TECHNICAL SPECIFICATION

ISO/TS 5594

First edition
2022-11

Soil and water quality — Guidance and requirements for designing an interlaboratory trial for validation of biotests

Qualité de l'eau et du sol — Recommandations et exigences relatives à la conception d'un essai interlaboratoires pour la validation des essais biologiques

Reference number
ISO/TS 5594:2022(E)

ISO copyright office
CP 401 • Ch. de Blandonnet 8
CH-1214 Vernier, Geneva
Phone: +41 22 749 01 11
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 190, *Soil quality*, Subcommittee SC 4, *Biological characterization*, in collaboration with Technical Committee ISO/TC 147, *Water quality*, Subcommittee SC 5, *Biological methods*.

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

Validation of an ISO biotest standard aims to estimate the uncertainty of results by means of an interlaboratory trial. This validation program typically involves three major steps.

a) Demonstration of readiness for testing: This 1st step is an opportunity for the lead laboratory to ensure that recommended instrumentation is in place and that test organism and cell cultures have been established using performance measures to ensure healthy organisms or cultures are used in testing. These conditions are typically part of the clause/subclause entitled "Test organisms" or "Test material" in an ISO standard.

b) Demonstration of laboratory capability and transferability of the biotest: In this 2nd step, the participating laboratories aim to achieve successful control performance during this preliminary interlaboratory testing round with a reference compound added either to a solid matrix or to a liquid medium according to the biotest method to be validated. The ability to conduct the testing standard is demonstrated by fulfilling the validity criteria (e.g. variability of controls expressed as the coefficient of variation for the number of juveniles in a reproduction test) and qualifies the laboratory for the final method validation step.

c) Method validation: The 3rd step involves only laboratories who have demonstrated the expertise in conducting the ISO standard under development (step b). In case of validation of an ecotoxicological or microbiological testing method, one or two rounds of interlaboratory method validation trial using a contaminated environmental sample (or samples) are conducted and the results from each round are used to calculate the intralaboratory and interlaboratory variability of the ISO testing standard as a demonstration of method precision (Annex A). If repeated testing runs of the biotest are feasible, repeatability is determined. For the validation of ecotoxicological methods it can be useful to evaluate the correctness of the measured effect – the measurement trueness. This holds true especially for methods of which the results are reported in terms of a quantitative measurement such as a biological equivalence concentration. Obtained results are used for confirming or adjusting the validity criteria.

For the validation study, representative samples should be selected according to the intended scope of the standard (e.g. contaminated soils, amended soils, soils after remediation, waste materials, wastewaters, eluates, surface water, groundwater, sediments and extracted samples).

An overall schema of the validation process can be found in Annex D.

Soil and water quality — Guidance and requirements for designing an interlaboratory trial for validation of biotests

1 Scope

This document aims to assist in designing and organizing trials for validation of biotests. The validation activities during the different steps of the standardization process are described. This document comprises the overall data evaluation and subsequent validation study conclusion.

This document is intended for the validation of biotests which can differ in their experimental design and endpoints. It is possible that some of the requirements of this document are not applicable to all test methods.

2 Normative references

There are no normative references in this document.

3 Terms and definitions

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

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

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

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

3.1
control
material and/or matrix that duplicates all factors that can affect results except the specific condition or treatment being studied

Note 1 to entry: In toxicity tests, the control should have all the same conditions as in the treatment exposure but without the toxicant.

[SOURCE: Environment Canada 2005]

3.2
endpoint
statistically derived toxicity threshold (e.g. EC50)

[SOURCE: Environment Canada 2005, modified — The recommendation not to use the term for observed variables, such as size, is deleted]

3.3
lead laboratory
laboratory responsible for organization of the interlaboratory validation study

3.4
measurement trueness
trueness of measurement
trueness
closeness of agreement between the average of an infinite number of replicate measured quantity values and a reference quantity value

Note 1 to entry: The requirement for an infinite number of replicate measurements has a theoretical background. In practice a large series of test results is used to estimate the measurement trueness.

[SOURCE: ISO/IEC Guide 99:2007 2.14, modified — Notes to entry have been replaced]

3.5
performance characteristics
measures of the performance of a test under specific conditions, including its reliability and accuracy

Note 1 to entry: Performance characteristics are an indication of the test's usefulness, limitations, and relevance.

[SOURCE: OECD 2005]

3.6
precision
closeness of agreement between indications or measured quantity values obtained by replicate measurements on the same or similar objects under specified conditions

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

Note 2 to entry: The 'specified conditions' can be, for example, *repeatability conditions* (3.10) of measurement, intermediate precision conditions of measurement, or *reproducibility conditions* (3.12) of measurement (see ISO 5725-1).

Note 3 to entry: Measurement precision is used to define measurement *repeatability* (3.9), intermediate measurement precision, and measurement *reproducibility* (3.11).

Note 4 to entry: Sometimes "measurement precision" is erroneously used to mean measurement accuracy.

[SOURCE: JCGM 200:2012]

3.7
prevalidation
initial phase(s) of a validation study

Note 1 to entry: A small-scale study intended to obtain preliminary information on the relevance and reliability of a test method. Based on the outcome of those studies, the test method protocol may be modified or optimized to reduce intra- and/or interlaboratory variability and increase accuracy in subsequent validation studies. If available, literature data may be used for this purpose.

[SOURCE: OECD:2005, modified — Reasons for performing prevalidation are not included]

3.8
reference compound
chemical for which the response of the test organism is known

3.9
repeatability
measurement *precision* (3.6) under a set of *repeatability conditions* (3.10) of measurement

[SOURCE: JCGM 200:2012]

3.10
repeatability condition
condition of measurement, out of a set of conditions that includes the same measurement procedure, same operators, same measuring system, same operating conditions and same location, and replicate measurements on the same or similar objects over a short period of time

Note 1 to entry: A condition of measurement is a repeatability condition only with respect to a specified set of repeatability conditions.

[SOURCE: JCGM 200:2012]

3.11
reproducibility
measurement *precision* (3.6) under *reproducibility conditions* (3.12) of measurement

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

3.12
reproducibility condition
condition of measurement, out of a set of conditions that includes different locations, operators, measuring systems, and replicate measurements on the same or similar objects

Note 1 to entry: The different measuring systems may use different measurement procedures.

Note 2 to entry: A specification should give the conditions changed and unchanged, to the extent practical.

[SOURCE: JCGM 200:2012]

3.13
interlaboratory method validation trial
interlaboratory validation study in which all laboratories perform the same biotest using the same material under testing and the same test protocol

Note 1 to entry: The purpose of the test is to determine inter laboratory and, whenever possible, intra laboratory variability.

3.14
transferability
ability of a test procedure to be accurately and reliably performed in independent, competent laboratories

[SOURCE: OECD:2005]

4 Principle

Validation of a proposed test method is achieved through the demonstration of reasonable coefficients of intralaboratory and interlaboratory variation, which have participated in the formal method validation testing round(s) (i.e. the 3rd validation step, 5.5). Repeatability should preferentially be determined. However, on occasion, the duration or workload of the test method prevents the analysis by individual laboratories from being repeated. In case of missing experimental data obtained under repeatability conditions, it is proposed to estimate repeatability from the confidence limits of the toxic metric that integrates the variability of repeated measures of the biological variable in each treatment/dilution (see Reference [67]). If appropriate, the measurement trueness of the method can be determined in line with the method validation, in particular for methods the results of which are reported in biological equivalence (BEQ) concentrations (see ISO 23196). The final decision to include the measurement trueness in the interlaboratory method validation trial has to be taken by the lead laboratory in consultation with other experts involved in the development of the method.

5 Requirements, design and organization of method validation testing round(s)

5.1 General

A prerequisite for the performance of an interlaboratory trial is the method development has been completed. As the aim of this kind of interlaboratory comparison is to evaluate the performance characteristics of a biotest, technical aspects are defined within an interlaboratory method validation trial procedure prepared by the lead laboratory and unique to each round of interlaboratory testing. Specific features which are recommended to be addressed by lead laboratories are summarized in Annexes F and G for terrestrial and water tests, respectively.

A lead laboratory should be able to demonstrate their level of competency for planning and conducting the interlaboratory method validation trial. For example, the laboratory proposing to lead the validation effort can state the years of experience of participating staff in conducting environmental toxicology or microbiological testing (e.g., 10 + years); provide proof of recognised quality service through a laboratory certification or accreditation program; provide proof that they have a track record in planning and conducting multi-laboratory testing studies; demonstrate in writing that their staff have a clear understanding of all steps the laboratory would be responsible for leading an inter-laboratory validation program; clearly state their willingness to provide the materials to culture or holding and acclimation organisms required to conduct the new biological testing standard prior to and during the full validation testing round; or, state their willingness to provide remote or on-site training to assist participating laboratories with preliminary steps needed to show their readiness to conduct the new testing procedure. Specific suggestions and further details on steps and capability of a lead laboratory and their staff are outlined in ISO 17043.

