INTERNATIONAL STANDARD

ISO 17994

Second edition
2014-02-15

Water quality — Requirements for the comparison of the relative recovery of microorganisms by two quantitative methods

Qualité de l'eau — Exigences pour la comparaison du rendement relatif des microorganismes par deux méthodes quantitatives

Reference number
ISO 17994:2014(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 ..... **iv**

Introduction ..... **v**

1 Scope **1**

2 Normative references **1**

3 Terms, definitions and symbols **1**
3.1 Terms and definitions 1
3.2 Symbols and abbreviated terms 3

4 Principle **3**

5 Basic requirements for a comparison study **4**
5.1 General 4
5.2 Description of methods 4
5.3 Types of samples 4
5.4 Number of samples and participating laboratories 5
5.5 Counting and confirming 7

6 Calculations **8**
6.1 Preliminary editing of the raw data 8
6.2 Basic relative differences 8
6.3 Half-width of the confidence interval 9

7 Evaluation **9**
7.1 Preliminary evaluations 9
7.2 Two-sided evaluation 10
7.3 One-sided evaluation 12

8 Test report **13**

Annex A (informative) **Flowchart** **14**

Annex B (informative) **Comparison studies** **15**

Annex C (informative) **Derivation of equation for calculation of the number of samples** **17**

Annex D (informative) **Example of a two-sided evaluation** **19**

Bibliography **22**

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.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular the different approval criteria needed for the different types of ISO documents should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2, www.iso.org/directives.

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. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received, www.iso.org/patents.

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation on the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the WTO principles in the Technical Barriers to Trade (TBT) see the following URL: Foreword - Supplementary information

The committee responsible for this document is ISO/TC 147, *Water quality*, Subcommittee SC 4, *Microbiological methods*.

This second edition cancels and replaces the first edition (ISO 17994:2004), which has been technically revised.

Introduction

This International Standard specifies criteria and procedures for comparing the average quantitative 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).

Water quality — Requirements for the comparison of the relative recovery of microorganisms by two quantitative methods

WARNING — Persons using this document should be familiar with normal laboratory practice. This document 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.

IMPORTANT — It is absolutely essential that tests conducted in accordance with this document be carried out by suitably qualified staff.

1 Scope

This International Standard specifies an evaluation procedure for comparing two methods with established performance characteristics according to ISO/TR 13843 and intended for the 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 quantitative methods against chosen criteria for the comparison. It does not provide data for assessment of the precision of the methods being compared. It is appropriate that the precision of methods is assessed as part of their performance characterization.

This International Standard does not provide methods for the verification of method performance characterization in a single laboratory.

2 Normative references

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

ISO 8199, *Water quality — General guidance on the enumeration of micro-organisms by culture*

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

3 Terms, definitions and symbols

3.1 Terms and definitions

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

3.1.1 General terms

3.1.1.1
comparison trial
interlaboratory method comparison that involves laboratories which perform paired measurements on several of their own samples with two different methods

3.1.1.2
not-different method
method considered quantitatively not different to another method when the mean difference between their confirmed counts and stipulated difference lie between predetermined stipulated limits, taking into account all sources of variation

Note 1 to entry: This difference can be assessed by the average relative difference of their confirmed counts.

3.1.1.3
predetermined stipulated limit
permitted average difference (based on a "confidence interval" designated $-2L$ to $+2L$) between results obtained by each method, based on professional practices or regulatory requirements

Note 1 to entry: Reference [1] suggests that, in international and interlaboratory method performance tests, a limit of $2L$ = 10 % for setting the "confidence interval" be the predetermined stipulated limit for drinking water, and this has been widely used. However, for environmental waters, such as bathing waters, Reference [2] proposes a predetermined stipulated limit of $2L$ = 20 %.

3.1.1.4
reference method
method of analysis internationally recognized by experts or by agreement between the parties

Note 1 to entry: As a rule, the reference method is a standard or a commonly used method.

