INTERNATIONAL STANDARD

ISO 5725-1

Second edition
2023-07

Accuracy (trueness and precision) of measurement methods and results —

Part 1:
General principles and definitions

Exactitude (justesse et fidélité) des résultats et méthodes de mesure —

Partie 1: Principes généraux et définitions

Reference number
ISO 5725-1:2023(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 document 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).

ISO draws attention to the possibility that the implementation of this document may involve the use of (a) patent(s). ISO takes no position concerning the evidence, validity or applicability of any claimed patent rights in respect thereof. As of the date of publication of this document, ISO had not received notice of (a) patent(s) which may be required to implement this document. However, implementers are cautioned that this may not represent the latest information, which may be obtained from the patent database available at www.iso.org/patents. ISO shall not be held responsible for identifying any or all such patent rights.

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 69, *Applications of statistical methods*, Subcommittee SC 6, *Measurement methods and results*.

This second edition of ISO 5725-1 cancels and replaces the first edition (ISO 5725-1:1994) which has been technically revised. It also incorporates the Technical Corrigendum ISO 5725-1:1994/Cor.1:1998.

The main changes are as follows:

— normative references have been revisited;

— some definitions have been deleted (observed value, cell in a precision experiment, collaborative assessment experiment) and others have been added (repeatability critical difference, reproducibility critical difference, intermediate precision conditions, intermediate precision standard deviation, intermediate precision critical difference, intermediate precision limit);

— the number of laboratories required for a precision study and Annex B presenting charts of uncertainties for precision measures have been moved in ISO 5725-2;

— guidance on the practical use of trueness and precision to evaluate uncertainty and the use of ISO 21748 was added.

A list of all parts in the ISO 5725 series can be found on the ISO website.

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

0.1 The general term accuracy is used in ISO 5725 (all parts) to refer to both trueness and precision.

The term accuracy was at one time used to cover only the one component now named trueness, but it became clear that to many persons it should imply the total displacement of a result from a reference value, due to random as well as systematic effects.

The term bias has been in use for statistical matters for a very long time, but because it caused certain philosophical objections among members of some professions (such as medical and legal practitioners), the positive aspect has been emphasized by the invention of the term trueness.

0.2 ISO 5725 (all parts) uses two terms "trueness" and "precision" to describe the accuracy of a measurement method. "Trueness" refers to the closeness of agreement between the arithmetic mean of a large number of test results and the true or accepted reference value. "Precision" refers to the closeness of agreement between test results obtained under stipulated conditions.

0.3 The need to consider "precision" arises because tests or measures performed on presumably identical test items in presumably identical circumstances do not, in general, yield identical results. This is attributed to unavoidable random errors inherent in every measurement procedure; the factors that influence the outcome of a measurement cannot all be completely controlled. In the practical interpretation of measurement data, this variability should be taken into account. For instance, the difference between a test result and some specified value may be within the scope of unavoidable random errors, in which case a real deviation from such a specified value has not been established. Similarly, comparing test results from two batches of product will not indicate a fundamental quality difference if the difference between them can be attributed to the inherent variation in the measurement procedure.

0.4 The general term for variability between replicate measurements is precision. Two conditions of precision, termed repeatability and reproducibility conditions, have been found necessary and, for many practical cases, useful for describing the variability of a measurement method. Under repeatability conditions, all factors that influence the measurement are considered constant and do not contribute to the variability, while under reproducibility conditions, some or all influential factors vary and do contribute to the variability of the test results. Thus repeatability and reproducibility are the two extremes of precision, the first describing the minimum and the second the maximum variability in results. Other intermediate conditions between these two extreme conditions also occur when one or more of the factors that influence the measurement are allowed to vary, and are used in certain specified circumstances. Precision is normally expressed in terms of standard deviations.

0.5 The purpose of ISO 5725 (all parts) is as follows:

a) to outline the general principles to be understood when assessing accuracy (trueness and precision) of measurement methods and results, and in applications, and to establish practical estimations of the various measures by experiment (ISO 5725-1);

b) to provide basic methods for estimating the two extreme measures of the precision of measurement methods by experiment, giving the circumstances in which they apply (ISO 5725-2);

c) to provide designs for obtaining intermediate measures of precision, giving the circumstances in which they apply and methods for estimating them and to provide some alternative designs to those given in ISO 5725-2, for determining the precision and trueness of measurement methods for use under certain circumstances (ISO 5725-3);

d) to provide basic methods for the determination of the trueness of a measurement method (ISO 5725-4);

e) to provide some alternatives to the methods, given in ISO 5725-2 to ISO 5725-4, for determining the precision and trueness of measurement methods for use under certain circumstances (ISO 5725-5);

f) to present some practical applications and use of these measures of trueness and precision (ISO 5725-6).

