TECHNICAL REPORT

ISO/TR 13843

First edition
2000-06-01

Water quality — Guidance on validation of microbiological methods

Qualité de l'eau — Lignes directrices pour la validation des méthodes microbiologiques

Reference number
ISO/TR 13843:2000(E)

ISO copyright office
Case postale 56 • CH-1211 Geneva 20
Tel. + 41 22 749 01 11
Fax + 41 22 734 10 79
E-mail copyright@iso.ch
Web www.iso.ch

Printed in Switzerland

Contents

Page

Foreword

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

International Standards are drafted in accordance with the rules given in the ISO/IEC Directives, Part 3.

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

In exceptional circumstances, when a technical committee has collected data of a different kind from that which is normally published as an International Standard ("state of the art", for example), it may decide by a simple majority vote of its participating members to publish a Technical Report. A Technical Report is entirely informative in nature and does not have to be reviewed until the data it provides are considered to be no longer valid or useful.

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

ISO/TR 13843 was prepared by Technical Committee ISO/TC 147, *Water quality*, Subcommittee SC 4, *Microbiological methods*.

Water quality — Guidance on validation of microbiological methods

1 Scope

This Technical Report deals with validation of microbiological methods, with particular emphasis on selective quantitative methods in which the quantitative estimate is based on counting of particles either directly, with the aid of a microscope, or indirectly, on the basis of growth (multiplication) into colonies or turbidity.

The principles and procedures within this scope are commonly known as the presence/absence (P/A), most probable number (MPN), colony count and direct (microscopic) count.

This Technical Report does not apply to the validation of the so-called rapid or modern methods which mostly depend on measuring products or changes due to microbial activity but do not address the detection of individual particles.

2 Terms and definitions

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

2.1
accuracy of measurement
closeness of the agreement between a test result and the accepted reference value

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

[ISO 3534-1:1993, 3.11]

2.2
analyte
measurand
particular quantity subjected to measurement

NOTE 1 See reference [5].

NOTE 2 In microbiology the analyte is ideally defined as a list of taxonomically defined species. In many cases, in practice the analyte can only be defined by group designations less accurate than taxonomic definitions.

2.3
analytical portion
test portion
volume of particle suspension inoculated into a detector unit

NOTE Examples of a detector unit are agar plate, membrane filter, test tube, microscopic grid square.

2.4
application range
range of particle concentrations routinely subjected to measurement by a method

2.5
categorical characteristic
method performance characteristic numerically expressed as a relative frequency based on P/A or +/– classification

2.6
CFU, deprecated
colony-forming unit, deprecated
CFP, deprecated
colony-forming particle, deprecated

NOTE 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** (2.27) and **germ** (2.13) are terms with similar meanings but convey the original idea better and apply not only to colony-count methods but also to MPN and P/A.

2.7
coefficient of variation
CV
relative standard deviation
for a non-negative characteristic, the ratio of the standard deviation to the average

NOTE 1 The ratio may be expressed as a percentage.

NOTE 2 The term "relative standard deviation" is sometimes used as an alternative to "coefficient of variation", but this use is not recommended.

[ISO 3534-1:1993, 2.35]

NOTE 3 In this Technical Report the term coefficient of variation (CV) is used when the relative standard deviation is expressed in percent (CV % = 100 RSD).

2.8
collaborative test
method or laboratory performance test where several laboratories join in an experiment planned and co-ordinated by a leader laboratory

NOTE 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.

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

2.9
confirmed [verified] colony count
x
presumptive colony count corrected for false positives

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

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.

2.10
control chart
two-dimensional scattergram for monitoring method performance with control values obtained by a Type A study

NOTE In control charts the horizontal axis is usually in the time scale or ordinate scale and the control variable is the mean or some precision measure (s, CV, RSD).

2.11
detector
particle detector
plate of solid matrix or a tube of liquid containing a nutrient medium for counting or detecting living microbial particles

2.12
detection set
detector set
combination of plates or tubes on which quantitative estimation of microbial concentration in a sample is based

NOTE The detection set is the set of plates or tubes utilized for numerical estimation of a single value.