3.1.1.5
standard measurement uncertainty
measurement uncertainty expressed as a standard deviation

[SOURCE: ISO/IEC Guide 99:2007 [3], 2.30]

3.1.1.6
trial method
any method which is to be tested for comparison with a reference method

3.1.2 Specific terms

3.1.2.1
count
observed number of objects

EXAMPLE Colonies or cells of microorganisms, plaques of bacteriophages

Note 1 to entry: In this International Standard, the result of an MPN estimation is also considered to be a count.

3.1.2.2
presumptive count
colony count or most probable number (MPN) estimate based on the number of colonies or fermentation tubes that have an outward appearance that is interpreted as typical of a target organism

3.1.2.3
confirmed count
presumptive count multiplied by the confirmation coefficient

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

3.2 Symbols and abbreviated terms

A	(symbol for the idea of) trial method
a	test result by method A
a_i	test result (confirmed count) of method A in sample i
B	(symbol for the idea of) reference method
b	test result by method B
b_i	test result (confirmed count) of method B in sample i
i	subscript indicating a series
k	coverage factor used for calculating the confidence interval
L	smallest microbiologically significant mean relative difference between the results by methods A and B
MPN	most probable number quantitative method
n	number of samples
s	experimental standard deviation of the relative difference (standard uncertainty)
s^2	experimental variance
$s_{\bar{x}}$	standard deviation of the relative difference (standard uncertainty) of the mean
W	half-width of the confidence interval
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 lower confidence limit, derived by subtracting the value of the half-width of the confidence interval from the mean
x_U	value of the relative difference at the upper confidence limit, derived by adding the value of the half-width of the confidence interval to the mean
X^2	experimental Poisson index of dispersion
y	conditional variable used in computing the number of samples for comparison testing

4 Principle

This International Standard is based on the principle of the paired t-test (see Annex C).

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 “not different” if the mean relative difference of the paired confirmed counts does not differ significantly from zero and the

confidence interval does not extend beyond the level of the predetermined stipulated limit. The decision rules based on the above principle are given in 7.2 and 7.3 and a flow chart is given in Annex A.

5 Basic requirements for a comparison study

5.1 General

Both methods shall have data on detailed performance characteristics, derived in accordance with the guidance outlined in ISO/TR 13843.

The most important basic requirement of comparison studies is a wide range of samples. Participation by a number of laboratories is preferable, allowing the expansion of the sample range over large geographical areas. Also the credibility of a general conclusion is commonly believed to require the participation of several laboratories. However, the inclusion of a wide range of sample types by a single laboratory will also be valid. The result of the comparison is generally valid only within the range of sample types studied. Advice on the conduct of comparison studies is given 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 Description of methods

It is important to note that the principles of operation of the two methods being compared should be well understood and that the significance of any differences in the methods on the outcome of the comparative assessment should be recognized. This is particularly important if the confirmed results from each method are based on different principles. Any differences should be detailed in the test report (see Clause 8).

Performance characteristics data shall be derived in accordance with ISO/TR 13843. Such data for the methods shall be compared in order to assess potential differences in performance.

EXAMPLE Methods for the enumeration of coliform bacteria based on possession of the enzyme β-galactosidase have been shown to produce higher counts than those based on the fermentation of lactose due to the detection of a greater range of coliform bacteria.

5.3 Types of samples

The requirements for method comparisons differ somewhat from the daily routine situation. It is useful and often necessary to preselect or prepare special samples. Samples for method comparisons shall contain enough target organisms that the likelihood of scoring a zero count is small.

Samples for method comparisons shall represent types that are included in the scope of both methods. Natural samples are ideal. Samples to be tested shall represent those water source types relevant to the geographical and environmental area where the method is applied. The water types to be tested shall be included to the scope of the methods under evaluation. 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 shall be considered the last resort.

To avoid the inhibition of the target organisms by other organisms, ensure that the concentration of total bacteria in a sample is not too high. Consult ISO 8199 to ascertain the ranges of the colony counts for different cultivation methods.

It can be appropriate to influence the microbial population of existing samples to simulate situations encountered in routine laboratory practice. Such modifications could be the applications of disinfectants (e.g. chlorine, ozone or UV, Reference [1]), different ranges of temperature or the influence of daylight, in order to simulate different environmental situations from where the samples for laboratories can originate.

