INTERNATIONAL STANDARD

ISO 22117

First edition
2019-02

Microbiology of the food chain — Specific requirements and guidance for proficiency testing by interlaboratory comparison

Microbiologie de la chaîne alimentaire — Exigences spécifiques et recommandations relatives aux essais d'aptitude par comparaison interlaboratoires

Reference number
ISO 22117:2019(E)

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

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

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO shall not be held responsible for identifying any or all such patent rights. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see www.iso.org/patents).

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

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

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

This first edition cancels and replaces ISO/TS 22117:2010, which has been technically revised. The following changes have been made:

— updates have been made to align the document with ISO 13528:2015.

Any feedback or questions on this document should be directed to the user's national standards body. A complete listing of these bodies can be found at www.iso.org/members.html.

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 [12]. 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 for a provider (and associated collaborators) of PT schemes for microbiological examinations to meet and be recognized as competent. This applies particularly to the specific technical requirements necessary to deal with microorganisms, such as sample homogeneity and stability, as well as with the interpretation of detection tests which is not covered by an existing document.

PT 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:

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

— estimation of uncertainty of test results, in conjunction with routine results, for quantitative (enumeration) methods (see ISO/TS 19036) and levels of detection for qualitative (detection) methods;

— demonstration of staff competence to perform a specific microbiological examination;

— evaluation or validation of a given method by the study of trueness, precision and robustness;

— identification of variability in test results between individual laboratories;

— assignment of a “target” value for a microorganism in a material in order to establish a reference material (see ISO 17034).

However, these aspects are not specifically covered in this document.

PT schemes are therefore designed to meet certain criteria and the testing programme (frequency, number of samples, number of repeats, etc.) to meet the requirements of the type of method used and commodity tested, to achieve the level of control required by all parties.

Microbiology of the food chain — Specific requirements and guidance for proficiency testing by interlaboratory comparison

1 Scope

This document specifies requirements and gives guidelines for the organization of proficiency testing (PT) schemes for microbiological examinations of

a) foods and beverages,

b) feeding animals,

c) environmental samples from food and feed production and handling, and

d) primary production stages.

This document is also 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 document relates to the technical organization and implementation of PT schemes, as well as the statistical treatment of results of microbiological examinations.

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

2 Normative references

The following documents are referred to in the text in such a way that some or all of their content constitutes requirements of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO 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 13528:2015, *Statistical methods for use in proficiency testing by interlaboratory comparison*

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.

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

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

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

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 data sets 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 document cover those EQA activities that meet the definition of proficiency testing.

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

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

3.3
matrix
all the components of the sample

[SOURCE: ISO 16140-1:2016, 2.38, modified — In the term, "(product)" has been removed.]

3.4
reference strain
microorganism obtained directly from an official culture collection or reference laboratory and defined to at least the genus and species level, catalogued and described according to its characteristics and preferably originating from food, food production areas, primary production stages, animals or water, as applicable

[SOURCE: ISO 11133:2014, 3.4.2, modified — In the term, "an official culture collection or reference laboratory" has replaced "a reference culture collection, i.e. a culture collection, which is a member of the World Federation of Culture Collections (WFCC) or the European Culture Collections' Organisation (ECCO)", and "food production areas, primary production stages, animals" has replaced "animal feed, the food or feed production environment".]

3.5
recovery percentage
proportion of the assigned value of the *target organism* (3.1) recovered by the participant

Note 1 to entry: The recovery percentage is calculated by multiplying by 100 the number of recovered colony-forming units (cfu) per volume or per mass and dividing by the assigned value.

Note 2 to entry: The recovery percentage can be significantly below 100 % when competitive microflora 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. This clause discusses areas requiring special consideration for microbiological PT schemes 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 biosafety levels 1 and 2, as appropriate (see ISO 7218).

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 Poisson distribution 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 sufficient, 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 (or quantification for viruses) of target microorganisms.

5 Technical requirements and guidance for sample design and content

5.1 Sources, characterization and traceability of organisms

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

Target viruses shall give expected results when tested by reference methods. Target parasites can be identified by microscopy or molecular methods depending on their size and/or other characteristics.