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

2.13
germ
living entity capable of producing growth in a nutrient medium

cf. **propagule** (2.27)

2.14
guidance chart
two-dimensional scattergram for presenting method performance data (quantity or precision) with arbitrary guide values or guide values obtained by Type B reasoning

NOTE In guidance charts, the horizontal axis is usually the colony count per detector.

2.15
heterogeneous Poisson distribution
distribution arising when the mean of a Poisson distribution varies randomly from occasion to occasion

NOTE 1 See reference [11].

NOTE 2 See also negative binomial distribution (2.19).

2.16
limit of detection
particle number x (per analytical portion) where the probability p_0 of a negative result equals 5 %

NOTE 1 Probability of a positive result $p(+) = 1 - p_0$.

NOTE 2 a) Calculation of x via Poisson distribution:

$$x = \ln\left(\frac{1}{p_0}\right) = \ln\left(\frac{1}{0,05}\right) = \ln(20) = 3,00$$

b) Calculation of x via negative binomial distribution:

$$x = \frac{\left(p_0{}^{-u^2} - 1\right)}{u^2} = \frac{0,05^{-u^2} - 1}{u^2} = \frac{20^{-u^2} - 1}{u^2}$$

2.17
limit of determination
lowest average particle concentration x per analytical portion where the expected relative standard uncertainty, equals a specified value (RSD)

NOTE a) Calculation of x via Poisson distribution:

$$x = \frac{1}{(\mathrm{RSD})^2}$$

b) Calculation of x via negative binomial distribution:

$$x = \frac{1}{(\mathrm{RSD})^2 - u^2}, \text{ given overdispersion factor} = u$$

2.18
linearity
linear dependence of the signal on concentration of the analyte

cf. **proportionality** (2.28)

2.19
negative binomial distribution
a particular "overdispersed" statistical distribution of counts

NOTE 1 Its variance can be expressed as

$$\sigma^2 = \mu + u^2 \mu^2$$

where μ is the mean.

NOTE 2 In this Technical Report the square of the overdispersion factor (u) is substituted for the inverse of the exponent ($1/k$) of the standard formula for the negative binomial distribution.

2.20
overdispersion
variation in excess of Poisson randomness

NOTE It is detected qualitatively by the Poisson index of dispersion, and measured quantitatively by estimating the parameter u (overdispersion factor) of the negative binomial distribution.

2.21
overdispersion factor
u
additional random uncertainty of determination in excess of the Poisson distribution, measured in terms of relative standard deviation

2.22
overlap error
crowding error
systematic depression of colony counts due to confluence of colonies

NOTE Quantitatively, overlap error depends primarily on the fraction of available growth space occupied by colonial growth.

2.23
parallel counts
particle or colony numbers in equal analytical portions drawn from the same suspension

2.24
Poisson distribution
fully random distribution of particle numbers when sampling a perfectly mixed suspension

NOTE The probability $P(k)$ of observing exactly k units in a test portion when the mean equals μ is calculated from

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

2.25
precision
closeness of agreement between independent test results obtained under stipulated conditions

NOTE Precision does not relate to the true value or the specified value. It is usually expressed in terms of imprecision and computed as a standard deviation of the test results.

2.26
primary validation
full validation
establishment of the specifications for the performance of a new method and/or experimental verification that a method meets theoretically derived quality criteria

2.27
propagule
a viable entity, vegetative cell, group of cells, spore, spore cluster, or a piece of fungal mycelium capable of growth in a nutrient medium

cf. **germ** (2.13)

2.28
proportionality
agreement of observed particle counts with the volume (or dilution) of a series of analytical portions from a common root suspension

NOTE Proportionality is computed for statistical evaluation as the log-likelihood ratio statistic G^2 with n–1 degrees of freedom.

2.29
qualitative method
method of analysis whose response is either the presence or absence of the analyte in a certain amount of sample

NOTE See reference [10].

2.30
recovery
general term 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 detector

2.31
relative accuracy
degree of correspondence between the response obtained by the reference method and the response obtained by the alternative method on identical samples

NOTE See reference [10].

2.32
relative difference
d
difference between two measured values divided by their mean

$$d = \frac{x_A - x_B}{\bar{x}} = \frac{2(x_A - x_B)}{x_A + x_B}$$

$$d\ \% = 100\ d$$

NOTE For all practical purposes, the same value results from the calculation $d = \ln(x_A) - \ln(x_B)$.

