INTERNATIONAL STANDARD

ISO 13843

First edition
2017-06

Water quality — Requirements for establishing performance characteristics of quantitative microbiological methods

Qualité de l'eau — Exigences pour l'établissement des caractéristiqes de performance des méthodes microbiologiques quantitatives

Reference number
ISO 13843:2017(E)

ISO copyright office
Ch. de Blandonnet 8 • CP 401
CH-1214 Vernier, Geneva, Switzerland
Tel. +41 22 749 01 11
Fax +41 22 749 09 47
copyright@iso.org
www.iso.org

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.

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 (see 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 (see 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 voluntary nature of standards, the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT) see the following URL: www.iso.org/iso/foreword.html.

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

This first edition of ISO 13843 cancels and replaces ISO/TR 13843:2000, which has been technically revised.

Introduction

Methods are considered microbiological when the quantitative estimate is based on counting of microbial particles either directly with the aid of a microscope or indirectly on the basis of growth (multiplication) into colonies, turbidity, a colour change or fluorescence. The principles and procedures within the scope of this document are commonly known as microscopic count, most probable number (MPN) and colony count. Most of the procedures for the determination of performance characteristics described in this document are applicable to all three types of method. However, where the procedures are not applicable, alternative suggestions are made within the body of the document or in Annexes D and E (for repeatability, reproducibility and uncertainty of counting).

Plaque counts of bacteriophages are in most respects similar to bacterial colony counts.

Some of the "newer" microbiological methods such as those utilizing fluorescent *in situ* hybridization (FISH) or polymerase chain reaction (PCR) can also be covered by this document. However, they may require special consideration, depending upon how they are used. The issues of importance in these situations include the mechanism of determining the numbers of microbes present (e.g. standard curve for qPCR or microscopic count for FISH) and the viability of the organisms detected. If such techniques are used for confirmation as part of a method then all sections of this document are relevant.

While not essential, during the characterization of microbiological methods it may be beneficial to generate data using stressed organisms. Various methods can be used to stress organisms, but the two that are most useful for water are disinfectant stress (usually chlorine injury) and nutrient depletion caused by organisms being in a low nutrient environment (i.e. drinking water and other oligotrophic waters) for a period prior to testing. The effect on some of the performance characteristics of "stressing" organisms is almost totally dependent on the type and degree of stress applied and it is inappropriate to include such detail in this document. However, there are descriptions in the literature that laboratories can follow in case they should wish to determine performance characteristics of a method with stressed cells.

Water quality — Requirements for establishing performance characteristics of quantitative microbiological methods

1 Scope

This document deals with characterization of microbiological methods. In terms of this document, characterization means the study of parameters that can be measured to describe how the method is likely to perform in a given set of conditions, which can be described as performance characteristics. The document describes procedures for the determination of performance characteristics which can be used for subsequent validation or verification of methods.

The emphasis is on selective quantitative methods and this document applies to all types of water. For methods that are not based upon direct microscopic count, colony count or most probable number, the applicability of the procedures described in this document should be considered carefully.

2 Normative references

The following documents are referred to in the text in such a way that some or all of their content constitutes requirements 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 17994:2014, *Water quality — Requirements for the comparison of the relative recovery of microorganisms by two quantitative methods*

3 Terms and definitions

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

ISO and IEC maintain terminological databases for use in standardization at the following addresses:

— IEC Electropedia: available at http://www.electropedia.org/

— ISO Online browsing platform: available at http://www.iso.org./obp

3.1
accuracy
measurement accuracy
closeness of agreement between a measured quantity value and an assigned quantity value of a measurand

Note 1 to entry: The concept 'measurement accuracy' is not a quantity and is not given a numerical quantity value. A measurement is said to be more accurate when it offers a smaller measurement error.

Note 2 to entry: 'Measurement accuracy' is sometimes understood as closeness of agreement between measured quantity values that are being attributed to the measurand.

[SOURCE: ISO/IEC Guide 99:2007, 2.13[16], modified — "...a true quantity value" replaced by "... an assigned quantity value; Notes 1 and 2 to entry added]

3.2
analyte
component represented in the name of a measurable quantity

Note 1 to entry: In water microbiology, the analyte is ideally defined as a list of taxonomically defined species. In most cases, in practice the analyte can only be defined by group designations less accurate than taxonomic definitions.

[SOURCE: ISO 17511:2003, 3.2[14]]

3.3
analytical portion
test portion
volume of particle suspension (sample) inoculated into a detector unit (agar plate, membrane filter, test tube, microscopic grid square)

3.4
bias
measurement bias
estimate of a systematic measurement error, or the systematic difference between the quantitative assigned value and the average of measurement replicate results

3.5
categorical characteristics
method performance characteristic numerically expressed as a relative frequency based on P/A or +/- classification

3.6
colony-forming unit
CFU
colony-forming particle
CFP
organism (or cluster of organisms) with the ability to form a colony under certain specified conditions

Note 1 to entry: The term was originally introduced to convey the idea that a colony may originate not only from a single cell but from a solid chain or aggregate of cells, a cluster of spores, a piece of mycelium, etc. It mistakenly equates the number of colonies observed to the number of living entities seeded on the medium. Growth unit, viable particle, propagule and germ are terms with the same meaning but convey the original idea better and apply not only to colony count methods but also to the most probable number (MPN).

3.7
collaborative method performance
method or laboratory performance test where several laboratories join in an experiment planned and co-ordinated by a leader laboratory

Note 1 to entry: Collaborative tests are mainly of two types. Intercalibration exercises are made to allow laboratories to compare their analytical results with those of other participating laboratories.

Note 2 to entry: Method performance tests produce precision estimates (repeatability, reproducibility) out of data accumulated when several participating laboratories study identical samples with a strictly standardized method.

3.8
confirmed colony count
verified colony count
presumptive colony count corrected for false positives

Note 1 to entry: Mathematically:

$$pc = \frac{k}{n} c$$

where

c is the presumptive count;

p is the true positive rate;

n is the number of presumptive positives isolated for confirmation;

k is the number confirmed.

3.9
corroborated count
count obtained when using a secondary confirmation procedure

3.10
detection level
minimum concentration of organisms that produce evidence of growth with a probability of $P = 0{,}95$ when inoculated into a specified culture medium and incubated under defined conditions

Note 1 to entry: The theoretical level that conforms to this definition is an average of three viable cells in an inoculum volume.

3.11
detection set
combination of plates or tubes on which quantitative estimation of sample microbial concentration is based

Note 1 to entry: The detection set is the set of plates or tubes utilized for numerical estimation of a single value.

EXAMPLE Parallel plates of a suspension, plates from consecutive dilutions, 3 × 5 tube MPN system, microtitre plate.

3.12
detector
particle detector
plate of solid matrix or a tube of liquid containing a nutrient medium for counting or detecting biologically active particles

3.13
efficiency
E
fraction of colonies that are correctly assigned as positives and negatives

Note 1 to entry: Mathematically:

$$E = \frac{a+d}{n}$$

where

a is the number of typical colonies confirmed as being the target organism (true positives);

d is the number of atypical colonies confirmed as not being the target organism (true negatives);

n is the total number of colonies tested for confirmation.

3.14
false negative
result indicated by the test method to be negative which has subsequently been shown to contain the target organism

3.15
false positive
result indicated by the test method to be positive which was subsequently shown not to contain the target organism

3.16
germ
entity capable of biological activity (e.g. respiration or reproduction in a nutrient medium)

3.17
limit of determination
lowest analyte concentration per analytical portion where the expected relative standard uncertainty, equals a specified value

3.18
method-defined count
count obtained by using only the procedures in the described method

3.19
negative binomial distribution
particular "overdispersed" statistical distribution of counts

Note 1 to entry: Its variance can be expressed as $s^2 = \bar{x} + u_0^2\,\bar{x}^2$ ($\bar{x}$ = mean).

Note 2 to entry: In this document, the square of the relative operational standard deviation (u_0) is substituted for the inverse of the exponent ($1/k$) of the standard equation for the negative binomial distribution.

3.20
outlier
member of a set of values which is inconsistent with other members of that set

Note 1 to entry: An extreme value which normally appears randomly in less than 1 % of repetitive tests, but more frequently if abnormal situations occur. Statistical test procedures can be used to quantify this probability.

3.21
over-dispersion
variation in excess of Poisson randomness

Note 1 to entry: Detected qualitatively by the Poisson index of dispersion and measured quantitatively by estimating the parameter u_0 (relative operational standard deviation) of the negative binomial distribution.

3.22
parallel counts
particle or colony numbers in equal analytical portions drawn from the same suspension

3.23
Poisson distribution
fully random distribution of particle numbers when sampling a perfectly mixed suspension

Note 1 to entry: The probability $P(k)$ of observing exactly k units in a test portion when the mean equals μ is calculated from

$$P(k) = \frac{u^k}{k!} e^{-\mu}$$

3.24
precision
measurement precision
closeness of agreement between indications or measured quantity values obtained by replicate measurements on the same or similar objects under specified conditions

Note 1 to entry: Measurement precision is usually expressed numerically by measures of imprecision, such as standard deviation, variance, or coefficient of variation under the specified conditions of measurement.

Note 2 to entry: The 'specified conditions' can be, for example, repeatability conditions of measurement, intermediate precision conditions of measurement, or reproducibility conditions of measurement (see ISO 5725-3[4]).

Note 3 to entry: Measurement precision is used to define measurement repeatability, intermediate measurement precision, and measurement reproducibility.

3.25
proportionality
agreement of observed particle counts with the volume (or dilution) of a series of analytical portions from a common root suspension

Note 1 to entry: Proportionality is evaluated as the log-likelihood ratio statistic G^2 with n-1 degrees of freedom.

3.26
recovery
general term used for the number of particles estimated in a test portion or sample, with the understanding that there is a true (although unknown) number of particles of which 100 % or less are "recovered" by the employed methodology

Note 1 to entry: Another similar term commonly used is productivity (see ISO 11133[12]).

3.27
relative recovery
ratio of colony counts obtained by two methods tested on equal test portions of the same suspension

3.28
relative operational standard deviation
u_0
operational variability, expressed as a relative standard uncertainty, associated with the technical steps of the analytical procedure

Note 1 to entry: The relative operational standard deviation is often expressed in percent.

3.29
relative operational variance
u_0^2
over-dispersion constant, the square of relative operational standard deviation

3.30
relative standard deviation
u_{rel}
estimate of the standard deviation of a population from a sample of n results divided by the mean of that sample

3.31
relative variance
u_{rel}^2
square of relative standard deviation

3.32
repeatability
measurement repeatability
measurement precision under a set of repeatability conditions of measurement

3.33
repeatability conditions
condition of measurement, out of a set of conditions that includes the same measurement procedure, same operators, same measuring system, same operating conditions and same location, and replicate measurements on the same or similar objects over a short period of time

3.34
reproducibility
measurement reproducibility
measurement precision under reproducibility conditions of measurement

Note 1 to entry: Relevant statistical terms are given in ISO 5725-1[2] and ISO 5725-2[3].

3.35
reproducibility conditions
condition of measurement, out of a set of conditions that includes different locations, operators, measuring systems, and replicate measurements on the same or similar objects

3.36
robustness
insensitivity of an analytical method to small changes in procedure

Note 1 to entry: To examine the robustness it is advisable to "abuse" the method in a controlled way.

3.37
sensitivity
fraction of the total number of positive cultures or colonies correctly assigned in the presumptive inspection

3.38
specificity
fraction of the total number of negative cultures or colonies correctly assigned in the presumptive inspection

3.39
standard uncertainty
uncertainty of the result of a measurement expressed as a standard deviation

3.40
uncertainty of counting
relative standard deviation of results of repeated counting of the colonies or particles of the same plate(s) or field(s) under stipulated conditions (same person, different persons in one laboratory)

3.41
verification
performance of a second characterization by a different laboratory to confirm the results of the original characterization

4 Basic concepts

4.1 General

As far as particle statistics are concerned, microscopic counts obey the same laws as viable counts but they are, with the exception of microcolony methods, free from the biological problems associated with growth. Differential stains, specifically labelled complexes or other agents used for finding the target

do not change the basic principles. The same principles as those used with selective colony methods can be applied. For a more detailed understanding of the theory and application of the formulae used in this document, the mathematical basis for the variation encountered in all of these types of method is described in Annex A.

4.2 Characterization

The characterization of a microbiological method is largely based upon the examination and expression of the performance characteristics of that method.

Characterization is a process of providing information about the likely performance of that procedure under a specific set of circumstances. It is not the intention of this document to provide guideline values for each of the specified performance characteristics but rather to give guidance on which parameters should be determined and how best to derive them for comparative purposes. Methods that have "poor" performance characteristics may still be useful.

Characterization is an exploratory process with the aim of establishing the likely set of performance characteristics of a new, modified or otherwise inadequately characterized method. It should result in numerical and descriptive specifications for the performance and include a detailed and unambiguous description of the target of interest (such as positive colony, tube or plaque). However the values generated should not be used as limits since they may change depending on the laboratory, matrix or even specific samples.

Characterization is performed by a single laboratory in the first instance to determine the likely performance of a test method in a specific laboratory.

A collaborative method performance study can be performed as an additional step to evaluate the interlaboratory performance characteristics.

NOTE A laboratory developing an in-house method or a variant of an existing standard could carry out the steps of characterization.

It is imperative that technicians involved in the characterization of a method have considerable experience with other microbiological methods.

The performance characteristics covered by this document are listed in Table 1.