Within the interlaboratory method validation trial, the lead laboratory should:

— be responsible for the organization and performance of the three validation steps (5.3, 5.4, 5.5) or interlaboratory testing rounds;

— prepare an interlaboratory method validation trial procedure for each round of testing, defining all technical aspects of the performance of the biotest;

— supply the participants with the necessary instructions, test samples (e.g. contaminated soils, amended soils, soils after remediation, waste materials, wastewaters, eluates, surface water, groundwater, sediments and extracted samples) and materials selected by the lead laboratory, for each testing round;

— be responsible for the analysis and reporting of results from each testing round;

— be responsible for requesting feedback from laboratory participants on the proposed interlaboratory method validation trial procedure prior to each testing round, through an organized tele- or video conference or e-mail exchange.

An example of different roles of personnel involved in organising of the interlaboratory method validation trial can be found in ISO 5725-2:2019, Clause 6.

Unless requested otherwise by the lead laboratory, it is highly recommended to transfer all results along with all test condition parameter measurements, immediately after completion of testing for each specific testing round.

5.2 Prevalidation — Minimum requirements of test method performance

The primary step of method validation should rely on a thorough evaluation of the peer-reviewed scientific literature and other relevant and credible reports and publications providing information about the performance of the test method. This literature review should aim at identifying experimental conditions which can have an impact on the outputs of a biotest. As a result, additional laboratory experiments can be performed to refine and optimize the procedure accordingly. Special attention should be paid to the ecological relevance of organisms and biological responses used in biological, ecological and ecotoxicological testing with respect to the properties of the matrix to be considered, e.g.

pH, organic matter, conductivity, clay content or turbidity. The interlaboratory method validation trial design should cover the range of properties appropriate for the testing matrix (e.g. water, sediment, soil) for which the biotest is relevant.

For the optimization of the biotest for ecotoxicity testing, controls have to be defined which differentiate effects of intrinsic properties of the test samples from those caused by contaminants. For example, the validation of a test with regards to its applicability in the risk assessment of contaminated sites, should include a contaminated as well as a reference soil sample. The reference soil sample should have similar intrinsic properties to the contaminated one, but contain no or negligible levels of the contaminants.

Test methods are often defined for two or more different type(s) of samples, for instance, contaminated sites, amended soils, soils after remediation, waste materials (e.g. dredged material, municipal sludge from a wastewater treatment plant, composed material, or manure, especially those for possible land disposal), wastewaters, eluates, surface water, groundwater, sediments, chemicals. Ideally, the test method is validated using the most applicable contaminant or contaminated media type. However, full validation of a method that incorporates a large variety of samples is typically not possible. It can be unnecessary to perform such a wide validation if a limited number of samples is, to the lead laboratory experts' opinion, considered representative of the most important fields of application. For example, the validation of a test method for evaluating the quality of surface waters (e.g. river water samples from a non-polluted upstream area and a downstream point affected by diffuse pollution or an effluent discharge), or the efficiency of wastewater treatment plants (e.g. using liquid samples of pre-treatment influent and/or final treated effluent) should consider the selection of samples appropriate for the application purpose.

5.3 First validation step — Demonstration of participating laboratories' readiness for testing

The first step is to demonstrate the readiness of participating laboratories to culture test organisms or cell cultures and to conduct the testing standard. It is recommended to evaluate readiness of the participating laboratories using a questionnaire or survey completed by each participant (Annex E). This questionnaire provides information to the validation trial coordinator of the lead laboratory regarding the available technical know-how, experience and resources for the culturing of test organisms and conductance of the test procedure.

5.4 Second validation step — Demonstration of laboratory capability and transferability of the biotest

The second step of interlaboratory testing is used to perform a preliminary assessment of the transferability and reliability of the test and to identify possible limitations of the test. Occasionally, interlaboratory method validation trial rounds involving experienced laboratory participants do not generate comparable data. In such a situation, the lead laboratory can further restrict method options to bring a higher degree of standardization prior to the next round of inter-lab testing or can conduct or sponsor additional method research to further standardize the culturing or testing parts of the methodology. Results of the second validation step can be used in the design of the future interlaboratory validation testing round.

At least one laboratory independent from the laboratory that developed the test method conducts the full biotest for an initial assessment and review of its interlaboratory transferability and preliminary reproducibility. Participating laboratories perform the biotest in control conditions and with a reference compound with known toxicity at a specific test concentration or known concentration range provided by the lead laboratory. Results are evaluated by the lead laboratory which can lead to optimization of the organisms' culturing and/or testing procedure. Only laboratories who pass the proposed validity criteria (e.g. variability of controls expressed as the coefficient of variation, sensitivity, biological response rate of the tested organisms in controls) should participate in the method validation testing round (see 5.5). Involvement of inexperienced laboratories is outlined in 5.5.2. Integration of these laboratories sometimes results in less precise conclusions leading to questions of test transferability. In cases where the test method fails to provide sufficient reproducibility, depending on the degree of failure, it can be considered for further optimization or can require further test method research. All

requirements specified for the interlaboratory method validation trial in 5.5.5 are valid for the second step.

5.5 Third validation step — Method validation

5.5.1 General

The third step involves a testing round (or rounds) to achieve method validation. The interlaboratory method validation trial procedure is designed to evaluate the interlaboratory variability of the results obtained. This is done by conducting the same measurements using the full test method in each participating laboratory. Repeatability (intralaboratory variability) can be estimated if the testing and data analyses are repeated by each laboratory. An alternative approach for the estimation of repeatability using the confidence limits of the toxic metric that integrates the variability of repeated measures of the biological variable in each treatment/dilution can be found in Annex B. If the output is not a numerical value (e.g. extraction of DNA), a specific approach is needed to evaluate reproducibility (see 5.5.6).

5.5.2 Participating laboratories

The number of participants has an influence on the reliability of the statistically calculated performance data. It is suggested to obtain valid datasets from a minimum of 6 laboratories located in 3 different countries. Therefore, it is strongly recommended that a greater number of test laboratories and countries be invited and participate in interlaboratory method validation trials. Annex C provides guidance on how to proceed if a lower number of valid datasets is achieved by calculating and reporting the uncertainty of the reproducibility variance.

Participation in this kind of trials is voluntary, and each participant laboratory should be given a code (that can be communicated to the respective laboratory) to maintain anonymous the source of the validation trial datasets. If the call for participants yields an insufficient number of laboratories that are experienced with the method, assistance in applying this method should be provided by the lead laboratory (e.g. offer to train personnel at inexperienced laboratories, host a training workshop, propose a planning teleconference).

An invitation to interested interlaboratory participants shall be circulated well in advance of the interlaboratory method validation trial launch. It is recommended that the lead laboratory organizing the interlaboratory program circulate the invitation to potential participating laboratories five months prior to the start of the first round of interlaboratory testing. Laboratories should be asked to express their interest in participating within four weeks of receiving the announcement of the interlaboratory trial.

If a full validation of a method is already present, the method may be adopted without a further interlaboratory method validation trial.

5.5.3 Samples

Sample matrices shall reflect the scope of the test method (see 5.2). No further information about the expected values of estimated parameters should be given to participants. Every participant has to perform the biotest in a number of replicates specified by the interlaboratory method validation trial procedure. Each laboratory needs to receive a sufficient quantity of sample or subsamples. Samples shall be clearly labelled (e.g. number of sample or subsample, participant name, matrix, date).

If the determination of the measurement trueness is included in the validation trial a true measurement value is required. Such a true value is of theoretical nature and can only be determined by using a suitable reference material, by a reference to another, validated measurement method or by the preparation of a known reference sample. In the latter case, the environmental matrix to be assessed, for example, surface water or an eluate, is spiked with a defined amount of a reference compound with a known toxic effect potency for the method under investigation. Both the spiked and the respective un-spiked samples shall be included in the interlaboratory method validation trial where the un-spiked

sample should produce no effect. Alternatively, synthetic or artificial environmental matrices such as synthetic sewage (see Reference [61]) can be used to produce a defined no-effect sample. In either case, the lead laboratory should verify the correct spiking and the stability of the reference compound in the environmental matrix by a pre-test.

Information on the stability and homogeneity of subsamples shall be provided along with the final date for testing initiation. Storage conditions (e.g. temperature range, protection from light) are important points to be outlined in the interlaboratory method validation trial procedure. Compliance with the critical conditions needs to be considered by the lead laboratory coordinator when shipping the samples to other countries. It is the responsibility of the organizer to choose a proper way of sample dissemination or distribution.

If the concentrations of a chemical substance are estimated within the biotest, standard solutions for checking calibration and additional information can be provided by the lead laboratory to help laboratory participants establish and conduct the biotest in their laboratories.

The lead laboratory should request that all participating laboratories ship a subsample of the test media to a common analytical laboratory under contract to the lead laboratory, for chemical confirmation of the nominal test concentration when a reference compound is used for the standard validation round.