5.4 Number of samples and participating laboratories

5.4.1 Number of laboratories

The number of laboratories participating in comparison trials shall be sufficient to obtain a representative result for the relative recovery of the two methods being tested. Factors that shall be considered when deciding on the number of participating laboratories in a comparison trial include:

a) whether the alternative method is being assessed as a replacement for the reference method;

b) whether the comparison trials are for statutory or verification purposes;

c) the need to cover the range of geographical areas and water types for which the alternative method may be used;

d) the need to consider seasonal variability in occurrence of the target organisms;

e) the number of test results needed for the assessment of relative recovery;

f) the number of laboratories with sufficient capacity and technical expertise available to participate.

It can be acceptable to have a limited number of participating laboratories analysing a wide range of water types rather than a higher number of laboratories analysing a narrower range of water types appropriate for the methods being compared.

NOTE Several successful comparisons have been achieved with three to six laboratories. In theory, it is possible that one laboratory is able to conduct a suitable comparison study provided they have access to a wide enough range of sample types for which the methods have been characterized.

5.4.2 Number of samples

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 “predetermined stipulated limit” and the half-width of the confidence interval. If the data are found inadequate for deciding that the methods are either “different” or “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 relative difference that corresponds to the predetermined stipulated limit chosen at the beginning of the trial.

The total number of samples, n, for a two-sided evaluation (7.2) that would be sufficient for the detection of a given average relative difference at about 95 % confidence depends on the experimental variance according to Formula (1):

$$n = \frac{4s^2}{L^2} \tag{1}$$

where

n is the number of samples required for the detection of a difference L;

L is the smallest microbiologically significant mean relative difference;

s is the experimental standard deviation.

EXAMPLE A rather frequently observed value for the experimental standard deviation of the relative difference is approximately s = 80. In order to detect an average relative difference of 10 % (L = 10 %), n = 25 600/100 = 256 samples is expected to be sufficient for a two-sided evaluation.

For a one-sided evaluation (7.3) the corresponding number of samples can be calculated according to Formula (2):

$$n = \frac{3s^2}{L^2} \tag{2}$$

The rationale for the derivation of Formulae (1) and (2) is presented in Annex C.

High variability of counts can be experienced due to irregular behaviour of either laboratories or the range of sample types analysed. Whether it is warranted to continue a comparison of methods study can be ascertained by an examination of the standard deviation of the mean relative difference. Valid comparisons are indicated by standard deviations of less than 100.

NOTE With some recent methods based on chromogenic substrates, it is possible to estimate two bacterial groups simultaneously. One of the groups can be 10 or more times as numerous as the other. It is possible that the number of samples sufficient for making a final decision of equivalence with the more numerous type is not sufficient for an organism present in low numbers.

5.4.3 Number of additional samples

Formula (3) can be applied to estimate the number of additional samples when the initial number has been found to be inadequate (see 7.2.4 and 7.3.5) (see also Annex D).

$$n = 4\left(\frac{s}{y}\right)^2 \tag{3}$$

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:

$$y_1 = \bar{x}$$
$$y_2 = |\bar{x}| - |2L|$$

in which

$2L$ is the predetermined stipulated limit from 0 in the case that the methods are "not different" in %;

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

NOTE In the case of a one-sided evaluation (7.3) using Formula (3), y_2 is calculated as $y_2 = \bar{x} - 2L$, with the algebraic signs of $2L$ and $\bar{x}$ taken into account.

5.5 Counting and confirming

5.5.1 Counts

Counts shall be recorded to the last digit observed. Do not round the counts to two significant figures, as often specified for reporting routine results. MPN results shall be rounded to the nearest whole number.

5.5.2 Confirmation

Most of the methods are of a type that include confirmation of the presumptive observations. Even though often overlooked in daily routine practice, 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 for confirming counts is from 10 to 30 target colonies per plate. Dilution of the samples should be chosen accordingly.

It is recommended to use dilutions of 3^{-1} to 2^{-1} or even less should be used. This will 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.

