INTERNATIONAL STANDARD

ISO
5725-1

First edition
1994-12-15

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

Contents

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.

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

International Standard ISO 5725-1 was prepared by Technical Committee ISO/TC 69, *Applications of statistical methods*, Subcommittee SC 6, *Measurement methods and results*.

ISO 5725 consists of the following parts, under the general title *Accuracy (trueness and precision) of measurement methods and results*:

— *Part 1: General principles and definitions*

— *Part 2: Basic method for the determination of repeatability and reproducibility of a standard measurement method*

— *Part 3: Intermediate measures of the precision of a standard measurement method*

— *Part 4: Basic methods for the determination of the trueness of a standard measurement method*

— *Part 5: Alternative methods for the determination of the precision of a standard measurement method*

— *Part 6: Use in practice of accuracy values*

Parts 1 to 6 of ISO 5725 together cancel and replace ISO 5725:1986, which has been extended to cover trueness (in addition to precision) and intermediate precision conditions (in addition to repeatability and reproducibility conditions).

Annexes A and B form an integral part of this part of ISO 5725. Annex C is for information only.

Introduction

0.1   ISO 5725 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.

0.2   The need to consider "precision" arises because tests performed on presumably identical materials 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 has to 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 material will not indicate a fundamental quality difference if the difference between them can be attributed to the inherent variation in the measurement procedure.

0.3   Many different factors (apart from variations between supposedly identical specimens) may contribute to the variability of results from a measurement method, including:

a)   the operator;

b)   the equipment used;

c)   the calibration of the equipment;

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

e)   the time elapsed between measurements.

The variability between measurements performed by different operators and/or with different equipment will usually be greater than the variability between measurements carried out within a short interval of time by a single operator using the same equipment.

0.4   The general term for variability between repeated 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, factors a) to e) listed above are considered

© ISO

constants and do not contribute to the variability, while under reproducibility conditions they 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 of precision are also conceivable, when one or more of factors a) to e) are allowed to vary, and are used in certain specified circumstances. Precision is normally expressed in terms of standard deviations.

0.5 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, for some measurement methods, 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 materials 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. Bias can arise, for example, in chemical analysis if the measurement method fails to extract all of an element, or if the presence of one element interferes with the determination of another.

0.6 The general term accuracy is used in ISO 5725 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.

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

Part 1:
General principles and definitions

1 Scope

1.1 The purpose of ISO 5725 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 a basic method for estimating the two extreme measures of the precision of measurement methods by experiment (ISO 5725-2);

c) to provide a procedure for obtaining intermediate measures of precision, giving the circumstances in which they apply and methods for estimating them (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 basic methods, given in ISO 5725-2 and 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 of these measures of trueness and precision (ISO 5725-6).

1.2 This part of ISO 5725 is concerned exclusively with measurement methods which yield measurements 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 correct result (trueness) or to replicate a given result (precision). Thus there is an implication that exactly the same thing is being measured, in exactly the same way, and that the measurement process is under control.

This part of ISO 5725 may be applied to a very wide range of materials, including liquids, powders and solid objects, manufactured or naturally occurring, provided that due consideration is given to any heterogeneity of the material.

2 Normative references

The following standards contain provisions which, through reference in this text, constitute provisions of this part of ISO 5725. At the time of publication, the editions indicated were valid. All standards are subject to revision, and parties to agreements based on this part of ISO 5725 are encouraged to investigate the possibility of applying the most recent editions of the standards indicated below. Members of IEC and ISO maintain registers of currently valid International Standards.

ISO 3534-1:1993, *Statistics — Vocabulary and symbols — Part 1: Probability and general statistical terms.*

ISO 5725-2:1994, *Accuracy (trueness and precision) of measurement methods and results — Part 2: Basic method for the determination of repeatability and reproducibility of a standard measurement method.*

ISO 5725-3:1994, *Accuracy (trueness and precision) of measurement methods and results — Part 3: Intermediate measures of the precision of a standard measurement method.*

ISO 5725-4:1994, *Accuracy (trueness and precision) of measurement methods and results — Part 4: Basic methods for the determination of the trueness of a standard measurement method.*

3 Definitions

For the purposes of ISO 5725, the following definitions apply.