5.5.4 Experimental design

Ecotoxicological methods allow different experimental designs which can lead to different types of statistical calculation for test endpoint estimation [e.g. LOEC (lowest observed effect concentration), ECx (effect concentration), combined approach), LID (lowest ineffective dilution) or a biological equivalence concentration (BEQ)]. The lead laboratory shall distribute detailed information about the experimental design (e.g. dilution series, number of test dilutions and controls, number of replicates per treatment) to each participating laboratory. An estimation of ECx is preferred to NOEC (no observed effect concentration) or LOEC. A statistician should be consulted at this stage but likely the lead laboratory conducts this consultation.

5.5.5 Supporting information for participants

The interlaboratory method validation trial procedure should include the following information:

— strict adherence to the validation study procedure (e.g. standard operating procedure) for the specific interlaboratory method validation trial round;

— instructions for safety precautions since potentially present contaminants in the sample can pose a risk to the laboratory staff;

— pretreatment and storage conditions of samples;

— additional instructions for performance of analysis, if necessary;

— description of the test system, exposure conditions, concentration/dilutions selection procedures, estimated variables;

— contact person (from the lead laboratory);

— schedule (deadlines for start of analysis and return of results);

— form for results of analyses (unit, number of significant digits, etc.);

— questionnaire on experimental details, especially when testing a multi-options protocol;

— list of data to be submitted, which should include: blank values, calibration data, original experimental data, culture organism health records, temperature logs, procedures used to calculate and express results, the use of controls and other performance checks or measures of validity criteria;

— a spreadsheet for recalculation of raw data into the final output, if relevant (e.g. for calculation of microbial activities from estimated concentrations).

5.5.6 Data analysis and statistical evaluation of interlaboratory testing

All results have to meet each criterion of test validity proposed. Before statistical evaluation, the results from the participants shall also be checked for erroneous data and explainable outlying datasets. If necessary, a laboratory participant may be asked to check for errors in the transferred data.

Statistical evaluation of interlaboratory testing consists of two steps – evaluation of the individual biological observations or calculated test endpoints (e.g. calculation of microbial activity or ECx), and calculation of intra-/interlaboratory variability. Participants submit the raw data to the lead laboratory, who assures its evaluation. If intended, individual calculations performed by participating laboratories can also be submitted. It should be noted that even where the same statistical procedure (e.g. logistic regression) is used, there can be slightly different results depending on the options set for calculation and the software used. Flowcharts of dose-response modelling for estimation of NOEC/LOEC and ECx can be found in ISO/TS 20281, Reference [64] and Reference [68].

For continuous data (e.g. length, weight), the statistical evaluation should be performed according to ISO 5725-2, which describes the analysis of data consistency and outlier detection, as well as the calculation of reproducibility and repeatability. If only one test is available per laboratory, repeatability cannot be calculated according to ISO 5725-2. In this case the repeatability can be calculated according to Annex B. The decision on withdrawing outlier test outcome or laboratories from interlaboratory data analysis should be based on statistical expert opinion or a discussion with the laboratory staff involved in testing. If data do not meet criteria given by ISO 5725-2, robust statistics described in ISO 5725-5 can be more suitable. Statistical procedures for small numbers of participants are given in ISO 13528:2015, Annex D. Results reported as "less than" values are not valid and have to be excluded from the statistical evaluation of test variability.

If data does not comply with a normal distribution [counts (e.g. number of juveniles) and binomial data (e.g. if an organism is alive or dead)], appropriate statistical calculations should be applied for the estimation of standard deviation and to meet the assumptions required by the statistical technique used. In some cases, transformation (e.g. arcsin for percentage data or data in log space) results in normalization of the data, which enables the use of standard statistical procedures. See Annex A for further information. Alternatively, formulas for calculation of variability of data from a given distribution (e.g. Poisson or binomial) can be used. Statistical procedures for evaluation of ecotoxicity data can be found in references ISO/TS 20281, Reference [64], Reference [68] and Reference [69]. It is strongly advised that a statistician be consulted when challenging datasets are submitted.

If estimated, the measurement trueness is expressed as a bias, i.e. a systematic error in terms of a percentage deviation from the true measurement value. Positive values reflect an overestimation and negative values reflect an underestimation of the true value by the measurement method. Under consideration of the variability of the measurement, it can be determined if an observed bias between the average of a large series of test results, for example, from a spiked sample, and the true value is statistically significant (e.g. by analyses of variance and corresponding post hoc tests).

Control data from the method validation round can be used to confirm or adjust, if necessary, the validity criteria.

It should be noted that this guideline cannot cover all types of data obtained using biotests. The lead laboratory is responsible for such evaluation, which characterizes interlaboratory variability.

6 Assessment

A minimum of 6 valid datasets (in accordance with 5.5.2) should be considered to estimate test repeatability and/or reproducibility of the test method. It is acknowledged that different biotests can lead to differing acceptable variation coefficients of reproducibility. Based on the prevalidation step and on the test method optimization outcome, the lead laboratory and participating experts can set the acceptable values for the coefficient of variation (CV). In any case, CV ≤ 30 % is a commonly stated

target of acceptable biotest variability (see Reference [68]). Nevertheless, some methods are inherently more variable and exceed a CV of 30 %, in which case an explanation shall be provided.

If variability cannot be expressed as CV (e.g. LOEC), the lead laboratory should specify the criteria for acceptance of the results for a given biotest. Regardless of the type of data, the results of the biotest have to comply with the validity criteria established in the method.

If included in the interlaboratory method validation trial, the measurement trueness of the method has to be assessed as well. If there is no significant bias (see 5.5.6) associated with the measurement, the respective method produces true results, i.e. shows an acceptable measurement trueness. If there is a significant bias, the acceptance of the measured values (results) has to be discussed by the expert group involved in the development of the method. In case of acceptance, a justification has to be included in the report of the validation data.

If the defined criteria are not fulfilled, the test procedure is not fit for purpose. It should be checked if those can be revised or the method should be abandoned. The reason should be identified and, if necessary, the draft method should be revised. This revision of the experimental procedure can necessitate further interlaboratory trials.

7 Documentation and reporting

A report which includes graphical and tabular presentations of all received results (which are generally made anonymous) is issued and sent to participants electronically, preferably as in PDF format. This report should specify for which type of sample the method has been validated.

Upon completion of the statistical evaluation a certificate of participation showing the laboratory´s results and the overall means of results for each sample should be sent to each participant.

Since it is impossible to fully standardize some properties of a biological testing system (e.g. artificial soil in ecotoxicity testing), the participating laboratories should store all relevant information from the interlaboratory trials (e.g. supplier, LOT number, if available).

If the determination of the measurement trueness is included in the interlaboratory method validation trial, information about the reference compound(s) used for the spiked sample, i.e. compound name, CAS-Nr., effect potency (e.g. EC50), spiked concentrations, stability of the spiked sample and if possible an analytical verification of the spike-level shall be documented.

All experimental data and information received from laboratory participants shall be archived for a minimum of 5 years.

For information, a summary of interlaboratory trials performed within the validation of terrestrial biotests can be found in Annex H and of water biotests in Annex I.

Annex A
(informative)

Determination of accuracy (trueness and precision) in case the results of ecotoxicity tests are expressed as toxicity metrics

A.1 General

This annex describes a method to determine the accuracy of a test method in terms of trueness (if applicable) and precision in case the results of the ecotoxicity test are expressed as toxicity metrics.

A.2 Background

In the context of environmental risk assessments, toxicity metrics are derived from ecotoxicity tests which were developed in consideration of either the toxicant's mode of action, the target environmental compartment or how toxicants, waste or waste water reach the environment. Toxicity metrics (TMs) include effect concentrations (EC), lethal concentrations (LC), no- and lowest observed effect concentrations (NOEC/LOEC) as well as least-ineffective dilutions (LID) of e.g. waste water. LC50, for example, gives the concentration which causes 50 % mortality of the test cohort, EC20 points to the effect strength that evokes a 20 % decrease in a biological variable, such as growth rate, body weight or length, and LID 8 says that the ineffective concentration of, for example, waste water was reached at the dilution step 8. Common to all the aforementioned TMs is that they are derived on a logarithmic scale (log concentration) using appropriate statistical methods. In dilution tests, the dilution steps follow a geometric pattern (e.g. 1:2, 1:4, 1:8, 1:16). Consequently, the TMs are assumed to be log-normal distributed. In addition, a TM is a statistical summary measure of a biotest comprising measurements of several treatments, including untreated or reference controls and dilutions.

According to ISO 5725-2, the accuracy of a test method is determined by its trueness and its precision. The trueness denotes the degree of deviation of a measured value from a reference value, whereas the precision is a measure of variability. The observed variability includes two components: the intralaboratory variability, s_r, called repeatability (combined from all laboratories), and the interlaboratory variability, s_L. Both variance components sum up to the so-called reproducibility variance according to Formula (A.1):

$$s_R^2 = s_L^2 + s_r^2 \tag{A.1}$$

where

s_R^2 is the reproducibility variance;

s_L^2 is the interlaboratory variability;

s_r^2 is the intralaboratory variability.