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

Recognized reference strains from international collections or reference laboratories should be used where they are most suitable for the scheme purpose; however, laboratory isolates or so-called "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.

NOTE Strains, particularly wild isolates, can adapt to culture media and environment unless the number of passages is kept to a minimum.

Spore suspensions can be used to inoculate samples intended to enumerate moulds as these help to improve stability and homogeneity. A method to prepare these is given in Annex B.

In all cases, the organisms used in PT scheme samples should be traceable to the relevant reference source 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, for example, PT schemes for non-cultivable organisms such as human noroviruses or parasites. Naturally contaminated samples may be used, if available, or clinical material can be used to contaminate ("spike") 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.

For distributions to be used with serological or molecular methods, it may not be necessary to distribute live, or even whole, microorganisms. Use of inactivated microorganisms, target antigens or nucleic acid sequences will often be safer and these may be more stable. Stability of such materials should be determined by the scheme provider and, in all cases, the targets should give expected results in reference methods.

5.2 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 microorganisms 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 main aim of examination for all pathogenic organisms (bacteria, viruses, parasites) is to prevent illness, but also to detect pathogenic bacteria at a very low level, before they can grow to a higher level.

For quantitative (enumeration) methods, the target level shall be appropriate for the levels routinely found in, and any specifications applicable to, the sample matrices used. The target level should also sometimes be near to 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 methods and dilutions, the expected mean number of organisms in a sample is fewer than 10 colonies per plate or less than 1 MPN/g (< 100 MPN/g for bivalve molluscs).

For traditional qualitative (detection) methods, target bacteria shall be at a sufficiently low level to provide a valid challenge to the methods and contribute validation data for performance criteria or verify levels of detection for individual participant laboratories.

5.3 Non-target organisms and interferences

The total microflora of PT samples, either naturally or artificially contaminated, is usually chosen to assess the ability of participants to detect and/or enumerate target bacteria in the presence of background flora. This background flora can include non-target strains typical of the sample matrix and presumptive target organisms which, without appropriate confirmation tests, can lead to false-positive results. However, basic schemes intended for specific purposes may provide samples containing only the target bacteria.

Any strains added to matrices to simulate background flora shall meet the requirements of 5.1 for characterization and traceability.

Determine any adverse effects of the background flora of artificially contaminated samples on the target bacteria (e.g. inhibition or other interference) before such samples are used.

5.4 Matrix selection and effects

All matrices shall be evaluated before use to check for any effects on spiked target and background flora, for example, whether a matrix reduces the recovery percentage of the spiked organisms. It may be useful to include information for participants on food matrices 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.

Include suitable and validated (or verified) preparation procedures for the proficiency samples in the information for participants.

Sample matrices used for microbiology PT schemes are often, but not necessarily, sterilized before use. Alternatively, the absence of the target is checked by other means (e.g. use of special sources of matrices).

Where natural, unsterilized sample matrices are distributed, the organizers shall determine the effect of any background microflora on the target organisms before use. Also, absence of target organisms in natural samples is usually required if these are to be artificially spiked. For example, in a PT scheme to detect parasitic *Anisakidae* larvae in fish fillets, use only freshwater fish fillets as the larvae are found only in sea fish.

6 Sample verification by the provider

6.1 General

General requirements for sample verification are given in ISO/IEC 17043 and ISO 13528. This clause expands the specific requirements and any particular issues 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, assess the test material for homogeneity, usually before but also at the time of testing for less stable fresh materials.

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

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 appropriate) is suggested.

A test material that 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 such 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.

For parasite detection, each test material is generally spiked with a known number of organisms and homogeneity checked by microscopy counts of each sample, by at least two operators, since homogeneous distribution of target parasites in bulk material is difficult to achieve.

For virus methods, the bacteriological terminology used in 6.3 and 6.4 may not be applicable, but similar principles apply to ensure sufficient homogeneity in distributed PT samples.

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, but alternative procedures may sometimes be required for microbiological materials.