Table 1 — Performance characteristics described in the document

Parameter	Definition
Sensitivity[a, b, c]	fraction of the total positives[e] correctly assigned in the presumptive count
Specificity[a, b, c]	fraction of the total negatives[f] correctly assigned in the presumptive count
False positive rate[a, b]	fraction of positive results (e.g. typical colonies) that are subsequently shown to be due to non-target organisms
False negative rate[a, b]	fraction of negative results (e.g. atypical colonies) shown to be target organisms
Selectivity[a, b, c]	ratio of the number of target colonies to the total number of colonies in the sample volume
Efficiency[a, b]	fraction of total colonies correctly assigned in the presumptive count

[a] Required for determination of the performance characteristics.

[b] Required for single laboratory verification.

[c] Guidance specification given.

[d] Methods for interlaboratory reproducibility and precision are described in Annex F. Use of these methods should be considered when interlaboratory performance is paramount, for example when methods are being developed for regulatory compliance.

[e] Positives may be colony counts, positive reaction vessels (MPN) or cell counts.

[f] Negatives may be atypical colonies, negative reaction vessels (MPN) or cells without the specific characteristics required.

Table 1 *(continued)*

Parameter	Definition
Upper limit[a]	upper end of the working range for which the method is useful (i.e. the maximum countable colonies per plate, or other detection systems)
Repeatability[a, b, c]	precision under repeatability conditions (same operators, same operating conditions, short period of time, ...)
Reproducibility[a]	precision under intralaboratory reproducibility conditions[d]
Robustness[a]	measure of the capacity of a test to remain unaffected by small but deliberate variations in testing conditions (e.g. temperature)
Relative recovery[a]	efficiency with which a method recovers target organisms from a sample when compared to another procedure (This comparison shall be done where an alternative method for the same organism exists. Comparison with an ISO reference method is preferred.)
Uncertainty of counting[a, b]	relative standard deviation of replicate counts of the target obtained by repeated counting (plates, fields, tubes, etc.) under stipulated conditions (same person, different person, same laboratory, etc.)

[a] Required for determination of the performance characteristics.

[b] Required for single laboratory verification.

[c] Guidance specification given.

[d] Methods for interlaboratory reproducibility and precision are described in Annex F. Use of these methods should be considered when interlaboratory performance is paramount, for example when methods are being developed for regulatory compliance.

[e] Positives may be colony counts, positive reaction vessels (MPN) or cell counts.

[f] Negatives may be atypical colonies, negative reaction vessels (MPN) or cells without the specific characteristics required.

While interlaboratory reproducibility and precision do not form part of the performance characteristics described in the body of this document, in certain situations knowledge of these parameters is highly desirable. Such situations include when methods are being used for regulatory compliance or when data from a variety of laboratories are being compared for any of a number of reasons. For this reason, suggested methods to determine interlaboratory reproducibility are described in Annex F.

4.3 Verification

Verification takes place when a laboratory proceeds to implement a method developed elsewhere. Verification focuses on gathering evidence that the laboratory is able to generate performance data similar to those established in primary characterization. It is not helpful to establish limits on the various components of method characterization since these can vary dependent on many aspects of the method, type of sample and performing laboratory. The verification data should be used to establish the type and quality of data likely to be generated by the laboratory with a given procedure and any given sample type.

Typically, verification uses selected and simplified forms of the same procedures used in method characterization, but possibly extended over a longer time. Natural samples are the optimal test materials and the work need only address those aspects of the method performance that are of interest to the laboratory. The requirements for single laboratory verification are described in Clause 7.

4.4 Method comparison

Method performance consists of many aspects. There is neither a single test of method comparison nor numerical criteria for it. One method may be superior in specificity but inferior in recovery. All the collective information about robustness, precision and specificity gained during characterization tests can be used for method comparison. The methods only need to be tested in parallel for recovery comparisons.

It is necessary to apply two methods in parallel on the same samples when developing an in-house method, and also when collecting information to justify the use of an alternative method. Relative recovery studies of an alternative method against a reference method organized according to ISO 17994 involve preferably a wide range of samples and participation by a number of laboratories allowing the expansion of the sample range over large geographical areas. However, sometimes it may be necessary to verify the result of an alternative method recovery study under ecological conditions or in a geographical area not represented in the earlier collaborative trial. When a laboratory only needs to confirm the comparison result of 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. Formula (3) given in ISO 17994:2014, 5.4.3 can be applied to estimate the recommended number of samples. However, whatever the result of the calculation, the number of samples should not be less than thirty.

A method giving the highest recovery of confirmed target organisms is obviously the best when confirmation is required for routine use. A method giving somewhat lower recovery but not requiring confirmation may be preferable. If high false negative rates or false positive rates observed in characterization cannot be corrected by more refined target colony definitions or other procedures, the method may be deemed invalid. The comparison of two microbiological methods should include a comparison of their performance characteristics (i.e. characterization) together with a side by side comparison of recovery, using naturally contaminated or spiked samples as specified in ISO 17994.

4.5 Samples

It is a popular concept that the characterization and comparison of methods should be performed with natural samples with natural concentrations of microbes. While conceptually this is a good idea, there are exceptions under some circumstances.

Artificial samples (reference materials and spiked samples) are used in internal and external quality assurance systems to ensure the basic proficiency of the laboratories participating in method characterization exercises.

Spiking may be useful and even necessary in verification or whenever it is difficult to find natural samples with target organisms. The optimal concentration range for the characterization of microbiological methods is narrower than the projected working range. High concentrations are unnecessary. Such samples resemble pure cultures and do not put the performance of the method or of the laboratory to test.

Samples with very low bacterial content need to be studied for public health reasons but are not ideally suited for method comparisons and other characterization exercises for statistical reasons. However, their use is unavoidable in many situations. In particular, where a method seeks to identify two types of target organisms on the same plate (e.g. total coliforms and *E. coli*), low numbers of organisms are usually unavoidable.

The number and variety of samples examined need to be appropriate. Without the help of statistics, there are no objective ways of making a decision. In some instances, the first sample studied might give the answer that the method is not good enough. Usually, however, more samples are needed. Choosing too few samples may not yield representative results.

Specific guidelines on the numbers and types of samples (together with their microbial content) are given in 6.1.

5 Specifications: some guideline values

Historically, standards have provided little help for laboratories seeking to make sure that they apply the methods well and obtain valid results. What seems to be lacking is a concise presentation of what laboratories should do to verify that the method also works in their hands properly and how to distinguish between good and bad performance.

A clause of performance characteristics shall be added to all ISO standards that deal with water microbiology methods and refer to defined microbes or groups of microbes.

The format for colony count methods might include the following:

a) **Sensitivity:** Generally greater than 90 %;

b) **Specificity:** Generally greater than 80 %;

c) **Selectivity:** Results are generally not valid if selectivity is less than 10 %.

d) **Uncertainty of counting:** Individual counting uncertainty (one person) remains normally below u_{rel} = ± 0,03. Intralaboratory uncertainty of counting is below u_{rel} = ± 0,05. Intralaboratory uncertainty of counting greater than 0,1 is a certain sign of problems or difficulties.

e) **Repeatability (parallel plating):** Variation is within the Poisson distribution. If not, the extent of over-dispersion should be given.

f) **Upper limit:** For membrane filtration methods, a range that has been quoted is 0 CFU to 80 CFU while for plate count methods using a 90 mm Petri dish, the range may be 0 CFU to 300 CFU. The surface area of a 47 mm membrane filter is approximately 25 % of that of a 90 mm Petri dish. These upper limits are dependent upon the extent of background (non-target) growth, the number of different types of target organisms (e.g. total coliforms and *E. coli*) and colony size.

6 Designs for determining performance characteristics of a method

6.1 General considerations

While it is generally better to use samples that are naturally contaminated with the target organisms, in some situations this does not facilitate the determination of the performance characteristic of interest. In such situations, the use of reference materials may be appropriate. Alternatively, samples can be "spiked" with known numbers of target organisms obtained from commercial sources.

Whichever procedure is used to prepare samples, attention should be paid to adequate mixing to facilitate a random distribution of organisms. The desired number of target organisms in the sample is dependent upon the method of interest but typically a range of 10 to 60 colony forming units per test portion will be practical. The number is a compromise between a desirable higher number and the number of colonies that is practical for a membrane or plate.

6.2 Determination of sensitivity, specificity, efficiency, selectivity, false positive rate and false negative rate

6.2.1 Type of samples to be used

Appropriate samples are prepared containing 20 to 80 total colony forming units (10 to 60 CFU target organisms) per test portion. The samples are then examined by the procedure being studied. Typical and atypical colonies (i.e. those having the typical appearance of the target organism **and** those not having the typical appearance of the target organism) are counted. All typical and atypical colonies are then identified using an appropriate procedure which could include commercially available microbial identification kits, DNA sequencing or other specified procedures.

The method of preparation of samples will vary depending on the method and the types of samples typically analysed. If naturally contaminated samples are available with an appropriate level of target organisms, then these should be used. However, in many cases naturally contaminated samples are not available (e.g. for methods designed for drinking water). In such cases, laboratory-prepared spike material (using appropriate sources of target organisms such as river water or sewage) can be used to prepare samples. This has the benefit of including non-target organisms that are likely to be encountered in contaminated drinking water in a similar ratio to what may be seen in the "real world".

The use of reference materials should be avoided because the choice of strains used directly influences the outcome of the experiment.

6.2.2 Number of samples

A minimum of twenty samples should be used from different sources. If using surface water or sewage to prepare the spike material, this should be obtained from at least three sources. It is important when performing this work to examine an adequate number of "atypical" colonies, although when the method is very selective these may be hard to find.

6.2.3 Procedure

Samples are incubated and all colonies "confirmed" according to the procedures appropriate to the method (e.g. for ISO 9308-1[10] all presumptive coliforms must be tested for the production of cytochrome oxidase). Where methods have no confirmatory procedures (e.g. ISO 9308-2[11]) described, results are recorded as described in the method with no confirmation. Colonies (or reaction vessels: wells, tubes, etc.) are recorded as positive or negative.

When determining the parameters sensitivity, specificity, false positive rate and false negative rate, it is necessary to apply a further confirmatory test to confirm or deny the results generated by the test method. This further confirmatory test is called secondary confirmation. Such methods might include commercially available identification kits, other phenotypic methods (e.g. tests for a certain trait or enzyme system), tests of chemical composition (e.g. MALDI-TOF) or molecular methods. The choice of method will influence the outcome of the data and care should be taken when selecting these procedures as they define the target. Wherever possible, the procedure used to give support to the original confirmatory test should be based on a taxonomically sound procedure or on tests that reflect the definition of the target organisms as described in the relevant standard. For example, in the characterization of a test for *Escherichia coli* isolates could be subjected to 16S ribosomal RNA sequencing to determine if they are truly *E. coli*. Alternatively, isolates could be examined by Gram staining and tested for the expression of functional cytochrome oxidase, β-D-galactosidase and β-D-glucuronidase. Organisms that are Gram negative rods, cytochrome oxidase negative and β-D-galactosidase and β-D-glucuronidase positive would be considered to be *E coli*.

6.2.4 Categorical performance characteristics

6.2.4.1 When a confirmation step is included in the method, the identification data can be divided into four categories:

a) number of typical colonies confirmed as being the target organism in the primary confirmatory test the identity of which is supported by the secondary identification test (true positives);

b) number of atypical colonies, or typical colonies that are negative in the primary confirmatory test identified as being the target organism by the secondary identification test (false negatives);

c) number of typical colonies confirmed as being the target organism by the primary confirmatory test which are subsequently shown to not be the target organism by the secondary identification test (false positives);

d) number of atypical colonies or typical colonies that are negative in the primary confirmation test which are shown by the secondary identification test to not be a target organism (true negatives).

6.2.4.2 For methods without a confirmatory procedure, the identification data can be divided into four categories:

a) number of typical colonies identified as being the target organism by an external identification test (true positives);

b) number of atypical colonies identified as being the target organism by an external identification test (false negatives);

c) number of typical colonies identified as not being the target organism by an external identification test (false positives);

d) number of atypical colonies identified as not being the target organism by an external identification test (true negatives).

6.2.4.3 The frequencies of these categories can be conveniently expressed in a 2 x 2 diagram:

		Presumptive count		
		+	–	
Confirmed count	+	*a*	*b*	*a* + *b*
	–	*c*	*d*	*c* + *d*
		a + *c*	*b* + *d*	*n*

The total number of tests is $a + b + c + d = n$.

The sensitivity, specificity, selectivity, false positive rate and false negative rates for the target organism can be calculated as follows:

Sensitivity = $a / (a + b)$

Specificity = $d / (c + d)$

False positive rate = $c / (a + c)$

False negative rate = $b / (b + d)$

Selectivity = a / n

A further parameter, efficiency (E), which gives the fraction of colonies correctly assigned, can be calculated as $E = (a + d) / n$.

The confirmed count relates to the organisms that have been shown by some method, described in the procedure to be the target organisms. Where the method itself includes a confirmation procedure (e.g. production of indole from tryptophan for *E. coli*) then this method will suffice. Where the procedure does not include a confirmation step (e.g. ISO 9308-2[11]) then some alternative confirmation (secondary confirmation) step can be used. Such methods might include commercially available identification kits, other phenotypic methods (e.g. tests for a certain trait or enzyme system), tests of chemical composition (e.g. MALDI-TOF) or molecular methods. The choice of method will influence the outcome of the test and care should be taken when selecting a confirmation procedure as this defines the target.

NOTE For MPN methods, the same approach can be applied. The term “colonies” can be changed into “aliquots”, “typical” into “positive” and “atypical” into “negative”.

