TECHNICAL SPECIFICATION

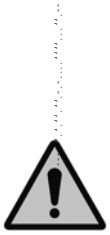
ISO/TS 22117

First edition
2010-11-01

Microbiology of food and animal feeding stuffs — Specific requirements and guidance for proficiency testing by interlaboratory comparison

Microbiologie des aliments — Exigences spécifiques et lignes directrices pour les essais d'aptitude par comparaison interlaboratoires

Reference number
ISO/TS 22117:2010(E)



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

Published in Switzerland

Contents

Page

Foreword

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

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

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

In other circumstances, particularly when there is an urgent market requirement for such documents, a technical committee may decide to publish other types of document:

— an ISO Publicly Available Specification (ISO/PAS) represents an agreement between technical experts in an ISO working group and is accepted for publication if it is approved by more than 50 % of the members of the parent committee casting a vote;

— ISO Technical Specification (ISO/TS) represents an agreement between the members of a technical committee and is accepted for publication if it is approved by 2/3 of the members of the committee casting a vote.

An ISO/PAS or ISO/TS is reviewed after three years in order to decide whether it will be confirmed for a further three years, revised to become an International Standard, or withdrawn. If the ISO/PAS or ISO/TS is confirmed, it is reviewed again after a further three years, at which time it must either be transformed into an International Standard or be withdrawn.

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

ISO/TS 22117 was prepared by Technical Committee ISO/TC 34, *Food products*, Subcommittee SC 9, *Microbiology*.

Introduction

General requirements for organization of proficiency testing (PT) schemes of all types are given through ISO/CASCO (Committee on Conformity Assessment) in ISO/IEC 17043; additionally, general guidance is available from the International Union of Pure and Applied Chemistry (IUPAC, see Reference [9]) and the International Laboratory Accreditation Cooperation (ILAC, see Reference [8]). However, these recommendations may not be directly applicable to all cases and should be interpreted specifically for different laboratory sectors where PT schemes are organized. For this reason, a document is needed to establish the criteria which a provider (and associated collaborators) of PT schemes shall meet in order to be recognized as competent to provide PT schemes for microbiological analysis. This applies particularly to the specific technical requirements necessary to deal with living microorganisms, such as sample homogeneity and stability, as well as with the interpretation of presence/absence (detection) tests which is not covered by an existing document.

Proficiency testing schemes for microbiology laboratories are mainly used to evaluate performance, particularly trueness (bias) and in some cases precision, of food microbiological examinations in specific laboratories.

Additionally, data from such PT schemes can be used:

a) to provide information to the organizations responsible for laboratory acceptance within an official control framework and to allow continuous monitoring;

b) to aid laboratory accreditation in a general framework of quality management;

c) to inform those responsible for quality in the participating laboratories as part of the educative elements of external quality assessment of trueness (bias).

Information from PT schemes may also be used for:

1) identification of the possible sources of errors, particularly the bias component of uncertainty, to improve performance;

2) estimation of measurement uncertainty for enumeration methods (see ISO/TS 19036[6]) and limits of detection for presence/absence methods;

3) demonstration of staff competence to perform a specific microbiological examination;

4) evaluation or validation of a given method by the study of trueness and precision;

5) identification of variability between individual laboratories;

6) assignment of a “target” value for an analyte in a material in order to establish a reference material.

However, these aspects are not specifically covered in this Technical Specification.

Proficiency testing schemes are therefore organized to meet certain criteria and the testing programme (frequency, number of samples, number of repeats, etc.) shall meet the requirements of the type of method used and commodity analysed, to achieve the level of control desired by all parties involved.

Microbiology of food and animal feeding stuffs — Specific requirements and guidance for proficiency testing by interlaboratory comparison

1 Scope

This Technical Specification gives requirements and guidance for the organization of proficiency testing schemes for microbiological examinations of:

a) food and beverages;

b) animal feeding stuffs;

c) food production environments and food handling;

d) primary production stages.

This Technical Specification is also potentially applicable to the microbiological examination of water where water is either used in food production or is regarded as a food in national legislation.

This Technical Specification relates to the technical organization and the implementation of proficiency testing schemes, as well as the statistical treatment of the results of microbiological examinations.

This Technical Specification is designed for use with ISO/IEC 17043 and ISO 13528, and deals only with areas where specific or additional details are necessary for proficiency testing schemes dealing with microbiological analyses for the areas specified in the first paragraph.

2 Normative references

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

ISO 3534-1, *Statistics — Vocabulary and symbols — Part 1: General statistical terms and terms used in probability*

ISO 3534-2, *Statistics — Vocabulary and symbols — Part 2: Applied statistics*

ISO 5725-1, *Accuracy (trueness and precision) of measurement methods and results — Part 1: General principles and definitions*

ISO 5725-5, *Accuracy (trueness and precision) of measurement methods and results — Part 5: Alternative methods for the determination of the precision of a standard measurement method*

ISO 7218, *Microbiology of food and animal feeding stuffs — General requirements and guidance for microbiological examinations*

ISO 13528, *Statistical methods for use in proficiency testing by interlaboratory comparisons*

ISO/IEC 17043:2010, *Conformity assessment — General requirements for proficiency testing*

3 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO 3534-1, ISO 3534-2, ISO 5725-1, ISO 13528, ISO/IEC 17043 and the following apply.

NOTE 1 Some terms used in the text have different meanings in microbiology and statistics, e.g. homogeneity, heterogeneity, test, sample, distribution. The context clarifies whether the terms refer to microbiological test samples or datasets used for statistical analysis.

NOTE 2 Some providers of proficiency testing use the term external quality assessment (EQA) to indicate schemes with broader application to all areas of operation of a laboratory and a particular educational remit. The requirements of this Technical Specification cover those EQA activities that meet the definition of proficiency testing.

3.1
target organism
microorganism which is the designated analyte for a proficiency testing sample

3.2
background flora
microorganisms included in a proficiency testing sample which can be introduced to compete with or mimic the target microorganism

3.3
reference strain
microorganism obtained directly from an official culture collection and defined to at least the genus and species level, catalogued and described according to its characteristics and preferably originating from food or water as applicable

[ISO/TS 11133-1:2009[3], 3.3.2]

3.4
recovery percentage
proportion of the assigned value of the target organism recovered by the participant

NOTE 1 The recovery percentage is calculated by multiplying by 100 the number of recovered colony forming units (cfu) per volume or per mass.

NOTE 2 The recovery percentage can be significantly below 100 % when competitive flora and matrix effects are present in a proficiency testing sample.

4 Scheme design and purpose

4.1 General

General requirements for designing PT schemes are given in ISO/IEC 17043; in this clause, only areas requiring special consideration for microbiological PT schemes are discussed in the context of these general principles.

4.2 Scheme objectives

The primary objective of any PT scheme is to provide information to enable laboratories to have confidence in the reliability of their results.

The detailed requirements for a documented plan of a PT scheme are covered in ISO/IEC 17043:2010, 4.4.1.3, and the plan should also include reference to any relevant legislation. An example of a plan for a typical microbiology food examination scheme is given in Annex A.

The studies required to establish a new PT scheme are extensive and shall be clearly defined in the scheme objectives. These should include, as a minimum, the requirements listed in Clause 5. Requirements for checking individual rounds of testing, including homogeneity and stability testing, should also be established in the scheme design and be appropriate for the scheme objectives.

4.3 Laboratory requirements for schemes

General requirements for appropriate laboratory facilities to handle all aspects of PT schemes are given in ISO/IEC 17043:2010, 4.3.1, and safety requirements are covered in ISO/IEC 17043:2010, 4.6.2.4.

For microbiology schemes, providers shall have a documented policy to bring hazards to the attention of participants and ensure that relevant safety advice is given (see Clause 7). For example, food microbiology laboratories shall have facilities for dealing with microorganisms of risk categories 1, 2, and 3, as appropriate (see ISO 7218:2007, 3.2).

4.4 Choice of test matrices

General requirements to document test matrices in the scheme plan are given in ISO/IEC 17043:2010, 4.4.1.3, and choice of the matrices to reflect routine sample types in ISO/IEC 17043:2010, 4.4.2.3.

The reasons for the choice of matrix type should be stated, e.g. to provide levels of sample stability and homogeneity that are fit for the intended purpose of the scheme.

The description of the test items shall specify the sample matrix (natural or simulated); whether artificially or naturally contaminated; the source and country of origin to comply with international transport regulations; and any method of preservation used, e.g. freeze-dried, air dried.

4.5 Information on test methods used by the PT provider

The general requirements for methods to be used by the PT provider are given in ISO/IEC 17043:2010, 4.4.1.3.

If the scheme is targeted at one or more tests specified in or required by legislation, the routine quality control tests on the scheme samples (e.g. homogeneity and stability) shall be undertaken in accordance with the methods stipulated in that legislation and this shall be stated (ISO/IEC 17043:2010, 4.5.1).

Participants shall be encouraged to use their routine methods but, where they are undertaking tests in accordance with legislation, some degree of guidance shall be given, e.g. reference to ISO methods, legislative texts, or peer-reviewed publications (ISO/IEC 17043:2010, 4.5.1).

4.6 Statistical design

General requirements for statistical design are given in ISO/IEC 17043:2010, 4.4.4.