Accuracy (trueness and precision) of measurement methods and results —

Part 1: General principles and definitions

1 Scope

1.1 This document

— introduces conditions, constraints and resources necessary to evaluate a measurement method or a result;

— defines an organizational scheme for the acquisition of trueness and precision data by study;

— provides the necessary definitions, statistical model and principles for ISO 5725 (all parts).

— is not applicable to proficiency testing or production of the reference item that has their own standards (ISO 13528, respectively and ISO Guide 35).

1.2 This document is concerned exclusively with measurement methods which yield results on a continuous scale and give a single value as the test result, although this single value may be the outcome of a calculation from a set of observations.

It defines values which describe, in quantitative terms, the ability of a measurement method to give a true result (trueness) or to replicate a given result (precision). Thus, there is an implication that exactly the identical item is being measured, in exactly the same way, and that the measurement process is under control.

This document may be applied to a very wide range of test items, including gas, liquids, powders and solid objects, manufactured or naturally occurring, provided that due consideration is given to any heterogeneity of the test item.

This document does not include methods of calculation that are described in the other parts.

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 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*

3 Terms and definitions

For the purpose of this document, the terms and definitions given in ISO 3534-1, ISO 3534-2 and the following apply.

The symbols used in ISO 5725 (all parts) are given in Annex A.

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 https://www.electropedia.org/

3.1
test result
value of a characteristic obtained by carrying out a specified test method

Note 1 to entry: The test method should specify that one or a number of individual observations be made, and their average or another appropriate function (such as the median or the standard deviation) be reported as the test result. It can also require standard corrections to be applied, such as correction of gas volumes to standard temperature and pressure. Thus a test result can be a result calculated from several observed values. In the simple case, the test result is the observed value itself.

Note 2 to entry: When measurement is used (for methods or results) in this document it means test or measurement (for methods or results).

[SOURCE: ISO 3534-2:2006, 3.4.1, modified — Note 2 to entry rephrased.]

3.2
accepted reference value
value that serves as an agreed-upon reference for comparison, which is derived as:

a) a theoretical or established value, based on scientific principles;

b) an assigned or certified value, based on experimental work of some national or international organization;

c) a consensus or certified value, based on collaborative experimental work under the auspices of a scientific or engineering group;

d) the expectation, i.e. the mean of a specified population of measurements when a), b) and c) are not available.

[SOURCE: ISO 3534-2:2006, 3.2.7]

3.3
level
<test item> general average of the *test results* (3.1) or *test results* (3.1) from all laboratories for one particular test item or test item tested

Note 1 to entry: The accuracy of a measurement method is defined at each level and can be different.

3.4
measurement
test item
sample which is prepared and can be presumed to be identical for the intended purpose

Note 1 to entry: Practical requirements are stated in the protocol of the intended purpose.

Note 2 to entry: Examples of test items: sample, product, artifact, reference test item, equipment, measurement standard.

3.5
accuracy
closeness of agreement between a *test result* (3.1) and the true value

Note 1 to entry: In practice, the accepted reference value is substituted for the true value.

Note 2 to entry: The term accuracy, when applied to a set of test results, involves a combination of random components and a common systematic error or bias component.

Note 3 to entry: Accuracy refers to a combination of trueness and precision.

Note 4 to entry: Common systematic error is called bias component.

[SOURCE: ISO 3534-2:2006, 3.3.1, modified — Note 4 to entry added.]

3.6
trueness
closeness of agreement between the expectation of *test results* (3.1) and a true value

Note 1 to entry: The measure of trueness is usually expressed in terms of bias.

Note 2 to entry: Trueness is sometimes referred to as "accuracy of the mean". This usage is not recommended.

Note 3 to entry: In practice, the accepted reference value is substituted for the true value.

[SOURCE: ISO 3534-2:2006, 3.3.3]

3.7
outlier
value from a set of values which is inconsistent with the other values of that set, identified by a statistical test

Note 1 to entry: ISO 5725-2 specifies the statistical tests and the significance *level* (3.3) to be used to identify outliers in trueness and precision experiments.