Each presumptive positive observation should be confirmed. Partial confirmation shall be justified and a sufficient number of confirmed observations needs to be undertaken to ensure that they truly reflect confirmation rates of the methods under evaluation.

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 (e.g. by transplantation of the membrane filter) give the confirmed count directly. In such cases, the presumptive results need not be reported.

Every presumptive positive observation in MPN 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 nevertheless be reported.

6.2 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) following Formula (4):

$$x_i = \left[\ln(a_i) - \ln(b_i)\right] \times 100\,\% \tag{4}$$

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 Formulae (5) and (6).

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

$$x_i = \ln(a_i + 1) \times 100\,\% \tag{5}$$

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

$$x_i = -\ln(b_i + 1) \times 100\,\% \tag{6}$$

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 are required to contain regular count data from both methods. Samples with zero results can have a great influence on the mean relative difference if the other method gives a high count. These data can be detected by the outlier test. The inclusion of such data should be carefully considered.

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

6.2.3 Mean relative difference

The average relative performance, $\bar{x}$, is estimated in percentage units according to Formula (7):

$$\bar{x} = \frac{\sum x_i}{n} \tag{7}$$

where

n is the number of samples;

x_i is the relative difference in sample i.

6.3 Half-width of the confidence interval

The standard uncertainty (standard deviation) of x is obtained from the conventional Formula (8) for experimental standard deviation:

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}} \tag{8}$$

The standard uncertainty of the mean (formerly "standard error") is computed from Formula (9):

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} \tag{9}$$

The half-width of the confidence interval (W) is derived from the standard uncertainty of the mean by using the coverage factor $k = 2$ resulting in Formula (10):

$$W = k s_{\bar{x}} = \frac{2s}{\sqrt{n}} \tag{10}$$

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

Lower limit:

$$x_L = \bar{x} - W$$

Upper limit:

$$x_U = \bar{x} + W$$

7 Evaluation

7.1 Preliminary evaluations

7.1.1 Examination by groups

The evaluation of comparative recoveries should begin with different groupings of the data to detect possible differences between laboratories, categories of samples, seasonality or geographical location. The analysis of variance on log-transformed data or its non-parametric equivalents are likely to be suitable methods.

If great differences between laboratories are apparent, 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. Large differences attributable to the origin of samples can lead to a recommendation to exclude certain sample types from the scope of the method.

NOTE 1 The term "method" refers to confirmed results by the method.

NOTE 2 Reference methods occupy no protected position. They can be deemed unsuitable for the types of samples studied when a trial method yields significantly higher confirmed results than the reference method.

7.1.2 Outlier detection

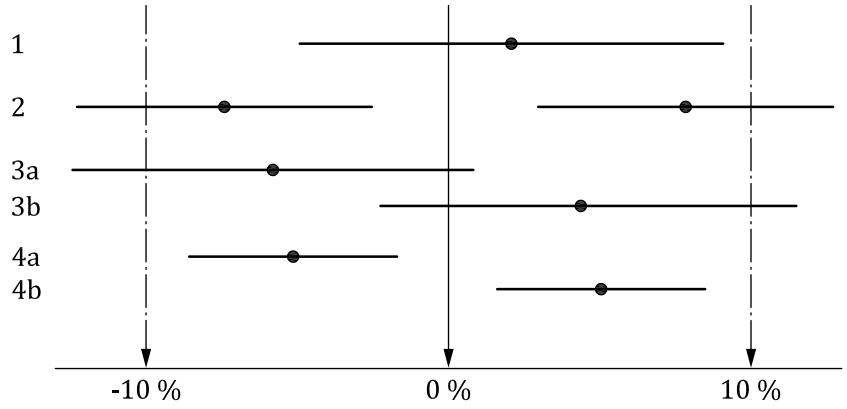
This International Standard provides no formal tests for the detection of outlier laboratories, unsuitable sample types, or individual outlying test results. It is left for the statistical analyst to decide upon the use of outlier tests. However, individual outlier values can usually be detected graphically by plotting $\ln(a_i)$ against $\ln(b_i)$ points. Omission of results may be suggested 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 outliers are excluded from data analysis, these should be listed in the test report together with the reasons for their exclusion (statistical or technical).

