INTERNATIONAL STANDARD

ISO 4259-1

First edition
2017-11

Petroleum and related products — Precision of measurement methods and results —

Part 1: Determination of precision data in relation to methods of test

Produits pétroliers — Fidélité des méthodes de mesure et des résultats —

Partie 1: Détermination des valeurs de fidélité relatives aux méthodes d'essai

Reference number
ISO 4259-1:2017(E)

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

Contents

Page

Foreword

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

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

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

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

For an explanation on 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 the following URL: www.iso.org/iso/foreword.html.

This document was prepared by Technical Committee ISO/TC 28, *Petroleum and related products, fuels and lubricants from natural or synthetic sources.*

This first edition of ISO 4259-1, together with ISO 4259-2, cancels and replaces ISO 4259, which has been technically revised.

A list of all parts in the ISO 4259 series can be found on the ISO website.

Introduction

For purposes of quality control and to check compliance with specifications, the properties of commercial petroleum products are assessed by standard laboratory test methods. Two or more measurements of the same property of a specific sample by a specific test method, or, by different test methods that purport to measure the same property, will not usually give exactly the same result. It is, therefore, necessary to take proper account of this fact, by arriving at statistically based estimates of the precision for a method, i.e. an objective measure of the degree of agreement expected between two or more results obtained in specified circumstances.

This document makes reference to ISO 3534-2[1], which gives a different definition of true value (see 3.23). This document also refers to ISO 5725-2. The latter is required in particular and unusual circumstances (see 5.3.1) for the purpose of estimating precision.

The two parts of ISO 4259 encompass both the derivation of precision estimates and the application of precision data. They combine the information in ASTM D6300[2] regarding the determination of the precision estimates and the information in ASTM D3244[3] for the utilization of test data.

A glossary of the variables used in this document and ISO 4259-2 is included as Annex I in this document.

Petroleum and related products — Precision of measurement methods and results —

Part 1: Determination of precision data in relation to methods of test

1 Scope

This document specifies the methodology for the design of an Interlaboratory Study (ILS) and calculation of precision estimates of a test method specified by the study. In particular, it defines the relevant statistical terms (Clause 3), the procedures to be adopted in the planning of ILS to determine the precision of a test method (Clause 4), and the method of calculating the precision from the results of such a study (Clauses 5 and 6).

The procedures in this document have been designed specifically for petroleum and petroleum related products, which are normally considered as homogeneous. However, the procedures described in this document can also be applied to other types of homogeneous products. Careful investigations are necessary before applying this document to products for which the assumption of homogeneity can be questioned.

2 Normative references

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

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

3 Terms and definitions

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

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

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

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

3.1
analysis of variance
ANOVA
technique that enables the total variance of a method to be broken down into its component factors

3.2
accepted reference value
ARV
agreed-upon reference value for a specific property of a material determined using an accepted reference method and protocol, e.g. derived from an ILS

3.3
between laboratory variance
component of the total variance attributable to the difference between the means of different laboratories

Note 1 to entry: When results obtained by more than one laboratory are compared, the scatter is usually wider than when the same number of tests is carried out by a single laboratory, and there is some variation between means obtained by different laboratories. These give rise to the between laboratory variance which is that component of the overall variance due to the difference in the means obtained by different laboratories.

Note 2 to entry: There is a corresponding definition for between operator variance.

Note 3 to entry: The term "between laboratory" is often shortened to "laboratory" when used to qualify representative parameters of the dispersion of the population of results, for example as "laboratory variance".

3.4
bias
<of a test method> difference between the population mean of test results from a very large number of different laboratories for the property of a material obtained using a specific test method versus the accepted reference value for the property where this is available

Note 1 to entry: See Note 1 to entry in 3.13 for an interpretation of "population mean of test results".

3.5
blind coding
assignment of a different number to each sample so that no other identification or information on any sample is given to the operator

3.6
check sample
sample taken at the place where a product is exchanged, i.e. where the responsibility for the product quality passes from the supplier to the recipient

3.7
degrees of freedom
divisor used in the calculation of variance

Note 1 to entry: The definition applies strictly only in the simplest cases. Definitions for more complex cases are beyond the scope of this document.