6.2.5 Worked example

The method ISO 9308-1[10] utilizes a medium containing chromogenic substrates for β-D-galactosidase and β-D-glucuronidase. Consequently, coliform colonies are coloured pink to red, *E. coli* colonies are blue to violet and non-target colonies should be colourless. No confirmation step is required for *E. coli* when using this method. For the primary characterization of the method for the recovery of *E. coli*, 16S rRNA sequencing could be used as the secondary confirmation procedure. During validation all colonies, blue/violet, pink/red and colourless would be examined using 16S rRNA sequencing.

In a primary characterization experiment, 300 blue/violet colonies were sequenced and 285 were shown to be *E. coli.* A total of 600 pink/red colonies were sequenced and 30 shown to be *E. coli.* A total of 300 colourless colonies were sequenced and none were shown to be *E. coli.*

Table 2 shows the results from the 20 samples analysed for the primary characterization of the above-cited method.

Table 2 — Tabulation of the counts for the categorical characteristics determination

Sample	*a*	*b*	*c*	*d*
1	15	3	1	42
2	8	0	0	33
3	4	1	0	26
4	15	3	1	50
5	16	1	0	45
6	12	5	0	48
7	6	0	1	38
8	10	1	1	29
9	14	2	0	53
10	18	0	2	51
11	17	2	0	45
12	19	0	1	63
13	13	2	2	40
14	11	3	1	39
15	13	0	0	35
16	25	3	2	33
17	21	1	0	54
18	16	0	1	55
19	15	1	2	40
20	17	2	0	51
Sum	**285**	**30**	**15**	**870**

In this scenario, the following data would be generated:

		Method-defined count		
		+	–	
Corroborated count	+	285	30	315
	–	15	870	885
		300	900	1 200

The method-defined count is the count obtained when following the methods described in the original procedure. The corroborated count is that obtained after secondary confirmation.

The total number of tests (n) is 1 200.

The sensitivity, specificity, selectivity, false positive rate and false negative rates for the target organism can be calculated as follows:

Sensitivity = a / (a + b), i.e. 285/315 = 90,5 %

Specificity = d / (c + d), i.e. 870/885 = 98,3 %

False positive rate = c / (a + c), i.e. 15/300 = 5,0 %

False negative rate = b / (b + d), i.e. 30/900 = 3,3 %

Selectivity = a / n, i.e. 285/1 200 = 23,8 %

A further parameter, efficiency (E), which gives the fraction of colonies correctly assigned, can be calculated as E = (a + d) / n, i.e. 1 155/1 200 = 96,3 %.

6.3 Determination of the upper limit and consideration of the lower limit of detection

6.3.1 Working range

The working range of a method is often specified in the original description of the procedure or defined by manufacturers when the test is a commercially available one. However, these figures may not always be accurate and can vary with different samples types.

For MPN-based methods, the working range will be determined by the sample size used and the number of reaction vessels used. The practical upper limit has been exceeded when all tubes of all dilutions are found positive.

6.3.2 Upper limit related to linearity

With colony count methods precision theoretically improves steadily with the number of target colonies observed in the detection set. The upper limits are defined by the space requirements and ensuing interactions of microbial colonies.

In practice colony count methods have an upper limit per detector which varies with the testing situation. The colony count detector (agar plate, membrane filter) becomes “clogged” or saturated for several reasons of which the number of target colonies is only one.

Every method has an upper reliable limit. It is not a clearly fixed number but a region of colony numbers where counts per plate become too uncertain to base a valid enumeration upon.

An experimental design based on a finely graded series of dilutions or volumes with replication of plating and counting provides the data for determining the upper working limit.

A carefully mixed liquid sample is prediluted to a density giving an expected colony number per plate that somewhat exceeds the assumed upper limit of the detector performance. A series of six or seven further dilutions with dilution steps 1:2 is continued from the starting suspension. Three parallel plates are seeded from each dilution.

Plates from dilutions averaging more than 20 colonies per plate are read.

The data are analysed for proportionality and over-dispersion of parallels assuming perfect randomness at every step as the basis of evaluation.

A single value of the G^2 test is of rather limited use. Performing similar proportionality tests on different samples helps determine the highest colony count where proportionality of the method is sufficient.

6.3.3 Type and number of samples to be used

The working range of a method can be determined by the use of naturally contaminated samples, pure cultures or samples spiked with contaminated material containing the target organism (e.g. sewage for enteric organisms). In the most practical choice, it should be a pure culture. Examination of a minimum of 20 samples is required.

When using sewage or surface water, a minimum of three sources should be employed.

NOTE 1 The determination of working range can be complicated for methods where there are two target organisms sought on the same plate, for example the simultaneous detection of total coliforms and *E. coli* on a single membrane filter. In such circumstances, the working range may be different for total coliforms and *E. coli* if *E. coli* are present in the sample together with other total coliforms. When determining the working range for these types of methods, the working range could be reported as the total number of target colonies present since the ratio of *E. coli* to other coliforms will vary from sample to sample.

NOTE 2 Another confounding factor in determining the working range is the selectivity of the procedure. For example, a membrane filtration-based method for the detection of total coliforms and/or *E. coli* which has poor selectivity may allow many other non-target organisms to grow. Growth of these organisms can obscure or inhibit the growth of target organisms. In such cases, the working range could be quoted in terms of the total number of target and non-target colonies on the membrane as the working range is significantly impacted by the growth of non-target organisms.

6.3.4 Worked example

6.3.4.1 Preparation

A natural sample was prediluted to suitable level and a dilution series of six successive steps of 1:2 was prepared.

Three parallel plates were made from each dilution using the surface spread technique. The colonies were counted after incubation and the results are shown in Table 3.

Table 3 — Tabulation of the counts in a linearity experiment

Dilution	Parallel counts			Sum	Mean	Relative volume	
				S_i	$\bar{x}$	R_i	S_i/R_i
2^{-1}	121	204	162	487	162,3	32	15,2
2^{-2}	109	128	148	385	128,3	16	24,1
2^{-3}	111	114	97	322	107,3	8	40,3
2^{-4}	56	60	68	184	61,3	4	46,0
2^{-5}	36	29	24	89	29,7	2	44,5
2^{-6}	11	13	17	41	13,7	1	41,0
			Total:	**1 508**		**63**	-

NOTE The formula used for the calculations is given in Annex C, Formula (C.1).

6.3.4.2 General proportionality test

The agreement of the sums of parallel colony numbers with the respective relative volumes 32/16/8/4/2/1 is calculated using sums of parallel counts and the general G^2 (Formula (C.1)) as follows:

$$G_5^2 = 2\,[487\ln(487/32) + 385\ln(385/16) + 322\ln(322/8) + 184\ln(184/4) + 89\ln(89/2) + 41\ln(41/1) - 1\,508\ln(1\,508/63)] = 292{,}526.$$

The test statistic has 6 – 1 = 5 degrees of freedom. The value of the index is compared with the χ^2 distribution with 5 degrees of freedom, which corresponds to 11,070 for 5 % and 15,086 for 1 %.

The calculated value exceeds the theoretical value for 1 % (15,086) which means that the general linearity of the results is poor.

6.3.4.3 Interpretation of data and further steps

The detector system was not linear in this sample in the colony count range from 41/3 = 14 to 487/3 = 162 colonies per plate. From the inspection of the counts per relative volume (S_i/R_i), it can be observed that the high colony numbers deviate the most from the expectation. In the four dilutions from 2^{-3} to 2^{-6} the counts per relative volume (S_i/R_i) was approximately constant, about 43 on the average. A sudden change occurred between dilutions 2^{-3} and 2^{-2}.

The detector has become non-functional even before 160 colonies per plate, as less colonies than expected have grown for dilutions 2^{-2} and 2^{-1}. The ratio 2:1 between successive dilutions is not seen in the colony counts at the higher concentrations, indicating that at high colony concentrations, the results are not linear.

The same analysis of proportionality can be repeated without the counts of the first dilution. The agreement of the sums for five dilutions is calculated:

G_4^2 = 2 [385ln(385/16) + 322ln(322/8) + 184ln(184/4) + 89ln(89/2) + 41ln(41/1) – 1 021ln(1 021/31)] = 81,933.

With four degrees of freedom the reference values for the test statistic are 9,488 for 5 % and 13,277 for 1 %. The proportionality was still seriously out of statistical control when the highest mean count was 128.

The next step is the proportionality test for four dilutions:

G_3^2 = 2 [322ln(322/8) + 184ln(184/4) + 89ln(89/2) + 41ln(41/1) – 636ln(636/15)] = 2,328.

The reference values for three degrees of freedom 7,815 for 5 % and 11,345 for 1 % are much higher than the observed 2,328. No signs of systematic deviation from proportionality remain in the four dilutions beginning with the mean count of 107.

It can be concluded that linearity is only seen as the sample becomes more dilute. The point at which linearity is not seen, soon after the number of colonies per plate becomes higher than about 100, determines the upper limit of the method.

A similar plan should be repeated with a minimum of twenty samples to determine the working range.

This approach is only applicable with colony methods.

6.3.5 The lower limit of detection

The lower limit of detection cannot reliably be determined by experimentation and is largely a matter of definition and sample volume analysed. A detailed explanation is given in Annex B.

6.4 Assessment of precision: Determination of repeatability and reproducibility

6.4.1 General

ISO 5725- series was developed as a guidance document for characterizing the variability of standard measurement methods. Two measures of variability (or precision), repeatability and reproducibility, are accepted in many disciplines as representative of data encountered in measurement processes.

The characterization of a new method should provide the initial values of its precision estimates. Other laboratories need this information for their verification of the method and subsequently for establishing the systems of analytical quality control.

Applied to water microbiological methods, ISO 5725-1[2], ISO 5725-2[3] and ISO 5725-3[4] need some adaptations because the basic principles originally applied to continuous data and not to discrete data (counts).

The main levels of precision are generally evaluated in two different conditions:

— the repeatability conditions which refer to the variability among measurements made on identical samples under identical circumstances. In ideal cases for microbiological determinations, these conditions of analysis are expected to follow a Poisson distribution. In practice, this is not always the case. Cases of over-dispersion can be detected by experiments between parallel counts (see Annex D and 6.4.2);

— the reproducibility conditions which refer to the variability among measurements made on identical materials under differing conditions by different laboratories following the same measurement method. Theoretically, reproducibility includes effects caused by differences among instruments, reagents, operators, laboratories, and environmental conditions. The assessment of this level of precision for microbiological methods leads to a relative operational variance which includes the maximum variability factors of uncertainty (see Annex F).

A third measure of precision of a method can be assessed in one laboratory with specific experimental planning (decomposition of the bias into elementary components such as operator, equipment, material effects, ...). It is known as intralaboratory reproducibility or intermediate precision (see 6.4.3 and Annex D for more details).

6.4.2 Repeatability

6.4.2.1 Design

The design for determining the repeatability performance of a method consists of 10 replicates of the same sample which are analysed in repeatability conditions, i.e. by the same technician on the same day, at the same approximate time and all samples incubated in the same incubator.

A minimum of three sets of repeatability data should be prepared using different sources of target organisms. Natural samples are preferably used.

6.4.2.2 Worked example: Tabulation of counts

Table 4 shows 3 series of 10 measurements obtained in repeatability conditions with a plate method.

Table 4 — Tabulation of the counts in 3 repeatability experiments

Sample	Repeated measurements (plates)									
1	63	65	77	59	69	61	55	65	33	90
2	47	60	40	57	24	39	57	52	35	54
3	21	16	20	24	21	34	23	26	18	14
NOTE The formulae used for the following calculations are given in Annex A, Formula (A.8) and Annex D, Formulae (D.1) and (D.2).										

6.4.2.3 Worked example: Detection of over-dispersion by applying the Poisson index of dispersion

For sample 1, the observed value of x^2_{r-1} = [10 × (63^2+65^2+77^2+...+33^2+90^2) / (63+65+77+...+33+90)] − (63+65+77+...+33+90) = 30,582.

For the first series:

Arithmetic mean ($\bar{x}$) = 63,7

Variance (S^2) = 216,456

Relative operational variance (u_0^2) = (216,456 – 63,7) / $63,7^2$ = 0,038.

In Table 5, the calculated values of the relative operational variance for the three samples analysed are shown.

Table 5 — Calculation of the relative operational variance

Sample	Arithmetic mean $\bar{x}$	Variance S^2	Observed value of χ^2_{r-1}	Relative operational variance u_0^2
1	63,7	216,456	30,582	0,038
2	46,5	136,278	26,376	0,042
3	21,7	31,789	13,184	0,021

According to the Chi square distribution, the critical 0,05 probability value for (10-1) degrees of freedom is: 16,919.

The observed value of χ^2_{r-1} is greater than the critical 0,05 probability value for samples 1 and 2, therefore significant over-dispersion is detected in the series of repeated measurements. The relative operational variance u_0^2 gives an order of magnitude of the operational variability for each series of repeated measurements.

For sample 3, the observed value of χ^2_{r-1} is lower than the critical 0,05 probability value. The observed variability between parallel counts complies with Poisson distribution. Even if not statistically significant, the calculated relative operational variance can be retained for sample 3 for further global evaluation.

The average relative operational variance for the 3 samples is equal to 0,034. The final expression of the repeatability relative standard deviation in % is the square root of the average relative operational variance (18,4 %). It corresponds to the performance of the method on the test material in repeatability conditions.

6.4.3 Intralaboratory reproducibility

6.4.3.1 Design

Intralaboratory reproducibility is also known as intermediate reproducibility (or intermediate precision). It is evaluated by performing sets of replicates in conditions as different as possible within a single lab (e.g. different technicians, different incubators and different batches of media…).