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

7.2 Two-sided evaluation

7.2.1 General

Whether or not one method can be named as reference, "not different" is normally understood to mean that neither method gives significantly higher or significantly lower results than the other. For special practical purposes, asymmetric predetermined stipulated limits may be agreed, i.e. different values for the lower ($-2L$) and higher ($+2L$) side. An example of a two-sided evaluation is given in Annex D. A graphical representation of potential outcomes of evaluations of data are presented in Figure 1.



Key

1	$-10\ \% \le x_L \le 0$ and $0 \le x_U \le +10\ \%$	Methods not different
2	$x_L > 0$ or $x_U < 0$	Methods different
3a	$x_L < -10\ \%$ and $x_U > 0$	Inconclusive
3b	$x_L < 0$ and $x_U > +10\ \%$	Inconclusive
4a	$x_L > -10\ \%$ and $x_U < 0$	Indifferent
4b	$x_L > 0$ and $x_U < +10\ \%$	Indifferent

Figure 1 — Graphical representation of the four potential outcomes of two-sided evaluations of comparison data where the value for 2*L* has been set at 10 %

7.2.2 Inconclusive

The data are insufficient for a decision when

$x_L < -2L$ and $x_U > 0$ or

$x_L < 0$ and $x_U > +2L$

More samples should be examined. The number of additional samples required can be estimated as shown in 5.4.3 (see also Annex D).

7.2.3 Methods “not different”

The methods are “not different” when

$-2L \le x_L \le 0$ and $0 \le x_U \le +2L$

7.2.4 Indifferent

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

$x_L > -2L$ and $x_U < 0$ or

$x_L > 0$ and $x_U < +2L$

It is an arbitrary choice which argument is considered decisive (statistical or practical). Although from a statistical point of view the methods can be considered different it can be acceptable from a practical point of view that the methods are considered to be not different.

7.2.5 Methods “different”

The methods are “different” when

$x_L > 0$ or $x_U < 0$

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 not different or higher than the reference method, see Reference [1]. In such cases, only the lower value of the predetermined stipulated limit ($-2L$) is of concern in the evaluation. The result categories differ somewhat from those in the two-sided evaluation.

7.3.2 Inconclusive

The data are insufficient for a decision when

$x_L < -2L$ and $x_U > 0$

More samples should be examined. The number required can be estimated as shown in 5.4.3 (see also the Note to 5.4.3 and Annex D).

7.3.3 Methods “not different”

Methods are “not different” when

$-2L \leq x_L \leq 0$ and $x_U > 0$

7.3.4 Indifferent

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

$x_L > -2L$ and $x_U < 0$

7.3.5 Trial method: higher recovery

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

$x_L > 0$

7.3.6 Trial method: lower recovery

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

$x_U < 0$

8 Test report

This test report shall contain at least the following information:

a) the test method used, together with a reference to this International Standard (ISO 17994:2014);

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

c) relevant descriptive details of the experiment (numbers of samples, participants, the stipulated acceptable difference);

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 (these data should be available in a form to allow analysis by other parties).

Annex A
(informative)