Some definitions are taken from ISO 3534-1.

The symbols used in ISO 5725 are given in annex A.

3.1 observed value: The value of a characteristic obtained as the result of a single observation.

[ISO 3534-1]

3.2 test result: The value of a characteristic obtained by carrying out a specified test method.

NOTE 1   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 may 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.

[ISO 3534-1]

3.3 level of the test in a precision experiment: The general average of the test results from all laboratories for one particular material or specimen tested.

3.4 cell in a precision experiment: The test results at a single level obtained by one laboratory.

3.5 accepted reference value: A value that serves as an agreed-upon reference for comparison, and 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) when a), b) and c) are not available, the expectation of the (measurable) quantity, i.e. the mean of a specified population of measurements.

[ISO 3534-1]

3.6 accuracy: The closeness of agreement between a test result and the accepted reference value.

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

[ISO 3534-1]

3.7 trueness: The closeness of agreement between the average value obtained from a large series of test results and an accepted reference value.

NOTES

3   The measure of trueness is usually expressed in terms of bias.

4   Trueness has been referred to as "accuracy of the mean". This usage is not recommended.

[ISO 3534-1]

3.8 bias: The difference between the expectation of the test results and an accepted reference value.

NOTE 5   Bias is the total systematic error as contrasted to random error. There may 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.

[ISO 3534-1]

3.9 laboratory bias: The difference between the expectation of the test results from a particular laboratory and an accepted reference value.

3.10 bias of the measurement method: The difference between the expectation of test results obtained from all laboratories using that method and an accepted reference value.

NOTE 6 One example of this in operation would be where a method purporting to measure the sulfur content of a compound consistently fails to extract all the sulfur, giving a negative bias to the measurement method. 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 may be different at different levels.

3.11 laboratory component of bias: The difference between the laboratory bias and the bias of the measurement method.

NOTES

7 The laboratory component of bias is specific to a given laboratory and the conditions of measurement within the laboratory, and also it may be different at different levels of the test.

8 The laboratory component of bias is relative to the overall average result, not the true or reference value.

3.12 precision: The closeness of agreement between independent test results obtained under stipulated conditions.

NOTES

9 Precision depends only on the distribution of random errors and does not relate to the true value or the specified value.

10 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.

11 "Independent test results" means results obtained in a manner not influenced by any previous result on the same or similar test object. Quantitative measures of precision depend critically on the stipulated conditions. Repeatability and reproducibility conditions are particular sets of extreme conditions.

[ISO 3534-1]

3.13 repeatability: Precision under repeatability conditions.

[ISO 3534-1]

3.14 repeatability conditions: Conditions where independent test results are obtained with the same method on identical test items in the same laboratory by the same operator using the same equipment within short intervals of time.

[ISO 3534-1]

3.15 repeatability standard deviation: The standard deviation of test results obtained under repeatability conditions.

NOTES

12 It is a measure of dispersion of the distribution of test results under repeatability conditions.

13 Similarly "repeatability variance" and "repeatability coefficient of variation" could be defined and used as measures of the dispersion of test results under repeatability conditions.

[ISO 3534-1]

3.16 repeatability limit: The value less than or equal to which the absolute difference between two test results obtained under repeatability conditions may be expected to be with a probability of 95 %.

NOTE 14   The symbol used is r.

[ISO 3534-1]

3.17 reproducibility: Precision under reproducibility conditions.

[ISO 3534-1]

3.18 reproducibility conditions: Conditions where test results are obtained with the same method on identical test items in different laboratories with different operators using different equipment.

[ISO 3534-1]

3.19 reproducibility standard deviation: The standard deviation of test results obtained under reproducibility conditions.

NOTES

15 It is a measure of the dispersion of the distribution of test results under reproducibility conditions.

16 Similarly "reproducibility variance" and "reproducibility coefficient of variation" could be defined and used as measures of the dispersion of test results under reproducibility conditions.

[ISO 3534-1]

3.20 reproducibility limit: The value less than or equal to which the absolute difference between two test results obtained under reproducibility conditions may be expected to be with a probability of 95 %.