Together with the overall mean (m) i.e., the mean of all laboratory means, the reproducibility standard deviation $\sqrt{s_R^2}$ forms the basis for all further interlaboratory method validation trial statistics, as amply described, for example, in ISO 5725-2. The computation s_r and s_L requires normally distributed data and thus, is not appropriate for toxic metrics showing lognormal distribution. Therefore, ISO 5725-2 cannot be used without adaptations when applied to toxicity metrics. This has been recognized in past interlaboratory method validation trials (e.g References [67] and [68]), although normative advice on how to handle lognormal distributed toxic metrics in ecotoxicity testing is still missing.

An additional problem is that ecotoxicity laboratories often do not perform under repeatability conditions due to the time and costs required for ecotoxicity testing. In such cases, when tests are only repeated once, a direct computation of s_r is impossible and thus, cannot be used to determine S_R. Donnevert et al. 2009[67] propose to estimate s_r using the confidence limits of the toxic metric (see Annex B). In a strict sense, this estimate is not obtained under repeatability conditions either, but it can be argued that repeated measures of the biological variable in each treatment/dilution contribute to variability within the confidence limits. In any case, this approach achieves a better estimate of S_R.

A.3 Procedure

A.3.1 General

The steps to conduct as interlaboratory method validation trial with toxic metrics are as follows:

— collect or calculate toxic metrics;

— perform log-transformation;

— compute interlaboratory method validation trial statistics;

— perform back transformation;

— present and report interlaboratory method validation trial results.

A.3.2 Collection or calculation of toxic metrics

Ideally, the statistical analysis of the ecotoxicity tests in the participating laboratories can be trusted and the interlaboratory method validation trial statistician only needs to collect the toxic metrics from repeated tests of all laboratories. In reality, however, laboratories often use various methods or software to determine toxic metrics, in particular if concentration-response models are used to determine ECx values (x: 10, 20,50, …). Therefore, it is advisable that the statistician determine the toxic metrics from raw data of each laboratory using the same statistical method and valid statistical software.

A.3.3 Performing log-transformation

The toxic metrics of each experiment under repeatability conditions from each laboratory are transformed to natural logarithms [$y' = \log_e(y)$; y: ECx, LCx, NOEC, LID, …].

A.3.4 Computing interlaboratory method validation trial statistics

In order to distinguish between log and normal scale, the statistics calculated on log scale are denoted with Greek letters. For the calculation of the overall mean, μ, and the repeatability, σ_r, and the interlaboratory variability, σ_L, use the log-values (y'). The overall arithmetic mean, μ, is then simply the geometric mean. All further statistics, e.g. Mandels-h and –k statistics, Grubbs test, Cochran test, as described in ISO 5725-2, are based on μ, σ_r, σ_L and σ_R.

A.3.5 Performing back transformation

For reporting the interlaboratory method validation trial results, statistics of toxic metrics have to be back transformed to original scale. The procedure for back transformation of μ and σ is summarized in Table A.1 (replace σ by σ_r, σ_L and σ_R if appropriate) giving the statistics on original scale.

Table A.1 — Formulae for the back transformation

Geometric mean, *m*	**EXP (μ)**
95 %-confidence limits	$\mathrm{EXP}\left(\mu \pm \frac{\sigma}{\sqrt{n}} \times t_{(0{,}025;\ \mathrm{df})}\right)$ with t: one-sided 2,5 %-quantile of the Student t-distribution and df = n-1
95 % and 99 % tolerance limits [a]	$\mathrm{EXP}(\mu + \sigma \times z)$ with z = 1,96 [b] for the 95 %- and z = 2,57 for the 99 %-tolerance limits
Expected value *EV(X)* [c]	$\mathrm{EXP}\left(\mu + \frac{1}{2} \times \sigma^2\right)$
Standard deviation, *sd(X)*	$EV(X) \times \sqrt{e^{\sigma 2} - 1}$
Coefficient of variation CV %	$\frac{sd(X)}{EV(X)} \times 100$

[a] Synonym “prediction interval” or “warning limits”.

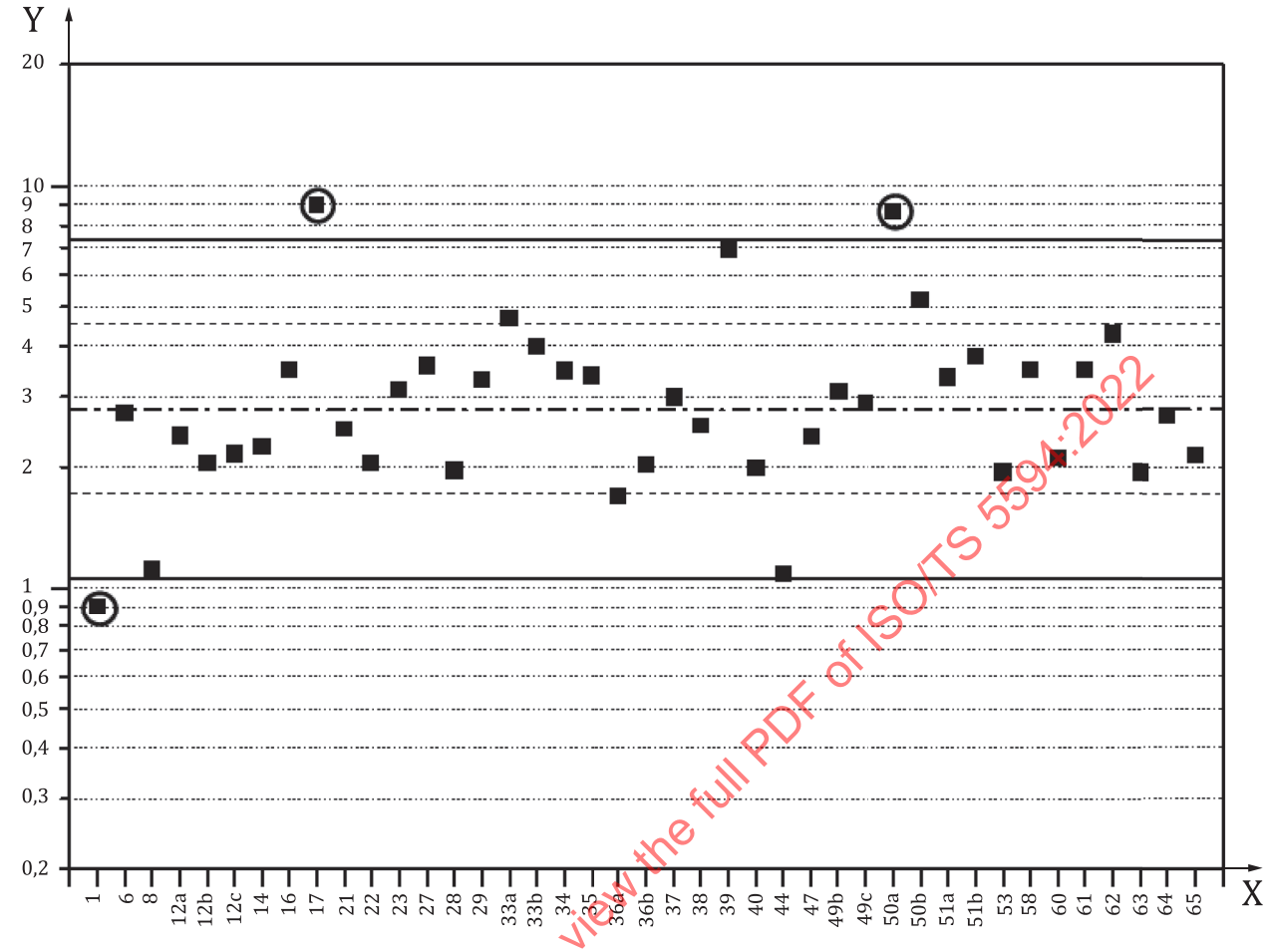
[b] In Canada and the US rounded to 2,0 (see Figure A.1).

[c] Expected value: location parameter of a probability distribution, i.e. the expected midpoint of the distribution to which the results aspire with increasing size of an experiment and increasing number of experiments. In normally distributed variables EV is identical to the arithmetic mean whereas in lognormally distributed variables the EV is always greater than the arithmetic mean.

The expected value (EV) is important for back transforming the standard deviation and the computation of the coefficients of variation.