3.8
determination
process of carrying out the series of operations specified in a test method, whereby a single value is obtained

3.9
interlaboratory study
ILS
study specifically designed to estimate the repeatability and reproducibility of a standard test method achieved at a fixed point in time by multiple laboratories through the statistical analysis of their test results obtained on aliquots prepared from multiple materials

3.10
known value
quantitative value for a property that can be theoretically derived or calculated by the preparation of the sample

Note 1 to entry: The known value does not always exist, for example for empirical tests such as flash point.

3.11
mean
sum of a set of results divided by the number of results

3.12
mean square
sum of squares divided by the degrees of freedom

3.13
normal distribution
probability distribution of a continuous random variable, x, such that, if x is any real number, the probability density is as shown in Formula (1):

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right],-\infty<x<\infty \quad (1)$$

Note 1 to entry: In the context of modelling a distribution of test results, μ is the population mean, or true value (see 3.23) of the property as determined by a specific test method; σ is the standard deviation of the normal distribution used to describe the distribution of an infinite number of test results obtained using the same test method by an infinite number of laboratories ($\sigma > 0$).

3.14
operator
person who normally and regularly carries out a particular test

3.15
outlier
result far enough in magnitude from other results to be considered not a part of the set

3.16
precision
closeness of agreement between the results obtained by applying the same test procedure several times on essentially the same materials and under prescribed conditions

Note 1 to entry: The smaller the random part of the experimental error, the more precise the procedure.

3.17
random error
component of measurement error that in replicate measurements varies in an unpredictable manner

3.18
repeatability
limiting value for the difference between two independent results obtained in the normal and correct operation of the same method, for test material considered to be the same, within a short interval of time, under the same test conditions, that is expected to be exceeded with a probability of 5% due to random variation

Note 1 to entry: Same test conditions are to be considered as same operator, same apparatus, same calibration and same laboratory.

Note 2 to entry: The representative parameter for the dispersion of the population that can be associated with these results is repeatability standard deviation or repeatability variance. Repeatability refers to the maximum difference attributable to random variation between two results obtained under the state of minimum random variability. Therefore, the period of time during which repeat results are to be obtained should be short enough to exclude time dependent variation, for example, variation caused by environmental changes, or variation associated with multiple calibrations".

Note 3 to entry: The term "repeatability" is not to be confused with the terms "between repeats" or "repeats".

3.19
reproducibility
limiting value for the difference between two independent results obtained in the normal and correct operation of the same method, for test material considered to be the same, under different test conditions, that is expected to be exceeded with a probability of 5 % due to random variation

Note 1 to entry: Different test conditions are to be considered as different operator, different apparatus, different calibration, and different laboratory.

Note 2 to entry: The representative parameter of the dispersion of the population that can be associated with these results is reproducibility standard deviation or reproducibility variance. Reproducibility refers to the maximum difference attributable to random variation between two results obtained under the state of maximum random variability.

3.20
result
final value obtained by following the complete set of instructions in a test method

Note 1 to entry: It is assumed that the result is rounded off according to the procedure specified in Annex G.

3.21
standard deviation
measure of the dispersion of a series of results around their mean, equal to the positive square root of the variance and estimated by the positive square root of the mean square

3.22
sum of squares
sum of squares of the differences between a series of results and their mean

3.23
true value
for practical purposes, the value towards which the average of single results obtained by *n* laboratories tends, as *n* tends towards infinity

Note 1 to entry: Such a true value is associated with the particular method of test.

Note 2 to entry: A different and idealized definition is given in ISO 3534-2[1].

3.24
variance
mean of the squares of the deviation of a random variable from its mean, estimated by the mean square

4 Stages in the planning of an interlaboratory study for the determination of the precision of a test method

4.1 General

The stages in planning an interlaboratory study (ILS) are as follows:

a) preparing a draft method of test;

b) planning a pilot study with at least two laboratories;

c) planning the ILS;

d) executing the ILS.

The four stages are described in turn in 4.2 to 4.5.