The experimental design described in ISO 29201[15] can be used. The whole analytical process is duplicated using maximum variation of the lab parameters. Natural samples shall be studied whenever possible. A minimum of 30 samples are recommended.

6.4.3.2 Worked example for colony counts

Table 6 shows the results of 10 samples analysed in 2 replicates x_1 and x_2 performed in conditions as different as possible within the laboratory (intralaboratory reproducibility conditions).

Table 6 — Tabulation of the colony counts for 10 samples analysed in duplicate

Sample	Replicates		Arithmetic mean	Variance	Relative operational variance
	x_1	x_2	$\bar{x}$	S^2	u_0^2
1	34	23	28,5	60,5	0,039
2	17	15	16	2	-0,055
3	11	27	19	128	0,302
4	40	21	30,5	180,5	0,161
5	42	25	33,5	144,5	0,099
6	43	38	40,5	12,5	-0,017
7	25	12	18,5	84,5	0,193
8	34	28	31	18	-0,014
9	58	39	48,5	180,5	0,056
10	37	48	42,5	60,5	0,010

NOTE 1 The formula used for the calculations in Table 6 is given in Annex D, Formula (D.2).

NOTE 2 Theoretically, variance can never be negative. However, when an estimate of variance is obtained by subtraction and the experimental variances are based on small numbers of replicates such things can happen.

The average relative operational variance from the set of 10 pairs of counts is 0,077. The square root of the average relative operational variance (27,8 %) corresponds to the performance of the method on the test material in intralaboratory reproducibility conditions.

6.4.3.3 Worked example for MPN systems

10 samples were analysed in duplicate using a MPN method. For each sample, the two repeated measurements were performed in intralaboratory conditions, in order to consider the maximum variability of analytical conditions within the laboratory. The data are shown in Table 7.

Table 7 — Tabulation of MPN results for 10 samples analysed in duplicate

Sample	Measurements		Lower confidence limit	Upper confidence limit	Lower confidence limit	Upper confidence limit	Overlap of confidence intervals in intralaboratory reproducibility conditions
	M_1	M_2	$T_{0,1}$	$T_{1,1}$	$T_{0,2}$	$T_{1,2}$	
1	600,1	176,1	419,3	858,9	97,2	319,1	no
2	2 086,6	1 148,4	1 560,4	2 790,4	850,7	1 550,3	no
3	1 885,3	1 362,8	1 413,0	2 515,5	1 017,3	1 825,5	yes
4	76,8	110,0	31,9	184,9	52,5	230,6	yes
5	1 672,6	2 094,8	1 254,0	2 230,9	1 566,3	2 801,6	yes
6	799,8	311,8	576,6	1 109,5	196,4	494,9	no
7	196,7	143,8	111,8	346,3	74,9	276,2	yes
8	1 202,0	1 316,6	892,5	1 618,7	981,6	1 765,8	yes
9	7 100,7	7 682,9	4 488,8	11 232,5	4 845,4	12 181,9	yes
10	7 682,9	3 421,3	4 845,4	12 181,9	2 450,4	4 777,0	no

Table 8 shows the relative operational variances for MPN results.

Table 8 — Calculation of relative operational variance for MPN results

Sample	Intralaboratory reproducibility u_R^2	Intrinsic variability for first replicate u_{d1}^2	Intrinsic variability for second replicate u_{d2}^2	Average intrinsic variability u_d^2	Relative operational variance u_0^2
1	0,752	0,034	0,092	0,063	0,689
2	0,178	0,022	0,023	0,023	0,156
3	0,053	0,022	0,022	0,022	0,031
4	0,065	0,201	0,143	0,172	-0,107
5	0,025	0,022	0,022	0,022	0,004
6	0,444	0,028	0,056	0,042	0,402
7	0,049	0,083	0,111	0,097	-0,048
8	0,004	0,023	0,022	0,023	-0,019
9	0,003	0,055	0,055	0,055	-0,052
10	0,327	0,055	0,029	0,042	0,285

NOTE The formulae used for the calculations in Table 8 are given in Annex D, Formulae (D.4), (D.5), (D.6), (D.7).

The average relative operational variance calculated from the set of 10 pairs is 0,134. Its square root (36,6 %) is the variability of the method observed in intralaboratory reproducibility conditions.

6.5 Robustness

6.5.1 General

Robustness means tolerance towards slight changes in procedure or towards unavoidable variations in conditions of the laboratory environment.

The determination of robustness varies according to the type of procedure being studied. Specific studies are required.

The purpose of such studies is to specify limits within which a method may be expected to be fit for the intended use. Robustness considerations may lead into limitations concerning the scope and conditions of the use of the method.

6.5.2 Experimental designs for effects due to time and temperature

For most microbiological methods, time and temperature of incubation are of importance. The interaction between membranes and media and/or media and reaction vessels may also be of importance.

For commercially prepared test kits, the shelf life of the product may also be a parameter of interest, as can the shelf life of prepared media.

When studying the robustness of a method duplicate samples of either spiked or naturally contaminated samples should be run at the extremes of the parameter being studied. Same plates can be read repeatedly, returning the plates for further incubation in the incubator after each reading.

For example, if a method procedure states the temperature range as (35 ± 1) °C then samples should be run at 34 °C and 36 °C. If the incubation period is 18 h to 22 h then the samples should be read at 18 h and 22 h. It may be preferable to produce robustness data as a matrix using (for example) the maximum and minimum incubation temperatures for the maximum and minimum duration of incubation.

Another plan is to prepare a series of parallel plates, place them in different parts of the incubator and remove them one at a time after different periods in the incubator. The second design also measures incubation space effects (see ISO 29201[15]). Should the developer of a method claim applicability of the

method over a considerable temperature range, special experiments involving several incubators set at different temperatures are required.

A minimum of 30 data points should be collected for each parameter.

Data should be collected using a minimum of three sources of target organisms whether these be reference cultures or laboratory-prepared spike material.

Graphical methods and standard statistical tests such as parametric or non-parametric tests of inference can be used for the analysis of the data whether these are generated by direct microscopical counts, MPN or colony counts.

6.6 Relative recovery

6.6.1 General

Trueness refers to the closeness of observed results to the true value, but the absolute trueness of a microbiological test result cannot be proven.

The best that can be done is to agree on a consensus true value. This may be the result determined by another method, the mean of a certified reference material given by the producer or the mean of observations in different laboratories in quality assurance work.

Recovery compared with the accepted reference, also called relative recovery, is the best that can be done for quantitative determination of bias (deviation from the true value).

When membrane filters are part of the method they need to be used in the recovery comparison. However, membrane filters themselves may create part of the bias and thus affect the relative recovery. For information on recovery comparisons, see ISO 11133[12] and ISO 7704[8]. Furthermore the interaction between membrane filters and nutrient media is of interest and may add significantly to the bias.

6.6.2 Determination of relative recovery

True recovery by a method can be approached with tests on pure cultures or spiked sterilized samples using non-selective methods as reference (see also ISO 11133[12]). Reference materials are also available for the purpose. These approximations, however, depend on the recovery efficiency of the methods used in the testing of the reference material or in determination of the reference value.

The recovery efficiencies of different microbiological methods vary considerably and may also be significantly affected by matrix effects. It is therefore prudent to determine the relative recovery of methods with the different matrices being tested.

For the determination of relative recovery, naturally contaminated samples, spiked samples or pure cultures can be used. For drinking water methods, when choosing the samples to be used for such studies the amount of target organisms added to the spiked sample should exceed those that typically occur naturally by at least an order of magnitude. For other matrices such as recreational waters, biosolids and wastewaters, the number of target organisms typically encountered is often suitable, but such samples may require dilution.

A minimum of three different organisms should be used for the spiking experiments and a minimum of thirty data points should be collected.

To be as realistic as possible, the relative recovery should then be studied using naturally contaminated samples rather than artificial test materials. Alternatively, samples spiked with naturally contaminated materials can be used.

To be free of subjective interpretation, the comparison must be based on confirmed counts generated by confirming all colonies.

The procedure described in ISO 17994 is recommended for determining the relative recovery performance of the method (both colony counts and MPN).

6.7 Uncertainty of counting

6.7.1 General

The baseline reliability of the counts is studied by repeated counting of the colonies of the same plates within a short time. The observations on counting uncertainty will give the first impression of potential problems with wide use of the method.

According to ISO 29201[15], repeatability and intralaboratory reproducibility of counting shall be evaluated, i.e. individual or collective numerical estimates of counting uncertainty expressed as relative standard deviation.

The uncertainty of counting serves as a base value when estimating other robustness features such as time-sensitivity.

For MPN-based methods, the intralaboratory uncertainty of reading by different operators can be studied according to ISO 29201[15], as well. The results of the same MPN system can be read by different operators. The uncertainty calculations are made using the MPN values obtained by each operator.

6.7.2 Experimental design for assessing the uncertainty of counting colonies

— Read the same plates repeatedly under uniform conditions, i.e. within a time interval clearly shorter than the assumed tolerance allowed for the method. In practice this means a maximum interval of one hour.

— The plates for repeated counting should be selected at random ignoring plates with less than 20 colonies and not selecting unusual ones. Otherwise, the plates picked should represent the whole working range of the method.

— For a reliable general estimate, at least 30 plates should be available.

6.7.3 Example of individual (or personal) uncertainty of counting colonies

A technician familiar with the microbiological method should read different plates twice within a short time interval (e.g. less than one hour). The duplicate counts denoted by x_1 and x_2 are shown in Table 9.

Table 9 — Uncertainty of counting colonies

Plate	x_1	x_2	x_1-x_2	x_1+x_2	$u^2_{rel,L}$
1	129	122	7	251	0,002
2	417	377	40	794	0,005
3	73	80	-7	153	0,004
4	49	52	-3	101	0,002
5	86	81	5	167	0,002
6	37	39	-2	76	0,001
7	112	115	-3	227	0,000
8	204	214	-10	418	0,001
9	66	71	-5	137	0,003
10	306	299	7	605	0,000
Sum	1 479	1 450			0,020
NOTE The formula used for the calculations in Table 9 is given in Annex E, Formula (E.2).					

The average estimate of the personal relative variance of counting is the mean value of $u^2_{rel,L}$ = 0,020/10 = 0,002. Its square root 0,045 thus indicated a 4,5 % relative standard uncertainty of the repeatability of counting by this person.

In this example of real data on total colony counts on non-selective media, the repeatability relative standard deviation was mostly larger than the "ideal" ($u_{rel,L}$ < 0,02) but remains below the boundary of 0,1.

If the value is high (larger than 0,1), it may be worth returning to the table to examine the individual $u_{rel,L}$ values in search for reasons. One accidental large value may be responsible or a trend on the mean colony count may be present. They are best illustrated graphically by plotting the $u_{rel,L}$ values against the colony count.

6.7.4 Example of intralaboratory uncertainty of counting colonies

Five technicians participated in a colony-counting session. Standard agar plates were picked from the available determinations and were read by each participant. Plates with less than 20 colonies were omitted. Results of six plates are shown in Table 10.

NOTE Six plates are far too few for a reliable general estimate but illustrate the computations.

The mean (m) and standard deviation (s) of each plate are first computed. From them, the values of $u_{rel,L} = s/m$ are obtained (last but one column).

Table 10 — Intralaboratory uncertainty of counting

Plate	A1	A2	B1	B2	B3	m	s	$u_{rel,L}$	$u^2_{rel,L}$
1	33	26	33	34	33	31,8	3,271	0,103	0,011
2	160	156	166	176	174	166,4	8,649	0,052	0,003
3	142	128	142	146	139	139,4	6,841	0,049	0,002
4	78	97	81	81	83	84,0	7,483	0,089	0,008
5	89	94	81	94	92	90,0	5,431	0,060	0,004
6	38	44	38	42	40	40,4	2,608	0,065	0,004
Sum	540	545	541	573	561				0,031
NOTE The formulae used for the calculations in Table 10 are given in Annex E, Formulae (E.1) and (E.3).									

The sum of $u^2_{rel,L}$ (sum of relative variances) is 0,031 and their mean is 0,005. Its square root 0,071 is the average relative intralaboratory uncertainty of counting with this method and group of operators (7,1 %).

6.7.5 Example of intralaboratory uncertainty of reading MPN

The results of 30 different samples were read by two different operators. The results of the five first samples are shown in Table 11:

Table 11 — Intralaboratory uncertainty of reading MPN

Sample	Analyst 1	Analyst 2	Relative variance
1	1 409,3	1 273,8	0,005
2	3 074,5	2 905,3	0,002
3	4 984,2	5 363,5	0,003
NOTE The formula used for the calculations of relative variance in Table 11 is given in Annex E, Formula (E.1) used in a squared form.			

Table 11 *(continued)*

Sample	Analyst 1	Analyst 2	Relative variance
4	1 114,0	1 047,1	0,002
5	651,1	778,3	0,016
		Mean	0,006

NOTE The formula used for the calculations of relative variance in Table 11 is given in Annex E, Formula (E.1) used in a squared form.

The square root of the mean relative variance is the average value of the relative uncertainty of reading MPN by different operators. In the numerical example, its value is 0,077 (7,7 %).

7 Designs for single laboratory verification of a method

7.1 General considerations

This section describes the procedures to be carried out in order to verify that the method is performing adequately in a given laboratory. The characteristics studies (see Table 12) are not as broad as those used for the initial determination of performance characteristics and the number of data points required is generally lower.