Flowchart

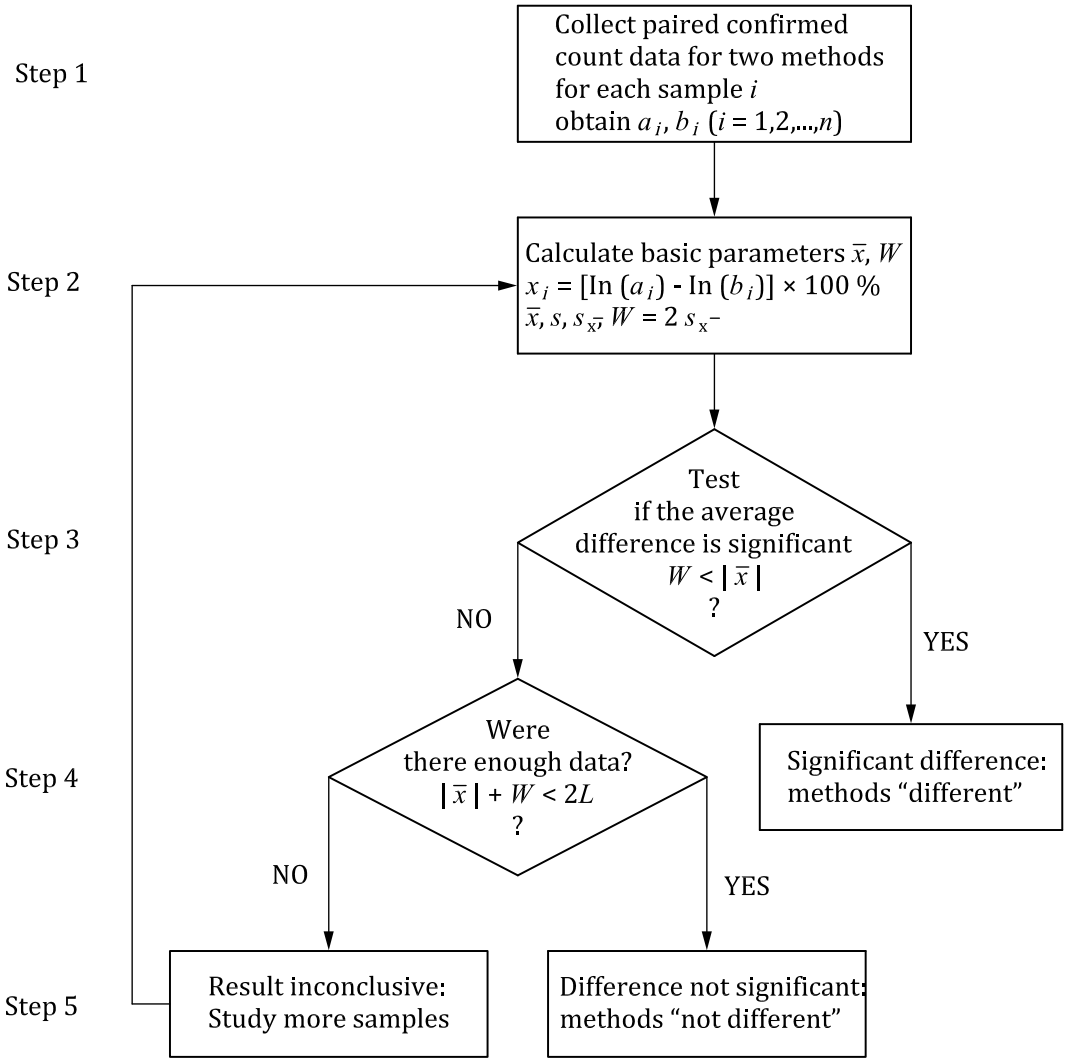


Figure A.1

Annex B
(informative)

Comparison studies

B.1 General

Collaborative comparison of methods studies 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 coordinate 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 predetermined stipulated limit 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 co-ordinating functions to enlist the laboratories and to draft a protocol covering the laboratory procedure and data reporting. The co-ordinator 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 collaborate in the writing of a report 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.

The detailed protocol provided by the co-ordinator of a collaborative trial shall be strictly followed. If a laboratory disagrees with the protocol, it should withdraw from the collaborative trial rather than stay with it and keep to its own ways.

Each laboratory should analyse its own samples reflecting the range of water types or quality usually analysed by that laboratory.

B.4 Data report sheet

The co-ordinator shall provide a data report form with space for the following information:

a) manufacturer, product name and lot number of the media and materials (e.g. membrane filters) used;

b) sample and site identification;

c) sample category (river, lake, sewage, sea, private well, etc.);

d) date of sampling, storage between sampling and analysis;

e) presumptive count by method A;

f) presumptive count by method B;

g) confirmed count by method A;

h) confirmed count by method B;

i) number isolated for confirmation (if different from the presumptive count);

j) number of isolates confirmed (when partial confirmation is practiced).