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. However, in general, a material is considered sufficiently homogenous if the between-unit standard deviation (on the appropriately transformed scale) is $\leq 0{,}3\sigma_{pt}$, where σ_{pt} is the target standard deviation used to assess the performance of laboratories (see ISO 13528).

Any alternative homogeneity test should meet the following criteria:

a) the probability of rejecting a sufficiently homogenous test material should be ≤ 5 %;

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

An 80 % probability of rejecting a material where between-unit variation is $1{,}5\sigma_{pt}$ is based on simulation studies of the duplicate analysis of 10 test units using a method with an analytical standard deviation of $0{,}5\sigma_{pt}$ (i.e. 0,125 log units) and a critical value for T_2 (see Annex C) 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 and analytical (repeatability) variance obtained under repeatability conditions. Suitable methods of testing for such variation are given in Annex C.

The analytical (repeatability) variance should be estimated from replicate analyses of the initial suspensions obtained from test portions[17][18]. This analytical variance can also be calculated from the number of counted colonies and the precision of analytical materials in use[13].

In microbiology, the between-unit variance shall be estimated under repeatability conditions (in one run). If that is not possible, the between-unit variance will include the within-laboratory reproducibility and can perhaps lead to 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 Formula (1):

$$\frac{\sigma_{an}}{\sigma_{pt}} < 0{,}5 \tag{1}$$

where σ_{pt} is the target standard deviation.

In such cases, the test for sufficient homogeneity proposed in Reference [14] should be used (see Annex C). 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 C).

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 replicates increases the statistical power of the analysis and can be a good indicator for successive rounds.

Where the PT material contains low numbers of cfu (e.g. 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 0,95 probability.

6.4 Homogeneity testing for qualitative methods

Similar principles apply to homogeneity testing for qualitative (detection) bacterial methods, but special consideration is required where the level of target organisms is low (see 8.4).

The homogeneity of qualitative samples for bacteriological methods may be tested by enumerating the spike. For levels of < 100 cfu/g use a most probable number (MPN) technique and for levels of 100 cfu/g and above use the appropriate plate count enumeration.

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. The number of samples tested should be proportional to the number of test materials produced.

These approaches may also be used with molecular methods (e.g. detection of viruses in foods). However, because there may be differences between performance characteristics of quantitative and qualitative methods for the same target, homogeneity should also be checked with a reference qualitative method.

6.5 Stability testing by the provider

6.5.1 General

Samples to be used for PT shall be stable enough to complete the study before 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 on the testing date) 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 testing.

For stability checks on qualitative samples containing low levels, a number of samples (e.g. 10) can be tested at different time intervals (and at different storage temperatures when needed) throughout the time period of the study. The number of positive samples after different storage times will give information on stability throughout the study.

Fresh natural samples spiked with target organisms are often not stable and need special consideration as discussed in 6.5.3.

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), determine the stability by checking the level of the organisms 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 for testing by participants.

Frequency of stability 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 to show ongoing stability across the whole batch (ISO 13528:2015, 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).

Perform a stability test initially, using different temperatures to reflect "worst case" transport conditions and maximum expected delays, 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.

For samples prepared using natural matrices, stability studies should be carried out for each matrix and target microorganism combination because even minor changes in either the matrix or the target can affect sample stability during distribution (e.g. different species of bivalve molluscs).

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, for example, 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 (see national, regional or international regulations), 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 for performing the examinations and sending results 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, and

   4) if the organizers request details of materials, methods, incubation conditions, the date and time of testing, etc., in report proformas, they should provide instructions for completing such proformas.

Instructions may be despatched in hard copy with samples or accessible electronically, for example, on the scheme website.

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. However, microbiology PT schemes may adopt procedures which differ from those commonly used in other sectors if they are appropriate to their particular schemes and the robustness of such procedures with respect to outliers is specified (see, for example, Reference [21] on evaluation of PT schemes for parasites).

8.2 Preliminary considerations

PT involves the regular distribution of test materials to participating laboratories for examination for specific targets (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 (see ISO/IEC 17043).