Table 12 — Minimum performance characteristics required for single laboratory verification

Parameter	Definition
Sensitivity	The fraction of the total positives[a] correctly assigned in the presumptive count
Specificity	The fraction of the total negatives[b] correctly assigned in the presumptive count
False positive rate	The fraction of positive results (e.g. typical colonies) that are subsequently shown to be due to non-target organisms
False negative rate	The fraction of negative results (e.g. atypical colonies) shown to be target organisms
Selectivity	The ratio of the number of target colonies to the total number of colonies in the sample volume
Efficiency	The fraction of total colonies correctly assigned in the presumptive count
Repeatability	Precision under repeatability conditions (same operators, same operating conditions, short period of time...)
Uncertainty of counting	The relative standard deviation of replicate counts of the target obtained by repeated counting (plates, fields, tubes, etc.) under stipulated conditions (same person, different person, same laboratory, etc.)

[a] Positives may be colony counts, positive reaction vessels (MPN) or cell counts.

[b] Negatives may be atypical colonies, negative reaction vessels (MPN) or cells without the specific characteristics required.

7.2 Calculation of sensitivity, specificity, efficiency, selectivity, false positive rate and false negative rate

7.2.1 Type of sample to be used

Samples containing 20 to 80 target organisms per volume examined are prepared. Naturally contaminated materials are preferred for the preparation of samples wherever possible. For faecal indicator bacteria, sewage polluted surface water or sewage effluent can be used. The samples are then examined by the procedure being studied. Typical and atypical colonies (i.e. those having the typical appearance of the target organism **and** those not having the typical appearance of the target organism) are counted. Both typical and atypical colonies are then identified using an appropriate procedure which could include commercially available microbial identification kits, DNA sequencing or other specified procedures.

The types of sample to be used vary with the method being studied and the target organism. For example, for methods for total coliforms drinking water spiked with surface water or sewage effluent is appropriate. Make dilutions in drinking water to obtain 20 to 80 target organisms per 100 ml. For membrane filtration methods that detect both total coliforms and *E. coli* on the same membrane, the target number of *E. coli* could be 10 colony forming units or possibly even less.

7.2.2 Number of samples

A minimum of five samples of drinking water should be spiked. The surface water or sewage used to prepare the spike material should be obtained from at least two sources.

7.2.3 Procedure for confirmation

Samples are incubated and all colonies confirmed according to the procedures appropriate to the method (e.g. for ISO 9308-1[10] all presumptive coliforms must be tested for the production of cytochrome oxidase). Where methods have no confirmatory procedures (e.g. ISO 9308-2[11]) described, then results are recorded as described in the method with no confirmation. Colonies (or reaction vessels: wells, tubes, etc.) are recorded as positive or negative.

When determining the parameters sensitivity, specificity, false positive rate and false negative rate it is necessary to apply a further confirmatory test to confirm (corroborate) or deny the results generated by the test method. Tests based on commercially available identification kits or other phenotypic methods (e.g. tests for a certain trait or enzyme system) are recommended, while the use of tests of chemical composition (e.g. MALDI-TOF) or molecular methods can be mainly used for primary characterization.

7.2.4 Categorical performance characteristics

7.2.4.1 When a confirmation step is included in the method, the identification data can be divided into four categories:

a) number of typical colonies confirmed as being the target organism in the primary confirmatory test the identity of which is supported by the secondary identification test (true positives);

b) number of atypical colonies, or typical colonies that are negative in the primary confirmatory test identified as being the target organism by the secondary identification test (false negatives);

c) number of typical colonies confirmed as being the target organism by the primary confirmatory test which are subsequently shown to not be the target organism by the secondary identification test (false positives);

d) number of atypical colonies or typical colonies that are negative in the primary confirmation test which are shown by the secondary identification test to not be a target organism (true negatives).

7.2.4.2 In the case of methods without confirmatory procedure:

a) number of typical colonies identified as being the target organism by an external identification test (true positives);

b) number of atypical colonies identified as being the target organism by an external identification test (false negatives);

c) number of typical colonies identified as not being the target organism by an external identification test (false positives);

d) number of atypical colonies identified as not being the target organism by an external identification test (true negatives).

7.2.4.3 The frequencies of these categories can be conveniently expressed in a 2 x 2 diagram:

		Presumptive count		
		+	–	
Confirmed count	+	*a*	*b*	*a* + *b*
	–	*c*	*d*	*c* + *d*
		a + *c*	*b* + *d*	*n*

The total number of tests is $a + b + c + d = n$.

The sensitivity, specificity, selectivity, false positive rate and false negative rates for the target organism can be calculated as follows:

Sensitivity = $a / (a + b)$

Specificity = $d / (c + d)$

False positive rate = $c / (a + c)$

False negative rate = $b / (b + d)$

Selectivity = a / n

A further parameter, efficiency (E), which gives the fraction of colonies correctly assigned, can be calculated as $E = (a + d) / n$.

NOTE For MPN methods, the same approach can be applied. The term "colonies" can be changed into "aliquots", "typical" into "positive" and "atypical" into "negative".

7.3 Determination of repeatability

The design for determining the repeatability performance of a method consists of 10 replicates of the same sample which are analysed in repeatability conditions, i.e. by the same technician on the same day, at the same approximate time and all samples incubated in the same incubator.

A minimum of three sets of repeatability data should be prepared using different sources of and levels of target organisms. Naturally contaminated samples are preferable. The three sets of data are then collected and examined using the procedure described in 6.4.2.2.

7.4 Uncertainty of counting

The reliability of the counts is determined by repeated counting of the colonies of the same plates, or positive tubes/wells of the same MPN system, within a short time. The observations on counting uncertainty will give an indication of potential problems with use of the method. Uncertainty of counting can be determined with single or multiple analysts. If multiple analysts routinely perform the test then uncertainty of counting should be determined with multiple analysts (see 6.7).

7.5 Procedure for single laboratory verification

The minimum characteristics required for a single laboratory verification are presented in Table 13.

Table 13 — Detailed requirements for the verification procedure

	Categorical performance characteristics	Repeatability [a]	Uncertainty of counting
Minimum number of samples, colonies/CFU and replicates	5 samples: 20 to 80 typical colonies /sample. 100 to 400 typical colonies (and associated atypical colonies) in the five samples. No more than 300 colonies per 90 mm plate or 80 per 47 mm membrane filter. No replicates.	3 samples: 1 times 10 replicates each – 20 CFU to 80 CFU	30 plates (preferably but not necessarily from different samples) Counts > 20 CFU. No more than 300 colonies per 90 mm plate or 80 per 47 mm membrane filter. One analyst: 30 samples x 2 counts. Multiple analysts: each analyst counts the 30 plates one time only.
Type of samples (in order of preference)	Naturally contaminated samples (real samples). Drinking water spiked with surface water or sewage effluent. (If spiked material, it should be from at least two sources).	Naturally contaminated samples (real samples). Drinking water spiked with surface water or sewage effluent. Reference materials. Water spiked with several strains (pure cultures) of typical and atypical colonies isolated in the laboratory.	Naturally contaminated samples (real samples). Drinking water spiked with surface water or sewage effluent. Reference materials. Water spiked with several strains (pure cultures) of typical and atypical colonies isolated in the laboratory.
Analysts	One or more analyst(s). If several analysts are working in the laboratory, several analysts are preferred.	Same analyst, same (or similar) time and same incubator per sample (but the different samples can be set up by different analysts). Repeatability should be calculated for a single analyst.	According to the laboratory routine, the uncertainty of counting may be calculated for one analyst or several (all) analysts (N). Repeated counts of the colonies in the same plates within a short time.
For the calculation of the parameter/s needed	To obtain pure cultures from typical and atypical colonies and confirm / identify the different strains using appropriate procedures.	To compare microbial counts.	To compare microbial counts.

[a] For MPN methods, an appropriate range should be selected based upon the specific design of the MPN procedure (number of tubes/wells/dilutions).

Table 13 *(continued)*

	Categorical performance characteristics	Repeatability [a]	Uncertainty of counting
Procedure	1) If the method has a confirmation system: confirm ALL colonies / strains. 2) If the method has not a confirmation system: identify ALL colonies / strains using appropriate systems. 4 kinds (and numbers)of colonies are obtained: Typical colonies confirmed / identified as being the target organisms (true positives). Atypical colonies confirmed / identified as being the target organisms (false negatives). Typical colonies confirmed / identified as not being the target organisms (false positives). Atypical colonies confirmed / identified as not being the target organisms (true negatives).	For each sample (10 replicates), calculate: Arithmetic mean $\bar{x} = \frac{\sum x_i}{10}$ Variance $s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$ Relative operational variance $u_0^2 = \frac{s^2 - \bar{x}}{\bar{x}^2}$ Poison index of dispersion (for 10 replicates, $r-1=9$) $\chi^2_{r-1} = \frac{10\sum x_i^2}{\sum x_i \sum x_i}$ Compare (using appropriate statistical tables) the observed χ^2_{r-1} value to the theoretical limits of χ^2 distribution with 9 degrees of freedom with: Critical values at 5 % $\chi^2_{r-1;0,05}$ Critical values at 1 % $\chi^2_{r-1;0,01}$	<u>ONE ANALYST</u> Calculate, for each plate, the addition ($x_1 + x_2$) and the difference ($x_1 - x_2$) of both replicate counts. The relative variance of each pair of counts (x_1 and x_2) is: $u^2_{\text{rel,L}} = 2\left(\frac{x_1 - x_2}{x_1 + x_2}\right)^2$ The average estimate of the personal (one analyst) relative variance of counting is the arithmetic mean of the relative variances of the different pairs of counts. Its square root $(u_{\text{rel,L}}) \times 100$ (expression in %) is the relative standard uncertainty of the repeatability of counting by this person. <u>*N* ANALYSTS</u> For each plate, calculate the relative variance: $u^2_{\text{rel,L}} = \frac{s}{m}$ where *m* and *s* are respectively the arithmetic mean and the standard deviation of the *N* replicate values (corresponding to *N* analysts) for the same plate. The average estimate of the relative variance of counting is the arithmetic mean of the relative variances of the different plates.

[a] For MPN methods, an appropriate range should be selected based upon the specific design of the MPN procedure (number of tubes/wells/dilutions).

Table 13 *(continued)*

	Categorical performance characteristics	Repeatability [a]	Uncertainty of counting
			The square root $\left(u_{\mathrm{rel,L}}\right)\times 100$ (expression in %) of the arithmetic mean of the relative variances of all the plates is the estimated intralaboratory uncertainty of counting
Results	**SENSITIVITY** a / $(a + b)$ **SPECIFICITY** d / $(c + d)$ **FALSE POSITIVE RATE** c / $(a + c)$ **FALSE NEGATIVE RATE** b / $(b + d)$ **SELECTIVITY** a / $(a + b + c + d)$ **EFFICIENCY** $(a + d)$ / $(a + b + c + d)$	1) If observed $\chi^2_{r-1} < \chi^2_{r-1;0,05}$ dispersion not significantly different from the one predicted by the Poisson distribution. 2) If $\chi^2_{r-1;0,05}$ < observed $\chi^2_{r-1} < \chi^2_{r-1;0,01}$ dispersion significantly greater that the one predicted by the Poisson distribution. 3)If $\chi^2_{r-1;0,01}$ < observed χ^2_{r-1} dispersion highly significantly greater that the one predicted by the Poisson distribution.	ONE ANALYST The ideal value of $u_{\mathrm{rel,L}}$ is < 0,02 but values < 0,1 are accepted. If the value is > 0,1, examine the individual $u^2_{\mathrm{rel,L}}$ values in search for reasons. *N* ANALYSTS The acceptable levels vary depending not only on the method but on the number of analysts. However, a guideline value for multiple analyst is 0,1.
Consult examples from primary characterization	See Worked example in 6.2.5	See Worked example in 6.4.2	See Worked examples in 6.7.3 (one analyst) and 6.7.4 (several analysts)

[a] For MPN methods, an appropriate range should be selected based upon the specific design of the MPN procedure (number of tubes/wells/dilutions).

Annex A
(informative)

Mathematical models of variation

A.1 General

The chance variation of particle numbers between parallel test portions is considerable even if the suspension is perfectly mixed (completely random) and no technical uncertainties of measurement are involved. This unavoidable intrinsic variation is a property of suspensions and the same for all microscopic and colony counting methods. It can be mathematically modelled by the Poisson distribution.

The Poisson distribution does not fully account for the intrinsic variation of MPN counts.

A.2 Intrinsic precision of colony counts

The variance of the Poisson distribution is equal to the mean. Equality of mean and variance does not prove that the data follow a Poisson distribution but inequality proves that they do not. Compatibility is traditionally tested by the use of the Poisson index of dispersion χ^2_{r-1} or the corresponding log-likelihood-ratio statistic G^2 (see also Annexes C and D).