An outline of the statistical design for PT schemes for microbiology shall indicate that the statistical tests to be used are influenced by the level of homogeneity of the test material which, in turn, is influenced by the random variation in distribution of the microorganisms.

Except for low numbers, a log-normal distribution is usually expected in quantitative testing data and suitable statistical analysis methods shall be used for such data [ISO/IEC 17043:2010, B.3.1.4 d)]. Where low numbers are required in quantitative test items (e.g. water or beverage examination), a random Poisson model is more applicable, as the variation in numbers of organisms between different units of material becomes relatively large and can mask variations in performance.

Sample homogeneity shall normally be such that it does not significantly influence the observed variation between laboratories.

Semi-quantitative enumeration tests and qualitative detection tests require different statistical methods to analyse data and these are discussed further in 8.3 and 8.4.

The scheme plan shall clarify distinctions between performance testing for methods for detection and those for enumeration of target microorganisms.

5 Technical requirements and guidance for sample design and content

5.1 Target organisms level

The target organisms shall be provided at levels suitable to show that examination methods are fit-for-purpose and to reflect levels likely to be found in the sample matrices being tested (ISO/IEC 17043:2010, 4.4.2.3). Where pathogenic bacteria are the target, the levels should also take account of and reflect the levels likely to cause hazard to human health and, if appropriate, any limits specified in microbiological criteria.

NOTE The level causing hazard to human health is not always known with accuracy and depends on the susceptibility of individuals. The aim of examination for pathogens is to prevent illness, and also to detect pathogens at a very low level, before those pathogens can grow to a higher level.

For quantitative (enumeration) methods, the target level shall be appropriate for the levels routinely found in and any statutory specifications applicable to the sample matrices used. The target level should also sometimes be used near the limit of quantification of routine methods to challenge the performance of the participants across the applicable range of the method. However, samples should not be dispatched with organism levels so low that, when using routine dilutions, the expected mean number of organisms in a sample is fewer than 10 colonies per plate.

For qualitative (detection) methods, the target organisms shall generally be required to be at a sufficiently low level to provide a valid challenge to the methodology and to contribute data for validation exercises to establish or verify limits of detection for individual participant laboratories.

5.2 Sources, characterization and traceability of organisms

The characteristics of the target organisms shall be established before use in order reliably to assess performance, especially in schemes where participants are permitted to use different methodologies.

Both typical and atypical strains should be considered and included in the scheme programme to challenge laboratory performance.

Recognized reference strains from international collections should be used where they are most suitable for the scheme purpose; however, laboratory isolates or “wild” strains isolated from the matrices used by PT schemes are useful to reflect routine situations more closely. Where these are used, they should be sufficiently characterized according to the appropriate International Standard reference methods, to ensure that any atypical reactions are apparent to the organizers before use.

In all cases, the organisms used in PT scheme samples should be traceable to the relevant culture collection or to valid characterization data held by the organizers.

Under certain circumstances, it is not possible to use reference cultures or materials from internationally recognized collections or cultured laboratory strains, e.g. for PT schemes for non-cultivable organisms such as human noroviruses. In such circumstances, clinical material can be used to contaminate a test matrix artificially, either through immersion, spraying or, in the case of bivalve shellfish, through bioaccumulation. The method of artificial contamination should be as close to the “natural” route of contamination as possible. Extreme care should be used when manipulating human clinical material, faecal or vomitus samples and these should be screened for additional pathogens before use. Target viruses should be fully characterized to strain level by conventional polymerase chain reaction followed by sequencing.

5.3 Background and competitive flora

The total flora of the samples, either naturally or artificially contaminated, shall be chosen to assess the ability of participants to detect and/or enumerate target organisms in the presence of non-target background flora (typical of the sample matrix) and presumptive target organisms which, without appropriate confirmation tests, can lead to false positive results.

Any strains used to simulate background flora shall meet the requirements of 5.2 for characterization and traceability. In naturally contaminated samples, the effects of any background flora on the target organisms shall be determined.

5.4 Matrix selection and effects

All matrices shall be evaluated before use to check for any effects on the target and background floras, e.g. where the matrices reduce the recovery of spiked organisms.

Participants shall be alerted to the nature of the food matrix where such matrices are known to affect recovery of microorganisms adversely (e.g. those which bind and retain cells, such as fatty materials) or have bactericidal or bacteriostatic properties. Suitable and validated preparation procedures shall be included for the information of participants.

Sample matrices used for microbiology PT schemes are often, but not necessarily, sterilized before use. Where natural, unsterilized samples are distributed, the organizers shall determine the effect of the background microflora of the samples.

6 Sample verification by the provider

6.1 General

General requirements for sample verification are given in ISO/IEC 17043 and ISO 13528 (for information, see also Reference [9]); this clause expands the specific requirements and problems for homogeneity and stability testing in materials containing living microorganisms.

6.2 Sample homogeneity testing — General considerations

(See also ISO/IEC 17043:2010, 4.4.3 and B.5.)

Proficiency tests may involve the preparation of a bulk test material, which is then subdivided into individual portions, as similar as possible to each other, for distribution to participants. Alternatively, test portions may be individually inoculated for distribution.

Whatever preparation method is used, the test material shall be assessed for homogeneity, usually prior to but also at the time of testing for unstable fresh materials.

A homogeneity test should be performed on each batch of samples, based on relevant statistical principles (ISO/IEC 17043:2010, 4.4.3.2 and B.5). Such tests are given in ISO 13528 or, as an alternative, Annex B.

It is also necessary to test for homogeneity if materials are to be stored for longer periods of time to ensure criteria are still met before use. The number of samples to be tested from each batch should also be sufficient to obtain ongoing information on the homogeneity of the batch; 10 samples (tested in duplicate if required) is suggested.

A test material which is less than sufficiently homogenous may still be used in a proficiency test round (ISO/IEC 17043:2010, 4.4.3.1 Note 3), provided suitable statistical principles are used to take account of the greater variance between samples (see ISO 13528). A statistical plan for heterogeneous materials, including replicate analysis of several samples (see ISO 5725-5), should be used to minimize the effects of lack of homogeneity on the evaluation of participant performance.

6.3 Homogeneity testing for quantitative (enumeration) samples

General requirements and procedures for testing homogeneity of quantitative proficiency test materials are given in ISO/IEC 17043:2010, 4.4.3 and B.5 and ISO 13528.

Materials that show between-unit variation large enough to affect the assessment of laboratory performance significantly should not be used in interlaboratory studies, unless special requirements and methods of data analysis apply, e.g. low numbers of microorganisms in drinking water and other samples.

The criterion for "sufficiently homogenous" is defined by the requirements of the interlaboratory comparison (see ISO 13528 and Annex B). However, in general, a material for which the between-unit standard deviation (on the appropriately transformed scale) is $\leqslant 0{,}3\sigma_p$, where σ_p is the target standard deviation used to assess the performance of laboratories, is considered sufficiently homogenous (see ISO 13528).

Any alternative homogeneity test should meet the following criteria (reproduced from Reference [9]):

a) the probability of rejecting a sufficiently homogenous test material should be $\leqslant$ 5 %;

b) the probability of rejecting a test material where between-unit variation is $1{,}5\sigma_p$, in which σ_p is the acceptable between-laboratory variation (expressed as a target standard deviation), is $\geqslant$ 80 %.

An 80 % probability of rejecting a material where between-unit variation is 1,5 σ_p is based on simulation studies of the duplicate analysis of 10 test units using a method with an analytical standard deviation of $0{,}5\sigma_p$ (i.e. 0,125 log units) and a critical value for T_2 (see Annex B) that meets criterion a) in the previous paragraph. It represents what is achievable with a reasonable amount of analytical effort.

Homogeneity tests are based upon estimations of between-unit variance and analytical (repeatability) variance obtained under repeatability conditions. Suitable methods of testing for such variation are given in Annex B.

The analytical (repeatability) variance should be estimated from replicate analyses of the initial suspensions obtained from test portions (References [15][16]). This analytical variance can also be calculated from the number of counted colonies and the precision of analytical materials in use (Reference [10]).

In microbiology, the between-unit variance shall be estimated under repeatability conditions (in one run). If that is impossible, the between-unit variance includes the within-laboratory reproducibility and can perhaps lead to the false rejection of a satisfactory material.

When the number of counted colonies is sufficiently high (more than 35 to 40 colonies per plate), the analytical standard deviation, σ_{an}, generally satisfies

$$\frac{\sigma_{an}}{\sigma_p} < 0{,}5$$

where σ_p is the target standard deviation, and the test for sufficient homogeneity proposed in Reference [11] should be used (see Annex B). If the number of counted colonies is low (fewer than 35 to 40 colonies per plate), the $T_1 - T_2$ test is recommended (see Annex B).

When replicated test units are provided to participating laboratories, the between-unit variability obtained by participants should be examined by the provider to assess the homogeneity of the material. Although this variability includes within-laboratory reproducibility, the higher number of participating laboratories increases the statistical power of the analysis and can be a good indicator for successive rounds.

When the number of counted colonies is low (say fewer than 20), the analytical (repeatability) variance is high. In that case, the provider should recommend that participating laboratories replicate enumerations of test portions to satisfy the condition (see ISO 13528):

$$\frac{\sigma_r}{\sqrt{n}} \leqslant 0,3\ \sigma_p$$

where

σ_r is the repeatability standard deviation;

σ_p is the target standard deviation;

n is the number of replicates.