Ongoing satisfactory performance in proficiency test rounds can provide reassurance to participants on 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 PT but objective interpretation methods are required.

The majority of participants taking part in microbiological PT schemes may not be familiar with statistics, so scheme organizers have a responsibility to demonstrate that the procedures used are sound.

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

a) validity;

b) explanatory information for participants;

c) reasons for selection;

d) consistency of interpretation of the data.

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 Assessment of 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.

Where participants are permitted to use a method of their choice, the PT provider shall choose a method of statistical analysis that is appropriate for the scheme (see ISO/IEC 17043:2010, 4.5.2) although this can be problematic if the methods used by all the participants are not known. A preferred method of analysis, for example, plate count enumeration rather than MPN may be given in the scheme plan.

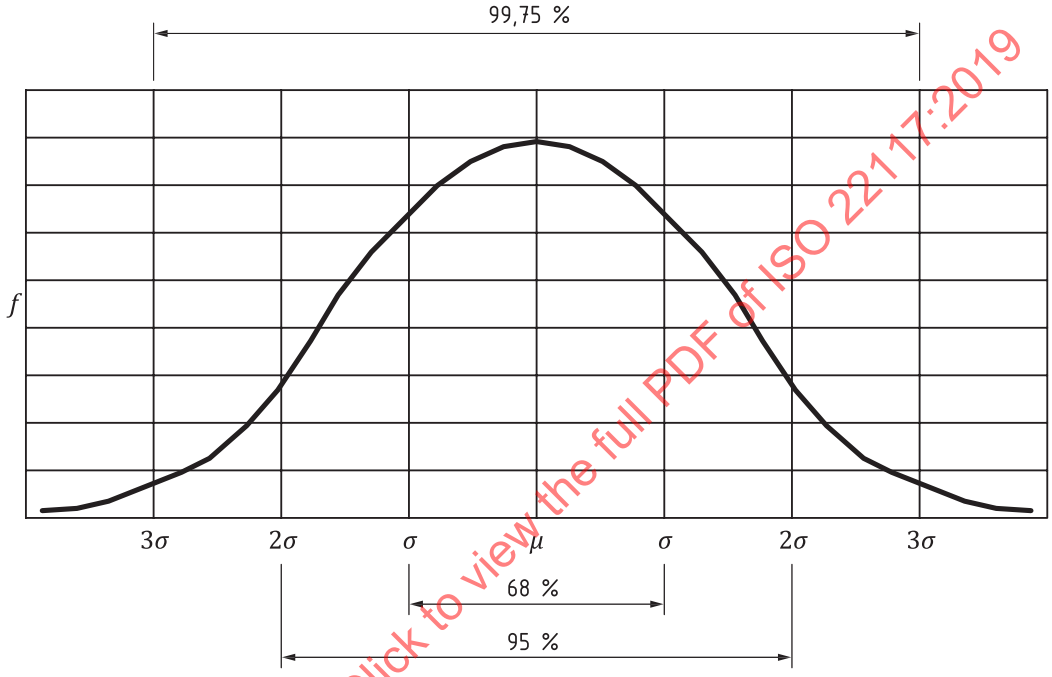
Where methods used are requested from the participants, results from each method can be separated and statistically analysed so any method effects can be determined and included in the report.

Different statistical parameters may be needed to assess participants' performance for different tests within a single scheme. 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. Handling of censored values, such as results of < 10 or < 100 cfu, is discussed at E.2 and in ISO 13528:2015, Annex E.

8.3.2 Distribution of data

Repeated microbiological counts usually follow a log-normal distribution so the data are transformed to logarithm to the base 10 values to produce a bell-shaped normal distribution curve as shown in Figure 1. However, for low target organism levels, 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 PT, the tests are not usually performed under repeatability, or even reproducibility conditions. Tests are performed by different personnel, at different test locations and at different times, using a variety of equipment, media, reagents and test 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 PT is to assess how proficient participants are in achieving the "correct" result. However, in many PT 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, median or mode 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. Each measure of consensus (robust mean, median or mode) has an associated uncertainty.