The relative variance of the Poisson distribution is in inverse relation to the mean count, or more generally total count, of colonies in the detector. It means that with colony methods, precision is not a constant performance characteristic. The intrinsic relative uncertainty can be reduced by increasing the number of colonies.

$$u_{\text{rel}} = \frac{s}{x} = \frac{\sqrt{x}}{x} = \sqrt{\frac{1}{x}} = \frac{1}{\sqrt{x}} \qquad u_{\text{rel}}{}^2 = \frac{s^2}{x^2} = \frac{x}{x^2} = \frac{1}{x} \tag{A.1}$$

where

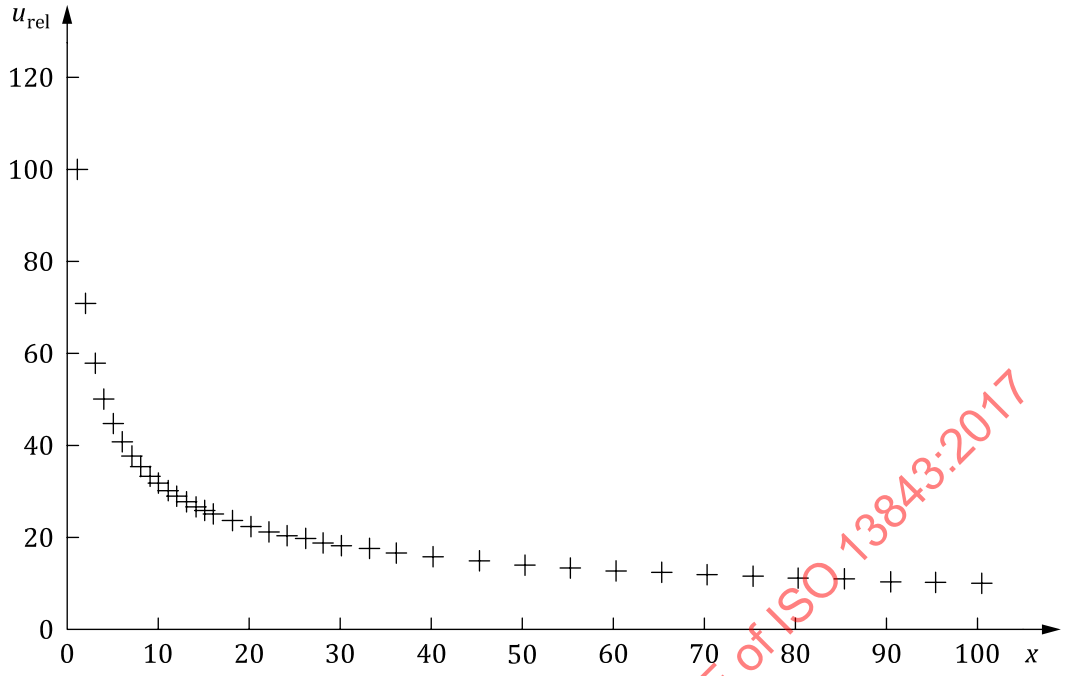
u_{rel} is the relative standard deviation;

s is the standard deviation;

x is the number of colonies observed.

EXAMPLE 1 A single plate count from a sample of a perfectly mixed suspension, say 48 colonies, has the theoretical relative precision of $u_{\text{rel}} = 1/\sqrt{48} = \pm 0{,}14$ (14 %).

The dependence of the relative precision (u_{rel}, relative standard deviation) on the particle count is illustrated in Figure A.1. Random relative uncertainty increases rapidly as the count decreases below twenty.



Key

x number of colonies observed

u_{rel} relative standard deviation

Figure A.1 — Relative standard deviation (expressed in %) of colony number in a perfectly mixed suspension following the Poisson distribution

The graph shows why colony numbers such as 20, 25, or 30 have been traditionally considered the lowest statistically reliable counts. In the count range below ten, which happens to be of considerable public health interest, single measurements are so imprecise that they can hardly be characterized as better than semi-quantitative. However, relative precision can be improved by carrying out repeated measurements.

EXAMPLE 2 Five parallel plates were inoculated with 1 ml test portions from the same laboratory sample. The number of colonies counted were: 6, 7, 11, 6, 9.

Sum of counts: 6 + 7 + 11 + 6 + 9 = 39

Mean of counts: 39/5 = 8

Theoretical relative precision of $u_{rel} = 1/\sqrt{39} = \pm 0{,}16$ (16 %).

The theoretical relative standard deviation of a single measurement of 8 colonies would be $u_{rel} = 1/\sqrt{8} = \pm 0{,}35$ (35 %).

The Poisson model can be used for estimating the lowest theoretical statistical uncertainty at any colony count and conversely for calculating the theoretical lowest count to reach a stipulated statistical precision (see also Annex B).

A.3 Intrinsic precision of MPN counts

A.3.1 General

The relative precision of MPN methods depends, in addition to the count itself, on the choice of the number of parallel tubes. Poisson distribution is assumed in every suspension but the presence-absence probability of positive reactions adds to the intrinsic variability.

A.3.2 Single dilution MPN

The relative precision of a single dilution series of n_t tubes in terms of the standard deviation in the natural logarithmic scale is expressed by Formula (A.2):

$$u(\ln M) = \frac{\left(1 - e^{-x}\right)}{x\sqrt{n_p e^{-x}}} \tag{A.2}$$

The x in the formula is the most probable number of organisms per tube which is estimated by Formula (A.3):

$$x = \ln\left(\frac{n_t}{n_t - n_p}\right) \tag{A.3}$$

where

M is the MPN value;

x is the number of organisms per tube;

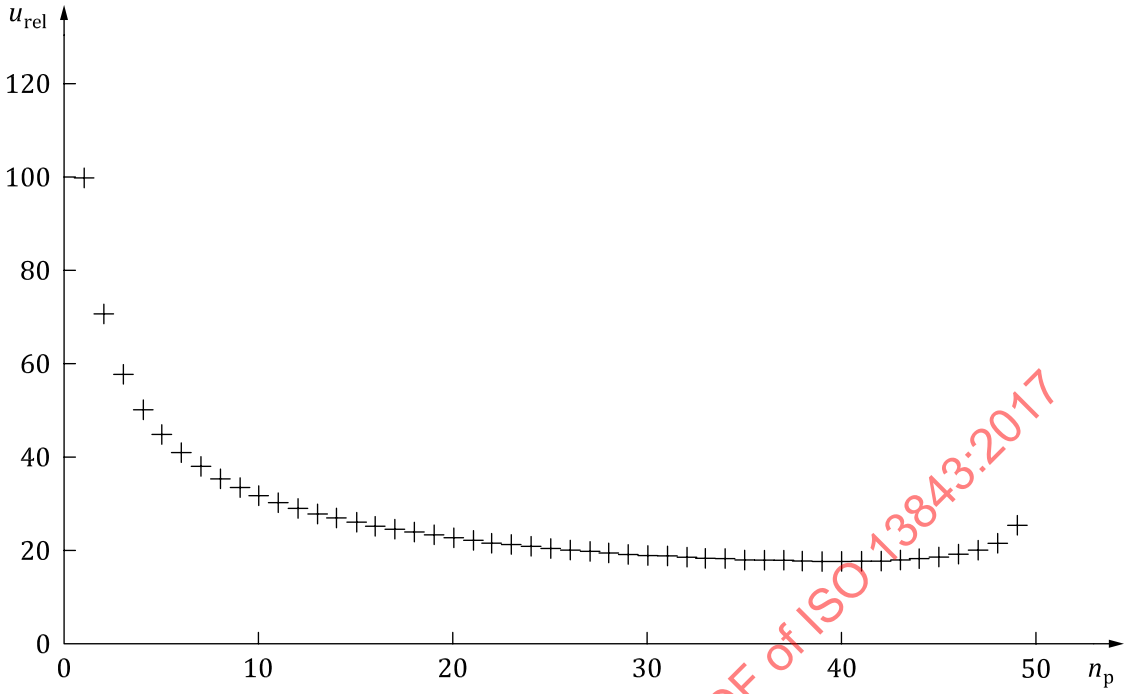
n_t is the number of tubes;

n_p is the number of positive tubes.

NOTE Formulae (A.2) and (A.3) can be combined, which gives a formula showing that the relative precision is completely defined by the total number of tubes and the number of positive tubes.

$$u_{rel}(M) \approx u\ln(M) = \frac{1}{\ln\left(\frac{n_t}{n_t - n_p}\right)} \cdot \sqrt{\frac{n_p}{n_t\left(n_t - n_p\right)}}$$

How the relative (logarithmic) precision of an MPN estimate varies with the number of positive tubes is illustrated in Figure A.2.



Key

n_p number of positive tubes

u_{rel} relative standard deviation

Figure A.2 — Relative standard deviation (expressed in %) of the single dilution MPN assay with 50 parallel tubes

Figure A.2 shows that the relative precision of MPN estimates has a minimum when 80 % of the tubes are positive. According to Formulae (A.2) and (A.3), the best achievable precision (smallest relative standard deviation) is therefore a function of the number of parallel tubes according to Formula (A.4):

$$u_{\min}(\ln M) = \frac{1{,}24}{\sqrt{n_t}} \tag{A.4}$$

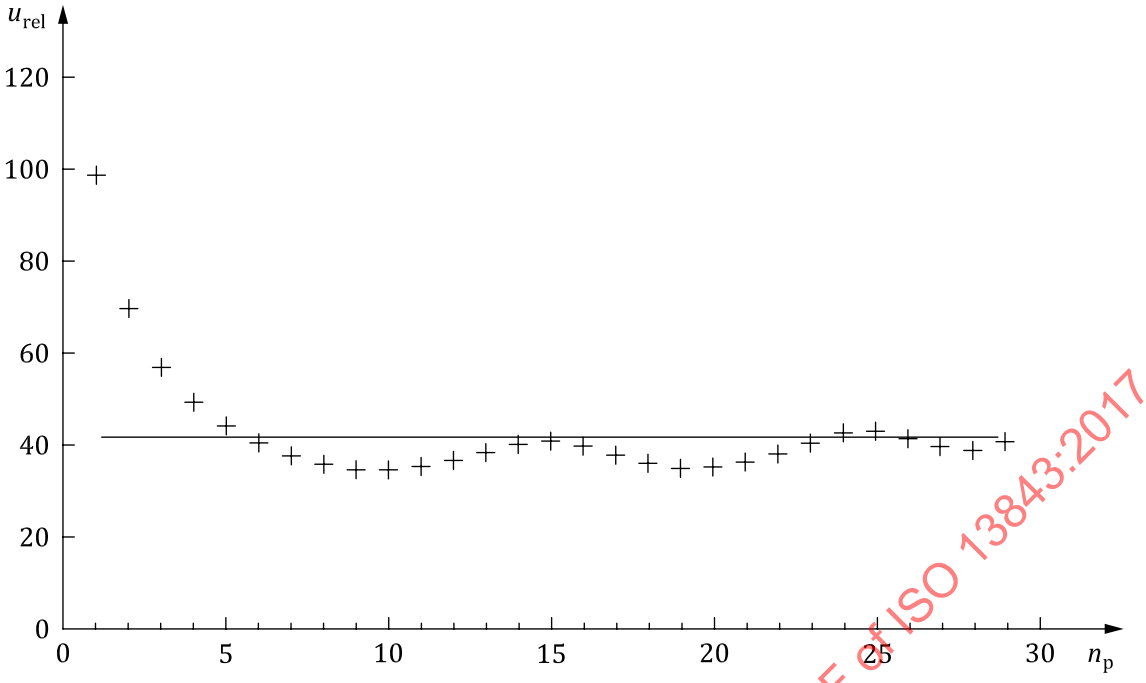
The uncertainty of an MPN estimate cannot become any smaller than this. When the proportion of positives differs from 80 % the relative standard uncertainty becomes greater, but as Figure A.2 shows there is a relatively long almost flat part of the precision curve between 60 % to 95 % positives.

A.3.3 Multiple dilution MPN

A.3.3.1 "Exact" precision

With multiple dilution MPN counts the relative precision is a wavy curve. It has as many local minima as there are dilution levels in the detector.

As an example, the relative standard deviation was calculated for the 3 × 10 tube MPN detector. Minima occur when the first, second, and third series of parallel tubes in turn become 80 % full of positives.



Key

n_p number of positive tubes

u_{rel} relative standard deviation

Figure A.3 — Relative standard deviation (expressed in %) of 3 × 10 tube MPN

The horizontal axis represents the number of positive tubes in the whole set. The undulating curve displays the calculated relative standard deviation.

The relative precision of a multiple dilution MPN depends mostly on the number of parallel tubes per dilution. Multiple dilutions are not, however, completely without effect. As an example, Figure A.3 shows that the lowest of the minima (the first one) for the 3 × 10 MPN systems has the value 0,350 (35 %) whilst the minimum relative standard deviation for a ten-tube single dilution MPN according to Formula (A.4) is 0,392 (39 %).

A.3.3.2 Approximate characteristic precision

Cochran proposed an approximate constant standard deviation for the entire MPN range. According to Cochran's equation, precision of the MPN estimate in logarithmic scale depends in a simple way on the number of tubes per dilution (n_t) and on the dilution factor (f) between consecutive dilutions.

$$s\lg(M) = 0{,}58\sqrt{\frac{\lg f}{n_t}} \tag{A.5}$$

The constant 0,58 was chosen "by the eye" for tenfold dilutions. If the dilution factor between dilutions is less than ten, then the constant 0,55 can be used.

Cochran's approximation is indicated by the horizontal line in Figure A.3. It is seen that the exact result deviates most from the approximation when most of the tubes in the set are negative.

The standard deviation of any individual MPN value is nowadays easily obtained by an appropriate computer program. Despite its approximate nature, Cochran's formula is useful in experimental

planning and method comparisons (see example below). It gives an idea of the characteristic precision of an MPN system.

EXAMPLE Assume a determination based on the MPN detection system of 3 × 32 wells in 96-well microtitre plate. Suppose further a dilution factor f = 3 between consecutive dilutions. The standard deviation of lg MPN, according to Cochran's approximate formula, is

$$s\lg(M) = 0{,}55\sqrt{\frac{\lg 3}{32}} = 0{,}55\sqrt{0{,}014\,9} = 0{,}067$$

Conversion to natural logarithmic scale gives $u\ \ln(M)$ = 2,303 × 0,067 = 0,155, approximately 15 % relative precision.