If that is impossible, the laboratory performance should be assessed cautiously.

Where the PT material contains low numbers of cfu (say fewer than 20), the between unit (i.e. between replicate samples) variation shall be measured to demonstrate that it does not exceed random (Poisson) variation in order to provide a meaningful assessment. The index of dispersion test on n samples (where n is a minimum of 10) should not exceed the χ^2_{n-1} value at 0,05 probability.

6.4 Homogeneity testing for qualitative methods

Similar principles apply to homogeneity testing for qualitative (presence/absence) methods, but special consideration is required where the level of spike is low (see 8.4).

The homogeneity of qualitative samples may be tested by enumerating the spike. It may also be possible to determine the contamination level by using a most probable number (MPN) technique. If enumeration is possible, the homogeneity tests detailed in 6.3 may be used, depending on how the samples are artificially contaminated and on how the spike is enumerated.

6.5 Stability testing by the provider

6.5.1 General

Samples to be used for proficiency testing shall be stable, at least for the period from preparation (by the provider) to the date of the study or the end of the time period allowed (see ISO/IEC 17043:2010, 4.4.3).

If the same type of sample with the same test strain(s) is always used, it is sufficient only to perform a verification (e.g. by checking after preparation and at the date of the study) on subsequent samples. The provider should also examine results obtained by the participating laboratories to check the stability of the material during the period allowed for the study.

6.5.2 Stability during storage conditions

For stable samples which are always stored at low temperatures (e.g. −70 °C, −20 °C, +5 °C), the stability shall be determined by checking the level of the analyte and the homogeneity of each batch at regular intervals during storage. The minimum period for stability testing should be the time between preparation of the materials and the specified date or time period of analysis.

Frequency of testing depends on the information already available for the batch of samples and the total period of time over which stability information is required. If the total storage time is, for instance, only two weeks, it may be necessary to test every two days, but if storage is required for one year, it may be sufficient to test monthly. For large batches of samples, a minimum of three samples should be tested on each occasion in order to show ongoing stability across the whole batch (ISO 13528:2005, Annex B). Whatever frequency of testing is used, this shall be justified and validated as acceptable by the scheme organizers.

6.5.3 Stability during transport conditions

In addition to information on stability during storage conditions gathered during validation studies for a new scheme, it is also important to test the effect of "abuse conditions" on the samples, e.g. long transport times at elevated temperatures (see ISO/IEC 17043:2010, 4.6.3.2).

A stability test using different temperatures to reflect "worst case" transport conditions and maximum expected delays, shall be performed initially to test the effect of such abuse on test samples. For example, samples of one batch are stored at the specified storage temperature (e.g. −20 °C), but also at +5 °C, +15 °C and +25 °C. Every day, five samples held at each storage temperature are analysed for a total period of one or two weeks.

The design of such stability experiments is variable but should be appropriate to obtain information on the effect of different storage temperatures on the samples and to establish any upper temperature limit for receipt by the laboratory. The information obtained can be used to choose the optimal distribution conditions for the scheme samples, e.g. whether it is necessary to cool the samples during transport using dry ice or ice packs, or whether ambient distribution is acceptable.

For temperature-critical distributions requiring controlled refrigeration with strict sample acceptance criteria, such as schemes in support of legislative testing for *E.coli* in bivalve shellfish, inclusion of individual temperature loggers in each sample box to record the sample temperature in transit is recommended. Checking the temperature data may enable the scheme provider to explain anomalous results returned by participants.

7 Sample handling

7.1 General

Requirements for general sample handling are detailed in ISO/IEC 17043:2010, 4.6.1 and 4.6.2, and only additional information relevant to microbiological samples is given in this clause.

7.2 Instructions to participants

For each study, each participating laboratory shall receive a clear set of instructions covering:

a) storage conditions for samples of all types, particularly information on the storage temperature, which should also appear on the outside of the transport packaging;

b) maximum temperature of the samples on receipt at the participant laboratory, if appropriate;

c) instructions on how to handle the samples — if reconstitution, dilution or other processing of the samples is required, this should be described clearly for each set and type of samples;

d) appropriate safety data sheets, which should be included with each distribution (an example of the detail required is given in Annex D);

e) other supplementary instructions, such as:

   1) the latest dates when the examinations should be performed and results sent to the organizers,

   2) the method(s) of examination (prescribed or participant choice, as required),

   3) how to report the results to the organizers, particularly the units of measurement,

   4) the organizers may request details on materials, methods, incubation conditions, etc., in report proformas — if this is the case, instructions for completing such proformas should be provided.

8 Performance evaluations

8.1 General

Wherever possible, the statistical principles used to evaluate performance in PT schemes should be based on those given in standards, such as ISO/IEC 17043:2010, 4.7 and ISO 13528 (for information, see also Reference [9]), although microbiology PT schemes may adopt procedures which differ from those commonly used in other sectors if they are appropriate to their particular schemes.

8.2 Preliminary considerations

Proficiency testing involves the regular distribution of test materials to participating laboratories for them to examine the test materials for specific measurands (in most cases microorganisms). The results of examination are then compared against those of other participants. Proficiency testing therefore provides an independent means of testing and comparing individual performance.

Ongoing satisfactory performance in proficiency test rounds can provide reassurance to participants of their laboratory processes, including methods of examination, analyst training, equipment, reagents, quality control procedures, interpretation of results, and reporting techniques.

However, unsatisfactory results imply that the performance of the individual participant was weak compared with the consensus value and this can raise a number of questions. These include: were the test samples all the same, how did other participants perform, could the methodology have biased results and have the results been interpreted correctly? The scheme organizers should therefore use suitable systems of performance evaluation to aid participants in answering such questions.

There are many different ways to interpret data from proficiency testing but methods for interpretation shall be objective.

The majority of participants taking part in microbiological PT schemes are not familiar with statistics and shall have confidence that the procedures used by the PT scheme organizers are sound.

The following considerations about the statistical principles applied shall be addressed:

a) validity;

b) explanatory information for participants;

c) reasons for selection;

d) consistency.

Clear and unbiased information shall be provided to participants to allow self-assessment and interpretation of results and to maximize the benefit from participation.

8.3 Quantitative methods

8.3.1 General

The statistical design for PT schemes for microbiological enumeration methods is influenced by the level of homogeneity of the test material, which in turn is influenced by the random variation in distribution of organisms. In addition, there are likely to be substantial differences between the participants in the precision required or expected of a test.

The choice of statistical method shall take into account the factors outlined in 8.1 and 8.2, together with other considerations such as the number of participants in a particular scheme. Parameters that contribute to deciding which statistical tests are to be used for analysing results should be stated. For example, specialized schemes may have fewer than 30 participants; results from 30 participants may be analysed in a different way from larger schemes with, for example, more than 100 participants.

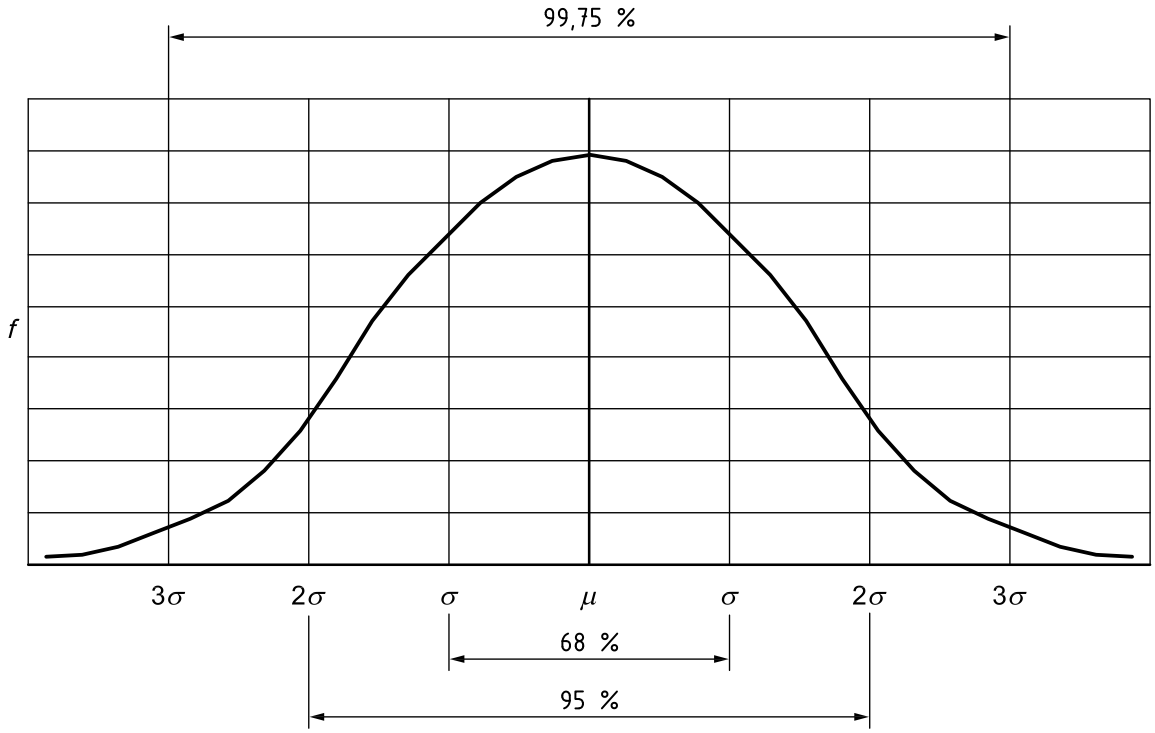
The chosen method of statistical analysis shall be appropriate for not only the number of participants undertaking an examination but also the method used. For example, results for an examination using colony count methods are analysed in a different way to those where determination of the MPN is required. This is because the inherent variation in the MPN method tends to be greater than that for colony count methods. Indeed, different statistical parameters may be needed to assess participants' performance for different tests within a single scheme (see ISO/IEC 17043:2010, 4.5.2). This shall be stated in the scheme design documents and generally accepted microbiological criteria shall be referenced in the scheme plan.