3.8
bias
difference between the expectation of the *test results* (3.1) and a true value

Note 1 to entry: Bias is the total systematic error as contrasted to random error. There can be one or more systematic error components contributing to the bias. A larger systematic difference from the accepted reference value is reflected by a larger bias value.

Note 2 to entry: The bias of a measuring instrument is normally estimated by averaging the error of indication over an appropriate number of repeated measurements. The error of indication is the ''indication of a measuring instrument minus a true value of corresponding input quantity''.

Note 3 to entry: In practice, the accepted reference value is substituted for the true value.

[SOURCE: ISO 3534-2:2006, 3.3.2]

3.9
bias of the measurement method
difference between the expectation of *test results* (3.1) obtained from all laboratories using the same method on identical test or measurement items and an *accepted reference value* (3.2)

Note 1 to entry: In practice, the bias of the measurement method is measured by the displacement of the average of results from a large number of different laboratories all using the same method. The bias of a measurement method can be different at different *levels* (3.3).

3.10
laboratory bias
difference between the expectation of the *test results* (3.1) obtained from a particular laboratory and an *accepted reference value* (3.2) under the conditions of a particular experiment

Note 1 to entry: It is assessed based on the performance of a particular laboratory.

3.11
laboratory component of bias B
difference between the *laboratory bias* (3.10) and the *bias of the measurement method* (3.9)

Note 1 to entry: The laboratory component of bias is specific to a given laboratory and the conditions of measurement within the laboratory, and also it can be different at different levels of the measurement method.

Note 2 to entry: The laboratory component of bias is relative to the overall average result, not the true or accepted reference value.

Note 3 to entry: Laboratory component of bias can be named effect of laboratory.

Note 4 to entry: The relationship between the laboratory bias, Δ, the bias of the measurement method, δ, and the laboratory component of the bias B is detailed in ISO 5725-4.

3.12
precision
closeness of agreement between independent *test results* (3.1) obtained under stipulated conditions

Note 1 to entry: Precision depends only on the distribution of random errors and does not relate to the true value or the specified value.

Note 2 to entry: The measure of precision is usually expressed in terms of imprecision and computed as a standard deviation of the test results. Less precision is reflected by a larger standard deviation.

Note 3 to entry: Quantitative measures of precision depend critically on the stipulated conditions. Repeatability and reproducibility conditions are particular sets of extreme conditions.

[SOURCE: ISO 3534-2:2006, 3.3.4]

3.13
repeatability
precision under *repeatability conditions* (3.14)

Note 1 to entry: Repeatability can be expressed quantitatively in terms of the dispersion characteristics of the results.

[SOURCE: ISO 3534-2:2006, 3.3.5]

3.14
repeatability conditions
observation conditions where independent *test results* (3.1) are obtained with the same method on identical test or measurement items in the same test or measuring facility by the same operator using the same equipment within short intervals of time

Note 1 to entry: Repeatability conditions include:

a) The same measurement procedure or test procedure;

b) The same operator;

c) The same measuring or test equipment used under the same conditions;

d) The same location;

e) Repetition over a short period of time.

[SOURCE: ISO 3534-2:2006, 3.3.6]

3.15
repeatability standard deviation
standard deviation of *test results* (3.1) obtained under *repeatability conditions* (3.14)

Note 1 to entry: It is a measure of dispersion of the distribution of test results under repeatability conditions.

Note 2 to entry: Similarly "repeatability variance" and "repeatability coefficient of variation" can be defined and used as measures of the dispersion of test results under repeatability conditions.

Note 3 to entry: Coefficient of variation should be used with caution. Variance or standard deviation is preferred.

[SOURCE: ISO 3534-2:2006, 3.3.7, modified — Note 3 to entry added.]

3.16
repeatability critical difference
value less than or equal to which the absolute difference between two final values, each of them representing a series of *test results* (3.1) obtained under *repeatability* (3.13), is expected to be with a specified probability

Note 1 to entry: Examples of final results are the mean and the median of the series of results; the series itself can consist of only one result.

[SOURCE: ISO 3534-2:2006, 3.3.8]

3.17
repeatability limit
r
repeatability critical difference (3.16) for a specified probability of 95 %

[SOURCE: ISO 3534-2:2006, 3.3.9]

3.18
reproducibility
precision under *reproducibility conditions* (3.19)

Note 1 to entry: Reproducibility can be expressed quantitatively in terms of the dispersion characteristics of the results.

Note 2 to entry: Results are usually understood to be corrected results.