A.4 Reporting

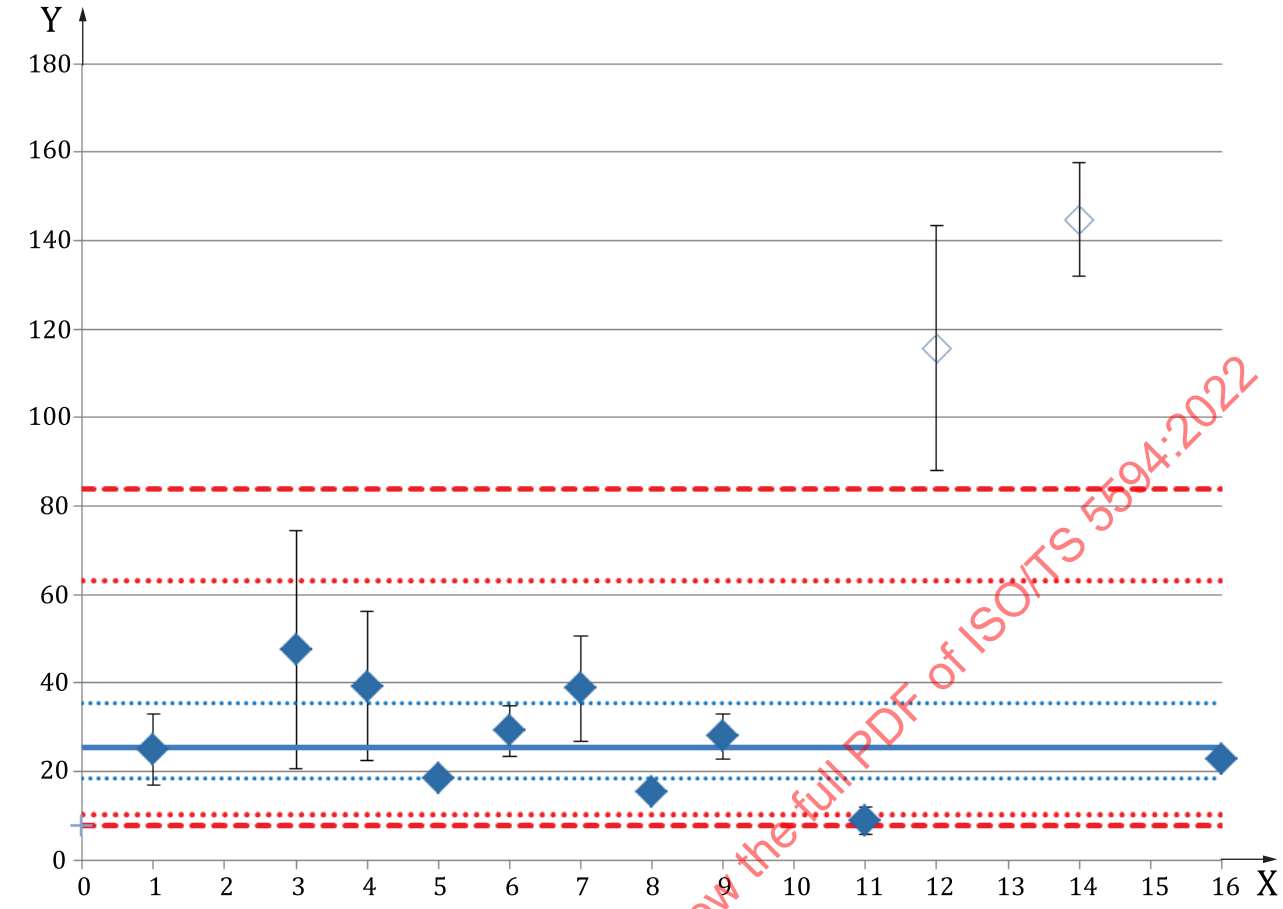
The precision measures s_r, s_L and s_R normally have to be given on original scale. The graphical representation of the interlaboratory method validation trial results can be performed on log scale (Figure A.1) or on original scale although the latter is preferred since it is more illustrative for the reader (Figure A.2). In the first case, σ_R forms the basis to calculate the shown tolerance limits, in the second case it is s_R. Please note that the tolerance limits are symmetrical on log scale but become asymmetrical on original scale due to the back transformation.



Key

X	lab code
Y	dilution [% of INC eluates]
■	laboratory specific values
◉	outlier
— · —	interlaboratory geometric mean
- - -	geometric mean ± 1 × *sd*
—	upper and lower warning limit [geometric mean ± 2 × *sd*]; 95 %-warning limits (the standard normal variable $z = 1{,}96$ is rounded to $Z = 2{,}0$)

Figure A.1 — Representation of interlaboratory method validation trial results of the EC50 of acute Daphnia tests as applied to test dilutions of a leachate of the waste material INC



Key

X	laboratory
Y	EC50 of reference (17 β-Estradiol [ng/l])
◆	laboratory specific expected values with standard deviation (whisker)
◇	outliers (expected values with standard deviation)
▬	interlaboratory geometric mean
····	95 % confidence interval of the interlaboratory geometric mean
···	95 % tolerance range of the interlaboratory geometric mean
– –	99 % tolerance range of the interlaboratory geometric mean

Figure A.2 — Representation of interlaboratory method validation trial results of the EC50 of the YES-Tests[70]

Annex B
(informative)

Estimation of the repeatability standard deviation

B.1 General

This annex describes a method to estimate the repeatability standard deviation if the interlaboratory trial was not performed under repeatability conditions.

B.2 Background

Donnevert et al. 2009[67] propose to obtain an estimate of the repeatability standard deviation by using the confidence limits of the toxicity metric (TM). They argue that repeated measures of the biological variable in each treatment/dilution contribute to variability within the confidence limits. Thus, the 95 %-confidence limits of toxicity metrics, e.g. the EC50/LC50, give an estimate of the internal variability as caused by the repeated preparation of the treatments/dilutions of the toxicity test. It can therefore be considered an approximation of the repeatability conditions.

B.3 Procedure

The repeatability standard deviation of toxicity metrics is normally calculated on the log-scale since the toxicity metrics are log-normally distributed (see Annex A). Therefore, the basis for estimating the repeatability standard deviation is Formula (B.1): for confidence limits:

$$\mathrm{EXP}(T_\mathrm{M} \pm s \times t_{(0{,}025;\ \mathrm{df})}) \tag{B.1}$$

where

T_M is the toxicity metric;

s is the standard error of the toxicity metric;

t is the one-sided 2,5 %-quantile of the Student t-distribution and df = n-1.

In most cases $t_{(0.025;\ \mathrm{df})}$ can be rounded off to 2; hence Donnevert et al. 2009[67] propose the simplified formulae:

$$\mathrm{EXP}(T_\mathrm{M} \pm s \times 2) \tag{B.2}$$

$$C_\mathrm{UL} = T_\mathrm{M} + s \times 2 \tag{B.3}$$

$$C_\mathrm{LL} = T_\mathrm{M} - s \times 2 \tag{B.4}$$

where

C_UL is the log upper 95 %-confidence limit;

C_LL is the log lower 95 %-confidence limit.

The difference is as follows:

$$C_{UL} - C_{LL} = T_M + s \times 2 - (T_M - s \times 2) = 4 \times s \quad \text{(B.5)}$$

Solve the formula for s as an estimate of s'_r:

$$s \cong s'_r = (C_{UL} - C_{LL})/4 \quad \text{(B.6)}$$

And apply to the formula to calculate the reproducibility variance, s_R^2:

$$s_R^2 = s_L^2 + s_r'^2 \quad \text{(B.7)}$$

Even though $s_r'^2$ does not represent the whole variability, s_r, under true repeatability conditions, it still provides a better estimate of the reproducibility variance, s_R^2 than with $s_r'^2$ omitted from Formula (B.7).

For reporting, $\sqrt{s_R{}^2}$ is back transformed to the linear scale according to Annex A.

Annex C
(informative)

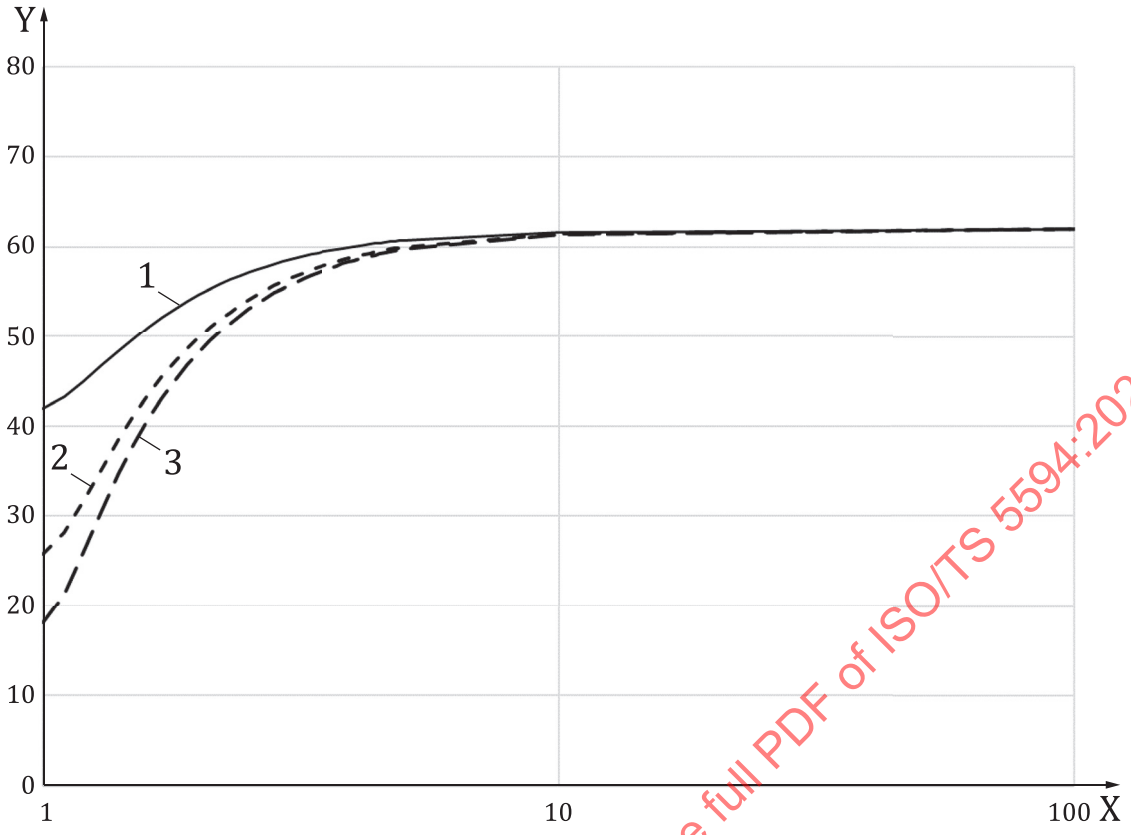
Calculation and reporting of uncertainty of s_R

C.1 General

This annex provides guidance on how to proceed if less than six valid datasets are obtained in line with the interlaboratory method validation trial by calculating and reporting the uncertainty of the reproducibility variance s_R.