Suitable statistical analyses and allocation of scores to participants are covered in ISO/IEC 17043:2010, Annex B.

8.3.2 Distribution of data

When an enumeration test is repeated several times, under repeatability conditions, the frequency distribution of the results would be expected to form a bell-shaped curve, called a normal distribution.

Microbiological counts usually follow a log-normal distribution and the data are converted to logarithm to the base 10 values to produce a normal distribution curve. However, for low bacterial numbers, actual counts may be used, or a square-root transformation may be applied.



Key

f frequency

μ mean

σ standard deviation

Figure 1 — Diagram of a normal distribution

Although it is the same material under test, the results are not all identical, as numerous small, independent variations are expected to occur during the different manipulations involved in performing the tests.

In proficiency testing, the tests are not usually performed under repeatability, or even reproducibility conditions. Tests are performed by different analysts, at different test locations and at different times, using a variety of equipment, media, reagents and analytical methods and this results in further variability which could be described as “super-reproducibility” or “over-reproducibility” conditions because the term reproducibility is usually used in the context of a single method only.

Despite such variations in test conditions, a (log-)normal distribution of the results is usually observed and the principles of statistics appropriate to (log-)normal distributions should be used to interpret the data, provided the distribution is roughly symmetrical and unimodal.

If the distribution of data does not appear (log-)normal, the possible reasons should be assessed and the data interpreted accordingly using other suitable tests.

8.3.3 Determining the assigned value

The purpose of proficiency testing is to assess how proficient participants are in achieving the “correct” result. However, in many proficiency testing schemes, it is not possible to know the “correct” result, as numerous analysts may all examine the same test material and all return a slightly different result.

Instead, an estimate of the “correct” or “true” result should be made and this is referred to as the “assigned value”. The assigned value may be determined in a number of ways, as described in ISO/IEC 17043:2010, B.2 and ISO 13528.

The most usual method for microbiological PT schemes is consensus values from participant laboratories. The assigned value is determined from the robust mean or the median of the results of all participants. The use of robust methods is intended to minimize the influence of outliers, so that such results need not be excluded from the data because “true” outliers may be difficult to identify.

The assigned value has a higher uncertainty than with other methods, but this is taken into account when assessing performance. The assigned value is deemed to be fair because all participants' results contribute to the calculation.

If a low overall median is produced where assigned values are set from participant consensus (e.g. because a large number of participants had difficulty isolating or identifying a particular organism), the scheme organizers should comment accordingly, so that the performance of participants whose results were not affected is correctly judged.

8.3.4 Uncertainty of the assigned value

The assigned value represents the best estimate of the true value. It has a standard uncertainty indicating the level of confidence in this estimate. If the standard uncertainty of the assigned value is too large in comparison to the standard deviation of the test round, then some participants receive action and warning evaluations, not because of their performance but instead due to the large uncertainty in the assigned value.

Criteria for the acceptability of an assigned value in terms of its uncertainty should therefore be established. A number of methods for estimating this uncertainty and determining acceptability are available and are described in ISO 13528.

8.3.5 Methods of assessing performance

The organizer of the PT scheme shall determine the assigned value and assess by how much the result(s) from each individual participant deviate(s) from that assigned value compared with the results from all other participants. Thus participant performance is judged not against the “correct” result, but against statistical estimates of the “correct” result derived from all of the submitted data. The larger the number of data, the more accurate such statistical estimates are likely to be.

Common methods of assessing performance and allocating scores are detailed in ISO/IEC 17043:2010, B.3 and ISO 13528.

8.3.6 Using *z*-scores

8.3.6.1 General

A common and widely accepted method used in proficiency testing is the *z*-score system, as this is relatively easy to calculate and interpret (see ISO/IEC 17043:2010, B.3). The *z*-score indicates how many standard deviations away from the mean a given value lies, e.g. a *z*-score of 2 represents a value which is 2σ, where σ is standard deviation, from the mean.

As depicted in Figure 1, data with a standardized normal distribution have 95 % of values within 2 σ of the mean and 99,7 % of values within 3 σ. Results with a *z*-score greater than 2 are therefore considered questionable because only 5 % of correct measurements are expected to be that different from the assigned value. Results with a *z*-score over 3 are considered unsatisfactory because only 0,3 % of correct measurements are expected to be that different from the assigned value (see ISO/IEC 17043:2010, B.4).

8.3.6.2 The target standard deviation for *z*-score calculations

The *z*-score calculation uses a target value for standard deviation. This target standard deviation defines the scale of acceptable variation among laboratories for each particular test. The same target standard deviation should be used over successive rounds of the proficiency test so that scores may be compared from round to round. There are a number of methods for establishing the target standard deviation, detailed in Reference [9], ISO/IEC 17043:2010, B.3 and ISO 13528.

8.3.6.3 Multiple results in *z*-score systems

Differences between participants' results arise from between-laboratory variation and also from within-laboratory variation. Within-laboratory variation or intra-laboratory variance is the variation between measurements made by the same laboratory on the same sample and is an inherent feature of microbiological examinations. Within-laboratory variation is measured by the repeatability variance.

When a participant laboratory reports multiple results for a single test material, this may potentially bias the remaining data from other participants. For example, if 10 analysts from the same laboratory all tested the same sample, which had been incorrectly diluted initially, all 10 results would be incorrect. This number of incorrect results becomes a subset within the bulk data, and may bias all the other results. To avoid such bias, only one reported result per laboratory should be included in the overall analysis of data for a distribution.

When a participant laboratory has obtained multiple results from a single PT sample, these results shall be reported separately and not as a mean value. Where only one result per participant is permitted by the scheme organizers, the individual result to be reported shall be chosen before the examination is undertaken and results are known.

8.3.7 Other methods of performance evaluation

8.3.7.1 General

Although *z*-scores are very commonly used for evaluation of PT scheme results from enumeration methods, other methods of scoring may be appropriate for particular schemes.

For example, with samples where low counts are sought (such as drinking water) the statistical assessment can be based on a model which predicts random variation. Thus “low” or “high” tail-end counts are defined using the Poisson formula. Low or high results can occur occasionally by chance in any laboratory, but an accumulation of tail-end results indicates poor performance. The advantages and disadvantages of a model versus a percentile approach to statistical assessment are discussed in Reference [17].

With schemes where high counts are expected, the 0,5 $\log_{10}$ rule or the percentile approach can be used, but the median absolute deviation method is required for schemes with low numbers of participants. These are briefly outlined in 8.3.7.2 to 8.3.7.4.

8.3.7.2 Using the 0,5 $\log_{10}$ rule

A scoring system based on the 0,5 $\log_{10}$ rule can be used for colony counts (adapted from Reference [13]). In summary, the 95 % confidence intervals around a mean colony count are generally not more than $\pm$0,5 $\log_{10}$ cfu. Internal quality control procedures for microbiology laboratories commonly require replicate counts to show agreement to not more than 0,5 $\log_{10}$ units to demonstrate good control. This is applied to participants' results such that all results within $\pm$0,5 $\log_{10}$ units of the participants' median are considered as acceptable and are allocated the maximum score. This rule allows participants' scores to improve over time if the overall quality of participants' enumerations also improves.

The 0,5 $\log_{10}$ rule is based on microbiological criteria but is also statistically valid because if the expected count on a plate is 10 colonies, and organisms are randomly distributed, then 95 % of results should show between 3 and 17 colonies. On a decimal logarithm scale, the expected median count is 1, the lower limit is 0,47 and the upper limit is 1,23, i.e. the lower and upper limits are within 0,5 $\log_{10}$ units. Therefore, a result of within $\pm$0,5 $\log_{10}$ units of the expected value should be deemed acceptable.

8.3.7.3 Using percentiles

Percentiles can be used to identify outlying counts for enumerations when $\geqslant$ 50 participants undertake an enumeration using a colony count procedure. This entails calculation of the 5th, 10th, 90th, and 95th percentiles of the distribution of participants' results (C5, C10, C90, and C95 respectively). C5 and C10 should be rounded down to the nearest 0,05 $\log_{10}$ unit (e.g. 2,23 rounds to 2,20), whereas C90 and C95 should be rounded up (e.g. 3,36 rounds to 3,40). An example of how scores might be allocated (e.g. for aerobic colony counts) is outlined below:

— Results between C10 and C90 (acceptable range): Score = 2

— Results between C5 and C10 or C90 and C95: Score = 1

— Results below C5 or above C95: Score = 0

Application of the 0,5 $\log_{10}$ rule may extend the acceptable range and therefore upgrade the scores allocated for some results in C5, C10, C90 and C95.