NOTE 17   The symbol used is R.

[ISO 3534-1]

3.21 outlier: A member of a set of values which is inconsistent with the other members of that set.

NOTE 18   ISO 5725-2 specifies the statistical tests and the significance level to be used to identify outliers in trueness and precision experiments.

3.22 collaborative assessment experiment: An interlaboratory experiment in which the performance of each laboratory is assessed using the same standard measurement method on identical material.

NOTES

19   The definitions given in 3.16 and 3.20 apply to results that vary on a continuous scale. If the test result is discrete or rounded off, the repeatability limit and the reproducibility limit as defined above are each the minimum value equal to or below which the absolute difference between two single test results is expected to lie with a probability of not less than 95 %.

20   The definitions given in 3.8 to 3.11, 3.15, 3.16, 3.19 and 3.20 refer to theoretical values which in reality remain unknown. The values for reproducibility and repeatability standard deviations and bias actually determined by experiment (as described in ISO 5725-2 and ISO 5725-4) are, in statistical terms, estimates of these values, and as such are subject to errors. Consequently, for example, the probability levels associated with the limits r and R will not be exactly 95 %. They will approximate to 95 % when many laboratories have taken part in the precision experiment, but may be considerably different from 95 % when fewer than 30 laboratories have participated. This is unavoidable but does not seriously detract from their practical utility as they are primarily designed to serve as tools for judging whether the difference between results could be ascribed to random uncertainties inherent in the measurement method or not. Differences larger than the repeatability limit r or the reproducibility limit R are suspect.

21   The symbols r and R are already in general use for other purposes; in ISO 3534-1 r is recommended for the correlation coefficient and R (or W) for the range of a single series of observations. However, there should be no confusion if the full wordings repeatability limit r and reproducibility limit R are used whenever there is a possibility of misunderstanding, particularly when they are quoted in standards.

4   Practical implications of the definitions for accuracy experiments

4.1   Standard measurement method

4.1.1   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 specimen should be obtained and prepared.

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

NOTE 22   The standard measurement method is discussed more fully in 6.2.

4.2   Accuracy experiment

4.2.1   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.

4.2.2   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 that the standard measurement

method is not yet sufficiently detailed and can possibly be improved. If so, this should be reported to the standardizing body with a request for further investigation.

4.3 Identical test items

4.3.1 In an accuracy experiment, samples of a specific material or specimens 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;

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 23   The selection of material is discussed more fully in 6.4.

4.4 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 those listed in 0.3 should be constant. In particular, the equipment should not be recalibrated 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 specimens 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 specimens are to be

measured, and presumably that order will be randomized so that all the "identical" 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 Participating laboratories

4.5.1 A basic assumption underlying this part of ISO 5725 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, any laboratory can, by carrying out a series of measurements under repeatability conditions, 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.

4.5.2 The quantities defined in 3.8 to 3.20 in theory apply 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 shall be repeated.

4.6 Observation conditions

4.6.1 The factors which contribute to the variability of the observed values obtained within a laboratory are listed in 0.3. They may be given as time, operator and equipment when observations at different times include the effects due to the change of environmental conditions and the recalibration of equipment between observations. 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 lab-

oratories in management and maintenance of the laboratory, stability checking of the observations, etc.

4.6.2 It may be useful on occasion to consider intermediate precision conditions, in which observations are carried out in the same laboratory but one or more of the factors time, operator or equipment are allowed to vary. In establishing the precision of a measurement method, it is very important to define the appropriate observation conditions, i.e. whether the above three factors should be constant or not.

Furthermore, the size of the variability arising from a factor will depend on the measurement method. For example, in chemical analysis, the factors "operator" and "time" may dominate; likewise with microanalysis the factors "equipment" and "environment", and with physical testing "equipment" and "calibration" may dominate.

5 Statistical model

5.1 Basic model

For estimating the accuracy (trueness and precision) of a measurement method, it is useful to assume that every test result, y, is the sum of three components:

$$y = m + B + e \qquad \ldots (1)$$

where, for the particular material 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.