A.4 Over-dispersion

A.4.1 General

In the Poisson model technical and other additional uncertainties of procedure are assumed absent. Preparation of the initial suspension, dilution, inoculation, incubation and counting of the colonies are however not entirely free of uncertainty. Every technical step adds to the total variability of the measurement. Parallel determinations involving the whole analytical procedure cannot always be expected to follow the Poisson distribution. Over-dispersion, i.e. variation greater than fully random (in the Poisson sense), between parallel observations can be observed. This additional variation is partly method-dependent.

Another factor that causes higher than Poisson variation is the use of partial confirmation. Over-dispersion due to partial confirmation depends on the sample and the number and type of colonies selected for confirmation, not on the method itself. It forms a part of the uncertainty of measurement (see ISO 29201[15]).

For these reasons, over-dispersion is the normal state of microbiological test results and Poisson distribution is a simplified approximation. The Poisson model nevertheless sometimes works. That happens when the uncertainty due to the intrinsic variation is very great due to low counts. The technical uncertainties are dwarfed in comparison.

A.4.2 The negative binomial model

The common causes of over-dispersion, apart from the spurious errors, have effects roughly proportional to the mean or actually to the number of colonies. Other reasons for the same pattern of over-dispersion have been described previously. As the intrinsic precision due to the random scatter of particles in suspension follows the Poisson distribution, the total variance can be written as

$$s^2 = \bar{x} + u_0^2\bar{x}^2 \tag{A.6}$$

where

$\bar{x}$ is the mean number of objects (particles, colonies…) counted;

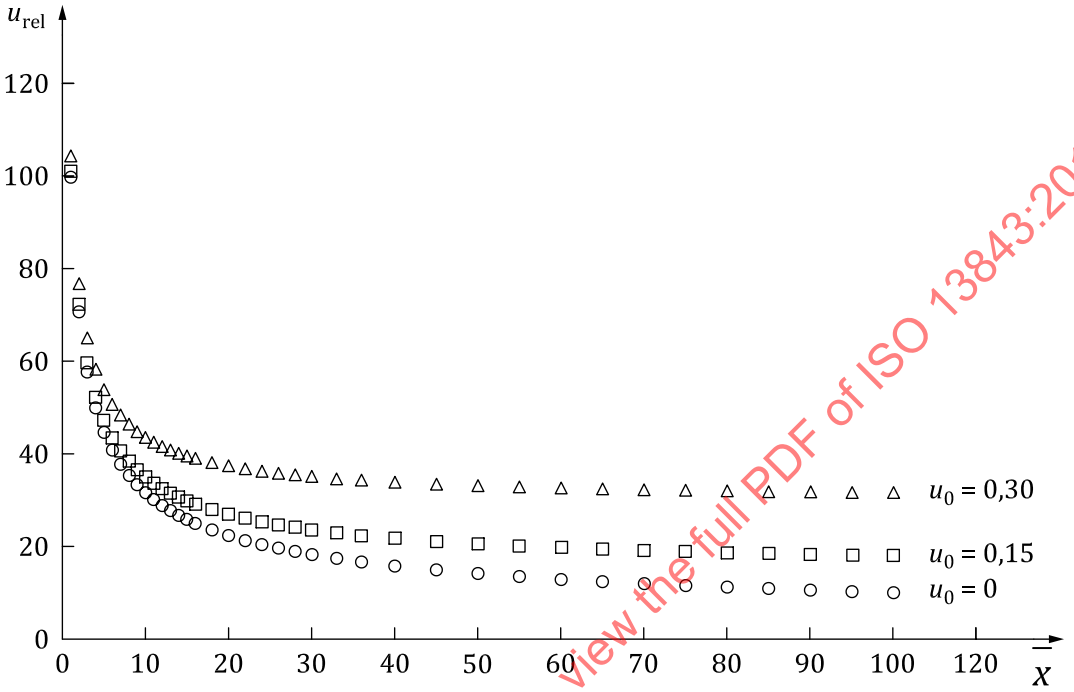
u_0 is the over-dispersion, the relative operational standard deviation.

The first part of the variance is due to the Poisson process, the rest is due to the combined effect of all the random over-dispersion factors. A statistical distribution with this model of variance is called a negative binomial distribution (other names are Gamma Poisson distribution and Pascal distribution).

The relative standard deviation of the distribution is

$$u_{\text{rel}} = \sqrt{\frac{1}{\bar{x}} + u_0^2} \tag{A.7}$$

Figure A.4 shows the effect of different degrees of over-dispersion on the total relative precision (relative standard deviation).



Key

$\bar{x}$ mean number of colonies

u_{rel} relative standard deviation

Figure A.4 — Effect of over-dispersion on the total relative precision (relative standard deviation expressed in %)

Lower curve: the Poisson model with no over-dispersion; upper curves: negative binomial with 15 % and 30 % over-dispersion (moderate additional variation u_0 = 0,15 and 0,30).

The average (or mean) colony number required to reach a given total relative precision is considerably higher in an over-dispersed situation than in the totally random (Poisson) case. It can be calculated by solving Formula (A.7) for the mean colony number $\bar{x}$.

EXAMPLE To achieve the relative standard deviation $u_{\text{rel}}^2 = 0{,}2$ when the over-dispersion is $u_0 = 0{,}15$ requires, according to Formula (A.7), the colony number $x = 1/(0{,}2^2 - 0{,}15^2) = 1/(0{,}04 - 0{,}022\,5) = 57$. The same precision is reached in a fully random (Poisson) situation with the colony number $x = 1/0{,}2^2 = 1/0{,}04 = 25$.

NOTE It is clear that total precision lower than the over-dispersion cannot be achieved within a single determination. The whole procedure could be repeated if better precision is required. With r parallel determinations the total relative standard deviation can be roughly estimated from

$$\sum x = \frac{1}{u_{\text{rel}}^2 - u_0^2}$$

where

$\sum x$ is the total number of colonies recorded;

u_0 is the over-dispersion constant, the relative operational standard deviation.

A.5 Detection of over-dispersion

Existence of statistically significant deviation from the Poisson distribution can be tested by applying the Poisson index of dispersion on a series of parallel counts

$$\chi^2_{r-1} = \frac{r\sum n_i^2 - \left(\sum n_i\right)^2}{\sum n_i} = \frac{r\sum n_i^2}{\sum n_i} - \sum n_i \tag{A.8}$$

or by the corresponding likelihood-ratio test statistic which can be computed from

$$G^2_{r-1} = 2\left[\sum\left(n_i \ln n_i\right) - \left(\sum n_i\right)\ln\left(\frac{\sum n_i}{r}\right)\right] \tag{A.9}$$

where in both formulae

r is the number of parallel observations;

n_i is the i^{th} observation (i = 1…r).

Both indices theoretically (asymptotically) follow the chi-square distribution, which enables statistical conclusions on the presence of over-dispersion (or under-dispersion).

NOTE Under-dispersion indicates 'too close' agreement of parallel counts. There are few natural reasons for it. It may occur when the test portions consume a large part of the test suspension. It has also been observed to happen when the technician knows which plates belong to the same parallel series and knowingly or inadvertently steers the counts towards each other. Under-dispersion disappears if parallel plates are blindly coded and mixed with other plates for counting.

A.6 Quantifying over-dispersion

A.6.1 Anscombe's method I

Anscombe's first method is applicable when enough independent observations on a single sample are available to base reliable estimation of the variance and mean on. In the range of mean values and relative operational variances that are of interest in the use and evaluation of colony-counting methods, Anscombe's method is efficient. It consists of solving Formula (A.6) for u_0^2

$$u_0^2 = \frac{s^2 - \bar{x}}{\bar{x}^2} \tag{A.10}$$

To be efficient, the observations should be in the optimal counting range. The mean should not be smaller than 20. This applies especially if the purpose is to estimate the operational variability in a single experiment. If several experiments are available, care will be taken to check the operational variability values obtained from low counts.

A.6.2 Regression approach

Whenever parallel observations on the same sample are available it is possible to calculate an estimate of the variance and of the mean. A solution, also based on Formula (A.6), makes use of replicate data from several samples. It is of advantage if the means cover a wide range. Series with more than two parallels are preferable. At least thirty samples should be studied.

Computing the variance-to-mean ratio (K) from several replicate sets provides a number of ($\bar{x}$, K) pairs. A regression line fitted on such data provides an estimate of the relative operational variance.

Dividing both sides of Formula (A.6) by the mean number of colonies yields an equation of a line. Its slope represents the relative operational variance.

$$K = \frac{s^2}{\bar{x}} = \frac{\bar{x} + u_0^2 \bar{x}^2}{\bar{x}} = 1 + u_0^2 \bar{x} \tag{A.11}$$

The random scatter is inevitably considerable if the estimates of mean and variance are based on small numbers of parallel determinations. The advantage of this approach is that the estimate of over-dispersion is based on a large selection of different samples. According to Formula (A.11), the y-intercept is expected to have the value 1. It may not appear so when the regression equation is fitted. If the y-intercept is significantly higher than 1, it indicates over-dispersion also at the detector level, i.e. in the test suspensions.

NOTE Formula (A.11) applies only when the mean is based on untransformed colony (or particle) numbers. Dilution factors and/or logarithms are not used.

A.6.3 Index of dispersion (χ^2 or G^2) approach

Calculation of the index of dispersion of parallel determinations is routinely applied for quality assurance purposes in many laboratories. As a consequence, large amounts of precision data suitable for over-dispersion calculations may be automatically accumulated. The third method of estimation makes use of such data.

The relationship between the mean, variance, and the index of dispersion is

$$\chi^2_{n-1} = \frac{s^2}{\bar{x}}(n-1) \tag{A.12}$$

The variance-to-mean ratio multiplied by the degrees of freedom gives the value of the (Poisson) index of dispersion. The value of the corresponding log-likelihood-ratio statistic $G^2_{(n\text{-}1)}$ can legitimately be substituted for it. Dividing both sides of Formula (A.12) by the degrees of freedom (n-1) gives the variance-to-mean ratio

$$K = \frac{\chi^2_{(n-1)}}{n-1} = \frac{s^2}{\bar{x}} \tag{A.13}$$

Assuming that the negative binomial model is the most likely description of statistical variation of parallel determinations, it is possible to insert the value of the variance-to-mean ratio from Formula (A.11):

$$K = \frac{s^2}{\bar{x}} = 1 + u_0^2 \bar{x} \tag{A.14}$$

Rearranging and solving for the relative operational variance

$$u_0^2 = \frac{K-1}{\bar{x}} \tag{A.15}$$

The mean of many estimates (respecting the algebraic sign) of relative operational variances from different samples provides a value for a general over-dispersion constant. In this approach Poisson distribution of test suspensions is taken for granted (see A.6.2).

A.6.4 Over-dispersion at detector level

It is a classical observation that parallel counts from a single suspension can vary more than the Poisson distribution accounts for. At this level, over-dispersion is caused by pipetting errors, uncertainty of counting and spurious errors ("accidents"), and possibly by the sample properties and incubation conditions. Over-dispersion at detector level is a useful quality assurance measure. It depends to some extent also on the method. It can be detected by the indices of dispersion (χ^2, G^2) (See A.6.3).

Annex B
(normative)

Assessment of the lower limits

B.1 General

The lower working limits of microbiological methods are to a large extent matters of definition.

At very low particle concentrations, all microbiological methods become essentially detection methods. The theoretical physical detection level for all methods is one particle of the target organism in the test portion or a detection system, such as MPN.

Because microbiological analytes consist of particles, there is a distinct statistical possibility that a microbe is absent from a test portion even though it is not absent from the laboratory sample.

The detection level is defined as the lowest analyte concentration that can be reliably detected (95 % probability of a positive result). The count that conforms to this definition is, on average, 3 particles per volume of material tested (see details in B.2).

The detection level is the property of suspensions and is the same for all colony count and MPN methods.

Alternatively, when a consensual relative standard deviation can be determined, the limit of determination can be used. It corresponds to the lowest analyte concentration where relative standard deviation equals the determined specified limit (see also B.3). For colony count methods, ISO 8199 mentions a limit of determination of 10 particles per test portion, corresponding to a relative standard deviation of around 32 %, in a fully random (Poisson) situation.

B.2 Detection level based on probability

B.2.1 Poisson model

The probability of a positive result p(+) when the Poisson distribution prevails can be calculated from

$$p(+) = 1 - e^{-x} \tag{B.1}$$

Solving the formula for x gives

$$x = -\ln\left[1 - p(+)\right] \tag{B.2}$$

where

e is the base of natural logarithms;

x is the number of particles per analytical portion.

One popular definition of the detection level is the concentration at which the probability of detecting the presence of the analyte equals 95 % [p(+) = 0,95].

According to the equation x = -ln(1-0,95) = -ln(0,05) = 3,0. Thus, at the average count of 3 (particles per test portion), the chances of detecting the presence of the analyte equals 0,95 (provided that the Poisson distribution prevails).

The detection level is the property of suspensions and does not distinguish one method from another. The detection level is the same for all colony count methods.

B.2.2 MPN methods

The detection level of MPN methods can be reasoned in the same way as for colony methods. The probability of detecting the presence of the analyte in the MPN system is given by Formula (B.1). Irrespective of the geometrical configuration the same average number of particles is needed in the system to ensure detection of the analyte with a chosen probability.