[SOURCE: ISO 3534-2:2006, 3.3.10]

3.19
reproducibility conditions
observation conditions where independent *test results* (3.1) are obtained with the same method on identical test or measurement items in different test or measurement facilities with different operators using different equipment

[SOURCE: ISO 3534-2:2006, 3.3.11]

3.20
reproducibility standard deviation
standard deviation of *test results* (3.1) obtained under *reproducibility conditions* (3.19)

Note 1 to entry: It is a measure of the dispersion of the distribution of test results under reproducibility conditions.

Note 2 to entry: Similarly a "reproducibility variance" and "reproducibility coefficient of variation" can be defined and used as measures of the dispersion of test results under reproducibility conditions.

[SOURCE: ISO 3534-2:2006, 3.3.12]

3.21
reproducibility critical difference
value less than or equal to which the absolute difference between two final values, each of them representing a series of *test results* (3.1) obtained under *reproducibility conditions* (3.19), is expected to be with a specified probability

Note 1 to entry: Instances of final results are the mean and the median of the series of test results, the series itself can consist of only one test result.

[SOURCE: ISO 3534-2:2006, 3.3.13]

3.22
reproducibility limit
R
reproducibility critical difference (3.21) for a specified probability of 95 %

[SOURCE: ISO 3534-2:2006, 3.3.14]

3.23
intermediate precision
precision under intermediate precision conditions

[SOURCE: ISO 3534-2:2006, 3.3.15]

3.24
intermediate precision conditions
conditions where *test results* (3.1) are obtained with the same method, on identical test or measurement items in the same test or measurement facility, under some different operating condition

Note 1 to entry: There are four elements to the operating condition: time, calibration, operator and equipment;

Note 2 to entry: A test house is an example of a test facility. A metrology laboratory is an example of a measurement facility;

Note 3 to entry: In addition to the four elements of the operating condition listed above, some more elements may be different as batch, preparation and others.

Note 4 to entry: The above conditions can change independently.

[SOURCE: ISO 3534-2:2006, 3.3.16, modified —Notes 3 and 4 to entry added.]

3.25
intermediate precision standard deviation
standard deviation of *test results* (3.1) obtained under *intermediate precision conditions* (3.24)

[SOURCE: ISO 3534-2:2006, 3.3.17]

3.26
intermediate precision critical difference
value less than or equal to which the absolute difference between two final values, each of them representing a series of *test results* (3.1) obtained under *intermediate precisions conditions* (3.24), is expected to be with a specified probability

[SOURCE: ISO 3534-2:2006, 3.3.18]

3.27
intermediate precision limit
intermediate precision critical difference (3.26) for a specified probability of 95 %

[SOURCE: ISO 3534-2:2006, 3.3.19]

4 General principles and practices of accuracy experiments

4.1 Accuracy experiment

4.1.1 The accuracy (trueness and precision) measures should be determined from a series of test results reported by the participating laboratories, organized under a panel of experts established specifically for that purpose.

Such an interlaboratory experiment is called an "accuracy experiment". The accuracy experiment may also be called a "precision" or "trueness experiment" according to its limited purpose. If the purpose is to determine trueness, then a precision experiment shall either have been completed previously or shall occur simultaneously.

The estimates of accuracy derived from such an experiment should always be quoted as being valid only for tests carried out according to the standard measurement method.

4.1.2 An accuracy experiment can often be considered to be a practical test of the adequacy of the standard measurement method. One of the main purposes of standardization is to eliminate differences between users (laboratories) as far as possible, and the data provided by an accuracy experiment will reveal how effectively this purpose has been achieved. Pronounced differences in the within-laboratory variances (see Clause 7) or between the laboratory means may indicate the inadequacy of the standard measurement method.

4.2 Standard measurement method

4.2.1 In order that the measurements are made in the same way, the measurement method shall have been standardized. All measurements shall be carried out according to that standard method. This means that there has to be a written document that lays down in full detail how the measurement shall be carried out, preferably including a description as to how the measurement item should be obtained and prepared.

4.2.2 The existence of a documented measurement method implies the existence of an organization responsible for the establishment of the measurement method under study.

NOTE The standard measurement method is discussed more fully in 6.2.

4.3 Requirements concerning test items

4.3.1 In an accuracy experiment, samples of a specific test item or samples of a specific product are sent from a central point to a number of laboratories in different places, different countries, or even in different continents. The definition of repeatability conditions (3.14) stating that the measurements in these laboratories shall be performed on identical test items refers to the moment when these measurements are actually carried out. To achieve this, two different conditions have to be satisfied:

a) the samples have to be identical when dispatched to the laboratories but at one or different levels;

b) they have to remain identical during transport and during the different time intervals that may elapse before the measurements are actually performed.