4.2 Preparing a draft method of test

This shall contain all the necessary details for carrying out the test and reporting the results. Any condition that could alter the results shall be specified.

The ILS shall be designed so that it covers the intended range of the test method (see also 6.5). A clause on precision is included in the draft method of the test at this stage only as a heading.

4.3 Planning a pilot study with at least two laboratories

A pilot study is necessary for the following reasons:

a) to verify the details in the operation of the test;

b) to find out how well operators can follow the instructions of the method, and thus of the ILS;

c) to check the precautions regarding samples;

d) to estimate approximately the precision of the test.

At least two samples are required, covering the range of results to which the test method is intended to apply; however, at least 12 laboratory/sample combinations shall be included. Each sample is tested twice by each laboratory under repeatability conditions. The samples should be equally distributed across the test method range, and should include major product groups covered in the test method scope. If any omissions or inaccuracies in the draft test method are revealed, they shall now be corrected. The results shall be analysed for precision, and bias for sample(s) with accepted reference values. If either is considered to be too large, then alterations to the test method shall be considered.

4.4 Planning the ILS

There shall be at least six participating laboratories, but it is recommended this number be increased to eight or more in order to ensure the final precision is based on at least six laboratories and to ensure the precision statement is more representative of the user population.

The number of samples shall be sufficient to adequately represent the types of materials to which the test method is to be applied, to cover the range of the property measured at approximately equidistant intervals, and to give reliability to the precision estimates. If precision is found to vary with the level of results in the pilot study, then at least five samples shall be used in the ILS. In order to correctly estimate precision versus level relationship, it is important that the choice of samples evenly covers the range and materials for the property measured, so that an estimated relationship is not too dependent upon the leverage of a sample with extreme property value.

It is strongly recommended that the leverage of each planned sample in the sample set design, lev_i, be assessed using Formula (2). No sample shall have a leverage exceeding 0,5. See Table D.11 for an example of leverage calculation (second column from the right under heading 'lev_i').

$$lev_i = \frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum_{k=1}^{n}(x_k - \bar{x})^2} \tag{2}$$

where

lev_i is leverage of sample i;

n is total number of planned samples;

x_i is Napierian logarithm, ln (p_i), with p_i being the planned property level for sample i;

$\bar{x}$ is grand average of all x_i.

In any event, it is necessary to obtain at least 30 degrees of freedom for both repeatability and reproducibility (see Annex B for the corresponding rationale). For repeatability, this means obtaining a total of at least 30 pairs of results in the ILS.

For reproducibility, Annex A, Table A.1 gives the minimum number of samples required in terms of L, P and Q, where L is the number of participating laboratories, and P and Q are the ratios of variance component estimates obtained from the pilot study. Specifically, P is the ratio of the interaction component to the repeats component and Q is the ratio of the laboratories component to the repeats component. Annex B gives the derivation of the formula used. If Q is much larger than P, then 30 degrees of freedom cannot be achieved; the blank entries in Table A.1 correspond to this situation (i.e. when more than 20 samples are required). For these cases, there is likely to be a significant bias between laboratories.

In the absence of pilot test program information to permit the use of Table A.1, the number of samples shall be greater than five, and chosen such that the number of laboratories times the number of samples is greater than or equal to 42.

When it is known or suspected that different types of materials exhibit different precision functional forms when tested by the test method, consideration should be given to conducting separate ILS for each type of material.

4.5 Executing the ILS

One person shall be responsible for the entire ILS, from the distribution of the texts of the test method and samples to the final appraisal of the results. This person shall be familiar with the test method, but shall not personally take part in the tests.

The text of the test method shall be distributed to all the laboratories in time to allow any queries to be raised before the tests begin. If any laboratory wants to practice the method in advance, than this shall be carried out with samples other than those used in the ILS.