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.

Highly deviant distributions can result if different methods are used by participants. In such cases, the scheme organizer shall acknowledge any effects of the methods used.

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 are given action and warning evaluations (see 8.3.6.1), 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 data set, the more accurate such statistical estimates are likely to be but the distribution should remain homogeneous with no multi-modality, for example, from use of different methods by participants.

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 PT 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 2σ of the mean and 99,7 % of values within 3σ. Results with a *z*-score greater than 2 are therefore considered questionable and should be used as "warnings" 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 and require "action" 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 (σ_{pt}). This target standard deviation defines the scale of acceptable variation among laboratories for each particular test and σ_{pt} of 0,35 or 0,25 are commonly used in microbiological PT schemes. 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 [12], ISO/IEC 17043:2010, B.3, and ISO 13528:2015, Clause 8.

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 results from 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 sample and/or laboratory is 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 laboratory is permitted by the scheme organizers, the individual result to be reported shall be chosen before the examination is undertaken and results are known to avoid bias.

8.3.6.4 Using scaled median absolute deviation (*MADe*) from the median values

The scaled *MADe* method is an alternative to the classical standard deviation in the calculation of the z-score; it is used to identify outlying counts when fewer than 50 participants undertake an enumeration (see ISO 13528).

MADe 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 the constant 1,482 6 to get a robust estimate of the standard deviation (*MADe* value), $\sigma MADe$.

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

— results within participants' median ±2$\sigma MADe$: Score = 2

— results between ±2$\sigma MADe$ and ±3$\sigma MADe$: Score = 1

— results outside ±3$\sigma MADe$: Score = 0

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 ±2σ*MADe* range and 1 % outside the ±3σ*MADe* range (assuming normality of the decimal logarithm counts without extreme outliers).

The *MADe* method should be used for PT schemes where there are fewer than 50 participants.

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 bacteriological schemes.

For example, with samples where low bacterial 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 distribution. 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 [20].

With schemes where high bacterial counts are expected, the 0,5 $\log_{10}$ rule or the percentile approach can be used. These are briefly outlined in 8.3.7.2 and 8.3.7.3.

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 [16]). In summary, the 95 % confidence intervals around a mean colony count are generally not more than ±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 ±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 ±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 ≥ 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, percentiles should not be used because fewer than 50 results provide insufficient data to calculate valid values for C5, C10, C90 and C95. In such cases, use scaled median absolute deviation (*MADe*) values (see 8.3.6.4) instead.

8.3.7.4 Poisson 95 % confidence interval (CI)

Poisson 95 % CI may be used when a sample contains low levels (≤ 20) of organisms for a specific parameter, for example, if the participants' median or the organizer's median is 1 cfu/ml in schemes for water samples.

This method is used to ensure that participants receive the maximum score for all counts that could be due to random variability of organisms in the sample (Poisson variability). Other methods may give ranges that are tighter than this which would be unreasonable. The rationale is similar to that of using the 0,5 $\log_{10}$ rule for other schemes.

The correct 95 % CI for Poisson data with a mean (median) of 1 is 0 to 3 so the 95 % confidence interval around the participants' or organizer's median may still be applied even at very low levels.

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 regional 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 derived using the MPN technique 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.

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 increase the variance by about 1,8 times (and hence the standard deviation by about 1,34 times).

Therefore, the limits of acceptability for participants' results for MPN determinations should be raised to $\pm 2{,}68\sigma$ and $\pm 4\sigma$ (see Table 1).

Table 1 — Limits of acceptability

Limit of acceptability	Three-by-three method	Three-by-five method
$\pm 2{,}68\sigma$	$\pm 0{,}86 \log_{10}$	$\pm 0{,}64 \log_{10}$
$\pm 4\sigma$	$\pm 1{,}28 \log_{10}$	$\pm 0{,}96 \log_{10}$

These limits may be changed based on inflation factors determined from further assessment of between-laboratory variability.

The 0,5 $\log_{10}$ rule should never 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.