In organizing accuracy experiments, both conditions shall be carefully observed.

NOTE The selection of test items is discussed more in details in 6.4.

4.4 Conditions for evaluation of repeatability (short intervals of time)

4.4.1 According to the definition of repeatability conditions (3.14), measurements for the determination of repeatability have to be made under constant operating conditions; i.e. during the time covered by the measurements, factors such as, by example, those listed should be constant:

a) the time elapsed between measurements;

b) the operator (experience, dexterity, ...);

c) the equipment or test bench used;

d) the calibration of the equipment;

e) the environment (temperature, humidity, air pollution, vibration, etc.);

f) the batch of reagent;

g) preparation of test item;

h) other influential factors.

These influent factors should be constant. In particular, the equipment should not be recalibrated or readjusted between the measurements unless this is an essential part of every single measurement. In practice, tests under repeatability conditions should be conducted in as short a time as possible in order to minimize changes in those factors, such as environmental, which cannot always be guaranteed constant.

4.4.2 There is also a second consideration which may affect the interval elapsing between measurements, and that is that the test results are assumed to be independent. If it is feared that previous results may influence subsequent test results (and so reduce the estimate of repeatability variance), it may be necessary to provide separate test items coded in such a way that an operator will not know which are supposedly identical. Instructions would be given as to the order in which those test items are to be measured, and presumably that order will be randomized so that all the "identical" test items are not measured together. This might mean that the time interval between repeated measurements may appear to defeat the object of a short interval of time unless the measurements are of such a nature that the whole series of measurements could all be completed within a short interval of time. Common sense must prevail.

4.5 Conditions for evaluation of trueness

The "trueness" of a measurement method is of interest when it is possible to conceive of a true value for the property being measured. Although the true value cannot be known exactly, it may be possible to have an accepted reference value for the property being measured; for example, if suitable reference test items or measurement standards are available, or if the accepted reference value can be established by reference to another measurement method or by preparation of a known sample. The trueness of the measurement method can be investigated by comparing the accepted reference value with the level of the results given by the measurement method. Trueness is normally expressed in terms of bias.

4.6 Participating laboratories

4.6.1 A basic assumption underlying this document is that, for a standard measurement method, repeatability will be, at least approximately, the same for all laboratories applying the standard procedure, so that it is permissible to establish one common average repeatability standard deviation which will be applicable to any laboratory. However, by carrying out a series of measurements under repeatability conditions, any laboratory can arrive at an estimate of its own repeatability standard deviation for the measurement method and check it against the common standard value. Such a procedure is dealt with in ISO 5725-6.

4.6.2 Values of the quantities defined in 3.10 to 3.26 apply theoretically to all laboratories which are likely to perform the measurement method. In practice, they are determined from a sample of this population of laboratories. Further details of the selection of this sample are given in 6.3. Provided the instructions given there regarding the number of laboratories to be included and the number of measurements that they carry out are followed, then the resulting estimates of trueness and precision should suffice. If, however, at some future date it should become evident that the laboratories participating were not, or are no longer, truly representative of all those using the standard measurement method, then the measurement should be repeated.

4.7 Influential factors (observation conditions)

4.7.1 The factors which contribute to the variability of the observed values obtained within a laboratory are listed in 4.4.1. They may be given as time; operator and equipment when observations at different times include the effects due to the change of environmental conditions, the calibration of equipment between observations and other factors. Under repeatability conditions, observations are carried out with all these factors constant, and under reproducibility conditions observations are carried out at different laboratories; i.e. not only with all the other factors varying but also with additional effects due to the difference between laboratories in management and maintenance of the laboratory, stability checking of the observations, etc.

4.7.2 Under intermediate precision conditions, observations are carried out in the same laboratory, but one or more of the influential factors are allowed to vary. In establishing the precision of a measurement method, it is very important to define the appropriate observation conditions, i.e. which influential factors should be constant or not.

NOTE The size (magnitude) of the variability arising from a factor will depend on the measurement method.

5 Statistical model

5.1 Basic model

In this standard the model is not a physical or chemical equation, but by a statistical model.

For estimating the accuracy (trueness and precision) of a measurement method, it is useful to consider the statistical model according to which every test result. So the basic model given by Formula (1) is used:

$$y = m + B + e \tag{1}$$

where, for the particular test item tested,

- m is the general mean (expectation);
- B is the laboratory component of bias under repeatability conditions;
- e is the random error occurring in every measurement under repeatability conditions.