The samples shall be accumulated, subdivided and distributed by the coordinator, who shall also keep a reserve of each sample for emergencies. It is most important that the individual laboratory portions be homogeneous and stable for the property of interest throughout the entire duration of the ILS. Prior to distribution, the ILS sample set shall be blind coded in a manner that preserves the anonymity of the nature of the test material and the expected value of the property. The following information shall be sent with the ILS sample set.

a) Agreed (draft) method of test.

b) Handling and storage requirements for the samples.

c) Order in which the samples are to be tested. A different random order for each laboratory is highly recommended. For large number of laboratories, several unique test orders may be randomly assigned to groups of laboratories, with no more than 4 laboratories per group.

d) For statistical reasons, it is imperative that the repeat results are obtained independently of each other, i.e. that the second result is not biased by knowledge of the first. This is achieved by blind coding where the repeat for each material in the ILS design is included in the test set sent to ILS participants without disclosing that it is a repeat, with an accompanying statement that a single result is to be obtained on each sample in the test set, in the specified testing order, by the same operator with the same apparatus within a short time. If this blind coding is regarded as infeasible to achieve, then the statement shall state that a pair of results associated with a sample shall be obtained by the same operator with the same apparatus within a short time, without disclosing the nature of the sample.

e) Period of time within which all the samples are to be tested.

f) Blank form for reporting the results. For each sample, there shall be space for the date of testing, the test results, and any unusual occurrences. The unit of accuracy for reporting the results shall be specified.

g) Statement that the test shall be carried out under normal conditions, using qualified operators who carry out this kind of test routinely and that the duration of the test shall be the same as normal.

h) A questionnaire requesting information on the conditions used in the application of the test method, e.g. apparatus details, reagents and materials, calibration and verification procedures, quality control procedure, any deviations from either the test method or the instructions supplied, observations and suggestions for future improvement of the test method.

Operators that participated in the pilot study may also participate in the ILS. If their extra experience in testing a few more samples produces a noticeable effect, it serves as a warning that the test method is not satisfactory. They shall be identified in the report of the results so that any effect can be noted.

NOTE For additional guidance on the planning and execution of an ILS, consult ASTM D7778[4] and ASTM D6300[2].

5 Statistical treatment of ILS results

5.1 General recommendation

Although the procedures described in Clauses 5 and 6 of this document are in a form suitable for hand calculation, it is strongly advised that these procedures be carried out using an electronic computer with appropriately validated software designed specifically to store and analyse ILS test results based on the procedures of this document. It is also highly recommended that these procedures be carried out under the guidance of a statistician.

NOTE A software package extensively used in the ISO and ASTM community is D2PP[13]. That software package does not include GESD or Cook's Distance assessment in line with this document.

In the clauses to follow, procedures are specified to achieve the following:

a) pre-screen the results as reported from the ILS on a sample-by-sample basis for grossly discordant results (outliers);

b) assess independence or dependence of precision and the level of results after pre-screening;

c) assess uniformity of precision from laboratory to laboratory by detecting the presence (or absence) of additional outliers using the detection power from the entire data set.

The procedures are described in mathematical terms based on the notation of Annex C.

Illustration of the procedures is provided in referenced Annexes.

For all the procedures, it is assumed that the results are either from a single normal distribution or capable of being transformed into such a distribution (see 5.3). Other cases (which are rare) require a different treatment that is beyond the scope of this document. See Reference [6] for a statistical test on normality.

5.2 Pre-screen using GESD technique

Prior to execution of 5.3 to 5.7, examine all information returned by ILS participants to determine compliance with agreed-upon test protocol and method of test. If the investigation disclosed no clerical, sampling or procedural errors, apply the Generalized Extreme Studentized Deviation (GESD) technique as outlined in this clause to results received for each ILS sample to identify unusual or extreme results. Investigation for causes associated with unusual results shall be conducted. If acceptable cause(s) is found during the investigation, the unusual results shall be either corrected, replaced, or rejected. Correction or replacement of the unusual results with a new set of results shall be approved by the ILS

coordinator in consultation with the ILS statistician. If no acceptable cause is found, the unusual or extreme results as identified by the GESD technique at the 99 % confidence level shall be rejected.

An overall summary of this GESD pre-screening technique is outlined below.

For each ILS sample, execute the following steps.

1) Calculate the sample mean using all results received for the sample.

2) Calculate difference for each pair of results as received from laboratories that have reported both results.

3) Identify outlier(s) in the data set of differences obtained from step 2) by following the methodology outlined in Annex H.