The percentile method is robust and does not depend on the actual distribution of decimal logarithm counts being normal. It enables a clear interpretation of the performance assessment.

If the number of participants in an established PT scheme falls to fewer than 50 laboratories or the number of participants reporting enumeration results for a particular parameter falls to less than 50, then median absolute deviation (MAD) values (see 8.3.7.4) should be used instead of percentiles.

8.3.7.4 Using median absolute deviation from the median values

The MAD method is used to identify outlying counts when fewer than 50 participants undertake an enumeration. Percentiles should not be used because fewer than 50 results provide insufficient data to calculate valid values for C5, C10, C90 and C95. MAD values provide a robust method for calculating the acceptable range when assessing participants' results and allocating scores. The analysis requires calculation of the median difference from the median for every result which is then multiplied by 1,482 6 to get a robust estimate of the standard deviation (MAD value), σ_{MAD}.

An example of how scores might be allocated using MAD values is outlined below:

— Results within participants' median $\pm 2\ \sigma_{MAD}$: Score = 2

— Results between $\pm 2\sigma_{MAD}$ and $\pm 2{,}58\ \sigma_{MAD}$: Score = 1

— Results outside $\pm 2{,}58\sigma_{MAD}$: Score = 0

As with percentiles, the lower limits should be rounded down to the nearest 0,05 $\log_{10}$ value and the upper limits up to the nearest 0,05 $\log_{10}$ value.

Note that, unless the 0,5 $\log_{10}$ rule is applied, approximately 5 % of laboratories should be outside the $\pm 2\ \sigma_{MAD}$ range and 1 % outside the $\pm 2{,}58\ \sigma_{MAD}$ range (assuming normality of the decimal logarithm counts without extreme outliers).

The MAD method should be used for new PT schemes where there are fewer than 100 participants.

8.3.7.5 Special considerations for most probable number methods

The test method used to determine an MPN value has greater inherent variability than colony count methods and is therefore often regarded as only semi-quantitative. However, it is sometimes required for the detection and estimation of levels when low levels of microorganisms are expected, especially when the microorganisms may be stressed (e.g. as a result of processing or freezing). Also, MPN methods are stipulated in legislation or criteria for the microbiological examination of certain products such as dairy products and live bivalve molluscs and other shellfish.

Any method for assessing participants' results for determining MPN values should allow for the inherent variability of the MPN, and assume that the sample is well mixed prior to testing.

For the three-by-five tube method, the standard deviation of a $\log_{10}$ MPN result is approximately 0,24, provided results do not show “extreme” tube combinations, e.g. tube combinations of 3, 0, 0 to 5, 5, and 2 (see ISO 7218).

For the three-by-three tube method, the standard deviation of the $\log_{10}$ MPN result is approximately 0,32, provided results do not show “extreme” tube combinations, e.g. tube combinations of 2, 0, 0 to 3, 3, and 1.

This means that in a perfect situation, with no excess between-laboratory variability, 95 % of results should be within $\pm 2\sigma$ and more than 99 % within $\pm 3\sigma$, where σ is standard deviation.

However, in practice there is some between-laboratory variability. Analysing a number of sets of data has shown this to inflate the variance by about 2,5-fold (and hence the standard deviation by about 1,58-fold).

Therefore the limits of acceptability for participants' results for MPN determinations should be raised to $\pm 3\sigma$ and $\pm 5\sigma$ (see Table 1).

Table 1 — Limits of acceptability

Limit of acceptability	Three-by-three method	Three-by-five method
$\pm 3\ \sigma$	$\pm 0{,}96\ \log_{10}$ (9,1-fold)	$\pm 0{,}72\ \log_{10}$ (5,2-fold)
$\pm 5\ \sigma$	$\pm 1{,}60\ \log_{10}$ (39,8-fold)	$\pm 1{,}20\ \log_{10}$ (15,8-fold)

The 0,5 $\log_{10}$ rule should never need to be applied to MPN results due to the inherent method variability.

It is also possible, for the MPN method, to check that the tube combinations and dilutions reported are consistent with the MPN reported using tables (see ISO 7218).

If the PT scheme or the legislation on which it is based requires MPN values to be determined in duplicate, the results can be compared and if the tube combinations are credible, then the difference between the two results should not differ, in terms of decimal logarithm units, by more than $2{,}58 \times \sqrt{2} \times 0{,}24 = 0{,}88$ for the three-by-five tube method and $2{,}58 \times \sqrt{2} \times 0{,}32 = 1{,}17$ for the three-by-three tube method.

If two distributions with two replicates per distribution are compared, then the mean of the two should not differ, in terms of decimal logarithm units, by more than $2{,}58 \times 0{,}24 = 0{,}62$ for the three-by-five tube method and $2{,}58 \times 0{,}32 = 0{,}83$ for the three-by-three tube method.

8.3.8 Long-term performance assessment

8.3.8.1 General

Performance assessment in proficiency testing schemes is generally confined to assessment of results from single rounds, but there are instances where assessment in the longer term may be beneficial. Whilst this generally applies to external quality assessment schemes, some guidance is given for the sake of completeness.

Any method of assessing long-term performance shall ensure that laboratories undertaking enumeration examinations are not identified as “poor performers” by chance due to the number of organisms in the samples they receive.

“Low” and “high” counts should be defined according to objective rules (e.g. Poisson model-based definition for low counts, percentiles or other methods for high counts), then used to determine those laboratories reporting such results more frequently over time than could be expected by chance.

Scheme organizers should encourage participants to exercise their own professional judgment to assess information supplied in reports, thereby self-assessing their performance. Organizers may suggest to their participants various ways of undertaking laboratory self-assessment but they should not dictate criteria for self-assessment by an individual laboratory; rather these should be set by each individual participating laboratory, based on what it believes to be microbiologically significant in the context of its own routine work and client requirements.

8.3.8.2 Low count assessments

If chance is the only factor involved, the “tail-end” counts should be distributed at random. The results may be scrutinized to determine the scatter of tail-end results, between laboratories over a series of samples (e.g. using Cochran's Q-test).

If they are not distributed at random, a second stage analysis may be performed to determine the expected distribution of tail-end counts, amongst those laboratories reporting them, if they were simply due to natural variation between samples and not to laboratory performance effect. Then, contrasting those expected numbers of laboratories with the number actually observed highlights those laboratories that may have experienced problems. An example of performance assessment for samples containing low numbers is given in Table 2.

Table 2 — Observed and expected numbers of sets of results assuming random distribution of low results (from a distribution of low levels of *Clostridium perfringens* in drinking water samples, where variation in numbers between test items may necessarily exceed laboratory performance variation)

No. of “lows”	Observed	Expected
0	58	58
1	32	48,56
2	18	16,94
3	14	3,15
4	3	0,33
5	2	0,02
Total	127	127

One or two low results could have been due to chance; three were possibly not due to chance; four or more were unlikely to be due to chance and procedures should be checked.

8.3.8.3 High count assessments

Assessment of higher counts relies on similar principles, but less variation in participants' results due to variation in sample content is expected.

Long-term performance with, for example, the percentile method of evaluation, can be assessed over 12 samples, allowing a maximum score of 24 points. Participants are set a performance target of at least 70 %. If the 0,5 $\log_{10}$ rule is discounted and an assumption is made that all participants are capable of delivering equivalent performance, then using multinomial theory, it can be shown that the probability of a participant obtaining a cumulative score that is less than 70 % of the maximum possible score is 5,2 %. This means that under these circumstances, approximately 1 in 20 participants may be identified incorrectly as “poor performers”. In reality performance is not equivalent and some laboratories do experience difficulties with the examinations that they undertake, so the probability of a satisfactory performance being incorrectly identified as “poor” is much lower than this. Furthermore, the use of the 0,5 $\log_{10}$ rule reduces this probability to less than 0,1 %.

A practical example of long-term performance assessment using spreadsheets is shown in Annex C.

8.4 Assessment of qualitative methods

8.4.1 General

For interlaboratory comparison studies in which one or more qualitative methods are used, the results are in fact black or white, yes or no, detected or not detected.

Methods of statistical analysis for this type of result are limited, but various options have been proposed, such as LOD_{50}, accordance or concordance assessments and percentage accuracy, and the optimal approach is still under consideration.

8.4.2 Performance of individual laboratories

A simple method for self-assessment by participant laboratories is to record the number of positive and negative results they have found, together with the number of positive and negative results which were expected. This information should be linked with records of the levels of target organisms in the samples to assess performance and also provide ongoing data on the limit of detection of the method in individual laboratories.

As the result of a presence/absence test is a yes or no, many samples need to be tested in order to assess the performance of the individual laboratories in a PT scheme. Each participant should test at least 18 samples in total. These 18 samples consist of six replicates of three different levels of contamination of the samples. The three levels are: negative (check on the occurrence of false positive results due, for example, to cross contamination); low level [meaning samples contaminated at or slightly above the detection limit for the method used, which ideally should be at the level where 50 % of the samples are found positive; and 50 % negative (LOD_{50})] and high level (this level should be 10 times higher than the low level and represents the level at which all samples tested should be found positive).