C.2 Background

The ISO 5725 series describes the theoretical background to determine the number of laboratories required for an estimation of precision. The aim is to balance a reduction in uncertainty of the reproducibility variance s_R^2 with resources available for the interlaboratory method validation trial. ISO 5725-1:1994, Formula (10) describes the dependency of s_R on the parameters p representing the number of laboratories providing valid data sets, n representing the number of replicas for one sample within each laboratory and γ representing the ratio of the reproducibility standard deviation to the repeatability standard deviation. ISO 5725-1:1994, Annex B provides a chart showing the percentage uncertainty of s_R against the number of laboratories p for various levels of n and γ. Increasing the number of laboratories and number of replicas decreases the uncertainty of s_R whereas the uncertainty of s_R increases with increasing γ. This is illustrated in Figure C.1 for $p = 6$.



Key

X γ describing the ratio of the reproducibility standard deviation to the repeatability standard deviation

Y percentage uncertainty of s_R

1 $n = 2$

2 $n = 5$

3 $n = 10$

Figure C.1 — Dependency of the uncertainty of s_R on γ for $p = 6$

From Figure C.1 it becomes evident that the percentage uncertainty of s_R raises to a maximal value independent from n. Thus, the maximal possible uncertainty in s_R can be estimated for each p. In case of $p = 6$ the maximal percentage uncertainty in s_R is 61,98 % (rounded to two decimals).

C.3 Procedure

If less than six valid datasets are available for the method validation, the uncertainty in s_R has to be calculated according to ISO 5725-1:1994, Formula (10). If the percentage uncertainty in s_R is below 61,98 %, the method validation can be accepted even though the number of valid data sets is not met. The values for p, n, γ and uncertainty in s_R have to be reported and a justification for the acceptance of the method has to be provided.

Annex D
(informative)

Validation schema

Figure D.1 shows the overall schema of the validation process.

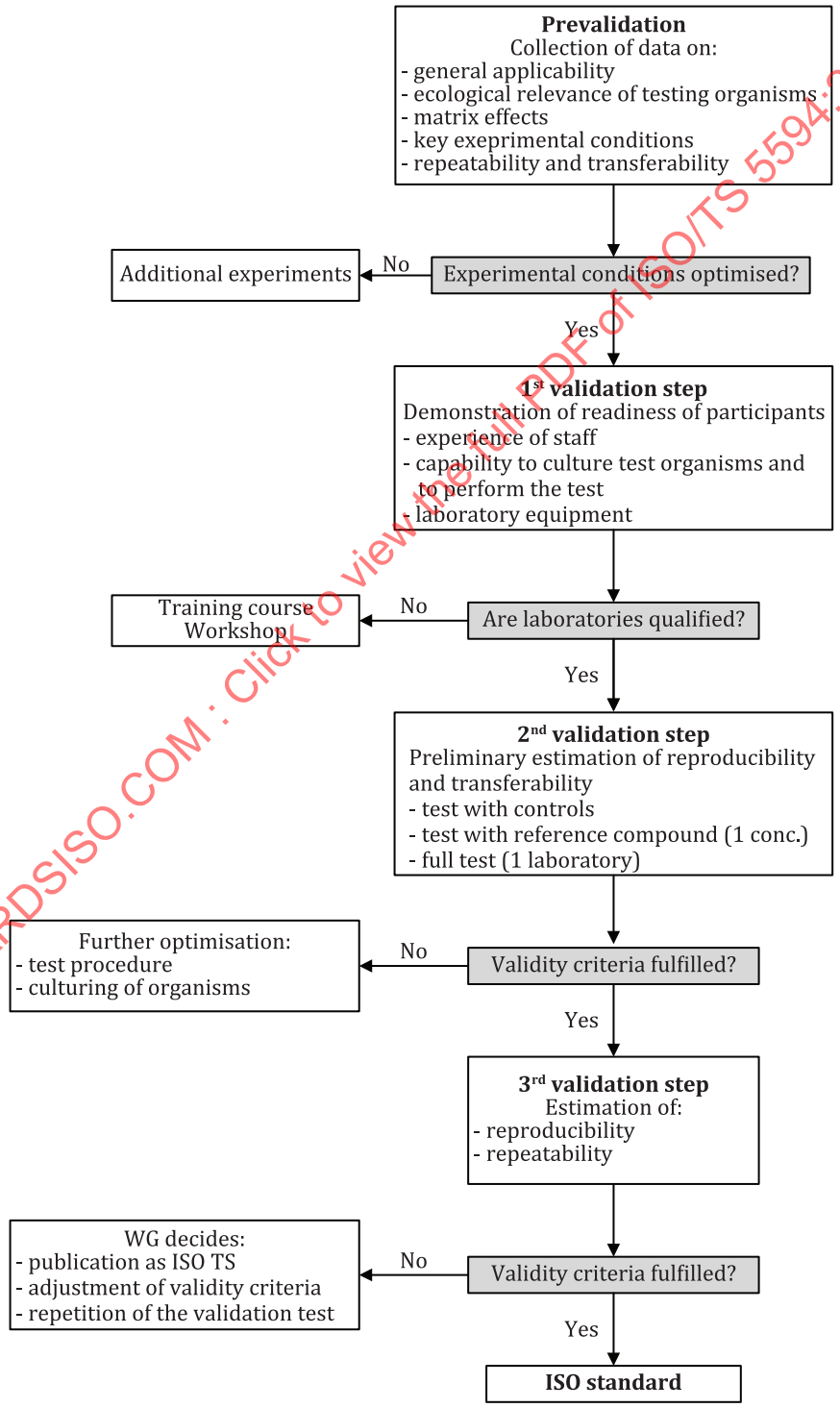


Figure D.1 — Validation schema

Annex E
(informative)

Questionnaire (example)

Information for participation in international interlaboratory trial for validation

Soil/Water quality – The name of a test

Contact information:

Laboratory main contact:	Name:
	Email address:
	Telephone number:
Validation study leader (if different from main contact)	Name:
	Email address:
	Telephone number:
Full mailing address (supplies will be mailed here)	Attn:

Test organisms:

Characterization of breeding stock:	(yes/no)	Description:
Breeding stock with verified taxonomy		
Breeding substrate		
Food		

Equipment on site checklist:

Type of equipment:	Do you have access to this equipment? (yes/no)	Model and specification (parameters are listed):
Analytical instrument (e.g. spectrophotometer, chromatograph)		
Autoclave		
pH meter		
Oven set at 105 °C		
Analytical balance		
Top-loading balance		
Appropriate glassware (volume, material)		
Enclosure, capable of being controlled to a temperature of (20 ± 2) °C		
Light source, capable of delivering a constant light intensity of 400 lx to 800 lx at the substrate surface at a controlled light: dark cycle of between 12 h:12 h and 16 h:8 h		

Type of equipment:	Do you have access to this equipment? (yes/no)	Model and specification (parameters are listed):
Phytotron, plant growth room, green house, climatic chamber (growth of higher plants)		

Equipment for extraction and counting of juveniles:

Water bath (50 °C to 60 °C) or 2 sieves (0,5 mm)		
Containers for counting (photographing) of juveniles		
Photographic equipment for close-up photo of water surface		

Annex F
(informative)

Specific features of terrestrial biotests which are recommended to be addressed by lead laboratories

Table F.1 summarizes the specific features which are recommended to be addressed by lead laboratories for terrestrial biotests.