NOTE 1 Methods described in ISO 5725-2 to evaluate precision parameters (variance of B and variance of e) are based on this basic statistical model. Alternative models are described in ISO 5725-3 and ISO 5725-4.

NOTE 2 The general mean, m, includes the bias of the measurement method. The relationship is described in ISO 5725-4.

5.1.1 General mean, m

5.1.1.1 For the particular test item, the general mean m is the level of a particular test item. Test items of different purities of a chemical, or different test items (e.g. different types of steel), will have different levels. In many technical situations the level of the test item is exclusively determined by the measurement method, and the notion of an independent true value does not apply. However, in some situations the concept of a true value μ of the test property may hold good, such as the accepted reference concentration of a solution that is being titrated. The level m is not necessarily equal to the true value or accepted reference value μ.

5.1.1.2 When examining the difference between results obtained by the same measurement method, the bias of the measurement method will have no influence and can be ignored. However, when comparing test results with a value specified in a contract or a standard where the contract or specification refers to the true value (μ) and not to the “mean of the test item ”(m), or when comparing results produced using different measurement methods, the bias of the measurement method will have to be taken into account. If a true value exists and a satisfactory reference test item is available, the bias of the measurement method should be determined as shown in ISO 5725-4.

5.1.2 Laboratory component of bias: term *B*

5.1.2.1 This term is considered to be constant during series of tests performed under repeatability conditions, but differs in value for tests carried out under other conditions (intermediate precision conditions/reproducibility conditions).

Using the basic model, it is considered that the repeatability conditions are the conditions for a given laboratory.

When other sources of variation are identified (e.g. assembly/disassembly of test items) they could be taken into account in the repeatability conditions. For example, with a more complicated model with a laboratory effect and an operator effect, these conditions apply for a given laboratory, a given operator and given sources of variation.

Coming back to the basic model, as there are several laboratories then a general distribution of laboratory components of bias must be considered.

The variance of the between-laboratory variance is the unknown variance of that distribution that must be estimated.

The procedures to evaluate precision parameters (variance of B and variance of e), given in ISO 5725-2 were developed assuming that the distribution of laboratory components of bias is approximately normal but in practice they work for most distributions that are unimodal and approximately symmetric.

NOTE When test results are always compared between the same two laboratories, it is necessary for them to determine their relative bias, either from their individual bias values as determined during an accuracy experiment, or by carrying out a private trial between themselves. However, in order to make general statements regarding differences between two unspecified laboratories or when making comparisons between two laboratories that have not determined their own bias, then a general distribution of laboratory components of bias must be considered. This was the reasoning behind the concept of reproducibility.

5.1.2.2 The variance of *B* is called the “between-laboratory” variance and is expressed as shown in Formula (2)

$$var(B) = \sigma_L^2 \tag{2}$$

where σ_L^2 includes at least the “between-operator” and “between-equipment” variabilities and generally all variabilities due to the change of an influential factor in a given laboratory or from one laboratory to another.

In the basic precision experiment described in ISO 5725-2, these components are not separated. Methods are given in ISO 5725-3 for measuring the variance component of *B*.

5.1.2.3 In general, *B* can be considered as the sum of both random and systematic components. No attempt is made to give here an exhaustive list of the factors that contribute to *B*, but they include different climatic conditions, variations of equipment within the manufacturer's tolerances, and even differences in the techniques in which operators are trained in different places.

5.1.3 Error term *e*

5.1.3.1 This term represents a random error occurring in every test result and the procedures given throughout this document were developed assuming that the distribution of this error variable was approximately normal, but in practice they work for most distributions provided that they are unimodal.

5.1.3.2 Within a single laboratory, its variance under repeatability conditions is called the within-laboratory variance and is expressed as shown by Formula (3):

$$var(e) = \sigma_W^2 \tag{3}$$

When the ANOVA model is used to estimate the repeatability standard deviation, that residual standard deviation takes into account all variabilities except those linked to the factors included in the model of analysis of variance.

5.1.3.3 It may be expected that σ_W^2 will have different values in different laboratories due to differences, such as in the skills of the operators, but in this document it is assumed that for a properly standardized or defined measurement method, such differences between laboratories should be small and that it is justifiable to establish a common value of "within laboratory" variance for all the laboratories using the measurement method.