5.1.1 General mean, m

5.1.1.1 The general mean m is the level of the test; specimens of different purities of a chemical, or different materials (e.g. different types of steel), will correspond to different levels. In many technical situations the level of the test is exclusively defined 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 true concentration of a solution that is being titrated. The level m is not necessarily equal to the true value μ.

5.1.1.2 When examining the difference between test 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 "level of the test" (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 material is available, the bias of the measurement method should be determined as shown in ISO 5725-4.

5.1.2 Term B

5.1.2.1 This term is considered to be constant during any series of tests performed under repeatability conditions, but to differ in value for tests carried out under other conditions. 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. The procedures 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 provided that they are unimodal.

5.1.2.2 The variance of B is called the between-laboratory variance and is expressed as:

$$\text{var}\,(B) = \sigma_L^2 \qquad \ldots (2)$$

where σ_L^2 includes the between-operator and between-equipment variabilities.

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 size of some of the random components 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 part of ISO 5725 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:

$$\text{var}\ (e) = \sigma_W^2 \qquad \ldots (3)$$

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 part of ISO 5725 it is assumed that for a properly standardized 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, which is estimated by the arithmetic mean of the within-laboratory variances, is called the repeatability variance and is designated by:

$$\sigma_r^2 = \overline{\text{var}\ (e)} = \overline{\sigma_W^2} \qquad \ldots (4)$$

This arithmetic mean is taken over all those laboratories taking part in the accuracy experiment which remain after outliers have been excluded.

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

$$\sigma_r = \sqrt{\text{var}\ (e)} \qquad \ldots (5)$$

and the reproducibility standard deviation

$$\sigma_R = \sqrt{\sigma_L^2 + \sigma_r^2} \qquad \ldots (6)$$

5.3 Alternative models

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

6 Experimental design considerations when estimating accuracy

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 standard 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.

a) Is a satisfactory standard available for the measurement method?

b) How many laboratories should be recruited to co-operate 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 materials 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 materials are measured in the same state in all laboratories?

These questions are considered in 6.2 to 6.4.

6.2 Standard measurement method

As stated in 4.1, 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 specimen 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

6.3.1 Choice of laboratories

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 measurement method. Volunteers might not represent a realistic cross-section. 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.

The number of laboratories to be recruited to participate in a cooperative interlaboratory experiment and the number of test results required from each laboratory at each level of the test are interdependent. A guide to deciding how many there should be is given in 6.3.2 to 6.3.4.

6.3.2 Number of laboratories required for an estimate of precision

6.3.2.1 The various quantities represented by the symbol σ in equations (2) to (6) of clause 5 are true standard deviations whose values are not known, an object of a precision experiment being to estimate them. When an estimate (s) of a true standard deviation (σ) is to be made, conclusions can be drawn as to the range about σ within which the estimate (s) can be expected to lie. This is a well-understood statistical problem which is solved by the use of the chi-squared distribution and the number of results from which the estimate of s was based. One formula frequently used is:

$$P\left[-A < \frac{s-\sigma}{\sigma} < +A\right] = P \qquad \ldots (7)$$

Often A is quoted in percentage terms, enabling a statement to be made that the estimated standard deviations (s) can be expected to be within A either side of the true standard deviation (σ) with a certain probability P.

6.3.2.2 For a single level of the test, the uncertainty in the repeatability standard deviation will depend on the number of laboratories (p) and the number of test results within each laboratory (n). For the reproducibility standard deviation, the procedure is more complicated as this is determined from two standard deviations [see equation (6)]. An extra factor γ is needed, representing the ratio of the reproducibility standard deviation to the repeatability standard deviation, that is:

$$\gamma = \sigma_R / \sigma_r \qquad \ldots (8)$$

6.3.2.3 Assuming a probability level P of 95 %, approximate equations for the values of A have been prepared and are given below. The equations are intended for the purposes of planning how many laboratories to recruit and deciding how many test results are to be required from each laboratory at each level of the test. These equations do not give confidence limits and so they should not be used during the analysis stage to calculate confidence limits. The equations are as follows.