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 PT schemes is generally confined to assessment of results from single rounds, but there are instances where assessment in the longer term may be beneficial. While 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 the likelihood of identifying laboratories undertaking enumeration examinations as "poor performers" by chance due to variations in the number of organisms in samples they receive is as low as possible.

"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 judgement 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. 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 E.

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} and percentage accuracy.

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 is linked with records of the levels of target organisms in the samples to assess performance and also provide ongoing data on the level of detection of the method in individual laboratories.

For example, in a PT scheme where participants receive two samples of which one, both or neither contains the target organism, the assessment is based on detected/not detected in both samples. An

overall percentage satisfactory assessment can therefore be provided for all participants and linked to the level of target organisms in the test materials provided in the report (see 6.4).

In certain cases more samples may be tested to verify laboratory performance at low level, in particular around the LOD_{50} for a qualitative method. An example based on 18 samples is given below, but other combinations of numbers of samples at each level can be used and justified.

Each participant should test at least 18 samples in total. These 18 samples consist of 6 replicates of 3 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 (samples contaminated at or slightly above the detection level for the method used, ideally at the level where 50 % of the samples are found positive and 50 % negative (LOD_{50}); and high level (10 times higher than the low level, representing the level at which all samples tested should be found positive).

Interpretation of the data is simple:

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

b) for the high level: all 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 (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 determine the average percentage of positives. Here, it is assumed to be 30 %, meaning that around 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.

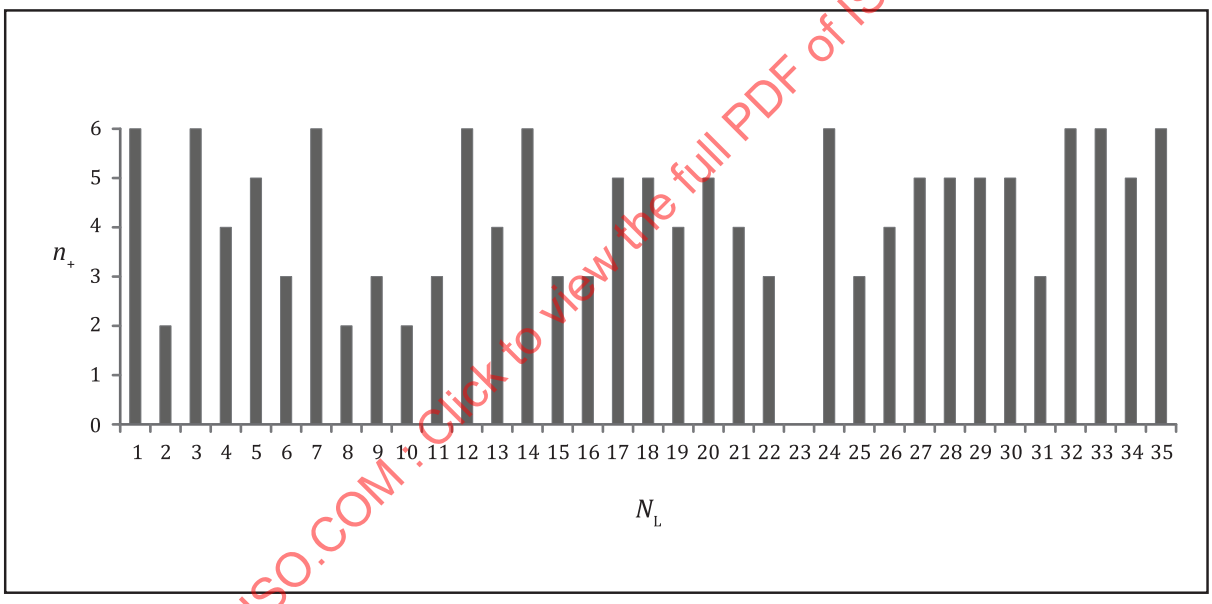
Table 3 shows that the chance of obtaining a set of six samples consisting of four positives and two negatives is 6 %. This is not likely to happen, but is still acceptable when a level of confidence of 95 % is assumed. This can be calculated as follows: the sum of the chances of finding 0, 1, 2, 3 or 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). Omitting four positives from this sum gives 93 %, which is below the 95 % confidence level.