Counts larger than 100 shall not be rounded to two figures. MPN counts shall be rounded to the nearest whole number.

The co-ordinator should ask the participants to report all relevant observations they have made on the methods in addition to the information that are required in the data report form.

B.5 Study report

Key aspects that should be addressed in the study report are as follows:

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, half-widths of confidence intervals;

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 difference between the methods;

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

Annex C
(informative)

Derivation of equation for calculation of the number of samples

C.1 Basic principle

This International Standard is based on the principle of the paired *t*-test. The test results are converted to logarithmic values before calculating their difference. The use of logarithms largely eliminates the effect of concentration and permits the use of samples from different populations with different mean concentrations in the same analysis. Applying natural logarithms justifies calling the test variable, $x = \ln a - \ln b$ the "relative difference". The relative difference is the best measure of the quality that is of interest, the relative performance of two methods.

Formally, the paired *t*-test is expressed by Formula (C.1):

$$t_\alpha = \frac{\bar{x}}{s_{\bar{x}}} \tag{C.1}$$

t_α is the value of Fisher's *t*-distribution for risk level α that corresponds to the risk of being wrong when saying that the mean does not differ from zero. If the value of the division exceeds the theoretical value t_α, the difference is considered significant. In this International Standard a confidence level of 95 % was chosen which corresponds to α = 5 %.

The test is dressed into a graphical examination of the points of the half-width *W* of the approximate confidence interval around the observed mean and a stipulated smallest significant difference *L*. The word "approximate" is important, because there is no certainty that the distribution of relative differences follows the normal distribution very well. Another reason is because the value for t_α is approximate.

The half-width of the confidence interval *W* is given by Formula (C.2):

$$W = t_\alpha s_{\bar{x}} = t_\alpha \frac{s}{\sqrt{n}} \tag{C.2}$$

The value of *W* depends on the confidence level chosen. The value of t_α depends also on the number of samples *n*, which is not known exactly beforehand. It is known, however, that the number of samples will be at least 30 or more. In that vicinity, the value of *t* (for a 5 % risk level) is approximately 2 for a two-sided test and 1,7 for a one-sided test.

C.2 Calculation of the number of samples

The smaller the differences that are to be detected are, the more samples need to be studied. The extreme testing situation occurs when the experimental mean relative difference equals the smallest significant difference *L*. To detect this situation with adequate confidence, the study of sufficient samples is required to make the half-width of the confidence interval equal to *L*.

By equating the smallest significant difference L with the value of the half-width W results in Formula (C.3):

$$L = W = t_{\alpha} \frac{s}{\sqrt{n}} \tag{C.3}$$

Solving for the number of samples n, Formula (C.4) is derived:

$$n = t_{\alpha}^{2} \left(\frac{s}{L} \right)^{2} \tag{C.4}$$

Even though the value of t_{α} depends on the unknown n, it is anticipated that the value (of n) would be rather large in every case. Therefore, the values of t of approximately 2 and approximately 1,7 as representative approximate values for the two-sided and one-sided testing situations are chosen. With $2^2 = 4$, the formula for the required number of samples for the two-sided case, thus, becomes Formula (C.5):

$$n = \frac{4s^2}{L^2} \tag{C.5}$$

For the one-sided testing situation ($t_{\alpha}^{2} \approx 1{,}7^2 \approx 3$) the formula is Formula (C.6):

$$n = \frac{3s^2}{L^2} \tag{C.6}$$

where

n is the number of samples required for the detection of a difference L;

L is the smallest significant mean difference;

s is the experimental standard deviation.

