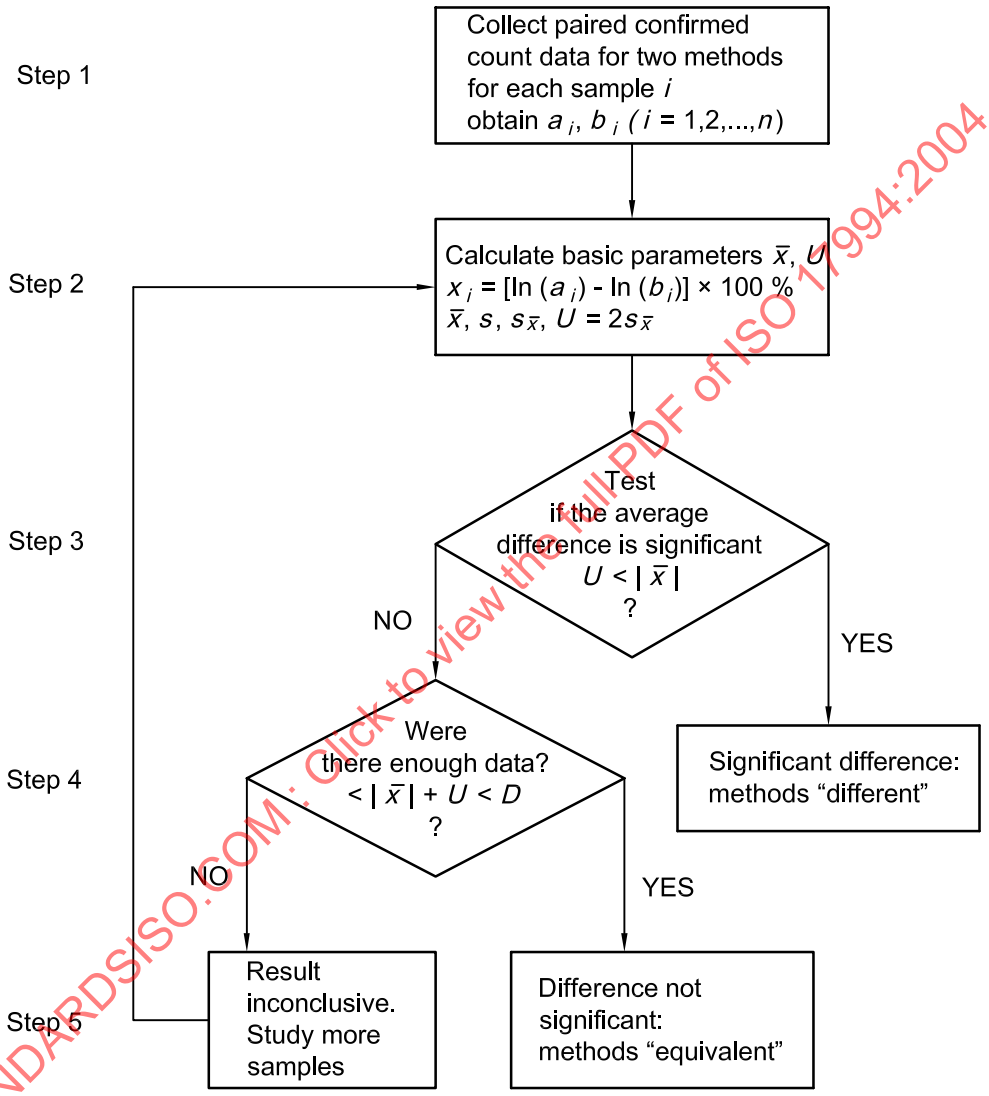

Figure A.1 — Flowchart

Annex B
(informative)

Collaborative equivalence trials

B.1 General

Collaborative trials are organized in order to expand the geographic and environmental scope of the samples. Involvement of different laboratories also indirectly tests the robustness of the method.

B.2 Panel

A panel of experts should be appointed to plan and co-ordinate the experiment. Some of the tasks of the panel are to choose the participating laboratories and to discuss the methodologies to be tested and used. Among its important tasks is to decide on the definition of the target organism(s) and the tests used for confirmation of presumptive positive results. If the quantitative criterion for the methods to be "not different" (the maximum acceptable deviation or the least significant average difference) is not set by higher authorities, the panel shall choose it.

The panel should include statistical expertise. The expert should be engaged at the planning stage to ensure a satisfactory experimental design.

The panel should appoint someone for the executive functions to enlist the laboratories and to draft a protocol covering the laboratory procedure and data reporting. The executive officer is also responsible for collecting the data and presenting it to the statistical expert.

The task of the statistical expert is to contribute to, and approve of, the design of the experiment, to analyse the data and to write a report to the panel. The specific tasks in the analysis of equivalence trials are:

a) critical examination of the data in order to identify outliers or other irregularities, to make recommendations about their handling;

b) analysis of the data grouped by laboratories, sample types and regions when possible;

c) calculation of the relative differences, means, standard deviations, expanded uncertainties;

d) decision on adequacy of the data, recommendation on the amount of additional samples when necessary, conclusions by groups or generally;

e) evaluation of the equivalence of the methods;

f) reporting the results of the analysis to the panel.

B.3 Practical aspects

The protocol of the collaborative experiment may differ in some respects from the daily practices of some laboratories. Whenever there is a conflict the written protocol shall be followed. It is essential to make sure the protocol is understood and agreed upon. This is best achieved by organizing training sessions for the responsible persons of the participating laboratories to learn and harmonize the details. It is essential to practice reading, confirming and reporting the results.