4) For each outlying difference identified, remove the member from the pair that is farthest from the sample mean calculated in 1) and replace it with the value of the remaining result.

5) For laboratories that have only reported one result, i.e. the other result is missing, assign the value of the single reported result to the missing result before proceeding to step 6).

6) Calculate the sum of the pair of the results for each lab. For laboratories that have reported both results and neither result has been rejected, this will be the sum of both reported results. In the case where one of the pair of results is missing (not reported) or rejected from step 4), this sum will be twice the single reported result since the missing result is assigned the same value as the reported result.

7) Identify outlier(s) in data set of sums as obtained from step 6) by following the methodology outlined in Annex H.

8) For each outlying sum of results, exclude both results from further statistical analysis.

9) For the pairs of results with sums that have not been rejected, retain both reported results for analysis if both results are as originally received from the laboratories. If one of the two results of the pair is an assigned value from step 4) or step 5), retain the reported result from the laboratories for analysis, and treat the other result as “missing”.

10) The data set remaining after completion of step 9) then constitutes the data set to be further analysed as per 5.3 to 5.7.

5.3 Transformation of data and outlier tests

5.3.1 General

In many test methods, the precision depends on the level of the test result, and thus the variability of the reported results is different from sample to sample. The method of analysis outlined in this document requires that this shall not be so and the position is rectified, if necessary, by a transformation.

The laboratories standard deviations, D_j, and the repeats standard deviations, d_j, for sample j (see Annex C for notation explanation) are calculated and plotted separately against the sample means, m_j, in accordance with Annexes D and E). If the points so plotted can be considered as lying about a pair of lines parallel to the m-axis, then no transformation is necessary. If, however, the plotted points describe non-horizontal straight lines or curves of the form $D = f_1(m)$ and $d = f_2(m)$, then a transformation is necessary.

The relationships $D = f_1(m)$ and $d = f_2(m)$ are not, in general, identical. The statistical procedures of this document require, however, that the same transformation be applicable both for repeatability and for reproducibility. For this reason, the two relationships are combined into a single dependency relationship $D = f(m)$ (where D now includes d) by including a dummy variable, T. This takes account of the difference between the relationships, if one exists, and provides a means of testing for this difference (see F.1).

The single relationship $D = f(m)$ is best estimated by a weighted linear regression analysis, even though in most cases an unweighted regression gives a satisfactory approximation. The derivation of weights is described in F.2, and the computational procedure for the regression analysis is described in F.3. Typical forms of dependence $D = f(m)$ are given in E.1. These are all expressed in terms of transformation parameters B and B_0.

The estimation of B and B_0, and the transformation procedure which follows, are summarized in E.2. This includes statistical tests for the significance of the regression (i.e. is the relationship $D = f(m)$ parallel to the m-axis), and for the difference between the repeatability and reproducibility relationships, based at the 5 % significance level. If such a difference is found to exist, or if no suitable common transformation exists, then the alternative sample by sample procedures of ISO 5725-2 shall be used. In such an event, it is not possible to test for laboratory bias over all samples (see 5.6) or separately estimate the interaction component of variance (see 6.2).

If it has been shown at the 5 % significance level that there is a significant regression of the form $D = f(m)$, then the appropriate transformation $y = F(x)$, where x is the reported result, is given by Formula (3):

$$F(x) = K \int \frac{dx}{f(x)} \tag{3}$$

where K is a constant.

In that event, all results shall be transformed accordingly and the remainder of the analysis carried out in terms of the transformed results. Typical transformations are given in E.1.

It is difficult to make the choice of transformation the subject of formalized rules. Qualified statistical assistance can be required in particular cases. The presence of outliers can affect judgement as to the type of transformation required, if any (see 5.7). That is why extremely discordant results shall be removed as described in 5.1 above prior to making a judgement on transformation(s).

The transformation and outlier procedure is described in the form of a flow chart in Figure 1. Note that the transformation process is an iterative procedure, requiring confirmation of the choice of transformation if outliers have been rejected. If the original transformation is found to be inadequate after outliers have been removed, then a different transformation will be required.