Annex D
(informative)

Example of a two-sided evaluation

Table D.1 — Results of a comparative study

Sample No.	Method A presumptive	Method B presumptive	Method A confirmed	Method B confirmed	Relative difference %
i			a_i	b_i	x_i
1	1	2	1	0	69,32[a]
2	0	0	—	—	—[b]
3	82	> 120	—	—	—[b]
4	0	3	0	1	−69,32[a]
5	0	5	0	2	−109,86[a]
6	1	4	1	1	0,00
7	4	2	3	1	109,86
8	2	3	1	2	−69,32
9	3	4	3	2	40,55
10	4	4	4	2	69,32
11	5	3	4	2	69,32
12	5	3	5	2	91,63
13	5	2	5	2	91,63
14	11	2	10	2	160,94
15	1	3	1	3	−109,86
16	8	3	8	3	98,08
17	11	6	10	3	120,40
18	12	5	11	4	101,16
19	3	6	1	5	−160,94
20	5	5	4	5	−22,31
21	10	10	8	5	47,00
22	3	8	3	6	−69,32
23	5	7	5	7	−33,65
24	6	7	5	7	−33,65
25	6	9	6	7	−15,42
26	10	8	8	7	13,35
27	5	8	5	8	−47,00
28	20	12	18	8	81,09
29	7	10	7	9	−25,13
30	10	10	10	10	0,00
31	16	15	14	11	24,12
32	8	15	8	13	−48,55
33	11	17	11	14	−24,12
Mean					**11,27**

Table D.1 *(continued)*

Sample No.	Method A presumptive	Method B presumptive	Method A confirmed	Method B confirmed	Relative difference %
Std. dev.					**78,32**
[a]	Calculated according to 6.2.2				
[b]	Deleted (6.1)				

In this example, 33 samples were analysed with the generation of confirmed counts by two methods. Method B is the reference method. All presumptive positives were tested for confirmation.

Two samples (numbers 2 and 3) were deleted because the counts were unfit for calculating the relative difference.

The half-width of the confidence interval (W) was calculated from the mean, standard deviation and the number of relevant pairs of confirmed counts:

$$W = \frac{2 \times 78,32}{\sqrt{31}} = 28,13 \qquad \text{(D.1)}$$

The "confidence limits" according to Clause 7 are:

Lower limit: $x_L = 11,27 - 28,13 = -16,86$

Upper limit: $x_U = 11,27 + 28,13 = 39,40$

Assuming that the predetermined stipulated limit has been chosen as $2L$ = 10 %, the evaluation in accordance with 7.2.2 is: *inconclusive*.

The number of samples was insufficient. More samples should be examined. The counts in general were quite low in the 33 samples. It is advisable to try to find samples with more target organisms in further comparisons.

To estimate how many samples might have been enough for a firm decision, the procedure described in 5.4.3 was applied using the information available:

$\bar{x}$ = 11,27 %;

$2L$ = 10 %;

s = 78,32 %

The two estimates of y are:

$y_1 = 11,27$

$y_2 = |11,27| - |10| = 11,27 - 10 = 1,27$

y_1, being the larger of the two, is inserted in Formula (D.2) for the number of samples:

$$n = 4\left(\frac{s}{y}\right)^2 = 4\left(\frac{78,32}{11,27}\right)^2 = 193 \qquad \text{(D.2)}$$

About 193 samples would have been a sufficient number to reach a firm decision, provided that the mean and standard deviation do not change markedly when the results of the new samples are included. Thus, 193 – 31 = 162, i.e. about 160 additional samples should suffice.

If the test had been one-sided, the basic evaluation would be the same: *inconclusive*.

The number of additional samples would be different (see 5.4.3).

$y_2 = 11{,}27 - (-10) = 21{,}27$

$y_1 = 11{,}27$

y_2 being the larger of the two,

$$n = 4\left(\frac{78{,}32}{21{,}27}\right)^2 = 54 \tag{D.3}$$

The number of additional samples to be analysed is 54 – 31 = 23.

Bibliography

[1] Drinking Water Inspectorate. *Comparison of methods for drinking water bacteriology — Cultural techniques*. London: Department of the Environment, Transport and the Regions, 2000

[2] Laboratory of Environmental Microbiology. *Escherichia coli methods for bathing waters — Results of a Finnish equivalence trial*. Helsinki: National Public Health Institute, 2008

[3] ISO/IEC Guide 99:2007, *International vocabulary of metrology — Basic and general concepts and associated terms (VIM)*

ICS 07.100.20;13.060.70

Price based on 22 pages