This common value, called the repeatability variance, σ_r^2, is estimated as the arithmetic mean of the "within laboratory" variance, as given by Formula (4):

$$\sigma_r^2 = \overline{var(e)} = \overline{\sigma_W^2} \tag{4}$$

This arithmetic mean is taken over all those laboratories taking part in the accuracy experiment which remains after outliers have been excluded (see ISO 5725-2).

NOTE Methods are given in ISO 5725-2 for estimating the precision.

5.2 Relationship between the basic model and the precision

5.2.1 When the basic model in 5.1 is adopted, the repeatability variance is measured directly as the variance of the error term *e*, but the reproducibility variance depends on the sum of the repeatability variance and the between-laboratory variance mentioned in 5.1.2.2.

5.2.2 Two quantities are required as measures of precision, the repeatability standard deviation, given by Formula (5):

$$\sigma_r = \sqrt{\overline{var(e)}} \tag{5}$$

and the reproducibility standard deviation, given by Formula (6):

$$\sigma_R = \sqrt{\sigma_L^2 + \sigma_r^2} \tag{6}$$

5.3 Bias of the measurement method

The bias of the measurement method $\delta = m - \mu$ is estimated by $\hat{\delta} = \hat{m} - \mu$ where $\hat{m}$ is an estimate of the general mean *m* and μ is the true value (if known) or the accepted reference value.

NOTE Methods to evaluate the bias are given in ISO 5725-4.

5.4 Alternative models

Extensions to the basic model are used when appropriate and are described in the relevant parts of ISO 5725.

6 Experimental design of an accuracy experiment

6.1 Planning of an accuracy experiment

6.1.1 The actual planning of an experiment to estimate the precision and/or trueness of a measurement method should be the task of a panel of experts familiar with the measurement method and its application. At least one member of the panel should have experience in the statistical design and analysis of experiments.

6.1.2 The following questions should be considered when planning the experiment (the list may not be exhaustive).

a) Is the measurement method completely defined or developed?

b) How many laboratories should be recruited to cooperate in the experiment?

c) How should the laboratories be recruited, and what requirements should they satisfy?

d) What is the range of levels encountered in practice?

e) How many levels should be used in the experiment?

f) What are suitable test items to represent these levels and how should they be prepared?

g) What number of replicates should be specified?

h) What time-frame should be specified for the completion of all the measurements?

i) Is the basic model of 5.1 appropriate, or should a modified one be considered?

j) Are any special precautions needed to ensure that identical test items are measured in the same state in all laboratories?

6.2 Standard measurement methods

As stated in 4.2, the measurement method under investigation shall be one that has been standardized. Such a method has to be robust, i.e. small variations in the procedure should not produce unexpectedly large changes in the results. If this might happen, there shall be adequate precautions or warnings. It is also desirable that in the process of developing a standard measurement method every effort has been made to remove or reduce bias.

Similar experimental procedures may be used to measure the trueness and precision of both established measurement methods and recently standardized measurement methods. In the latter case, the results obtained should be regarded as preliminary estimates, because the trueness and precision could change as laboratories gain experience.

The document setting out the measurement method shall be unambiguous and complete. All essential operations concerning the environment of the procedure, the reagents and apparatus, preliminary checking of equipment, and the preparation of the test item should be included in the measurement method, possibly by references to other written procedures that are available to the operators. The manner of calculating and expressing the test result should be precisely specified, including the number of significant figures to be reported.

6.3 Selection of laboratories for the accuracy experiment

From a statistical point of view, those laboratories participating in any experiment to estimate accuracy should have been chosen at random from all the laboratories using the same measurement method.

Volunteers might not represent a realistic cross-section, and they should represent all those using the standard measurement method. However, other practical considerations, such as a requirement that the participating laboratories be distributed over different continents or climatic regions, may affect the pattern of representation.

The participating laboratories should not consist exclusively of those that have gained special experience during the process of standardizing the method. Neither should they consist of specialized “reference” laboratories in order to demonstrate the accuracy to which the method can perform in expert hands.

A guide for deciding on how many laboratories and replications are required for an estimate of precision is given in ISO 5725-2 and ISO 5725-3; and correspondingly "on estimating bias in ISO 5725-4.”

If the laboratories participating were not, or are no longer, sufficiently representative of all those using the measurement method defined, then the experiment or measurement should be repeated.