2.33
relative recovery
ratio (A/B) of colony counts obtained by two methods tested on equal test portions of the same suspension, where B is the reference (when applicable)

2.34
relative standard deviation
RSD
estimate of the standard deviation of a population from a sample of n results divided by the mean of that sample

$$\text{RSD} = \frac{s}{\bar{x}}$$

cf. **coefficient of variation** (2.7)

2.35
repeatability
closeness of the agreement between the results of successive measurements of the same measurand carried out under the same conditions of measurement

NOTE 1 See *Guide to the expression of uncertainty in measurement* [6].

NOTE 2 Repeatability is computed as $r = 2{,}8 s_r$, where s_r is the repeatability standard deviation.

2.36
reproducibility
closeness of the agreement between the results of measurements on the same measurand carried out under changed conditions of measurement

NOTE 1 See *Guide to the expression of uncertainty in measurement* [6].

NOTE 2 Reproducibility is computed as $R = 2{,}8\ s_R$,

where

s_R is the reproducibility standard deviation usually compounded from the between-laboratories standard deviation s_L and repeatability standard deviation s_r:

$$s_R = \sqrt{s_L^2 + s_r^2}$$

2.37
robustness
insensitivity of an analytical method to small changes in procedure

NOTE 1 See reference [23].

NOTE 2 To examine the robustness it is advisable to “abuse” the method in a controlled way.

2.38
secondary validation
demonstration by experiment that an established method functions according to its specifications in the user's hands

2.39
apparent selectivity
F
ratio of the number of target colonies to the total number of colonies in the same sample volume

$F = \lg(t/n)$

where

t is the apparent concentration of presumptive target types estimated by counting colonies;

n is the concentration of total colonies.

2.40
sensitivity
fraction of the total number of positive cultures or colonies correctly assigned in the presumptive inspection

2.41
specificity
fraction of the total number of negative cultures or colonies correctly assigned in the presumptive inspection

2.42
standard uncertainty
uncertainty of the result of a measurement expressed as a standard deviation

NOTE See reference [5].

2.43
type A evaluation
⟨of uncertainty⟩ method of evaluation of uncertainty by the statistical analysis of a series of observations

EXAMPLE Observations may be e.g. standard deviation, relative standard deviation.

NOTE 1 See references [5] and [6].

NOTE 2 Repeatability and reproducibility are often estimated by carrying out collaborative method performance tests where several laboratories study "identical" samples provided by a central organizer [15].

2.44
type B evaluation
⟨of uncertainty⟩ method of evaluation of uncertainty by means other than the statistical analysis of series of observations e.g. from assumed probability distributions based on experience or other information

NOTE See references [5] and [6].

2.45
uncertainty
⟨of measurement⟩ parameter, associated with the result of a measurement, that characterizes the dispersion of the values that could reasonably be attributed to the measurand

NOTE See reference [6].

2.46
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

EXAMPLE Stipulated conditions may be e.g. the same person or different persons in one laboratory, or different laboratories.

2.47
validation range
range of mean number of particles per analytical portion for which obeyance of validation specifications (particularly linearity) have been acceptably demonstrated

NOTE It is usually expressed as the range of "reliable" colony counts.

3 Arrangement of the document

The first part (clauses 4 to 8) of this Technical Report contains informative material on basic principles, characteristics and limitations of microbiological methods, as well as on general aspects of validation. The second half (clauses 9 to 11) is the actual validation document, containing specifications and recommended procedures for their determination.

Old and new concepts and principles are not completely defined in the body of this Technical Report. Three annexes are attached. Annex A details the statistical formulae most relevant to this document, annex B contains numerical examples and annex C gives detailed plans for two validation experiments.

Statistical tests in the ordinary sense are not central to the ideas. Mathematical calculations are used mainly for the purpose of providing convenient summaries of data and statistical distributions provide guidance values. A table of the χ^2 distribution is the guide most frequently consulted.

The two BASIC programmes given in annex A are easily copied into desk-top computers or programmable pocket calculators to help with the basic calculations most frequently needed.