B.2.3 Negative binomial model of over-dispersion

The detection level, when defined in terms of probability can also be calculated from the probability of a negative result. Quoting Anscombe, but changing the symbols to the ones used in this document the probability of a negative result (probability of zero) is given by Formula (B.3):

$$\mathrm{p}(-)=\left(1+u_0^2\bar{x}\right)^{-1/u_0^2} \tag{B.3}$$

Solving for $\bar{x}$ gives the detection level when the probability (relative frequency) of negatives and the relative operational variance have been given.

$$\bar{x}=\frac{\mathrm{p}(-)^{-u_0^2}-1}{u_0^2} \tag{B.4}$$

As is evident from Figure A.4 (A.4.2), the detection level is rather little affected by moderate over-dispersion (see example below).

EXAMPLE The bacterial concentration required in order to achieve a 95 % probability of a positive result in an overdispersed situation depends on the relative operational variance. Assume a relative operational standard deviation u_0 = 0,30. Direct substitution of the probability of a negative result p(-) = 1 – p(+) = 1- 0,95 = 0,05 in Formula (B.4) yields $x = (0{,}05^{-0{,}09} - 1)/0{,}09 = 3{,}44$. The corresponding estimate with the Poisson distribution (no over-dispersion) would be $x = -\ln(0{,}05) = 3{,}00$ (see B.2.1).

NOTE There is no doubt that some causes of over-dispersion might also affect parallel MPN results. No experimental data or mathematical models for the over-dispersed MPN case seem to be available at present.

B.3 Limit of determination based on precision

B.3.1 General

The average concentration necessary for a specified relative uncertainty can be calculated by solving the precision formulae for the number of colonies.

B.3.2 The Poisson model

Solving Formula (A.1) for the number of colonies

$$x=\frac{1}{u_{\mathrm{rel}}^2} \tag{B.5}$$

where

x is the number of colonies observed in the detection system;

u_{rel} is the targeted relative precision (relative standard deviation).

EXAMPLE In a fully random (Poisson) situation, relative precision of 20 % is reached when the number of colonies observed in the detection system equals $x = 1/0{,}2^2 = 1/0{,}04 = 25$.

B.3.3 The negative binomial model

Solving Formula (A.7) for the number of colonies

$$\bar{x} = \frac{1}{u_{\text{rel}}^2 - u_0^2} \tag{B.6}$$

where

$\bar{x}$ is the mean number of colonies;

u_{rel} is the targeted relative precision;

u_0 is the relative operational standard deviation.

EXAMPLE To achieve the relative standard deviation $u_{\text{rel}} = 0{,}2$ when the relative operational standard deviation is $u_0 = 0{,}15$ requires, according to Formula (B.6) the mean colony number $\bar{x} = 1/(0{,}2^2 - 0{,}15^2) = 1/(0{,}04 - 0{,}0225) = 1/0{,}017\,5 = 57$. (Compare with the example in B.3.2).

B.3.4 Design detection levels

Irrespective of the analytical technique, method, or target organism, the detection level defined in terms of probabilities varies very little. Only extreme degrees of over-dispersion might change the picture slightly. The examples presented previously showed that e.g. 95 % probability of detecting the target would require about 3 particles of the analyte per analytical portion, on the average. The analytical portion in this case means the total volume of test suspension seeded in the detector.

A method that can handle a test portion of 100 ml has a 95 % probability of detecting (5 % probability of missing) the target when the average number is approximately three in the test portion. Another method that can only handle a test portion of 10 ml can detect the analyte at the same probability only at the average density of thirty per 100 ml. Such design detection levels vary between methods. It is possible to doubt the validity of a method for a purpose if it is structurally limited so that the probability of detection is not sufficient at concentrations that commonly occur.

Annex C
(normative)

Assessment of the upper limit

C.1 General

The number of target organisms per test portion where the linearity begins to deteriorate is considered as the upper limit of the method.

Linearity means a straight-line relationship of the observed result with concentration of the analyte. In microbiological contexts linearity means linear response of the count with the volume of test portion. Linearity is usually good at the low end of the scale, near the detection level.

C.2 Statistical evaluation of the upper limit

The statistical calculations are based on the G^2 index procedure.

Let n colony counts x_1, x_2,..., x_n be obtained from the study of the same test suspension in volumes or dilutions that are related as the numbers R_1, R_2,..., R_n.

The log-likelihood ratio estimate of the proportionality (linearity) of the counts can be calculated from

$$G^2_{n-1} = 2\left[x_1 \ln\frac{x_1}{R_1} + x_2 \ln\frac{x_2}{R_2} + \ldots + x_n \ln\frac{x_n}{R_n} - \left(\sum x\right) \times \ln\left(\frac{\sum x}{\sum R}\right)\right] \quad \text{(C.1)}$$

A guide value can be obtained by referring to tables of χ^2 with n-1 degrees of freedom. Values exceeding the tabulated value indicate departure from proportionality at the chosen probability level.

The number of germs per test portion where the linearity is lost can be regarded as the upper boundary.

NOTE Linearity is an aspect of trueness. Deviation from linearity develops gradually as the number of colonies increases.

Annex D
(normative)

Determination of the operational variability in repeatability and intralaboratory reproducibility conditions

D.1 General case: Statistical evaluation in a repeatability experiment

A worked example with the minimum required number of data is presented in 6.4.2.2. If more data are included in the repeatability experiment, Tables D.1 and D.2 and Formulae (D.1) to (D.3) can be used.

Table D.1 — Tabulation of the counts in a repeatability experiment

Repeated measurements			
n_1	n_2	...	n_i

a) Detection of over-dispersion by applying the Poisson index of dispersion on each series of parallel counts

$$\chi^2_{r-1} = \frac{r\sum n_i^2 - \left(\sum n_i\right)^2}{\sum n_i} = \frac{r\sum n_i^2}{\sum n_i} - \sum n_i \qquad \text{(D.1)}$$

where

r is the number of parallel observations;

n_i is the i^{th} observation (i = 1...r).

The Poisson index of dispersion theoretically (asymptotically) follows the chi-square distribution, which enables statistical conclusions on the presence of over-dispersion (or under-dispersion).

b) Carry on comparing the observed χ^2_{r-1} value to the theoretical limits of χ^2 distribution with r-1 degrees of freedom.

For a one-sided evaluation, the theoretical limits are:

— critical value at 5 %: $\chi^2_{r-1;0,05}$

— critical value at 1 %: $\chi^2_{r-1;0,01}$

c) Depending on the position of the observed value χ^2_{r-1} in relation to the theoretical limits, determine for each analytical series the significance of the observed dispersion regarding the dispersion predicted by the Poisson distribution.

Table D.2 — Statistical assessment of the data set in a repeatability experiment

Case	Position of observed χ^2_{r-1} in relation to the theoretical limits	Conclusion on the difference between the observed dispersion and the dispersion predicted by the Poisson distribution
1	observed $\chi^2_{r-1} < \chi^2_{r-1;0,05}$	Dispersion not significantly different from the one predicted by the Poisson distribution
2	$\chi^2_{r-1;0,05} <$ observed $\chi^2_{r-1} < \chi^2_{r-1;0,01}$	Dispersion significantly greater than the one predicted by the Poisson distribution
3	$\chi^2_{r-1;0,01} <$ observed χ^2_{r-1}	Dispersion highly significantly greater than the one predicted by the Poisson distribution

If cases 2 or 3 are observed, the relative operational variance u_0^2 is calculated using Anscombe's first method.

$$u_0^2 = \frac{s^2 - \bar{x}}{\bar{x}^2} \tag{D.2}$$

where

s^2 is the variance of parallel observations;

$\bar{x}$ is the arithmetic mean of parallel observations.

If case 1 occurs, u_0^2 is not significantly different from 0. For global assessment, the calculated relative operational variance as described above can be used.

To be efficient, the observations should be in the optimal counting range. The mean should not be smaller than 20. This applies especially if the purpose is to estimate the operational variability in a single experiment. If several experiments are available, care will be taken to check the operational variability values obtained from low counts.

When the different sets of repeatability data are statistically processed, the arithmetic mean of the different relative operational variances is calculated. Hence the average relative operational variability corresponds to the repeatability performance of the method.

The final expression of repeatability in % can be deduced:

$$u_0 = \sqrt{u_0^2} \times 100 \tag{D.3}$$

D.2 General case: Statistical evaluation in an intralaboratory reproducibility experiment for colony count method

A limited worked example (with a restricted number of data points) is presented in 6.4.3.1. Table D.3 can be used for the data processing of the intralaboratory reproducibility experiment.

Table D.3 — Tabulation of the colony counts in an intralaboratory reproducibility experiment

Sample No	Measurements	
1	x_{11}	x_{12}
2	x_{21}	x_{22}
...	...	...
q	x_{q1}	x_{q2}

a) Determine the relative operational variance u_0^2 using Anscombe's first method for each pair of counts.

b) Calculate the average relative operational variance from the set of q pairs of counts from the different tested samples. The final expression of intralaboratory reproducibility in % is the square root of the average relative operational variance, multiplied by 100.

D.3 General case: Statistical evaluation in an intralaboratory reproducibility experiment for MPN methods

A worked example is presented in 6.4.3.2. Tables D.4 and D.5 and Formulae (D.4) to (D.7) can be used for the data processing of the intralaboratory reproducibility experiment.

Table D.4 — Tabulation of MPN results in an intralaboratory reproducibility experiment

Sample	Measurements		Lower confidence limit	Upper confidence limit	Lower confidence limit	Upper confidence limit	Overlap of confidence intervals in intralaboratory reproducibility conditions
	M_1	M_2	$T_{0,1}$	$T_{1,1}$	$T_{0,2}$	$T_{1,2}$	
1	M_{11}	M_{12}	$T_{0,1,1}$	$T_{1,1,1}$	$T_{0,2,1}$	$T_{1,2,1}$	yes / no
2	M_{21}	M_{22}	$T_{0,1,2}$	$T_{1,1,2}$	$T_{0,2,2}$	$T_{1,2,2}$	yes / no
...	...	...	...	...	...	...	...
q	M_{q1}	M_{q2}	$T_{0,1,q}$	$T_{1,1,q}$	$T_{0,2,q}$	$T_{1,2,q}$	yes / no

Table D.5 — Calculation of relative operational variance for MPN results

Sample	Intralaboratory reproducibility	Intrinsic variability for first replicate	Intrinsic variability for second replicate	Average intrinsic variability	Relative operational variance
	$u_{R'1}^2$	u_{d1}^2	u_{d2}^2	u_d^2	u_0^2
1	$u_{R'1}^2$	$u_{d1,1}^2$	$u_{d2,1}^2$	$u_{d,1}^2$	$u_{0,1}^2$
2	$u_{R'2}^2$	$u_{d1,2}^2$	$u_{d2,2}^2$	$u_{d,2}^2$	$u_{0,2}^2$
...	...	...	...	...	...
q	$u_{R'q}^2$	$u_{d1,q}^2$	$u_{d2,q}^2$	$u_{d2,q}^2$	$u_{0,q}^2$

$$u_{R'}^2 = \frac{\left(\ln M_1 - \ln M_2\right)^2}{2} \quad \text{(D.4)}$$

$$u_{d1}^2 = \left(\frac{\left(\ln T_{1,1} - \ln T_{0,1}\right)}{2\times 1,96}\right)^2 ; u_{d2}^2 = \left(\frac{\left(\ln T_{1,2} - \ln T_{0,2}\right)}{2\times 1,96}\right)^2 \quad \text{(D.5)}$$

$$u_d^2 = \frac{u_{d1}^2 + u_{d2}^2}{2} \quad \text{(D.6)}$$

$$u_0^2 = u_{R'}^2 - u_d^2 \quad \text{(D.7)}$$

a) Determine the relative operational variance u_0^2 for each pair of counts.

b) Calculate the average relative operational variance from the set of q pairs of MPN results from the different tested samples. The final expression u_0 in intralaboratory reproducibility conditions (in %) is the square root of the average relative operational variance, multiplied by 100.

Annex E
(normative)

Uncertainty of counting

E.1 General

Methods should not be unduly sensitive to the operator. Operator effects are included in reproducibility tests (see D.2) and collaborative method performance study (see Annex F) used for precision assessment purposes. Global estimates of measurement uncertainty also include various operator effects (ISO 29201[15]).

It is useful to assess clearly identified operator effects. The most important are the between-operators differences in interpreting the presumptive count.

Differences in interpretation can be studied efficiently by blind parallel reading of the colonies of the same plates by the same or different operators (see E.2 and E.3). The exercise should produce individual or collective numerical estimates of counting uncertainty, expressed as relative standard deviation. The relative standard deviation calculated from the results of different persons is more informative than that calculated from duplicate counts by one person. (One person can usually duplicate the count, however wrong, with considerable precision.)

E.2 Statistical determination of the uncertainty of counting colonies and reading MPN

Assume a plate with the (unknown) number x of characteristic colonies, and denote with $x_1, x_2, \ldots, x_n$ the numbers observed in counting of the plate by the same person repeatedly or by different persons.

Counting uncertainty $u_{rel,L}$ is expressed as the relative standard deviation:

$$u_{rel,L} = \frac{s(x)}{\bar{x}} \tag{E.1}$$

where

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

NOTE The same formula in a squared form is used for the determination of the uncertainty of reading MPN.

E.3 Individual (or personal) uncertainty of counting colonies

The estimate corresponds to a summary of the results of one person reading the same plates twice.

The relative variance of each pair is calculated using Formula (E.2):

$$u^2_{rel,L} = 2\left(\frac{x_1 - x_2}{x_1 + x_2}\right)^2 \tag{E.2}$$

The average estimate of the squared personal uncertainty of counting is the arithmetic mean of the relative variances of the pairs of counts.