Table F.1 — Features of terrestrial biotests

	Criteria or unique features to be specifically addressed	Tests targeting individual organisms/responses	Methods targeting microbial community responses	Genomic/Molecular biology methods
	Examples of standards	ISO 11267 ISO 11268-2 ISO 11269-2 ISO 16387 ISO 18187 ISO 21285 ISO 22030	ISO 14240-2 ISO 14238 ISO 15685 ISO 16072 ISO 20130	ISO 11063 ISO 17601
Scope	Specific additions required?	The sample types for which the method has been demonstrated to be appropriate should be included in the scope. Information on which sample types the validation has been performed should be specified in the scope. Moreover, the scope should mention the known limitations of the method at the date of publication of the standard.		
Pre-validation	Specific interfering factors to consider?	pH soil organic matter texture nutrients (plants) water holding capacity colour of "soil extracts"		Humic acid substances are often co-extracted with nucleic acid substances and interfere with Taq polymerase thereby impairing the quantification of the abundance of microbial groups; also heavy metals and other substances interfere with PCR.

Table F.1 *(continued)*

	Criteria or unique features to be specifically addressed	Tests targeting individual organisms/responses	Methods targeting microbial community responses	Genomic/Molecular biology methods
	Specific representative samples?	To consider types of samples for which a method is intended with respect to ecological valence of a tested organism. Samples should cover as much as possible the range of typical soil properties (pH, texture, soil organic matter content) and typical soil types (sandy, clayey).	Samples should cover the range of typical soil properties (pH, texture, soil organic matter content). If the method is to be used also for organic horizons of forest soil, a representative sample should be included too.	Samples should cover the range of typical soil properties (pH, texture, soil organic matter content). If the method is to be used also for organic horizons of forest soil, a representative sample should be included too. Sampling strategies should take spatial and temporal variabilities into account.
1st step	Specific criteria for assessing labs readiness?	Equipment for culturing of test organisms and performing the test (e.g. control of temperature, humidity, light conditions), some experience with testing using similar organisms. EC50 of reference compound(s).	Instruments are key factors and their specifications should be checked if there are requirements not common for commercial products.	Enzymes (Taq polymerase), primers (specific for a given microbial guilds) and instruments (disruptor, fluorimeter, spectrophotometer, qPCR machine) should be carefully checked.
2nd step	Number of labs at the 2nd validation step, and selection criteria for integrating the 3rd step.	Fulfilment of validation criteria concerning mainly the minimum number of offspring in controls indicates readiness for testing. Successful breeding of cultures is a key factor.	No specific criteria are mostly required, measurement of a calibration curve will be helpful in some cases.	A calibration has to be established for the quantification of soil DNA extracts and for the quantification of the microbial groups by qPCR.
3rd step	Any particular suggestion?	Control soil (e.g. artificial soil) is a critical factor having an impact on final results. It is highly recommended to distribute control soil for preparation of dilution series to filter out these effects.	Storage of soil (freezing, cooling, drying) is critical and requirements differ among methods depending mainly on physiology. For example, drying will have much smaller impact on activity of exoenzymes than on respiration where such treatment has to be avoided.	Soil DNA can be extracted from fresh samples or from samples kept frozen at -20 ° (DNA extraction) or -80 °C (RNA extraction).
Labs	Minimum number of valid datasets?	6	6	6
	Minimum number of participating countries?	3	3	3
	Participation of unexperienced labs acceptable?	Yes	Yes	Yes

Table F.1 *(continued)*

	Criteria or unique features to be specifically addressed	Tests targeting individual organisms/responses	Methods targeting microbial community responses	Genomic/Molecular biology methods
Data analysis	2-step statistical evaluation applicable: (i) biological parameters/endpoints; (ii) variability?	Yes	Yes. Many methods generate normally distributed data therefore evaluation can be done according to ISO 5725-2 in these cases.	Yes
	Trueness assessment?	No	No	
	Point estimates (ECx, LID, LOEC, MIC, BEQ)?	NOEC, LOEC, Ecx.	Activity.	DNA yield, sequence number per ng of DNA.
	Repeatability analysis applicable?	It depends on duration of the test. Repeatability can be estimated for short-term tests. Surrogate estimation of repeatability using confidence limits of toxicity metric is suggested.	Yes	Yes on biological repeats.
	Reproducibility assessment based on CV ≤ 30 % or other metrics?	Yes	Yes	Yes
Other	Terrestrial plant tests.	The draft standard under validation should provide soil nutrient data and, if available, the nutrient needs of the plant species used in the interlaboratory testing rounds.		Microbial diversity assessment using high throughput sequencing is largely used by microbial ecologist but not yet standardized because the methods (both sequencing and bioinformatics) are under constant evolution. Similarly metagenomic analysis relying on direct sequencing of environmental DNA (without PCR step) is not yet standardized for the same reasons. As in most cases kits are used for most purposes of nucleic acid based analysis (and compounds of kits change often without notice) it is essential to define MOCK communities which should be included in any analysis of microbial diversity.

Annex G
(informative)

Specific features of water biotests which are recommended to be addressed by lead laboratories

Table G.1 summarizes the specific features which are recommended to be addressed by lead laboratories for water biotests.

Table G.1 — Features of water biotests

	Criteria or unique features to be specifically addressed	In vitro tests	Tests targeting individual organisms/responses	Methods targeting microbial community responses
	Examples of standards	The ISO 19040 series ISO 11350 ISO 13829 ISO 21427-2	ISO 15088 ISO 7346-1 ISO 16303 ISO 10872 ISO 6341	ISO 13641-1 ISO 9509
Pre-validation	Specific interfering factors to consider?	acute toxic effects masking specific effects conductivity turbidity, colour pH	conductivity pH oxygen level ammonia	pH Anaerobic methods are subject to error from oxygen contamination. This interference is minimized by the use of strictly anaerobic handling techniques.
	Specific representative samples?	Focus waste water assessment, range of DOC, turbidity.	Focus waste water assessment, range of DOC, turbidity.	Focus waste water assessment, specific requirements for biological activity of sludge.
1st step	Specific criteria for assessing labs readiness?	EC50 of a reference compound, LOQ of reference compound, (positive controls) and negative controls	EC50 of a reference compound, (positive controls) and negative controls	EC50 of a reference compound, (positive controls) and negative controls
2nd step	Number of labs at the 2nd validation step, and selection criteria for integrating the 3rd step	Not mandatory, suggestion: proof of transferability with minimum three labs.	Not mandatory, suggestion: proof of transferability with minimum three labs.	Not mandatory, suggestion: proof of transferability with minimum three labs.
3rd step	Any particular suggestion?	Sample storage and stability, small pre-ringtest, initial characterization by lead laboratory.	Sample storage and stability, small pre-ringtest, initial characterization by lead laboratory if test is not too time demanding.	Age of sludges < 24 h after sampling, sludges grown in laboratory can be used. Storage of samples < 4 °C, testing as soon as possible, reference to ISO 5667-16.

Table G.1 *(continued)*

	Criteria or unique features to be specifically addressed	In vitro tests	Tests targeting individual organisms/responses	Methods targeting microbial community responses
Labs	Minimum number of valid datasets?	6, Guidance by "Guide to Method Validation for Quantitative Analysis in Chemical Testing Laboratories" by the Irish National Accreditation Board and ISO 5725-1:1994, 6.3.4 suggests 8 valid data sets.	6, Guidance by "Guide to Method Validation for Quantitative Analysis in Chemical Testing Laboratories" by the Irish National Accreditation Board and ISO 5725-1:1994, 6.3.4 suggests 8 valid data sets.	6, Guidance by "Guide to Method Validation for Quantitative Analysis in Chemical Testing Laboratories" by the Irish National Accreditation Board and ISO 5725-1:1994, 6.3.4 suggests 8 valid data sets.
	Minimum number of participating countries?	3	3	3
	Participation of unexperienced labs acceptable?	Yes (experience of lab reported)	Yes (experience of lab reported)	Yes (experience of lab reported)
Data analysis	2-step statistical evaluation applicable: (i) biological parameters/endpoints; (ii) variability?	Data in log-scale, transformation of data required, (for AMES arc-sin transformation).	Data in log-scale, transformation of data required.	Data in log-scale, transformation of data required.
	Trueness assessment?	Yes, in part via biological equivalence concentration.	No	No
	Point estimates (ECx, LID, LOEC, MIC, BEQ)?	ECx, LID, BEQ, induction factor	ECx, LCx, LID	ECx
	Repeatability analysis applicable?	Yes	Depending on test duration/effort.	Depending on test duration/effort.
	Reproducibility assessment based on CV CV ≤ 30 % or other metrics?	Yes (in general); consider dependency of CV and effect strength.	Yes (in general); consider dependency of CV and effect strength.	Reported variation on EC50 in ISO 9509 are in the range of 60 %.

Annex H
(informative)

Summary of interlaboratory trials performed within the validation of terrestrial biotests

Table H.1 summarizes the interlaboratory trials performed within the validation of terrestrial biotests.