For repeatability

$$A = A_r = 1{,}96\sqrt{\frac{1}{2p(n-1)}} \qquad \ldots (9)$$

For reproducibility

$$A = A_R = 1{,}96 \sqrt{\frac{p[1 + n(\gamma^2 - 1)]^2 + (n - 1)(p - 1)}{2\gamma^4 n^2 (p - 1)p}}$$

. . . (10)

NOTE 24   A sample variance which has ν degrees of freedom and expectation σ^2 may be assumed to have, approximately, a normal distribution with variance $2\sigma^4/\nu$. Equations (9) and (10) were derived by making this assumption about the variances involved in the estimation of σ_r and σ_R. The adequacy of the approximation was checked by an exact calculation.

6.3.2.4   The value of γ is not known, but often preliminary estimates are available of the within-laboratory standard deviations and the between-laboratory standard deviations obtained during the process of standardizing the measurement method. Exact values of the uncertainty percentages for repeatability and reproducibility standard deviations with different numbers of laboratories (p) and different numbers of results per laboratory (n) are given in table 1 and are also plotted in chart form in annex B.

6.3.3   Number of laboratories required for the estimate of bias

6.3.3.1   The bias of the measurement method, δ, may be estimated from:

$$\hat{\delta} = \bar{\bar{y}} - \mu \qquad \text{. . . (11)}$$

where

$\bar{\bar{y}}$ is the grand mean of all the test results obtained by all the laboratories at a particular level of the experiment;

μ is the accepted reference value.

The uncertainty of this estimate can be expressed by the equation:

$$P \left[\delta - A\sigma_R < \hat{\delta} < \delta + A\sigma_R \right] = 0{,}95 \qquad \text{. . . (12)}$$

which shows that the estimate will be within $A\sigma_R$ of the true measurement method bias with a probability of 0,95. In terms of the factor γ [see equation (8)]:

$$A = 1{,}96 \sqrt{\frac{n(\gamma^2 - 1) + 1}{\gamma^2 pn}} \qquad \text{. . . (13)}$$

Values of A are given in table 2.

6.3.3.2   The laboratory bias, Δ, at the time of the experiment may be estimated from:

$$\hat{\Delta} = \bar{y} - \mu \qquad \text{. . . (14)}$$

where

$\bar{y}$ is the arithmetic mean of all the results obtained by the laboratory at a particular level of the experiment;

μ is the accepted reference value.

The uncertainty of this estimate can be expressed by the equation:

$$P \left[\Delta - A_W\sigma_r < \hat{\Delta} < \Delta + A_W\sigma_r \right] = 0{,}95 \qquad \text{. . . (15)}$$

which shows that the estimate will be within $A_W\sigma_r$ of the true laboratory bias with a probability of 0,95. Here the within-laboratory uncertainty is:

$$A_W = \frac{1{,}96}{\sqrt{n}} \qquad \text{. . . (16)}$$

Values of A_W are given in table 3.

Table 1 — Values showing the uncertainty of estimates of the repeatability and reproducibility standard deviations

No. of laboratories	A_r			A_R								
				$\gamma = 1$			$\gamma = 2$			$\gamma = 5$		
p	$n = 2$	$n = 3$	$n = 4$	$n = 2$	$n = 3$	$n = 4$	$n = 2$	$n = 3$	$n = 4$	$n = 2$	$n = 3$	$n = 4$
5	0,62	0,44	0,36	0,46	0,37	0,32	0,61	0,58	0,57	0,68	0,67	0,67
10	0,44	0,31	0,25	0,32	0,26	0,22	0,41	0,39	0,38	0,45	0,45	0,45
15	0,36	0,25	0,21	0,26	0,21	0,18	0,33	0,31	0,30	0,36	0,36	0,36
20	0,31	0,22	0,18	0,22	0,18	0,16	0,28	0,27	0,26	0,31	0,31	0,31
25	0,28	0,20	0,16	0,20	0,16	0,14	0,25	0,24	0,23	0,28	0,28	0,27
30	0,25	0,18	0,15	0,18	0,15	0,13	0,23	0,22	0,21	0,25	0,25	0,25
35	0,23	0,17	0,14	0,17	0,14	0,12	0,21	0,20	0,19	0,23	0,23	0,23
40	0,22	0,16	0,13	0,16	0,13	0,11	0,20	0,19	0,18	0,22	0,22	0,22