4 Basic concepts

4.1 General

As far as particle statistics is 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 metrological principles. The same validation principles as with selective colony methods can be applied.

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

4.2 Validation

4.2.1 General

Validation means a process providing evidence that a method is capable of serving its intended purpose: to detect or quantify a specified microbe or microbial group with adequate precision and accuracy. The total count methods do not have a definable target group and can only be validated in relation to other methods or against theoretical expectations of precision.

Validation is classified as primary or secondary according to its purpose.

4.2.2 Primary validation

Primary validation is an exploratory process with the aim of establishing the operational limits and 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 (positive colony, tube or plaque).

Primary validation characteristically proceeds by the use of specially designed test schemes.

A laboratory developing an in-house method or a variant of an existing standard should carry out the steps of primary validation.

It is imperative that technicians involved in primary validation have considerable experience with other microbiological methods.

4.2.3 Secondary validation

Secondary validation (also called verification) takes place when a laboratory proceeds to implement a method developed elsewhere. Secondary validation focuses on gathering evidence that the laboratory is able to meet the specifications established in primary validation. Presently, specifications are not available for most of the methods. Results of external quality assurance (see 4.2.8) may have to be used as the first step towards complete secondary validation.

Typically, secondary validation uses selected and simplified forms of the same procedures used in primary validation, but possibly extended over a longer time. Natural samples are the optimal test materials and the work need only address the procedure within the operational limits set by primary validation.

4.2.4 Analytical quality control (AQC)

Application of valid methods in their specified reliable limits does not automatically ensure valid results. Analytical quality control (AQC) used in connection with daily routine analyses is necessary. It controls the ability to use a method successfully.

AQC is a continuous process. Guidance charts, with limits derived from method specifications (from primary validation) or from theoretical considerations are the principal tools.

The methods of AQC are extensions of the routine analytical process, e.g. replications at different levels, or simply calculations not ordinarily performed on routine data. In addition, reference materials, intercalibrations and spiked samples are used.

Analytical quality control is needed in connection with primary and secondary validation. Only results reliable in the AQC sense should be used for derivation of validation criteria and performance characteristics.

International and national working groups have produced numerous documents on analytical quality control of microbiological methods (e.g. references [1, 2, 3, 4, 7, 8, 20, 21, 24, 26]). Standards manuals also contain sections on that subject. Although vital to validation, the methods of analytical quality control are not detailed in this Technical Report. In everything that follows, it is assumed that laboratories have the appropriate analytical controls and internal and external quality assurance systems in operation.

4.2.5 Equivalent methods

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.

Method performance consists of many aspects. There is no single test of method equivalence, 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 validation tests can be used for method comparisons (examples B.2, B.3, B.4 in annex B). The methods only need to be tested in parallel for recovery comparisons.

A method giving the highest recovery of confirmed target organisms is obviously the best, unless confirmation is always 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 primary validation cannot be corrected by more refined target colony definitions, the method should be deemed invalid.

4.2.6 Test materials

It is a popular notion that validation should simulate routine as much as possible. Natural samples with natural concentrations of microbes should therefore be the main test materials. There are exceptions under some circumstances.

Artificial materials (certified 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 validation excercises.

Spiking may be useful and even necessary in secondary validation whenever it is difficult to find natural samples with target organisms. Laboratory personnel will be able to familiarize themselves with the target.

Negative samples (blanks) should be limited to internal quality assurance. Their inclusion among samples studied for method equivalence may lead to a false impression of a good correlation between methods. If it were possible to know in advance which natural samples contain no target organisms, they would be a suitable selection for testing false positives in actual validation excercises.

The optimal concentration range for the validation of microbiological methods is narrower than the projected application range. High concentrations are unnecessary. Such samples resemble pure cultures and do not put the performance of the method or the laboratory to test.

Samples with very low bacterial content need to be studied for public health reasons, but are ill suited for method comparisons and other validation exercises for statistical reasons. The problem is mostly avoidable, because microbiological methods are generally not concentration-sensitive at the low end of the scale. Each individual germ reacts with the nutrient medium almost independently of other particles in the sample. If a method has a low recovery compared to another, the fact is more readily discovered with twenty or thirty colony-forming particles per plate than with one or a few.