Table H.1 — Interlaboratory trials performed within the validation of terrestrial biotests

Standard	Year	Organism	N. of lab.	Tested samples	Reference compound	Repetitions	Statistical evaluation	Comments
ISO 10872	2019	*Caenorhabditis elegans*	11	2 soils	Yes (BAC-C16) 15 mg/l	Yes (1-3 times); one lab 5 times	Significance of the observed inhibitions were statistically tested by Student t-test (one sided, alpha = 0,05) after testing for normal distribution (Shapiro-Wilk test) and variance homogeneity (Levene test). Obtained percentages of inhibitions for each variable and test soil were tested for outliers using the Hampel test.	Calculation of repeatability (intra-laboratory variability), inter-laboratory variability and overall variability (reproducibility)
ISO 11063	2009	Soil DNA	9	6 different soils collected in France (3), Germany (1), Finland (1), and Sweden (1)	-	5 aliquots from each soil were analysed by each lab (30 samples per lab)	The obtained percentages of inhibitions for each variable and test soil were tested for outliers using the Hampel test.	
ISO 11268-2	1995	*Eisenia fetida*	6	2 pesticides: Derosal® (a.i. carbendazim); E605 forte (a.i. parathion)	Yes (first reference compound)	No	No information	Two ringtests were performed in 1990/1991 and 1991/1992 in Germany.
ISO 11267	1995	*Folsomia candida*	10	2 pesticides: Betanal plus (a.i. phenmedipham); E605 forte (a.i. parathion)	Yes (first reference compound)	No	No information	A ringtest was also performed in 1993.
ISO 15952	2000	*Helix aspersa aspersa Müller (also known as Cantareus aspersus and Cornu aspersum)*	4	1 compound: Cadmium chloride	Yes	No	EC50 calculations (regtox)	Test performed in Canada, outside the framework of SC4/WG2.

Table H.1 *(continued)*

Standard	Year	Organism	N. of lab.	Tested samples	Reference compound	Repetitions	Statistical evaluation	Comments
ISO 17512-1		*Eisenia fetida*	4	1 compound: Boric acid (one soil Alberta Black Chernozem soil)	Yes	No	Environment Canada guidelines	Test performed in Canada, outside the framework of SC4/WG2.
ISO 17601	2013	Soil microorganisms (DNA)	6	6 different soils collected in France (2), Greece (2), Italy (1), and Portugal (1)	-	3 DNA extraction replicates per soil sample. qPCR assays performed in triplicate on 3 different plates by each partner. In total, each partner analysed 162 samples per target gene (i.e. 6 soils x 3 different extraction x 3 repeats per plate x 3 plates)	Student's t test, Fisher's F test, and Kruskal-Wallis to test for differences in the quantification of microbial sequences among labs and assess the repeatability of the qPCR assays.	
ISO 18187	2013-2014	*Arthrobacter globiformis*	5 to 9	4 soils and 4 wastes (plus 5 compounds)	Yes	No	Data consistency, repeatability and reproducibility analyzed according to ISO 5725- with some amendments (following Environment Canada guidelines).	
ISO 18763	2011	*Lepidium sativum Sinapis alba, Sorghum saccharatum*	28	1 compound: Boric acid (one concentration in artificial soil)	Yes	No	Repeatability and reproducibility of the interlaboratory comparison, according to the ISO 5725-2.	Test performed outside the framework of SC4/WG3.
ISO 20130	2016	Enzymatic activities	8	6 soils (2 grasslands and 4 culture soils)	Not relevant		The intra-laboratory and inter-laboratory coefficients of variation.	

Table H.1 *(continued)*

Standard	Year	Organism	N. of lab.	Tested samples	Reference compound	Repetitions	Statistical evaluation	Comments
ISO/TS 20131-1	2016	Denitrifying enzymes	4	4 soils (from a 100 km area around Orleans)	Yes (standard N_2O gas)	Yes (standard corrections considered in the repetition)	Non parametric statistics (*r* value). No information on data consistency analysis and dataset validation strategy/method.	
ISO/TS 20131-2	2016	N_2O reduction	3	4 soils (from a 100 km area around Orleans)	Yes (standard N_2O gas)	Yes (standard corrections considered in the repetition)	Non parametric statistics (r_{max} value). No information on data consistency analysis and dataset validation strategy/method.	
ISO 21286	2013	*Eisenia sp.* DNA	5	144 earthworm specimens supplied by 28 eco-toxicological test laboratories from 15 countries.	-	5 (each earthworm sample was divided into 5 pieces, one for each laboratory). All 144 individuals were analysed in at least two DNA barcoding laboratories. Replicate analyses of each sample at each participant laboratory were not performed.	None. Inter-laboratory variability was assessed as: 1) discrepancies and base calling differences in the DNA barcode sequences obtained [91 differences in 647,321 (0.00014) pairwise site comparisons] and 2) consistency of species identification (100 % agreement in the taxonomic assignment of the 144 specimens analysed).	
ISO 23265	2019-2022	Organic Matter Decomposition	14	1 compound: $AgNO_3$ and 3 soils	Yes	Yes (standard corrections considered in the repetition)	Data consistency, repeatability and reproducibility analyzed following Environment Canada guidelines)	Three rounds of inter-lab testing

Table H.1 *(continued)*

Standard	Year	Organism	N. of lab.	Tested samples	Reference compound	Repetitions	Statistical evaluation	Comments
ISO 23266	2020	Oribatid mite	9	1 compound: Boric acid and 4 soils	Yes	Yes (standard corrections considered in the repetition)	Data consistency, repeatability and reproducibility analyzed following Environment Canada guidelines)	Inter-lab results also used to validate the EC test method for mite reproduction inhibition (STB 1/RM/61).
OECD 220		*Enchytraeus albidus*	29	2 compounds: Derosal (a.i. carbendazim); 4-nitrophenol	Yes	No, but 2 test designs NOEC: ECx	NOEC and Ecx; comparison of approaches. The reproducibility of data derived from various laboratories was determined by using the h index, which is the standardized measure of the distance between the result of a single laboratory and the mean of the results from all laboratories (Weyers et al. (2002)).	Same method as ISO 16387. Test performed outside the framework of SC4/WG2.
OECD 226	2006	*Hypoaspis aculeifer*	12	2 compounds: Boric acid; dimethoate	Yes	No, but 2 test designs NOEC: ECx	ECx and NOEC; minimum detectable treatment effect (MDTE); minimum detectable increase in juveniles (MDJJ)	Same method as ISO 21285. Test performed outside the framework of SC4/WG2.

Table H.1 *(continued)*

Standard	Year	Organism	N. of lab.	Tested samples	Reference compound	Repetitions	Statistical evaluation	Comments
OECD 232	2006	*Folsomia candida; Folsomia fimetaria*	14	3 compounds: Boric acid; copper chloride; dimethoate	Yes	No	The inter-laboratory variability is evaluated by calculating h, the standardized difference of a toxicity test result observed for one laboratory from the mean toxicity values (Weyers et al., 2002). ANOVA, F-test for the comparison of LC50 of the two species.	Same method as ISO 11267. Test performed outside the framework of SC4/WG2.

Annex I
(informative)

Summary of interlaboratory trials performed within the validation of water biotests

Table I.1 summarizes the interlaboratory trials performed within the validation of water biotests.

Table I.1 — Interlaboratory trials performed within the validation of water biotests

Standard	Year	Organism	Effect	Type	N. of lab.	N. of samples	Tested samples, ref. compound	(Toxicity) measure	Precision data	Comments
ISO 10253	1989/ 1990, 2014 with cuvettes	Algae *Phaeodactylum tricornutum*	growth inhibition	in vivo	10, 2014: 11	2, 2014:1	$K_2Cr_2O_7$, 3,5-dichlorophenol, 2014: $K_2Cr_2O_7$	ErC50	mean, SD, CV, 2014: mean, sr, CV %, sR, CV %, UCL, LCL	BAUDO, R. Report on the International Interlaboratory Comparison of the Marine Algaltoxkit. 2015. In: Publications — Reports on the website https://www.microbiotests.com/
ISO 10710	2006	Macroalga *Ceramium tenuicorne*	growth inhibition	in vivo	6	2	$ZnSO_4 \cdot 7H_2O$, 3,5-dichlorophenol	ErC20, ErC50	mean, SD, CV	ISO/TC147/ SC5WG5 N219, Presentation of Ceramium Ring test results, by Britta Eklund, at the ISO/TC147/ SC5/WG5 meeting in Cape Town, Sept. 2006
ISO 10712	1989	*Pseudomonas putida*	growth inhibition	in vivo	21	1	3,5-DCP	EC10, EC50: turbidity (FNU)	mean, VCr	inhibition reproduction, turbidity (FNU)
ISO 11348-1	1999	*Vibrio fischeri*	Acute Tox	in vivo	10-15	4	3,5-DCP, zinc sulfate heptahydrate, potassium dichromate, cetyltrimethylammonium bromide	EC20, EC50: luminescence	mean, Sr, CVr, standard deviation of reproducibility	inhibition luminescence
ISO 11348-2	1999	*Vibrio fischeri*	Acute Tox	in vivo	14-20	4	3,5-DCP, zinc sulfate heptahydrate, potassium dichromate, cetyltrimethylammonium bromide	EC20, EC50: luminescence	mean, Sr, CVr, standard deviation of reproducibility;	luminescence inhibition