The interpretation of the data is simple:

a) for the negatives: all samples should be found negative;

b) for the high level: all the samples should be found positive;

c) for the low level: it can be calculated (see Table 3) using the binomial distribution and the percentage of samples found positive (can be obtained from a reference value from the organizer or as a best estimate from the results of all participants) at a 95 % confidence level.

Table 3 — Chance of finding a certain number of positives out of six samples tested as a function of the average percentage of positive samples (binomial distribution)

Number of positives out of six samples	Average percentage of positives								
	10 %	20 %	30 %	40 %	50 %	60 %	70 %	80 %	90 %
0	53,1 %	26,2 %	11,8 %	4,7 %	1,6 %	0,4 %	0,1 %	0,0 %	0,0 %
1	35,4 %	39,3 %	30,3 %	18,7 %	9,4 %	3,7 %	1,0 %	0,2 %	0,0 %
2	9,8 %	24,6 %	32,4 %	31,1 %	23,4 %	13,8 %	6,0 %	1,5 %	0,1 %
3	1,5 %	8,2 %	18,5 %	27,6 %	31,3 %	27,6 %	18,5 %	8,2 %	1,5 %
4	0,1 %	1,5 %	6,0 %	13,8 %	23,4 %	31,1 %	32,4 %	24,6 %	9,8 %
5	0,0 %	0,2 %	1,0 %	3,7 %	9,4 %	18,7 %	30,3 %	39,3 %	35,4 %
6	0,0 %	0,0 %	0,1 %	0,4 %	1,6 %	4,7 %	11,8 %	26,2 %	53,1 %

EXAMPLE To use Table 3, first the average percentage of positives has to be known. Here it is assumed to be 30 %, meaning that about one out of three samples contains the target organism. As only 30 % of the samples contain the target organism, it is likely that some of the samples might not contain the target when, for example, six samples are tested.

Table 3 shows that the chance of obtaining a set of six samples consisting of four positives and two negatives is 6 %. This is not likely to happen, but is still acceptable when a level of confidence of 95 % is assumed. The sum of the chances of finding 0, 1, 2, 3 or 4 positives is 99,0 % (11,8 + 30,3 + 32,4 + 18,5 + 6,0). In this case, 99 % is the smallest value, which is above 95 % (the confidence limit). Leaving out four positives results in 93 %, which is below the 95 % confidence level.

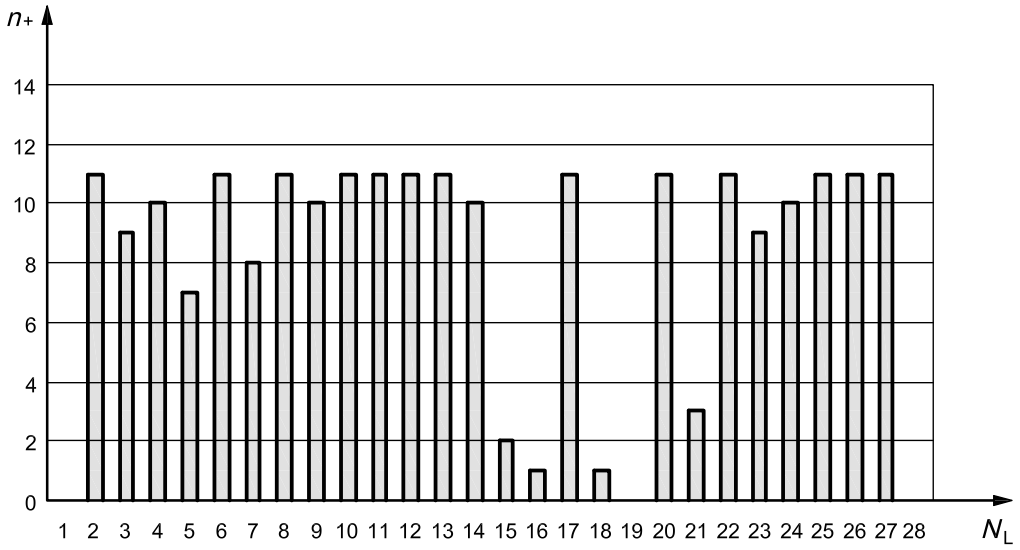
Another reasoning is that the chance of finding six out of six positives is only 0,1 %. This is very unlikely to happen purely by chance. As the uncertainty is set at a maximum of 5 % (100 % – 95 % confidence level), this falls within this limit. The same occurs for the situation when five or six samples out of a total of six are found positive. The sum of these chances is 1,1 % which is still below the 5 % limit. Only when the situation where four out of six are positives (6 %) is added does the uncertainty exceed the 5 % limit. As the maximum is set at 5 %, the situation of five or six positives out of a total of six tested is regarded as an unexpected result.

NOTE At LOD_{50}, 50 % of the samples are found positive. In theory [taking into account that the method used is capable of detecting a single microorganism in a sample and assuming a homogeneous (Poisson) distribution between the samples] the average contamination level of the samples is expected to be 0,7 microorganisms per sample in order to reach the LOD_{50}. In practice, the ideal homogeneous distribution is often not reached and relatively more of the samples do not contain any (viable) microorganism than could be expected for a true homogeneous distribution. In order to obtain 50 % of positive samples, an increase in the average level of contamination is required, based on experience with the material used in the study.

8.4.3 Scheme comparisons of laboratory performance

To compare the performance of one laboratory against other participating laboratories, the scheme organizers may calculate the numbers (or percentages) of positives for the levels contaminated with the target organism found per test sample by each laboratory (reported by laboratory code). An example of such data is given in Figure 2. In this study 28 laboratories participated (indicated as lab codes on the N_L-axis of Figure 2). Each laboratory analysed 22 chicken faeces samples, artificially contaminated with reference materials containing two different *Salmonella* serovars at four different contamination levels (varying from 10 to 500 cfu per sample). In Figure 2 the results are summarized for the sample with the low level of contamination ($n = 14$). The number of positives found in the participating laboratories varied from 0 to 11 samples (n_+-axis in Figure 2).

In addition to this more descriptive way of presenting the results, it is possible to calculate specificity rates, sensitivity rates and accuracy rates per level of contamination of the samples (ISO 16140[4]). These rates may be calculated for each laboratory and for the results from all laboratories.



Key

N_L lab code

n_+ No. of positives

Figure 2 — Number of positive isolations per laboratory code for all tested low-level materials ($n = 14$)

The specificity rate, r_{SP}, is given by:

$$r_{SP} = \frac{n_-}{E(n_{-\,tot})} \times 100\ \%$$

where

n_- is the number of negative results found;

$E(n_{-\,tot})$ is the total number of expected negative samples.

The sensitivity rate, r_{SE}, is given by:

$$r_{SE} = \frac{n_+}{E(n_{+\,tot})} \times 100\ \%$$

where

n_+ is the number of positive results found;

$E(n_{+\,tot})$ is the total number of expected positive samples.

The accuracy rate, r_{AC}, is given by:

$$r_{AC} = \frac{n_+ + n_-}{n_{tot}} \times 100\ \%$$

where n_{tot} is the total number of samples.

This assessment is only meaningful if linked with the number of target organisms present.

Annex A
(informative)

Example of details to be included in a PT scheme plan

PT scheme plan — Summary of scheme

Scheme name:	Food examinations scheme
Scheme type:	Food microbiology
Aims:	To provide external quality assessment samples for general routine examinations undertaken by food microbiology laboratories
Criteria for selection of participants:	Food microbiology laboratories with laboratory facilities adequate for dealing with pathogenic microorganisms of risk categories 1 and 2
Target participants:	Food microbiology laboratories in the private and public sectors
Legislation:	EU Regulation 882/2004 concerning the official control of foodstuffs
Sample type:	Freeze-dried microorganisms in evacuated glass vials
Examinations:	**Presence/absence:** *Campylobacter* spp. *Escherichia coli* O157 *Salmonella* spp. **Enumerations:** Aerobic colony count *Bacillus cereus* *Clostridium perfringens* Coliforms Enterobacteriaceae *Escherichia coli* *Listeria monocytogenes* Coagulase positive staphylococci
Criteria for sample content:	Realistic microflora simulating that of real foods and providing a realistic challenge to routine food microbiology procedures
Target number of participants:	More than 200
Number of distributions per year:	Six
Number of samples per distribution:	Two
External subcontractors:	None

PT scheme plan — Summary of scheme *(continued)*

Technical experts:	Named individuals
Quality control testing of sample batch:	Name of provider laboratory and standard methods used 25 samples from every batch examined for all tests specified
Statistical methods:	Consensus median (enumerations) Percentiles to identify outliers
Allocation of scores:	Yes
Criteria for scores:	Three scores allocated per sample for: i) pathogen examinations; ii) aerobic colony counts; iii) indicator organisms
Continuous performance assessment:	Yes
Criteria for identifying “poor performers”:	Less than 70 % maximum possible score over six distributions
Proactive approach by organizers to “poor performance”:	Yes
Method assessment:	Yes — for indicators only

Scheme co-ordinator or deputy to sign/date:

Approved by steering group:	Date
Promotional literature authorized:	Date
Costing approved:	Date
Accreditation dated:	Date
Other comments:	Review at steering group meeting

Annex B
(informative)

Methods of testing for variation between portions of test materials

B.1 $T_1 - T_2$ test

This test is recommended for cases where low numbers of organisms are present in portions of test materials (TMs) at levels up to 35 to 40 cfu per plate or where no "target standard deviation" can be assigned for assessing sufficient homogeneity.