6.4 Selection of test items to be used for an accuracy experiment

6.4.1 The test items to be used in an experiment to determine the accuracy of a measurement method should represent fully those to which the measurement method is expected to be applied in normal use. A smaller number might be appropriate in the first investigation of a recently developed measurement method when it is suspected that modifications to the method may be necessary, followed by further accuracy experiments.

6.4.2 When the measurements have to be performed on discrete test items that are not altered by measuring, they could, in principle at least, be carried out using the same set of test items in different laboratories. This, however, would necessitate circulating the same set of test items around many laboratories often situated far apart, in different countries or continents, with a considerable risk of loss or damage during transport. If different test items are to be used in different laboratories, then they shall be selected in such a way as to ensure that they can be presumed to be identical for practical purposes.

6.4.3 When selecting the test items to represent one or the different levels, it should be considered whether test items should be specially homogenized before preparing the samples for dispatch, or whether the effect of the heterogeneity of the test items should be included in the accuracy values.

6.4.4 When measurements have to be performed on solid test items that cannot be homogenized (such as metals, rubber or textile fabrics) and when the measurements cannot be repeated on the same test item, inhomogeneity in the test item will form an essential component of the precision of the measurement and the idea of identical test item no longer holds good. Precision experiments can still be carried out, but the values of precision may only be valid for the particular test item used and should be quoted as such. A more universal use of the precision as determined will be acceptable only if it can be demonstrated that the values do not differ significantly between test items produced at different times or by different producers. This would require a more elaborate experiment than has been considered in ISO 5725 (all parts).

6.4.5 In general, where destructive testing is involved, the contribution to the variability in the test results arising from differences between the test items on which the measurements are performed shall either be negligible compared to the variability of the measurement method itself, or else shall form an inherent part of the variability of the measurement method, and thus be truly a component of precision.

6.4.6 When the test items under measurement might change with time, the overall time-scale of the experiment should be chosen to take this into account. It might be appropriate in some cases to specify the times at which the test items are to be measured.

6.4.7 In all above, reference is made to measuring in different laboratories, with the implication of transportation of the test items to the laboratory, but some test items are not transportable, such as an oil storage tank. In such case, measuring by different laboratories means that different operators are sent with their equipment to the test site. In other cases, the quantity being measured may be transitory or variable, such as water flow in a river, when care shall be taken that the different measurements are made under, as near as possible, the same conditions. The guiding principle must always be that the objective is to determine the ability to repeat the same measurement.

6.4.8 The establishment of precision values for a measurement method presupposes that the precision either is independent of the test item being tested, or depends on the test item in a predictable manner. With some measurement methods it is possible to quote the precision only in relation to one or more definable classes of test items. Such data will be only a rough guide to the precision in other applications. More often it is found that the precision is closely related to the level of the test, and determination of the precision then includes the establishment of a relationship between precision and level. Therefore, when publishing precision values for a standard measurement method, it is recommended that the test item used in the precision experiment should be clearly specified along with the range of test items to which the values can be expected to apply.

6.4.9 For the assessment of trueness, at least one of the test items used should have an accepted reference value. If it is likely that trueness varies with level, test items with accepted reference values will be needed at several levels.

7 Utilization of accuracy data

7.1 Publication values of trueness and precision

7.1.1 When the aim of a precision experiment is to obtain estimates of the repeatability and reproducibility standard deviations under the conditions defined in 3.14 and 3.19, then the basic model of 5.1 shall be used. ISO 5725-2 then provides an appropriate method of estimating these standard deviations, or an alternative may be found in ISO 5725-3. When the aim is to obtain estimates of intermediate measures of precision, then the alternative model and the methods given in ISO 5725-3 shall be used.

7.1.2 Whenever the bias of the measurement method has been determined, it should be published with a statement regarding the reference against which that bias was determined. Where the bias varies with the level of the test, publication should be in the form of a table giving the level, the bias as determined, and the reference used in that determination.

7.1.3 When an interlaboratory experiment has been performed for estimating trueness or precision, each participating laboratory should be informed of its laboratory component of bias relative to the general mean as determined from the experiment. This information could be of value in the future if similar experiments are performed, but should not be used for calibration purposes or any corrections.

7.1.4 The repeatability and reproducibility standard deviations for any standard measurement method shall be determined as laid down in ISO 5725-2, ISO 5725-3 and ISO 5725-5, and should be published as part of the standard measurement method under a section entitled precision. This section may also show the repeatability and reproducibility limits (r and R). When precision does not vary with level, single average figures can be given in each case. Where precision varies with the level of the test, the publication should be in the form of a table, such as Table 1, and may also be expressed as