Looking at the data the other way around, the chance of finding six out of six positives is only 0,1 %, which is very unlikely to occur purely by chance. With the uncertainty set at a maximum of 5 % (100 % to 95 % confidence level) this falls within the 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, 35 laboratories participated (indicated as lab codes on the N_L-axis of Figure 2). Each laboratory analysed 18 minced chicken meat samples, artificially contaminated with a *Salmonella* serovar at 3 different contamination levels (0 cfu, 5 cfu and 55 cfu per sample). In Figure 2, the results are summarized for the sample with the low level of contamination (n = 6). The number of positives found in the participating laboratories varied from 0 to 6 samples (n_+-axis in Figure 2).



Key

N_L lab codes

n_+ no. of positives

Figure 2 — Number of positive isolations per laboratory code for all tested low-level materials (n = 6)

In addition to this more descriptive way of presenting the results, it is possible to calculate specificity and sensitivity rates per level of contamination of the samples (see ISO 16140-2). These rates may be calculated for each laboratory and for the results from all laboratories.

The specificity rate, r_{SP}, is given by Formula (2):

$$r_{SP} = \frac{n_-}{E(n_{-\,tot})} \times 100\ \% \tag{2}$$

where

n_- is the number of negative results found;

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

The sensitivity rate, r_{SE}, is given by Formula (3):

$$r_{SE} = \frac{n_+}{E(n_{+\,\mathrm{tot}})} \times 100\ \% \tag{3}$$

where

n_+ is the number of positive results found;

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

Table A.1 — PT scheme plan — Summary of scheme

Scheme name:	Food examinations scheme
Scheme provider:	Name of organization
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 biosafety levels 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:	**Qualitative (detection):** — *Campylobacter* spp. — *Escherichia coli* O157 — *Salmonella* spp. **Quantitative (enumeration):** — 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 testing
Target number of participants:	More than 200
Number of distributions per year:	Six
Number of samples per distribution:	Two
External subcontractors:	None
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 (enumeration tests) Percentiles to identify outliers
Allocation of scores:	Yes

Table A.1 *(continued)*

Criteria for scores:	Three scores allocated per sample for: a) pathogen examinations; b) aerobic colony counts; c) 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)

Preparation of fungal spore suspensions

B.1 General

Cultures of filamentous fungi used in PT schemes are preferably distributed as spore suspensions to improve homogeneity and stability of samples during distribution and testing. This annex describes a method of preparing such spore suspensions for guidance. It is recognized that other methods exist.

B.2 Procedure

B.2.1 General

Manipulate all fungal cultures with suitable containment to avoid spread, for example, in a safety cabinet, and only prepare one culture at a time to minimize cross-contamination.

Use only well-characterized strains (see 5.2) and ensure cultures on plates or slopes are free from obvious contamination before preparing spore suspensions.

B.2.2 Preparing spore suspensions from agar plates

Use a swab moistened with sterile distilled water to rub growth from the surface of the plates and suspend in 1,5 ml of sterile distilled water. Transfer the harvest to a larger volume (e.g. 500 ml) of sterile distilled water and mix the bulk suspension well to improve homogeneity.

Carry out a purity check on the suspension by surface inoculation of a suitable agar and incubate at the appropriate temperature (e.g. Sabouraud Dextrose Agar incubated at 25 °C for 48 h to 72 h).

If the bulk suspension is pure, carry out a spore count by microscopy to estimate spore numbers so the PT samples can be inoculated with the desired number of spores.

B.2.3 Preparing spore suspensions from agar slopes

Preparation from agar slopes is similar to that from agar plates, except that spores are best harvested by adding 1 ml sterile distilled water to the slope and rubbing the growth with a sterile swab to release the spores.

B.2.4 Storage

Store all fungal spore suspensions at ambient temperature (18 °C to 27 °C) in the dark in a designated and labelled laboratory cupboard.