The variation between analytical portions from one (reconstituted) unit of TM, T_1, and that between analytical portions from different (reconstituted) units of one batch of TM, T_2, is tested in different ways. Details can be found in Reference [12], but a summary is given here.

For the determination of the variation between analytical portions of one (reconstituted) unit of an TM (replicate testing), the T_1 test statistic is applied:

$$T1 = \sum_i \sum_j \left[\frac{\left(z_{ij} - z_{i+}/J \right)^2}{\left(z_{i+}/J \right)} \right] \tag{B.1}$$

where

z_{ij} is the number of cfu in one analytical portion j of unit i;

z_{i+} is the sum of numbers of cfus in all analytical portions of unit i

$$z_{i+} = \sum_j z_{ij} \tag{B.2}$$

J is the number of analytical portions per unit.

For the determination of the variation between analytical portions from different (reconstituted) units of one batch of TM, the T_2 test statistic is applied:

$$T2 = \sum_i \left[\frac{(z_{i+} - z_{++}/I)^2}{(z_{++}/I)} \right] \tag{B.3}$$

where

z_{++} is the sum of numbers of cfus in all analytical portions of the tested units of one batch of TMs

$$z_{++} = \sum_i (\sum_j z_{ij}) \tag{B.4}$$

I is the number of units tested.

If the Poisson distribution applies, T_1 and T_2 follow a χ^2-distribution with $I(J-1)$ and $I-1$ degrees of freedom, respectively. In this case, the expected values of T_1 and T_2 are the same as the number of degrees of freedom. Hence, $T_1/I(J-1)$ and $T_2/(I-1)$ are expected to be equal to one.

For the variation between units of one batch of TM, the Poisson distribution is the theoretical smallest possible variation that could be achieved. However, overdispersion is expected and $T_2/(I-1)$ is mostly larger than 1 (Reference [12]). An acceptable variation between units of a batch of TM is $T_2/(I-1) \leqslant 2$.

EXAMPLE

Given the following data:

unit:	(duplicate) counts:	
1	$z_{11} = 45$	$z_{12} = 49$
2	$z_{21} = 33$	$z_{22} = 42$
3	$z_{31} = 40$	$z_{32} = 42$
$I = 3$ (three units)		
$J = 2$ (two replicates)		
	$z_{1j}/J = (45 + 49)/2 = 94/2 = 47$	
	$z_{2j}/J = (33 + 42)/2 = 75/2 = 37{,}5$	
	$z_{3j}/J = (40 + 42)/2 = 82/2 = 41$	

$$T_1 = \frac{(45-47)^2}{47} + \frac{(49-47)^2}{47} + \frac{(33-37{,}5)^2}{37{,}5} + \frac{(42-37{,}5)^2}{37{,}5} + \frac{(40-41)^2}{41} + \frac{(42-41)^2}{41}$$
$$= 0{,}085 + 0{,}085 + 0{,}54 + 0{,}54 + 0{,}024 + 0{,}024$$
$$= 1{,}298$$

T_1 should follow a χ^2-distribution with $I(J-1) = 3 \times (2-1) = 3$ degrees of freedom.

Tested two-sided at the 95 % confidence level, the lower and upper limit for this distribution are, with 3 degrees of freedom, 0,22 and 9,3, respectively. The calculated T_1 value (1,298) follows these criteria.

$$\sum z_{ij} = 45 + 49 + 33 + 42 + 40 + 42 = 251$$

$$\sum z_{ij}/I = 251/3 = 83{,}7$$

$$T_2 = \frac{(94-83{,}7)^2}{83{,}7} + \frac{(75-83{,}7)^2}{83{,}7} + \frac{(82-83{,}7)^2}{83{,}7}$$
$$= 1{,}268 + 0{,}904 + 0{,}034$$
$$= 2{,}206$$

Accepted variation for the batch is: $T_2/(I-1) \leqslant 2$.

Here $T_2/(I-1) = 2{,}206/(3-1) = 1{,}103$ and thus follows the criteria for acceptability of the batch.

B.2 Test for sufficient homogeneity

This test is recommended for cases where larger numbers of organisms (more than 35 to 40 cfu per plate) are present in portions of the test material and a target standard deviation, σ_p, that describes the performance expected of PT scheme participants is available. It is based on the "sufficient homogeneity" test of Reference [11].

Given a set of test material portions analysed in duplicate with results expressed in log units, the test is passed if the between-portion variance, s^2_{sam}, satisfies Condition (B.5):

$$s^2_{sam} \leqslant F_1\left(0{,}3\sigma_p\right)^2 + F_2 s^2_{an} \tag{B.5}$$

where s^2_{an} is the analytical variance. This test is less likely than that of $T_1 - T_2$ to lead to the rejection of a test material that is not perfectly homogenous (analytical results follow the Poisson distribution), but is sufficiently homogenous to be used in a proficiency test round with a target standard deviation of σ_p. This is because materials are accepted unless it is shown with high confidence (95 %) that the fitness for purpose criterion of $\sigma_{sam} > 0{,}3\ \sigma_p$, where σ_{sam} is the standard deviation of which s_{sam} is an estimate.

EXAMPLE

Given quantitative results from the duplicate analysis of 10 test material portions, calculate the difference (D) and sum (S) and the square of D (D^2) of the decimal logarithm of each set of results (Table 1).

Portion	Analysis 1	Analysis 2	Log analysis 1	Log analysis 2	D	S	D^2
1	35	51	1,544 1	1,707 6	–0,163 5	3,251 6	0,026 733
2	52	46	1,716 0	1,662 8	0,053 2	3,378 8	0,002 835
3	35	33	1,544 1	1,518 5	0,025 6	3,062 6	0,000 653
4	53	38	1,724 3	1,579 8	0,144 5	3,304 1	0,020 878
5	30	40	1,477 1	1,602 1	–0,124 9	3,079 2	0,015 610
6	33	30	1,518 5	1,477 1	0,041 4	2,995 6	0,001 713
7	41	60	1,612 8	1,778 2	–0,165 4	3,390 9	0,027 346
8	35	55	1,544 1	1,740 4	–0,196 3	3,284 4	0,038 532
9	68	67	1,832 5	1,826 1	0,006 4	3,658 6	0,000 041
10	52	60	1,716 0	1,778 2	–0,062 1	3,494 2	0,003 862

Calculate the sum of D^2 and divide it by twice the number of portions. This is the mean sum of squares of within sample variation, $\overline{S}_w$.

In this case, $\overline{S}_w = 0{,}138\ 2/20 = 0{,}006\ 91$.

Calculate the variance of $\overline{S}_w$ and divide it by two. This is equal to $\overline{S}_b$.

In this case, $\overline{S}_b = 0{,}042\ 24/2 = 0{,}021\ 12$

Then $s^2_{an} = \overline{S}_w$ and $s^2_{sam} = \left(\overline{S}_b - \overline{S}_w\right)/2$

In this case, $s^2_{an} = 0{,}006\ 91$ and $s^2_{sam} = \left(0{,}021\ 12 - 0{,}006\ 91\right)/2 = 0{,}007\ 104$

Values for F_1 and F_2 depend on the number of portions examined in the test. For 10 portions $F_1 = 1{,}88$ and $F_2 = 1{,}01$ (values for other numbers of portions may be found in Reference [11]).

If the target standard deviation to be applied to the proficiency test results, σ_p, is equal to 0,25 $\log_{10}$ units then

$$F_1\left(0{,}3\sigma_p\right)^2 + F_2 s^2_{an} = 1{,}88 \times \left(0{,}3 \times 0{,}25\right)^2 + 1{,}01 \times 0{,}006\ 91 = 0{,}017\ 55$$

This value is larger than s^2_{sam} (0,007 104). Hence Condition (B.5) is satisfied and the test material is sufficiently homogenous.

Annex C
(informative)

A practical method to assess long-term performance of participants in PT schemes using enumeration methods

C.1 Preparing data for analysis

Four columns of a spreadsheet are required to analyse enumeration results:

a) the count as reported by the participant (n_R-count);

b) the count to be used for charts and histograms and/or for scoring (n_S-count);

c) the count for analysis (n_A-count);

d) a comment column (text field) (n_C-count) for recording any notes about participants' results.

The n_R-count should normally be entered as a numerical field (e.g. 1 100) or a scientific number (e.g. 1.1e3). If the result is reported as a censored value, then a text field shall be used and a “less than” or “greater than” symbol entered in front of the number (e.g. < 10, $> 1\,100$, etc.).

The n_R-count may also be entered as other text such as NE (not examined), ND (not detected), and UA (unassessable). Entries shall be made for all fields in the column.

If the n_R-count cannot be analysed (e.g. entered as NE) then there is no further analysis or allocation of score.

If the n_R-count cannot be included in the statistical analysis, e.g. the result was reported as a censored value, but a score is to be allocated and/or the result is to be plotted, then a numerical value shall be assigned to the n_S-count.

This value shall ensure that the correct score is allocated and that the result is plotted correctly. For example, scores may not be allocated to results reported as “not detected” but it may be helpful to include those results on a chart. In this case, the n_S-count may be entered as -99 and the n_A-count should be left blank.