Table 2 — Values of A, the uncertainty of an estimate of the bias of the measurement method

No. of laboratories	Value of A			
	$\gamma = 0$	$\gamma = 1$		
p	all n	$n = 2$	$n = 3$	$n = 4$
5	0,88	0,76	0,72	0,69
10	0,62	0,54	0,51	0,49
15	0,51	0,44	0,41	0,40
20	0,44	0,38	0,36	0,35
25	0,39	0,34	0,32	0,31
30	0,36	0,31	0,29	0,28
35	0,33	0,29	0,27	0,26
40	0,31	0,27	0,25	0,25

Table 3 — Values of A_W, the uncertainty of an estimate of the within-laboratory bias

No. of test results n	Value of A_W
5	0,88
10	0,62
15	0,51
20	0,44
25	0,39
30	0,36
35	0,33
40	0,31

6.3.4 Implications in the choice of laboratories

The choice of the number of laboratories will be a compromise between availability of resources and a desire to reduce the uncertainty of the estimates to a satisfactory level. From figures B.1 and B.2 in annex B it can be seen that estimates of the repeatability and reproducibility standard deviations could differ substantially from their true values if only a small number ($p \approx 5$) of laboratories take part in a precision experiment, and that increasing the number of the laboratories by 2 or 3 yields only small reductions in the uncertainties of the estimates when p is greater than 20. It is common to choose a value of p between 8 and 15. When σ_L is larger than σ_r (i.e. γ is larger than 2), as is often the case, little is to be gained by obtaining more than $n = 2$ test results per laboratory per level.

6.4 Selection of materials to be used for an accuracy experiment

6.4.1 The materials 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. As a general rule, five different materials will usually provide a sufficiently wide range of levels to allow the accuracy to be established adequately. 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 objects that are not altered by measuring, they could, in principle at least, be carried out using the same set of objects in different laboratories. This, however, would necessitate circulating the same set of objects around many laboratories often situated far apart, in different countries or continents, with a considerable risk of loss or damage during transport. If different 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 In selecting the material to represent the different levels, it should be considered whether the material should be specially homogenized before preparing the samples for dispatch, or whether the effect of the heterogeneity of the material should be included in the accuracy values.

6.4.4 When measurements have to be performed on solid materials that cannot be homogenized (such as metals, rubber or textile fabrics) and when the measurements cannot be repeated on the same test piece, inhomogeneity in the test material will form an essential component of the precision of the measurement and the idea of identical material no longer holds good. Precision experiments can still be carried out, but the values of precision may only be valid for the particular material 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 materials produced at different times or by different producers. This would require a more elaborate experiment than has been considered in ISO 5725.

6.4.5 In general, where destructive testing is involved, the contribution to the variability in the test results arising from differences between the specimens 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 materials 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 samples are to be measured.

6.4.7 In all the above, reference is made to measuring in different laboratories, with the implication of transportation of the test specimens to the laboratory, but some test specimens are not transportable, such as an oil storage tank. In such cases, 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 material being tested, or depends on the material 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 material. 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 material used in the precision experiment should be clearly specified along with the range of materials to which the values can be expected to apply.

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

7 Utilization of accuracy data

7.1 Publication of trueness and precision values

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 3.18, 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 5725-5. 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.

7.1.4 The repeatability and reproducibility standard deviations for any standard measurement method shall be determined as laid down in parts 2 to 4 of ISO 5725, 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, publication should be in the form of a table, such as table 4, and may also be expressed as a mathematical relationship. Intermediate measures of precision should be presented in a similar form.

Table 4 — Example of method of reporting standard deviations

Range or level	Repeatability standard deviation s_r	Reproducibility standard deviation s_R
From .... to .... From .... to .... From .... to		

7.1.5 The definitions of repeatability and reproducibility conditions (3.14 and 3.18) shall be given in the precision clause. When intermediate measures of precision are given, care should be taken to state which of the factors (time, operators, equipment) have been allowed to vary. When the repeatability and reproducibility limits are given, some statement should be added linking them to the difference between two test results and the 95 % probability level. Suggested wordings are as follows.