Methods found valid at concentrations sufficient for validation are trusted to work also at low analyte concentrations.

4.2.7 Samples — Representativeness and sufficiency

Statistical theory provides solutions for calculating the number of samples required for different testing or estimation situations [3, 13]. To be able to make use of the theory, the size of real effects of importance and the power for their detection should be defined. An estimate of the uncertainty (precision) of the determination should be available and random sampling should be practiced.

Many or all of the above requirements are difficult to meet in advance planning and execution of microbiological method performance tests. Statistical techniques, if used at all, become rough guidelines.

The number and variety of samples examined ought to be sufficient to be convincing. Without the help of statistics, there are no exact ways of deciding. In some instances, the first sample studied might give the answer that the method is not good enough. Usually, however, more samples are needed. It may take a thousand samples to "prove" that two P/A methods are not equivalent. Choice of too few examples may be a waste of time.

4.2.8 External quality assurance and other collaborative tests

Participation of several laboratories in studies of "homogeneous" material are considered essential tests of both method and laboratory performance. (After outliers have been recognized and deleted, the remaining data are thought to provide the necessary information on method performance and proficiency.)

Collaborative tests have been developed into a tool widely applied for testing precision characteristics of chemical methods [34]. It seems somewhat premature to fully recommend the same in microbiology. It is assumed that all the participating laboratories have several years of experience with the methods tested and a proven ability to use them. The present experience is that collaborative experiments intended for method performance testing tend to turn into laboratory proficiency tests and training excercises.

A number of microbiological methods have been in use for decades (e.g. Endo agar for total coliforms, mFC for thermotolerant coliforms, m-*Enterococcus* agar for intestinal enterococci) by hundreds of laboratories. These methods would therefore theoretically be suitable objects for collaborative method performance testing.

When making collaborative proficiency tests for specific target organisms with selective media, the samples almost necessarily should be spiked with pure cultures or mixtures of organisms. Another solution is to use certified reference materials. This is a simplified and artificial situation. Major difficulties experienced by different laboratories in the routine use of methods on natural samples may be missed. As long as the detailed performance characteristics of a microbiological method have not been expressed quantitatively, these types of external quality assurance (EQA) schemes may nevertheless be the most satisfactory means towards secondary validation (verification) of a method.

4.3 Detectors

4.3.1 General

It is often convenient to call the nutrient medium in its container a detector (2.11). Two types of detector, solid and liquid, are employed in different microbiological method variants. They are also mostly associated with different enumeration or detection principles: liquid with P/A and MPN, and solid with colony counts.

All forms of validation in microbiology focus on the performance of the detectors.

The set of tubes (MPN) or the series of (countable) plates used for analysis is called a detection set or a detector set (2.12). Each individual MPN tube is a P/A detector.

EXAMPLE An individual well of a microtitre plate is a P/A detector. The whole plate when used as an MPN system is the detection set.

4.3.2 Detector comparisons

For most colony-count methods a liquid counterpart with the same chemical composition but without the solid matrix (agar, membrane filter) can be produced. The effect of the solid environment can be evaluated by comparing colony counts with the equivalent MPN estimate provided that the reaction for target recognition is the same on both types of detectors and the number of parallel tubes is large enough for adequate precision. Also the sensitivity of P/A detectors can be evaluated by similar liquid-solid comparisons.

4.4 Performance characteristics

Performance characteristics should be quantifiable and testable to be of use in validation.

The terminology on performance characteristics in this Technical Report mostly follows the chemometric usage. Because the original definitions of the terms do not always fit microbiological methods perfectly, they have been modified and adapted as necessary.

The performance characteristics dealt with in this Technical Report are related to scope (list of situations and sample types where the method is applicable), precision, linearity, recovery, working limits in terms of lowest and highest recommendable colony number per plate, selectivity, specificity and robustness (ruggedness). Definitions of these and other terms can be found in clause 2.

4.5 Specifications

Specifications are either numerical or qualitative expressions of performance characteristics or of working limits derived from them. Primary validation should provide the following:

a) morphological identification of the (presumptive) target;

b) statements regarding incubation conditions (temperature, time, gas atmosphere, moisture) and media characteristics (pH, stability);