If a reported result is to be allocated, a score is to be included in the statistical calculations; then the n_A-count $= n_S$-count.

C.2 Dealing with censored data

All low censored results are analysed in one of the following ways.

a) All results are assigned an n_S-count value of 0,2, but are not included in the analysis (n_A-count $= 0$). This arises when reports of a low censored value or reports of zero or “not detected” are clearly due to laboratory error because the level of the target organism or group in the sample was relatively high.

b) All low censored values and results of zero or "not detected" are assigned an n_S-count of 0,2[1)] and included in the analysis because the level of the target organism or group was relatively low and those results may have arisen by chance (n_S-count = n_A-count = 0,2). There are exceptions, such as when the reported low censored value ($<x$) shows an inappropriate level of detection and the value of x is actually higher than the median (on initial calculation). In these exceptional cases the n_A-count should be left blank.

c) Scores are not allocated to low censored values. In general, scores should be allocated unless there is a microbiological reason for not doing so. In this case, the n_A-count and n_S-count fields are left blank.

High censored values are entered as 1,0 $\log_{10}$ above the maximum reported score. If, for any reason, the results are to be excluded from the charts and analysis, and no scores are to be allocated, then the n_S-count and n_A-count fields should be left blank. If a result is reported as $>x$ where the x-value is less than the median then the n_A-count field should be left blank.

C.3 Plotting results

Plots are based on the n_S-count values. There are two main types of plots used for PT scheme results: histograms (or bar charts) and scatter plots. Points to be considered when choosing which type of plot to use include the type of examination (MPN, colony counts, etc.) and the number of participants undertaking the examination.

Histograms are produced from the decimal logarithm value of the n_R-count grouped as follows:

$<$0, (0 to 0,05), (0,05 to 0,1), (0,1 to 0,15), (max. $\log_{10}$ n_R-count), (max. $\log_{10}$ n_R-count + 0,05).

Any non-numeric or special cases to be plotted (e.g. –99) should be given their own bar.

Alternatively, the histogram may be produced based on the decimal logarithm value of the n_R-count rounded to the nearest 0,05. In this case the groupings are as follows.

0, 0,05, 0,1, 0,15 …. max. ($\log_{10}$ n_R-count).

In this case, the bar 0,1 includes all results from 0,075 to 0,124.

Scatter plots can be produced using the n_S-count values directly then converting the y-axis to a decimal logarithmic scale. Non-numeric data are excluded from scatter plots.

Bin-sizes (groups) of 0,05 $\log_{10}$ are normally used for the histograms so the ranges for allocation of scores are also rounded to 0,05 $\log_{10}$.

C.4 Allocation of scores

Where PT scheme results are assessed using scores, the criteria for allocation of scores should be listed in scheme protocols or reports. The score for the enumeration result is entered in a score field (n_C-score); this may be manually amended where necessary.

Note that the final score for a result may be reached from the n_C-score, but other factors may also require consideration.

1) The value of 0,2 is used because 0 cannot be assigned a decimal logarithm value.

Annex D
(informative)

Example of a safety data sheet

Safety data sheet for food PT scheme samples – Freeze-dried

Effective date: dd-mm-yy

Review date: dd-mm-yy

Issued to: All scheme participants

Identification of the product and the establishment

Product: Simulated food samples for general microbiological examinations

Establishment: Full address and contact details for scheme organizer

Composition or information on ingredients

Glass vials of freeze-dried material containing a mixture of bacteria of hazard group 2 as defined by national and international legislation. A hazard group 2 organism may cause human disease and may be a hazard to laboratory workers, but is unlikely to spread to the community.

Hazard identification

Physicochemical hazard: Not applicable

Health hazard: Minimal risk of infection, provided good laboratory practice is observed

Environmental hazard: Not applicable

First aid measures

If accidental contact with material occurs, laboratory staff shall follow local first aid procedures that are normally applied following exposure to an equivalent food sample. Following exposure to the material, medical advice shall be sought.

Fire fighting measures

Not applicable.

Accidental release measures

Cover the area with absorbent material and flood with a suitable disinfectant. The area shall be left undisturbed for 30 min before the spill is mopped up with an excess of absorbent material. Wear appropriate personal protective equipment.

Handling and storage

Store at room temperature in the dark. Samples shall be processed in a laboratory environment which, as defined by national regulations or guidelines, is suitable for the practice of microbiology. Staff handling the material should have been trained in the handling of infectious biological material. The material should be treated with the same degree of care as would be exercised with equivalent food samples. Hand-to-mouth contact should be avoided while working with the samples and normal hand-washing procedures relating to the handling of routine samples shall also be observed with PT samples.

Exposure controls and personal protection

Use good laboratory practice and wear appropriate laboratory coats, gloves and eye protection. Removal of vials from packaging and reconstitution should be carried out in an exhaust protective cabinet.

Physical and chemical properties

Inert odourless dry material.

Stability and reactivity

Storage is unlikely to increase or decrease the risks of infection associated with handling the material.

Toxicological information

Not applicable.

Ecological information

Not applicable.

Disposal considerations

The used material shall be disposed of using an autoclave as for food products containing infectious microorganisms and in accordance with all local and national regulations.

Transport information

Refer to national and international regulations for transport of bacteria in hazard group 2 (biological substance, category B; UN3373).

Regulatory information

EC Biological agent, hazard category/risk group 2.

CAUTION — This safety data sheet does not constitute the user's own assessments of workplace risk as required by health and safety legislation.

Other information

In the event of an accident involving exposure of staff to the material contained in the samples, contact the PT scheme organizers.

For further safety information concerning this product, participants are advised to read the instruction sheet accompanying the samples.

Bibliography

[1] ISO 5725-2, *Accuracy (trueness and precision) of measurement methods and results — Part 2: Basic method for the determination of repeatability and reproducibility of a standard measurement method*

[2] ISO 5725-4, *Accuracy (trueness and precision) of measurement methods and results — Part 4: Basic methods for the determination of the trueness of a standard measurement method*

[3] ISO/TS 11133-1:2009, *Microbiology of food and animal feeding stuffs — Guidelines on preparation and production of culture media — Part 1: General guidelines on quality assurance for the preparation of culture media in the laboratory*

[4] ISO 16140, *Microbiology of food and animal feeding stuffs — Protocol for the validation of alternative methods*

[5] ISO/IEC 17025, *General requirements for the competence of testing and calibration laboratories*

[6] ISO/TS 19036, *Microbiology of food and animal feeding stuffs — Guidelines for the estimation of measurement uncertainty for quantitative determinations*

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

[8] ILAC-G13, *ILAC Guidelines for the requirements for the competence of providers of proficiency schemes*. Available (2010-02-10) at: http://www.ilac.org/documents/ILAC_G13_08_2007.pdf

[9] THOMPSON, M., ELLISON, L.R., WOOD, R. for IUPAC. The international harmonized protocol for the proficiency testing of analytical chemistry laboratories. *Pure Appl. Chem*. 2006, **78**, pp. 145-196. Available (2010-10-08) at: http://www.iupac.org/publications/pac/2006/pdf/7801x0145.pdf

[10] AUGUSTIN, J.C., CARLIER, V. Lessons from the organization of a proficiency testing program in food microbiology by interlaboratory comparison: Analytical methods in use, impact of methods on bacterial counts and measurement uncertainty of bacterial counts. *Food Microbiol.* 2006, **23**, pp. 1-38

[11] FEARN, T., THOMPSON, M. A new test for "sufficient homogeneity". *Analyst* 2001, **126**, pp. 1414-1417

[12] HEISTERKAMP, S.H., HOEKSTRA, J.A., VAN STRIJP-LOCKEFEER, N.G.W.M., HAVELAAR, A.H., MOOIJMAN, K.A., IN 'T VELD, P.H., NOTERMANS, S.H.W., MAIER, E. A. ; GRIEPINK, B. *Statistical analysis of certification trials for microbiological reference materials*. Luxembourg: Commission of the European Communities, 1993. 41 p. (Report EUR 15008 EN.)

[13] JARVIS, B. Sampling for microbiological analysis. In: LUND, B.M., BAIRD-PARKER, A.C., GOULD, G.W., editors. *The microbiological safety and quality of food*, Vol. 2, pp. 1691-1734. Gaithersburg, MD: Aspen, 2000

[14] JARVIS, B. *Statistical aspects of the microbiological examination of foods,* 2nd edition. Amsterdam: Academic Press, 2008. 306 p.

[15] JARVIS, B., HEDGES, A.J., CORRY, J.E.L. Assessment of measurement uncertainty for quantitative methods of analysis: Comparative assessment of the precision (uncertainty) of bacterial colony counts. *Int. J. Food Microbiol.* 2007, **116**, pp. 44-51

[16] JARVIS, B., CORRY, J.E.L., HEDGES, A.J. Estimates of measurement uncertainty from proficiency testing schemes, internal laboratory quality monitoring and during routine enforcement examination of foods. *J. Appl. Microbiol*. 2007, **103**, pp. 462-467

[17] TILLETT, H.E., LIGHTFOOT, N.F. EATON, S., PLACE, B.M. External quality assessment of microbial counts from water: To score or not to score for proficiency. *J. Chart. Inst. Water Environ. Manage*. 2000, **14**, pp. 304-308

ICS 07.100.30

Price based on 28 pages