The difference between two test results found on identical test material by one operator using the same apparatus within the shortest feasible time interval will exceed the repeatability limit (r) on average not more than once in 20 cases in the normal and correct operation of the method.

Test results on identical test material reported by two laboratories will differ by more than the reproducibility limit (R) on average not more than once in 20 cases in the normal and correct operation of the method.

Ensure that the definition of a test result is clear, either by quoting the clause numbers of the measurement method standard that have to be followed to obtain the test result or by other means.

7.1.6 In general, a brief mention of the accuracy experiment should be added at the end of this precision section. Suggested wording is as follows.

The accuracy data were determined from an experiment organized and analysed in accordance with ISO 5725- (part) in (year) involving (p) laboratories and (q) levels. Data from ( ) laboratories contained outliers. The outliers were not included in the calculation of the repeatability standard deviation and the reproducibility standard deviation.

A description of the materials used in the accuracy experiment should be added, especially when the trueness or precision depend on the materials.

7.2 Practical applications of trueness and precision values

Practical applications of trueness and precision values are covered in detail in ISO 5725-6. Some examples are as follows.

7.2.1 Checking the acceptability of test results

A product specification could require repeated measurements to be obtained under repeatability conditions. A repeatability standard deviation may be used in these circumstances to check the acceptability of the test results and to decide what action should be taken if they are not acceptable. When both a supplier and a purchaser measure the same material and their results differ, repeatability and reproducibility standard deviations may be used to decide if the difference is of a size that is to be expected with the measurement method.

7.2.2 Stability of test results within a laboratory

By carrying out regular measurements on reference materials, a laboratory can check the stability of its results and produce evidence to demonstrate its competence, with respect to both the bias and the repeatability of its testing.

7.2.3 Assessing the performance of a laboratory

Laboratory accreditation schemes are becoming increasingly widespread. Knowledge of the trueness and precision of a measurement method allows the bias and repeatability of a candidate laboratory to be assessed, either using reference materials or an interlaboratory experiment.

7.2.4 Comparing alternative measurement methods

Two measurement methods may be available for measuring the same property, one being simpler and less expensive than the other but less generally applicable. Trueness and precision values may be used to justify the use of the less expensive method for some restricted range of materials.

12

Annex A

(normative)

Symbols and abbreviations used in ISO 5725

a — Intercept in the relationship

$$s = a + bm$$

A — Factor used to calculate the uncertainty of an estimate

b — Slope in the relationship

$$s = a + bm$$

B — Component in a test result representing the deviation of a laboratory from the general average (laboratory component of bias)

B_0 — Component of B representing all factors that do not change in intermediate precision conditions

$B_{(1)}$, $B_{(2)}$, etc. — Components of B representing factors that vary in intermediate precision conditions

c — Intercept in the relationship

$$\lg s = c + d \lg m$$

C, C', C'' — Test statistics

C_{crit}, C'_{crit}, C''_{crit} — Critical values for statistical tests

CD_P — Critical difference for probability P

CR_P — Critical range for probability P

d — Slope in the relationship

$$\lg s = c + d \lg m$$

e — Component in a test result representing the random error occurring in every test result

f — Critical range factor

$F_p(v_1, v_2)$ — p-quantile of the F-distribution with v_1 and v_2 degrees of freedom

G — Grubbs' test statistic

h — Mandel's between-laboratory consistency test statistic

k — Mandel's within-laboratory consistency test statistic

LCL — Lower control limit (either action limit or warning limit)

m — General mean of the test property; level

M — Number of factors considered in intermediate precision conditions

N — Number of iterations

n — Number of test results obtained in one laboratory at one level (i.e. per cell)

p — Number of laboratories participating in the interlaboratory experiment

P — Probability

q — Number of levels of the test property in the interlaboratory experiment

r — Repeatability limit

R — Reproducibility limit

RM — Reference material

s — Estimate of a standard deviation

$\hat{s}$ — Predicted standard deviation

T — Total or sum of some expression

t — Number of test objects or groups

UCL — Upper control limit (either action limit or warning limit)

W — Weighting factor used in calculating a weighted regression

w — Range of a set of test results

x — Datum used for Grubbs' test

y — Test result