B.2.5 Quality control checks

Perform quality control checks on each spore suspension to ensure homogeneity and stability are maintained during storage. This is carried out regularly (e.g. weekly) by inoculating the water suspensions onto selective media to check viability. The growth is then examined macroscopically and microscopically to check stability of the suspensions.

Annex C
(informative)

Methods of testing for variation between portions of test materials

C.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 cfu 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 [15], but a summary is given here.

For the determination of the variation between analytical portions of one (reconstituted) unit of a TM (replicate testing), the T_1 test statistic is applied, as shown by Formula (C.1):

$$T_1 = \sum_i \sum_j \left[\frac{\left(z_{ij} - z_{i+} / J\right)^2}{\left(z_{i+} / J\right)} \right] \tag{C.1}$$

where

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

z_{i+} is the sum of numbers of cfu in all analytical portions of unit i, as shown by Formula (C.2):

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

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, as shown by Formula (C.3):

$$T_2 = \sum_i \left[\frac{\left(z_{i+} - z_{++} / I\right)^2}{\left(z_{++} / I\right)} \right] \tag{C.3}$$

where

z_{++} is the sum of the numbers of cfu in all analytical portions of the tested units of one batch of TMs, as shown by Formula (C.4):

$$z_{++} = \sum_i \left(\sum_j z_{ij} \right) \tag{C.4}$$

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, over-dispersion is expected and $T_2/(I-1)$ is mostly larger than 1[15]. An acceptable variation between units of a batch of TM is $T_2/(I-1) \leq 2$.

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 limits 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: $T_2/(I-1) \leq 2$.

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

C.2 Test for sufficient homogeneity

This test is recommended for cases where larger numbers of organisms (more than 35 cfu to 40 cfu per plate) are present in portions of the test material and a target standard deviation, σ_{pt}, which describes the performance expected of PT scheme participants, is available. It is based on the "sufficient homogeneity" test of Reference [14].

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 Formula (C.5):

$$s^2_{sam} \leq F_1 \left(0{,}3\,\sigma_{pt}\right)^2 + F_2 s^2_{an} \quad \text{(C.5)}$$

where s^2_{an} is the analytical variance and values for F_1 and F_2 depend on the number of portions examined in the test (see below). 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 (test results follow the Poisson distribution), but is sufficiently homogenous to be used in a proficiency test round with a target standard deviation of σ_{pt}. This is because materials are accepted unless it is shown with high confidence (95 %) that the fitness for purpose criterion of $\sigma_{sam} > 0{,}3\ \sigma_{pt}$, where σ_{sam} is the standard deviation of which s_{sam} is an estimate.

EXAMPLE

Given quantitative results from the duplicate (t = 2 analysis of g = 10 test material portions, calculate the difference and sum (S) and the square of the difference of the decimal logarithm of each set of results (Table 1). Note that the square of the difference is twice the between portion variance, that is $2w_t^2$.

Item (*g samples*)	Count 1 (Test portion 1)	Count 2 (Test portion 2)	Log Count 1	Log Count 2	Difference (Log C1 − Log C2	Sum = Z (Log C1 + Log C2)	Difference squared [$2w_t^2$]
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
∑						32,900	0,138 203
Mean						3,290 0	0,013 820
Variance						0,042 24	

Calculate the "within sample" variance $S_{an}^2 = \sum (2w_t^2)/2g$ where g is the number of samples.

In this case, $S_{an}^2 = 0{,}138\,2/20 = 0{,}006\,91$.

Determine the variance of the summed counts (Z_t)

$$Var(Z) = \frac{1}{(g-1)} \sum_{t=1}^{g} (Z_t - \bar{Z})$$

where

Z_t is the summed value for sample t;

$\bar{Z}$ is the mean value ($\sum Z_t/g$) (which is twice the overall average $\bar{\bar{x}}$);

g is the number of samples.

Next, determine the variance of the sample averages (S_x^2) by dividing the calculated variance by two, i.e. $S_x^2 = Var(Z)/2$.